CONTENTS

2 Evaluation of the standard design flood method in selected basins in South Africa
O J Gericke, J A du Plessis

15 Catchment parameter analysis in flood hydrology using GIS applications
O J Gericke, J A du Plessis

27 What are the legal remedies available to contractors and consultants to enforce payment?
M J Maritz, D C Robertson

36 A probabilistic approach for modelling deterioration of asphalt surfaces
T F P Henning, D C Roux

45 The wave climate on the KwaZulu-Natal coast of South Africa
S Corbella, D D Stretch

55 Coastal defences on the KwaZulu-Natal coast of South Africa: a review with particular reference to geotextiles
S Corbella, D D Stretch

65 Adjudication as an alternative dispute resolution method in the South African construction industry
N C Maiketso, M J Mantz

71 Influence of mica on unconfined compressive strength of a cement-treated weathered granite gravel
M R Mshali, A T Visser

78 A step towards standardising accelerated corrosion tests on laboratory reinforced concrete specimens
G Malumbela, P Moyo, M Alexander

86 The effects of placement conditions on the quality of concrete in large-diameter bored piles
G C Fanourakis, P W Day, G R H Grieve

94 The discontinuity required at an air valve or vent for effective pipeline de-aeration
S J van Vuuren, M van Dijk

101 Definition and application of a cohesive crack model allowing improved prediction of the flexural capacity of high-performance fibre-reinforced concrete pavement materials
E Denneman, E P Kearsley, A T Visser

112 Prediction of the debonding/slip load of composite deck slabs using fracture mechanics
J Mahachi, M Dundu

117 Estimating soil plasticity properties from pedological data
G C Fanourakis

126 The importance of plane end-bearing surfaces when measuring the strengths of concrete core specimens
T Bugai, D Kruger, R G D Rankine
Evaluation of the standard design flood method in selected basins in South Africa

O J Gercke, J A du Plessis

Design flood estimates display relatively wide confidence bands of uncertainty around all estimates of flood magnitude-frequency relationships. Taking cognisance of this, and the fact that most of the available design flood estimation methods in South Africa were developed in the 1970s and have not been updated since, led to the development of the Standard Design Flood (SDF) method (Alexander 2002a; 2002b; 2003). In this study, the SDF method was evaluated by establishing the accuracy of the regionalised SDF runoff coefficients, taking both the areal extent and homogeneous hydrological catchment responses into consideration. The SDF runoff coefficients were evaluated, calibrated and verified at a quaternary catchment level in SDF basin 9 (primary catchment area) and in 19 of the other 29 SDF basins in South Africa (secondary study areas) by establishing catchment parameters and evaluating the ratios between the results obtained through the SDF method and probabilistic analysis. The results showed that the original SDF method overestimated the magnitude and frequency (return period) of flood peaks in all the basins under consideration, while the verification results confirmed that the calibrated/verified SDF method, based on quaternary runoff coefficients, significantly improves the accuracy in comparison with the probabilistic analysis results. The result confirmed that the probabilistic-based approach of the original SDF method does not have the ability to overcome the deficiencies evident in the other design flood estimation techniques used in South Africa. Revision of the runoff coefficients at a quaternary catchment level is proposed.

INTRODUCTION

The last decade in southern Africa was characterised by several damaging flood events, especially the February 2000 floods in the northeastern part of South Africa, Zimbabwe and Mozambique (Alexander 2002a). On considering the economic and environmental impact of these and other flood events, the importance of flood frequency analysis becomes evident (Smithers 2011). However, reliable estimates of flood frequency in terms of peak flows and volumes remain a constant challenge in hydrology (Cameron et al 1999). Cordery & Pilgrim (2000) also highlighted that the international demands for improved design flood estimations have not been met with any increased understanding of the fundamental hydrological processes. According to Van der Spuy & Rademeyer (2010), this is also the case in South Africa where the search continues for a universally applicable method for design flood estimation.

In essence, the failures of civil engineering structures (e.g. bridges, culverts, dam spillways and drainage canals) caused by floods are largely due to the immense variability in the flood response of catchments to storm rainfall, which is intrinsically variable in its own right. Consequently, flood estimations for design purposes can be expected to display relatively wide confidence bands of uncertainty around all estimates of flood magnitude-frequency relationships (Alexander 2002a; 2002b; 2003). Thus, both the occurrence and the frequency of flood events, along with the uncertainty involved in the estimation thereof, as well as the lack of updated design flood estimation methods in South Africa since the 1970s, indicate that there is an urgent need to revise existing methods or develop alternative design flood estimation methods by using about 40 years of additional observed data.

The developmental effort in this regard by Alexander (2002a) led to the development of a numerically calibrated version of the Rational Method (RM), known as the Standard Design Flood (SDF) method, which incorporates engineering factors of safety to accommodate the uncertainties in hydrological analyses at a regional level (Alexander 2002a; 2002b; 2003). In this study, the SDF method was evaluated in specific areas by establishing the accuracy of the regionalised SDF runoff coefficients, taking both the areal extent and homogeneous...
hydrological catchment responses into consideration. The question of whether or not the probabilistic-based approach of the SDF method has the ability to overcome some of the deficiencies evident in the other techniques used for design flood estimation, was investigated and alternative revisions at a quaternary catchment scale were proposed.

The development of the original SDF method is reviewed in the next section. The purpose of the study is discussed and explained in the section thereafter, followed by an overview of the spatial distribution and characteristics of the study areas. The methodologies involved in assessing the paper’s purpose and objectives are then expanded on in detail, followed by the results and discussion, conclusions and recommendations.

DEVELOPMENT OF THE ORIGINAL STANDARD DESIGN FLOOD METHOD

In the Introduction the wide confidence bands of uncertainty around all estimates of flood magnitude-frequency relationships were emphasised. However, Alexander (2002a; 2002b; 2003) indicated that these uncertainties cannot be satisfactorily accommodated, and necessitate a new, single approach to the estimation of the design flood, namely the SDF method.

The identification of representative, homogeneous flood-producing regions, which followed the boundaries of the drainage regions as depicted by the Department of Water Affairs and Forestry (DWAF) (1995), was a major step in the development of the SDF method. These regions are referred to as SDF basins and a total of 29 basins in South Africa were identified (Alexander 2002a; 2002b; 2003). Thereafter, Alexander (2002a; 2002b; 2003) reviewed all the different design flood estimation methods (deterministic, empirical and probabilistic) in use in South Africa and selected the conventional RM as the basis for the SDF method to be used in the delineated basins. At least one hydrological flow-gauging station and one representative daily rainfall station were selected for each of the 29 basins in the development of the SDF method. Probabilistic analyses of the annual maximum series (AMS) were conducted at 152 flow-gauging stations to calibrate and verify the SDF runoff coefficients at a quaternary catchment level in SDF basin 9 (primary study area) and in 19 of the other 29 SDF basins in South Africa (secondary study areas) by establishing the catchment parameters and the ratios between results obtained with the SDF method and those obtained through the probabilistic analysis of the observed flow data (SDF(probability distribution ratios). These newly calibrated quaternary runoff coefficients were then compared with the existing regional SDF runoff coefficients. The work done by Van Bladeren (2005) during the compilation of the South African National Roads Agency Limited (SANRAL) Drainage Manual resulted in a proposal that adjustment factors be used to balance the tendency of the SDF method to provide over-conservative results. The runoff coefficient adjustment factors as proposed by Van Bladeren (2005) were therefore also evaluated as part of this study.

The secondary study areas were purposely evaluated to enhance the understanding of the results obtained from the primary study area, as well as to illustrate the relevance thereof in a South African context. This served as clarification of the influence of different climatic regions (Highveld as opposed to Mediterranean/southern coastal regions), types of weather systems (summer convective as opposed to winter/all year orographic/frontal rainfall) and rainfall occurrence frequencies on the depth, area, duration and movement of storm rainfall, which in turn influence the magnitude and frequency of floods as estimated by the SDF method.

Firstly, it was hypothesised that the calibration of the SDF method at a quaternary catchment level will improve the accuracy and practical use thereof. Secondly, it was hypothesised that the extent of the current delineated SDF basins is too large, with associated non-homogeneous flood-producing characteristics. Thirdly, it was hypothesised that the runoff coefficients are essentially functions of the return period and time of concentration. The fourth hypothesis was that the Log-Normal (LN), Log-Pearson Type III (LP3) and the General Extreme Value (GEV) probability distributions, or a combination thereof, are the most suitable for flood frequency analyses at a single site in South Africa.

PURPOSE OF STUDY

The purpose of this study was to evaluate, calibrate and verify the SDF runoff coefficients at a quaternary catchment level in SDF basin 9 (primary study area) and in 19 of the other 29 SDF basins in South Africa (secondary study areas) by establishing the catchment parameters and the ratios between results obtained with the SDF method and those obtained through the probabilistic analysis of the observed flow data (SDF(probability distribution ratios). These newly calibrated quaternary runoff coefficients were then compared with the existing regional SDF runoff coefficients. The work done by Van Bladeren (2005) during the compilation of the South African National Roads Agency Limited (SANRAL) Drainage Manual resulted in a proposal that adjustment factors be used to balance the tendency of the SDF method to provide over-conservative results. The runoff coefficient adjustment factors as proposed by Van Bladeren (2005) were therefore also evaluated as part of this study.

Figure 1 Location of primary and secondary study areas within the SDF basins
Modder River catchments. The primary study area is characterised by 99,1% rural areas, 0,7% urbanisation and 0,2% water bodies (CSIR 2001). The natural vegetation is dominated by Grassland of the Interior Plateau, False Karoo and Karoo (light bush). Cultivated land is the largest human-induced vegetation alteration in the rural areas, while residential and suburban areas dominate the urban areas. The topography is gentle (slopes between 2,4% and 5,5%) and water tends to pond easily, thus influencing the attenuation and translation of floods. The Mean Annual Precipitation (MAP) is 424 mm, ranging from 275 mm in the west to 685 mm in the east. It is characterised as highly variable and unpredictable. The rainy season starts early September and ends mid-April with a dry winter (Midgley et al 1994). The Modder and Riet Rivers are the main river reaches and discharge into the Orange-Vaal River drainage system (Seaman et al 2001). Like most inland rivers in South Africa, these rivers were traditionally seasonal rivers, but due to the construction of significant storage dams, they now resemble permanent rivers.

The secondary study area catchments ranged in size from 126 km² to 33 277 km² and are located within basins 1, 2, 4 - 11, 16, 18, 21 - 23, 26, 28 and 29. The locations of these catchments are shown in Figure 1.

**METHODOLOGY**

This section provides the detailed methodology followed during this study, which focuses on the evaluation and calibration of SDF runoff coefficients and flood peaks, the verification of calibrated runoff coefficients and flood peaks, and the statistical assessment of results.

**Evaluation and calibration of runoff coefficients and flood peaks**

To evaluate and verify the original (Alexander 2003), adjusted (Van Bladeren 2005) and calibrated (this study) SDF methods at a quaternary catchment level in SDF basin 9 (primary study area) and the other randomly selected SDF basins in South Africa (secondary study area), the following procedures were followed: establishment of physical catchment parameters, probabilistic analyses, selection of representative rainfall stations, and numerical calibration of runoff coefficients.

These procedures are discussed in the following sub-sections:

**Establishment of physical catchment parameters**

All the required catchment input parameters of SDF basin 9 (average main course water length and slope, time of concentration, design rainfall intensities, average number of days per year during which thunder was heard (R) and areal reduction factors (ARFs)) were determined in a similar fashion as for the original SDF method (Alexander 2002a; 2002b; 2003). In all the other SDF basins under consideration, the catchment areas and time of concentration were based on previous research conducted by Petras & Du Plessis (1987) and Parak & Pegram (2006).

**Probabilistic analyses**

Probabilistic analysis of the AMS was conducted at a representative flow-gauging station in each catchment under consideration to summarise the observed flood peak data, estimate parameters and select appropriate theoretical probability distributions. The observed flood peak data was summarised by ranking the AMS in a descending order of magnitude. The Cunnane plotting position, based on a general plotting formula, was used to assign a probability to the flood peaks. The parameter estimation was largely based on the Method of Moments (MM), although the usefulness of Linear-Moments (LM) estimation to fit the General Logistic (GLO) probability distribution was also investigated.

Several of the AMS data sets were characterised by insufficient record lengths (e.g. missing data, low outliers and flood peaks exceeding the hydraulic capacity of flow-gauging structures), which made it impossible to conclusively select a single probability distribution that could consistently provide flood frequency estimates for return periods much greater than the period of record. To overcome this limitation of single site analyses, a Mean Logarithm Value Approach (MLVA), based on the mean values of the logarithms of two or more probability distributions, were used as the most suitable combined probability distribution at a single site or flow-gauging station. The MLVA, as expressed in Equation 1, is also used as a standard method in the DWA (Directorate: Flood Studies) (Van der Spuy & Rademeyer 2010).

\[
Q_p = 10 \exp \left[ {\log\left( {\frac{Q_i}{Q_i^{1 + 1} - Q_i^{1 + 1}}} \right)} \right],
\]

where:

- \(Q_p\) = peak flow based on the MLVA (m³/s)
- \(Q_i\) = peak flows based on a recognised theoretical probability distribution, with a minimum of two probability distributions used in combination (m³/s), and
- \(N\) = number of probability distributions used.

The individual peak flows (\(Q_i\)) can either be based on the combination of two or more theoretical probability distributions, e.g. Extreme/Log-Extreme Value Type I (EV1/LEV1), LN, LP3, GEV and/or GLO distributions. Statistical properties, visual inspection of the plotted values and Goodness-of-Fit (GOF) statistics were used to select the most suitable single probability or combined probability distribution in Equation 1. Both the EV1 and LEV1 probability distributions have a fixed skewness of 1,14; hence the limited use thereof in flood hydrology. The LN distribution was only used where the logarithms of the observed data have near symmetrical distribution or where the skewness coefficients are close to zero. In all other asymmetrical data sets, the LP3 distribution was used instead. The GEV distributions were used at asymmetrical data sets characterised by either positive (Extreme Value Type II) or negative (Extreme Value Type III) skewness coefficients.

A total of 44 catchments were evaluated of which 12 fall within the primary study area. The probabilistic analysis results of eight of the catchments were based on the GEV distribution, as obtained from previous research conducted by Parak & Pegram (2006). These results were therefore not analysed again. However, the nature and record length of the observed flood peak data sets initially used by these authors are unknown.

**Selection of representative rainfall stations**

The design rainfall information used in this study was based on the Regional L-Moment Algorithm South African Weather Service n-day design point rainfall database (RLMA-SAWS) (after Smithers & Schulze 2000b), as opposed to the original SDF method which used the TR102 information. The RLMA-SAWS database contains design rainfall information of 3 946 daily rainfall stations up to the year 2002. The SAWS contributed the majority (82,2%) of the data of these daily rainfall stations, while the Institute for Soil, Climate and Water (ISCW), the South African Sugar Association Experiment Station (SASEX) and private individuals provided the remaining daily rainfall data (Smithers & Schulze 2000b). In contrast, the TR102 daily design rainfall database has ± 20 years less information available and is limited to 1 946 rainfall stations (Adamson 1981). The selection of the single rainfall station, to be used in the calibrated SDF method at a quaternary catchment level, was based on the following criteria:

- **Average meteorological conditions:** The MAP and design rainfall depths associated with return periods ranging from 2 to 200 years of the selected station must have a
Table 1: Probabilistic analysis results in SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Station</th>
<th>Return period</th>
<th>Design flood based on probability distributions (m³/s)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Return period:</td>
<td>1.25 – 200 years LP3/MM distribution</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.25 – 20 years GEV/MM distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 – 200 years LP3/MM distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5R001</td>
<td>2</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>145</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>231</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>327</td>
<td>266</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>477</td>
<td>401</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>611</td>
<td>641</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>766</td>
<td>884</td>
</tr>
<tr>
<td></td>
<td>Qₚ</td>
<td>(Eq 1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C5R002</td>
<td>2</td>
<td>243</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>765</td>
<td>604</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1 201</td>
<td>1 089</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1 704</td>
<td>1 770</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2 506</td>
<td>3 029</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>3 242</td>
<td>4 405</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>4 115</td>
<td>6 130</td>
</tr>
<tr>
<td>C5R003</td>
<td>2</td>
<td>126</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>323</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>465</td>
<td>401</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>609</td>
<td>635</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>810</td>
<td>1 064</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>972</td>
<td>1 502</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 143</td>
<td>2 058</td>
</tr>
<tr>
<td></td>
<td>Qₚ</td>
<td>(Eq 1)</td>
<td></td>
</tr>
<tr>
<td>C5R004</td>
<td>2</td>
<td>302</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>637</td>
<td>643</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>893</td>
<td>1 114</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1 168</td>
<td>1 755</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1 571</td>
<td>2 925</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1 913</td>
<td>4 112</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>2 293</td>
<td>5 617</td>
</tr>
<tr>
<td></td>
<td>Qₚ</td>
<td>(Eq 1)</td>
<td></td>
</tr>
<tr>
<td>C5R005</td>
<td>2</td>
<td>46</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>103</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>146</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>189</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>251</td>
<td>306</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>301</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>355</td>
<td>528</td>
</tr>
<tr>
<td></td>
<td>Qₚ</td>
<td>(Eq 1)</td>
<td></td>
</tr>
<tr>
<td>C5R006</td>
<td>2</td>
<td>182</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>481</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>689</td>
<td>546</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>897</td>
<td>884</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1 179</td>
<td>1 521</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1 400</td>
<td>2 182</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 629</td>
<td>3 038</td>
</tr>
<tr>
<td>C5H003</td>
<td>2</td>
<td>239</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>698</td>
<td>730</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>925</td>
<td>1 291</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1 147</td>
<td>2 068</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1 438</td>
<td>3 514</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1 659</td>
<td>5 003</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 882</td>
<td>6 914</td>
</tr>
<tr>
<td>C5H004</td>
<td>2</td>
<td>240</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>730</td>
<td>736</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>995</td>
<td>1 291</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1 169</td>
<td>2 068</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1 438</td>
<td>3 514</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1 659</td>
<td>5 003</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 882</td>
<td>6 914</td>
</tr>
</tbody>
</table>
high degree of association with the MAP and design rainfall depths as obtained by the Thiessen polygon method using the total number of rainfall stations within the catchment(s) under consideration.

- Record length: Where possible, a minimum record length of 50 years must be used in order to enable the selection of a distribution that could consistently provide adequate rainfall frequency estimates for return periods much greater than the period of record. However, shorter records (between 40 and 47 years) were used in 11% of the secondary study areas to evaluate the SDF method due to the limited number of suitable rainfall stations available in five of these catchments.

### Numerical calibration of runoff coefficients

The numerical calibration of the runoff coefficients used in the SDF method followed the probabilistic analyses and the selection of a single rainfall station in all the catchments under consideration in both the primary and secondary study areas. The purpose of the calibration was to fit the results obtained by the probabilistic analysis with those of the SDF method in order to establish the SDF/probability distribution ratios and adapted quaternary SDF runoff coefficients for return periods, ranging from 2 to 200 years. This was accomplished by determining the $C_2$ (2-year return period) and $C_{100}$ (100-year return period) runoff coefficients in Equation 2 (Alexander 2003) in such a way that the calibrated runoff coefficients ($C_T$) for a range of return periods resulted in the best fit between the design values of flood peaks based on the probabilistic analysis and the SDF method (Equation 3; Alexander 2003).

\[
C_T = C_2 \left( \frac{Y_T}{2.33} \right) + \left( \frac{C_{100} - C_2}{2} \right) \frac{100}{100}
\]

\[
Q_T = 0.278C_TA
\]

where:
- $C_T$ = calibrated runoff coefficient
- $Q_T$ = design flood peak (m$^3$/s)
- $A$ = catchment area (km$^2$)
- $C_2$ = 2-year return period runoff coefficient
- $C_{100}$ = 100-year return period runoff coefficient
- $I_T$ = average design rainfall intensity (mm/h), and
- $Y_T$ = return period factor.

These coefficients ($C_2$ and $C_{100}$) were changed manually until an appropriate fit with the probabilistic analysis results was achieved, after which values of $C_T$ were

### Table 2 MAP of selected catchments in SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>MAP (mm)</th>
<th>Number of rainfall stations ($N_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study area</td>
<td>439</td>
<td>424</td>
</tr>
<tr>
<td>C5R001</td>
<td>492</td>
<td>488</td>
</tr>
<tr>
<td>C5R002</td>
<td>421</td>
<td>420</td>
</tr>
<tr>
<td>C5R003</td>
<td>553</td>
<td>549</td>
</tr>
<tr>
<td>C5R004</td>
<td>530</td>
<td>518</td>
</tr>
<tr>
<td>C5R005</td>
<td>642</td>
<td>660</td>
</tr>
<tr>
<td>C5H003</td>
<td>553</td>
<td>549</td>
</tr>
<tr>
<td>C5H012</td>
<td>448</td>
<td>444</td>
</tr>
<tr>
<td>C5H015</td>
<td>530</td>
<td>518</td>
</tr>
<tr>
<td>C5H016</td>
<td>440</td>
<td>429</td>
</tr>
<tr>
<td>C5H018</td>
<td>479</td>
<td>461</td>
</tr>
<tr>
<td>C5H022</td>
<td>686</td>
<td>660</td>
</tr>
<tr>
<td>C5H054</td>
<td>542</td>
<td>523</td>
</tr>
</tbody>
</table>

### Table 3 Selected single SDF rainfall stations used in SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>Station number</th>
<th>MAP (mm)</th>
<th>Record length ($N_i$ years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5R001</td>
<td>0261750W</td>
<td>497</td>
<td>488</td>
</tr>
<tr>
<td>C5R002</td>
<td>0230168W</td>
<td>420</td>
<td>420</td>
</tr>
<tr>
<td>C5R003</td>
<td>0232123W</td>
<td>555</td>
<td>549</td>
</tr>
<tr>
<td>C5R004</td>
<td>0261423W</td>
<td>518</td>
<td>518</td>
</tr>
<tr>
<td>C5R005</td>
<td>026123W</td>
<td>649</td>
<td>660</td>
</tr>
<tr>
<td>C5H003</td>
<td>0232123W</td>
<td>555</td>
<td>549</td>
</tr>
<tr>
<td>C5H008</td>
<td>0230168W</td>
<td>420</td>
<td>420</td>
</tr>
<tr>
<td>C5H012</td>
<td>0231395W</td>
<td>454</td>
<td>444</td>
</tr>
<tr>
<td>C5H015</td>
<td>0261423W</td>
<td>518</td>
<td>518</td>
</tr>
<tr>
<td>C5H016</td>
<td>0291899W</td>
<td>433</td>
<td>429</td>
</tr>
<tr>
<td>C5H018</td>
<td>026013W</td>
<td>461</td>
<td>461</td>
</tr>
<tr>
<td>C5H022</td>
<td>026734W</td>
<td>649</td>
<td>660</td>
</tr>
<tr>
<td>C5H054</td>
<td>026123W</td>
<td>518</td>
<td>523</td>
</tr>
</tbody>
</table>

### Table 4 Design information applicable to SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>Method</th>
<th>SDF design information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_2$</td>
<td>$C_{100}$</td>
</tr>
<tr>
<td>C5R002</td>
<td>SDF$_{Original}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Calibrated}$</td>
<td>9</td>
</tr>
<tr>
<td>C5R003</td>
<td>SDF$_{Original}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Calibrated}$</td>
<td>9.5</td>
</tr>
<tr>
<td>C5R004</td>
<td>SDF$_{Original}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Calibrated}$</td>
<td>18</td>
</tr>
<tr>
<td>C5R005</td>
<td>SDF$_{Original}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Calibrated}$</td>
<td>11</td>
</tr>
</tbody>
</table>
coaxially plotted with the regional SDF runoff coefficients from Alexander (2003) against the return period. This exercise was repeated with the runoff coefficient adjustment factors as proposed by Van Bladeren (2005) to validate these runoff coefficient adjustment factors.

**Verification of calibrated runoff coefficients and flood peaks**

Verification tests were conducted in 16 catchments that were not used in the calibration exercise to establish whether the calibrated runoff coefficients were predictable and to confirm that the method was reliable. In order to verify the \( C_T \) values achieved during calibration, it was necessary to find some physical or regional descriptors with which to relate the runoff coefficients and to enable the use and extension of the calibrated runoff coefficients to ungauged catchments. Several regional descriptors, such as MAP, average catchment slope and land-use distribution, were taken into consideration in combination with the \( C_T \) values to establish whether or not any relationship existed on which to regress the runoff coefficients.

The 16 catchments used in the verification exercise, and for which AMS were available, were selected based on the fact that their physical or regional descriptors were similar to the catchments used during the calibration exercise, in order to result in near homogeneous hydrological responses. These catchments ranged in size from 26 km² to 23 067 km², with eight catchments within SDF basin 9 (primary study area) and eight catchments within SDF basins 6, 7, 10, 18 and 21 (secondary study areas).

**Statistical assessment of results**

Regression (coefficient of determination) and descriptive (Chi-square) statistics were used to evaluate the GOF of the probabilistic analyses (fitted probability distributions). The coefficient of determination \( r^2 \) calculations were based on the full record length where the ranked observed values, with their associated probability or return period, were compared with the theoretical probability distributions. The Chi-square statistics were evaluated by making use of the concept of contingency tables, consisting of margin totals, which were used to establish the expected estimated values. The comparisons between the probabilistic analyses and the calibrated and verified versions of the SDF method were evaluated in the same manner.

**RESULTS AND DISCUSSION**

The results based on the methodology used in this study are discussed next. The evaluation and calibration of SDF runoff coefficients and flood peaks are discussed first, followed by a discussion on the verification of calibrated runoff coefficients and flood peaks. In conclusion, the statistical assessment of results are highlighted.

**Evaluation and calibration of runoff coefficients and flood peaks**

**Probabilistic analyses**

The MLVA not only overcame the limitation of AMS data sets at a single site characterised by insufficient record lengths, but also took cognisance of the strong evidence that in South Africa most of the high flood peaks are a result of rare and severe meteorological phenomena. Alexander (2012) also confirmed that the AMS of these floods could consist of a mixture of two or more statistical populations with different parameter values and associated flood peak frequency relationships, particularly if preceding severe rainfall storms occur in close succession. Alexander (2012) also emphasised that: “It is a serious mistake to assume that a single probability distribution method can be applied to all the data at the site.”

The MLVA inclusive of the LP3-GEV/MM distributions dominated the probabilistic analyses in 42% of the catchments, followed by the MLVA inclusive of the LP3-GEV-LN/MM distributions in 28% of the catchments. The LP3/MM distribution was the only distribution which was used as a single most suitable distribution in 11% of the catchments. The remaining 19% of the catchments under consideration were characterised by a different combination of the above-mentioned distributions. However, when Equation 1 was used to combine two or more probability distributions, it was impossible to indicate which probability distribution would be the best suited for a specific return period range in all the catchments. The SDF basin 9 (primary study area) results for return periods ranging from 2 to 200 years based on the most relevant distributions are listed in Table 1.

**Selection of representative rainfall stations**

The different results obtained using the arithmetic mean and Thiessen polygon methods respectively to calculate the average catchment design rainfall are listed in Table 2. Table 3 provides the selected single rainfall stations in each catchment under consideration in SDF basin 9 in comparison with the catchment design rainfall calculated using the Thiessen polygon method.

The number of rainfall stations used for averaging the design rainfall varied from catchment to catchment with an overall average of one station per 100 km². It was observed that the arithmetic mean values slightly exceeded the Thiessen polygon values in each catchment, but the coefficient of determination \( r^2 \) of 0.98 confirmed the high degree of association, and highlighted the even areal distribution of the rainfall stations and the relatively flat topography.

**Numerical calibration of runoff coefficients**

The original, adjusted and calibrated \( C_2 \) and \( C_{100} \) runoff coefficients applicable to SDF basin 9 are presented in Table 4. The corresponding MAP, two-year one-day rainfall \( (M) \) and \( R \) values are also shown. The following notations are applicable to Tables 4 to 7:

- \( SDF_{Original} \) Values as used in the original SDF method (Alexander 2003)
- \( SDF_{Adjusted} \) Values as used in the adjusted SDF method based on the adjustment factors (Van Bladeren 2005)
- \( SDF_{Calibrated} \) Values as used in the calibrated SDF method at a quaternary catchment level based on the methodology of this study
- \( SDF_{Verified} \) Verified SDF method at a quaternary catchment level used to evaluate the calibrated version of the SDF method.

The original and calibrated \( C_2 \) and \( C_{100} \) runoff coefficients of SDF basin 9 presented in Table 4 are characterised by large proportional differences between the values. The differences between the original and adjusted runoff coefficients were found to be 45% (60 – 15) in all cases, since the adjustment factors as proposed by Van Bladeren (2005) are only used to adjust the final \( C_T \) coefficients.

In the case of the calibrated runoff coefficients, the proportional differences tended to decrease with an increase in the MAP, with a 45% difference for MAP less than 500 mm, a 38% to 41% difference for MAP ranging from 500 to 600 mm, and a 22% difference for MAP exceeding 600 mm. It is important to note that the original MAP (376 mm) in SDF basin 9 is also less than 500 mm, confirming the trend identified. It indicated that the antecedent soil moisture status in the quaternary catchment(s) under consideration introduces additional variability into the rainfall-runoff process and that the hydrological response in each quaternary catchment will be different. This can be ascribed to the difference in soil permeability which controls the infiltration rate and consequently the balance of rainfall that constitutes surface runoff and contributes to the flood peak. Table 5 provides a summary...
of the results obtained during the numerical calibration of the SDF runoff coefficients in SDF basin 9. The original, adjusted and calibrated runoff coefficients ($C_T$), as calculated using Equation 2, are shown in Table 5.

The original runoff coefficients of SDF basin 9 ranged from 0.150 (2-year return period) to 0.642 (200-year return period), while the calibrated runoff coefficients ranged from 0.090 (2-year return period) to 0.534 (200-year return period). The adjusted runoff coefficients based on the adjustment factors proposed by Van Bladeren (2005) ranged from 0.103 (2-year return period) to 1.878 (200-year return period). According to Van Bladeren (2005), the latter runoff coefficient exceeding unity is justified, since it is used when the SDF method overestimates the more frequent events and underestimates the extreme events in a particular basin.

However, the question arises whether or not the adjustment factor can successfully describe a meaningful relationship between the regional descriptors (average catchment slope, catchment area, land-use distribution and MAP) and the $C_T$-coefficients as shown in Equation 4 (Van Bladeren 2005).

$$CT_{T2} = \frac{C_{T1}}{F} \tag{4}$$

where:

- $CT_{T2}$ = adjusted runoff coefficient (Van Bladeren 2005)
- $A$ = catchment area (km²)
- $C_{T1}$ = original runoff coefficient (Alexander 2003)
- $F$ = adjustment factor ($F = xA^y$)
- $x$ = regional descriptor (multiplier), and
- $y$ = regional descriptor (exponent).

Based on the results listed in Table 5, it was interesting to note that the runoff coefficients adjusted by Equation 4 had a tendency to decrease in magnitude with increasing recurrence interval. This was especially the case in SDF basins 11 and 17, but the results are not presented here. In SDF basin 9 the 20-year adjusted runoff coefficients exceeded the 50-year runoff coefficients. Similar results were also evident in SDF basin 18.

The adjusted runoff coefficients, which exceeded unity and decreased in magnitude with increasing recurrence interval, deviated from the norm. Runoff coefficients exceeding unity indicate that more than a 100% runoff can occur, but this is physically impossible. An increase in the $C_T$-coefficients with return period is necessary to accommodate the known effects which also increase with return period, but the increase is not accounted for in Equation 4. The coxially plotted values of $C_T$ (original, adjusted and calibrated) against the return period applicable to SDF basins 6, 7, 9, 10, 18 and 21 are shown in Figures 2 to 7.

It is evident from Figure 2 that the calibrated $C_T$-coefficients in the quaternary catchments obtained from the primary study area are spread around those of Alexander (2003), but are generally lower in magnitude. Similar trends were also witnessed in the other SDF basins evaluated (Figures 4 to 6), except for SDF basins 6 and 21 (Figures 3 and 7). The curves representing the calibrated runoff coefficients had similar growth curves as a function of the recurrence interval in most of the basins under consideration. The decreasing trend in the adjusted runoff coefficients with an increase in recurrence interval is clearly evident from Figures 2 and 6, with the adjusted 20-year runoff coefficients highly questionable, since they are larger than the 50- and 100-year runoff coefficients.

### Evaluation of calibrated flood peaks

The comparison between the design flood peak values based on the probabilistic analyses and the original, adjusted and calibrated versions of the SDF method in SDF basin 9 (primary study area) is listed in Table 6. The average SDF/probability distribution ratios are also shown in the same table.

### Table 5 Calibrated runoff coefficients ($C_T$) in SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>Method</th>
<th>Calibrated runoff coefficients ($C_T$) for return period (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5R002</td>
<td>SDFOriginal ($C_T$)</td>
<td>0.150, 0.312, 0.397, 0.467, 0.546, 0.600, 0.648</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($C_T$)</td>
<td>0.155, 0.383, 0.719, 1.338, 1.130, 1.422, 1.878</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($C_T$)</td>
<td>0.090, 0.253, 0.338, 0.408, 0.487, 0.540, 0.588</td>
</tr>
<tr>
<td>C5R003</td>
<td>SDFOriginal ($C_T$)</td>
<td>0.150, 0.312, 0.397, 0.467, 0.546, 0.600, 0.648</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($C_T$)</td>
<td>0.125, 0.287, 0.446, 0.870, 0.753, 0.902, 1.137</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($C_T$)</td>
<td>0.095, 0.232, 0.304, 0.364, 0.430, 0.475, 0.516</td>
</tr>
<tr>
<td>C5R004</td>
<td>SDFOriginal ($C_T$)</td>
<td>0.150, 0.312, 0.397, 0.467, 0.546, 0.600, 0.648</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($C_T$)</td>
<td>0.148, 0.361, 0.653, 1.226, 1.042, 1.299, 1.696</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($C_T$)</td>
<td>0.180, 0.328, 0.406, 0.470, 0.542, 0.590, 0.634</td>
</tr>
<tr>
<td>C5R005</td>
<td>SDFOriginal ($C_T$)</td>
<td>0.150, 0.312, 0.397, 0.467, 0.546, 0.600, 0.648</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($C_T$)</td>
<td>0.103, 0.224, 0.294, 0.597, 0.529, 0.607, 0.733</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($C_T$)</td>
<td>0.110, 0.190, 0.231, 0.266, 0.304, 0.330, 0.354</td>
</tr>
</tbody>
</table>

### Table 6 Calibrated SDF flood estimation results in SDF basin 9 (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>Method</th>
<th>Design flood (m³/s) for return period (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5R002</td>
<td>$Q_p$</td>
<td>218, 616, 1,098, 1,577, 2,327, 2,990, 3,746</td>
</tr>
<tr>
<td></td>
<td>SDFOriginal ($Q_p$)</td>
<td>351, 1,059, 1,692, 2,417, 3,590, 4,678, 5,929</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($Q_p$)</td>
<td>362, 1,298, 3,065, 6,926, 7,438, 11,085, 17,183</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($Q_p$)</td>
<td>204, 773, 1,221, 1,703, 2,404, 3,010, 3,712</td>
</tr>
<tr>
<td>C5R003</td>
<td>$Q_p$</td>
<td>75, 225, 440, 655, 810, 972, 1,143</td>
</tr>
<tr>
<td></td>
<td>SDFOriginal ($Q_p$)</td>
<td>90, 291, 470, 677, 992, 1,271, 1,565</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($Q_p$)</td>
<td>75, 267, 528, 1,261, 1,368, 1,912, 2,744</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($Q_p$)</td>
<td>64, 237, 389, 559, 812, 1,024, 1,253</td>
</tr>
<tr>
<td>C5R004</td>
<td>$Q_p$</td>
<td>290, 646, 935, 1,253, 1,807, 2,366, 3,075</td>
</tr>
<tr>
<td></td>
<td>SDFOriginal ($Q_p$)</td>
<td>236, 710, 1,134, 1,618, 2,402, 3,129, 3,966</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($Q_p$)</td>
<td>233, 821, 1,866, 4,251, 4,585, 6,766, 10,387</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($Q_p$)</td>
<td>290, 709, 1,033, 1,376, 1,878, 2,356, 2,866</td>
</tr>
<tr>
<td>C5R005</td>
<td>$Q_p$</td>
<td>35, 86, 139, 198, 287, 368, 461</td>
</tr>
<tr>
<td></td>
<td>SDFOriginal ($Q_p$)</td>
<td>38, 133, 221, 320, 469, 594, 726</td>
</tr>
<tr>
<td></td>
<td>SDFAdjusted ($Q_p$)</td>
<td>26, 95, 163, 410, 454, 601, 822</td>
</tr>
<tr>
<td></td>
<td>SDFCalibrated ($Q_p$)</td>
<td>32, 92, 147, 209, 300, 375, 455</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>$Q_p/Q_p$</td>
<td>1.18, 1.41, 1.35, 1.37, 1.43, 1.45, 1.45</td>
</tr>
<tr>
<td></td>
<td>$Q_p/Q_p$</td>
<td>1.05, 1.42, 1.79, 2.95, 2.25, 2.54, 3.04</td>
</tr>
<tr>
<td></td>
<td>$Q_p/Q_p$</td>
<td>0.93, 1.12, 1.04, 1.02, 1.03, 1.02, 1.00</td>
</tr>
</tbody>
</table>
Figure 2: Comparison of runoff coefficients in SDF basin 9 (Gericke 2010)

Figure 3: Comparison of runoff coefficients in SDF basin 6 (Gericke 2010)

Figure 4: Comparison of runoff coefficients in SDF basin 7 (Gericke 2010)

Figure 5: Comparison of runoff coefficients in SDF basin 10 (Gericke 2010)

Figure 6: Comparison of runoff coefficients in SDF basin 18 (Gericke 2010)

Figure 7: Comparison of runoff coefficients in SDF basin 21 (Gericke 2010)
According to Van Bladeren (2005), the original SDF method tends to overestimate the more frequent design floods for return periods of up to 20 years in SDF basin 9, while the extreme events are underestimated. However, the results contained in Table 6 indicate that on average the original SDF method overestimated all the probabilistic flood peaks. The overestimation varied between 18% and 45%. Except for the 2-year return period, the adjusted SDF method overestimated all the flood peaks as well, with average overestimations of up to 204%. The calibrated version of the SDF method proved to be the most accurate, with the 2-year return period being underestimated on average by 7%, while the maximum overestimation was limited to 12%.

Apart from the above-mentioned results obtained in SDF basin 9, the original SDF method also overestimated the probabilistic flood peaks in all the other SDF basins under consideration, except in SDF basins 6 (200-year), 8 (2-year), 16 (2-year), 23 (100- and 200-year) and 26 (all return periods). On average, the ratio between the original SDF method and probability distributions varied between 0.43 and 5.77. The best results were evident in SDF basin 16 (Figure 10), with the overestimation limited to ± 12% for the return periods ranging from 10 to 200 years.

The overall results (calibration and verification) of the adjusted SDF method were only better in 26% of all the basins under consideration when compared to those estimated by the original SDF method. The adjusted SDF method demonstrated the most acceptable results in SDF basin 16 (Figure 10), and either significantly over- or underestimated the flood peak values in the remaining basins. On average, the adjusted SDF/ probability distribution ratios varied between 0.25 and 6.58, which seems improper.

The calibrated version of the SDF method proved to be the most accurate in all the basins under consideration, except in...
SDF basin 6 (Figure 8), where it proved to be slightly less accurate than the adjusted SDF method. On average, the calibrated SDF/probability distribution ratios varied between 0.91 and 1.30, while at some basins and individual return periods, less accurate results were evident.

**Verification of calibrated runoff coefficients and flood peaks**

In the methodology it was highlighted that the selection of verification catchments within the same basin was based on the fact that their physical and regional descriptors were similar. In other words, the catchments are situated within a smaller portion/number of quaternary catchments within the larger group of quaternary catchments used in the calibration exercise. The flood peaks estimated for verification purposes with calibrated runoff coefficients are referred to as $SDF_{\text{Verified}} (Q_i)$. The comparison between
the design flood peak values based on the probabilistic analyses and the original, adjusted and calibrated/verified versions of the SDF method in SDF basin 9 is listed in Table 7. The flow-gauging station numbers in brackets in column 1 are the stations used for the verification.

The verification results listed in Table 7 indicate that the original SDF method demonstrated the same trends of overestimating all the probabilistic flood peaks in SDF basin 9. However, the magnitude of overestimation of the 2- and 5-year return period floods were slightly larger, while the flood peaks for the remaining return periods showed some improvement with the overestimation being limited to ±30%. The adjusted SDF method results also improved slightly and were characterised by overestimations of up to 163%. The verification results confirmed that the calibrated SDF method was the most accurate, and similar trends were evident. On average, the verified SDF/probability distribution ratios varied between 1.01 and 1.18, which is considered to be acceptable.

Figures 12 and 13 provide a visual measure of performance showing the average SDF/probability distribution ratios in the primary study area (SDF basin 9) and SDF basins 6, 7, 10, 18 and 21 obtained during the verification exercise.

Apart from the above-mentioned verification results obtained in SDF basin 9, varied the verified SDF/probability distribution ratios in the primary study area (SDF basin 9) and SDF basins 6, 7, 10, 18 and 21 obtained during the verification exercise.

Apart from the above-mentioned verification results obtained in SDF basin 9, varied the verified SDF/probability distribution ratios on average between 0.80 and 1.20. However, the 5- to 20-year return period flood peaks were overestimated by 41% to 56% in SDF basins 6 (Figure 12) and 21 (Figure 13). The calibrated/verified SDF method remains the preferred method in the latter basins, based on the higher degree of association and accuracy obtained. The original and adjusted versions of the SDF method also demonstrated similar trends as established during the calibration exercise, although some individual return period flood peaks were characterised by either a slightly improved or worse estimation. All the verification tests also confirmed that the calibrated runoff coefficients behaved in a probabilistic manner as anticipated, since the verification results showed that the calibrated/verified SDF method is the most accurate, and similar trends were evident in all the basins under consideration.

### Statistical assessment of results

The coefficients of determination ($r^2$) results were indicative of a high degree of association between the observed AMS data and the theoretical probability distributions, with 0.85 and 0.79 respectively as the poorest correlations in the primary (SDF basin 9) and secondary study areas.

In all the primary study area catchments, except C5R003, C5H003 and C5H018, the

---

<table>
<thead>
<tr>
<th>Catchment description</th>
<th>Method</th>
<th>Design flood (m³/s) for return period (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>C5R002 (C5R001)</td>
<td>$Q_p$</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Original}$ ($Q_1$)</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$ ($Q_2$)</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Verified}$ ($Q_3$)</td>
<td>39</td>
</tr>
<tr>
<td>C5R002 (C5H008)</td>
<td>$Q_p$</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Original}$ ($Q_1$)</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$ ($Q_2$)</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Verified}$ ($Q_3$)</td>
<td>40</td>
</tr>
<tr>
<td>C5R003 (C5H003)</td>
<td>$Q_p$</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Original}$ ($Q_1$)</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$ ($Q_2$)</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Verified}$ ($Q_3$)</td>
<td>89</td>
</tr>
<tr>
<td>C5R004 (C5H015)</td>
<td>$Q_p$</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Original}$ ($Q_1$)</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Adjusted}$ ($Q_2$)</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>SDF$_{Verified}$ ($Q_3$)</td>
<td>292</td>
</tr>
<tr>
<td>Average</td>
<td>$Q_d/Q_p$</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>$Q_d/Q_p$</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>$Q_d/Q_p$</td>
<td>1.02</td>
</tr>
</tbody>
</table>
Chi-square statistic was less than the limiting critical value and the confidence level larger than the significance level. In other words, the null hypothesis could be accepted. This was also the case in 60% of the secondary study area catchments used to verify the calibrated SDF method. However, acceptance of the null hypothesis at low confidence levels (< 50%), highlighted the likelihood of differences to be present, especially at C5R001, C5R002 and C5H015 in SDF basin 9. Similar detectable differences were present in 12% of the secondary study area catchments.

The GOF statistic results of the original, adjusted, calibrated and verified versions of the SDF method were evaluated within the context of their SDF/probability distribution ratios. In other words, pair values of the coefficient of determination and these ratios were evaluated in combination to get a true reflection of the accuracy. Typically, the calibrated and verified versions of the SDF method proved to be the most accurate, i.e. SDF/probability distribution ratios between 0.80 and 1.30 along with a high degree of association ($r^2$ values > 0.9). On the other hand, the original and adjusted SDF methods also demonstrated a high degree of association, but it must be evaluated within the context of their poor SDF/probability distribution ratios, which were occasionally different by up to a factor of 3 or more.

The influence of different variables on the calibration and verification results, as discussed, demonstrates that the probabilistic-based approach of the SDF method in its current format requires further refinement at a quaternary catchment scale. The conclusions and recommendations in the following section will synthesise the results in order to address these requirements.

**CONCLUSIONS AND RECOMMENDATIONS**

The main purpose of this study was to establish whether or not the probabilistic-based approach of the SDF method has the ability to overcome some of the deficiencies evident in the other design flow estimation methods used in South Africa. Although the study was limited to only 19 of the 29 SDF basins in South Africa, the overall results are considered to be representative in a South African context, while all the hypothesis statements were investigated and confirmed.

The conclusions and recommendations pertaining to the results obtained from this study can be summarised as follows:

- **Probabilistic analyses**: The results confirmed that the LN, LP3 and GEV distributions or a combination thereof using Equation 1, are the most suitable probability distributions for flood frequency analysis in South Africa at a single site (Hypothesis 4). The selection and use of probability distribution pair combinations for specific return periods can only be based on the statistical properties, visual inspection of the plotted values and GOF statistics. However, regional analyses of pooled-AMS could also be conducted to overcome the problem of short flow record lengths at single sites in addition to the MLVA.

- **Selection of representative rainfall stations**: The use of average meteorological conditions and record length as criteria to select single representative rainfall stations in each catchment under consideration confirmed Hypothesis 2, which stated that the flood-producing characteristics within the current delineated SDF basins are non-homogeneous. The newly identified single rainfall stations proved to be a much better representation of the hydrological response at a smaller scale or catchment level.

- **Numerical calibration of runoff coefficients**: It was evident from this study that the calibrated $C_T$ coefficients at a quaternary catchment level significantly improved the accuracy of the flood peak estimation using the SDF method (Hypothesis 1). The effort made by Van Bladeren (2005) to regionalise the runoff coefficients requires further improvement, since some of the adjusted runoff coefficients exceeded unity and other had a tendency to decrease in magnitude with increasing recurrence interval. The relationships established between the parameters (multiplier and exponent) of the power-law function (Equation 4) fitted to the $C_T$ coefficients as a function of return period and regional descriptors, are also questionable, because the likelihood that a catchment is to be more saturated at the start of a storm with a longer recurrence interval was ignored.

These results are in agreement with similar studies conducted on the RM in South Africa (Parak & Pegram 2006) and Australia (Pilgrim & Cordery 1993), which confirmed that no relationship can be successfully established between the regional descriptors and the $C_T$ values in order to regress the runoff coefficients. It also confirms that the $C_T$ coefficients are essentially functions of the return period and time of concentration as conjectured (Hypothesis 3).

- **Calibrated and verified flood peaks**: The calibrated/verified version of the SDF method proved to be the most...
accurate at a quaternary catchment level, thus enhancing the accuracy and practical use of the original SDF method (Hypothesis 1). The degree or extent to which the original SDF method overestimated the magnitude and frequency of flood peaks varied from basin to basin. Apart from the previously discussed factors, this is also due to the influence of different climatic regions, types of weather and rainfall occurrence frequencies on the depth, area, duration and movement of storm rainfall. The original SDF/probability distribution ratios were the highest in the northern inland regions (SDF basins 1, 2 and 4), the Highveld (SDF basins 6 and 7), Lesotho (SDF basins 10 and 11) and southern coastal regions (SDF basins 22 and 23) with summer convective rainfall. In these regions the flood-peak ratios were occasionally different by up to a factor of 3 or more. The southern coastal regions (SDF basins 16 to 18) with winter orographic/frontal rainfall demonstrated the best flood peak ratios with flatter growth curves as a function of the recurrence interval and varied between 0.8 and 1.6.

- **Statistical assessment of results:** Both the coefficient of determination and Chi-square statistic can be satisfactorily used to evaluate the GOF of theoretical probability distributions and other design flood estimation methods. However, the Chi-square statistic proved to be more sensitive towards short record lengths, inconsistency, non-homogeneity and non-stationarity of data.

**Further research on the SDF method could focus on the following:**

- **Review of the current regional boundaries of the SDF basins by increasing the number of SDF basins based on single or multiple quaternary catchment boundaries.** The availability of hydrological (flow) and meteorological (rainfall) data, as well as the extent of the hydrological homogeneity within the identified catchments, will have an influence on the identification and delineation of the new basins.

- **Improvement and extension of the data pool of hydrological and meteorological**

**gauging sites by updating the data sets to ensure that periods of observation are as long as possible. All available historical information of flood peaks should be included in and made available from a central database.**

- **Utilisation of the Regional Linear Moment Algorithm and Scale Invariance (RLMA&SI) approach (Smithers & Schulze 2000a) to estimate design rainfall.**

- **Establishment of physical or regional descriptors on which to regress the calibrated runoff coefficients to enable the extension thereof to ungauged catchments.**

**IN CONCLUSION**

All these results emphasised that there is no single design flood estimation method that is superior to all other methods used to address the wide variety of flood magnitude frequency problems that are encountered in practice. Design engineers still have to apply their own experience and knowledge to these particular problems until the search for a universally applicable design flood method in South Africa produces a method by which to overcome all the inherent uncertainties present in flood hydrology.

**REFERENCES**


Alexander, W J R, Professor Emeritus, University of Pretoria, personal communication, 17 February 2012.


DWAF (Department of Water Affairs and Forestry) 1995. GIS data: Primary drainage regions of South Africa. Pretoria: DWAF.


Catchment parameter analysis in flood hydrology using GIS applications

O J Gerick, J A du Plessis

The use of Geographical Information Systems (GIS) has permeated almost every field in the engineering, natural and social sciences, offering accurate, efficient, reproducible methods for collecting, viewing and analysing spatial data. GIS do not inherently have all the hydrological simulation capabilities that complex hydrological models do, but are used to determine many of the catchment parameters that hydrological models or design flood estimation methods require. The purpose of this study was to perform catchment parameter analysis using GIS applications available in the ArcGIS™ environment. The paper will focus on the deployment of special GIS spatial modelling tools versus conventional manual methods used in conjunction with standard GIS tools to estimate typical catchment parameters, e.g. area, average catchment and watercourse slopes, main watercourse lengths and the catchment centroid. The manual catchment parameter estimation methods with GIS-based input parameters demonstrated an acceptable degree of association with the special GIS spatial modelling tools, but proved to be sensitive to biased user-input at different scale resolutions. GIS applications in an ArcGIS™ environment for the purpose of catchment parameter analyses are recommended to be used as the standard procedure in any proposed hydrological assessment.

INTRODUCTION

The use of Geographical Information Systems (GIS) has permeated almost every field in the engineering, natural and social sciences, offering accurate, efficient, reproducible methods for collecting, viewing and analysing spatial data. These spatial data sets represent the key components in the hydrological response of catchments to storm rainfall and the resulting runoff. GIS do not inherently have the hydrological simulation capabilities that complex hydrological models do, but are used to determine many of the catchment parameters that hydrological models or design flood estimation methods require.

In hydrological catchment parameter analyses, a Digital Elevation Model (DEM) and spatial data sets represent the two fundamental data sets initially required. The DEM contains raster information of the catchment and surrounding areas, while the spatial data sets contain the spatial information which originates from other sources than the DEM. The DEM is used to do a complete catchment parameter analysis, including the determination of flow directions, catchment areas, land surface and river channel characteristics. The spatial data sets contain layers of combined spatial information used to analyse the spatial distribution and associated attributes of geology, soil, land use and vegetation.

In addition to catchment parameter analysis, GIS also provide a powerful data management framework with a consistent, intuitive platform for organising and analysing relationships amongst the spatial variables and information associated with those variables encountered in the field of flood hydrology. Various GIS software packages exist. This paper will, however, only refer to the Environmental Systems Research Institute (ESRI) GIS software in the form of ArcGIS™ 9.3.

The purpose of the study is discussed and explained in the next section, followed by an overview of the study area’s spatial distribution and characteristics. In the section thereafter, the methods used in South Africa to estimate catchment parameters are reviewed in detail. The methodologies involved in assessing the paper’s purpose and objectives are then expanded on in detail, followed by the results, discussion and conclusions.

PURPOSE OF STUDY

The purpose of this study was to perform catchment parameter analysis using GIS applications available in the ArcGIS™ environment. The focus was on the deployment of special GIS spatial modelling tools versus conventional manual methods used in conjunction with standard GIS tools to estimate typical catchment parameters.
It was hypothesised that the accuracy of conventional manual procedures used in flood hydrology to establish typical catchment parameters could be improved by using automated GIS input processing functionalities, since manual inputs are regarded as insufficiently accurate and outdated. It was further hypothesised that the spatial distribution of slope classes, used as primary input data to the deterministic flood estimation methods, are not sufficiently representative of the specific conditions under evaluation. Many practitioners in the field of flood hydrology typically ignore the importance thereof and follow a “thumb-suck” approach. In addition, hydrologists and engineers are frequently doubtful when deciding on, or determining, the position of the catchment centroid.

**STUDY AREA**

The study area covers 34 795 km² between 28°25’ and 30°17’ South and 23°49’ and 27°00’ East, and comprises the C5 secondary drainage region. The tertiary drainage regions of concern are C51 (Riet River Catchment (RRC)) and C52 (Modder River Catchment (MRC)), covering an area of 17 435 km² and 17 360 km² respectively. The MRC and RRC consist of eleven and twelve quaternary catchments respectively (Midgley et al 1994).

The topography is gentle (average quaternary catchment slopes between 2.4% and 5.5%), while the mean altitude above sea level varies between 997 m and 2 122 m (NASA 2002). Twelve catchments with contributing catchment areas consisting of either single or multiple tertiary and quaternary catchments are presented in the Table 1. The catchment areas are based on their contributing areas consisting of either single or multiple tertiary and quaternary catchments. The study area encompasses the C5 secondary drainage region, and the primary drainage regions C51 and C52, as shown in Figure 1. The C5 secondary drainage region is delineated by the catchment descriptor (flow-gauging station) and study area and comprises the C51 and C52 tertiary drainage regions. The primary drainage regions consist of the C51 and C52 tertiary drainage regions, as shown in Figure 2. The study area is delineated by the catchment descriptor (flow-gauging station) and study area and comprises the C51 and C52 tertiary drainage regions. The primary drainage regions consist of the C51 and C52 tertiary drainage regions, as shown in Figure 2.

![Figure 1 Location of study area in relation to the primary drainage regions of South Africa](image1.jpg)

![Figure 2 Hydrology toolset and associated Watershed tool](image2.jpg)

![Figure 3 Average catchment slope using the Grid method (Alexander 2001)](image3.jpg)
multiple C51 or C52 quaternary catchments were evaluated individually in the study area. A Department of Water Affairs (DWA) flow-gauging station is situated at the outlet of each of these catchments. The flow-gauging station numbers were therefore used as the catchment descriptor for easy reference in all the Tables and Figures included in this paper. The general information applicable to these catchments is listed in Table 1, while the location thereof within the study area and in relation to the primary drainage regions of South Africa is shown in Figure 1.

REVIEW OF CATCHMENT PARAMETER ESTIMATION METHODS

To provide the background for further discussion, the manual and automated methods used in design flood estimation to establish catchment parameters, will now be discussed briefly.

Catchment area

The standard maps recommended to manually determine the catchment areas for use in flood hydrology are either the 1:50 000 scale topographical maps and/or 1:10 000 scale orthophotos. The latter are normally used if the catchment area under consideration is less than 10 km². The manual procedure to determine the catchment area entails the demarcated catchment boundary on the map is copied onto graph paper, after which the number of squares within the catchment are counted by including squares more than halfway into the catchment. A conversion factor is then used to convert the number of squares to the catchment area in km² (Alexander 2001). Planimeters are also still in use to measure the manually demarcated catchment areas.

Alternatively, the aforementioned standard maps in an electronic format can be imported to a suitable Computer-Aided Design (CAD) environment as a picture file, after which standard CAD functions are used for the demarcation and area calculation respectively. The use of Google Maps as alternative is also worthwhile to consider.

In an ArcGIS™ 9.3 environment, the Watershed tool contained in the Hydrology toolset of the Spatial Analyst Tools toolbox (Figure 2) can be used to identify catchment areas for specified pour points representative of the catchment outlet. However, a hydrologically correct and depressionless DEM must be prepared for these calculations, using most of the tools contained in the Hydrology toolset.

Average catchment slope

Slopes, whether gentle or steep, influence the catchment response time and hence the duration of critical rainfall intensity and resulting peak discharges and volumes (Alexander 2001). The average catchment slope (S) can be determined by using any one of the Grid, Empirical or Neighbourhood methods in conjunction with standard tools available in the ArcGIS™ 9.3 environment.

Grid method

A grid of at least 50 squares must be superimposed over the catchment area. At each grid intersection point, the horizontal (shortest) distance between the contour intervals which straddle the grid point along a line that passes through the grid point, is measured. The average catchment slope is consequently defined as the average slope perpendicular to the nearest contour line at each grid point. This is presented diagrammatically in Figure 3 and expressed by Equation 1 (Alexander 2001).

\[ S = \frac{\Delta H}{\sum_{i=1}^{N} L_i} \]  

where:

- \( S \) = average catchment slope (m/m)
- \( \Delta H \) = contour interval (m)
- \( L_i \) = horizontal distance between consecutive contours (m), and
- \( N \) = number of grid points.

Empirical method

According to Schulze et al. (1992), the average catchment slope can be determined by making use of the following empirical relationship (Equation 2):

\[ S = MAH^{10^{-2}} \]  

where:

- \( S \) = average catchment slope (m/m)
- \( A \) = catchment area (km²)
- \( \Delta H \) = contour interval (m), and
- \( M \) = total length of all contour lines within the catchment (m).

Equation 2 is not widely used, especially due to the tedious task to determine the \( M \) values manually. However, the use of Equation 2 in its more rudimentary form (derived from first principles), in conjunction with standard functions in ArcGIS™, will be highlighted further in the Methodology.

Neighbourhood method

This method is also known as the Average Maximum Technique (Equation 3) and is included as the standard slope algorithm in the ArcGIS™ environment to generate slope rasters from raw DEM and/or point elevation GIS data sets to enable the determination of average catchment slopes and steepness frequency distributions. The slope raster generation is based on a cell matrix approach which represents the maximum change in elevation over the distance between the cell and its eight neighbouring cells. Typically, in a 3 x 3 search window (grid network with nine cells, \( C_1 \) to \( C_9 \)), eight grid points from the surrounding cells are used to calculate the average slope of the central cell (\( C_s \)) using unequal weighting coefficients, which are proportional to the reciprocal of the square of the distance from the central cell (Jones 1998; ESRI 2006b).

\[ S = \sqrt{\frac{\Delta x^2}{\Delta y^2} + \frac{\Delta z^2}{\Delta y^2}} \]  

where:

- \( S \) = average catchment slope (m/m)
- \( \Delta x \) = rate of change of the slope surface in a horizontal direction from centre cell
- \( \Delta y \) = rate of change of the slope surface in a vertical direction from centre cell

\[ \frac{C_{j \neq s} - C_s}{(Ny_C)} \]  

\( C_{j \neq s} \) = surrounding cells
- \( C_s \) = centre cell
- \( N \) = number of grid points or cells
- \( x_C \) = horizontal cell size, and
- \( y_C \) = vertical cell size.

Length and average slope of main watercourses

The main watercourse is a defined flow path along which water will travel the longest time to reach the catchment outlet from a point on or near the catchment boundary. This distance can be measured manually on orthophotos or topographical maps by using dividers set at a predefined incremental distance which is a function of the map scale (Alexander 2001). The average main watercourse slope can be determined manually by using the following methods (Alexander 2001; Van der Spuy & Rademeyer 2010):

Equal-area method

An average slope line is drawn or positioned in relation to the longitudinal profile of the main watercourse in such a way that the area above \( A_L \) this line equals the area below \( A_L \) the line. This relationship is expressed by Equation 4 and illustrated in Figure 4.

\[ S_{CHI} = \frac{H_L - H_{GL}}{L} \]  

where:

- \( H_L \) = elevation of the longitudinal profile
- \( H_{GL} \) = elevation of the grid profile
- \( L \) = grid line length in km

\[ C_{j \neq s} = \]
where:

\[ S_{CH1} = \text{average main watercourse slope (m/m)} \]

\[ A_i = \frac{(H_i + H_{i+1}) - H_B}{2} L_i \]

\[ H_T = \frac{\sum_{i=1}^{N} A_i}{L} + H_B \]

\[ H_B = \text{height at catchment outlet (m)} \]

\[ H_i = \text{specific contour interval height (m)} \]

\[ L = \text{length of main watercourse (m)}, \text{and} \]

\[ L_i = \text{distance between two consecutive contours (m)}. \]

**10-85 method**

This method was developed by the United States Geological Survey (USGS) and is the most widely used in South Africa (SANRAL 2006). This relationship is expressed by Equation 5 and illustrated in Figure 5.

\[ S_{CH2} = \frac{H_{0.85L} - H_{0.10L}}{750L} \] (5)

where:

\[ S_{CH2} = \text{average main watercourse slope (m/m)} \]

\[ L = \text{length of main watercourse (km)} \]

\[ H_{0.85L} = \text{height (m) of main watercourse at length 0.85L, and} \]

\[ H_{0.10L} = \text{height (m) of main watercourse at length 0.10L}. \]

**Taylor-Schwarz method**

This method is preferred by the Department of Water Affairs (DWA) and the Natural Environment Research Council (NERC 1975). The latter also proposed the use thereof in the United Kingdom Flood Studies Report (UK FSR 1975) (Van der Spuy & Rademeyer 2010). The main watercourse profile is subdivided into sub-reaches of which the velocities are related to the square root of the slope. The index is equivalent to the slope of a uniform channel with the same length as the longest watercourse and an equal travel time. This relationship is expressed by Equation 6 and illustrated in Figure 6.

\[ S_{CH3} = \left( \frac{L}{\sum_{i=1}^{N} L_i / S_i} \right) \] (6)

where:

\[ S_{CH3} = \text{average main watercourse slope (m/m)} \]

\[ L = \text{length of main watercourse (m)} \]

\[ L_i = \text{distance between two consecutive contours (m)}, \text{and} \]

\[ S_i = \text{slope between two consecutive contours (m/m)}. \]

In the ArcGIS™ environment, both fully and semi-automated methods are available.
to estimate the main watercourse length. The *Longest Flow Path* tool which forms part of the *ArcHydro* toolbox automatically determines the longest watercourse. However, a hydrologically correct and depressionless DEM based on extensive input rasters with increased computing time is required. The use of semi-automated methods in conjunction with Equations 4 to 6 will be expanded on in detail in the Methodology.

**Distance to catchment centroid**

According to Alexander (2001), an eyeball estimate of the location of the catchment centroid is adequate. In practice, the distance to the centroid can be determined manually by using a cut-out of the catchment area to hang freely from a pin inserted close to a border of the catchment area. A string with a weight attached to the bottom, attached to the pin and hanging vertically under gravity from the pin, provides a guideline on the paper cut-out of the catchment.

With the guideline drawn on the catchment, the pin is then moved to another position (approximately rotated 90° from the first position) close to the boundary of the catchment. The intersection of the two guidelines on the catchment provides the approximate position of the centroid.

In the *ArcGIS* environment, the location of the catchment centroid can be automatically determined by making use of the *Mean Center* tool available in the *Measuring Geographic Distributions* toolset of the *Spatial Analyst* toolbox, which will be expanded on in more detail in the Methodology.

**METHODOLOGY**

To evaluate the deployment of special GIS spatial modelling tools versus conventional manual methods used in conjunction with standard GIS tools to estimate typical catchment parameters, the following procedures were followed:

**Projections and catchment geometry calculations**

All the relevant GIS and catchment related data were obtained from the DWA (Directorate: Spatial and Land Information Management), which is responsible for the acquisition, processing and digitising of the data. These data sets were normally presented as geographical coordinate systems; in other words, the position of a geographical location on the earth’s surface is described by using spherical measures of latitude and longitude (in degrees) from the centre of the earth to a point on the earth’s surface.

<table>
<thead>
<tr>
<th>Parameter description</th>
<th>Modified (original) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>False easting</td>
<td>0 (0)</td>
</tr>
<tr>
<td>False northing</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Central meridian</td>
<td>24 (25)</td>
</tr>
<tr>
<td>Standard parallel 1</td>
<td>-18 (-20)</td>
</tr>
<tr>
<td>Standard parallel 2</td>
<td>-32 (-23)</td>
</tr>
<tr>
<td>Latitude of origin</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Linear unit</td>
<td>metre</td>
</tr>
</tbody>
</table>

These geographical input data sets need to be transformed to a projected coordinate system, which portrays the curved surface of the earth on a flat surface, during which the distance, area, shape and direction, or a combination thereof, might be distorted (ESRI 2006a).

The Africa Albers Equal-Area projected coordinate system, with modification, was used during this study. This approach is best suited for land masses extending in an east-to-west orientation (as in the case of the study area), rather than those lying north-to-south. This conic projection uses two standard parallels to reduce some of the distortion of a projection with one standard parallel. Although neither shape nor linear scale is truly correct, the distortion of these properties is minimised in the region between the standard parallels. All areas are proportional to the same areas on the earth, while distances are most accurate in the middle latitudes (ESRI 2006a).

The standard parallels were established by using the *one-sixth rule* by determining the range in latitude (degrees) north to south, divided by six. The first standard parallel is positioned at one-sixth the range above the southern boundary and the second standard parallel minus one-sixth the range below the northern boundary (ESRI 2006a). These modifications are listed in Table 2.

The specific GIS data features classes (lines, points and polygons) applicable to the study area, and individual sub-catchments were extracted and created from the original GIS data sets by using the *Clip* tool available from the *Extract* toolset contained in the *Analysis Tools* toolbox. The *Clip* tool cuts out a piece of one feature class using one or more of the features in another feature class as a cookie cutter. Either the tertiary or quaternary drainage region polygons were used as clip feature classes, since a clip feature class has to be a polygon. The data extraction was followed by data projection and transformation, editing of attribute tables and redetermination of catchment geometry (areas, perimeters and distances).

**Digital Elevation Model**

The Shuttle Radar Topography Mission (SRTM) elevation data for southern Africa at 90 metre resolution (NASA 2002) was extracted, projected and transformed for the study area and used as the DEM. An alternative DEM was also generated by making use of point elevation and/or contour data as the input features. The *Interpolation* tool contained in the *Spatial Analyst* toolbox was used to generate rasters for the DEM interpolation process. The input features (contours or point elevations) were selected, the *Output Surface Raster* was specified and *Tolerance 1* was set to a value of 10, which is equal to half the contour interval, or set to zero if point elevations are predominately used. The *Output Cell Size*, which specifies the output raster cell size, was then selected. A smaller cell size increases the amount of cells in the raster matrix with both an increased accuracy and computing time. A trade-off between time and accuracy was used in selecting the output cell size.

**Average catchment slope**

The average catchment slope of the study area, as well as of individual catchments, was determined by using the following manual methods with GIS-based input parameters:

**Grid method**

The *Create Vector Grid* tool available in the *Sampling* toolset of the *Hawth’s Analysis Tools* toolbox was used to superimpose a grid over the catchment areas. Refer to Figure 7 for the *Create Vector Grid* data input screen. In Figure 7, the *Extent* selection was in accordance with the extent of the catchment boundary under consideration, while polygon features were selected as the required *Output*, since this option enables geometry (area) calculations. Shapefiles containing the polylines as feature type were created in *ArcGIS*, via the *Sketch* tool accessible from the *Edit* toolbar, to represent the horizontal distances measured at each grid intersection point between two consecutive contours (e.g. Figure 8). The attribute table of each developed shapefile was edited and the length of each polyline was determined by making use of the *Calculate Geometry* function. These attribute tables were then exported to Microsoft Excel for further computations.
It is important to note that the Hawth's Analysis Tools Version 3.27 (Beyer 2004) is not a standard toolbox available in ArcGIS™ 9.3, but it can be downloaded from either www.ESRI.com or www.spatial ecology.com/htools. However, this toolbox is only compatible with ArcGIS™ 9.3 or earlier versions, since the ArcGIS™ programming interface (ArcObjects) changed with the update to ArcGIS™10. In this new version of ArcGIS™ the Hawth’s Analysis Tools was replaced with a toolbox known as the Geospatial Modelling Environment (GME). The GME incorporates most of the functionality of its predecessor, but has a greater range of analysis and modelling tools, supports batch processing, offers new graphing functionality, automatically records work-flows for future reference and supports geodatabases (Beyer 2009).

Empirical method

The Sum Line Lengths in Polygons tool (Figure 9) in the Analysis Tools toolset contained in the Hawth’s Analysis Tools toolbox was used to calculate the total length of all contour lines \( M \) within each catchment, after which it was used as an input variable for Equation 2. The other input variables, area \( A \) and the contour interval \( \Delta H \), were obtained from the relevant developed feature classes of the study area.

Neighbourhood method

A slope raster was generated from the raw DEM data using the Slope tool available from the Surface toolset contained in the Spatial Analyst Tools toolbox. The generated slope raster is based on a cell matrix approach, which represents the maximum change in elevation over the distance between the cell and its eight neighbouring cells, thus the maximum slope for each cell. The Zonal Statistics as Table tool

in the Zonal toolset contained in the Spatial Analyst Tools toolbox (Figure 10) was applied on the slope raster to generate a summary table containing the statistical information about the input data or raster for a defined zone within

![Figure 7 Create Vector Grid data input screen](image)

![Figure 8 Example of horizontal distances at grid intersection points](image)

![Figure 9 Sum Line Lengths in Polygons data input screen](image)

![Figure 10 Zonal Statistics as Table data input screen](image)

![Figure 11 Summary table of average slopes in each quaternary catchment](image)

![Figure 12 Example of reclassified Summary table with slope frequency distribution classes](image)
the data frame, thus the average slope for each catchment (Figure 11). The slope raster was converted to a feature class (polygons) and reclassified into four slope frequency distribution classes, e.g. 0-3%, 3-10%, 10-30% and >30% as required by the deterministic flood estimation methods (SANRAL 2006; Van der Spuy & Rademeyer 2010) to establish the surface slope coefficients associated with different Mean Annual Precipitation (MAP) ranges. This conversion was done by using the Raster to Polygon tool in the Conversion Tools toolbox of ArcToolbox, while the Reclass tool in the Reclass toolset contained in the Spatial Analyst Tools toolbox was used for the reclassification. The reclassified summary table is shown in Figure 12.

Length and average slope of main watercourses

The main watercourse in each catchment was manually identified in ArcMap. A new shapefile containing polyline feature classes representative of the identified main watercourse was created by making use of the Trace tool in the Editor toolbar. Each identified main watercourse was traced using the polyline feature classes of the 20 m interval contour shapefile as the specified offset or point of intersection, resulting in chainage distances between two consecutive contours. The attribute table of each shapefile was then edited by using the Add Field function to include the reduced heights of the contour intervals, and the length of each polyline was determined by making use of the Calculate Geometry function. These attribute tables (e.g. Figure 13) can then be exported to Microsoft Excel for further computations and used as input data for the deterministic and empirical methods used in design flood estimation.

Distance to catchment centroid

The centroid of each catchment under consideration was determined by making use of the Mean Center tool in the Measuring Geographic Distributions toolset contained in the Spatial Statistics Tools toolbox (Figure 14). Only the input polygon feature class representative of each catchment has to be selected to result in a point output feature class and associated attribute table representative of the x and y coordinate of the geometric centroid of each catchment (e.g. Figure 15). The length of the identified main watercourse in each catchment to a point opposite the identified centroid within the catchment was established by using the Measure tool in ArcMap. This measured length \( L_C \) represents the distance along the main watercourse between the outlet and the point closest to the centroid of the catchment (e.g. Figure 15).

RESULTS AND DISCUSSIONS

The results based on the methodology used during this study will now be discussed.

Projections and catchment geometry calculations

The frequency distribution of the altitude-above-sea-level classes present in the study area is summarised in Table 3 (a), while the slope-frequency-distribution classes based on the developed DEM (slope raster) are listed in Table 3 (b). The class-to-class variation and frequency distribution of the altitude-above-sea-level classes are indicative that the topography is relatively flat and that flood peaks will be attenuated and translated both in magnitude and duration respectively. The developed DEM for the study area is illustrated in Figure 16.
Average catchment slope

The results of the average catchment slope calculations based on the Neighbourhood method (DEM data), Grid method and Empirical method as used in the specific catchments, are listed in Tables 4 to 6. The scatter plots are shown in Figures 17 and 18. The developed DEM data was used as the baseline data for the evaluation of and/or comparisons with the two other methods.

According to Alexander (1990) there must be at least 50 grid points within a catchment, while Van der Spuy & Rademeyer (2010) suggested that the minimum number of grid points in catchments smaller or larger than 10 km² must be 20 and 50 respectively. The number of grid points used varied from 50 to 7200, with an overall average of 0.45 grid points per km². The results indicated that either an increase or decrease in the number of grid points per km² does not necessarily guarantee higher accuracies when compared with the Neighbourhood method (DEM data). For comparison purposes, the average catchment slopes (as %) for all the catchments were plotted as a scatter plot using the Neighbourhood method slopes against the Grid method slopes. The results are illustrated in Figure 17.

The Grid method underestimated the average catchment slope in all the catchments under consideration compared to the Neighbourhood method. The underestimation varied between 16.7% (0.48 grid points/km²) and 32.5% (0.34 grid points/km²). Thus, if the DEM data based on the Neighbourhood method are accepted as true and accurate, then the average slope calculation using the Grid method with GIS-based input parameters must be increased with a value of between 17% and 33%. The inverse is also true. No definite relationship between the catchment area and these underestimations could be established.

The coefficient of determination ($r^2$) of 0.88 is indicative of a high degree of association between the two methods. The Grid method is also useful for the development of slope frequency distribution classes used in the deterministic flood estimation methods. The Grid method is, however, time-consuming and sensitive to biased user input at different scale resolutions, extent of catchment areas and contour intervals used.

The results (Figure 18), based on the Empirical method (Equation 2), compared well with the Neighbourhood method. Since Equation 2 is a function of the catchment area ($A$), contour interval ($\Delta H$) and total length of all contour lines within the catchment ($M$), the influence of each variable was evaluated. The results (Table 6) were indicative that there is only a direct relationship between $M$ and $A$ for slopes steeper than 4%, since flatter slopes will result in a lower

Table 4 Average catchment slope based on Neighbourhood method (DEM data) (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Area ($A$, km²)</th>
<th>Average slope ($S$, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5R001</td>
<td>921.6</td>
<td>3.054</td>
</tr>
<tr>
<td>C5R002</td>
<td>10 259.9</td>
<td>4.369</td>
</tr>
<tr>
<td>C5R003</td>
<td>936.7</td>
<td>5.044</td>
</tr>
<tr>
<td>C5R004</td>
<td>6 330.9</td>
<td>4.186</td>
</tr>
<tr>
<td>C5R005</td>
<td>116.4</td>
<td>5.501</td>
</tr>
<tr>
<td>C5H003</td>
<td>600.0</td>
<td>5.044</td>
</tr>
<tr>
<td>C5H012</td>
<td>2 366.3</td>
<td>4.771</td>
</tr>
<tr>
<td>C5H015</td>
<td>6 009.0</td>
<td>4.186</td>
</tr>
<tr>
<td>C5H016</td>
<td>33 277.2</td>
<td>3.598</td>
</tr>
<tr>
<td>C5H018</td>
<td>17 360.3</td>
<td>3.211</td>
</tr>
<tr>
<td>C5H022</td>
<td>38.0</td>
<td>5.501</td>
</tr>
<tr>
<td>C5H054</td>
<td>687.8</td>
<td>3.659</td>
</tr>
</tbody>
</table>

Table 5 Average catchment slope based on Grid method (Gericke 2010)

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Area ($A$, km²)</th>
<th>Proposed number of grid points ($N_p$, Alexander 1990)</th>
<th>Actual number of grid points used ($N$)</th>
<th>Average slope ($S$, %)</th>
<th>% Difference compared to Neighbourhood method</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5R001</td>
<td>921.6</td>
<td>≥ 50</td>
<td>250</td>
<td>2.072</td>
<td>32.2</td>
</tr>
<tr>
<td>C5R002</td>
<td>10 259.9</td>
<td>≥ 50</td>
<td>3 400</td>
<td>3.060</td>
<td>30.0</td>
</tr>
<tr>
<td>C5R003</td>
<td>936.7</td>
<td>≥ 50</td>
<td>450</td>
<td>4.123</td>
<td>18.3</td>
</tr>
<tr>
<td>C5R004</td>
<td>6 330.9</td>
<td>≥ 50</td>
<td>2 220</td>
<td>2.919</td>
<td>30.3</td>
</tr>
<tr>
<td>C5R005</td>
<td>116.4</td>
<td>≥ 50</td>
<td>50</td>
<td>3.713</td>
<td>32.5</td>
</tr>
<tr>
<td>C5H003</td>
<td>1 650.0</td>
<td>≥ 50</td>
<td>450</td>
<td>4.200</td>
<td>16.7</td>
</tr>
<tr>
<td>C5H012</td>
<td>2 366.3</td>
<td>≥ 50</td>
<td>1 030</td>
<td>3.610</td>
<td>24.3</td>
</tr>
<tr>
<td>C5H015</td>
<td>6 009.0</td>
<td>≥ 50</td>
<td>2 220</td>
<td>2.850</td>
<td>31.9</td>
</tr>
<tr>
<td>C5H016</td>
<td>33 277.2</td>
<td>≥ 50</td>
<td>7 200</td>
<td>2.461</td>
<td>31.6</td>
</tr>
<tr>
<td>C5H018</td>
<td>17 360.3</td>
<td>≥ 50</td>
<td>3 300</td>
<td>2.211</td>
<td>31.1</td>
</tr>
<tr>
<td>C5H022</td>
<td>38.0</td>
<td>≥ 50</td>
<td>50</td>
<td>3.720</td>
<td>32.4</td>
</tr>
<tr>
<td>C5H054</td>
<td>687.8</td>
<td>≥ 50</td>
<td>305</td>
<td>2.479</td>
<td>32.2</td>
</tr>
</tbody>
</table>

Figure 17 Neighbourhood method versus Grid method (Gericke 2010)

$y = 1.07x + 1.01$

$r^2 = 0.88$
This trend was particularly evident for catchment areas exceeding 15 000 km². The Empirical method underestimated the average catchment slope in all the catchments under consideration, except in catchment C5R005, where the average catchment slope result agreed with that of the Neighbourhood method.

\( M : A \) ratios of less than 1 500 resulted in an underestimation of between 20% and 30.2%, while \( M : A \) ratios between 1 700 and 2 750 were associated with underestimations between 18.7% and 0%. Thus, the higher the \( M : A \) ratios, the more accurate Equation 2 becomes. The coefficient of determination \( (r^2) \) of 0.97 is also indicative of a high degree of association.

The visual comparison of results can be highly subjective. Therefore, the data pairs in each catchment under consideration were compared and evaluated using an array of conservation and regression statistics. Values of the \( y \)-intercept (\( a \)), slope (\( b \)), coefficients of efficiency (\( EC \)) and determination (\( r^2 \)), which provide quantitative amplification of the results discussed above, are presented in Table 7.

The conservation statistics percentage differences in Table 7 reflect the differences between the average results as obtained with the Grid and Empirical methods compared respectively to the Neighbourhood method results. In both cases, the objective function (OF) is to minimise these percentage differences, of which the Empirical method’s OF proved to be the minimum, with, on average, the underestimation limited to 14.3%. The \( y \)-intercept (\( a \)) and slope values (\( b \)) of the Grid and Empirical methods showed that these two methods could have different predictive abilities at flat and steep slope classes respectively. In the case of the Grid method, the positive \( y \)-intercept (1.01) is indicative of a possible overestimation of flatter slopes, while the slope value (\( b \)) which slightly exceeded unity (1.07), highlighted that the overestimation of steeper slopes is neither excluded nor impossible. The Empirical method’s positive \( y \)-intercept value (1.58) highlighted that this method is even more likely to overestimate flatter slopes, while the slope value (\( b \)) less than unity (0.74) is associated with the underestimation of steeper slope classes.

### Length and average slope of main watercourses

The main watercourse average slope results based on the Equal-area, 10-85 and Taylor-Schwarz methods are listed in Table 8, while the scatter plots are shown in Figures 19 to 21.
The degree of association between these methods was very high, since the coefficient of determination varied between 0.995 and 0.998. In the past, preference was given to the 10-85 method, since the Equal-area method is largely a graphical procedure and the use of the Taylor-Schwarz method is not widely known in South Africa.

**Distance to catchment centroid**

The results contained in Table 9 are indicative that the length of the watercourse to a position closest to the centroid \(L_{C}:L\) is influenced by the size and shape of the catchment, but more importantly, influenced by the average catchment slope. It is clearly evident from Table 9 that an increase in the average catchment slope is associated with a decrease in the \(L_{C}:L\) ratio, which varied between 0.48 and 0.62.

**CONCLUSIONS AND RECOMMENDATIONS**

**Projections and catchment geometry calculations**

The DEM developed from the SRTM elevation data for southern Africa at 90 metre resolution proved to provide highly accurate raster information which can be used to calculate various catchment parameters (area, length and slope).

**Average catchment slope**

The developed DEM data based on the Neighbourhood method was assumed to be the most accurate representation of the actual average catchment slope and was therefore used as the baseline data to evaluate the Grid and Empirical methods. The Grid method underestimated the average catchment slope in all the catchments under consideration, while the results were indicative that either an increase or decrease in the number of grid points per km² does not necessarily guarantee higher accuracies when compared with the DEM data. The use of at least 50 grid points in catchments up to 10 km² is recommended; thereafter additional grid points at a grid density of 0.5 grid points/km² in catchments exceeding 10 000 km².

The Empirical method also underestimated the average catchment slope in all the catchments under consideration, except in catchment CSR005, where the average catchment slope result agreed with that of the Neighbourhood method. The results were indicative that there is a direct relationship between the area (A) and the total length of all contour lines within the catchment (M).

**Table 8 Average main watercourse slopes (Gericke 2010)**

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Main watercourse length ((L, \text{km}))</th>
<th>Equal-area method</th>
<th>10-85 method</th>
<th>Taylor-Schwarz method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS001</td>
<td>86.44</td>
<td>0.197</td>
<td>0.229</td>
<td>0.225</td>
</tr>
<tr>
<td>CS002</td>
<td>201.69</td>
<td>0.113</td>
<td>0.131</td>
<td>0.108</td>
</tr>
<tr>
<td>CS003</td>
<td>53.80</td>
<td>0.272</td>
<td>0.273</td>
<td>0.266</td>
</tr>
<tr>
<td>CS004</td>
<td>186.70</td>
<td>0.102</td>
<td>0.131</td>
<td>0.113</td>
</tr>
<tr>
<td>CS005</td>
<td>16.20</td>
<td>0.723</td>
<td>0.895</td>
<td>0.819</td>
</tr>
<tr>
<td>CH003</td>
<td>71.18</td>
<td>0.195</td>
<td>0.232</td>
<td>0.195</td>
</tr>
<tr>
<td>CH012</td>
<td>86.96</td>
<td>0.203</td>
<td>0.269</td>
<td>0.222</td>
</tr>
<tr>
<td>CH015</td>
<td>166.95</td>
<td>0.099</td>
<td>0.139</td>
<td>0.103</td>
</tr>
<tr>
<td>CH016</td>
<td>430.72</td>
<td>0.091</td>
<td>0.078</td>
<td>0.081</td>
</tr>
<tr>
<td>CH018</td>
<td>375.39</td>
<td>0.073</td>
<td>0.079</td>
<td>0.075</td>
</tr>
<tr>
<td>CH022</td>
<td>7.91</td>
<td>1.316</td>
<td>1.687</td>
<td>1.493</td>
</tr>
<tr>
<td>CH054</td>
<td>68.04</td>
<td>0.252</td>
<td>0.261</td>
<td>0.283</td>
</tr>
</tbody>
</table>

**Table 9 Catchment centroid distances (Gericke 2010)**

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Main watercourse length ((L, \text{km}))</th>
<th>Centroid distance ((L_{C}, \text{km}))</th>
<th>(L_{C}:L) ratio</th>
<th>Average slope ((S_{A}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS001</td>
<td>86.44</td>
<td>53.18</td>
<td>0.62</td>
<td>3.054</td>
</tr>
<tr>
<td>CS002</td>
<td>201.69</td>
<td>96.72</td>
<td>0.48</td>
<td>4.369</td>
</tr>
<tr>
<td>CS003</td>
<td>53.80</td>
<td>31.11</td>
<td>0.58</td>
<td>5.044</td>
</tr>
<tr>
<td>CS004</td>
<td>186.70</td>
<td>113.02</td>
<td>0.61</td>
<td>4.186</td>
</tr>
<tr>
<td>CS005</td>
<td>16.20</td>
<td>7.90</td>
<td>0.49</td>
<td>5.501</td>
</tr>
<tr>
<td>CH003</td>
<td>71.18</td>
<td>41.18</td>
<td>0.58</td>
<td>5.044</td>
</tr>
<tr>
<td>CH012</td>
<td>86.96</td>
<td>47.62</td>
<td>0.55</td>
<td>4.771</td>
</tr>
<tr>
<td>CH015</td>
<td>166.95</td>
<td>101.06</td>
<td>0.61</td>
<td>4.186</td>
</tr>
<tr>
<td>CH016</td>
<td>430.72</td>
<td>237.14</td>
<td>0.55</td>
<td>3.598</td>
</tr>
<tr>
<td>CH018</td>
<td>375.39</td>
<td>232.99</td>
<td>0.62</td>
<td>3.211</td>
</tr>
<tr>
<td>CH022</td>
<td>7.91</td>
<td>3.86</td>
<td>0.49</td>
<td>5.501</td>
</tr>
<tr>
<td>CH054</td>
<td>68.04</td>
<td>33.05</td>
<td>0.49</td>
<td>3.659</td>
</tr>
</tbody>
</table>

**Figure 19 10-85 method versus Equal-area method (Gericke 2010)**

**Table 8 Average main watercourse slopes (Gericke 2010)**

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Main watercourse length ((L, \text{km}))</th>
<th>Equal-area method</th>
<th>10-85 method</th>
<th>Taylor-Schwarz method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS001</td>
<td>86.44</td>
<td>0.197</td>
<td>0.229</td>
<td>0.225</td>
</tr>
<tr>
<td>CS002</td>
<td>201.69</td>
<td>0.113</td>
<td>0.131</td>
<td>0.108</td>
</tr>
<tr>
<td>CS003</td>
<td>53.80</td>
<td>0.272</td>
<td>0.273</td>
<td>0.266</td>
</tr>
<tr>
<td>CS004</td>
<td>186.70</td>
<td>0.102</td>
<td>0.131</td>
<td>0.113</td>
</tr>
<tr>
<td>CS005</td>
<td>16.20</td>
<td>0.723</td>
<td>0.895</td>
<td>0.819</td>
</tr>
<tr>
<td>CH003</td>
<td>71.18</td>
<td>0.195</td>
<td>0.232</td>
<td>0.195</td>
</tr>
<tr>
<td>CH012</td>
<td>86.96</td>
<td>0.203</td>
<td>0.269</td>
<td>0.222</td>
</tr>
<tr>
<td>CH015</td>
<td>166.95</td>
<td>0.099</td>
<td>0.139</td>
<td>0.103</td>
</tr>
<tr>
<td>CH016</td>
<td>430.72</td>
<td>0.091</td>
<td>0.078</td>
<td>0.081</td>
</tr>
<tr>
<td>CH018</td>
<td>375.39</td>
<td>0.073</td>
<td>0.079</td>
<td>0.075</td>
</tr>
<tr>
<td>CH022</td>
<td>7.91</td>
<td>1.316</td>
<td>1.687</td>
<td>1.493</td>
</tr>
<tr>
<td>CH054</td>
<td>68.04</td>
<td>0.252</td>
<td>0.261</td>
<td>0.283</td>
</tr>
</tbody>
</table>

**Figure 19 10-85 method versus Equal-area method (Gericke 2010)**

**Table 9 Catchment centroid distances (Gericke 2010)**

<table>
<thead>
<tr>
<th>Catchment descriptor</th>
<th>Main watercourse length ((L, \text{km}))</th>
<th>Centroid distance ((L_{C}, \text{km}))</th>
<th>(L_{C}:L) ratio</th>
<th>Average slope ((S_{A}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS001</td>
<td>86.44</td>
<td>53.18</td>
<td>0.62</td>
<td>3.054</td>
</tr>
<tr>
<td>CS002</td>
<td>201.69</td>
<td>96.72</td>
<td>0.48</td>
<td>4.369</td>
</tr>
<tr>
<td>CS003</td>
<td>53.80</td>
<td>31.11</td>
<td>0.58</td>
<td>5.044</td>
</tr>
<tr>
<td>CS004</td>
<td>186.70</td>
<td>113.02</td>
<td>0.61</td>
<td>4.186</td>
</tr>
<tr>
<td>CS005</td>
<td>16.20</td>
<td>7.90</td>
<td>0.49</td>
<td>5.501</td>
</tr>
<tr>
<td>CH003</td>
<td>71.18</td>
<td>41.18</td>
<td>0.58</td>
<td>5.044</td>
</tr>
<tr>
<td>CH012</td>
<td>86.96</td>
<td>47.62</td>
<td>0.55</td>
<td>4.771</td>
</tr>
<tr>
<td>CH015</td>
<td>166.95</td>
<td>101.06</td>
<td>0.61</td>
<td>4.186</td>
</tr>
<tr>
<td>CH016</td>
<td>430.72</td>
<td>237.14</td>
<td>0.55</td>
<td>3.598</td>
</tr>
<tr>
<td>CH018</td>
<td>375.39</td>
<td>232.99</td>
<td>0.62</td>
<td>3.211</td>
</tr>
<tr>
<td>CH022</td>
<td>7.91</td>
<td>3.86</td>
<td>0.49</td>
<td>5.501</td>
</tr>
<tr>
<td>CH054</td>
<td>68.04</td>
<td>33.05</td>
<td>0.49</td>
<td>3.659</td>
</tr>
</tbody>
</table>
The higher the $M : A$ ratios, the more accurate the results calculated using the Empirical method. Both the Grid and Empirical methods demonstrated high degrees of association with the DEM data and can be used along with suitable tools in the ArcGIS™ environment to estimate the average catchment slope. The Grid method is especially useful for the development of slope frequency distribution classes, but the method is sensitive to biased user-input at different scale resolutions, extent of catchment areas and contour intervals used.

On the other hand, the Empirical method in its more rudimentary form (derived from first principles), in conjunction with standard functions in ArcGIS™, proved to be quicker and more accurate, while it is also very suitable for the development of slope frequency distribution classes. The higher accuracy was reflected by the higher $r^2$ value (0.97) and the balance in tendency to either over- and underestimate the flat and steep average catchment slopes respectively. The results conclusively confirmed the preferential use thereof in conjunction with standard tools in the ArcGIS™ environment.

**Average main watercourse slope**

The high degree of association between the Equal-area, 10-85 and Taylor-Schwarz methods proved that any of these methods can be used satisfactorily and with confidence in design flood estimation. However, this high degree of association between these methods does not necessarily guarantee the correctness thereof when used to estimate the time of concentration ($T_C$). In essence, the use of the average main watercourse slope as a suitable predictor variable for $T_C$ estimation can only be justified when compared to $T_C$ estimates based on the temporal distribution of rainfall (observed hyetographs) and runoff (observed hydrographs). In such a case, the validity of the established empirical relationship is also limited to the catchments or regions of original development.

**Distance to catchment centroid**

The average $L_C : L$ ratio of 0.56 obtained from this study is indicative that the general assumption of using a $L_C : L$ ratio of between 0.5 and 0.6 times the distance along the main watercourse is sufficiently accurate in most cases to be used in the various design flood estimation methods (Rademeyer 2012; Van der Spuy 2012). This is also a more definite guideline than the eyeball estimate thereof as proposed by Alexander (2001). However, practitioners are advised to evaluate each catchment individually using the tools available in ArcGIS™, before just using the proposed $L_C : L$ ratios.
REFERENCES


DWAF (Department of Water Affairs and Forestry) 1995. GIS data: Drainage regions of South Africa. Pretoria: DWAF.


Van der Spuy, D, Directorate of Flood Studies, Department of Water Affairs, personal communication, 16 February 2012.


Rademeyer, P F, Directorate of Flood Studies, Department of Water Affairs, personal communication, 16 February 2012.


Van der Spuy, D, Directorate of Flood Studies, Department of Water Affairs, personal communication, 16 February 2012.
Smooth cash flow ensures the effective delivery of projects and is fundamental to develop and sustain a healthy, professional and competitive construction industry. The adverse effect of late or non-payment of contractors and consultants are well known to all in the construction industry. Late and non-payment problems have forced countries like the United Kingdom (UK), Singapore, New Zealand and Australia to introduce legislation to regulate the payment of contractors and consultants in terms of a building or a construction contract. From South African specific surveys conducted by the Construction Industry Development Board (CIDB) and Consulting Engineers South Africa (CESA) it appears that local building and construction contractors and consultants have the same problems as their international counterparts (if not more so) when it comes to payment of work completed or services rendered. In light of the above this article investigates the legal remedies available to enforce right of payment for work completed or services performed, to determine the effectiveness of the said remedies, and to suggest what possible solutions there are in order to improve payment practices in the South African building and construction industry.

INTRODUCTION

The adverse effects of non-payment and/or late payment by employers or contractors and consultants are well known to all in the construction industry. Several related studies have been conducted in developed countries which addressed the problems related to payment issues in the construction industry. Examples are the Latham Report (Latham 1994) and the Egan Report (Egan 1998). Both reports were in response to, inter alia, the problems experienced due to late or non-payment in the construction industry of the United Kingdom (UK). Late and non-payment problems have forced countries like the UK, Singapore, New Zealand and Australia to legislate their construction-specific statutory payment security regime. These legislations purposely enact provisions to address issues on prompt payment in the construction industry to eliminate poor payment practices and to improve the contractor’s cash flow.

From a South African perspective it appears from a Construction Industry Development Board (CIDB) survey, conducted by Marx (2009) and Consulting Engineers South Africa (CESA), that South African contractors and consultants have the same problems as their international counterparts when it comes to payment of work completed or services provided. In the light of the CIDB and CESA findings and the growing international trend to implement construction-specific legislation in order to, inter alia, ensure/facilitate prompt payment practices in the construction industry, it has been decided to address, for this article, the following problem statement:

“What are the legal remedies available to the South African building and civil engineering contractors and consultants to enforce their right of payment for work completed or services performed, and how effective are they in enforcing said right of payment?”

The research for this article was delimited as follows:

The study was limited to selected South African building and civil engineering contractors and consultants. Although extensive use was made of international literature for the literature survey, only local contractors and consultants were interviewed.

The legal remedies to enforce payment in terms of the following CIDB-endorsed forms of contract for construction and building work were researched:

published by the South African Institution of Civil Engineering [GCC 2004].


The legal remedies to enforce payment in terms of the following CIDB-endorsed forms of contract for the provision of professional services were researched:

- CIDB Standard Professional Services Contract [CIDB PSC]

**LITERATURE SURVEY**

**Marx Report (2009)**

Construction Industry Indicators (CIIs) have been developed by the Department of Public Works and the CIDB, with assistance from the Council of Scientific and Industrial Research (CSIR), to play a useful role in developing a sustainable industry and to be adopted as a tool for improving performance in the South African construction industry.

The CIDB CIIs measure the performance of the construction industry by measuring client satisfaction with:

1. the project milestones achieved
2. construction costs versus budget
3. contractors’ performance
4. consultants’ performance
5. the quality of materials used.

The CIDB CIIs have been captured since 2003, and are currently being captured in partnership with the Department of Quantity Surveying and Construction Management of the University of the Free State. A full report was published in March 2009 on the results of the 2008 survey for projects completed in 2007.

Regarding payment delays experienced by contractors for the years 2004 to 2007, the following was reported:

- There was a decrease from 24% to 9% in the number of all projects where payments were made timeously within 14 days, if the 2004 results are compared with the 2007 results (Marx 2009, Tables 15 and 17). In 2007 the private sector clients were the worst early payers, with payments made within 30 days on only 35% of their projects. The best performing client categories with 59% and 56% of project payments made within a month were the public private partnerships and provincial departments respectively. The percentage of projects with payments that took more than 30 days increased from 2004 to 2007 from 43% to 56%. In 2007 the contractors for 20% of all public corporation projects and 21% of all private sector and provincial department projects were only paid after 60 days. There was an encouraging reduction in the percentage of payments done later than 120 days from 13% to 3% if the 2004 and 2007 projects are compared. It is of great concern that only 44% of all contractors in 2007 were paid on time within 30 days (Marx 2009, Tables 15 and 17). With regard to consultants, the following was reported:
  - The consultants’ fees were paid within 30 days for only 45% to 51% of all projects completed between 2004 and 2007 (Marx 2009, Tables 37 and 39). In 2007 the provincial and national departments were the slowest payers of fees, with fees only paid after more than 60 days on 30% and 22% of all their projects respectively. This was followed by the regional/district councils and public private partnership client categories, where the consultants were only paid after three months on 14% of all their projects. On 14% of all public private partnership projects the consultants were only paid four months after submission of fee accounts (Marx 2009, Table 39). The tendency for late payment of consultants has grown if the 2007 results are compared with the 2006 results (Marx 2009, Tables 37 and 39).
  - In 2009 only 52% of all contractors were paid on time, within 30 days, with the metropolitan and regional/district councils being the worst performers (Marx 2011).


In response to the Marx report, the MBSA conducted a survey amongst its members to ascertain the prevalence of delayed or non-payment, as well as the possible causes of the delayed or non-payment. The survey included projects in the civil construction, residential building and non-residential building sectors.

From the draft report issued by the MBSA the findings from the survey can be summarised as follows:

- Of all the projects surveyed, those in the Free State reported the most frequent delays in payments (93%), followed by the Northern Cape (74%). The province reporting the least frequent delays was Gauteng (24%). There were also three provinces in which some projects were listed as “never experiencing delays”. These projects were in Gauteng (43%); KwaZulu-Natal (29%) and Mpumalanga (9%).

- Nationally, across all clients (all projects included in the survey), 54% of projects were paid within 30 days, 26% between 31 and 60 days, 13% between 61 and 90 days, 3% between 91 and 120 days and 4% after 120 days.

- Projects handled by the national government and private sector seem to have the best payment record with 85% and 79% of payments respectively made within 30 days. The remaining 15% of national government projects are paid between 31 and 60 days, while some payments in the private sector are delayed for more than 120 days.

- Payments for local and provincial government projects mostly occur between 31 and 60 days. (45% and 44% respectively), with only 38% and 23% of payments respectively made within 30 days from date of invoice. The remaining 15% of local government projects are only paid between 90 and 120 days, and 2% only after 120 days.

- Sub-contractors are also affected by delayed payments, as only 50% of payments were made within 30 days from date of invoice.

**CESA Report (June 2009)**

An Economic and Capacity Survey is conducted by CESA every six months. The purpose of this survey is to report on the prevailing conditions in the consulting engineering industry. The survey addresses aspects such as financial indicators, human resources, capacity utilisation and competition in tendering and pricing. Questionnaires are distributed to all member firms of CESA.

According to the survey of June 2009, consulting engineers reported a percentage fee income outstanding for 90 days or more of 9.5%. The comparative figure for June 2007 was 10.3%, 11.3% for December 2007, 11.1% for June 2008 and 12% for December 2008 (CESA 2009, Table 15).

Relevant to the employers, the situation is as follows:

- For June 2009, 7.3% of fee claims submitted to central government were outstanding for 90 days or more. For provincial government the figure was 3.8%, local government 13.2%, state-owned enterprises 1.4%, private sector 11.9% and foreign employers 13% (CESA 2009, Table 15).

**Maritz Paper (2007)**

The purpose of this paper was to provide an overview of the development of adjudication
as an alternative dispute resolution process in South Africa and its effectiveness in solving disputes in the local construction industry. The following findings are relevant to this research:

- Of the respondents 63% and 26% respectively agree and strongly agree that "there exists a chronic problem of delayed and non-payment in the South African construction industry affecting the entire delivery chain" (Maritz 2007, Table 1).
- Of the respondents 50% and 13% respectively agree and strongly agree that "allowing all disputed matters to come before adjudication would also reduce payment disputes" (Maritz 2007, Table 2).
- Of the respondents 39% and 30% respectively agree and strongly agree that "South Africa should introduce a Construction Industry Payment and Adjudication Act similar to those in the UK, Australia, New Zealand and Singapore" (Maritz 2007, Table 3).

Maiketso and Maritz Paper (2009)
The purpose of this research was to investigate what the requirements are for the South African construction industry to fully utilise and benefit from adjudication. The researcher, inter alia, reviewed the contractual, institutional and legislative framework for adjudication in South Africa. The following findings are relevant to this paper:

- Of the respondents 75% agreed that "South Africa needs a Payment and Adjudication Act similar to that in the UK". This finding correlates with the Maritz paper as discussed above (Maiketso et al 2009, Table 2).
- Of the respondents 60% agreed that "such legislation should address minimum pay-ment terms, 90% agreed with statutory adjudication, and 95% agreed with remedy in case of non-payment" (Maiketso et al 2009, Table 2).

Common-law position of building and civil engineering contractors
Building and civil engineering contracts are species of the genus locatio conductio operis (letting and hiring of work). Locatio conductio operis is a mutual agreement between one party (the employer) and the other (the contractor), where the contractor undertakes to make his services available with regard to a physical material matter to an employer, for payment. A contractor who accepts work as a result of such a contract is under the obligation to build or repair, as the case may be, for payment, without working under the direct supervision of an employer (Joubert 2003, Vol 13(1) par 113). The contractor is bound to perform the work within the time fixed by the contract of work or within a reasonable time where no time has been specified. When the end product is the erection of a building or a job of work of similar nature, the agreement is commonly described as a building contract, and when it has a significant civil engineering component, it is referred to as a civil engineering contract (Joubert 2003, Vol. 2(1) par 457).

The general principles of the South African law apply to building and construction contracts. In the case of standard construction contracts and where contracts with identical or similar wording have been interpreted by the courts, the courts will consider previous decisions in its judgements.2

In general the following principles apply where a contractor claims for payment for work done in terms of a locatio conductio operis (Harms 1998). The contractor needs to allege and prove:

- The terms of the contract relied upon.
- The work that had to be performed: It is usually an implied term of the contract that the contractor will use materials that are suitable for the purpose of the works.3
- Another implied term of the contract is that the contractor will perform the work in a workmanlike fashion. The level of skill and diligence to be employed is that possessed and exercised by other members of the trade to which the contractor belongs.4
- The remuneration applicable: The contractor must allege and prove (1) that the remuneration was, in terms of the contract, payable, and (2) the amount of the remuneration payable. If the contract is silent with regard to remuneration, remuneration will be payable and should be fair and reasonable (quantum meruit).
- Performance: The contractor must allege and prove that he has done all that was required to be done in terms of the contract.5

Statutory position of building and civil engineering contractors

The CIDB Act 38 of 2000 and its regulations

The CIDB Act 38 of 2000 was passed in October 2000. The Act provided for the establishment of the CIDB to implement an integrated strategy for the reconstruction, growth and development of the construction industry. Further, the Act creates a register of contractors linked to a best practice contractor recognition scheme, and a register of projects linked to a best practice project assessment scheme. Both these registers are central to the implementation of the integrated strategy.

Payment legislation
South Africa does not have construction-specific legislation to address the need for prompt payment of building and civil engineering contractors and consultants. The Public Finance Management Act of 1999 (PFMA) determines that all contractual obligations (and accounts) must be settled within 30 days from its receipt [section 38(1) (f) read with Part 4, Regulation 8.2.3 of the Regulations]. These provisions are mandatory, and an accounting officer of the guilty official may be found guilty of an offence in terms of the PFMA.

In several other countries acts, addressing this need for prompt payment, were endorsed. Acts, and the respective countries and states which enacted them to address the problem of late and non-payment, are:

- Housing Grants, Construction and Regeneration Act 1996 – UK
- Construction Contracts Act 2002 – New Zealand

Remedies to enforce payment in terms of the CIDB-endorsed standard building and construction contracts

Right to interim and final payment certificates

A contractor’s obligation to complete the work is generally indivisible. The mere completion of a specific subdivision of the work does not entitle a contractor for payment of the work done. In the absence of contractual provisions that allow for interim payments, a claim for partially completed work done would be met with the exceptio non adimpleti contractus.6 Only upon completion of the work as a whole would the contractor be entitled to payment.

As a rule the average contractor does not have/command the necessary resources to complete a construction contract before requiring payment for the work completed. In order to provide the contractor with the necessary cash flow to complete the work, most construction contracts provide for the issue of interim payment certificates. In such a certificate the employer’s representative records his reasonable, but only approximate, assessment of the total of work executed and materials supplied up to a given date.

This certificate entitles the contractor to payment of the amount certified within a set...
number of days. Failing payment, the contractor may sue the employer on the strength of the certificate, and the strength of the certificate alone. The claim would be one based on the express terms of the contract. It is not an enrichment claim, even though the amount may be certified as a “reasonable estimate of the total of the work and materials.”

From the comparison made of the four CIDB-endorsed construction contracts, the following general observations were made:

- Payment certificates are certified by independent persons.
- The frequency of interim payment certificates are defined in the contract.
- It is clear when the interim payment certificates should be issued.
- Payment of materials on site is made and only the GCC 2004 does not expressly allow for payment of material off-site.
- It is clear when the interim payment certificates should be paid.
- It is clear when the final payment certificates should be issued.
- It is clear when the final payment certificates should be paid.

Right to interest on late payments
If the employer fails to pay money due under the contract the contractor may elect to charge interest on the amount due. The easiest way to recover interest would be in the case where the contract has express provisions that provide for the payment of interest in specific circumstances at a quantified rate.

From a comparison of the four CIDB-endorsed construction contracts, the following general observations can be made relevant to interest on late payment:

- All of the CIDB-endorsed contract documents provide for the payment of default interest (“finance charges”).
- The circumstances when default interest may be charged are defined.
- The time from when interest accrues is defined.
- The rate of interest chargeable is defined.

Payment guarantee
Relevant to the South African construction industry, a payment guarantee could be defined as a contractual undertaking by a third party (the guarantor) towards the contractor, that the guarantor will pay to the contractor the amount of works done under the construction contract, up to the guaranteed amount or a percentage of the price of the works done, in case the employer defaults in its payment obligations.

Of the four CIDB-endorsed contract documents, only the FIDIC Red and the JBCC PBA contracts expressly provide for the use of payment guarantees. See clause 3.1, JBCC PBA, and the example clause on page 17 of the guidance notes of the FIDICIC Red. Both contracts have pro forma payment guarantee forms that could be used by the parties.

Right to terminate
In the case where the work is only partially complete, the contractor’s claim for interim payment of the partially completed work could be met with a counter claim from the employer based on exceptio non adimpleti contractus. Following from this common law position, a contractor cannot abandon site if the employer fails to pay the contractor for partially completed work.

All four the CIDB-endorsed contract documents contain provisions that allow for the suspension of work and/or the cancellation of the contract in the case of failure by the employer to pay interim payment certificates.

The following aspects should be considered in the case where a contractor wants to leave site or terminate the contract as a result of the employer’s failure to make payment for work completed.

- When the party wishes to enforce a termination clause, the conditions for its implementation have to be strictly complied with.
- In the case where it is required by the contractor to give the employer notice of his intention to terminate the contract as a result of the employer’s failure to make the required payment, the notice to be given should be an express, extra-judicial announcement, and such notice cannot be implied or given by notice of motion.
- In the absence of a contractual termination clause, a contractor will not be able to terminate a contract if an employer fails to make an interim payment. The rationale for this is as follows:
  - In the case of an interim certificate, the contractor has not completed the work in total, and until he has completed the work the contractor has not performed in terms of the contract.
  - If the contractor abandons site, as a result of the non-payment by the employer, the contractor will be in material breach of his obligations to deliver the completed work to the employer.
  - If the contractor terminates the contract, his termination may be held to be a repudiation of the contract, in other words, an indication that he no longer intends to be bound by the terms of the contract, and this would afford the employer the right either to accept such repudiation, bringing the contract to an end, or to refuse to accept the repudiation in which case the contract remains alive and both parties are obliged to continue to honour their obligations to each other. In either event, the employer would be entitled to such damages as he could show he has sustained as a consequence of the repudiation.

Other remedies to enforce payment
Evidence from the employer regarding financial arrangements for the project
The FIDIC Red provides for evidence to be provided by the employer to the contractor whereby, inter alia, the employer proves that it has access to or has the funds necessary to pay the contract price. Clause 2.4: Employer’s Financial Arrangements, reads as follows:

“The Employer shall submit, within 28 days after receiving any request from the Contractor, reasonable evidence that financial arrangements have been made and are being maintained which will enable the Employer to pay the Contract Price (as estimated at that time) in accordance with Clause 14 [Contract Price and Payment]. If the Employer intends to make any material change to his financial arrangements, the Employer shall give notice to the Contractor with detailed particulars.”

The mechanism for the provision of evidence by the employer is technically not a remedy to enforce prompt payment by the employer, but it can certainly be regarded as a mechanism that will assist the contractor to identify, upfront, any possible risks pertaining to the capability of the employer to pay for work completed by the contractor.

Similar provisions could not be found in the JBCC PBA, NEC3 ECC and GCC 2004 documents.

Contractor’s lien
A jus retentionis (right of retention) entitles the holder of that right to retain possession of property until expenditure of money or monies’ worth incurred by him in respect of that property is reimbursed to him.

Relevant to a contractor, the contractor has two kinds of liens at his disposal: enrichment liens or debtor and creditor liens. Where the contractor’s expenditure preserved the property or enhanced its market value the contractor has, to the extent of the true owner’s enrichment, an enrichment lien valid against all comers, including the employer. Otherwise the contractor may rely on the debtor and creditor lien. Commonly this lien is referred to as the contractor’s lien (Finsen 2005). A contractor’s lien is his legal right to retain possession of a construction site until the employer has paid to him monies which are lawfully due to him. The lien is
designed to buttress the contractor’s claim for payment and is not a cause of action in itself, but a course of resistance should the employer demand repossess the premises without tendering payment for the work done on it. A contractor’s *lien* is separate from and does not cover a retention fund.16

**Provisional sentence**

Provisional sentence, as provided for by Rule 8 of the High Court Rules (the Rules), is an extraordinary procedure which is available to a creditor (the plaintiff) who has liquid documentary proof of his claim against his debtor (the defendant).

This procedure is designed to give a plaintiff who is armed with a liquid document, and who accordingly has strong *prima facie* proof of his claim, a speedy provisional judgement without the expense and delay which an ordinary trial action would entail (Erasmus 2007, p. B1-62).

**Summary judgement**

Rule 32 of the Rules is a procedure which enables a plaintiff with a clear case to obtain the swift enforcement of his claim against a defendant who has no real defence to that claim. The courts have stressed the fact that the remedy provided by this rule is an extraordinary and stringent remedy, because it makes inroads into a defendant’s rights to have his case heard and that, if summary judgement is granted, the effect of the order is to close the doors of the court to the defendant. It is therefore only accorded to a plaintiff who has an unanswerable case because the defendant has no defence to it (Erasmus 2007, p. B1-206).

**Order of court**

It is common practice in South Africa to make an arbitration award an order of the court. An arbitration award can be made an order of the court of competent jurisdiction by any party.17 An award that has been made an order of the court can be enforced in the same way as any judgement or order to the same effect. After an award has been made an order of the court, the party enforcing its rights can, for example, issue a writ of execution to be executed by the sheriff of the court.

A contractor or consultant armed with an order of the court, resulting from a successful arbitration award or any other procedure, can enforce the order by applying for the following:

- A finding and order of contempt of court by and committal of the defaulting employer, or
- A writ of execution followed by an attachment of assets of the defaulting employer, and sale thereof.

**QUESTIONNAIRE SURVEY**

**Introduction**

In addition to the literature study for this article, a questionnaire survey was conducted amongst randomly selected consultants and contractors in the South African construction industry. The questionnaire was designed to be brief, concise and straightforward to encourage a high response rate from the potential respondents.

**Population size and response**

Two different sets of questionnaire forms were used in the survey; one for consultants and the other for contractors. The sampling geographic area was limited to level 5 to level 9 contractors registered with the CIDB and all consultants registered at CESA. An e-mail explaining the purpose of the questionnaire, together with the relevant questionnaire, was mailed to all contractors and consultants in the sampling area. In addition, attorneys, advocates and other legal advisors with expert knowledge in the field of construction law and related matters were also contacted and requested to participate in the survey.

Table 1 summarises matters pertaining to population size and response.

**Questionnaire design**

In conjunction with, and under the guidance of, the University of Pretoria’s Department of Statistics, a survey was developed to answer the research problem statement or to test the research hypothesis.

The questionnaire was developed through the following processes to ensure accuracy:

- Reviewing the related academic literature and articles, as well as previous relevant researches to identify pertinent variables to the study
- Drafting the questionnaire based on the identified variables
- Submitting the draft to the study leader and the Department of Statistics for comment and possible recommendations
- Pre-testing the questionnaire to ensure that the questionnaire is understandable to the respondents

**Structure of the questionnaire**

Both the questionnaires for consultants and contractors comprised five distinct sections, as follows:

- **Section 1** established the background information of the respondent.
- **Section 2** established the levels of use and knowledge of the respective CIDB-endorsed contracts. For the consultants’ questionnaire the respondents were requested to rate their knowledge and use of the CIDB PSC and the PROCSA 2009. For the contractors’ questionnaire respondents were requested to rate their knowledge and use of the JBCC PBA, GCC 2004, NEC3 ECC and FIDIC Red.
- **Section 3** was used to rate the sufficiency of the remedies in terms of the CIDB-endorsed contracts. For the consultants’ questionnaire the respondents were requested to rate the perceived sufficiency of certain contractual clauses to ensure prompt payment of professional fee accounts. Clauses pertaining to interim monthly accounts, interest on late payments, written proof of funding available for on-going requirements, payment guarantees, suspension of services and termination of services were rated. For the contractors’ questionnaire clauses pertaining to issues of interim payment certificates, payment of interest on late payments, payment guarantees, suspension and/or termination of work were rated.
- **Section 4** rated the attitude and perceptions of the respondents regarding the effectiveness of litigation in securing payment for professional services and construction work duly executed.
- **Section 5** proposed possible solutions on how to improve current payment practices in the South African construction industry. For both the consultants’ and the contractors’ questionnaires the respondents’ opinions regarding the introduction of statutory prompt payment

<table>
<thead>
<tr>
<th>Sampling group</th>
<th>Total contacted</th>
<th>Successful</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractors’ questionnaires to level 5 – 9 CIDB registered contractors</td>
<td>569</td>
<td>48</td>
<td>8.4%</td>
</tr>
<tr>
<td>Consultants’ questionnaires to CESA registered consultants</td>
<td>274</td>
<td>28</td>
<td>10.2%</td>
</tr>
<tr>
<td>Contractors’ questionnaires to experts in the field of construction law</td>
<td>5</td>
<td>4</td>
<td>80.0%</td>
</tr>
<tr>
<td>Consultants’ questionnaires to experts in the field of construction law</td>
<td>5</td>
<td>4</td>
<td>80.0%</td>
</tr>
</tbody>
</table>
provisions were measured. In addition the respondents were also requested to indicate what the prompt payment process should provide for as a minimum. Considering that there would be a wide range of expected or possible responses, questions that were open-ended were avoided. For most of the questions a 5-point Likert scale ranging from ‘very low sufficiency’ to ‘very high sufficiency’ or ‘strongly agree’ to ‘strongly disagree’ were used. The questionnaire was accompanied by a covering letter which explained the reasons for and background of the research.

Data analysis
Completed questionnaires were collected and submitted to the Department of Statistics at the University of Pretoria. The data was subsequently analysed statistically and a content analysis was employed for qualitative results.

FINDINGS OF THE QUESTIONNAIRE SURVEY
Tables 2 to 9 present a summary of the findings of the questionnaire survey.

CONCLUSIONS AND RECOMMENDATIONS
Selected conclusions from the questionnaire survey conducted
Some of the most relevant trends indicated by the questionnaire survey are:

- Of the consultants and the contractors surveyed 72% and 74% respectively responded that they never or rarely charge interest on late payments.
- Of the consultants and the contractors surveyed 86% and 76% respectively responded that they never or rarely insist on the provision of payment guarantees.
- Both groups of consultants and contractors surveyed regard litigation in South Africa as ineffective in securing payment for professional services and construction work duly executed.
- Of the consultants surveyed 12% disagreed with the statement that statutory prompt payment provisions will improve late payment practices in the South African construction industry.

Table 2 Interest on late payment/finance charges provisions and provision of payment guarantees (consultants)

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>How often do you/your company charge interest on late payment of professional fee accounts.</td>
<td>No</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>36</td>
<td>36</td>
<td>20</td>
</tr>
<tr>
<td>A2</td>
<td>How often do you/does your company insist on the provision of payment guarantees from the client?</td>
<td>No</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>43</td>
<td>43</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 3 Interest on late payment/finance charges provisions and provision of payment guarantees (contractors)

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Rarely</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>How often do you/your company charge interest on late payment of professional fee accounts.</td>
<td>No</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>37</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>A2</td>
<td>How often do you/does your company insist on the provision of payment guarantees from the client?</td>
<td>No</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>38</td>
<td>38</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 4 Attitude and perceptions regarding the effectiveness of litigation (consultants)

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Litigation takes a long time and a successful verdict may often come too late to prevent financial harm to your company.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>A2</td>
<td>Because of the high non-recoverable costs of litigation, a successful verdict may often be a paper victory (a worthless judgement).</td>
<td>No</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>A3</td>
<td>State departments and municipalities often ignore an order of court and therefore a successful verdict together with an order of court may often be a paper victory (a worthless judgement).</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>22</td>
</tr>
<tr>
<td>A4</td>
<td>Once you / your company have/has instituted litigation against a party (including private companies, state departments and municipalities), chances are slim that you will get further work from that party in future.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 5 Attitude and perceptions regarding the effectiveness of litigation (contractors)

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Litigation takes a long time and a successful verdict may often come too late to prevent financial harm to your company.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>A2</td>
<td>Because of the high non-recoverable costs of litigation, a successful verdict may often be a paper victory (a worthless judgement).</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>A3</td>
<td>State departments and municipalities often ignore an order of court and therefore a successful verdict together with an order of court may often be a paper victory (a worthless judgement).</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>58</td>
</tr>
<tr>
<td>A4</td>
<td>Once you / your company have/has instituted litigation against a party (including private companies, state departments and municipalities), chances are slim that you will get further work from that party in future.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>48</td>
</tr>
</tbody>
</table>
Of the contractors surveyed 100% agreed with the statement that statutory prompt payment provisions will improve late payment practices in the South African construction industry.

Of the consultants and the contractors surveyed 68% and 100% respectively agreed with the statement that a commission should be established to investigate errant payments.

Of the consultants surveyed 22% disagreed with the statement that councils or professional bodies for professional consultants in the South African construction industry should be enabled to suspend the licences/memberships of defaulting main consultants (main consultants that do not promptly pay sub-consultants).

Of the contractors surveyed 100% agreed with the statement that the CIDB should be enabled to suspend the registration of defaulting main contractors (main contractors that do not promptly pay sub-contractors).

---

### Table 6 Possible solutions to improve current payment practices in the construction industry (consultants)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Description</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Statutory prompt payment provisions will improve late payment practices in the South African construction industry.</td>
<td>No</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>6</td>
<td>6</td>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>A2</td>
<td>A commission should be established to investigate errant payments.</td>
<td>No</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>14</td>
<td>14</td>
<td>5</td>
<td>46</td>
</tr>
<tr>
<td>A3</td>
<td>Councils/professional bodies for professional consultants in the South African construction industry should be enabled to suspend the licences/memberships of defaulting main consultants (main consultants that do not promptly pay sub-consultants).</td>
<td>No</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>11</td>
<td>11</td>
<td>26</td>
<td>29</td>
</tr>
</tbody>
</table>

### Table 7 Possible solutions to improve current payment practices in the construction industry (contractors)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Description</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Statutory prompt payment provisions will improve late payment practices in the South African construction industry.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>A2</td>
<td>A commission should be established to investigate errant payments.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>A3</td>
<td>The CIDB should be enabled to suspend the registration of defaulting main contractors (main contractors that do not promptly pay sub-consultants).</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
</tbody>
</table>

### Table 8 Possible prompt payment provisions (consultants)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Description</th>
<th>To a very small extent</th>
<th>To a small extent</th>
<th>To an average extent</th>
<th>To a large extent</th>
<th>To a very large extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Statutory adjudication or a similar dispute resolution mechanism to ensure swift dispute resolution of payment disputes.</td>
<td>No</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>7</td>
<td>7</td>
<td>52</td>
<td>26</td>
</tr>
<tr>
<td>A2</td>
<td>A right to regular payment.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>A3</td>
<td>A right to a defined time frame for payment.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>A4</td>
<td>A right to interest on late payments.</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>A5</td>
<td>A restriction of the right to set-off or withhold sums due.</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>A6</td>
<td>Provision for a mechanism that will ensure that a client cannot withhold payment from a consultant unless he has given an effective notice of his intention to withhold such payment.</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>A7</td>
<td>A right to suspend services coupled with the right to reimbursement and additional time as a result of the suspension.</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>A8</td>
<td>Prohibition of &quot;pay-when-paid&quot; clauses.</td>
<td>No</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>24</td>
</tr>
</tbody>
</table>
Recommendations

The following recommendations are proposed:

- A commission should be established to investigate errant payments.
- The South African construction industry should embark on a process of drafting and implementing prompt payment legislation.
- From the questionnaire survey it appears that said legislation should provide for, *inter alia*, the following:
  - Protection of both the contracting and consulting fraternities
  - Statutory adjudication or a similar dispute resolution mechanism to ensure swift dispute resolution of payment disputes
  - A right to regular payment
  - A right to a defined time frame for payment
  - A right to interest on late payments
  - The provision of escrow accounts, or similar trust accounts, to the benefit of the contractor and for retention money retained from the contractor
  - A restriction of the right to set-off or to withhold sums due
  - Provision for a mechanism that will ensure that an employer cannot withhold payment from a contractor unless he has given an effective notice of his intention to withhold such payment
  - Statutory provision for a contractor’s *lien*.
  - A right to allow for stage payments for material in advance of their arrival on the construction site
  - A right to suspend services coupled with the right to reimbursement and additional time as a result of the suspension
  - Prohibition of “pay-when-paid” clauses.

Further research

Some of the findings of this study provide possible directions for further research in the following areas:

- The impacts that late or non-payment may have on sub-contractors and sub-consultants were not investigated. Further research should be conducted to ascertain to what extent sub-contractors and sub-consultants make use of the CIDB-endorsed contract documents. Failure to use the documents could mean that sub-contractors and sub-consultants will not have access to standard contractual remedies available in the case of late or non-payment of professional fees and payment certificates.
- The Consumer Protection Act (CPA) will have a major impact on the South African construction and building industry. For the purposes of this study the impact of the CPA and the extent thereof were not investigated. It is proposed that further research should be conducted in order to determine the impact of the CPA and the extent thereof on the South African construction and building industry.
- For this study a first order comparison was made between existing prompt payment legislation in the UK, Australia, New Zealand and Singapore. More in-depth research should be conducted in order to identify lessons learned from countries that have already implemented prompt payment legislation. The lessons could provide useful guidance to South African legislators if prompt payment legislation is considered.

<table>
<thead>
<tr>
<th>Table 9 Possible prompt payment provisions (contractors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>To a very small extent</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>A1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A6</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A7</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A8</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A9</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A10</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A11</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
NOTES

1 At the time of the study (2009) the GCC 2004 was researched. Since then, the second edition of the GCC was published in 2010 (GCC 2010). Since the clauses pertaining to non- or late payment are similar to the ones in the GCC 2004, it is the opinion of the authors that the findings relevant to the GCC 2004 are also relevant to the GCC 2010.

2 Smith v Mouton 1977 (3) SA 2 at 12

3 Colin v De Guisti 1975 (4) SA 223

4 Randaree NNO v WH Dixon & Associates 1983 (2) SA (1)

5 BK Tooling (Edms) Bpk v Scope Precision Engineering (Edms) Bpk 1979 (1) SA 391

6 Qwa Qwa Regeringsdiens v Martin Harris & Sons OVS 2000 (3) SA 339

7 Simmons v Bantoesake Administrasieraad (Vaalrechtoonkantoor) 1979 1 SA 940 (T)

8 Another cause of action for a claim for outstanding interest would be that the defendant was placed in mora on the date from which the interest is claimed. See Standard Bank of SA Ltd v Lotze 1950 (2) SA 698 (C).

9 Hauman v Nortje 1914 AD 293 at 296.

10 De Wet NO v Lys NO en andere 1998 (4)

11 Shrobree NO Simou 1999 (2) SA 488 (SE). See also clauses 55.1 and 56.1 of the GCC 2004 and clauses 36.3 and 38.2 of the JBCC PBA

12 Australis Estates (Pty) Ltd v Rix 1984 (1) SA 500

13 Goudini Chrome (Pty) Ltd v MCC Contracts (Pty) Ltd 1993 SA 77 (A) 85

14 Brooklyn House Furnishers (Pty) Ltd v Knoetze & Sons 1970 (3) SA 264 (A)

15 Ploughall (Edms) Bpk v Rae 1971 (1) SA 887

16 LIP Construction v Cousins 1985 (1) SA 297 (C) 299

17 Section 31 of the Arbitration Act 42 of 1965

REFERENCES

Consulting Engineers South Africa (CESA), 2009. CESA Biannual Economic and Capacity Survey. Johannesburg: CESA.


A probabilistic approach for modelling deterioration of asphalt surfaces

TF P Henning, D C Roux

This paper details findings from the New Zealand Transport Agency’s research project by Henning and Roux (2008). It forms part of the overall New Zealand Long-term Pavement Performance (LTTP) programme. The paper documents the development of prediction models for dense-graded asphalt (AC) surfaces and open-graded porous asphalt (OGPA) surfaces. Two models were developed including crack initiation on AC surfaces and raveling initiation on OGPA surfaces. Continuous probabilistic models were utilised for both crack and raveling initiation in order to predict the probability of the defect occurring. Models developed during this research use data which is readily available on network level databases, and can therefore be applied to asset management applications such as the New Zealand (NZ) dTIMS system (NZ’s nationally adopted pavement management system (PMS)). Although a crack initiation model was also developed, it was not as robust as the raveling model. Further work required includes refining of the models based on the LTTP data which includes bitumen property data. Although the developments are solely based on NZ data, there are a number of aspects applicable to the South African context. Firstly it presents a novel way of modelling the performance of asphaltic surfaces. Secondly it demonstrates some practical implications of maintenance practices that are sometimes considered for South African conditions.

INTRODUCTION

Background
The entire New Zealand (NZ) road network comprises approximately 120 000 lane km of sealed roads. Although the majority of these roads consist of thin flexible chip-sealed pavements, a large portion of vehicle kilometres travelled occurs on asphalt-surfaced pavements. On average any road carrying more than 10 000 vehicles per day would normally have an asphalt surface constructed on cemented granular pavement layers.

The New Zealand Transport Agency (NZTA) is responsible for funding 100% of the state highways and between 40–60% of the local authority roads, depending on the population size of the local council. Cost-effective management practices of the road networks are not only required by the NZTA, but are also legally required through the Local Government Act 2002 (LGA 2002).

It is realised that PMSs are essential in ensuring cost-effective maintenance planning processes. Pavement deterioration models form an integral part of these PMSs. They provide the predictive capability to forecast future maintenance needs and consequential road conditions. As a result, there has been a strong focus and investment into the NZ pavement model development in order to satisfy advanced asset management requirements specified by the NZ Local Government Act 2002. The NZ LTTP programme has been established primarily to facilitate pavement deterioration model development.

This research in context
Overlaying on existing asphalt layers is a popular maintenance treatment, given that it is more cost-effective and easier to construct compared to full-depth pavement rehabilitation. Overlay asphalt surface layers are expected to perform well, provided that the supporting granular layers ensure sufficient foundation for the asphalt surface. It has been recognised in NZ that there is a need to improve the decision process for the rehabilitation of highways, given the variable performance results from asphalt surface overlays. Based on network analysis, Henning & Roux (2008) demonstrated that asphalt surfaces last for as little as two to four years, suggesting that these layers simply are not cost effective on sub-performing pavement layers. In contrast, where asphalt surface overlays are applied on appropriate pavement conditions, the surfaces last up to 12 to 16 years.

Although international best practice advice is not to overlay open-graded porous asphalt (OGPA) surfaces (FHWA 2005), this is a common practice in NZ. The South
African guidelines do not recommend a specific approach, but emphasise the importance of life-cycle costing considerations during the decision of appropriate rehabilitation options (SABITA 1995). Again, where an overlay treatment is successful, significant savings are experienced in the construction cost, especially in traffic management. However, similar to dense-graded asphalt (AC) surfaces, OGPA surfaces have significant variations in service lives, depending on conditions such as traffic and stiffness of underlying layers.

Given that NZ authorities specify mix properties for asphalt surfaces, the actual properties are not recorded in the asset management system. Decisions on the rehabilitation and maintenance of these surfaces are made based on the condition, traffic and pavement strength data. Therefore, there is a great need to develop performance models that could forecast the behaviour of these surfaces based on the available data. In particular, these models have to be able to take account of the maintenance history, such as the thickness of existing layers.

Objectives of the research

The purpose of this research was to develop pavement deterioration models for asphalt-surfaced pavements in NZ. The models would differentiate between AC surfaces and OGPA surfaces. The models developed as part of this research included:

- Cracking initiation
- Raveling initiation.

Specific goals set for this research included:

- Given that limited data for asphalt surfaces existed on the LTPP programmes, this research had to assess the feasibility for developing the models on the basis of network level PMS data. Earlier work (Henning et al 2009) suggested that some models, especially initiation models, could be developed based on network level PMS data. However, for progression models such as rutting and roughness, more accurate LTPP data is required.
- The models had to reflect network conditions and maintenance practices that significantly influence expected performance.
- The robustness of the models had to be tested to determine whether there is a need for more fundamental or mechanistic asphalt performance models.
- Ultimately, the models have to be adopted within the national PMS in NZ. Therefore, the model format and data requirements had to fit into the framework adopted.

According to the World Bank’s Highway Design and Maintenance model (HDM) definition, crack initiation occurs when a surface displays cracks on more than 0.5% of its area (Watanatada et al 1987). The cracked area is calculated by multiplying the length of the crack by the width of the affected area (for line cracks the affected width is assumed to be 0.5 m). The same definition was adopted for raveling. This paper covers the development of cracking and raveling initiation models. The development of the rutting model has been published elsewhere (Henning et al 2009).

THE STATUS OF PERFORMANCE PREDICTION OF ASPHALT SURFACES

Crack initiation

Crack mechanisms

Cracking, in particular crack initiation, is of concern for road engineers as it is one of the early signs that a road is starting to fail. It is also important to keep the cracks sealed as pavements are significantly damaged by water ingress resulting from the cracks. There are a number of cracking mechanisms, and some of the most commonly found mechanisms on asphalt surfaces include:

- Transverse thermal or visco-elastic fractures, as described by Molenaar (2004)
- Reflective cracking, as described in NIDL (1995)
- Load-associated cracking (AUSTROADS 2004).

For this research, the load-associated cracking was of greatest importance, as it is associated with most of the maintenance costs around asphalt surfaces. Load-associated cracks can manifest in a number of ways, but would normally be located within the wheel path. In addition, they will not be straight and often are referred to as alligator cracks (or crocodile cracks in South Africa (SA)).

According to most mechanistic design methods, failure of an asphalt layer is associated with crack initiation. At this point, the magnitude of tensile strain at the bottom of the asphalt layer is exceeded. Factors that affect the strain conditions at the bottom of the asphalt layer include:

- Stiffness of the asphalt layer
- Thickness of the asphalt layer
- Support from underlying layers, including the subgrade.

Existing modelling approaches for cracking

Three main crack initiation prediction model types are described in the literature review, and these can be categorised according to:

- Deterministic models, which are based on performance data
- Mechanistic design models
- Probabilistic-Mechanistic models.

The most widely used deterministic models are the HDM-III and HDM-4 crack initiation models (Watanatada et al 1987). Separate crack initiation model forms exist in HDM-4 for stabilised and granular bases, and for original/new surfaces and resurfaced surfaces.

The mechanistic models are used for design purposes and predict the overall life expectancy of a pavement, and not necessarily the performance of the surface over time. A typical example of such a model is presented in the AUSTROADS design guide (2004).

Bouwmeester et al (2004) used the traditional risk approach of likelihood and consequence, combined with a stochastic model, to define the risk of cracking in asphalt pavements. The advantage of this approach is that it allows for variability in loading and in the existing strength of the asphalt pavement. A graphic presentation of this approach is given in Figure 1. The graph shows a probability of failure, and does not necessarily reflect initiation. In addition, it is observed that the probabilities are low, with the 50th percentile not reached after an age of 50 years.

![Figure 1 Increased failure probability with age of asphalt pavements (Bouwmeester et al 2004)](image-url)
Ravelling initiation

Ravelling mechanism
Ravelling of OGPA is primarily caused by the failure of the inter-particle bonding provided by the bitumen film. This may result as normal fatigue of the layer due to repetitive vehicle loading or due to premature failure as a result of poor design or construction processes. For example, if the stiffness of the bitumen is too high, it is expected to be brittle and therefore more prone to breaking. Figure 2 illustrates a meso-scale model of OGPA in an attempt to model ravelling by investigating the bitumen bonding between particles. It can be observed from Figure 2(b) that the contact area between particles is relatively small, thus the allowable tensile stresses have to be relatively low. It was concluded that ravelling is caused by a combined effect of binder condition, climate and traffic loading.

Modelling ravelling initiation
Ravelling models developed specifically for OGPA pavements are limited. Three approaches in particular were identified in the literature:

- The World Bank HDM ravelling model is given by NDLI (1995). It is a deterministic model format that provides a generic ravelling model for all surfaces. Each surface type has associated model coefficients that give the relative performances for the surface types.

- Mo et al (2008) developed a complex model to predict ravelling of OGPA surfaces by analysing the stresses of the inter-particle bitumen film (see Figure 2). A finite element analysis model was developed to determine the stresses and strains resulting from loading on the OGPA layer. Resulting from this process, the Von Mises stress in the material was developed for the various loading times (Mo et al 2008).

- In her paper, Miradi (2004) explains a process for modelling OGPA layers using neural networks. A neural network is a self-learning system that considers all factors in predicting an outcome of a variable. These predictions are based on historical outcomes that were caused by a certain contribution of the factors considered. With the introduction of more data, the system learns more about the decision process and the main drivers causing the various outcomes. The main benefit of this system is that it can accurately provide relative weightings and the relative importance of each factor that may influence an outcome of the system. The neural network process, however, does not provide a prediction model that can be used in an external system, with the relation of each factor to the independent variable remaining unclear. Based on this study, the most significant factors influencing ravelling were surface age, asphalt density, bitumen properties and void space.

Conclusions from the review
The main conclusion from the review was that the probabilistic modelling approach considered for NZ (Henning & Roux 2008) was consistent with some of the international trends in deterioration modelling. It is clear that there is a common need to understand risk of failure more than just the condition deterioration over time.

It was also observed that most models, especially for OGPA surfaces, relate to design aspects and data, whereas the intent to develop a model for NZ can be applied to asset management application using in-service data. Studies on OGPA pavements revealed that most factors influencing ravelling are related to the asphalt make-up. It seems that these studies do not include any long-term performance factors such as traffic, or the pavement underlying these layers. Therefore the challenge with this research was to establish the appropriateness of models based on network level data only, or whether more fundamental information would also be required.

<table>
<thead>
<tr>
<th>Structural number (snp_mech)</th>
<th>Annual daily traffic (adt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1.560</td>
</tr>
<tr>
<td>1st Qu</td>
<td>3.290</td>
</tr>
<tr>
<td>Median</td>
<td>3.900</td>
</tr>
<tr>
<td>Mean</td>
<td>4.251</td>
</tr>
<tr>
<td>3rd Qu</td>
<td>4.910</td>
</tr>
<tr>
<td>Max</td>
<td>8.000</td>
</tr>
<tr>
<td>NAs</td>
<td>1 672.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New surface age (hnew)</th>
<th>Total surface thickness (htot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>10.00</td>
</tr>
<tr>
<td>1st Qu</td>
<td>25.00</td>
</tr>
<tr>
<td>Median</td>
<td>30.00</td>
</tr>
<tr>
<td>Mean</td>
<td>35.63</td>
</tr>
<tr>
<td>3rd Qu</td>
<td>35.00</td>
</tr>
<tr>
<td>Max</td>
<td>300.00</td>
</tr>
</tbody>
</table>

RESEARCH DATA
For reasons stated the full data set used in this research consisted of typical network management data. The primary focus was on OGPA surfaced pavements used on the NZTA State Highways in both the Auckland and Wellington regions. Approximately 2 500 sections (data points) were taken from both these networks with typical section lengths being between 200 to 600 m. The network summary is presented in Table 1. Observations from the table include:

- The traffic volume typically ranges between 3 000 (25th percentile) vehicles per day to 16 000 (75th percentile). This is much lower traffic volumes compared to typical traffic volumes on urban motorways in SA.
- The mean modified structural number (SNP) is approximately four, which is typical for a flexible granular pavement with asphalt surfaces. It is, however, much lower compared to the expected SNP for asphalt-base pavements typically used for motorways in SA.

Typical surface configurations on these pavements normally include a dense-graded asphalt surface of between 30 to 50 mm with an open-graded top surface of between 30 to 40 mm. One would therefore expect that the total surface thickness should vary between 60 to 90 mm for sections where no overlay treatment has been applied. With the median thickness being approximately 80 mm, it thus suggested that a substantial portion of the network had been overlaid, especially since the 75th percentile thickness is 100 mm. Note that an isolated number of asphalt-base pavements were excluded from the data set.
Data compilation and validation. Given the significant cost of traffic control on motorways, overlay treatments are used more frequently than mill-and-replace treatments. There is an intuitive view that these surfaces are short-lived, especially when the previous surface was already cracked. Figure 4 confirms this view.

Similar plots were developed for most of the other variables and resulted in the following observations:

- The traffic loading in equivalent standard axles did not have a significant influence on crack initiation time. However, when the traffic loading and structural strength were considered as a combined factor there was a stronger relationship.

- Another factor that showed a potential relationship included the radius of curvature, which is derived from the falling weight deflectometer’s (FWD) deflection bowl.

**Predicting Crack Initiation of AC Surfaces**

**The analysis approach**

The model development consisted of the following main steps:

- Data compilation and validation. Given that these analyses relied on network-level PMS data, there was significant manual data processing and validation involved. For example, maintenance records were compared with performance data to ensure that surface ages and composition items were accurate.

- Statistical exploratory analysis. This was performed in order to gain a better understanding of the over-all data composition, such as distributions and apparent relationships between factors. Although the final model form and factor...
where:

\[ P = \frac{1}{1 + \exp(-a - Bx)} \]  

(1)

\[ P_{CIAC} = \frac{1}{1 + \exp(-0.228 \cdot AGF_2 + kciac \cdot \text{Surfnum} + 3.9 \cdot \text{PCA}_1 - 0.001 \cdot R \cdot \log(\text{ESA}) + 0.003 \cdot \text{PCA}_0 + 0.678 \cdot \text{PCA}_1 + 0.02 \cdot \text{SI rut})} \]  

(2)

The following observations are made from the resulting model and model regression results:

All factors, except (R), had a significance of less than 0.001.

The calibration coefficient kciac is a user-defined constant that allows for varying climatic regions within NZ.

The model was capable of correctly predicting the cracked status of pavements for 62% of cases. Figure 5 suggests that an asphalt overlay on top of an existing cracked surface will not take long before cracking again. On average, new surfaces will take more than ten years before crack initiation. These results correspond well with intuitive ranges as reported from practitioners managing NZ networks.

Figure 6 was developed in an attempt to test the ability of the model to correctly predict crack initiation. Positive means that the predicted and actual status of the surface is cracked. True negative means that both the predicted and actual instances the surface is un-cracked. In total, the model was capable of correctly predicting the cracked status of pavements for 62% of cases. This figure therefore suggests a weak, though acceptable result for the AC surfaces, although it indicates that some improvement would be desirable. It is believed that this improvement would be possible through considering more detailed data on LTPP sections including some bitumen properties. The success rate on other networks was comparable with this result; on some networks the cracked status was predicted correctly for as high as 70% of the cases. It is, however, not as high as the model on chip seals where success rates of up to 85% were achieved.

Table 2 Model regression results for cracking model

| Estimate/ coefficient | Std error | z value (sample variance) | Pr(>|z|) (confidence interval) | Significance |
|-----------------------|-----------|---------------------------|-------------------------------|--------------|
| (Intercept)           | -2.277    | 0.4350                    | -5.230                        | 1.70E - 07   | ***           |
| Hnew                  | 0.008     | 0.0060                    | 1.284                         | 0.19914      | *             |
| factor(PCA)1          | 3.900     | 0.4620                    | 8.431                         | < 2e – 16    | ***           |
| AGE2                  | 0.228     | 0.0160                    | 14.227                        | < 2e – 16    | ***           |
| R                     | 0.001     | 0.0000                    | 2.075                         | 0.03801      | *             |
| factor(PCA)0: Surfnum | -0.003    | 0.1250                    | -0.024                        | 0.98102      | *             |
| factor(PCA)1: Surfnum | -0.678    | 0.0856                    | -7.913                        | 2.51E – 15   | ***           |
| Log(ESA): SI rut      | -0.020    | 0.0070                    | -2.836                        | 0.00457      | **            |

Notes: Significance codes: "***" = 0; "**"= 0.001; "*"= 0.01; "." = 0.05 and ";" = 0.1

Figure 5 Predicting crack initiation for AC surfaces

Figure 6 was developed in an attempt to test the ability of the model to correctly predict crack initiation. The broken line represents the probability of cracking for new pavement surfaces or overlays on un-cracked surfaces. The lighter lines represent the 95th percentile confidence level for both these instances. The cracked status before resurfacing: Cracked Not cracked

Note: Mid-range values from the data set were used to develop the graph, including: Hnew (30 mm), Surfnum (3), ESA (900), R (80), and SI rut (4)
Motivation for developing a ravelling model

Most urban motorways are surfaced with OGPA surfaces in order to reduce noise pollution and improve driving conditions when raining. Ravelling of OGPA surfaces is one of the primary drivers of maintenance decisions on motorway networks (SABITA 1995). Ravelling in itself may start as a small defect, but it is the signal of more serious defects to follow. Secondary defects of ravelling include a higher degree of ravelling, delamination and potholes.

For this reason, the occurrence of severe ravelling on OGPA surfaces would trigger an overlay or mill-and-replace within one year. Without a doubt the modelling of ravelling on OGPA surfaces is essential in the maintenance decision support system for motorway networks.

Note that a crack initiation model was also considered for OGPA surfaces. However, the resulting model was not accepted, given that it forecasted results outside of expected ranges. Closer investigations revealed that the visually rated cracking was inconsistent between years.

Exploratory statistics

Figure 7 illustrates the distribution of ravelling initiation for OGPA surfaces. It is observed that more than half of sections start to ravel before a surface age of five years. It is accepted that ravelling is a strong function of the construction quality. In addition to that, the open matrix of OGPA surfaces poses a high probability of ravelling, regardless of the bitumen properties and construction practices. It is further observed that ravelling initiation has an approximately logarithmic relation with the surface age. This suggests that similar to crack initiation, ravelling can also be modelled using a Logit model format.

The exploratory statistic revealed that the significant factors predicting ravelling initiation included:
- Surface age
- Traffic loading
- Pavement strength
- Cracking.

Figure 8 illustrates the relationship between traffic loading and ravelling initiation for pavements classified into two strength categories, high and low strength, based on FWD filed measurements. It is observed that there is not much difference in ravelling initiation times for the two strength classes. However, it was of concern to observe a positive trend between traffic loading and time to ravelling initiation. This means that roads will ravel faster for lower volume roads compared to higher volume roads. This is opposite to what would be expected.

It has been demonstrated in other studies that asphalt has a ‘self-healing’ characteristic in terms of the way it ravels (Mo et al 2008). For increased traffic volumes, the bitumen would be in constant elasto-plastic motion and would not undergo hardening, which happens with more infrequent loading. For the purpose of this research, it was assumed that the traffic/ravelling initiation trend is valid for the intended model.

Resulting model

A model similar to that for crack initiation modelling was developed for ravelling. The model is given by (refer to Table 3 for the model regression results):
The continuous probabilistic model

The model format provides significant

triggered soon after ravelling initiation.

face to be approximately six years (Henning

the average life expectancy of an OGPA sur-

Auckland motorway network which suggests

are consistent with statistical analysis on the

expected (50th percentile) ravelling initiation

confidence limits. For non-cracked surfaces the

Note: Mid-range values from the data set were used to develop the graph including: H new (30 mm),

ESA (3000), SI rut (4)

Figure 9 Predicting ravelling initiation for OGPA surfaces

APPLICATION OF CONTINUOUS
PROBABILISTIC MODELLING

Earlier model development in NZ high-

lighted a number of benefits of probabilistic

modelling, including (Henning 2009):

■ The continuous probabilistic model

format in a broad sense is more robust in

its approach, because, instead of predict-

ing a discrete point of defect initiation, it

predicts a distribution of the defect

initiation. Sometimes, understanding the

defect initiation distribution itself assists in

understanding network trends. For example, if one assumes that crack

initiation occurs on all surface ages at a
given percentage, and that all surfaces

are cracked at an age of say ten years, a

discrete model may suggest that the aver-
age crack age is at say six years. However, the probabilistic model would suggest that the probability of cracking for ages

more than six years would be higher than

50%, but at a surface age of ten years, all

surfaces could be expected to be cracked.

 Naturally, the latter approach provides

more useful information in a decision

process.

■ The model format provides significant

flexibility to the PMS modelling process

as it gives a probability of failure through-

out the life of the pavement. For example, it is possible to predict different network

level risk profiles as a result of different

maintenance investment strategies. Also,

it is possible to intervene (schedule main-

tenance actions) at different probability

levels for different road classes. In other

words, a different risk appetite can be

adopted for different road classes.

The continuous probabilistic models, in

particular the Logit model, have some limita-

Note: Significance codes: ‘***’ = 0; ‘**’= 0.001; ‘*’= 0.01; ‘.’ = 0.05 and ‘ ‘ = 0.1

Table 3 Model regression results for porous asphalt ravelling model

<p>| Estimate/  | Std error  | z value (sample | Pr(&gt;|z|) (confidence  |</p>
<table>
<thead>
<tr>
<th>coefficient</th>
<th></th>
<th>variance)</th>
<th>interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.801</td>
<td>0.309</td>
<td>-9.062</td>
</tr>
<tr>
<td>age2</td>
<td>0.237</td>
<td>0.009</td>
<td>25.546</td>
</tr>
<tr>
<td>factor(crack_bln)</td>
<td>1.359</td>
<td>0.090</td>
<td>15.122</td>
</tr>
<tr>
<td>SIrut</td>
<td>0.030</td>
<td>0.026</td>
<td>1.131</td>
</tr>
<tr>
<td>log(esa)</td>
<td>0.139</td>
<td>0.044</td>
<td>3.177</td>
</tr>
<tr>
<td>CravOGPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More useful information in a decision

process.

■ The model format provides significant

flexibility to the PMS modelling process

as it gives a probability of failure through-

out the life of the pavement. For example, it is possible to predict different network

level risk profiles as a result of different

maintenance investment strategies. Also,

it is possible to intervene (schedule main-

tenance actions) at different probability

levels for different road classes. In other

words, a different risk appetite can be

adopted for different road classes.

The continuous probabilistic models, in

particular the Logit model, have some limita-

Notes: Significance codes: ‘***’ = 0; ‘**’= 0.001; ‘*’= 0.01; ‘.’ = 0.05 and ‘ ‘ = 0.1

Table 3 Model regression results for porous asphalt ravelling model

<p>| Estimate/  | Std error  | z value (sample | Pr(&gt;|z|) (confidence  |</p>
<table>
<thead>
<tr>
<th>coefficient</th>
<th></th>
<th>variance)</th>
<th>interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.801</td>
<td>0.309</td>
<td>-9.062</td>
</tr>
<tr>
<td>age2</td>
<td>0.237</td>
<td>0.009</td>
<td>25.546</td>
</tr>
<tr>
<td>factor(crack_bln)</td>
<td>1.359</td>
<td>0.090</td>
<td>15.122</td>
</tr>
<tr>
<td>SIrut</td>
<td>0.030</td>
<td>0.026</td>
<td>1.131</td>
</tr>
<tr>
<td>log(esa)</td>
<td>0.139</td>
<td>0.044</td>
<td>3.177</td>
</tr>
<tr>
<td>CravOGPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More useful information in a decision

process.

■ The model format provides significant

flexibility to the PMS modelling process

as it gives a probability of failure through-

out the life of the pavement. For example, it is possible to predict different network

level risk profiles as a result of different

maintenance investment strategies. Also,

it is possible to intervene (schedule main-

tenance actions) at different probability

levels for different road classes. In other

words, a different risk appetite can be

adopted for different road classes.

The continuous probabilistic models, in

particular the Logit model, have some limita-

Notes: Significance codes: ‘***’ = 0; ‘**’= 0.001; ‘*’= 0.01; ‘.’ = 0.05 and ‘ ‘ = 0.1

Table 3 Model regression results for porous asphalt ravelling model

<p>| Estimate/  | Std error  | z value (sample | Pr(&gt;|z|) (confidence  |</p>
<table>
<thead>
<tr>
<th>coefficient</th>
<th></th>
<th>variance)</th>
<th>interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.801</td>
<td>0.309</td>
<td>-9.062</td>
</tr>
<tr>
<td>age2</td>
<td>0.237</td>
<td>0.009</td>
<td>25.546</td>
</tr>
<tr>
<td>factor(crack_bln)</td>
<td>1.359</td>
<td>0.090</td>
<td>15.122</td>
</tr>
<tr>
<td>SIrut</td>
<td>0.030</td>
<td>0.026</td>
<td>1.131</td>
</tr>
<tr>
<td>log(esa)</td>
<td>0.139</td>
<td>0.044</td>
<td>3.177</td>
</tr>
<tr>
<td>CravOGPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More useful information in a decision

process.

■ The model format provides significant

flexibility to the PMS modelling process

as it gives a probability of failure through-

out the life of the pavement. For example, it is possible to predict different network

level risk profiles as a result of different

maintenance investment strategies. Also,

it is possible to intervene (schedule main-

tenance actions) at different probability

levels for different road classes. In other

words, a different risk appetite can be

adopted for different road classes.

The continuous probabilistic models, in

particular the Logit model, have some limita-

Notes: Significance codes: ‘***’ = 0; ‘**’= 0.001; ‘*’= 0.01; ‘.’ = 0.05 and ‘ ‘ = 0.1

Table 3 Model regression results for porous asphalt ravelling model

<p>| Estimate/  | Std error  | z value (sample | Pr(&gt;|z|) (confidence  |</p>
<table>
<thead>
<tr>
<th>coefficient</th>
<th></th>
<th>variance)</th>
<th>interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.801</td>
<td>0.309</td>
<td>-9.062</td>
</tr>
<tr>
<td>age2</td>
<td>0.237</td>
<td>0.009</td>
<td>25.546</td>
</tr>
<tr>
<td>factor(crack_bln)</td>
<td>1.359</td>
<td>0.090</td>
<td>15.122</td>
</tr>
<tr>
<td>SIrut</td>
<td>0.030</td>
<td>0.026</td>
<td>1.131</td>
</tr>
<tr>
<td>log(esa)</td>
<td>0.139</td>
<td>0.044</td>
<td>3.177</td>
</tr>
<tr>
<td>CravOGPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More useful information in a decision

process.

■ The model format provides significant

flexibility to the PMS modelling process

as it gives a probability of failure through-

out the life of the pavement. For example, it is possible to predict different network

level risk profiles as a result of different

maintenance investment strategies. Also,

it is possible to intervene (schedule main-

tenance actions) at different probability

levels for different road classes. In other

words, a different risk appetite can be

adopted for different road classes.

The continuous probabilistic models, in

particular the Logit model, have some limita-

Notes: Significance codes: ‘***’ = 0; ‘**’= 0.001; ‘*’= 0.01; ‘.’ = 0.05 and ‘ ‘ = 0.1

Table 3 Model regression results for porous asphalt ravelling model

<p>| Estimate/  | Std error  | z value (sample | Pr(&gt;|z|) (confidence  |</p>
<table>
<thead>
<tr>
<th>coefficient</th>
<th></th>
<th>variance)</th>
<th>interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.801</td>
<td>0.309</td>
<td>-9.062</td>
</tr>
<tr>
<td>age2</td>
<td>0.237</td>
<td>0.009</td>
<td>25.546</td>
</tr>
<tr>
<td>factor(crack_bln)</td>
<td>1.359</td>
<td>0.090</td>
<td>15.122</td>
</tr>
<tr>
<td>SIrut</td>
<td>0.030</td>
<td>0.026</td>
<td>1.131</td>
</tr>
<tr>
<td>log(esa)</td>
<td>0.139</td>
<td>0.044</td>
<td>3.177</td>
</tr>
<tr>
<td>CravOGPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
normal over-all goodness of fit techniques becomes problematic.

- Practitioners are initially reluctant to accept a probability forecast, given that their frame of reference may still refer to typical time outcomes such as expected life or time to cracking initiation. Referring to a probability seems to be less tangible to them. However, users of the PMSs normally embrace the concept, as it has significant advantages to the traditional descriptive models.

In adopting the initiation models developed through this research, it is recommended that users calibrate these to their respective networks. This calibration involves two parts, including:

- The model has to reflect the behaviour of a given network. The calibration coefficients provided have to be adjusted in order to maximise the model accuracy for a given network. Standard calibration coefficients were developed for NZ.

- The model adoption also has to be calibrated to local practices. For example, in some regions with more moisture-sensitive subgrades, resurfaces may be considered at lower crack initiation probabilities than in other regions with more stable subgrades.

CONCLUSIONS AND RECOMMENDATIONS
The objective of this research was to develop pavement deterioration models for application on asphalt pavements. In addition to that, it was aimed to differentiate between the performance of AC and OGPA surfaces. The research successfully achieved the development of the following models:

- An empirical-probabilistic model that predicts the likelihood of cracking of an AC pavement. The testing of this model suggests that further refinement to this model would be required.

- An empirical-probabilistic model forecasting the probability of ravelling of an OGPA pavement. This model was more robust compared to the crack initiation model, and adoption into the PMS was recommended.

This research also attempted to develop a crack initiation model for OGPA surfaces, but the results were not satisfactory. Visually-rated cracking data on OGPA surfaces was inconsistent between years.

Based on the findings of this research it is recommended that the robustness of the models could be enhanced significantly by including the binder types or classification between modified and un-modified binders. It is recommended further that this information becomes a standard data item for network data sets. Naturally, with the data becoming available, the models need to be reviewed to reflect this change in data.

Given the significant cost of traffic control on motorways, overlay treatments are used more frequently than mill-and-replace treatments. There is an intuitive view that these surfaces are short-lived, especially when the previous surface was already cracked. This research highlighted a number of performance and cost issues related to the overlay practices used on OGPA pavements. Best-practice guidelines should therefore be developed that will ensure life-cycle cost efficiencies from these maintenance practices. For example, there should be some criteria related to the maximum number of OGPA surfaces. There also need to be some criteria on minimum stiffness of the pavement before any OGPA surface is considered.

Most of the developments in this research had to rely on network data since the LTPP data on asphalt pavements was limited. The expansion of the LTPP programme to include more asphalt pavements is essential. Although satisfactory results were obtained from this research, it would be necessary to validate the findings based on accurate long-term performance data. Once more appropriate data becomes available the model needs to be refined, especially a crack initiation model for OGPA surfaces. The literature review has indicated that the inclusion of bitumen/mix properties would greatly enhance the robustness of the models.

Despite the identified limitations, the model tests confirmed that the models could greatly enhance decision processes, and it is recommended that the models be incorporated in the NZTIIMS system. It greatly enhances the appropriateness of the maintenance decision-making based on life-cycle costing considerations.

ACKNOWLEDGEMENT
The authors acknowledge the New Zealand Transport Agency for funding this research project.

REFERENCES


The wave climate on the KwaZulu-Natal coast of South Africa

S Corbella, D D Stretch

The east coast of South Africa has been the subject of numerous coastal developments over recent years. The design of such developments requires a thorough analysis of the local wave climate. Richards Bay and Durban’s Waverider data are two relatively long east coast data sets (18 years). These data sets have not been formally reviewed since Rossouw (1984) analysed existing wave data for South African and Namibian coastal waters. This paper aims to provide a formal analysis of the KwaZulu-Natal wave data.

Seasonal exceedance probability plots, wave roses and typical wave parameter statistics are presented. Return periods for extreme waves are estimated from the generalised extreme value distribution, and the associated limitations are discussed.

The average peak period on the east coast of South Africa is 10.0 seconds, the average significant wave height is 1.65 m and the average wave direction is 130 degrees. Autumn has the most frequent and the largest wave events while summer is the only season unlikely to produce either large or frequent events. The recurrence interval of the largest recorded significant wave height (8.5 m) was estimated to be between 32 and 61 year.

INTRODUCTION

The estimation of statistical return periods (average recurrence interval) of storm events is imperative for coastal managers and design engineers. An average recurrence interval \( T_R \) is the average time (usually expressed in years) between the realisations of two successive events. If the risk of engineering failure due to an event of a specified recurrence interval is not acceptable, it should be redesigned or relocated accordingly. In light of recent developments, from promenade and harbour upgrades to a prospective port and small craft harbour being undertaken in vulnerable coastal zones, the accurate estimation of design waves of specified return periods has become increasingly important.

The KwaZulu-Natal coastline on the east coast of South Africa (Figure 1) experienced its largest recorded wave event in March 2007. The storm coincided with the March equinox (highest astronomical tide of the year) and had devastating effects on the shoreline. Considering coincidence of tide and significant wave height, Theron & Rossouw (2008) (cited by Wright (2009) and Smith et al (2010)) referred to the event as having a 500 year recurrence interval. Phelps et al (2009) found the recurrence interval of the significant wave height to be between 34 and 85 years, but noting that a 35 year occurrence was more likely. CSIR (2008) estimated the significant wave height return period of the storm to be 10 to 35 years, but noted that it was probably closer to a 10 year return period. Apart from the 500 year recurrence interval that considers the coincidence of the tide and storm, the analysis of the significant wave height return period therefore ranges from 10 to 85 years. This wide range further highlights the need for additional research on the characteristics of design waves for the east coast of South Africa.

Once a coastal project has been designed in consideration of a specific return period, the construction or operation of the project becomes the point of focus. Construction and operation of a development often depends on the exceedance statistics of a given wave parameter (see METHODS). Exceedance graphs are a tool used to identify the percentage of time a parameter will be exceeded. Exceedance statistics are not very useful to the design engineer, as the probability of exceedance does not preclude dependent or related recordings of the same event. Therefore this does not yield a recurrence interval estimate of independent storm events. Exceedance graphs are, however, of value during coastal construction work as a management tool. It allows the contractor, resident engineer or project manager to estimate how often work will be disrupted. For example, if a specific height of a cofferdam is installed, exceedance statistics may be used to determine the probable number of days that the temporary works will be overtopped.
The wave climate on the east coast of South Africa has not been formally reviewed since Rossouw (1984) analysed existing wave data for South African and Namibian coastal waters. Rossouw concluded that only the Waverider data (refer METHODS) is reliable enough to consider for design purposes. The relatively long records of data (18 years) making up the current east coast record are from Durban and Richards Bay. Rossouw’s analysis was of a time when no wave recording buoys were operational in Durban. Durban’s reliable data has been analysed by various South African consultants and non-commercial authors (examples include Van der Borch van Verwelde (2004) and Rossouw (2001)). This paper provides a re-analysis and update of the KwaZulu-Natal wave recording data. It also places the analysed data into a formal design reference that is readily accessible.

From a coastal design point of view there was a need to identify what data was available for design applications and how representative it was, since Durban’s record was made up of three different instruments at three different locations. Fortunately Richards Bay has a continuous wave data set from its Waverider buoy that could be used to verify the results.

Storm waves are generated off the KwaZulu-Natal coast by tropical cyclones, cold fronts or cut-off lows. Cold fronts move from west to east and generally exist closer to the coast than cut-off lows and cyclones. Cold fronts occur more regularly than the other forcings and produce relatively smaller wave heights and wave periods with southerly direction. Tropical cyclones are rarely responsible for extreme waves in Durban – between 1962 and 2005 only seven cyclones affected the eastern parts of South Africa (Kruger et al 2010). Generally tropical cyclones produce north-easterly swells. Cut-off lows have been associated with the largest wave events on the KwaZulu-Natal coast (March 2007). They form further offshore than cold fronts and are generally associated with large south-easterly waves with long wave periods. For a detailed description of South African weather conditions the reader is referred to Hunter (1987), Preston-Whyte & Tyson (1993), and Taljaard (1995).

This paper aims (1) to determine the reliability of the Durban and Richards Bay Waverider data, and to use it to establish return periods of wave heights for the east coast of South Africa; (2) to present exceedance statistics of wave heights and peak period and to provide other typical wave statistics; and (3) to analyse wave height return periods by different methods to illustrate the uncertainties and risks of basing designs on a short wave record.

METHODS

The methods of analysis, as well as definitions of the wave parameters considered, are described under METHODS. We then present the exceedance statistics and other typical wave parameter statistics with seasonal variations. A discussion of multivariate return periods is given prior to summarising the conclusions.

Figure 1 Map of South Africa showing KwaZulu-Natal with locations of Waverider buoys and ADCP

The first phase of the analysis was verifying the validity of the available data. Analysis of the wave climate could then be performed with respect to seasonal distributions, exceedance graphs, typical statistics, and a univariate statistical analysis of extreme wave heights.

The wave parameters analysed included the significant wave height, Hs, which, in deep water, is equal to 4√πmo where mo is the area under the wave spectrum; the maximum wave height, Hmax, is the largest wave recorded in a recording period; the peak period, Tp, is the period at which the maximum energy density occurs and is the inverse of the peak energy frequency fp; and the wave direction is the mean wave direction measured from true north.

Hs should be used to model coastal process and shoreline response while Hmax is more appropriate to calculate wave loading on structures. Tp is used to define the surf similarity parameter and is consequently used to quantify wave run-up, scour and forces on structures (the larger the period, the larger the wave run-up and forces on structures). An increase in period has also been shown to increase erosion (Van Gent et al 2008; Van Thiel de Vries et al 2008).

Validity of the wave data

Durban’s 18 years of wave records are a combination of three different wave recording instruments at three different locations (Table 1), two Waverider buoys and an acoustic doppler current profiler (ADCP). Waverider, which is the trade name of Datawell’s wave recording buoy, is a spherical accelerometer buoy that calculates wave heights from accelerations. The ADCP is located on the ocean floor and uses sonar to measure wave heights.

The different locations were a concern because of the shoaling and refraction effects of the different water depths. Diedericks (2009) found that the Richards Bay data has a good correlation with Durban’s data. Diedericks’ findings were verified by finding a Pearson correlation coefficient and a ratio between the Durban Waverider buoy and the Richards Bay Waverider buoy, and between the Durban ADCP and the Richards Bay Waverider buoy (Table 5).

There was still a concern that the ADCP data was not representative enough of deep-water wave conditions and so the recorded waves were classified as either deep water, transitional or shallow water by considering the range of their depth over wave length ratio (Table 2). Newton’s method was used to iteratively solve the wave length L using the peak wave period Tp, depth h and gravitational acceleration g (Equation 1).

$$x_2 = x_1 - \frac{y'(x_1)}{y'(x_1)} = x_1 - \frac{x_1 - D\coth(x_1)}{1 + D\coth^2(x_1-1)}$$

where

$$D = \frac{4\pi^2h}{gT_p^2}$$ and $$x = \frac{2nh}{L}$$ is the wave number

It was decided that since the ADCP did not record the 2007 event, in addition to being in much shallower water than the other instruments, this data set would be replaced by the Richards Bay data which had a strong correlation to the Durban Waverider buoys. The Richards Bay data is a continuous...
Table 1 Historical wave recording instruments, their operating periods and water depth

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Date</th>
<th>Depth (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durban Waverider</td>
<td>1992–2001</td>
<td>42</td>
</tr>
<tr>
<td>Durban ADCP</td>
<td>2002–2006</td>
<td>15</td>
</tr>
<tr>
<td>Durban Waverider</td>
<td>2007–2009</td>
<td>30</td>
</tr>
<tr>
<td>Richards Bay Waverider</td>
<td>1992–2009</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 2 Classification of water waves by the ratio of water depth d to the wave length L (Adapted from U.S. Army Corps of Engineers, 2006).

<table>
<thead>
<tr>
<th>Classification</th>
<th>d/L (m/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep water</td>
<td>1/2 to ∞</td>
</tr>
<tr>
<td>Transitional</td>
<td>1/20 to 1/2</td>
</tr>
<tr>
<td>Shallow water</td>
<td>0 to 1/20</td>
</tr>
</tbody>
</table>

Table 3 Seasonal definition of months

<table>
<thead>
<tr>
<th>Season</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>12 to 2</td>
</tr>
<tr>
<td>Autumn</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Winter</td>
<td>6 to 8</td>
</tr>
<tr>
<td>Spring</td>
<td>9 to 11</td>
</tr>
</tbody>
</table>

Set from a constant location, and so it was also analysed to confirm and compare the Durban results. Unfortunately Richards Bay was not without its limitations and, although it recorded the 2007 event, it did not record the second and third largest events. These events had to be incorporated into the Richards Bay data from the Durban records.

Seasonal distribution of wave parameters

Each data set was analysed independently to establish if there were any inconsistencies or biases. The sets were analysed annually and seasonally. The months were divided into seasons using the meteorological convention seasons. The months were divided into biases. The sets were analysed annually and established if there were any inconsistencies or differences from the Durban data as a result of different local wind conditions.

Exceedance graphs

Supplementing Durban’s data with Richards Bay’s data created an 18 year data set for Durban. Exceedance graphs were created for Hs, Hmax and TP, for each of the four seasons. The exceedance graphs provided an initial idea of event occurrences and allowed an Hs value to be selected for the peak-over-threshold method.

The exceedance graphs were created by binning the parameter in question and then calculating the frequency of occurrence per bin. The frequencies were then used to find the exceedance of events exceeded each bin. The exceedance frequencies were then divided by the total number of data points and expressed as a percentage exceedance. The parameters were plotted against the percentage on a log scale to produce the exceedance graphs. A best fit line was then used to interpret the percentage of time a given wave height is equaled or exceeded.

Wave climate variation and typical statistics

The following parameters were extracted from the data set annually and seasonally: the maximum Hmax, Hs andTp, and the average Hs, Tp, and wave direction. Comparing the parameters seasonally illustrated the degree of seasonal variation.

The average wave direction was calculated, as well as the significant wave height weighted average direction. The results differed negligibly, so only the weighted average directions are presented.

Since minor events had the potential of dampening major events in specific seasons, the analysis of the Hs data was also done only considering events exceeding 3.5 m wave heights.

Univariate statistical analysis of extreme waves

The average recurrence interval or return period of independent wave events can be estimated by fitting a theoretical probability distribution to the data and using it to extrapolate to the event of interest. There are many available probability distributions, and the use of an appropriate one is important to accurately model the data and to realistically estimate the probability of rare events by extrapolation.


<table>
<thead>
<tr>
<th>Probability density functions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
</tr>
<tr>
<td>[ y = \alpha x^{\beta-1} e^{-(\alpha x)^{\beta}} ]</td>
</tr>
<tr>
<td>[ : 0 \leq x &lt; \infty ]</td>
</tr>
<tr>
<td>Extreme value (GEV or Gumbel)</td>
</tr>
<tr>
<td>[ y = \sigma^{-1} e^{\frac{-(x-\mu)}{\sigma}} e^{\left(\frac{-</td>
</tr>
<tr>
<td>[ : -\infty &lt; x &lt; \infty ]</td>
</tr>
<tr>
<td>Lognormal</td>
</tr>
<tr>
<td>[ y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{-(x-\mu)^2}{2\sigma^2}} ]</td>
</tr>
<tr>
<td>[ : 0 &lt; x &lt; \infty ]</td>
</tr>
<tr>
<td>Generalised extreme value</td>
</tr>
<tr>
<td>[ y = \left( \frac{1}{\sigma} \right)^{\gamma} \left[ 1 + \left( \frac{x-\mu}{\sigma} \right) \right]^{\gamma} \left[ 1 + \left( \frac{x-\mu}{\sigma} \right) \right]^{-\gamma - 1} ]</td>
</tr>
<tr>
<td>[ : 1 + k \left( \frac{x-\mu}{\sigma} \right) &gt; 0 ]</td>
</tr>
</tbody>
</table>
### Table 4 The percentage of different water waves recorded by the various recording instruments

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Water Depth (m)</th>
<th>Deep Water Waves (%)</th>
<th>Transition Water Waves (%)</th>
<th>Shallow Water Waves (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durban Waverider (1992–2001)</td>
<td>42</td>
<td>22.7</td>
<td>77.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Durban ADCP (2002–2006)</td>
<td>15</td>
<td>0.2</td>
<td>99.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Durban Waverider (2007–2009)</td>
<td>30</td>
<td>10.1</td>
<td>89.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Richards Bay Waverider (1992–2009)</td>
<td>22</td>
<td>2.2</td>
<td>97.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### Table 5 Pearson correlation, standard deviation and ratio between different instrument-recorded Hs

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Correlation (Pearson)</th>
<th>Average Ratio of Hs</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durban Waverider (1992–2001) vs Richards Bay Waverider (1992–2009)</td>
<td>0.84</td>
<td>1.08</td>
<td>0.25</td>
</tr>
<tr>
<td>Durban ADCP (2002–2006) vs Richards Bay Waverider (1992–2009)</td>
<td>0.77</td>
<td>0.85</td>
<td>0.28</td>
</tr>
</tbody>
</table>

### Figure 2 Comparison of Richards Bay’s Waverider (x) and Durban’s Waverider (●) during May 1998

### Figure 3 Comparison of Richards Bay’s Waverider (x) and Durban’s ADCP (●) during July 2002

There is a large number of fitting methods available, but probably the most popular is the maximum likelihood. The method maximises the probability of observing the data set that has been observed in the sample. This intuitive approach has led to the method being referred to as the most popular and best technique for deriving estimators (Casella & Berger 1990; Montgomery & Runger 2003). The maximum likelihood method is popular with statisticians as its characteristics can be examined mathematically (Goda 2008). It shows a small amount of negative bias, but seems to have the smallest degree of deviation (Goda 2008). The maximum likelihood method is therefore used in this study. The Akaike information criterion (Equation 2) was used to determine the best fitting probability distribution.

\[
AIC = 2k - 2ln(L) \quad (2)
\]

where \( k \) is the number of parameters in the probability distribution and \( L \) is the maximised value of the likelihood function for the estimated parameters.

The length of the wave data record was only 18 years and so it was decided to statistically analyse the \( H_{\text{max}} \) and \( H_s \) wave heights with both the annual maxima method and peak-over-threshold method (POT). The peak-over-threshold method was only applied to the \( H_s \) data for a threshold of 3.5 m. When performing the POT method it is imperative that only independent events are considered. To ensure this, data was divided into events using the following definition: a storm event commences when \( H_s \) exceeds 3.5 m and ends when \( H_s \) falls below 3.5 m, and remains below for approximately one month, based on the decay time of the autocorrelation. The Richards Bay data was similarly analysed.

The 95% confidence intervals were found for the return periods using bootstrapping. Bootstrapping is a resampling technique with replacement. The bootstrapped samples were used to calculate the critical t statistic, which was in turn used to bound the estimated return intervals. For a given value \( \mu \) of a sample there is a probability \((1-\alpha)\) of selecting a sample for which the confidence interval will contain the true value of \( \mu \). The \( 100(1-\alpha) \) percent confidence interval for the t distribution is given by Equation 3.

\[
\frac{\bar{x} - t_{\alpha/2,n-1}s}{\sqrt{n}} \leq \mu \leq \frac{\bar{x} + t_{\alpha/2,n-1}s}{\sqrt{n}} \quad (3)
\]

where \( \bar{x} \) is the mean of the bootstrapped sample, \( s \) is the standard deviation, \( n \) is the number of samples and \( t_{\alpha/2,n-1} \) is the upper 100\( \alpha/2 \) percentage point of the t distribution with \( n-1 \) degrees of freedom.

### RESULTS

The Richards Bay data is shown to be a representative measure of the Durban wave conditions. The two data sets are used in conjunction to establish exceedance probabilities, typical wave parameter statistics,
seasonal trends and average recurrence intervals of wave heights along the east coast of South Africa.

Wave data validity
The wave data showed that the Richards Bay data was an acceptable supplement to the Durban wave data. The waves recorded from all the recording instruments were largely transitional water waves (Table 4). The Durban Waveriders, being in deeper water, recorded the most deep water waves and, although the Richards Bay Waverider data consisted of only 2% deep water waves, it was still ten times larger than the ADCP, making Richards Bay’s recorded waves more similar to that of the Durban Waveriders than the ADCP.

Richards Bay’s Waverider showed a stronger correlation between the Durban Waverider than the ADCP (Table 5). When comparing the average ratios of significant wave heights, the Richards Bay data showed a 1.08 ratio with Durban’s Waverider data, while only a 0.85 with the ADCP data.

The final justification in replacing the ADCP data is shown in Figures 2 and 3. These time series plots of the largest wave events (overlapping the data sets) illustrate that the Richards Bay data is more representative of the Durban Waverider than the Durban ADCP.

Figure 4 shows a comparison of the wave roses for the entire data sets of the Durban Waverider (2007–2009), the Durban ADCP (2002–2006) and the Richards Bay Waverider (1997–2009). The Durban and Richards Bay Waveriders show a similar southerly distribution reaffirming the strong representation of one another. The Durban ADCP has a dominant easterly component and is essentially the result of refraction occurring at the ADCP’s shallow depth.

Exceedance probabilities and wave roses
As previously mentioned, exceedance graphs are not useful in a design application, but are valuable in project planning.

The exceedance graphs are shown seasonally. Figure 5 shows an exceedance graph of significant wave height (Hs) and Figure 6 shows an exceedance graph of maximum wave height (Hmax). Wave direction barely shows a seasonal variation and it is presented as wave roses in Figure 8.

Table 6 Intercepts and slopes of significant wave height exceedance regression lines for summer, autumn, winter and spring, and their associated R² values. The bracketed values show the 95% confidence intervals

<table>
<thead>
<tr>
<th>Season</th>
<th>Intercept</th>
<th>Slope</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>1.21 (1.01; 1.42)</td>
<td>−0.37 (−0.41; −0.33)</td>
<td>0.99</td>
</tr>
<tr>
<td>Autumn</td>
<td>0.82 (0.54; 1.09)</td>
<td>−0.68 (−0.73; −0.64)</td>
<td>0.99</td>
</tr>
<tr>
<td>Winter</td>
<td>1.25 (1.04; 1.46)</td>
<td>−0.45 (−0.49; −0.41)</td>
<td>0.99</td>
</tr>
<tr>
<td>Spring</td>
<td>1.24 (1.01; 1.46)</td>
<td>−0.45 (−0.50; −0.41)</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 5 Significant wave height (Hs) percentage exceedance for summer (■), autumn (▲), winter (×) and spring (+) (refer Table 6 for regression parameters)
Figures 5 and 6 show that autumn experiences the largest waves followed by winter and spring and then summer. Autumn, with regard to wave height exceedance (Hs and H\text{max}), is the only season that shows a significant statistical difference from the other seasons at a 95% confidence limit. Based on the available data, wave heights will exceed the 2007 event (Hs = 8.5 m, H\text{max} = 12.4 m) 0.01% of the time. However, from the regression line the Hs exceedance of 8.5 m is 0.0015% of the time and the H\text{max} exceedance is 0.005%. The event was evidently rare, relative to the data set. Tables 6 and 7 define the regression lines for Hs and H\text{max} respectively.

Figure 7 and Table 8 show that the peak period does not exhibit a statistically significant seasonal variation. The important result is that 90% of the peak periods fall between 10 and 20 seconds.

Figure 8 shows the seasonal wave direction roses for summer, autumn, winter and spring. The dominant wave angle is approximately south-east and is consistent with the south–north littoral drift as expected.

The wave parameters were compared over the entire data set annually and seasonally.

Referring to Figure 9 the highest wave height occurred in 2007. The next highest waves were in 2001. The year 2001 also had the highest average wave height, indicating a particularly rough year in terms of sea conditions. The average Hs for the entire data set was 1.65 m with an average direction of 130 degrees. The maximum T\text{p} occurred in 2008.

Figures 10 to 13 are identical to Figure 9 except that they show the seasonal results as opposed to the entire data set.

Summer’s maximum T\text{p} occurred in 1999 and summer’s largest Hs\text{max} occurred in 2001. Its largest average Hs occurred in 1997. The average Hs for summer is 1.58 m, the average peak period is 9.52 s and the average direction is 135 degrees.

Figure 11 highlights that the largest T\text{p}, H\text{max} and Hs\text{max} of autumn correspond to the 2007 event, while the largest average Hs was significantly higher in 2001 than in the other years. Autumn of 2001 had the second highest Hs\text{max} and the third highest H\text{max}. The average Hs was 1.65 m, the average peak period is 10.4 s and the average wave direction was 132 degrees.

Figure 12 shows that H\text{max}, Hs\text{max} and the maximum average Hs of winter all occurred in 2001. This further enforces the expectation of 2001 being a particularly rough year. The average Hs of winter is 1.64 m, the average peak period is 10.8 s and the average direction is 124 degrees.
The largest $H_{\text{max}}$ and $H_{\text{smax}}$ of spring (Figure 13) occurred in 1993, while the largest average $H_s$ occurred in 1996. The average $H_s$ for spring is 1.72 m, the average peak period is 9.56 s and the average direction is 129 degrees.

The data illustrates that 2001 had particularly rough sea conditions. It also demonstrates that in terms of average $H_s$, $T_p$ and direction there is not much seasonal variation. The above statistics are only those of the combined Durban and Richards Bay data sets.

Seasonal Trends
Seasonal trends, with regard to large wave heights, were identified by considering only the events that exceeded a significant wave height threshold of 3.5 m. Table 9 shows the seasonal percentage of events, the maximum and minimum $H_s$, and the average $H_s$ for the events exceeding a wave height of 3.5 m.

Table 9 shows that autumn has the highest frequency of events, followed by spring and winter and then summer. Summer is definitely the calmest season having the lowest frequency and smallest $H_{\text{smax}}$, $H_{\text{min}}$ and average $H_s$. It is important to note that autumn still experienced the highest $H_s$ of 6.3 m when not considering the 2007 event.

The results show that large events most frequently occur in autumn, as well as the largest events. Winter and spring have very similar events and event occurrences, while summer appears to be the only season unlikely to produce either large or frequent events.

Wave height return periods
For the estimation of average recurrence intervals of independent extreme wave events, Borgman & Resio (1977) suggest that a data set should not be extrapolated to more than three times the extent of the data set.

The results can also vary extensively based on the distribution used, as well as the data selected from the data set. These two limitations were considered by using numerous probability distributions and by applying the annual maxima method, as well as the POT method of sampling. The GEV was determined to be the best-fitting probability density function for all the data sets based on the Akaike information criterion.

Table 10 demonstrates the variations in the different methods. The annual maxima method of both $H_s$ and $H_{\text{max}}$ have the largest return periods, estimated for the 2007 event, of 48 and 61 years respectively. The 95% confidence intervals are a function of...
the number of data points. Since the annual maxima method only uses 18 data points, the confidence intervals are relatively large, ranging between 37 and 60 years for Hs, and 49 and 76 for $H_{\text{max}}$. It should be noted that the $H_{\text{max}}$ values and the Hs values do not always coincide with the same event, evident by the different results.

The POT method yields significantly lower return period estimates and confidence intervals. The Hs POT estimated the event to have a recurrence interval of 32 years, with a 95% confidence interval of 28 to 35 years. The estimates using the Richards Bay data were comparable (Table 10).

The variations in the estimates are indicative of the short data set. The estimates are limited to conclude that the event was between a 32 and 61 year event. This is similar to the 35 to 85 year return period that was determined by Phelps et al (2009). It should be noted that similar wave heights were experienced during Cyclone Imboa in 1984 (prior to the wave record analysed herein). The 23 year period between these major events suggests that the actual return period of the 2007 event is at the lower end of the estimated range. Figures 14 to 16 have been created to allow easy estimation of return periods using any of the two methods, considering the associated uncertainty demonstrated in Table 10.

**DISCUSSION OF MULTIVARIATE RETURN PERIODS**

We have demonstrated that the estimation of average recurrence intervals is dependent on the probability distribution used for estimation and the threshold used to sample wave heights. Apart from the analysis limitations, the estimation of a univariate return period is not a true estimate of the storm risk. The 2007 event’s wave height occurrence was estimated as a 32 year return period, but its coincidence with the highest astronomical tide (HAT) would make the combined event far rarer. Considering two independent events, the probability of both events being exceeded is the product of the exceedance probability of each event. In the case of the 2007 event, coincidence of the HAT (an 18.6 year return period) and wave height (a 32 year return period) yields an average recurrence interval of 595 years.

This extreme return period is actually incorrectly defined, as it assumes that the HAT is a random process which has equal probability of occurrence each year. The HAT is deterministic and the coincidence of a wave height needs be described by the probability of a wave height exceedance for that period of heightened water level. Furthermore the 595 year return period is not a useful measure of risk, since the HAT only exceeds mean high water springs by approximately 30 cm. This demonstrates that the event characteristics should be related to their contribution to the risk of failure. For
example, the same amount of damage may have occurred at any highest astronomical tide of the year for the given wave heights, but would have resulted in a significantly shorter return period estimate.

The estimation of risk becomes more complicated when events are interdependent and requires more advanced statistics. The Gumbel mixed model (Yue et al 1999), the Gumbel logistic model (Yue 2001) and copulas (De Michele et al 2007) are examples of multivariate models that may be appropriate for considering event dependencies in the estimation of return periods. Depending on the requirements of the risk estimation, the multivariate analysis can be extended to include storm duration, wave direction, peak wave period and any other parameters that may contribute to a storm’s damage potential.

**CONCLUSION**

We have re-analysed 18 years of reliable wave data for the KwaZulu-Natal coast and provided a timely update to the existing statistics. Typical statistics of wave parameters are now available without having to re-analyse the integrity of the data sets. The average peak period of the data set is 10.0 seconds, the average significant wave height is 1.65 m and the average wave direction is 130 degrees. Exceedance curves are now available to aid the programming and risk identification for coastal and marine projects. Autumn has been shown to be responsible for the most frequent and the largest amplitude wave events, while winter and spring are similar. Summer is the only season where large events are infrequent.

Five probability distributions have been fitted to the extreme wave events of which the generalised extreme value distribution best modelled the available data. Design waves are now available for coastal projects and the return periods of future events can be quickly estimated. The largest wave event on record occurred in autumn and had an 8.5 m significant wave height, with an estimated return period between 32 and 61 years. Given past records, which have not been considered in the analysis, it is most probable that the average recurrence interval is at the lower end of the range. The Richards Bay return periods were found to be larger, so it is recommended that the more conservative return periods calculated for Durban’s data be used in design.

The 32 year return period of the 2007 event would suggest that it was not very extreme. This return period highlights the limitations of risk analysis when only considering a single variable. Coastal storm damage

**Table 9** Seasonal exceedance and maximum, minimum and average Hs of conditionally sampled significant wave heights using a 3.5 m Hs threshold as the condition

<table>
<thead>
<tr>
<th>Season</th>
<th>Percentage events exceeding an Hs of 3.5 m (%)</th>
<th>Max Hs (m)</th>
<th>Min Hs (m)</th>
<th>Average Hs (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>13.2</td>
<td>4.55</td>
<td>3.52</td>
<td>4.01</td>
</tr>
<tr>
<td>Autumn</td>
<td>30.2</td>
<td>8.50</td>
<td>3.59</td>
<td>4.64</td>
</tr>
<tr>
<td>Winter</td>
<td>28.3</td>
<td>5.47</td>
<td>3.53</td>
<td>4.12</td>
</tr>
<tr>
<td>Spring</td>
<td>28.3</td>
<td>5.64</td>
<td>3.50</td>
<td>4.02</td>
</tr>
</tbody>
</table>

**Table 10** Comparison of the wave height recurrence intervals for the 2007 event. The results of Durban’s data is un-bracketed and Richards Bay’s data is bracketed

<table>
<thead>
<tr>
<th>Method</th>
<th>Distribution</th>
<th>Wave Height (m)</th>
<th>Return Period (years)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hs Annual Maxima</td>
<td>GEV</td>
<td>8.5</td>
<td>48 (58)</td>
<td>37 (47) - 60 (70)</td>
</tr>
<tr>
<td>Hmax Annual Maxima</td>
<td>GEV</td>
<td>12.0</td>
<td>61 (53)</td>
<td>49 (43) - 76 (63)</td>
</tr>
<tr>
<td>Hs POT (Hs&gt;3.5 m, one month)</td>
<td>GEV</td>
<td>8.5</td>
<td>32 (46)</td>
<td>28 (40) - 35 (53)</td>
</tr>
</tbody>
</table>

**Figure 14** Extreme wave height, $H_{max}$, return periods with a 95% confidence interval (---) and the 2007 event (●) for the annual maxima method

**Figure 15** Significant wave height, $H_s$, return periods with a 95% confidence interval (---) and the 2007 event (●) for the annual maxima method
is caused by a combination of high waves, long duration storms, sea levels, and possibly other factors. In order to fully assess the risks from the 2007 event, the probability of the event’s wave heights coinciding with the highest astronomical tide, as well as other characteristics such as the storm duration, should be accounted for.

REFERENCES


Diedericks, H 2009. Sintifc wave periods with a 95% confidence interval (---) and the 2007 event (●) for the peak-over-threshold method. Events defined by one month below the threshold.

Figure 16 Significant wave height, Hs, return periods with a 95% confidence interval (---) and the 2007 event (●) for the peak-over-threshold method. Events defined by one month below the threshold.

Flood frequency analysis.


Van der Borch van Verwolde, E 2004. Characteristics of extreme wave events and the correlation between atmospheric conditions along the South African coast. MSc dissertation, Cape Town: University of Cape Town.


Coastal defences on the KwaZulu-Natal coast of South Africa: a review with particular reference to geotextiles

S Corbella, D D Stretch

Modern coastal defences have to satisfy economic, environmental and sustainability criteria. The balancing of these criteria can make the implementation of coastal defences socially, environmentally and politically complicated. Durban’s local authority, the eThekwini Municipality, has had experience with numerous forms of coastal defences in its attempts to balance the operations of a port and associated beach erosion problems. In March 2007 the KwaZulu-Natal coastline suffered severe damage from an extreme storm event which necessitated the installation of additional coastal defences. This paper evaluates Durban’s experiences of coastal defences, and details the successes and failures to provide practical insight to those faced with similar circumstances or considering the implementation of coastal defences.

INTRODUCTION

The implementation of coastal defences has historically been dependent on the value of the hinterland and the nature of the coastline. In addition, environmental and sustainability considerations have become more prevalent in the implementation of coastal defences and are often the governing factors when determining an appropriate defence (Zanuttigh et al 2005; Airoldi et al 2005; Moschella et al 2005). Consideration of these three factors often leaves municipalities and coastal engineers struggling to find a solution that optimally satisfies all the requirements.

Durban is an important port city located on the east coast of South Africa (Figure 1). The local government was faced with a challenge concerning coastal defences when an extreme event in March 2007 devastated the KwaZulu-Natal coastline. This paper reviews the defences implemented before, during and after the event and identifies their successes and failures. The description of these successes and failures is intended to aid authorities and interested and affected parties to make insightful decisions when undertaking a coastal defence. The review also identifies the importance of monitoring with respect to coastal management. Durban’s beach profiles have been recorded since 1973 and have played an integral part in the municipality’s coastal management.

Gilbert and Vellinga (1990) identified five alternative ways to mitigate the damage of coastal storms, namely accommodation, protection, beach nourishment, retreat and the do-nothing alternative. These solutions can be further divided into two major categories of “hard” and “soft” engineering solutions. Hard solutions typically result in permanent structures that have continual effects on the environment. Soft solutions are the environmentally preferred options and do not involve permanent structures. Durgappa (2008) defined groynes, breakwaters, seawalls, revetment, etc, as hard solutions, and only beach nourishment as a soft solution. We suggest that the soft category should include accommodation, retreat and the use of geotextile sand-filled containers (GSC). Although soft solutions are always a priority they are often difficult to implement effectively. Accommodation is essentially part of a management or planning programme to reduce risk; an example being the establishment of setback lines or designing structures to accommodate occasional flooding. Retreat is the relocation of existing structures to a less vulnerable area, and is often seen as a last resort because of its socio-economic complications (French 2001).

While the eThekwini Municipality (Durban’s local government authority) did implement retreat and the do-nothing approach in the aftermath of the March 2007 event, the experiences from these solutions were trivial and did not warrant inclusion in this review. Minor public-owned structures
were relocated outside of vulnerable zones following their failure, without any public objections. Private properties were not protected with public funds and many of them recovered naturally. Setback lines were successful in mitigating damage during the March 2007 event and are considered an essential part of coastal planning. This review is limited to physical defences and precludes setback lines.

The review aims to compare Durban’s coastal defence experiences with international experiences in an attempt to highlight and recommend successful practices. The review commences with a brief history of Durban’s beach protection and then describes Durban’s coastal defence experiences with respect to: groynes, beach nourishment, loffelstein walls, geotextile sand-filled containers, geotextile tubes and geotextile wraps. The conclusion recommends coastal defences based on the Durban experience.

A BRIEF HISTORY OF DURBAN’S BEACH PROTECTION

Durban’s history of beach protection has largely revolved around efforts to effectively operate a port. There is evidence of the beaches being stable for a period of almost 100 years prior to the commencement of harbour works (Kinmont 1954; CSIR 1976).

In 1857 construction began on the north and south breakwaters of the harbour. It soon became necessary to deepen the channel to cater for vessels with larger draughts, and dredging operations commenced in 1895. Dredging operations intensified from 1897 onwards and this marked the onset of erosion (Kinmont 1954). The channel dredging remained at approximately 650 000 m³/year, which was about equal to the longshore drift estimated at that time, meaning no sand was reaching the beaches north of the harbour.

In response to the beach erosion the eThekwini Municipality started pumping sediment as beach nourishment to the south and central beaches in 1935 (Barnett 1999). What followed were a series of different schemes implemented to discharge dredged material to the beaches. These schemes included a failed fixed bypass scheme attempted between 1950 and 1953.

The so-called Paterson Groynes were built between 1954 and 1956. These two groynes were constructed to stabilise the central beaches in light of the fact that maintenance dredging was not providing the required quantities. The groynes did little to alleviate the problem.

Further effort was made to mitigate beach erosion in 1966 when construction of the underwater mound commenced. It was aimed at protecting the central and northern beaches against storm waves. The mound was never completed to its design height (Barnett 1999), and in 1977 the CSIR found it more beneficial to pump the sand available for the mound directly to the beaches.

The long-standing sediment supply problem was eventually solved by the sand bypass scheme, completed in 1982, and the new groynes were completed in 1985. These groynes, called the Bay of Plenty and North Beach Piers (Figure 2), replaced the Paterson Groynes. In 1989 a third groyne, the Dairy Beach Pier (Figure 2), was constructed (Mather et al 2003). The sand bypass scheme consisted of a concrete hopper and a series of four booster stations, each approximately 700 m apart and connected with a 400 mm diameter high-density polyethylene pipe. The hoppers could receive 5 000 m³ of fluidised sand. This sand would then be re-dug by a fluidising and pumping mechanism at the hopper station and could then be pumped to various outlets between Vetch’s (Figure 1) and Bay of Plenty beach, totalling approximately 3.5 km.

The widening of the Durban harbour in 2007 necessitated the demolition of the hopper station. A new design has been completed and the scheme is due to be operational by 2013. Until such time there is a temporary scheme, consisting of a bund in the northern breakwater and operating similarly to the original scheme.

Table 1 provides an inventory of all the physical defences that have been installed in the eThekwini Municipality, and shows the date of installation, the length of coast defended and comments on the performance of these defences. The length of the defence is only an estimation and does not include ad hoc defences that have been installed by private home owners. The eThekwini coastline is approximately 100 km and it is estimated that 11% of it is defended. Almost 90% of these defences are made up of rock revetments, loffelstein walls and geotextile sand bags. The bulk of the defences are loffelstein walls, which were installed in the Bluff, Umfolozi and Umhlanga Rocks in the 1980s. These retaining walls make up 34% of all eThekwini’s coastal defences. The oldest form of coastal defence is rock revetments. Almost the entire stretch of Durban’s central beaches is protected by rock revetments. The

Figure 1 Map of South Africa showing KwaZulu-Natal and Durban, and a map showing the Durban harbour and the relevant local areas
The rock that was installed in the 1900s is now permanently covered by sand, but makes up 29% of eThekwini’s coastal defences. The first geotextile sand bags were installed in 2007, but they already make up 24% of the eThekwini coastal defences.

### Table 1 Coastal defences along the eThekwini coastline, their installation dates, physical characteristics and encountered issues

<table>
<thead>
<tr>
<th>Defence</th>
<th>Locations</th>
<th>Installation date</th>
<th>Physical characteristics</th>
<th>Encountered issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groynes</td>
<td>Bay of Plenty</td>
<td>1985</td>
<td>214 m long</td>
<td></td>
</tr>
<tr>
<td></td>
<td>North Beach</td>
<td>1985</td>
<td>214 m long</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dairy Beach</td>
<td>1989</td>
<td>214 m long</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In 1998, following a large storm event, the seaward southern piles of the Bay of Plenty Groyne kicked from excessive scour. The structure required repair.</td>
</tr>
<tr>
<td>Beach nourishment</td>
<td>Vetch’s to South Beach</td>
<td>1982</td>
<td>650 000 m³</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>250 000 m³</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>250 000 m³</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Affects 1 000 m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High beach scarp</td>
</tr>
<tr>
<td>Gabion baskets</td>
<td>South Beach</td>
<td>1982</td>
<td>100 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No known issues</td>
</tr>
<tr>
<td>Gabion baskets filled</td>
<td>Ansteys Beach</td>
<td>2007/2008</td>
<td>20 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with 20 kg sand bags</td>
<td></td>
<td></td>
<td>The gabion baskets deformed from the sand bag movement. They look untidy but have performed well.</td>
</tr>
<tr>
<td>Loffestien walls</td>
<td>Brighton Beach</td>
<td>1980s</td>
<td>1 000 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Umhlanga Rocks</td>
<td>1980s</td>
<td>2 000 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Umdloti</td>
<td>1982</td>
<td>850 m</td>
<td></td>
</tr>
<tr>
<td>Fibreglass sheet</td>
<td>Ansteys Beach</td>
<td>2008</td>
<td>70 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>piles</td>
<td></td>
<td></td>
<td>No issues to date</td>
</tr>
<tr>
<td>Geotexile sand bags</td>
<td>Vetch’s Beach</td>
<td>2007</td>
<td>120 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ushaka Beach</td>
<td>2007</td>
<td>300 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>200 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addington Beach</td>
<td>2010</td>
<td>200 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Battery Beach</td>
<td>2010</td>
<td>250 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thekwini Beach</td>
<td>2012</td>
<td>110 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country Club Beach</td>
<td>2012</td>
<td>100 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blue Lagoon</td>
<td>2012</td>
<td>300 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ansteys Beach</td>
<td>2008</td>
<td>100 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Umdloti</td>
<td></td>
<td>300 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>300 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Umhlanga Rocks</td>
<td>2008</td>
<td>300 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total 2 660 m</td>
</tr>
<tr>
<td>Geotexile tubes</td>
<td>Amanzimtoti</td>
<td>2008</td>
<td>50 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No issues to date</td>
</tr>
<tr>
<td>Geotexile wrap</td>
<td>Numerous private homes</td>
<td>2008</td>
<td>500 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inappropriate materials (such as non-woven polyester/bidim) suffer from severe abrasion and pull apart. The wrap was also used as a secondary defence behind the bags at Umdloti and Addington.</td>
</tr>
<tr>
<td>Rock revetments</td>
<td>Umdloti sewer pump station</td>
<td>2007</td>
<td>30 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amanzimtoti</td>
<td>2007</td>
<td>30 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Bluff</td>
<td>1900s</td>
<td>1 200 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central beaches (South Beach to Bay of Plenty Beach)</td>
<td>1900s</td>
<td>2 000 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Social issues but are very effective.</td>
</tr>
<tr>
<td></td>
<td>The mound</td>
<td>Bay of Plenty Beach</td>
<td>1966</td>
<td>1 000 m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Never completed. Waves of sufficient height break on the mound.</td>
</tr>
</tbody>
</table>

**The March 2007 Event**

The March 2007 storm event refers to the storm’s climax on the 19th and 20th of March 2007. Approximately 350 km of the KwaZulu-Natal (KZN) coastline was subjected to severe erosion (Breetzke et al. 2008).

An extreme high tide cycle of 18.6 years and an offshore storm coincided to produce wave heights of up to 8.5 m. The combination of equinox tide level and peak storm-wave setup resulted in a water level of almost 2.7 m above chart datum (Phelps et al. 2009).
What followed were a multitude of coastal defences, many implemented under emergency conditions. The remainder of this paper reviews the successes and failures of defences before and after the 2007 event.

**GROYNES**

Groynes are shoreline stabilisation structures that retard the natural flow of sediment causing accretion. They are constructed perpendicularly to the shoreline and are designed to provide a minimum beach width.

Durban’s groyne field was constructed between 1985 and 1989 (Figure 2). This groyne field, in conjunction with the sand bypass scheme, has been successfully maintaining a stable beach over their existence. The groynes are semi-permeable rock groynes, making the beach width dependent on the rock elevation. The rock levels are monitored annually by the eThekwini Municipality and were adjusted to their design levels in 2009 with a combination of rock and geotextile sand bags.

The piers are constructed on precast friction piles, so the elevation of the sand determines the stability of the structure. The dynamic environment necessitates monitoring, and scour levels around the piles are determined every six months.

Corrosion of reinforcing steel is a major concern for all coastal structures. Although the groynes are still in relatively good condition it is worth noting corrosion observations. When the Bay of Plenty and North Beach Groynes were constructed in 1985 the handrail posts were reinforced with ordinary high-tensile steel bars with a concrete cover of 50 mm (a minimum cover of 25 mm was specified). The Dairy Beach Groyne was constructed four years later in 1989 with handrail posts consisting of hot-dipped galvanised high-tensile steel reinforcing bars, but with only 25 mm concrete cover. By 2010 all of the Dairy Beach handrail posts had to be replaced with polymer concrete posts as a result of severe concrete spalling. The superstructures and piles of all three piers were constructed similarly, but the galvanised Dairy Beach Pier was in a far superior condition than the two older groynes. The fact that the steel in the handrail posts was galvanised, had less concrete cover and was installed more recently implies that concrete cover cannot be neglected because the reinforcing is galvanised. A 20 mm concrete cover is inadequate to protect a galvanised bar under highly corrosive conditions (Yeomans 2004), and SANS 10100-2:1992 specifies 60 mm concrete cover in extreme environments for normal density concrete. The eThekwini Municipality Coastal Department use galvanised steel in all their coastal structures with a minimum cover of 60 mm.

Corrosion of reinforcing steel is a major concern for all coastal structures. Although the groynes are still in relatively good condition it is worth noting corrosion observations. When the Bay of Plenty and North Beach Groynes were constructed in 1985 the handrail posts were reinforced with ordinary high-tensile steel bars with a concrete cover of 50 mm (a minimum cover of 25 mm was specified). The Dairy Beach Groyne was constructed four years later in 1989 with handrail posts consisting of hot-dipped galvanised high-tensile steel reinforcing bars, but with only 25 mm concrete cover. By 2010 all of the Dairy Beach handrail posts had to be replaced with polymer concrete posts as a result of severe concrete spalling. The superstructures and piles of all three piers were constructed similarly, but the galvanised Dairy Beach Pier was in a far superior condition than the two older groynes. The fact that the steel in the handrail posts was galvanised, had less concrete cover and was installed more recently implies that concrete cover cannot be neglected because the reinforcing is galvanised. A 20 mm concrete cover is inadequate to protect a galvanised bar under highly corrosive conditions (Yeomans 2004), and SANS 10100-2:1992 specifies 60 mm concrete cover in extreme environments for normal density concrete. The eThekwini Municipality Coastal Department use galvanised steel in all their coastal structures with a minimum cover of 60 mm.

**BEACH NOURISHMENT**

Beach nourishment is the supply of sand to beaches usually from offshore dredging and is an environmentally preferred method of shore protection (Belkessa et al 2008). This technique is used worldwide usually in combination with a shoreline stabilisation technique.

It accentuated erosion on either side of the groyne, so it had to be removed (Kinmont 1954). A strong rip current is often induced by a groyne, which can be precarious to bathers and requires the presence of lifeguards. Their effective and safe functionality necessitates a fair amount of monitoring, and consequentially requires a good management structure.

The groyne field did not prevent the 2007 event from overtopping the promenade and damaging the adjacent commercial node. The wave heights also exceeded the soffit of the deck and broke numerous precast concrete slabs. With such an extreme event it is not expected that the groynes will completely prevent damage, but they did minimise the impacts by providing a beach buffer between the promenade and the ocean.

In Durban’s case the groynes have been worth the expense and effort, providing not only stable beaches but a recreational attraction to the public. The groynes are currently exceeding their 20 year design life and, other than some minor concrete spalling, are still in a safe operating condition.
Europe has adopted beach nourishment as central to its soft engineering strategy (Hamm et al. 2002; Hanson et al. 2002). The additional sediment on the beach essentially shifts the wave run-up further away from inland infrastructure, creating a buffer.

To protect dunes from erosion the dry beach has to be flat and wide enough to approximate to the Bruun-type equilibrium profile at the raised water level (Dette & Raudkivi 2002). This profile shape, and thus a nourishment volume, are difficult to estimate. The so-called equilibrium profile develops from predominant wave action, so a storm profile will be different from a calm weather profile. This is also true when widening a beach, as the wave conditions may change seaward. A way of avoiding this is to base the nourished profile on a desired historic profile. This in itself is erroneous since a beach that is being nourished is typically eroding and therefore is not at an equilibrium profile. A numerical beach response model is usually used to predict changes in the nourished profile. When attempting to reclaim beach for recreational purposes the inability to predict a fill volume can result in a failed project. The recently acquired dry beach can adjust under wave action and thus a nourishment volume, are difficult to estimate. The so-called equilibrium profile develops from predominant wave action, so a storm profile will be different from a calm weather profile. This is also true when widening a beach, as the wave conditions may change seaward. A way of avoiding this is to base the nourished profile on a desired historic profile. This in itself is erroneous since a beach that is being nourished is typically eroding and therefore is not at an equilibrium profile.

A 900 mm diameter pipe was laid 1.4 km from the north breakwater of the harbour to Addington beach (Figure 3). An offshore borrow site that had previously been used to reclaim Berths D to G within the harbour was used. The site was surveyed and sediments sampled to ensure that the grading was suitable. Sediment grading is important, as fines produce plumes and increase the sediment’s erosion susceptibility. At the same time the sediment cannot be too coarse as traversing the beach can become an uncomfortable barefoot experience. A good way of determining a suitable grading is to compare the grading of a popular and stable recreation beach with that of the borrow site. It was found that the borrow site was slightly coarser (a mean of 304 μm) than the destinations, Vetch’s and Addington Beach (Figure 1). The project was undertaken in two phases, the first in 2009 and the second in 2010, each phase contributing approximately 250 000 m³ of sediment to the beaches.

A project of such a nature is technically trivial, but management intensive. The dredger is chartered on an hourly rate, so the more sand that is pumped the smaller the cubic metre cost. At the time the dredger’s standby rate was R36 000 an hour, necessitating that no delays were incurred at the discharge pipe. The success of the project was largely indicative of good project management.

The project was successful as it introduced new high-quality course sediment (about 500 000 m³) into the system, aiding a correction in the sediment budget. The project was unfortunately not a complete success. The importance of creating a suitable beach profile was neglected in the first phase of pumping, largely due to time constraints. The resulting profile was quickly corrected to an equilibrium profile which produced unexpected earthworks costs to counteract the 3 m scarp that had formed. The borrow site was originally surveyed with a single scan sonar, as a higher-resolution multi-beam was considered an unnecessary expense since the site had been dredged extensively in the past. The initial dredging phase was without incident, except dredging of small ammunitions. The second dredging phase saw the dredging of steel elements (Figure 4) resulting in the cracking of the dredger impeller.

The cracking of the impeller was a major setback in the project and resulted in a large insurance claim. In hindsight an expensive multi-beam survey may have been more economical should the insurance not have covered the delays.
RETTAINING WALLS

Retaining walls in the coastal context are different from seawalls. In Durban a large number of dry-stacking, interlocking retaining walls have been used inappropriately or have developed into an inappropriate situation as a result of chronic erosion. Although there are various types of dry-stacking interlocking walls, Durban’s coastal retention structures are loffelstein walls.

Water-loffel is a variation of loffelstein, having interlocking wings. They are commonly used for hydraulic applications and had been extensively used as seawalls at Brighton Beach, Umhlanga and Umdloti. Seawalls are the most common form of coastal defence and the physical barrier between the land and sea is often considered most desirable by residents (French 2001). Unfortunately they can create a static coast and are one of the least environmentally acceptable solutions. Loffelstein walls are essentially coastal retention structures that are constructed at the backshore and are not intended to withstand direct wave action. Unfortunately, due to chronic erosion, the walls at Umhlanga and Umdloti are exposed to wave attack fairly regularly. Although in these situations a more substantial defence, such as the fibre-glass sheet piles at Ansteys Beach, is preferable, the loffelstein walls have performed relatively well.

A large portion of these walls failed during the 2007 event. The failures were a result of water down-rush, and overtopping washing sediment out from behind the walls. This combination of sediment loss caused the walls to collapse (Figure 5). This failure mechanism highlights the need to have substantial drainage and filtration behind the walls. A geotextile filter layer ensures that water can drain from behind the wall while the filter retains the sediment. Some of the walls that had a geotextile filter parallel to the wall still failed, while none of the walls that had a filter parallel and perpendicular to the wall failed. It is felt that these perpendicular geotextile tiebacks (Figure 6) limit the likelihood of sediment escaping through gaps in the parallel filter. This was the only failure mechanism experienced during the event, as none of the walls were undermined. This was a consequence of the walls being either founded on rock or on a bed of gabion mattresses below the lowest scour profile.

Although loffelstein walls are not a favourable solution, the majority of the loffels that failed during the event were reinstated. Only certain sections of the walls had failed and it was more economical to retain a continuous loffel wall. Where walls were severely damaged they were replaced with geotextile sand bags.

GEOTEXTILE SAND-FILLED CONTAINERS (GSC)

The use of GSCs was initiated in the USA, the Netherlands and in Germany more than 50 years ago (Saathoff et al. 2007). GSCs have become increasingly popular because of their multitude of applications, as well as their environmental benefits. GSCs are often spoken of as a soft engineering solution. This is not entirely correct because a soft solution is one that does not impede the natural morphology of the coast. The GSCs prevent erosion, so can develop a static shoreline. They are considered a soft solution because, if an unforeseen environmental impact ensues, they can easily be sliced open and removed, spilling sand back onto the beach.

Allan & Komar (2002) observed the effectiveness of an artificial dune for shore protection by surveying a dune constructed with sand-filled geotextile bags covered by loose sand and dune vegetation from 1999 to 2002. They reported that the dune survived fairly extreme conditions, which included overtopping, but noted that it was still to be seen if it would cope with the more severe storms.

Heerten et al (2008) did extensive research into the effectiveness of GSCs to mitigate coastal erosion. They described the successful use of GSCs on the island Sylt in Germany where geotextile cushions were covered with sand and sand trap fences. The geotextile was exposed after the second largest storm surge, yet it had prevented a 2.5 m above normal water level and waves exceeding 5 m from eroding the dune.

Recio & Oumeraci (2008) also did extensive research on geotextile bags. Through rigorous model testing they were able to consider all the forces acting on the containers, as well as the effects of container deformation. The impact of wave action and submergence causes sand to be moved inside the bag from the back to the exposed face. This movement
has two negative effects. It decreases the contact area between bags, thus reducing the friction forces, and it increases the surface area in the front of the bag, making it more susceptible to drag forces.

GSCs are a relatively new technology and have only recently found application in South Africa. Their advantages include being cost effective and easily transported, which make them ideal for emergency work. The geotextile can be easily cut and removed if required, but at the same time permanent containers are susceptible to vandalism. This issue has been combatted by a composite vandal-deterrent geotextile which traps 3 kg of sand per square metre within the geotextile. Although this significantly increases the resilience and durability of the container (Saathoff et al 2007) it has little effect on the penetration of a knife. GSCs used to protect dunes should always be covered with sand and vegetated to protect them from vandals and to restore a natural appearance to the coastline. Vegetation has the advantage of mitigating blown sediment and stabilising backshore morphology (Udo et al 2002).

Based on the documented success of GSCs in Australia (Restall et al 2002; Saathoff et al 2007), the municipality decided to pursue their installation as an emergency and permanent measure. The use of GSCs or geotextile sand bags became the eThekwini Municipality’s favoured form of sea defence after the 2007 event. Their extensive use and associated experience warrant an extended section dedicated to their application.

Manufacture

Kaytech Engineered Fabrics was approached to manufacture and supply the GSCs. The bag dimensions were initially 2 x 2.5 x 0.5 m which resulted in a fill-weight of approximately 4 tons. These dimensions were based on a geotextile container of 2.6 x 1.9 x 0.58 m used in Australia (Hornsey et al 2011). After the emergency production of bags had subsided Kaytech refined their manufacture and optimised the bag size to 2.1 x 1.8 x 0.55 m. The bags consist of a double layer, an inner geotextile and an outer UV-stabilised staple filament polypropylene. The two fabrics are bonded together and stitched into a bag. The bag contains two chutes that can be extended from the bag creating a conduit to convey sand into the bag. Since the pioneering of the first bags numerous manufacturers have entered the market.

The filling of the bags required a steel frame (Figure 7). The manufacture of this frame should be governed by the geometry and layout of the bag. This proved to be an issue, as new bag manufacturers entered the market and the manufacture of the bag evolved. When the bag’s chute diameter varied the frame’s funnel diameter remained constant, and this necessitated the removal of some stitching to allow installation of the bag into the frame. This is problematic as the stitching has a tendency to run, potentially causing the bag to pull open.

The stitching of the bags is done in the factory by sewing machines and the seam forms 80% of the bag material strength. The chutes, however, have to be sealed on site and this is done by hand-stitching with nylon string. This stitch is therefore the weakest part of the bag and has been mitigated by placing the bags with the hand-stitched portion facing landwards. Supervision and quality control of the hand-stitching is essential, as labour has a tendency to fluctuate quality, which may lead to the leaking of sand. A manufacturing technique which has proved to successfully increase quality is the pre-punching of holes. This means the spacing of hand-stitches is pre-defined and so ensures more consistency with regard to the quality of stitching. All that still has to be ensured then is that the nylon is knotted correctly. A handheld sewing machine was initially used to stitch the bags, but was abandoned during the emergency work following associated installation delays. The reintroduction of the sewing machine has not been supported, due to cost implications, as well as there being no present evidence of the adverse effects of hand-stitching.

Installation

The bags were originally filled to 80% of their capacity (based on the German construction technique (Oumeraci et al 2003)), ensuring that the sand is sufficiently compacted by flooding with water. If the bags are filled any more it becomes difficult to stitch them closed, jeopardising the quality of the stitch. Overfilling the bags also causes rounding. Since some of the bag’s stability is determined by its mass and friction, it was hypothesised that the more rounded it is the less contact each bag would have with the surrounding bags, lowering the stability. Not filling the bags to capacity allows them to be levelled for the next bag layer, as well as providing a large contact area. It must be noted that the findings of Hornsey et al (2011) contradict this theory, showing that the Australian practice of filling bags to capacity is more stable.

In certain circumstances there may be uncertainty as to where the lowest scour level is or additional confidence is required in minimising the undermining risk. This was accommodated by providing a Dutch toe (self-healing toe). A Dutch toe is a row of bags in front of the wall’s toe and tied back into the bottom bags. The theory is that, as the beach profile approaches the founding level of the bag wall, the Dutch toe will settle giving the structure an additional 2 m (length of one bag) scour resistance.

Slope

The slope of the bag-protected dune is still debated, with engineers designing slopes from 30° to 45°. The stability of the bag wall is dependent on the friction forces that develop between the bags, which is a function of their roughness, the net normal force (weight above the contact area) and the contact area. Recio & Oumeraci (2008) identified, from flume tests, that the friction between containers affects the hydraulic
stability much more than assumed in past and present literature. In order for the bags to be stable this frictional force must be equal to or greater than the active soil force behind the bags (Figure 8). Therefore, although the friction increases as the bag wall approaches the vertical, so does the active soil pressure. A balance of these two forces was used to calculate an optimum bag slope. Figure 9 shows the number of stable bags through a range of slopes. The calculation includes a safety factor used on the soil properties. For comparativeness, results shown ignore cohesion, place the water table below the bag wall, and use a constant soil weight of 18 kN/m³ and a friction angle of 30°. Active pressures were calculated using Coulomb’s active earth pressure coefficient. The bag dimensions and weight are as previously stated.

Figure 9 illustrates that the angle range of 18° – 26° can retain the most soil – the flatter the slope of the wall, the less the wave loading, but the greater the wave run-up. This simple calculation does not consider the hydraulic stability of the bags, and because the sand is displaced to the bags’ front face by the lifting and dropping of the exposed portion, it is thought that the less bag length that can be lifted the more stable the bags would be. It is felt that the increased restriction of this movement is what makes the bags more stable at steeper slopes (Hornsey et al 2011) and not the increased friction.

The bags are prone to vandalism and degradation by ultraviolet radiation. They also look untidy once people have traversed them and they have been subject to wave action. The bags should therefore be covered with sand and vegetated. Since beach sand generally cannot maintain a slope greater than 30° it also dictates that the bags should follow a similar slope. The ground conditions, space restrictions and retained height all influence the bag wall slope, and therefore a standard orientation cannot be specified. Each situation should be considered independently.

**Geotextile sand bag performance**

The bags endured their first substantial test on 26 July 2011 from a significant wave height of 5 m, and all the issues identified in the literature were realised. Vandalised bags leaked sand and created weak spots in the walls. The lower layers of bags shifted forward making the lower wall face steeper. This appears to be the combination of three factors: the bags not being filled sufficiently, the bags leaking sand and the bags’ geotextile elongation. All these factors enable the sand to move to the bag’s exposed face, lowering the friction forces and increasing the drag forces. In an extreme case a bag was completely removed from the lower portion of a wall (Figure 10). We propose that our local bags need some refinement in terms of elongation and that stringent quality controls are required during installation. The significance of bag deformation is perhaps more evidence that the Australian method of filling the bags to capacity and using a 45° slope is more appropriate.

**GEOTEXTILE TUBES**

Geotextile tubes have been used all over the world and have been particularly successful in the construction of artificial reefs. The geotextile tube was experimented with, as it was potentially faster, cheaper and more
structurally sound than using geotextile sand bags. The theory was that the 1.4 m diameter by 25 m tube could be laid in position and then pumped full of sand (Figure 11). This, however, was not as simple in practice.

Pumping slurry into the tube caused air to be trapped in the tube which had limited venting points. The slurry was also pumped at a ratio of about 30% sand to 70% water. The geotextile drained slowly, causing the tube to fill with water and air faster than it could expel them, resulting in the slurry discharge pipe being forced out the tube inlet. This issue was overcome by having scuba divers inside the tube directing the pump discharge. The tube also had to be braced every 5 m.

The tube was successfully installed and has been in place for almost four years without any issues. Structurally the tube is more stable than individual bags, as it is continuous, weighs more and only has one piece of hand stitching. The tube has the added advantage of being able to be placed and filled in the water. With all the complications associated with the tube it ended up costing twice as much as installing the geotextile sand bags. Although it was discontinued due to its difficult installation and associated costs, its use has been successful in other applications (Cantré 2002; Shin & Oh 2007; Alvarez et al. 2007). We propose that similar success is possible if more appropriate equipment is used.

**GEOTEXTILE WRAP**

The geotextile wrap was used as an alternative to the geotextile sand bags where access was limited, and has proved to be a reliable alternative (Yasuhara & Recio-Molina 2007).

A 5.3 m by 25 m geotextile fabric was used to create an in-situ sand bag or tube. The geotextile is laid flat, half is topped with sand and the other half is then folded over and stitched on the landward side. This method allows all the work to be done by hand. If the wrap will not be exposed (always to be covered by sand or by geotextile bags) bidim may be used. If the bidim will be exposed to sunlight and wave action it should be replaced with 1200 g/m² ultraviolet-treated (UV) geocontainer fabric, as the bidim is not UV-protected and pulls apart under wave attack. The long continuous hand-stitching is the main weakness of the wraps, but also makes the installation cost effective and has become increasingly popular amongst private homeowners along the KwaZulu-Natal coast.

**CONCLUSION**

Durban has had a long history of beach protection and some of its recent experiences have been shared in this review. From these experiences it is recommended that soft solutions, primarily as a combination of coastal setback lines, beach nourishment and GSCs be prioritised. Admittedly soft solutions are not always practical or appropriate.

Durban’s groyne field has been a valuable investment aiding in successfully stabilising the central beaches since their construction.
Experience has shown that concrete cover should be seriously considered as corrosion mitigation, even when providing galvanised reinforcing. It is recommended that all structures, including seawalls and GSCs, be founded on rock or at a depth that ensures structural stability when the lowest historical scour level is exceeded.

Beach nourishment and geotextile sand bag seawalls are the eThekwini Municipality’s preferred soft protection to be implemented in conjunction with coastal setback lines. A successful beach nourishment project can be executed by substantial preliminary research of the borrow site, as well as good project management. For a geotextile sand bag defence it is recommended that the bags are filled to capacity and installed at a slope of 45°, covered in sand, and vegetated. Although geotextile wraps have proven their reliability as coastal protection (Yasuhara & Recio-Molina 2007) the municipality’s experiences have only found them to be an appropriate substitute for the bags in severely restricted areas. In areas of high vulnerability it is recommended that a bidim wrap be installed and draped with bidim prior to overlaying it with one or two layers of geotextile sand bags. The bidim wrap acts as a secondary defence against extreme events, as well as a substantial filter layer, which has proved to be significant. A Dutch toe should also be installed as additional risk mitigation of undermining. The bags’ elongation at breaking point still needs to be refined and parity has to be achieved on the filling percentage of the bags’ capacity.

Many situations require more robust solutions than geotextile sand bags or beach nourishment. Such situations need to be individually accessed, but rock revetments and sheet-piled seawalls have been successfully installed in Durban.

A good monitoring system is essential for successful coastal management. The beach profiles recorded in Durban since 1973 have been instrumental in the design of all its coastal defences from seawalls to beach nourishment. To the authors’ best knowledge the record is the most extensive in South Africa and is the core of the eThekwini Municipality’s coastal management and defences.

The March 2007 event gave the eThekwini Municipality the opportunity to be innovative. It is hoped that other organisations will be able to convert the failures into successes and use the success stories to improve the sustainability of defending our coast.

REFERENCES


Adjudication as an alternative dispute resolution method in the South African construction industry

N C Maiketso, M J Mantz

Adjudication has recently been introduced to the South African construction industry as an alternative dispute resolution mechanism. This study investigates what the requirements are for the industry to realize the full potential of adjudication. To this end the study reviews the necessary contractual, institutional and legislative framework, discusses relevant skills and available training, and establishes what impact all these have on the current practice of adjudication.

A literature review was conducted, covering the local and international practice of adjudication. A structured interview was conducted with adjudicators, and those who were out of geographic reach were sent a survey questionnaire. The results obtained were statistically analysed.

Adjudication appears to have found acceptance in the South African construction industry, but it was found that the industry is not yet able to realize the full potential of adjudication, the main reason for this being a lack of knowledge.

FREQUENTLY USED ABBREVIATIONS
ADR Alternative Dispute Resolution
JBCC Joint Building Contracts Committee
CIDB Construction Industry Development Board (SA)
GCC General Conditions of Contract (SAICE)
FIDIC Federation Internationale des Ingenieurs-Conseils
NEC New Engineering Contract (ICE)

INTRODUCTION
Adjudication has recently been introduced into the four CIDB-endorsed forms of contract (JBCC, GCC, FIDIC and NEC) as the standard method of dispute resolution. As with almost everywhere else, the South African (SA) construction industry is more familiar with earlier forms of dispute resolution, namely mediation, arbitration and litigation. Adjudication is a relatively new concept and is not well understood. It also faces challenges in application, as most adjudicators are trained and/or experienced in these other forms of dispute resolution and not in adjudication per se. Those meant to be served by it, i.e. clients, consultants and contractors, also have limited understanding of the process or how best to make use of it.

The purpose of the paper is to investigate what the requirements are for the construction industry to fully utilise adjudication. To facilitate this, the research reviews the necessary contractual, institutional and legislative framework and other enabling factors, discusses relevant skills and available training, assesses whether or not these are in place in the SA construction industry, and establishes what impact the whole situation has on the current practice of adjudication. Recommendations are then made based on the findings.

Problem statement
What are the requirements for the SA construction industry to fully utilise adjudication?

The main problem was elaborated through the following sub-problems:

• How does the SA construction industry understand adjudication, how is it distinguished from other forms of dispute resolution, and what makes it attractive?
• Is adjudication adequately provided for in the contractual, institutional and legislative framework?
• Are there enough adjudicators in South Africa? Is there an established set of skills for adjudicators, and is relevant training available on adjudication?
• What impact does the state of affairs established above have on the realization of the full potential of adjudication in SA, and what can be done about it?

Hypothesis
The SA construction industry does not realize the full potential of adjudication because it is not sufficiently understood nor appropriately practised.
This was also broken down further into corresponding sub-hypotheses as follows:

- Adjudication is not well understood, and in practice it is not sufficiently distinguishable from the other forms of dispute resolution. It is, however, attractive because it is seen as quick and cheap.
- Adjudication does not enjoy sufficient institutional support, as there is neither legislation nor a voluntary association for adjudication. The four CIDB-endorsed forms of contract now all make provision for adjudication, but they can all be improved.
- There are not enough adjudicators in SA. There is no established set of skills for adjudicators, as there is neither regulation nor organisation for the practice of adjudication. There are therefore no universally accepted minimum training or skills requirements.
- The impact of the status above (as established through the findings of the research) on the realization of the full potential of adjudication is negative. Recommendations are made based on the findings.

### LITERATURE REVIEW

#### Definition
The term “adjudicate” is found in general usage to mean “to give a ruling” or “to judge”. In more recent times, a specialised use of the term “adjudication” appears as a form of alternative dispute resolution (ADR) available to the construction industry. Its definition in this context is not universally agreed, it being more often defined by what it is not than by what it is, but the following characteristics are reflected by most definitions (after CIDB 2004):

- Object is to reach a fair, rapid and inexpensive decision.
- Adjudicator is to act impartially and in accordance with rules of natural justice.
- Adjudication is neither arbitration nor expert determination, but adjudicator may rely on own expertise.
- Adjudicator’s decision is immediately binding (finality is dependent on whether it is challenged within the allotted time, in which case finality may be reached through arbitration, litigation or by agreement).

#### Origins
Differing views have been expressed regarding the origins of adjudication in construction (Gould 2006), but it is a commonly held view that its primary aim was to secure timely payment, having recognised that one of the most notorious inefficiencies of the construction industry is non- or late payment of contractors/sub-contractors by employers/contractors respectively (see for example Maritz 2007). This is possibly why adjudication is so closely associated with legislation of the form “Security of Payment”, and why it has been characterised by the adage “pay now, argue later” (Uff 2005).

An earlier form of adjudication was used in the United Kingdom (UK) in the 1970s, focusing on the payment problem between contractor and sub-contractor. In the United States of America, dissatisfaction with rising costs of arbitration and litigation in the construction industry led to the appearance of dispute boards in the 1960s, and this started to take root in the 1970s (Gaitskell 2005). Of perhaps greater significance is the recent questioning of the quasi-judicial role of the principal agent. One of the principles of natural justice – that one cannot be judge in one’s own cause – appears to have played a major role in this latter development, and this also features prominently in adjudication.

In their 1999 white paper to the Minister of Public Works, the CIDB recommended the use of ADR, as arbitration and litigation were seen as costly and time-consuming (CIDB PGC3 2005). The Latham report (UK 1996) is referred to as a point of departure. The CIDB went further and made it mandatory for the SA construction industry to adopt adjudication before referring disputes to arbitration or litigation (CIDB PGC3 2005).

<table>
<thead>
<tr>
<th>Year</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>ICE reviews contracting strategies</td>
</tr>
<tr>
<td>1993</td>
<td>ICE issues NEC (engineer separated from adjudicator)</td>
</tr>
<tr>
<td>1994</td>
<td>Latham Report issued (adjudication by contract and legislation)</td>
</tr>
<tr>
<td>1995</td>
<td>World Bank adopts DB in its procurement guidelines</td>
</tr>
<tr>
<td>1996</td>
<td>FIDIC introduces Dispute Adjudication Board (DAB) in orange book</td>
</tr>
<tr>
<td>1998</td>
<td>Part II of UK Housing Grants, Construction and Regeneration Act 1996 becomes effective (including mandatory adjudication provisions)</td>
</tr>
<tr>
<td>1999</td>
<td>FIDIC introduces DAB into new suite: all books (no longer optional)</td>
</tr>
<tr>
<td>2004</td>
<td>GCC (SA) introduces adjudication into 2004 edition</td>
</tr>
<tr>
<td></td>
<td>CIDB issues adjudication procedure and recommends use of GCC, JBCC, FIDIC, NEC</td>
</tr>
</tbody>
</table>

#### Adjudication within ADR
The rise in the modern use of ADR procedures appears to be due to the following factors (Uff 2005; Butler & Finsen 1993), which to a large degree used to be claimed for arbitration as its strong points in the past (in comparison to litigation):

- expertise of facilitator
- lower cost and shorter duration
- convenience and flexibility
- privacy and informality
- voluntary or customised dispute resolution process (can be made mandatory by agreement/contract).

Butler & Finsen (1993) observed that arbitration had become more formal and legalistic, and expressed the hope that the advent of ADR would rekindle arbitration and provide it with appropriate techniques to sustain its use. More than ten years later Uff (2005) observed that positive developments like the “100-day arbitration procedure” had grown out of the lessons learned from adjudication.

Many authors, however, view all dispute resolution methods as constituting a continuum or spectrum, with each method having its rightful place (see for example M’khomazi & Talukhaba 2004). Indeed, for enforceability if nothing else, ADR has had to form an alliance with the formal court system (Maritz 2007).
Table 2 Three tiers of application of adjudication (After Maiketso 2008)

<table>
<thead>
<tr>
<th>Tier of application</th>
<th>Elements reviewed</th>
<th>Summary findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forms of contract</td>
<td>JBCC 2005, GCC 2004, FIDIC ‘99 (‘red book’) NEC 3 (‘black book’)</td>
<td>Adjudicator’s (or DB’s) appointment: by the parties, otherwise by a named authority</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjudicator’s conduct: impartial, independent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inquisitorial: can ascertain the facts and the law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjudicator not liable and not called as witness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dispute scope: anything under contract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decision: immediately binding</td>
</tr>
<tr>
<td>Institutional guidelines</td>
<td>JBCC, CIB, DRBF, AAA, ICC, World Bank, CUB*</td>
<td>More detail than forms of contract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Procedural and administrative aspects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Funding institutions may prescribe</td>
</tr>
<tr>
<td>Legislation</td>
<td>UK, New Zealand, Queensland (Australia) Singapore</td>
<td>Conditional payment clauses outlawed (e.g. pay-when-paid)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Establishing minimum payment terms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Establishing statutory adjudication system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remedies available for non-payment</td>
</tr>
</tbody>
</table>

* AAA – American Arbitration Association; DRBF – Dispute Resolution Board Foundation; ICC – International Chamber of Commerce; CUB – Construction Umbrella Bodies (UK)

Table 3 Sampling summary

<table>
<thead>
<tr>
<th>Sampling group</th>
<th>Total contacted</th>
<th>Successful</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview – adjudicators</td>
<td>30</td>
<td>18</td>
<td>60%</td>
</tr>
<tr>
<td>Completing questionnaire – adjudicators</td>
<td>17</td>
<td>6</td>
<td>35%</td>
</tr>
<tr>
<td>Completing questionnaire – general sample</td>
<td>9</td>
<td>5</td>
<td>56%</td>
</tr>
<tr>
<td>Totals</td>
<td>56</td>
<td>29</td>
<td>52%</td>
</tr>
</tbody>
</table>

Adjudication in practice
The practice of adjudication was reviewed through its three tiers of application, namely standard forms of contract, institutional guidelines and legislation. See Table 2.

Level of use and knowledge
The work of the Adjudication Reporting Centre (Kennedy 2005) appears to represent best practice in monitoring the use of adjudication. The centre issues regular reports based on information obtained from adjudicator nominating bodies in the UK. The reports include:
- number and discipline of adjudicators
- trends in adjudications (growth, decline, fluctuations)
- performance of adjudication (dissatisfaction or otherwise).

Generally, this reporting shows adjudication to be successful.

Various levels of acceptance and use of adjudication in all its various forms have been recorded from elsewhere. Dispute boards (DBs) continue to grow in use in the form of Dispute Review Boards, Dispute Adjudication Boards or Combined Boards (DRBF 2007). The World Bank, along with other development banks, is playing a significant role in this aspect, more recently with the help of FIDIC harmonised conditions of contract (MDB). Povey’s research (2005), whilst focusing on mediation, also revealed that SA mediators tended to conduct themselves more like the modern adjudicator. Van Langelaar (2001) confirms that the international trends discussed above apply to southern Africa, including the observation that the DB role was not always understood or agreed between project participants. Van Langelaar (2001) further notes that, although the system appeared to have been successful, the knowledge base needs to be expanded.

Skills and techniques
A comparison was drawn between information on adjudication skills and training from selected institutions, namely the CIDB, Institution of Civil Engineers (ICE), Chartered Institute of Arbitrators (CIArb), DRBF, American Arbitration Association (AAA) and FIDIC. The following major findings emerged:
- Formal training is common, varying from workshops to formal tuition and assignments.
- Formal assessment and accreditation are also common, including examinations and peer reviews, used in different formats and to varying degrees of intensity.
- Continuing Professional Development (CPD) as an on-going requirement has become universal.

Thus the right mix has to be found which would be suitable for SA conditions. Whilst one does not necessarily want to “kill it with too much science”, there could be legitimate cause for concern that sub-standard levels of skill may not do justice to adjudication, or be able to exploit its full potential for the benefit of the construction industry.

RESEARCH METHODOLOGY

Population size and sampling
Due to limited numbers of people with knowledge of the subject, purposive or target sampling was adopted. Panels of dispute resolution practitioners were sourced from relevant organisations (Association of Arbitrators Southern Africa (AASA), Consulting Engineers South Africa (CESA), South African Institution of Civil Engineering (SAICE), and NEC Users Group), within which adjudicators were targeted. See Table 3 for summary.

Research design
The research design adopted was generally quantitative, but made provision for qualitative data in the form of comment. A survey questionnaire was developed and administered to answer the sub-problems or test the sub-hypotheses. The questionnaire design and administration incorporated considerations of threats to validity and research ethics.

The questionnaire was divided into the following categories (about 25 questions): 1. Adjudicator background 2. Level of use and knowledge
3. Forms of contract, institutional guidelines, legislation
4. Skills and techniques
5. Impact
The data was analysed statistically, and content analysis was employed for qualitative results.

RESULTS
Graphs 1, 2 and 3 were selected for illustrative purposes from question groups 2, 3 and 4 above, pertaining respectively to distinguishing features, sufficiency of adjudication provisions and useful techniques when using adjudication. A short summary is presented below. More detailed results are presented in Appendix 1.

Summary of graphs
- From Graph 1 the respondents agreed that the most distinguishing feature of adjudication was the speed within which the process is concluded.
- Graph 2 illustrates that respondents agreed that the four forms of contract were sufficient in their provisions for adjudication, with FIDIC scoring the highest.
- From Graph 3 the respondents consider the “inquisitorial” approach to be the most useful technique when conducting an adjudication.

FINDINGS
The results appear to reveal the following on the research problem:
- The first sub-hypothesis was disproved as far as adjudication practitioners are concerned – their understanding appears to be quite high, and is in keeping with generally accepted characteristics of adjudication. However, the same cannot necessarily be said of the rest of the construction industry.
- The second sub-hypothesis was disproved in the first part – contractual provisions were generally considered sufficient in all standard forms of contract, with the possible exception of GCC. Lack of organisation and visibility was a recurring theme. Thus the other part of the second sub-hypothesis was confirmed in that it was generally agreed that institutional support was lacking. Regularisation was suggested along the lines that the practice of arbitration is organised under AASA.
- The third sub-hypothesis was confirmed – there were not enough adjudicators, and although there was no established set of skills or minimum training requirements for adjudicators, there was general agreement on relevant skills, useful techniques and desirable personal attributes. There was also broad agreement on the possible content of an “adjudication qualification”
if it were to be implemented, from the acquisition of knowledge and experience to the assessment and accreditation of competence.

The fourth sub-hypothesis was confirmed – the SA construction industry was generally considered not to be able to realize the full potential of adjudication in the current circumstances, and the main reason for this was considered to be lack of knowledge.

**CONCLUSION**

Based on the findings above, it can be concluded that adjudication has found acceptance in the SA construction industry. However, it still has some way to go before its potential can be realized in full. The main challenge appears to be lack of knowledge. Other challenges range from the contractual, institutional and legislative framework, to matters of skills and training. It is with this in mind that the recommendations below are made.

**RECOMMENDATIONS**

In keeping with the conclusion and findings, the following recommendations are made:

- Increase knowledge and understanding of adjudication by the construction industry (full treatise available at libraries of SAICE, AASA).
- Improve the wording of standard forms of contract, strengthen provisions for adjudication, and standardise the process as far as possible.
- Organise the practice of adjudication, either through an existing organisation (e.g. AASA, CIDB, DRBF local chapter, etc) or by establishing a dedicated one.
- Introduce legislation to support the process of adjudication.

**SELECTED REFERENCES**

(The full list of references is available from the authors on request.)


**APPENDIX 1: MORE DETAILED RESULTS**

<table>
<thead>
<tr>
<th>Question</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Background</td>
<td>58% of the respondents practised in engineering construction and 34% in building construction</td>
</tr>
<tr>
<td>1.2</td>
<td>65% of the respondents held a qualification in engineering, 13% in architecture, and 10% in each of quantity surveying and legal</td>
</tr>
<tr>
<td>2. Level of use and knowledge</td>
<td>46% rated their knowledge of adjudication very high, 39% high and 14% average</td>
</tr>
<tr>
<td>2.2</td>
<td>34% each use adjudication rarely and often, 20% regularly and less than 10% each for “never” and “always”</td>
</tr>
<tr>
<td>2.3</td>
<td>total of 96% agreed that adjudication was quicker, 86% for cheaper, 80% for providing interim relief, 72% for immediately binding, 69% for expertise of adjudicator, 55% for enforceable, and 53% for consensual</td>
</tr>
<tr>
<td>2.4</td>
<td>total of 80% of respondents had had satisfactory experience with adjudication</td>
</tr>
<tr>
<td>3. Adjudication in practice</td>
<td>contractual provisions for adjudication were considered sufficient by a total of 55% of respondents for JBCC, 48% for GCC, 68% for FIDIC, and 62% for NEC</td>
</tr>
<tr>
<td>3.2</td>
<td>Institutional guidelines for adjudication were considered adequate by 50% of respondents for JBCC, and between 60% and 90% of respondents were not familiar with other (international) guidelines</td>
</tr>
<tr>
<td>3.3</td>
<td>legislation for adjudication was considered effective by 50% of respondents for UK, and over 75% of respondents were not familiar with legislation from other countries</td>
</tr>
<tr>
<td>3.4</td>
<td>other enabling factors appeared in the order of (from most suggested) skills, party relations, court support and publicity</td>
</tr>
<tr>
<td>4. Skills and techniques</td>
<td>65% of respondents considered that there were not enough adjudicators in the SA construction industry</td>
</tr>
<tr>
<td>4.2</td>
<td>total of 90% of respondents agreed that both technical expertise and legal knowledge were relevant skills for adjudicators, and 70% agreed with project management skills</td>
</tr>
<tr>
<td>4.3</td>
<td>96% of respondents agreed that the inquisitorial approach was useful in an adjudication, 60% disagreed with the adversarial approach, 80% agreed with the facilitative approach, and 90% agreed with the evaluative approach</td>
</tr>
</tbody>
</table>
## APPENDIX 1: MORE DETAILED RESULTS (continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4</td>
<td>total of 70% agreed that age was a desirable personal attribute in an adjudicator, 96% agreed with experience, 60% agreed with professional registration, 40% did not agree with professional accomplishments, 45% agreed with corporate seniority, 93% agreed with fairness, 84% agreed with procedural approach, and 90% agreed with availability</td>
</tr>
<tr>
<td>4.5</td>
<td>total of 80% agreed that participating in an adjudication was important to acquire knowledge and experience, 90% agreed with conducting an adjudication, 80% agreed with self-study, 72% agreed with attending seminars, 84% agreed with taught courses and 72% agreed with assignments</td>
</tr>
<tr>
<td>4.6</td>
<td>62% agreed that examination was important to assess competence, 80% agreed with interview/peer review, 65% agreed with mock adjudication, and 45% considered that a certificate of attendance was a nice-to-have</td>
</tr>
<tr>
<td>4.7</td>
<td>respondents were roughly split equally on regulating the practice of adjudication, but majority believed it should be better organised (similar to AASA role in arbitration)</td>
</tr>
<tr>
<td>5. Impact</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>respondents were roughly equally split on whether or not SA is able to realize the full potential of adjudication</td>
</tr>
<tr>
<td>5.2</td>
<td>75% believed the factors discussed had an impact on the practice of adjudication</td>
</tr>
<tr>
<td>5.3</td>
<td>50% considered lack of knowledge as the single most important contributing factor</td>
</tr>
<tr>
<td>5.4</td>
<td>suggestions for improvement appeared in the order of (from most suggested) skills and training, promoting adjudication, improving contracts, work-shopping lessons learned, introducing legislation and providing institutional support</td>
</tr>
<tr>
<td>6. Legislation</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>total of 75% agreed that SA needs a &quot;Payment and Adjudication Act&quot; similar to that in the UK and other countries</td>
</tr>
<tr>
<td>6.2</td>
<td>total of 60% agreed that such legislation should address minimum payment terms, 90% agreed with statutory adjudication, and 95% agreed with remedy in case of non-payment</td>
</tr>
<tr>
<td>6.3</td>
<td>95% agreed that scope for such law should cover all disputes under the contract, and there was a split opinion on professional liability as well as on special provisions for emerging contractors</td>
</tr>
<tr>
<td>6.4</td>
<td>80% agreed that such law should have an international component</td>
</tr>
<tr>
<td>7. Interest</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>96% of research respondents wished to see the results of the study</td>
</tr>
</tbody>
</table>
Influence of mica on unconfined compressive strength of a cement-treated weathered granite gravel

M R Mshali, A T Visser

The road construction industry faces a shortage of naturally occurring gravel materials that meet the requirements for base or even at times sub-base quality. This situation is exacerbated in some cases by the occurrence of mica in soils. This is reported to significantly affect the engineering properties of materials, including plasticity index and compacted density. The objective of this paper is to investigate the influence of mica on the unconfined compressive strength (UCS) and volumetric changes of a cement-treated gravel material. Free mica (muscovite) was added in predetermined percentages by mass to neat gravel (G5) and specimens subjected to a series of standard laboratory tests. The results show that UCS of greater than 3 MPa is achievable by stabilising less than 5% mica content gravel material with at least 4% cement. Mica content beyond 10% results in very low UCS, even for cement content greater than 6%.

INTRODUCTION

The road construction industry faces a shortage of naturally occurring gravel materials that meet the requirements for base or even at times sub-base quality. Construction economics, social and environmental factors surrounding sourcing and haulage of suitable materials at times justify the stabilisation of local material with additives such as cement. However, suitability of soil for stabilisation depends on factors such as grading characteristics, chemical and mineralogical composition, as well as conditions under which they occur on site.

Natural gravel materials vary in nature, composition and properties depending on their geological formation and weathering environment. The content of free mica minerals in gravel, particularly muscovite, is reported to significantly affect such engineering properties as plasticity index, compacted density and strength (Tubey & Bulman 1964; Stewart et al 1971; Weinert 1980; Balogun 1984; Gogo 1984 and Clayton et al 2004). This problem has been reported in several countries in Africa, for example Ghana, Nigeria (Gogo 1984; Gidigasu & Mate-korley 1980), Zimbabwe (Mitchell et al 1975), South Africa (Paige-Green & Semmelink 2002) and Malawi (Netterberg et al 2011) where some road projects traverse micaceous soils.

Mica is a phyllosilicate mineral with a common basic crystal structure, platy morphology and perfect basal cleavage (Fleet 2003). Micaceous materials include minerals such as muscovite and biotite, which are known to contain significant quantities of mica.

True micas are platy and highly elastic minerals that have been reported to influence the Atterberg limits, density and compactability of road building soils, whereas biotite is known to have less effect on the engineering properties of the soil (Weinert 1980). Micaceous rocks, such as granite (containing 2% – 5%), sedimentary rocks and certain metamorphic rocks, such as mica schist, gneiss and sandstones (Harvey 1982 and Dapples 1959).

Literature review revealed that a limited number of studies have reported the quantitative effects of mica on the unconfined compressive strength (UCS) of the cemented gravel soils that could assist in contextual assessment and deciding whether to consider stabilising micaceous gravel soils for use in base or sub-base layers or not. Ballantine & Rossouw (1989), TRH 14, TRH13, and DoT (1993) state that if mica can be easily seen, the quantity of mica is likely to cause problems and the soil should preferably not be stabilised. Weinert (1980) suggests that soils containing more than 10% of mica, especially muscovite, should be avoided for use in pavement layers, whereas Mitchell et al (1975) recommend that, where materials are adjudged to be very micaceous, the acceptable plasticity limits should be lowered by 33% besides meeting the strength specifications.

Influence of mica on soil properties

Weinert (1980) notes that mica affects soil properties such as liquid limits, plastic limits, density and compaction ability. Casagrande...
(1947) points out that micaceous soils have substantially greater liquid limit than a similar soil without the mica. Mitchell et al (1975) also reported that the presence of mica reduces the apparent plasticity as measured in the Atterberg tests, but increases the effective plasticity, making the material weaker and difficult to compact.

In a study on the influence of decomposed mica schist on compaction and strength of major soil groups in Ghana, Gogo (1984) found that the presence of mica at about 13.5% contributed to the relatively low compaction densities and the high sensitivity to moisture changes. Ballantine & Rossouw (1989) state that compaction problems associated with micaceous soils are due to the springy action and high water demand of the mica mineral. Tubey & Webster (1978) concluded from their investigation on the effects of mica on physical properties of china clay sand as a road-making material that the resilience of mica plates reduces the degree of compaction achievable for a given compaction effort by about 0.007 Mg/m³ and 0.12 Mg/m³ per one percent of fine (<0.425 mm) and coarse mica, respectively.

A study of micaceous sandy silts by Tubey & Bulman (1964) showed that the relation between soil strength in terms of CBR and equilibrium moisture content was relatively poor. CBR values of the micaceous soils at the same compaction effort, but from different climatic environments, affected the established correlation between California Bearing Ratio (CBR) and pavement thickness. The soils are noted to be permeable, and their field strength rapidly reduces by entry of water. Gogo (1984) notes that predicting CBR strengths of soil with mica content greater than 13% and at moisture content greater than 15% could be quite difficult.

Clayton et al (2004) carried out experiments on a mixture of sand and mica and demonstrated that the addition of 10% or more of mica by mass leads to suppression of any dilation, high levels of pore pressure during shear, and low un-drained shear strengths. In addition, it was noted that the mica particles significantly prevent close packing of the sand particles, resulting in a drop in void ratio and a decrease in dry density.

SANRAL (2004) recommends that crushed stone base aggregates containing mica, such as granite, mica schist, pegmatite and sandstone, shall not contain more than 2% by mass of free mica, especially muscovite, when assessed by visually separating the particles, or more than 4% by volume when assessed by means of microscopic slides.

**Stabilisation of micaceous soils**

Cement is recommended for stabilisation of micaceous material in order to improve the material strength, as well as suppress the effects of mica on the plasticity index and compacted density (Mitchell et al 1975; Stewart et al 1971). Cement stabilisation is also reported to reduce swell and increase the soaked CBR strength of the material (Gidigasu & Mate-korley 1980).

Stability of highly micaceous soils is achievable with cement or lime stabilisation. However, Tubey & Bulman (1964) pointed out the need for comprehensive laboratory and field tests in order to relate actual performance of the highly micaceous soils in road construction to the results of laboratory tests. Limited information is available that show a trend relation between occurrence of free mica in percentage by mass of gravel material to strength of the stabilised material. Reports are available (Netterberg et al 2011) that link failure of road and airport pavements to occurrence of mica, but limited information exists that relate percentage by mass of mica cement content and field performance levels.

In view of the above, the objective of this paper is to investigate the influence of mica on the unconfined compressive strength of cement-treated weathered granite gravel material. Other strength-related properties, such as compaction, are also investigated. In addition the effect of mica on the volumetric changes of the cement-treated gravel is investigated.

**MATERIALS AND METHODS**

This study used dry ground muscovite sourced from Phalaborwa mines, G5 weathered granite gravel (potentially problematic with less than 0.5% free mica content) from Midrand quarry, fresh CEM II/B-V 32.5R Portland fly ash cement from Pretoria Portland Cement (PPC) Ltd, and tap water. UCS is the main criterion for assessing suitability of the treated gravel material for use in base and sub-base layers (TRH 14, 1985). Thus, two variables (mica and cement content) were considered to influence the UCS, and hence the use of a factorial design for the experiment was adopted. Difficulty in establishing reliable percentage of naturally occurring mica content in soils has been reported by several researchers (Tubey & Bulman 1964; Weinert 1980; Gogo 1984). In this regard, controlled addition of a known amount of mica to a regular gravel material was considered in order to eliminate this problem and ensure that variation in mica and cement effects are not overshadowed by other factors.

Free mica was added to G5 gravel material in predetermined percentages of 0, 2, 5, 10 and 15% by mass so that subtle trends in the effects of the mica content on UCS and other properties could be investigated. Figure 1 shows particle size distribution of free mica, neat gravel and the prepared specimens. Based on the ICC (2% after one hour) of the gravel material 2, 4, 6...
and 8% cement was added and then each specimen was compacted to 100% Mod AASHTO density. Specimen preparation and testing were conducted in accordance with the standard methods of testing road construction materials (TMH1 1986).

RESULTS AND DISCUSSIONS

Table 1 shows the Atterberg limits and linear shrinkage for 0, 2 and 5% mica content gravel. Difficulties in determining reliable Atterberg limits for 10% and 15% mixes were noted, and hence not recorded. Similar difficulties in the replication of Atterberg limits results were also reported (Tubey & Bulman, 1964; Ruddock 1967).

The results in Table 1 show that the addition of mica reduces the PI from 7 to NP, and hence confirms cement as an appropriate stabilising agent for all the specimens.

Figure 2 gives selected scanning electron microscope (SEM) images of gravel and mica samples. Gravel particles are noted to be cubical and with rough faces, whereas mica particles are noted to be platy and with very smooth faces. Figure 3 shows an SEM image of a compacted sample with high (15%) mica content and schematic presentation of the gravel particles and mica plates during the compaction process. It is postulated in this figure that the platy mica particles restrain smaller gravel particles from filling the voids in the coarse gravel particle fabric. This could be one of the reasons for the difficulties reported by many in compacting high-mica content gravels.

Stabiliser demand of each mica-gravel design mix was determined using the modified DoT (Department of Transport) method (Netterberg 2007a & b). Initial consumption of cement (ICC) was averaged as 2% in light of the reasonably constant readings being obtained at pH greater than 12.4. Maximum dry density (MDD) and optimum moisture content (OMC) were determined as 2 154 kg/m$^3$ and 6.2% respectively. All compaction specimens were prepared

![Figure 2 SEM images of +0.425 mm and -2 mm neat gravel soil particles (left) and mica (muscovite) particles (right)](image)

![Figure 3 SEM image of 15% mica –gravel treated with 2% cement (top) and schematic figure of mica plates restraining soil grains from filling voids during compaction (bottom)](image)

![Figure 4 Effect of moisture content on the CBR and density of compacted specimens](image)
using minus 19 mm gravel, with greater gravel particles crushed and mixed in and compacted using Mod AASHTO compaction effort. Figure 4 shows that the addition of 10% mica reduces compacted density by almost 5% from 2 154 kg/m³ to 2 069 kg/m³ and increases OMC from 6.2% to 8.3%. It is interesting to note that the 10% mica gravel has high soaked CBR values at moisture content less than 6.2%.

At optimum moisture content (OMC), neat gravel had a CBR of 60% at 95% Mod AASHTO compaction effort. This confirmed the TRH 14 soil classification of G5 for the neat gravel sample. However, at 100% Mod AASHTO, the material had a CBR of 101%, which is well above the CBR of 58% for 10% mica–gravel compacted at the same energy and moisture content. This suggests that the addition of 10% mica to neat gravel results in a drop in soil strength, and by extrapolation this material could probably be G6 quality material at 97% Mod AASHTO density.

Figure 5 shows the relation between dry density and CBR strength of the specimen compacted using Mod AASHTO effort. Indications are that the CBR strength of the 10% mica–gravel decreases with increase in moisture content despite the increase in dry density. As is evident from the SEM, the specimen with more than 10% by mass of free mica particles does not stack in flat face to flat face, but rather randomly crisscross in the voids and between the larger granite soil particles. The flat surfaces of mica plates, together with the crisscrossing packing in the gravel particles fabric, result in increased void ratio and ability of the soil to absorb more water and reduce gravel particle interlocking and friction force at contact points. It is thus important to control moisture content during compaction of micaceous gravel materials.

**Volumetric changes**

Table 2 gives volume changes of the compacted specimens before and after soaking for four hours of seven-day and 28-day cured specimens.

Figures 6 and 7 show percentage changes in volume before and after soaking the specimens for four hours. The figures show that an increase in mica content at all levels of cement content results in expansive volumetric changes. Specimens with less than 10% mica content and 6% cement content recorded less than 1% volumetric change for seven-day specimens. However, it is interesting to note that specimens containing 15% mica content and 6% cement content or more expanded by more than 1% of the original volume.

With the exception of the 15% mica content, the general observation of the volumetric changes is that there is minimal change

### Table 2: Average percentage change in volume of compacted specimens

<table>
<thead>
<tr>
<th>Mica content (%)</th>
<th>Volume change: 7-day curing</th>
<th>Cement content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0.672</td>
<td>-0.332</td>
</tr>
<tr>
<td>2</td>
<td>-0.292</td>
<td>-0.209</td>
</tr>
<tr>
<td>5</td>
<td>0.549</td>
<td>0.014</td>
</tr>
<tr>
<td>10</td>
<td>0.724</td>
<td>0.199</td>
</tr>
<tr>
<td>15</td>
<td>0.702</td>
<td>0.545</td>
</tr>
</tbody>
</table>
in volume of the specimens after soaking for four hours, regardless of the increase in mica and cement content. This implies that, after the initial volume change that occurs soon after extrusion of the specimens, the compacted specimens attain stability against volume change in response to variation in moisture conditions. This may be as a result of cement hardening, binding the soil and mica particles together.

### Unconfined compressive strength

Table 3 provides a summary of average UCS results for seven-day and 28-day specimens. Analysis of the variance of the data in the table indicated significant influence of mica and cement content on the UCS values.

Figure 8 presents the relation between mica content and UCS for seven-day specimens at different levels of cement content. The design strength class indicated on the graphs does not in any case classify the treated material, but rather gives an indication in UCS range in which a specific material falls. Figure 8 shows that the addition of 2% mica content results in an increase in strength of 2% cement content specimens by almost 1 MPa before the UCS drastically drops at 5% mica. It is considered that small quantities of mica fill in the void spaces between gravel particles, thereby increasing the dry density. In addition, SEM images show that mica tends to align itself to flatter faces around the larger soil particles, which may suggest that a smooth mica surface provides a plane over which adjoining soil particles slide during compaction, resulting in an increase in density and UCS.

For mica content less than 5%, an increase in mica content results in a minimal low unit rate of reduction in strength of specimens, as evident from the flat plot lines. However, the steep slopes of plot lines for mica content greater than 5% indicate the greater negative effect of mica on UCS. The effects of inherent properties of free mica on the weathered gravel are noticeable at mica content greater than 5%. However, as was shown in Figure 1, it is noted that the addition of free mica changed the particle size distribution of the specimens to some extent, particularly for <2 mm particle size range. Particle size distribution has an effect on the engineering properties of the road building materials. This aspect of material properties was not further investigated in this study, but has been taken into consideration when analysing results from specimens with different amounts of added free mica content. The effect of mica properties, at 10% mica content and above, on the material strength dominates over the binding effect of as high as 8% cement content.

Figure 8 also shows a wide gap between UCS plot lines for 2% and 4% cement content.
for mica content less than 5%. This implies that 4% cement content gives the best UCS gain, and further addition of cement results in less increase in UCS per unit percentage of cement.

Figure 9 shows UCS plot lines for 28-day specimens. An average of 5.5 MPa was obtained for 4, 6 and 8% cement content, giving an increase of almost 1 MPa over and above the 2% cement content UCS. The flat gradient of plot lines implies that the addition of 2% mica has no effect on 4 – 8% cement-stabilised specimens. However, 2% mica negatively affects the strength of 2% cement content specimens. This indicates that optimum gain in strength is obtained at 4% cement stabilisation of gravel with less than 2% mica content.

Furthermore, Figure 9 shows a drastic drop in UCS of specimens with mica content greater than 2% and stabilised with 4% cement content. It is thus noted that 6% cement content provides constant UCS for as much as 5% mica content gravel. Further increase in free mica content beyond 5% results in a drastic decrease in UCS for all levels of cement content. This gives an indication of serious potential problems in strength that could be associated with soils greater than 5% free mica content.

Comparing strength plot lines in Figures 8 and 9, it is noted that there is a substantial and steady decrease in strength of the 5% mica content gravel stabilised with 4% cement from 4 MPa for seven-day to 3.4 MPa for 28-day. Without further investigation, it can only be speculated that the cause of the drop in strength could be that at content greater than 5% the properties of the elastic, smooth-faced and flaky mica dominate over the physical properties of the minus 2 mm component of the soil fabric (in-fill) of the original neat granite gravel. This also triggers questions as to the performance and durability of the treated material with time and under traffic loading.

**Relation between cement content and UCS**

Figure 10 shows the relation of cement content on seven-day UCS for 0, 2, 5, 10 and 15% mica content specimens. UCS results for 2% cement content give an indication of the quality of original micaceous gravel. The low strength achieved at 2% cement for all levels of mica indicates the failure of cement to suppress the effects of mica and improve the strength of the material. This relates well with the results from ICC tests that indicated 2% cement content as the minimum requirement for modification of the material, and not strength gain.

Figure 10 also indicates that at 2% cement and less than 5% free mica content the treated material achieves the C3 design strength, and even C2 for 4% cement content. The addition of more than 4% cement to less than 5% mica content gravel material yields less rate of gain in UCS. The strength gain difference between 5% and 10% free mica specimens, when compared with the difference between 10% and 15% free mica specimens, is an indication that the influence of free mica is pronounced for free mica quantities greater than 10% free mica content. This concurs with most researchers who have cautioned against stabilisation of road building materials with free mica content greater than 10% by mass.

The results confirm that, with gravel with 10% or more of mica, one achieves insignificant gain in UCS, even at a cement content greater than 8%. TRH 14 (1985) recommends that design UCS should be obtained with no more than 5% by mass stabiliser at optimum OMC and specified density in order to guard against the use of unnecessarily high and uneconomic stabiliser content in cemented layers. It follows then that use of cement as the only stabilising agent for the gravel material with greater than 10% free mica content is not a feasible option. Alternatively, further investigation that combines cement stabilisation with other stabilisation could be considered as possible options.

**Development of a model**

Linear regression analysis is utilised to develop a model based on the seven-day UCS. Coefficients derived from the linear regression showed significant correlation between mica content and cement content and the UCS. Considering the t-statistic results, the following regression model is proposed:

\[
UCS = 3.46 - 0.26\times MC + 0.30\times CC
\]
Increased mica content results in significantly higher UCS. A Seven-day UCS greater than 1.5 MPa is feasible. However, this is feasible for mica content levels less than 10%–10% mica content. The results suggest that stabilising a free mica content of greater than 10% to obtain strengths for sub-base and base layers might not be viable.

Increase in mica content up to 15% caused less than 2% volumetric increase. This is a marginal change in volume to warrant concerns regarding density rebound effects. Increase in volume caused by plus 10% mica content overshadowed shrinkage usually associated with cement-treated material.

ACKNOWLEDGEMENTS

We would like to acknowledge the guidance provided by Dr F Netterberg and Dr P. Paige-Green, and provision of muscovite by Ingwe Mica Industries, cement by Pretoria Portland Cement and gravel by Afrisam. Furthermore, the technical assistance and laboratory facilities provided by the CSIR are acknowledged.

REFERENCES


INTRODUCTION

Steel corrosion causes the most damage in in-service RC structures near the marine environment. However, in laboratory terms, the process of natural steel corrosion is very slow, needing tens of years to cause reasonable structural damage. For example, François & Arliguie (1998), Castel et al. (2003), Vidal et al. (2007) and Zhang et al. (2009a, b, 2010), who allowed their laboratory specimens to corrode naturally, had to wait for four years for steel corrosion to start and an additional two years for first cracking to occur. They only obtained reasonable structural damage after 20 years. These times are not often afforded in laboratory tests. Researchers, understandably, have and continue to use various techniques to accelerate steel corrosion so as to shorten the needed testing time. In doing so they anticipate that structural damage under accelerated tests is proportional to damage caused by natural steel corrosion.

It should be pointed out that results obtained by researchers on laboratory specimens that are subjected to accelerated corrosion tests are often passed on to structural engineers and asset managers to apply them to real RC structures which corrode in the field. If they are not applicable to those structures then there is likelihood for engineers to authorise repairs of corroding RC structures at dangerous levels of steel corrosion or when load-bearing capacities of structures are still adequate. For the safety of occupants of corroding RC structures, as well as to minimise costs from unnecessary repairs, there is need to understand well how to apply (if at all applicable) results from accelerated laboratory tests to in-service structures.

This paper discusses various techniques that are often used in research laboratories to accelerate steel corrosion. It then compares conditions and results between the procedures, and where possible, associates them with those from in-service conditions. Finally, it proposes and points out needed research to establish a standard procedure that should be used in laboratories to study behaviour of corroding RC structures. The focus of the paper is on steel corrosion caused by chloride attack. It is also aimed at steel corrosion carried out with the intention to understand structural behaviour of corroding RC members. Issues regarding effects of accelerated corrosion on the electrochemical nature of RC elements were discussed in detail by Poursaee & Hansson (2009). In their discussion, they strongly discouraged accelerating steel corrosion for the reason that it harms the electrochemical nature of concrete. If concern is limited to the electrochemistry of concrete then adequate results can be obtained within a reasonable time frame, even when corrosion is natural. For example, a period of four years which François & Arliguie (1998), Castel...
Figure 1 Transverse and vertical strains before cracking of cover concrete (Malumbela et al 2011)

et al (2003), Vidal et al (2007) and Zhang et al (2009a,b; 2010) had to wait for their specimens to start corroding is achievable in laboratory tests. However, and as previously mentioned, if interest is on structural behaviour then much longer testing periods are required. Accelerated corrosion is therefore often used to reduce this time of testing. The following section discusses various procedures used to accelerate steel corrosion and how they affect structural behaviour.

ACCELERATED DEPASSIVATION OF STEEL
Concrete normally has an alkaline environment that protects embedded steel from corrosion. This environment can be destroyed by carbonation or by chloride attack. As already mentioned, it may take some years for sufficient chlorides to ingress cover concrete and de-passify steel. To hurriedly depassify it, some researchers opted to mix concrete with chlorides ranging from 1% (Mangat & Elgarf 1999) to 5% (El Maaddawy & Soudki 2003) by weight of cement. Others immersed their cured samples in tanks with NaCl solution with concentration from 3% (Cairns et al 2008) to 5% (Cabrera 1996) by weight of the solution. Levels of concentration of chlorides were often selected to simulate chloride concentration of seawater which has a salt concentration of about 3.5%. Note that both procedures above were used by some researchers (Azad et al 2007; Mangat & Elgarf 1999; Cairns et al 2008). According to Poursae & Hansson (2009), if chlorides are added to a concrete mix, de-passivation of steel is immediate. Therefore the time required for steel to depassivate, which is often used in service life models, does not exist (Tuutti 1980). Understandably, Poursae & Hansson (2009) strongly discouraged this procedure. They rather recommended that steel firstly be allowed to passivate before introducing chlorides to break the passive film.

One important element not discussed by Poursae & Hansson (2009), but which emphasises their recommendation, is that adding chlorides to concrete results in uniform distribution of corrosion agents around the steel. Under natural steel corrosion, however, limited faces of a structure are often exposed to chloride attack. In addition, chlorides and other deleterious compounds are purposely excluded from concrete mixes in practice.

In an attempt to better represent natural steel corrosion, some researchers contaminated selected faces of their cured specimens with chlorides. This was achieved by either building NaCl ponds on surfaces of specimens to be contaminated (Yoon et al 2000; Malumbela et al 2009) or by selectively spraying them with salt solution (Zhang et al 2009a,b; Zhang et al 2010; Rio et al 2005). Under this selective contamination of RC specimens with chlorides, Malumbela et al (2011), Yuan & Ji (2009) and Yuan et al (2007) found steel corrosion to be localised within the direction of ingress of corrosion agents. Its implication was that compared to non-contaminated faces, larger tensile strains (especially prior to cover cracking) were measured on contaminated faces of concrete as shown in Figure 1 (Malumbela et al 2011).

In modelling time to cover cracking, Yuan & Ji (2009) proposed that the remaining bar diameter should be taken as elliptical shaped rather than circular, as used by many other researchers (Liu & Weyers 1998; El Maaddawy & Soudki 2007; Bhargava et al 2006). Jang & Oh (2010) and Malumbela et al (2011) demonstrated that assuming uniform loss of steel underestimates pressure applied by expansive corrosion products and hence overestimates resistance of the cover concrete to cracking that is observed under partial surface steel corrosion. Following discussions by Poursae & Hansson (2009), Yuan & Ji (2009), Yuan et al (2007), Jang & Oh (2010) and Malumbela et al (2011), it is recommended that in accelerated corrosion tests:

i. Steel should be allowed to passivate before adding chlorides to concrete. This is equivalent to saying chlorides should be added externally and not be mixed with concrete.

ii. Only selected faces of concrete elements should be contaminated with chlorides. Specimens should not be submerged in salt solutions. This point will be further discussed later.

IMPRESSED CURRENT DENSITY
Corrosion of steel embedded in concrete occurs by an oxidation-reduction reaction. Loss in steel occurs at the anode where electrons are also produced and transferred to the cathode. This flow of electrons produces a small current which is often divided by the surface area of an anode to give current density. According to Andrade & Alonso (2001) and Alonso et al (1998), current density due to natural steel corrosion is often between 0.1 and 10 μA/cm² but occasionally reaches 100 μA/cm². Researchers make use of this current to speed up laboratory corrosion tests. They apply a larger direct current and adjust it such that reinforcing steel bars which they need to corrode are connected to a positive terminal, and an artificial steel bar/plate is connected to a negative terminal. Reinforcing steel bars therefore become the anode and the artificial steel bar/plate becomes the cathode. A salt electrolyte (aqueous NaCl or CaCl₂) is used to provide electrical contact between the anode and the cathode. This procedure increases electrons that flow around the circuit. Bear in mind that from basic chemistry each reaction (oxidation/reduction) should always be balanced. It is clear that to balance increased electrons from the impressed current, more cations and anions are respectively produced at the anode and at the cathode. At the anode, this is achieved by an increased rate of loss of steel.

The level of impressed current density has varied greatly from 3 μA/cm² (Alonso
levels from accelerated tests would result in losses significantly. Malumbela et al (2010b) found that, at high sustained loads, when steel corrosion is firstly accelerated and then allowed to run naturally, the rate of the widening of corrosion cracks does not change, but the rate of steel corrosion reduces significantly. Malumbela et al (2010b) attributed this to natural steel corrosion producing dryer products that are not easily exuded to the exterior faces of concrete.

Results by El Maaddawy & Soudki (2003), Mangat & Elgarf (1999), Malumbela et al (2010a,b) and Alonso et al (1998) indicate that the effect of the level of current density on structural behaviour is contentious. Further research to clarify this is therefore necessary.

**TYPE OF CATHODES**

Whilst anodes are simply steel bars that are required to corrode, various types of cathodes have been used in accelerated corrosion tests. In some research, metal bars embedded in concrete were used (El Maaddawy et al 2005b, El Maaddawy & Soudki 2003; Badawi & Soudki 2005). In others, metal bars were placed on external surfaces and inside a chloride salt electrolyte (Malumbela et al 2009; Ballim et al 2001). These bars were of different

---

80

Journal of the South African Institution of Civil Engineering • Volume 54 Number 2 • October 2012
dimensions and made from different metals. The most used metals, such as stainless steel, copper and titanium, had good electrical conductivity. Rather than using bars, other researchers chose to use metal plates or mesh that covered the external faces of concrete (Azad et al. 2007; Rio et al. 2005; Gadve et al. 2009; Fang et al. 2004).

Placing cathodes inside concrete means hydroxyl ions are produced inside the concrete and then moved to the anodic steel. The rate of corrosion here is dependent on how well oxygen can penetrate the concrete, as well as how well hydroxyl ions penetrate the concrete to reach the anodic steel. When cathodes are placed externally, hydroxyl ions are no longer produced inside the concrete. It was pointed out by Poursaee & Hansson (2009) that, from an electrochemical viewpoint, this is not acceptable, as hydroxyl ions under natural corrosion are produced inside the concrete. In the situation where cathodes are placed externally, the rate of corrosion is dependent on how well hydroxyl ions can penetrate the concrete. How cathodes should be designed to better represent in-service conditions is unclear and most certainly requires further research. It is, however, reasonable to follow a recommendation by Poursaee & Hansson (2009) that they should be placed inside the concrete.

**TYPE OF CORROSION PRODUCTS DURING STEEL CORROSION**

One more parameter that needs discussion in designing corrosion tests in laboratory is the type of corrosion products. Researchers have detected various corrosion products in corrosion-affected RC structures, all with different densities and volume expansion as shown in Figure 2 (Liu & Weyers 1998; Roberge 1999; Jaffer & Hansson 2009).

The type of corrosion product was found to be primarily dependent on pH and availability of oxygen (Roberge 1999; Jaffer & Hansson 2009; Broomfield 1997). These factors (pH and quantity of oxygen) are very variable and difficult to quantify in a corroding RC structure. Many researchers contend that for corrosion of steel that is embedded in concrete, ferrous hydroxide is the fundamental corrosion product (Liu & Weyers 1998; El Maaddawy & Soudki 2007; Bhargava 2006; Roberge 1999). With an increase in the supply of oxygen (especially after cracking of the cover concrete), more stable corrosion products such as haematite and magnetite are formed.

Varying procedures of accelerated corrosion tests is therefore likely to influence types of corrosion products formed. For example, when specimens are fully immersed in NaCl solution, Hussain (2010) has shown that moisture blocks the pores of concrete, and hence prevents oxygen from diffusing into the concrete to reach the anode. More soluble products, such as ferrous hydroxide, are therefore expected. In addition, when the rate of steel corrosion is high (as in accelerated corrosion tests), the rate of ingress of oxygen into the concrete might not be adequate to produce stable compounds. This helps to explain why in accelerated corrosion tests where specimens are immersed in salt solution, corrosion products that exude the concrete are often greenish-black in colour, indicating a large presence of ferrous hydroxide (Malumbela et al. 2010b,c). On reaching the surface, they immediately turn reddish-brown, indicating a conversion to the more stable compounds such as haematite and magnetite. When steel corrosion is slow and concrete is drier, oxygen is expected to be in abundance to form the stable products. Reddish-brown products, indicating a large presence of stable corrosion compounds, are often found in in-service structures, as well as in laboratory specimens where steel corrosion is natural (François & Arliguie 1998; Castel et al. 2003; Vidal et al. 2007; Zhang et al. 2000; Zhang et al. 2010; Malumbela et al. 2010b). Since these products are of different volume densities, the rate of widening of corrosion cracks is expected to be greatly influenced by the procedure used to accelerate steel corrosion. It is important to observe that large densities belong with more soluble products. Therefore, at the same level of steel corrosion, specimens that exhibit unstable corrosion products are expected to be more severely cracked than those with more stable products. This is in agreement with results from El Maaddawy & Soudki et al. (2003). On the same note, soluble products easily exude the corroding area and therefore relieve the cover concrete of applied pressure. However, drier products do not easily egress the corrosion region and hence sustain the pressure. This argument is in agreement with results from Malumbela et al. (2010b) and from Alonso et al. (1998). Research on the chemical composition of corrosion products from accelerated corrosion and from natural corrosion tests, and how they affect cover cracking, is needed.

**ACTUAL LOSS OF STEEL DURING CORROSION TESTS**

As already mentioned, structural engineers and asset managers often rely on measurable parameters of corroding RC structures, such as corrosion crack widths and stiffness, to predict levels of steel corrosion, as well as their residual load-bearing capacities. This involves using relations developed by researchers such as a relation between corrosion crack widths and mass loss of steel. To confirm these relations, some researchers have measured the actual level of steel corrosion at the end of accelerated corrosion tests. This was done by removing corroded steel bars from concrete specimens, cleaning them, and measuring levels of steel corrosion as mass losses of steel or as corrosion pit depths. In real structures, however, it is uncommon for corroded steel bars to be removed from structures. Faraday’s Law is therefore often used to estimate the level of steel corrosion. It is also extensively used in modelling other parameters of corroding RC structures, such as time to first cover cracking (El Maaddawy & Soudki 2007) and stiffness of corroded structures (El Maaddawy et al. 2005a). To relate measurable parameters of RC structures with the level of steel corrosion accurately, the suitability of Faraday’s Law to estimate the level of steel corrosion needs to be understood.
Figure 3 shows a plot of mass loss of steel measured at the end of corrosion tests with predicted mass loss of steel from Faraday’s Law. The data in Figure 3 was obtained from various researchers in the literature. A summary of conditions of the experiments used by the researchers are in Table 1. As expected from the variation of conditions for accelerated corrosion from various researchers, Figure 3 shows a large scatter. The difference between mass loss of steel predicted from Faraday’s Law and actual mass loss ranged from -6.7 to 23.9% with a mean of 1.3% and a standard deviation of 3.6%. Note that this excludes results from Malumbela et al. (2010c) which are also shown in the figure but will be discussed in detail later. Despite the scatter, there was a trend (R^2 = 0.82) that measured mass loss was linearly related to predicted loss. It is evident from the figure that at mass losses of steel above 8%, the majority of data points were below the line of equality. This indicates that at large mass losses of steel (>8%), Faraday’s Law tends to underestimate the level of steel corrosion. The trend-line shows the predicted loss to be around 18% larger than the measured loss. Some researchers believe this to be caused by corrosion products building up around the reinforcing bar surface and thus forming a physical barrier to the ingress of corrosion agents (Liu & Weyers 1998; Badawi & Soudki 2005). From the previous discussion, it is expected that more soluble products which occupy larger volume will form a large barrier and hence significantly retard the corrosion process.

Despite the trend discussed above, it is worth pointing out that the measured mass loss of steel presented in Figure 3 is an average mass loss of steel over the entire corroded length of a bar. If the level of steel corrosion varies along the bar, average mass loss of steel, and hence Faraday’s Law, may underestimate the maximum level of steel corrosion. Rather than measuring average mass loss of steel, some researchers opted to measure maximum pit depths (Torres-Acosta & Martinez-Madrid 2003). Torres-Acosta et al. (2007) and Torres-Acosta & Martinez-Madrid (2003) tried to correlate maximum pit depths with average penetration depth (calculated from average mass loss). They found them to be linearly related, but maximum pit depth was about eight times larger than average penetration depth. Similar results were found by Rodriguez et al. (1997). This is important information which suggests the need to evaluate the accuracy of Faraday’s Law to predict maximum mass loss of steel.

Malumbela et al. (2010c) researched on the relation between maximum mass loss by Rodriguez & Torres-Acosta (2007) and Malumbela et al. (2010b). They found them to be linearly related, but maximum pit depth was about eight times larger than average mass loss. Similar results were found by Rodriguez et al. (1997). This is important information which suggests the need to evaluate the accuracy of Faraday’s Law to predict maximum mass loss of steel.

Table 1 Various procedures used to accelerate steel corrosion in RC specimens

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Procedure for accelerated steel corrosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badawi &amp; Soudki (2005)</td>
<td>Concrete mixed with 2.25% chlorides by weight of cement. Specimens placed in 100% humidity chamber during corrosion. Current density = 150 μA/cm^2. Stainless steel bar embedded in concrete was used as a cathode.</td>
</tr>
<tr>
<td>El Maaddawy &amp; Soudki (2003)</td>
<td>Concrete mixed with 5% chlorides by weight of cement. Specimens wrapped with burlap sheets and wetted daily with fresh water during corrosion. Used current densities that ranged from 100 to 500 μA/cm^2. Stainless steel bar embedded in concrete was used as a cathode.</td>
</tr>
<tr>
<td>Cabrera (1996)</td>
<td>Concrete mixed with 2% chlorides by weight of cement. Specimens immersed in 5% NaCl solution during corrosion. Used current densities that ranged from 76 to 764 μA/cm^2. Used a stainless steel plate immersed in NaCl solution as a cathode.</td>
</tr>
<tr>
<td>El Maaddawy et al. (2006)</td>
<td>Concrete mixed with 3% NaCl by weight of cement. Specimens placed in a humidity chamber and constantly sprayed with fresh water mist during the corrosion process. Applied constant voltages of 15 and 60 V. Stainless steel bar embedded in concrete was used as a cathode.</td>
</tr>
<tr>
<td>El Maaddawy et al. (2010c) (2-day drying cycles)</td>
<td>Concrete mixed with 2.15% chlorides by weight of cement. Specimens subjected to 2½-day wet (with water) and 1-day dry cycles. Current density = 140 μA/cm^2. Used a stainless steel bar embedded in concrete as a cathode.</td>
</tr>
<tr>
<td>Azad et al. (2007)</td>
<td>Concrete mixed with 2% NaCl by weight of cement. Specimens immersed in 5% NaCl solution during corrosion. Specimens subjected to 2½-day wet (with water) and 1-day dry cycles. Current density = 1214 μA/cm^2. Stainless steel bar embedded in concrete as a cathode.</td>
</tr>
<tr>
<td>Ballim et al. (2001, 2003)</td>
<td>Specimens were carbonated at a pressure of 80 kPa for six days. They were then immersed in 3% NaCl solution during corrosion. Current density = 400 μA/cm^2. Cathode was a steel rod immersed in NaCl solution.</td>
</tr>
<tr>
<td>Yoon et al. (2000)</td>
<td>Tensile face of specimens constantly wetted with 3% NaCl solution. Current density = 370 μA/cm^2. Copper plate immersed in NaCl solution was used as a cathode.</td>
</tr>
<tr>
<td>El Maaddawy et al. (2005b)</td>
<td>Concrete mixed with 2.25% chlorides by weight of cement. Concrete sprayed with mist during the corrosion process. Current density = 150 μA/cm^2. Cathode was a stainless steel bar embedded in concrete.</td>
</tr>
<tr>
<td>Malumbela et al. (2010c)</td>
<td>Tensile face of specimens cyclic wetted (for four days) with 5% NaCl solution and dried (for two days in some and four days in others). Current density = 189 μA/cm^2. Used stainless steel bar immersed in NaCl solution as a cathode.</td>
</tr>
</tbody>
</table>
of steel and Faraday’s Law. Their corrosion process involved a current density of 189 μA/cm² and two different cycles of wetting of cover concrete with 5% NaCl solution, and natural drying under laboratory conditions. One accelerated process entailed four-day wetting followed by two-day drying cycles whilst in the other, cycles were all four days. They only contaminated the tensile face of RC beams with salt solution. Their target mass loss of steel from Faraday’s Law was 10%. This meant 44 wetting days when the current was impressed. At the end of the test, they measured both average mass loss of steel and maximum loss.

It is clear from Figure 3 that for beams with two-day drying cycles, maximum mass losses of steel were largely greater than predicted losses. The largest loss in those beams was 12.1% compared to 10% from Faraday’s Law. Despite these larger mass losses of steel, Figure 3 shows that they are still within the range of values that were observed by other researchers who measured average mass loss. However, their consistency, which did not exist in results from other researchers, points to the need to be cautious when predicting maximum mass loss of steel using Faraday’s Law.

Mass losses of steel in beams with four-day drying cycles were certainly much larger than corresponding losses in beams with two-day drying cycles. The most obvious reason that can be attributed to beams with longer drying cycles having larger mass losses of steel, is that longer drying cycles could have allowed for more natural corrosion to occur because of the extended time required to reach the desired time of electrolysis. As already mentioned, for beams to have the target level of steel corrosion of 10%, beams corroded using two-day drying cycles were tested for 64 days (44 wetting days + 20 drying days). Beams tested with four-day drying cycles were, however, tested for 80 days (44 wetting days + 36 drying days). This implies that beams under the four-day drying cycle had 16 days of additional natural corrosion compared to beams under the two-day drying cycles. It was later shown by Malumbela et al (2010b) that the natural corrosion rate in beams was too low (30.4 μA/cm²) to have resulted in the recorded large mass losses of steel in beams with four-day drying cycles. The large differences in mass losses here could be ascribed to the set-up with two-day drying cycles not allowing the complete dryness of the concrete cover at the corroded rebar depth. Therefore, after the drying period, less stable products such as ferrous hydroxide (which according to Figure 2 occupy a larger volume than dryer products) would still be available within the corrosion region. On the other hand, more stable products, such as haematite and magnetite (which occupy less volume) could have formed during the four-day drying periods.

The formation of these lesser volumetric compounds could have allowed for more access of corrosion agents to the rebar, which could have led to larger corrosion rates. This notion is in agreement with discussions by Hussain (2010) on the effect of moisture variation on rate of steel corrosion.

Figure 3 clearly indicates that Faraday’s Law is not adequate to predict levels of steel corrosion, particularly where sufficient drying of cover concrete is permitted. Since natural steel corrosion often occurs under drier conditions than most accelerated corrosion tests, Faraday’s Law is likely to under-estimate levels of steel corrosion in in-service structures. It is therefore recommended that further research be carried out to model the interaction between dryness of cover concrete and rate of steel corrosion.

### INFLUENCE OF CORROSION TEST ON LOAD-BEARING CAPACITY

Mangat & Elgarf (1999) found that, at mass losses of steel due to steel corrosion up to 7%, the level of current density had little effect on the load-bearing capacity of RC beams. However, at mass losses of 10% and beyond, load-bearing capacity of RC beams decreased significantly with increase in the level of the impressed current density. For example, at a mass loss of steel of 20%, current density of 1000 μA/cm² induced a loss of load-bearing capacity of 60% compared to 78% when a current density of 4000 μA/cm² was used. They attributed this to a larger loss in the interfacial bond at the steel/concrete interface caused by the high corrosion rates. Contrary to findings by Mangat & Elgarf (1999), Azad et al (2007) reported that it was not the current density that caused a larger reduction in load-bearing capacities at higher levels of steel corrosion, but rather the product of current density with time. They further asserted that a higher value of corrosion current density for a lesser period of time would be as damaging as a lesser value of current density for a longer corrosion period.

Where Mangat & Elgarf (1999) and Azad et al (2007) agreed was that, at large mass losses of steel (>10%), calculated values of load-bearing capacity, using measured average mass losses of steel, had little relation with experimental results. For example, according to Mangat & Elgarf (1999), a mass loss of steel of 1% corresponded to a predicted loss in load-bearing capacity of 20%. The measured loss in the load-bearing capacity was, however, found to be 78%. Azad et al (2007) found average mass loss of steel of 1% to relate to loss in load-bearing capacity of 1.4%. The corresponding relation between mass loss of steel and theoretical load-bearing capacity varied with the level of steel corrosion. At a corrosion level of 31%, theoretical load-bearing capacity exceeded the measured capacity by 30%, but at lower levels of corrosion (around 5%) theoretical capacity was found to be similar to the measured capacity. The researchers (Mangat & Elgarf 1999; Azad et al 2007) attributed the poor predictions of ultimate capacity of beams at high mass losses of steel to losses in the bond between corroded steel bars and the surrounding concrete. They therefore developed necessary correction factors. According to Azad et al (2007), the residual load-bearing capacity of corroded RC beams should be calculated using Equations 1 and 2. In line with their experimental findings, these Equations indicate that the needed-correction-factor, α, reduces with an increase in the level of steel corrosion (it).

\[ M_{u, actual} = \alpha M_{u, theoretical} \]  (1)

\[ \alpha = \frac{14.7}{d(i)^{0.5}} \leq 1 \]  (2)

Where

- \( M_{u, actual} \) = measured capacity of beams (kN·m)
- \( M_{u, theoretical} \) = theoretical capacity of beams based on reduced average cross-sectional area of steel (kN·m)
- \( a \) = correction factor
- \( d \) = bar diameter (mm)
- \( i \) = corrosion current density (mA/cm²)
- \( t \) = duration of corrosion (days)

Torres-Acosta et al (2007) found a poor relation between average penetration depth on steel bars (calculated from average mass loss of steel), due to steel corrosion and the residual capacity of RC specimens. A cross-sectional loss of steel of 1% was found to be equivalent to a loss in capacity of 1.6%. This relation is similar to the relation found by Azad et al (2007) where average mass loss of steel was used. Torres-Acosta et al (2007), however, found a good relation (\( R^2 = 1 \)) between the load-bearing capacity and maximum pit depths. From this relation, it can be shown that a 1% maximum loss in area of steel yields 0.6% loss in load-bearing capacity. Note that Torres-Acosta et al (2007) presented their results using radius loss instead of loss in cross-sectional area of steel. They were converted here to allow them to be
compared with those from other researchers. Interestingly, the relation found by Torres-Acosta et al. (2007) is similar to a theoretical relation between loss of steel cross-sectional area and load-bearing capacity of RC beams developed by Ting & Nowak (1991). Even more intriguing, no correction factors, as recommended by Mangat & Elgarf (1999) and Azad et al. (2007) were needed in Ting & Nowak’s model. It therefore suggests that the correction factors are limited to theoretical models of load-bearing capacity which use average mass loss of steel. More importantly, it implies that the loss in bond between steel and concrete may not be the reason for the failure of the theoretical models.

Malumbela et al. (2010c) showed that load-bearing capacity of corroded RC beams against maximum mass loss of steel was closely related to theoretical results from a basic model of load-bearing capacity of RC beams. This was without applying factors of bond between steel and concrete, as suggested by Azad et al. (2007). Malumbela et al. (2010c) further demonstrated that the use of average mass loss to predict load-bearing capacity of RC beams at high mass losses of steel will overestimate it.

Results from Torres-Acosta et al. (2007) and from Malumbela et al. (2010c) against those from Azad et al. (2007) and from Mangat & Elgarf (2007) suggest that the load-bearing capacity of corroded RC beams is not related to the level of current density, but to the actual maximum mass loss of steel. However, more test results are needed to confirm this.

**CONCLUSIONS**

1. This paper discussed various procedures that are often used to accelerate steel corrosion in laboratory tests of RC specimens. It clearly pointed out that, to hurriedly de-passify steel, researchers should avoid adding chlorides to concrete mixes or fully immersing their samples in salt solutions. These procedures result in uniform steel corrosion that unfortunately underestimates the effects of partial surface steel corrosion, which is often observed in in-service structures. It was instead recommended in the paper that limited faces of specimens should be contaminated with chlorides. This can easily be achieved in laboratories by building ponds on surfaces to be contaminated or by selectively spraying them with chlorides.

2. To accelerate steel corrosion, continuous immersing of specimens in salt solution was shown to underestimate the rate of steel corrosion that is observed when corrosion occurs on a drier cover concrete. Since corrosion in in-service structures involves long drying periods, it was recommended that laboratory corrosion tests should also entail long drying periods. Probably more research is needed to standardise the duration of drying periods.

3. Various types of cathodes are often used when accelerating steel corrosion embedded in concrete. Placing cathodes on exterior surfaces of concrete was shown not to represent natural steel corrosion well. It was recommended that cathodes should be placed inside the concrete. Further research on this was, however, recommended.

4. The level of impressed current density to be used to accelerate steel corrosion was found to be contentious between researchers. For example, El Maaddawy & Soudki (2003) and Mangat & Elgarf (1999) found that, at the same level of steel corrosion, higher current densities cause more structural damage than lower densities, while Alonso et al. (1998) and Malumbela et al. (2010b) found that a lower current density caused more structural damage. It was therefore recommended that this should be researched further.

5. Except for results from Mangat & Elgarf (1999), many researchers found the load-bearing capacity of corroded RC specimens to be related to actual loss in area of steel and not to the level of current density used. Torres-Acosta et al. (2007) and Malumbela et al. (2010c) further showed that load-bearing capacity of corroded RC beams was related more to maximum mass loss of steel than to average loss. More data to confirm this is, however, needed.

**ACKNOWLEDGEMENT**

The support of the Concrete Materials & Structural Integrity Research Unit (CoMSIRU) at the University of Cape Town is greatly acknowledged.

**REFERENCES**


The effects of placement conditions on the quality of concrete in large-diameter bored piles

G.C. Fanourakis, P.W. Day, G.R.H. Grieve

In South Africa, concrete in large-diameter bored piles is generally placed by discharging a high-flow concrete mix directly from the truck mixer and allowing the concrete to fall freely to the base of the pile hole. While certain site practices have been used for piling contractors for years, many engineers are not convinced of their acceptability.

In order to assess the effects of free-fall concrete placement, a series of tests were undertaken in which the properties of concrete placed in this manner were compared with the properties of conventionally placed concrete. The tests included an assessment of the effect of water and spoil in the pile hole at the time of casting, as well as poor placement techniques.

The results of this investigation indicate that casting of concrete in 50 mm and 400 mm of water in the bottom of the pile hole significantly reduced the compressive strength by approximately 50% and 80%, respectively. Furthermore, the effect of spoil at the bottom of the pile hole was dependent on the amount of water present.

Finally, a separate investigation, at a bridge site, indicated the free-fall placement technique to be at least as effective as the tremie technique.

INTRODUCTION

Large-diameter (750 mm – 2 000 mm) bored (augered) piles are ideally suited to the stable residual soil profiles and deep water table conditions frequently encountered in the inland regions of South Africa. In many areas of the country, open holes can be augered without the need for temporary casing. The holes can safely be cleaned by hand and inspected in situ prior to the insertion of the reinforcing cage and placement of concrete. Typically, piles are cast by discharging high-flow concrete directly from the chute of the truck mixer using the deflector flap at the end of the chute to direct the concrete down the centre of the reinforcing cage in a continuous stream.

Most piling specifications and construction drawings clearly specify the class of concrete and the nature of the founding material for cast in situ bored piles. However, in most instances, little or no attention has been paid to site practices which can have a significant effect on the integrity of the pile. These include the method of concrete placement, the amount of water in the pile hole at the time of casting and the cleanliness of the pile socket. In an investigation by Alexander (1983), the unacceptable quality of pile concrete was attributed to the presence of excessive amounts of water at the bottom of the hole at the time of pouring the concrete.

In August 1991, the Research and Development Advisory Committee (RDAC) of the South African Roads Board commissioned a study of quality control during concrete placement in bored piles, with the intention of formulating rational guidelines for use by contractors and site supervision staff.

The main objectives of the programme were to investigate:

- whether the free-fall or slow-pour placement methods result in a loss of strength or in segregation,
- the extent to which the presence of water in the pile hole affects the strength of the concrete, and
- what happens to any spoil remaining in the bottom of the pile hole during concrete placement and how this affects the integrity of the pile shaft.

A total of 20 trial "piles" were cast with varying amounts of water and/or spoil. Concrete cores from these piles were tested to determine their compressive strength, actual density and aggregate–binder ratios. In addition, the percentage excess voids was visually assessed. The effect of spoil on the contact at the end of the pile was also visually assessed.

This paper describes the procedures used during the tests and summarises the results obtained. Further details are given in the RDAC (1995) Report.
The effectiveness of the free-fall placement method, relative to the tremie method, was further assessed by comparing cores taken from piles of a particular bridge cast by either of these two placement techniques.

**PRACTICE IN THE PILING INDUSTRY**

In order to ascertain the current practice within the piling industry, a survey was conducted amongst four of the larger piling contractors on the Witwatersrand.

The contractors were unanimous that, where possible, large-diameter auger holes should be hand-cleaned as a matter of course. The purpose of hand-cleaning is to remove all loose material, as this may settle. Furthermore, the concrete should not be cast into more than 100 mm depth of standing water at the bottom of the pile hole. The minimum hole diameter which can be cleaned by hand was given as 750 mm, although hand-cleaning of a 600 mm diameter hole is possible in exceptional circumstances. Most contractors would recommend a reduction in the allowable end-bearing stresses for small diameter piles which are cleaned using a cleaning bucket and cannot be inspected in situ.

The amount of spoil that contractors would allow at the bottom of an end-bearing bored pile hole at the time of casting varied from nothing (if the hole was hand-cleaned) to 150 mm if the hole could not be cleaned. In the latter case the remaining material would be compacted by means of a tamper on the Kelly bar (the drill stem attached to the auger flight).

With regard to the method of concrete placement, contractors deemed it acceptable to cast the concrete by free-fall either straight from the chute of the truck mixer or using a short centralising tube. Only in the case of raking piles would the fall be limited, by means of trunking (pouring the concrete through a tube inserted within the reinforcing cage), to near the top of the concrete. None of the piling contractors would tremie concrete into a dry hole in order to prevent segregation. The upper few metres of the shaft would normally be vibrated on completion of the pour.

Most of the contractors interviewed felt that the requirements imposed by site supervision staff for the use of trunking or tremie tubes in dry vertical pile holes, and insistence on vibration of the concrete over the full length of the pile shaft, were unnecessarily stringent.

On the basis of the above survey, it would appear that the piling contractors were generally in full agreement with one another on the requirements for concrete placement. The differences in requirements from site to site appear therefore to be largely due to the opinions of site supervision staff. The fact that these findings were not based on measured results warranted this investigation.

**CODE OF PRACTICE REQUIREMENTS**

**South African Codes**

Most codes of practice for structural concrete lay down strict requirements for the general placing of concrete. Many of these requirements are aimed at preventing segregation and ensuring adequate compaction of the concrete. SANS 1200 G (1982) requires that concrete shall not be allowed to fall freely through a height of more than 3 m, unless otherwise approved and that compaction of the concrete is carried out by mechanical vibration. These requirements frequently find their way into piling specifications where completely different practices prevail.

SANS 1200 F (1983) (Piling) specifies a concrete slump of between 75 mm and 175 mm for various conditions, depending on the method of placement, spacing of reinforcement and diameter of the pile hole. The code recommends that internal vibrators should not be used, that concrete should be placed in the dry or by means of a tremie, that concrete should be placed in such a way that segregation does not occur, and advocates the use of a chute extending far enough into the hole to ensure that the concrete drops vertically when leaving the chute. In the case of raking piles, the chute is required to extend to the leading edge of the newly placed concrete.

Read together, these clauses from SANS 1200 F (1983) imply that the free-fall placement of concrete is permitted in vertical pile holes provided that the concrete is permitted to fall unobstructed down the centre of the pile.

**ACI Manual of Concrete Practice: Concrete Piles**

The 1973 version of the ACI Manual permits the placement of pile concrete at a continuous and rapid rate from the top of the hole, but only through a funnel hopper having a discharge opening smaller than the smallest pile section. Furthermore, the pile hole is to be free of all foreign matter, including “appreciable quantities of water”. Vibration of concrete is recommended for reinforced piles.

Of all the manuals and specifications, the 1973 ACI Manual accords closest with common practice in the South African piling industry. The only significant difference is the recommendation that the concrete in reinforced piles be compacted by means of vibration.

In a later edition of this document (ACI 2000), and subsequent revisions, these clauses have been moved and re-titled. Unfortunately the clauses relating to methods of placement have been omitted in the 2000 and later versions of the ACI document.

**EXPERIMENTAL METHODS OF THIS INVESTIGATION**

**Construction of piles**

During the field work phase of the programme, which was carried out on 11 and 12 July 1991, 20 trial "piles" were cast using free-fall placement of concrete with various amounts of water and/or spoil at the bottom of the pile hole, as schematically indicated in Figure 1. The "piles" consisted of 200-litre...
steel drums placed at the bottom of a 6 m deep, 1.5 m diameter auger hole. The drums had a diameter of 560 mm and a depth of 870 mm. A 50 mm thick concrete blinding layer was cast at the bottom of each drum, to provide a solid base onto which the “pile” concrete could be cast. Before lowering the drums into the hole, measured amounts of water and/or spoil were placed into the drums so that concrete could be cast. Before lowering the drums into the hole, measured amounts of water and/or spoil were placed into the drums so that concrete could be cast. Before lowering the drums into the hole, measured amounts of water and/or spoil were placed into the drums so that concrete could be cast.

Two types of spoil were used, one slightly cohesive and the other granular. The first was a silty andesite taken from the spoil of other pile holes being drilled on the site. This material had a diameter of 560 mm and a depth of 6.8 m to the bottom of the “pile”. Height was 6.8 m to the bottom of the “pile”. Height was 6.8 m to the bottom of the “pile”. Height was 6.8 m to the bottom of the “pile”.

Concrete was discharged into the drums through a 500 mm diameter light-weight steel casing inserted about 100 mm into the top of each drum in turn, as shown in Figure 1. The hole was large enough to accommodate four drums at its bottom, and these were filled in turn. On completion of the pour, the drums were lifted from the hole and left to cure on the surface.

During casting, the concrete was directed down the centre of the casing using the deflector flap at the end of the chute of the truck mixer. The main stream of concrete reached the bottom of the “pile” without impinging on the sides of the casing. The casing simulated the side walls of an in situ pile hole. The rate of pour was rapid and the drop height was 6.8 m to the bottom of the “pile”.

In the penultimate test, holes were cut through the steel casing and reinforcing bars were inserted horizontally across the casing to act as barriers to the free fall of the concrete and to encourage segregation. The concrete was discharged at the same rate as was used for the other tests.

In the final test, the concrete was poured slowly, falling from the chute of the truck mixer as individual blobs (the flow was not rapid and continuous). The rate of discharge was not quantified.

### Mix design

The mix proportions of the concrete, which was supplied by a ready-mix company, are indicated in Table 1.

The tests made use of four batches of the concrete delivered to site by separate truck mixers over the two-day period. Control samples of concrete were taken from each truck by casting concrete into drums on the surface and compacting the concrete by mechanical vibration.

### Table 1 Concrete mix design and characteristics

<table>
<thead>
<tr>
<th>Characteristic strength</th>
<th>25 MPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target slump (actual slump range)</td>
<td>100 mm (50–200 mm)</td>
</tr>
<tr>
<td>Sand (dry)</td>
<td>795 kg</td>
</tr>
<tr>
<td>Stone (19 mm)</td>
<td>1 090 kg</td>
</tr>
<tr>
<td>50/50 CEM I/Slag</td>
<td>335 kg</td>
</tr>
<tr>
<td>Water</td>
<td>200 litres</td>
</tr>
</tbody>
</table>

### Sampling, testing and visual inspection

Approximately two weeks after casting of the “piles”, the drums were turned over and 100 mm diameter, 300 mm long core samples were drilled (vertically) through the bottom of each drum. After visual inspection and photography, the cores were submitted to an accredited commercial laboratory for testing.

Compressive strength tests were carried out on all core specimens at an age of 37 or 38 days after casting. In addition, the actual density (water-soaked density of the uncapped core) was determined and the percentage of excess voids was assessed visually. These tests were carried out in accordance with the recommendations contained in CSTR (1987).

### Table 2 Summary of concrete core test results

<table>
<thead>
<tr>
<th>Test ref</th>
<th>Sample ref</th>
<th>Concrete batch</th>
<th>Compressive strength (MPa)</th>
<th>Compressive strength (% of control)</th>
<th>Actual density (kg/m³)</th>
<th>Excess voids (%)</th>
<th>Aggregate / binder ratio</th>
<th>Test conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7B*</td>
<td>C1</td>
<td>51.0</td>
<td>100</td>
<td>2 450</td>
<td>0.0</td>
<td>9.5</td>
<td>Control test, vibrated</td>
</tr>
<tr>
<td>C2</td>
<td>7T*</td>
<td>C2</td>
<td>39.0</td>
<td>100</td>
<td>2 583</td>
<td>0.0</td>
<td>Control test, vibrated</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>18B</td>
<td>C3</td>
<td>43.0</td>
<td>100</td>
<td>2 620</td>
<td>0.0</td>
<td>Control test, vibrated</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>18T</td>
<td>C4</td>
<td>40.5</td>
<td>100</td>
<td>2 634</td>
<td>0.0</td>
<td>Control test, vibrated</td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td>3B</td>
<td>C1</td>
<td>48.5</td>
<td>95</td>
<td>2 540</td>
<td>1.0</td>
<td>9.2</td>
<td>Free fall, dry</td>
</tr>
<tr>
<td>W2</td>
<td>13B</td>
<td>C4</td>
<td>48.5</td>
<td>120</td>
<td>2 611</td>
<td>1.5</td>
<td>11.9</td>
<td>Free fall, dry</td>
</tr>
<tr>
<td>W3</td>
<td>1B</td>
<td>C1</td>
<td>37.5</td>
<td>74</td>
<td>2 393</td>
<td>0.5</td>
<td>9.3</td>
<td>Free fall, 50 mm water</td>
</tr>
<tr>
<td>W4</td>
<td>6B</td>
<td>C2</td>
<td>38.0</td>
<td>97</td>
<td>2 490</td>
<td>0.5</td>
<td>9.0</td>
<td>Free fall, 50 mm water</td>
</tr>
<tr>
<td>W5</td>
<td>2B</td>
<td>C1</td>
<td>25.0</td>
<td>49</td>
<td>2 517</td>
<td>0.5</td>
<td>7.6</td>
<td>Free fall, 100 mm water</td>
</tr>
<tr>
<td>W6</td>
<td>8B</td>
<td>C2</td>
<td>23.5</td>
<td>60</td>
<td>2 508</td>
<td>1.0</td>
<td>12.2</td>
<td>Free fall, 100 mm water</td>
</tr>
<tr>
<td>W7</td>
<td>4B</td>
<td>C1</td>
<td>9.0</td>
<td>18</td>
<td>2 454</td>
<td>3.0</td>
<td>13.2</td>
<td>Free fall, 200 mm water</td>
</tr>
<tr>
<td>W8</td>
<td>9B</td>
<td>C2</td>
<td>8.5</td>
<td>22</td>
<td>2 428</td>
<td>4.0</td>
<td>14.4</td>
<td>Free fall, 200 mm water</td>
</tr>
<tr>
<td>W9</td>
<td>5B</td>
<td>C2</td>
<td>7.0</td>
<td>18</td>
<td>2 434</td>
<td>10.0</td>
<td>16.9</td>
<td>Free fall, 400 mm water</td>
</tr>
<tr>
<td>W10</td>
<td>12B</td>
<td>C3</td>
<td>10.0</td>
<td>23</td>
<td>2 407</td>
<td>15.0</td>
<td>Free fall, 400 mm water</td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>10B</td>
<td>C3</td>
<td>50.0</td>
<td>116</td>
<td>2 580</td>
<td>0.5</td>
<td>Free fall, 50 mm silts, dry</td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>10T</td>
<td>C3</td>
<td>42.0</td>
<td>98</td>
<td>2 546</td>
<td>1.5</td>
<td>Free fall, 50 mm silts, dry</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>11B</td>
<td>C3</td>
<td>21.5</td>
<td>50</td>
<td>2 540</td>
<td>1.0</td>
<td>Free fall, 50 mm silts, 100 mm water</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>11T</td>
<td>C3</td>
<td>22.0</td>
<td>51</td>
<td>2 522</td>
<td>1.0</td>
<td>Free fall, 50 mm silts, 100 mm water</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>17B</td>
<td>C4</td>
<td>48.5</td>
<td>120</td>
<td>2 546</td>
<td>1.5</td>
<td>Free fall, 50 mm silts, 50 mm water</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>14B</td>
<td>C4</td>
<td>50.5</td>
<td>125</td>
<td>2 564</td>
<td>2.0</td>
<td>Free fall, 50 mm c. dust**, dry</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>15B</td>
<td>C4</td>
<td>46.0</td>
<td>114</td>
<td>2 506</td>
<td>1.5</td>
<td>Free fall, 50 mm c. dust, 50 mm water</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>16B</td>
<td>C4</td>
<td>31.0</td>
<td>77</td>
<td>2 569</td>
<td>1.0</td>
<td>Free fall, 50 mm c. dust, 100 mm water</td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>19B</td>
<td>C4</td>
<td>25.5</td>
<td>63</td>
<td>2 518</td>
<td>1.5</td>
<td>Free fall, with rebar, 100 mm water</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>20B</td>
<td>C4</td>
<td>20.0</td>
<td>49</td>
<td>2 637</td>
<td>1.5</td>
<td>Free fall, slow pour, 100 mm water</td>
<td></td>
</tr>
</tbody>
</table>

Notes: * T indicates top of drum, i.e. about 800 mm above bottom of “pile”
** B indicates bottom of drum, i.e. at bottom of “pile”
** c. dust indicates crusher dust (sandy fines from crushed aggregate)
The aggregate–binder ratios were determined on nine samples of concrete cast through various depths of water and on one of the control samples using the soluble silica test method, as detailed in BS 1881: Part 124: 1988.

After the concrete cores had been taken, the bottom of the drums containing spoil were cut away to observe the extent to which the spoil at the bottom of the “pile hole” had been displaced by the falling concrete. The thickness of the remaining spoil was measured at 100 mm intervals around the perimeter of the “pile”. After removal of the layer of blinding concrete, the area of intimate contact between the “pile” concrete and the bottom of the hole (i.e. the area over which the spoil had been completely displaced) was estimated.

RESULTS

Table 2 summarises the conditions under which the various “piles” were concreted, and the laboratory test results.

DISCUSSION

This section of the paper discusses each of the research objectives (listed earlier) in turn.

Segregation due to free-fall placement

Figure 2 shows cores drilled through the bottom of the “piles” cast through various depths of water in the “pile hole” at the commencement of the pour. The sample reference numbers are indicated on these cores (left to right: 5B, 4B, 2B, 1B, 3B, 7T and 7B).

In this figure, the bottom of the “pile” is facing away from the reader. The contact between the 50 mm blinding concrete cast in the drums and the “pile” concrete is visible in some of the cores.

In all these cores, there was an even distribution of coarse aggregate, despite the higher void content of the concrete for greater water depths. A similar, even distribution of aggregate was observed in Test R1 which simulated the effect of allowing concrete to impinge on the reinforcing cage during free fall into 100 mm of water. The only case where segregation was evident was where the concrete was poured slowly into 100 mm of water as shown in Figure 3.

In Figure 3 the bottom of the core is to the left of the picture. The disc of blinding concrete has separated from the “pile” concrete. The “pile” concrete shows classical signs of segregation, with unbonded aggregate at the toe of the “pile” and decreasing aggregate content with the accumulation of fines and laitance towards the top of the pour.

From these observations it was concluded that, despite the variance in slumps (50 mm to 200 mm), the pouring of concrete at the normal (rapid) rate resulted in sufficient turbulence (and mixing) at the bottom of the hole to prevent segregation of the concrete mix, even where the fall of the concrete was interrupted by impact with the reinforcing steel. This confirms the findings of separate studies carried out by STS Consultants (1994) and Turner (1979), where the free fall heights were 18 m and 15 m respectively. However, where the concrete was poured slowly, this turbulence was absent and no re-mixing occurred at the bottom of the hole. These discharging practices may be compared to opening a tap and discharging water into a half-full basin of water. If the stream of water is continuous, the waters mix. If the water flow is drop by drop, little mixing occurs.

Where water was present, the fine aggregate and cement paste were removed by the upward percolation of water through the concrete, leaving un-bonded coarse aggregate at the bottom of the hole.

Effect of water in “pile hole”

The effect of the depth of water in the “pile hole” prior to commencement of concreting on the strength of the concrete is shown in Figure 4. An exponential curve was fitted to the data.
Figure 4 clearly demonstrates that the free fall of concrete in dry holes did not affect the strength of concrete. However, free-fall casting of concrete into water adversely affected concrete strength. As little as 100 mm of water in the bottom of the "pile hole" resulted in an approximately 50% decrease in the strength of the concrete. For water depths in excess of 200 mm, the concrete strength was reduced by approximately 80%. Further increases in the amount of water in the bottom of the "pile hole" appear to have little effect. This was probably caused by the concrete not being able to absorb the excess water which was carried upwards during pouring of the concrete.

The samples containing crusher dust "spoil" generally achieved higher strengths at particular water depths. This was probably the result of the crusher dust mixing with the concrete and any water present, on impact, as the concrete was poured, hence reducing the formation of voids. However, this was not the case in the samples containing silt "spoil".

With reference to Table 2, tests were performed on cores taken from both the top and bottom of drums 10 and 11. It is interesting to note that the strength of the top and bottom cores from drum 11 differed only by 0.5 MPa. However, in the case of drum 10, the strength of the bottom core exceeded that of the top core by 8 MPa. The latter result was generally expected due to bleeding of the concrete. With the methodology employed, it was not possible to investigate the persistence of this effect up the length of the "pile" shaft or to assess any increase in bleed of the concrete with the increase in the amount of water in the hole.

The inclusion of reinforcing bars as obstructions in the casing appeared to have no effect on the concrete strength, whilst the slow pouring resulted in a slight reduction in strength.

No trend was identified when comparing the results deriving from each of the control samples. The declining trend shown in Figure 4 was not improved when normalising the compressive strength by expressing it as a percentage of the compressive strength of the relevant control sample.

Figure 5 indicates the effect of water on the actual density of concrete in the cores. It is evident from this figure that the actual density of the concrete was, on the whole, adversely affected by the depth of water in the "pile hole".

As is evident from Figure 5, 200 mm of water in the "pile hole" reduced the concrete actual density by approximately 90 kg/m³. Increasing the amount of water to 400 mm had an insignificant additional effect. It is interesting to note that the relatively wide range of densities for water depths of 0 and 50 mm are not reflected in the compressive strength results shown in Figure 4. Furthermore, the results of tests on samples C1, W3 and R2 are "outliers" as they fall out of the band of one standard deviation (s = 80) either side of a linear regression line. There is no apparent reason for these "outliers". The exclusion of the "outlier" samples resulted in an increase in the linear correlation coefficient, $r^2$ (from 0.369 to 0.701). This correlation was not improved by normalising the actual density by expressing it as a percentage of the density of the relevant control sample.

The relationship between actual density and compressive strength of concrete yielded a poor correlation coefficient ($r^2 = 0.2$).

Figure 6 indicates that the percentage excess voids increases with depth of water. The control samples were excluded from this relationship as they were vibrated. An exponential curve was fitted to the data. Referring to Figure 6, for water depths of less than 100 mm, the excess voids were...
Typically less than 2% and are thought to be due to the entrapment of air. However, as the depth of water increased, the excess voids increased to between 10% and 15%. The reason for this trend is that the water at the bottom of the "pile hole", due to its relatively low density, rose into the concrete (during pouring), displacing the mortar and creating voids. Hence, the greater the volume of water in the "pile hole", the greater the volume of excess voids in the concrete.

In contrast to Figures 4 and 5, however, no plateauing was evident with water depths in excess of 200 mm. The increase in voids with increasing water depth is clearly visible in the photograph in Figure 2 where, for water depths of 200 mm and 400 mm in particular, the matrix to the coarse aggregate appears to have been eroded during the drilling operation, giving visual confirmation of the low strength of the paste.

No correlation was found to exist between the percentage excess voids and actual density.

Figure 7 shows the correlation between the depth of water in the "pile hole" and the aggregate–binder ratio. For water depths of less than 100 mm, the average aggregate–binder ratio was of the order of 10. However, this increased to as much as 17 where the concrete was placed through 400 mm of water. This trend is attributable to the upward displacement of mortar, from amongst the coarse aggregates in the lower section of the "pile hole", caused by the upward movement of water through the concrete during pouring. Hence, the more water present, the more mortar displaced.

Table 3: Area of base of "pile" in intimate contact with the bottom of the "pile hole"

<table>
<thead>
<tr>
<th>Test</th>
<th>Test conditions</th>
<th>Percentage base contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Free fall, 50 mm silt, dry</td>
<td>0</td>
</tr>
<tr>
<td>S2</td>
<td>Free fall, 50 mm silt, 100 mm water</td>
<td>60</td>
</tr>
<tr>
<td>S3</td>
<td>Free fall, 50 mm silt, 50 mm water</td>
<td>40</td>
</tr>
<tr>
<td>S4</td>
<td>Free fall, 50 mm c.dust, dry</td>
<td>10</td>
</tr>
<tr>
<td>S5</td>
<td>Free fall, 50 mm c.dust, 50 mm water</td>
<td>60</td>
</tr>
<tr>
<td>S6</td>
<td>Free fall, 50 mm c.dust, 100 mm water</td>
<td>50</td>
</tr>
</tbody>
</table>

Displacement of spoil

As shown in Table 2, the bottom of "pile holes" for tests S1 to S6 contained 50 mm of spoil. By cutting away the bottom of these drums, the percentage of contact between the "pile" concrete and the blinding was estimated and the distribution of spoil was observed.

Table 3 shows the percentage of the area of the base of the "pile" which was in intimate contact with the bottom of the "pile hole", that is the area over which the spoil had been displaced. Higher percentage contacts are more favourable from a founding point of view.

In the case of both the silty spoil material and the crusher dust, casting of concrete onto 50 mm of dry spoil resulted in total separation between the "pile" concrete and the base of the "pile hole". However, the contact area increased to between 40% and 60% in the tests where 50 mm or 100 mm of water was added to the base of the "pile hole" together with the spoil.

Figure 8 shows the contact between the blinding concrete at the base of the drum (representing the in situ founding material) and the "pile" concrete for "piles" cast onto a 50 mm layer of crusher dust at the bottom of the "pile hole". The dry crusher dust (0 mm water – left core sample) was trapped between the "pile" concrete and the bottom of the "pile hole", resulting in a total lack of contact of the "pile" with the founding material. With 50 mm of water in the "pile hole", the crusher dust over the middle of the hole was displaced by the falling concrete and this material was assimilated into the "pile" concrete as a result of the remixing of the concrete as it falls to the bottom of the hole. With 100 mm of water in the hole, the contact over the central portion of the "pile" was visually tight. However, the strength of the "pile" concrete had reduced to 77% of that of the control sample (40.5 MPa to 31 MPa) and
More recent comparative observations of concrete quality in piles placed using free-fall or tremie methods

A number of questions typically arise when the free-fall method of placement is proposed by a contractor, questions which may not adequately be answered by the findings in the early research described above. For example, the engineer might enquire whether the use of tremie techniques might not produce better results. Similarly there might be concerns as to whether the relatively good compaction observed in the cores taken from the free-fall trials would extend right to the interface with the side of the pile excavation.

In an attempt to answer such questions, access was obtained to cores taken from the construction of the widening of Garsfontein Bridge, which was part of the widening of the N1 between Atterbury and Rigel Avenue off-ramps in Pretoria. On this project some of the piles were placed in wet conditions, and in such cases use was made of tremie tubes. The majority of the pile holes were, however, dry and these were placed using free-fall placement.

In all cases, three 76 mm diameter pipes were cast into the concrete for each pile, for subsequent integrity testing. These pipes were tied to the inside of the reinforcement cage and thus were situated about 70 mm from the interface between the pile and the soil. These pipes were sealed at the base with a steel plate to prevent intrusion of concrete. Subsequent to the successful integrity testing with ultrasonic techniques, cores were drilled through the base of one of these pipes per pile, through the concrete below the bottom of the cage and into the end-bearing in situ material as an additional integrity check. Typically these cores were taken about 7 m below ground level. Figure 9 shows details of the integrity testing arrangement of these piles.

Visual inspection was done of a random selection of a number of such lengths of core, both from free-fall-placed, as well as tremie-placed concrete piles. Generally the degree of compaction as assessed from the visible void content for the free-fall placing was at least equal to that of the tremie-placed concrete, although the concrete with the poorest compaction was generally from piles cast using tremie placing. Figure 10 illustrates that the free-fall concrete was as well compacted as the tremie-placed concrete.

CONCLUSIONS

On the basis of the above experimental data, the following conclusions were reached:

- No segregation of the concrete (in the sense of an accumulation of aggregate at the base of the pour) was observed when the concrete was discharged from the truck mixer at a rapid rate, even when the concrete was permitted to impinge on the reinforcing “cage”. Clear signs of segregation were evident when the concrete was poured slowly into 100 mm of water. It appears that the rapid discharge of concrete results in “remixing” of the concrete in the bottom of the “pile hole”.
- Free-fall placement of concrete into dry “pile holes” had no apparent effect on the compressive strength of the concrete compared to that of the four control samples.
- Casting of concrete through 50 mm of water at the bottom of the “pile hole” reduced the compressive strength by an average of approximately 15%.
- Casting of concrete through 100 mm and 400 mm of water in the bottom of the “pile hole” significantly reduced the compressive strength of the concrete by approximately 50% and 80% respectively.
- In addition to having an adverse effect on the strength of the concrete, casting of concrete into more than 100 mm of water was detrimental to the actual density of the concrete, the percentage excess voids and the aggregate–binder ratio.
- As little as 50 mm of dry spoil at the bottom of the “pile hole” negated all direct contact between the “pile” concrete on the underlying founding stratum. Wet spoil was more readily displaced by the concrete, but still resulted in significant reductions in base bearing area mainly around the perimeter of the “pile” base.
- Interruption of the free fall of the concrete by a moderate amount of reinforce- ment appeared to have a negligible effect on the quality of the concrete, provided the rate of pour was reasonable.
- The observations of the relative compac- tion of concrete taken from piles placed...
using free fall versus those placed using tremie techniques, demonstrated that free-fall placing over a range of depths was an acceptable, if not preferable, technique.

On the strength of this limited research, which included concrete mixes of different slumps and compressive strengths, it was concluded that the current practice of free-fall placement of concrete in clean, dry pile holes has no detrimental effect on the quality of the concrete. It is, however, recommended that such techniques should not be used when the depth of water at the bottom of the pile hole exceeds 75 mm.

ACKNOWLEDGEMENTS

The research was carried out under the auspices of the then Research and Development Advisory Committee of the South African Roads Board. The fieldwork was sponsored by specialist geotechnical contractors Esor. Both sponsorships are gratefully acknowledged.

REFERENCES


INTRODUCTION

Engineering skill and knowledge of physical laws are required for the design of cost-effective de-aeration facilities for water conveyance systems. Misconceptions regarding the operation of air valves and their positioning have led to ineffective de-aeration of water pipeline systems. An important aspect of air valve installation that has been largely neglected is the dimensioning of the discontinuities in the pipe at which free air is intercepted, conveyed to the air valve and released under operating conditions. The inability of discontinuities to intercept air can result in situations where free air remains trapped in the pipeline and reduces the hydraulic capacity to far less than the design capacity. Another possible consequence of the inability to intercept free air is the uncontrolled release of the air that could lead to high dynamic pressures.

There are basically two types of air release valves, the so-called large- and small-orifice air valves. The differences between the two types are the operating conditions and manner in which they function. Large-orifice valves release and admit air during the charging and draining of water mains respectively, while small-orifice air valves mainly release air under operating conditions. Air valve manufacturers produce a range of pressure classes, functional layouts and sizes. Effective de-aeration of a pipeline requires that:

- Free air is transported hydraulically in the pipeline to a position where it can be released.

EXPERIMENTAL SETUP

The purpose of the experimental investigation was to determine a relationship between the operational parameters of the pipeline and the size of the discontinuity required to intercept free air that is hydraulically transported. Tests were conducted to optimise the following parameters that influence the efficiency of air interception:

- Mean velocities of the fluid and air bubbles
- Relative air bubble sizes
- Pipe slope
- Size of the discontinuity (off-take) that leads to the air valve.

The experimental setups consisted of pipelines that were supplied with water from a constant head tank with the flow rate regulated by a gate valve. Two transparent pipes are-aeration of pipelines, required size of the discontinuity to capture free-moving air, air accumulator
of different diameters were tested: Test 1 comprised a 110 mm nominal diameter (ND) PVC pipe and Test 2 a 160 mm ND PVC pipe. Figure 1 shows a schematic layout of the experimental setups and highlights some differences between the setups for Test 1 and Test 2.

**Test 1 experimental setup**

110 mm ND pipe

The 110 mm pipeline supplying water to the experimental setup was fitted with an ultrasonic flow meter. The 110 mm diameter transparent PVC pipe made it possible to track the movement of the air bubbles from the point of injection to the discontinuity where a Perspex air valve was placed on a stub, as indicated in Figure 2. Tests were conducted with various discontinuity (orifice) sizes drilled into the pipe to intercept the air.

A defined quantity of air was introduced into the system through the custom-built “Air Induction Box” (AIB) shown in Figure 3. The AIB had the facility to set the bubble size (BUB), based on a specific pressure difference, by setting the opening time of a solenoid valve in milliseconds. The time delay between consecutive air bubbles could also be altered (GAP), as well as the number of air bubbles injected (QTY) for a specific test. Figures 4a to 4c give a quantitative indication of the various air bubble sizes that were used in the Test 1 setup (defined as small, medium and large measured in normal millilitres).

As indicated in Figure 1, a collector box was placed at the end of the pipeline to capture the air that bypassed the discontinuity. The volume of air intercepted by the discontinuity was calculated as the mass balance difference between the volume introduced and the volume captured by the interceptor box.

In the Test 1 setup, the pipe gradient was varied between 0° (horizontal) and 15° (downwards in the direction of flow) in steps of 2.5°. For each of the gradients, the size of the discontinuity was varied between 10% and 35%

---

**Figure 1: Schematic layout of the experimental setup**

**Figure 2: Discontinuity and air valve on 110 mm pipeline for Test 1 experimental setup**

**Figure 3: Air Induction Box (AIB) assembly**
of the 110 mm pipeline’s diameter in steps of 5%. The flow velocities investigated for each of the setups were 0.5 m/s, 1.0 m/s and 1.5 m/s.

Test 2 experimental setup (160 mm ND pipe)
The layout of Test 2 was similar to that of Test 1, except that a 160 mm diameter transparent PVC pipe was used. A similar series of tests was conducted for various discontinuity diameters, pipe slopes, air bubble sizes and flow velocities to determine the effectiveness of interception of the air for a range of operating conditions.

For Test 2 the capturing device shown in Figure 5 replaced the air valve of the Test 1 experimental setup. This comprised a 20 cm long equal Tee closed by the discontinuity. Discontinuity sizes of 20 mm and 67.8 mm were tested. This change in the layout of the discontinuity made the Test 2 layout more representative of the layout of normal air valve installations.

After each test the captured air was released in a controlled manner into a measuring container in which the (normal) volume of air could be determined.

The main differences between the Test 1 and Test 2 experimental setups were:
- The pipeline diameter was increased from 110 mm for Test 1 to 160 mm for Test 2.
- The discontinuity for Test 1 functioned as an orifice and was changed to a Tee off-take in the experimental layout of Test 2. This change represents the typical layout of an air valve installation.
- For the Test 1 experimental setup the air that passed the discontinuity was captured and the portion of the induced air that was intercepted was calculated. For the Test 2 experimental setup, the volume of air that was intercepted was captured and measured directly.

Three different bubble sizes were evaluated with the Test 2 experimental setup. The AIB was programmed to induce a certain size bubble by setting the differential pressure and the time that the solenoid valve remained open to allow a certain volume of air into the system. Tables 1 and 2 contain details of the AIB settings for the different air bubble sizes that were tested in the Test 1 and Test 2 experimental setups.

The variables that were considered in the experimental setups were the air bubble size, the water flow velocity, the pipe gradient and the size of the discontinuity.

The pipe gradients of the pipe in the tests conducted for the Test 2 experimental setup varied from $0^\circ$ to $15^\circ$ in steps of $2.5^\circ$. The flow velocities investigated were 0.5 m/s, 1.0 m/s and 1.5 m/s. The sizes of the discontinuity that were evaluated were 20 mm and 67.8 mm, i.e. 13% and 44% of the internal diameter of the 160 mm pipeline.

Since no literature could be sourced that defines the required size of discontinuity to intercept transported air, the Test 2 experimental setup was also modelled with a numerical Computational Fluid Dynamics (CFD) model. The results of the tests on the two experimental setups and of the numerical modelling are compared in this paper.

**NUMERICAL MODELLING (COMPUTATIONAL FLUID DYNAMICS)**
The objective of the numerical modelling was to determine the volume of air intercepted at a discontinuity for comparison with the results obtained from the experimental physical modelling. The commercial software package, FLUENT, which is widely used for incompressible and compressible, laminar and turbulent fluid flow problems, was used.

The modelling of multi-phase flow (air and water) was simulated by incorporation of the Volume of Fluids Model (VOF), which assumes that the different fluids (or phases) are not interpenetrating. The fields for all variables and properties are shared by

---

### Table 1 Bubble sizes introduced into the 110 mm pipe for experimental setup Test 1 (photos of all three bubble sizes are shown in Figures 4a to 4c)

<table>
<thead>
<tr>
<th>Description</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIB settings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bubble size (ms)</td>
<td>25</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Pressure (Bar)</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Quantity</td>
<td>400</td>
<td>200</td>
<td>90</td>
</tr>
<tr>
<td>Pressure setting</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Regulator in use</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated bubble size (nml)</td>
<td>25</td>
<td>75</td>
<td>490</td>
</tr>
</tbody>
</table>

### Table 2 Bubble sizes introduced into the 160 mm pipe for experimental setup Test 2

<table>
<thead>
<tr>
<th>Description</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIB settings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bubble size (ms)</td>
<td>25</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Pressure (Bar)</td>
<td>3.5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Quantity</td>
<td>400</td>
<td>200</td>
<td>90</td>
</tr>
<tr>
<td>Pressure setting</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Regulator in use</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated bubble size (nml)</td>
<td>61</td>
<td>221</td>
<td>547</td>
</tr>
</tbody>
</table>
Finally, during the period when the air pocket was introduced, a steady state solution in the pipeline was initialised prior to the introduction of any air pockets.

Thereafter, an air pocket was introduced into the flow field. This represented the unsteady flow condition and was solved over defined time steps, which were in the order of 0.0005 seconds for the small bubbles and in the order of 0.076 seconds for larger bubbles.

The unsteady flow computation continued with similar time steps, until such time as the air pocket reached the discontinuity/outlet. During this time, the bubble would assume an equilibrium state, influenced by the mean flow velocity and buoyancy forces, introduced on account of the inclination of the gravity vector.

Finally, during the period when the air pocket would start to rise into the pipe connected to the discontinuity, smaller time step sizes were implemented. These smaller time step sizes ensured that the solution at each time step size converged correctly in accordance with certain criteria. Computation was slower in this step, due to two major factors, namely the use of tetrahedral grid cells, and an increase in the number of free surfaces due to turbulence and air bubble break-up into smaller bubbles in the region of the discontinuity. Time step sizes for this step in the simulation ranged between a high of 0.0002 seconds and a low of 0.00005 seconds. Air intercepted at the discontinuity would continue to move up the "riser" pipe into the small box, while un-intercepted air pockets would continue past the discontinuity and would be released via the outlet boundary.

After the initial CFD modelling had been completed, further analyses were conducted in which the geometry and the grid generation were altered. In these evaluations, which are reviewed below, and thereafter the numerical modelling results are discussed.

### COMPARATIVE RESULTS
Firstly, the experimental results for Test 1 (110 mm pipe) and Test 2 (160 mm pipe) are reviewed below, and thereafter the numerical modelling results are discussed.

#### Results of Test 1 setup for 110 mm ND pipeline
The experimental Test 1 results indicated that the effectiveness of a specific discontinuity is inversely proportional to the slope and the flow velocity. The results suggested that a discontinuity opening (d) of at least 35% of the pipe diameter (D) is required to provide an air removal efficiency of more than 95% for negative slopes up to 15° and a maximum flow velocity of 1.5 m/s.

Table 5 illustrates the efficiencies of air removal for a discontinuity opening (d) equal to 35% of the pipe diameter (D), a flow velocity of 1.0 m/s and a medium air bubble (75 nml).

The results of the Test 1 experimental setup for velocities of 1.0 and 1.5 m/s are presented graphically in Figure 7, which clearly indicates that for a constant velocity the efficiency of air removal decreases as the pipeline gradient is increased. It was also established that the efficiency reduces to improve the formation of the flow patterns before reaching the discontinuity. A further geometrical alteration was made for a single case to model the Test 1 experimental setup where the discontinuity was an orifice.

For Phase 2, hexagonal grid cells were generated with the objective of improving the accuracy and reducing the solution time. An effort was also made to ensure cell volume uniformity, as this is considered important when performing free surface simulations. The cell count for the simulations ranged from 350 to 400 thousand grid cells for the entire volume of the pipe modelled.

Table 4 shows the details of the different analyses undertaken as part of the Phase 2 numerical modelling.

### Table 3 Fluid properties for numerical modelling with FLUENT

<table>
<thead>
<tr>
<th>Fluid property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (water) (kg/m³)</td>
<td>1.000</td>
</tr>
<tr>
<td>Density (air) (kg/m³)</td>
<td>1.225</td>
</tr>
<tr>
<td>Viscosity (water) (kg/m-s)</td>
<td>0.001003</td>
</tr>
<tr>
<td>Viscosity (air) (kg/m-s)</td>
<td>1.7894 e-05</td>
</tr>
<tr>
<td>Surface tension (N/m)</td>
<td>0.0728</td>
</tr>
<tr>
<td>Operating pressure (kPa)</td>
<td>187</td>
</tr>
</tbody>
</table>

### Table 4 Numerical analyses conducted with FLUENT

<table>
<thead>
<tr>
<th>Slope (degrees)</th>
<th>Off-take diameter as % of pipe diameter *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35%</td>
</tr>
<tr>
<td>5°</td>
<td>–</td>
</tr>
<tr>
<td>10°</td>
<td>–</td>
</tr>
<tr>
<td>15°</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: * The outlet from the pipe was modelled as a pipe.

the phases and represent volume-averaged values. Thus the variables and properties in any given cell are either purely representative of one of the phases, or representative of a mixture of the phases, depending upon the volume fractions.

The geometry of a section of the experimental setup was generated in FLUENT’s GAMBIT pre-processor, as shown in Figure 6 (Gambit 2001). It was assumed that the model is symmetrical along a vertical plane in the direction of flow, and that the swirl effects are negligible. The geometry shown in Figure 6 represents a pipeline length of 1 000 mm, a pipeline diameter of 110 mm and a discontinuity diameter equal to 35% of the pipeline diameter.

The effectiveness of the numerical simulation is largely influenced by the selected structure of the finite element grid. The grid was divided into quadrilateral and tetrahedral type cells. Cell density in the region of the discontinuity was higher to accommodate the expected high velocity gradients in the flow field. This boundary layer grid structure contributed to an accurate solution of the flow profile near the wall, where a zero velocity or no-slip condition exists at this boundary.

The boundaries used in this model consisted of the symmetrical vertical boundary plane which cuts the pipe in two halves, and the inlet- and outlet-boundaries. The fluid properties for the air and water phases are reflected in Table 3.
for larger air bubbles on account of higher turbulence at the nose of the bubble, which creates smaller bubbles that are torn from the larger air bubbles. The small bubbles are seldom transported at the crown of the pipeline and hence reduce the efficiency of interception of these small air bubbles.

It was visually observed that smaller bubbles enter the discontinuity (orifice) more easily than the larger bubbles, due to the tendency of larger bubbles to block the discontinuity, making it impossible for the liquid to be displaced by the smaller air bubbles.

Results of Test 2 experimental setup (160 mm ND pipeline)

In the Test 2 experimental setup the discontinuity was modelled as a “standpipe or riser pipe” which closely resembles the normal installation of air valves.

The results for a discontinuity of 67.8 mm (44% of the pipe diameter) and a small air bubble (25 nml) for velocities of 1.0 and 1.5 m/s are graphically illustrated in Figure 8.

Numerical modelling results

Table 6 provides the results of the numerical modelling of unsteady flow in a 110 mm diameter pipe with a large air bubble (240 nml), a flow velocity of 1.2 m/s and various slopes. The table shows the air capture efficiency.

<table>
<thead>
<tr>
<th>Slope (degrees)</th>
<th>Percentage air captured (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5°</td>
<td>93.1%</td>
</tr>
<tr>
<td>10°</td>
<td>71.1%</td>
</tr>
<tr>
<td>15°</td>
<td>77.3%</td>
</tr>
</tbody>
</table>

Note: * The outlet from the pipe was modelled as a pipe

Figure 7 Effectiveness of discontinuities in intercepting medium air bubbles (75 mm) hydraulically transported in the 110 mm pipeline for experimental setup Test 1

Figure 8 Effectiveness of discontinuities in intercepting small air bubbles (61 ml) hydraulically transported in the 160 mm pipeline for experimental setup Test 2
Table 7 shows the numerical modelling results for various off-take ratios (d/D) and slopes where the discontinuity was modelled as an orifice similar to that of the experimental layout of Test 1. Comparing the results for the off-take ratio of 0.35 it is evident that the discontinuity, which was modelled as a stub pipe (Table 6), would intercept air bubbles more effectively than the layout modelled as an orifice (Table 7).

The presence of air in a pipe results in the unsteady flow of air and water in the pipe flow, which was simulated with the numerical model FLUENT. Figure 9 indicates the variations in time of the proportions of water and air when small and large air bubbles are transported past a particular cross-section of the pipe. The results in Figure 9 are shown for three slopes and for an average water flow velocity of 1.2 m/s. The speed at which large and small air pockets pass a particular cross-section can also be inferred from Figure 9. Figure 9 indicates the decrease in the velocity of air pocket movement down the pipe with increasing pipeline slope. The volume–fraction–time relationship represents the changing shape of the bubbles due to the increasing effects of buoyancy and pipeline slope. The increase in small air pocket breakaway from the fronts of the large air pockets with increasing slope is also reflected in Figure 9.

Comparison of results of the numerical and experimental modelling

Figure 10 shows a comparison of the experimental laboratory results with the results from the CFD modelling. The air bubble sizes are indicated in Figure 10 by the (circle) sizes.

It seems that the efficiencies of air removal determined by the numerical modelling (CFD) are lower than the efficiencies measured in the experimental modelling. This difference can perhaps be attributed to the different sizes of the air bubbles which were used for this comparison.

CONCLUSIONS

The experimental work indicated that air bubbles do not always travel at the top of the pipe, and that a disturbance can create turbulence which breaks up the air pockets and mixes the air throughout the entire cross-section. This reduces the efficiency of interception of air by the discontinuity, since fewer of the air bubbles have the opportunity to enter the discontinuity.

Tests conducted on the 160 mm pipe (Test 2) indicated that a discontinuity of more than 44% of the pipe diameter was required to intercept the hydraulically transported large air bubble (490 nml) in the pipeline for a slope of less than 15° and an average flow velocity of 1.5 m/s.

The experimental results also indicated that the following aspects should be considered when a de-aeration system is designed:

- Firstly, the size of the discontinuity should be sufficient to intercept the air bubbles, and
- Secondly, the air which is intercepted at a discontinuity should be collected in an “accumulator” from where it is released via an air valve.

All tests on discontinuities were undertaken with negative pipe slopes. In the case of positive slopes the bubble velocity would be much greater and hence for a specific discontinuity a smaller portion of the air would be captured. Therefore a larger size discontinuity would be required for positive pipe slopes.

The complex nature of air movement in pipelines and of the factors influencing efficient air removal suggest that a conservative approach should be used when dimensioning a discontinuity to effectively intercept air bubbles in a pipeline.

RECOMMENDATIONS

Based on the understanding of the complexities associated with two-phase water and air flow, it is recommended that the experimental results are conservatively applied within the practical and financial constraints of water pipeline design. Accordingly the following recommendations are made for sizing a discontinuity for effective de-aeration during pipeline operation:

- The minimum discontinuity required for small pipes diameters (D ≤ 300 mm) should be set equal to the diameter of the pipe. An equal T-piece is a standard pipe fitting for these diameters.
- For diameters between 300 mm and 1 500 mm the discontinuity should be...
For pipes with diameters in excess of $\pi d$

Figure 11: Graphical representation of installation details for small orifice air valves

<table>
<thead>
<tr>
<th>Variable</th>
<th>Small orifice function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>As topography requires (Van Vuuren et al 2004b)</td>
</tr>
<tr>
<td>$h \geq 1.0D$</td>
<td></td>
</tr>
<tr>
<td>$h \geq 150$ mm</td>
<td></td>
</tr>
<tr>
<td>$d = D$ for $D \leq 300$ mm</td>
<td></td>
</tr>
<tr>
<td>$d = 0.6D$ for $300 &lt; D \leq 1500$ mm (minimum 300 mm)</td>
<td></td>
</tr>
<tr>
<td>$d \geq 0.35D$ for $D &gt; 1500$ mm (minimum 900 mm)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12: Graphical presentation of installation details for large orifice air valves

<table>
<thead>
<tr>
<th>Variable</th>
<th>Large orifice function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential pressure for the calculation of the outlet orifice</td>
<td>$P_a - P_i \leq 40$ kPa $P_i - P_a \leq 5$ kPa*</td>
</tr>
<tr>
<td>Location</td>
<td>Between shut-off valves or to allow free draining</td>
</tr>
<tr>
<td>$h \geq 1.0D$</td>
<td>$h \geq 150$ mm</td>
</tr>
<tr>
<td>$h \geq 1.0D$</td>
<td>$h \geq 150$ mm</td>
</tr>
<tr>
<td>$d$</td>
<td>$\frac{mdh}{4} \geq \frac{mdh}{4}$</td>
</tr>
</tbody>
</table>

Note: * To prevent dynamic closure of the large orifice. Refer to manufacturers details.

Figure 13: Air valve arrangement to ensure sequential closing and the provision of an air cushion

ACKNOWLEDGEMENTS
The authors wish to thank the Water Research Commission for funding the research project, as well as the two students who contributed to the study (Miss A Kotze and Miss M du Toit).

Input from personnel at the CFDlab, University of Pretoria, who assisted with the CFD modelling (Dr D de Kock and Mr T Kingsley) is appreciated. The authors also wish to thank Dr Mike Shand, who provided valuable contributions to improve this paper.

BIBLIOGRAPHY
Definition and application of a cohesive crack model allowing improved prediction of the flexural capacity of high-performance fibre-reinforced concrete pavement materials

E Denneman, E P Kearsley, A T Visser

In conventional concrete pavement design methods the design parameters are determined using linear elastic analysis. Concrete is subject to significant size effect and as a result linear elastic design concepts, such as the modulus of rupture determined for a beam, have limited reliability in the design of elements of different size and geometry. The objective of this paper is to demonstrate that, in contrast to the modulus of rupture, fracture mechanics material parameters can be used to accurately and precisely predict the flexural capacity of elements of a different size and geometry. The experimental framework includes two high-performance fibre-reinforced concrete mix designs, used to produce beams of different sizes tested in three-point bending configuration, as well as centrally loaded round panels. The fracture energy of the material is determined from the flexural beam tests. An adjusted tensile splitting test procedure is used to determine the tensile strength. The flexural tests on the beams and panels are simulated numerically using two finite element implementations of a cohesive crack approach. The numerical simulation yields satisfactory prediction of the flexural behaviour of the beam and disk specimens. It is concluded that using a fracture mechanics approach, the flexural behaviour of structural elements of different size and/or geometry can be reliably predicted.

INTRODUCTION

South Africa boasts an extensive and mature road network. At present the bulk of pavement design activities are aimed at preserving and upgrading the existing road infrastructure. Innovative methods of pavement rehabilitation are required to increase the service life of wearing courses and to reduce the need for traffic hampering maintenance activities. To this aim, the South African National Roads Agency Limited (SANRAL) has sponsored the development of the ultra-thin continuously reinforced concrete pavement (UTCRCP). UTCRCP is intended as an overlay strategy for existing roads. The innovative UTCRCP system are currently based on conventional mechanistic-empirical concrete pavement design methodologies. These methodologies make use of linear elastic (LE) mechanics to determine the stress in the pavement slab. In these models the material strength is characterised by the modulus of rupture (MOR), using a standard test method such as SANS 5864:2006. The MOR is also calculated under the assumption of LE material behaviour. The non-linear, non-elastic post-fracture behaviour of concrete is not taken into consideration. The ratio between the MOR and the stress in the pavement is used to predict the fatigue life of the pavement.

The first objective of this paper is to demonstrate that the fibre-reinforced concrete material under study exhibits a strong size effect due to its high post-crack stress capacity and that this limits the reliability of the MOR obtained for a specific specimen size and geometry, as a predictor of the peak load of elements of a different size and or geometry.
The second objective is to show that, in contrast to the MOR, fracture mechanics material parameters can be used to accurately and precisely predict the peak load, and importantly, the post-peak flexural behaviour of elements of different sizes and geometry.

Theoretical background on the size effect observed in flexural tests on concrete elements is provided in the next section. Following that, a cohesive crack fracture mechanics model for the FRC material is introduced. The material parameters required to define the model are obtained from experiments on beams and cylinders, as discussed in the section on the experimental work performed for this study. The final fracture mechanics model, defined using test results for a single beam size, is used to numerically predict the flexural performance of beams of different sizes, as well as of centrally loaded round panels. The main contribution of this paper is the generalisation of the cohesive crack model for FRC, previously used in two-dimensional space to simulate experiments on beams, to three-dimensional space for tests on centrally loaded disks. The paper presents a complete methodology, simple, yet effective, to determine the fracture parameters of FRC, and applies these to predict the flexural behaviour FRC structural elements.

### SIZE EFFECT
The MOR, or flexural strength, is a design parameter often used in civil engineering designs. The MOR is obtained from flexural tests on beams and represents the stress in the extreme fibre of the specimen, calculated under the assumption that an LE stress distribution is present at the peak load condition. It is assumed that this material strength parameter obtained from laboratory experiments can be reliably generalised to predict the failure stress in full-size structural elements. Researchers have, however, long established that for concrete the MOR is not a true material property, because its value changes with specimen size (Reagel & Willis 1931; Kellerman 1932). In flexure tests, large beams fail at lower maximum tensile stress than small beams of the same material.

Studies at the University of Pretoria have shown the high-performance fibre-reinforced concrete material used in UTCRCP to have significantly increased post-crack load carrying capacity when compared to plain Concrete.
concrete (Kearsley & Elsaigh 2003; Elsaigh 2007). Subsequent studies found the MOR value of the fibre-reinforced concrete material to be subject to significant size effects (Denneman et al 2010a,b).

To illustrate the size effect, the average load-displacement \((P – \delta)\) curves of three-point bending (TPB) tests, performed by Denneman et al (2010b) on specimens of different sizes, are shown in Figure 1a. The TPB test configuration is shown in Figure 2a. The specimens have similar geometry, i.e. the span \((a)\) to height \((b)\) ratio, and span to notch depth \((a)\) ratio are kept constant. To explore the size effect in the experiments, the nominal stress \(\sigma_{N1}\) in the section is calculated assuming a linear elastic stress distribution using Equation 1:

\[
\sigma_N = \frac{3Ps}{2b(h – a)^2}
\]

where \(P\) is the load, and \(b\) the beam width.

The stress at the peak load condition \(\sigma_{N1}\) represents the flexural strength, or MOR, parameter. An appreciation of the size effect can be obtained by plotting \(\sigma_{N1}\) against the deflection \(\delta\) as a ratio of the effective beam height \((h)\) as shown in Figure 1b. The figure shows the effect of stress not only on the peak stress, but also on the post-crack behaviour of the material.

The modulus of rupture is typically determined from an FPB test configuration, which is shown in Figure 2b. For this configuration \(\sigma_{N}\) is obtained from:

\[
\sigma_N = \frac{Ps}{bh^2}
\]

The effect of size on MOR values in FPB tests on high-performance fibre-reinforced concrete material is shown in Figure 3.

It is evident that MOR results are highly dependent on the height of the tested specimen. The main implication of the findings from size-effect studies is that the MOR is unsuitable as a design parameter; as the results obtained for a certain specimen size cannot be used to reliably predict the peak load of a specimen with the same geometry, but a different size. Much less can it be used to predict the flexural behaviour of elements of a different geometry.

The study presented in this paper is limited to high-performance fibre-reinforced concrete pavements. The findings are, however, relevant to plain concrete pavements as well, as plain concrete exhibits size effect similar in magnitude (Denneman 2011).

The main source of the size-effect phenomenon, is the fracture mechanics size effect (Bažant & Planas 1997). The fracture mechanics size effect is caused by the fact that concrete is a quasi-brittle material, and at the peak-load condition cracks would already have formed in the specimen. Due to the presence of a crack, LE stress distribution exists in the beam specimen. Different sizes of specimens, different amounts of fracture energy are released into the crack front, giving rise to the observed size effect.

Similarly, an LE stress distribution will not be present in a pavement slab loaded to failure, because it, too, would have cracked. Size effect in plain concrete has been well documented and can be predicted using fracture mechanics (Bažant & Planas 1997). In this paper a cohesive crack model is used to improve the prediction of the flexural behaviour of FRC elements.

DEFINING A COHESIVE CRACK FUNCTION

The complex behaviour of FRC composites calls for the use of advanced damage models. This has been the topic of many studies internationally. Local research in this field was performed by Kearsley & Elsaigh (2003), Boshoff & Van Zijl (2007), Shang & Van Zijl (2007), Elsaigh (2007) and Van Zijl (2009).

A fracture mechanics method for the analysis of crack propagation in concrete, favoured by various researchers for implementation in finite element analysis, is the fictitious crack model. The model, nowadays commonly referred to as the cohesive crack model, was introduced by Hillerborg et al (1976). According to the cohesive crack model, the material behaves linear elastically until the tensile stress reaches the tensile strength of the material. At this point a crack is induced. After crack nucleation, stresses are still transferred over the crack according to a softening relation. The crack bridging stress \(\sigma\) is written as a function of the crack width \(w\):

\[
\sigma = f(w)
\]

For plain concrete a bi-linear shape for the softening function is often used. For the more complex softening of fibre-reinforced concrete, tri-linear softening functions have been proposed (Lim et al 1987; Pereira et al 2004; RILEM 2003). Denneman et al (2011a) proposed a softening function which combines crack tip singularity with exponential softening. The softening behaviour is shown schematically in Figure 4. The softening function was proposed based on evaluation of direct tensile test results performed on material with similar fibres by Lim et al (1987). The softening model seeks to simulate the initial rapid reduction of stresses transferred across the crack as a crack is formed in the cement aggregate matrix. As the crack width increases, the steel fibres are activated. The point at the base of the crack tip singularity, where the fibres are activated, is represented in Figure 4 by stress \(\sigma_1\) and crack width \(w_1\). The values of \(\sigma_1\) and \(w_1\) are obtained through calibration. The function is defined by the following equations:

\[
\sigma = f_1 \left( \frac{\sigma - \sigma_1}{w_1} \right) \quad \text{for} \quad 0 < w < w_1
\]

\[
\sigma = \sigma_1 \exp \left( -\frac{\sigma}{G_{f1}} - (w - w_1) \right) \quad \text{for} \quad w_1 < w < \infty
\]

with

\[
G_{f1} = G_f \left( \frac{f_1 + \sigma_1}{2} \right) w_1
\]

where \(G_f\) is the specific fracture energy of the material, which is equal to the area under the softening curve, and \(f_1\) is the tensile strength, representing the stress at which a crack is formed, of the material need to be determined.
A method to obtain $G_f$ for fibre-reinforced concrete from TPB tests was developed by Denneman et al. (2011a). A best estimate of the tensile strength is determined using an adjusted tensile splitting test as described by Denneman et al. (2011b). A summary of the methodologies to determine $G_f$ and $f_t$ is provided in the following two sections.

Determining specific fracture energy

In order to determine $G_f$ from TPB tests, the work of fracture ($W_f$) required to completely break the beam specimen needs to be determined. Figure 5a shows the load-displacement curve for a TPB test; the area underneath the curve represents $W_f$. TPB tests on fibre-reinforced concrete specimens will invariably be stopped short. Near the end of the test the crack would have grown to the top of the beam. However, not all fibres are completely pulled out, and therefore not all work of fracture has been recorded. In the standard test configuration it is physically impossible to run the test up to the high deflection required to pull out the fibres at the top of the beam. To determine $W_f$ the load-displacement tail would have to be extrapolated as shown in Figure 5a. Denneman et al. (2011a) proposed a method to model the tail of the curve, based on an earlier methodology proposed for plain concrete by Elices et al. (1992).

The methodology makes use of the kinematic model in Figure 5b. It was shown that at large deflections, when the crack has propagated to the top of the beam, the two halves of the beam act as rigid parts, rotating around a hinge point. The angle of rotation ($\phi$) of the individual parts around the hinge point is a function of the deflection ($\delta$) at the centre of the span. The missing part of the tail can be modelled by determining the remaining moment capacity around the hinge point using the cohesive crack relation. Denneman et al. (2011a) showed that the work of fracture under the missing tail ($W_{tail}$) can be determined from:

$$W_{tail} = \frac{bA}{4\delta_{end}}$$  \(7\)

where $\delta_{end}$ is the deflection at the last recorded data point.

$A$ is a parameter corresponding to the slope of a graph plotting the moment due to external loading ($M$) divided by the width of the beam ($b$) against $2\delta^2$. At large deflections, parameter $A$ becomes a constant as shown in Figure 6 for TPB results on specimens of different sizes tested as part of this study. $W_{tail}$ which may represent up to 20% of the total area (Denneman 2011) is added to the area under the recorded part of the curve to obtain the total $W_f$ required to completely break the specimen. From $W_f$, the fracture energy dissipated per unit fractured surface can be calculated:

$$G_f = \frac{W_f}{b(h - a)}$$  \(8\)

The value of $G_f$ thus obtained represents the area under the softening function in Figure 4. The remaining material parameter to be determined for the definition of the softening function is the tensile strength of the material.

Determining the tensile strength

In standard test methods for the cylinder splitting test, such as the ASTM C496 (ASTM 2008a), the maximum tensile strength is calculated from the peak load in the tests using the continuum mechanics solution for a circle loaded with two equal and opposed point loads, offered by Timoshenko & Goodier (1970). However, the loading is actually introduced to the specimen by means of loading strips with a certain width instead of through load points. Tang (1994) corrected the linear elastic plane stress solution for the stress distribution in the specimen for the effect of the width of the load strips.

Denneman et al. (2011b) showed that, due to the high post-crack stress capacity of FRC, the ultimate peak load ($P_f$) recorded in the splitting test is not related to the assumed uniform
LE tensile stress condition along the loading plane. Through measurement of the transversal deformation of the specimen during the test it was found that the load-deformation curve has two distinct peaks, as shown schematically in Figure 7c. The first peak ($P_I$) is related to principal crack formation along the loading plane, as shown in Figure 7a. After this initial crack has formed, stresses in the specimen redistribute and once again load increases until secondary cracking takes place at the edges of the specimen at $P_u$, as shown in Figure 7b.

For the experiments reported on in this paper, the transversal deformation of the cylinders was measured using linear variable displacement transducers (LVDTs) on fixtures mounted to the base plate, as shown in Figure 8. Measurements were taken at the centre line of the specimen at the centre of its length. In later experiments, the transversal deformation was measured using LVDTs mounted to datum points which were glued to the specimens at either end of the cylinder (Denneman et al 2011b).

If the transversal deformation is measured in this way, the tensile strength $f_t$ can be obtained using Equation 9, which includes the correction for the width of the load strip proposed by Tang (1994).

$$f_t = \frac{2P_I}{\pi D} \left[ 1 - \left( \frac{b_1}{D} \right)^2 \right]^{\frac{1}{2}}$$

where $D$ is the diameter of the specimen, $b_1$ is the width of the load strip and $P_I$ is the initial peak load as identified from the load-transversal deformation curve.

![Figure 6 Determining parameter A for TPB specimens](image)

![Figure 7 (a) Principal tensile crack formation; (b) Secondary cracking; and (c) Schematic load-transversal deformation curve](image)

![Figure 8 Tensile splitting test with LVDTs to monitor transversal displacement](image)

**Table 1 Mix components by mass**

<table>
<thead>
<tr>
<th>Component</th>
<th>Type</th>
<th>Mix A (kg/m$^3$)</th>
<th>Mix B (kg/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cement</td>
<td>Cem I 42.5 R</td>
<td>450.3</td>
<td>448.0</td>
</tr>
<tr>
<td>Coarse aggregate</td>
<td>Quartzite (4.75 mm – 6.7 mm)</td>
<td>930.6</td>
<td>925.9</td>
</tr>
<tr>
<td>Fine aggregate</td>
<td>Quartzite (0.00 mm – 4.75 mm)</td>
<td>725.5</td>
<td>721.8</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td>170.7</td>
<td>169.8</td>
</tr>
<tr>
<td>Steel fibres</td>
<td>Bekaert Dramix (30 mm x 0.5 mm)</td>
<td>80.1</td>
<td>119.5</td>
</tr>
<tr>
<td>Polypolyene fibre</td>
<td>(12 mm)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Admixture</td>
<td>P100</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Admixture</td>
<td>O100</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Silica fume</td>
<td>Witbank</td>
<td>65.0</td>
<td>64.7</td>
</tr>
<tr>
<td>Fly ash</td>
<td>Lethabo</td>
<td>80.1</td>
<td>79.6</td>
</tr>
</tbody>
</table>
Through numerical simulation of the experiments Denneman *et al.* (2011b) confirmed that a close estimate of the true tensile strength of FRC may be obtained from this adjusted tensile splitting test procedure. The tensile splitting methodology for fibre-reinforced concrete proposed by Denneman *et al.* (2011b) provides a relatively simple alternative to the more complex direct tensile testing approach. Direct tensile testing yields more detail on the post-cracking behaviour of the composite material, but if only a measure of the tensile strength is required, the presented tensile splitting methodology will suffice.

**EXPERIMENTAL WORK**

The fracture experiments for this study were performed at the University of Pretoria. The components by mass for the concrete mix designs are shown in Table 1. The designs are typical for the material used in UTCRCP. One mix was prepared with 80 kg/m³ steel fibres, the other with a steel fibre content of 120 kg/m³.

For the TPB tests, beam specimens of various sizes and geometry were prepared from both mix designs. To investigate fracture in a three-dimensional test, a centrally loaded round panel on three support points was used. The panels were cast in two different sizes. The test configuration for the tests on panels is shown in Figure 2c. The dimensions of the beam and panel specimens are shown in Table 2.

The procedure for TPB to determine fracture properties as recommended by RILEM technical committee 162-TDF (RILEM 2002) was used as the point of departure for the TPB tests. Besides the standard recommended beams of 150 x 150 mm² cross section with a 550 mm length and a 25 mm notch, a number of other specimen sizes and geometries were used. Specimens with and without a notch were included to investigate the suitability of the eventual numerical models for fracture simulation for cases with and without a pre-formed crack.

During the TPB tests the vertical displacement at mid-span was recorded by means of LVDTs at either side of the beam. The reference frame for the displacement was mounted at half the height of the beam specimen. Mid-span displacement was measured relative to reference points above the supports.

The tests on the concrete panels were performed in accordance with ASTM standard test method C 1550 – 05 (ASTM 2005). In this test the load-displacement response of a centrally loaded concrete panel supported on three pivot points is recorded. The vertical displacement is measured with an LVDT placed under the disk in line with the position of the loading point at the top of the disk.

All tests were run in displacement control, at the loading rates prescribed in the respective standard test methods.

Compressive strength tests were performed on the material in accordance with British Standard BS 1881 (BSI 1983). The static modulus of elasticity (\(E\)) and Poisson’s ratio (\(\nu\)) were determined in accordance with the standard procedures contained in ASTM C469-02 (ASTM 2008b). The average results for \(f_c\) and \(E\) and \(\nu\) are shown in Table 3. The table also shows the value of \(f_t\) determined in accordance with the adjusted procedure

<table>
<thead>
<tr>
<th>Table 2: Specimen dimensions</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Length (L)</th>
<th>Diameter (D)</th>
<th>Height (h)</th>
<th>Width (b)</th>
<th>Span (s)</th>
<th>Notch (a)</th>
<th>Number cast</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPB-A</td>
<td>550</td>
<td>150</td>
<td>150</td>
<td>500</td>
<td>25</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TPB-B</td>
<td>550</td>
<td>125</td>
<td>150</td>
<td>500</td>
<td>–</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Disk-A</td>
<td>600</td>
<td>55</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Disk-B</td>
<td>800</td>
<td>70</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

| Table 3: Experimentally determined material properties |

<table>
<thead>
<tr>
<th>Property</th>
<th>Mix A</th>
<th>Std dev</th>
<th>Mix B</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_c) (MPa)</td>
<td>108.90</td>
<td>7.40</td>
<td>115.50</td>
<td>4.90</td>
</tr>
<tr>
<td>(f_t) (MPa)</td>
<td>6.29</td>
<td>0.25</td>
<td>6.39</td>
<td>0.33</td>
</tr>
<tr>
<td>(E) (GPa)</td>
<td>49.60</td>
<td>0.50</td>
<td>46.30</td>
<td>0.30</td>
</tr>
<tr>
<td>(\nu)</td>
<td>0.14</td>
<td>0.023</td>
<td>0.16</td>
<td>0.012</td>
</tr>
</tbody>
</table>

**Figure 9** Split tensile results for (a) Mix A and (b) Mix B
Figure 10 Comparison between experimental and simulated load-displacement results for various specimen types.
for the tensile splitting test presented in the previous section. The load-transversal deformation curves for the split cylinder tests on the mixes are shown in Figure 9.

Table 4 shows the values of $\sigma_{Nu}$ obtained for the TPB specimens, as well as the fracture properties in terms of $A$, $W_f$ and $G_f$ determined from the TPB results in accordance with the procedure discussed earlier in the paper. The load-displacement curves for the TPB tests are shown in Figure 10. The results of the flexural tests on panels are shown in Figure 11. Also shown in the figures are the results of the numerical simulation performed using the material parameters in Table 3 and Table 4. The numerical models are discussed in the next section.

### NUMERICAL SIMULATION OF FRACTURE

Two different finite element method (FEM) software frameworks were used for the simulation of fracture in the beams and panels. The numerical simulation of the beams can be reduced to a two-dimensional problem and was performed using the embedded discontinuity method (EDM). The EDM was implemented in the open-source FEM framework OpenSees (2008) by Wu et al (2009). The flexural tests on panels were simulated in three-dimensional space using the commercial software Abaqus (2009); this was necessary as the EDM implementation in OpenSees is at present limited to two-dimensional space.

#### Simulation of fracture in beams with EDM

The implementation of the embedded discontinuity method by Wu et al (2009) is based on work by Sancho et al (2007). An advantage of EDM over other cohesive crack FEM models is that it allows cracks to propagate through elements, independent of nodal positions and element boundaries. In the earlier work by Wu et al (2009), a simple exponential softening function was used for damage evolution. The softening function was found suitable to predict the fracture behaviour of a plain concrete pavement material (Denneman et al 2010c). The simple

<table>
<thead>
<tr>
<th>Specimen type</th>
<th>$\sigma_{Nu}$ (MPa)</th>
<th>Std dev (MPa)</th>
<th>$A$</th>
<th>$W_f$ (Nmm)</th>
<th>$W_{tens}$ (%)</th>
<th>$G_f$ (N/mm)</th>
<th>Std dev (N/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mix A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPB1-A</td>
<td>13.3</td>
<td>1.08</td>
<td>9.74</td>
<td>1.23E+05</td>
<td>17.6%</td>
<td>6.57</td>
<td>0.96</td>
</tr>
<tr>
<td>TPB2-A</td>
<td>13.5</td>
<td>1.27</td>
<td>4.96</td>
<td>8.54E+04</td>
<td>12.0%</td>
<td>4.56</td>
<td>0.69</td>
</tr>
<tr>
<td>TPB3-A</td>
<td>11.3</td>
<td>1.76</td>
<td>4.59</td>
<td>3.70E+04</td>
<td>22.2%</td>
<td>4.93</td>
<td>1.10</td>
</tr>
<tr>
<td>TPB4-A</td>
<td>13.9</td>
<td>0.89</td>
<td>2.97</td>
<td>2.86E+04</td>
<td>16.5%</td>
<td>3.82</td>
<td>0.88</td>
</tr>
<tr>
<td>Mix B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPB1-B</td>
<td>13.9</td>
<td>1.45</td>
<td>8.36</td>
<td>9.97E+04</td>
<td>15.7%</td>
<td>5.32</td>
<td>0.31</td>
</tr>
<tr>
<td>TPB2-B</td>
<td>13.4</td>
<td>1.83</td>
<td>4.34</td>
<td>4.13E+04</td>
<td>21.0%</td>
<td>5.51</td>
<td>1.21</td>
</tr>
<tr>
<td>TPB3-B</td>
<td>14.7</td>
<td>1.43</td>
<td>5.22</td>
<td>4.23E+04</td>
<td>21.2%</td>
<td>5.64</td>
<td>1.37</td>
</tr>
</tbody>
</table>

* The LVDT displacement recording malfunctioned during this set of disk tests. The results plotted are the actuator displacement vs load. The actuator LVDT has limited accuracy, resulting in the aberrant load-displacement curves.

---

**Figure 11** Comparison between experimental and numerical results for flexural tests on panels
exponential shape, however, over-predicts the peak load of fibre-reinforced concrete (Denneman et al. 2010a). The EDM code was updated by Denneman et al. (2011a) to allow for the more suitable exponential softening function with crack tip singularity described earlier in this paper.

The two-dimensional numerical model consists mainly of triangular elastic bulk elements. These elements require the modulus of elasticity and Poisson’s ratio as input. A narrow band of triangular-shaped embedded discontinuity elements is provided at the notch facilitating a vertical crack path to the top of the specimen. It is possible to run the analysis using EDM elements for the entire model. In the numerical simulation, a crack will invariably form at the position with the highest stress. To make the calculation more efficient, a narrow vertical band of triangular EDM elements was provided in the ligament area at the mid-span position of the beam specimens. An impression of the deformed meshes for the TPB tests at high displacements near the end of the test is provided with the simulated load-displacement curves for the various beam types in Figure 10.

Figure 12 shows the calibrated softening curves for Mix A and Mix B. As $G_f$ is determined from the TPB results and $f_t$ is obtained from tensile splitting tests, $w_1$ and $\sigma_f$ are the only unknowns to be calibrated in the model. The softening curves were developed based on a parameter study aimed at achieving the best fit for both the flexural beam and tensile splitting tests. $w_1$ was initially set to 0 for the simulation of flexural beam tests. It was later found that if a small displacement $w_1$ is used, a better fit of the model can be obtained for the simulation of tensile splitting tests (Denneman et al. 2011b). For the simulation of the flexural beam tests under study, the difference between $w = 0$ and $w = 0.005$ is insignificant. The main calibration is therefore the value of $\sigma_f$, which is chosen based on a parameter study for a single beam size after which the fit for other sizes is checked. As shown in Figure 10, the simulation of the TPB tests using the softening curves in Figure 12 results in a satisfactory, size-independent reflection of the physical measurements. As part of this study it will be investigated whether the softening behaviour can be generalised to a geometrically different test setup, i.e. the centrally loaded panel test.

**Simulation flexural tests on panels with Abaqus**

The commercial FEM software package Abaqus includes a number of fracture mechanics models that can be applied to concrete. The software includes a brittle cracking model intended specifically for the simulation of cracking in concrete. Damage evolution takes place using the cohesive crack principle for damage evolution in the principle tensile stress direction, the softening functions developed for the material in Figure 12 can be used unaltered.

In contrast to the EDM in which both shear and tension are handled using a crack width softening function, the brittle cracking model in Abaqus has a separate strain softening function for shear. To ensure that the response of the Abaqus model in shear was similar to that in the EDM simulation, the shear strain softening function was defined in accordance with the relation:

$$ w = h_s e_f $$  \hspace{1cm} (10)

This relation relates the softening as a result of the fracture strain ($e_f$) over a certain width ($h_s$) in smeared crack models to the softening as function of crack width ($w$) in cohesive crack models.

To verify that the fracture simulation using the Abaqus brittle cracking model is equivalent to the results obtained from the EDM in OpenSees, both approaches were applied to simulate a TPB test using the same softening relation. The results are shown in Figure 13. A characteristic element size of 1 mm was used for both the triangular OpenSees elements and the Abaqus quadrilateral elements in the ligament area above the notch. At the far edges of the mesh, the characteristic size of the elements was set

![Figure 12 Calibrated softening curves for mixes under study](Image)

![Figure 13 Comparison of numerical simulation using OpenSees and Abaqus](Image)
to 25 mm. The results obtained from the simulations are almost equivalent. Only at large displacements do the load displacement curves diverge slightly. This was taken as proof that the Abaqus brittle cracking model was correctly configured, allowing its use for the simulation of the flexural disk tests.

The layout of the disk tests is shown schematically in Figure 14. Owing to symmetry, only one sixth of the disk needs to be modelled as indicated in the figure. The model and boundary conditions are shown in Figure 15a. Fracture elements with the brittle crack damage model were used for the entire model, with exception of the area around the support. In this area linear elastic bulk elements were used to prevent cracks forming due to unrealistic stress concentrations. The support point is connected to a surface corresponding to the area of the transfer plate under the panel using a kinematic rigid coupling. Standard Abaqus 8 node brick elements of the C3D8R type were used for the mesh. The mesh is shown in Figure 15b. The characteristic size of the elements throughout the model is 5 mm, with exception of the area around the support where a size of 15 mm was used.

As the model consists mostly of fracture elements, multiple cracks will occur when stress redistribution after initial cracking leads to the development of new highly stressed areas. The results of the numerical simulation for the panels are shown in Figure 11. The results provide an accurate prediction of the pre-peak, peak load and early post-peak behaviour for the specimens. At larger deflections the crack tends to get locked, leading to unrealistic high stresses in the material. As a result, the load carried by the slab does not decrease at the pace observed in the experiments. The simulation was aborted when crack-locking started to occur. The simulation, however, yields satisfactory results in predicting the peak load for the panels.

To compare the results of the fracture mechanics-based analysis to a linear elastic design approach, the fracture elements in the panels were substituted for elements with linear elastic material behaviour. A load equal to the experimentally determined peak load per specimen type was applied to the models. Figure 16 shows the results in terms of the nominal tensile stress along the symmetry line midway between two supports from the centre of the panel to the edge. The maximum stresses at the centre of the disk are in the order of 22 MPa, almost twice the $\sigma_{Nu}$ values determined from beam tests shown in Table 4. These results indicate that if $\sigma_{Nu}$ (or MOR) was used to predict the peak load condition of the panels, as is done in pavement design, the flexural capacity of the panels would have been significantly underestimated – the error of the prediction would have been in the order of 70 to 100 per cent due to size effect.

CONCLUSIONS

The results in this paper show that, due to size effect, the MOR has limited reliability as a predictor of the peak load of FRC elements of a different size and/or geometry than the specimen for which the MOR was determined.

It was shown that the cohesive softening function with a crack tip singularity and exponential tail can be used to reliably predict the flexural behaviour of beams of different sizes and also of centrally loaded panels. The softening function was defined using a simple, but effective experimental methodology as presented in this paper. It is concluded that, in contrast to the MOR, the fracture mechanics models can be used to generalise the parameters obtained for a certain specimen size to reliably predict the flexural behaviour of specimens with a different size or geometry.

REFERENCES

concrete (using centrally loaded round panel). West Conshohocken: ASTM International.


Prediction of the debonding/slip load of composite deck slabs using fracture mechanics

J Mahachi, M Dundu

The aim of this paper is to develop equations that can be used to predict the load at which debonding or slip occurs in composite deck slabs, failing as a result of shear bond rupture. Debonding is the separation of the bonded steel plate from the concrete. The expressions are based on end-slip of the shear span occurring prior to ultimate load. Shear bond failure is considered to be a result of breakdown of mechanical and frictional resistance force between the steel and the concrete interface. Linear elastic fracture mechanics (LEFM) is assumed and the eccentric axial force transmitted by the steel deck is calculated using rotational congruence. The theoretical debonding load results are found to be comparable with experimental results, and to be of use in formulating the response of composite slabs when subjected to fatigue load.

INTRODUCTION

Composite steel deck floor slabs consist of concrete cast on top of cold-formed profiled steel sheets (Krige & Mahachi 1995). The steel deck is made by cold-forming structural grade sheet steel into a repeating pattern of parallel ribs, and the concrete, which may be either lightweight or normal weight, is then poured onto the decking, usually by pumping, to make up the composite system. Metal decking acts both as permanent formwork for the concrete, eliminating the need to provide props, and as tensile reinforcement for the slab. The integral composite action between the steel deck and the concrete is provided by mechanical interlocking devices (embossments) capable of resisting horizontal shear and preventing vertical separation of the steel/concrete interface. This form of slab construction is particularly popular for multi-storey buildings and bridge construction, when rapid construction is required.

Tests carried out by a number of researchers (Ekberg & Schuster 1968; Schuster & Ekberg 1970; Luttrell & Davison 1973; Porter & Ekberg 1976; Wright et al 1987; Mahachi 1997) have shown that generally the two materials exhibit complete interaction at relatively low loads. The load vs vertical mid-span deflection is linear up to the point when loss of interaction between the two materials occurs. The loss of interaction has been attributed to loss in shear bond. Determination of the debonding load is important, since it has been observed in fatigue tests (Mahachi 1994, 1995) that subjecting composite slabs to fatigue loads above the debonding load generally results in shear bond rupture. The debonding load is also important for serviceability, since deflection of the slab should be such that slippage does not occur at working loads. This paper will focus on the development of the debonding load for two steel deck profiles manufactured in South Africa, i.e. the Bond-Lok and the Bond-Dek.

The Bond-Lok and Bond-Dek floor slabs are one-way spanning slabs and are designed to carry uniformly distributed loads. They provide permanent formwork/shuttering to the wet concrete floor with an attractive flat ceiling on the underside and are able to span up to 3 m, unsupported, thus saving on the props utilised. The Bond-Lok decks have plain cold-rolled profiled steel sheets with male and female ribs that interlock. The deck has some form of re-entrant or dove tail angle, which prohibits vertical separation of concrete and steel deck due to the interlocking shape. When used with concrete, the system forms a composite slab, with the ribs bonded to the slab. A chemical bond is also formed with the ribs and flat section of the unit. The Bond-Lok tested in this investigation had a cover of 320 mm, gauge thickness of 0.92 mm and a trough depth of 50 mm (Figure 1(a)). To prevent rusting, galvanised surface coating with an average thickness of 0.035 mm (Z275) was bonded on both surfaces of the metal deck.

The steel deck profile for the Bond-Dek is trapezoidal in cross-section, as shown Figure 1(b). This figure also shows the dimensions of the Bond-Dek specimen that was tested. Both surfaces of the profile were galvanised with a 0.015 coating to prevent rusting. The shear connection in Bond-Dek is more important than that of the Bond-Lok, because...
of the embossments that project from the corrugations. The embossments run across the webs and provide good mechanical shear connection. The profile of the Bond-Dek is further stiffened by longitudinal grooves that run along the steel deck flanges to reduce the effects of buckling during construction loading. Both composite systems have been fire-tested by the Council of Scientific and Industrial Research (CSIR) in South Africa and have qualified for a rating of two hours.

For normal applications of Bond-Dek steel floors, no additional reinforcing other than a light mesh for shrinkage control is required, typically 193 mesh.

ANALYSIS
The behaviour and strength of composite slabs are often governed by the horizontal shear load at the interface of the steel deck and the concrete. Shear connection between steel and concrete is normally achieved by chemical adhesion, frictional resistance and mechanical interlock, collectively. The bond thus attained depends on panel geometry, thickness of the steel sheeting and concrete, surface conditions of the steel deck and types of embossments or rolled dimples that project from the steel deck into the concrete.

In this analysis, equations to predict the load \( P_d \) at which debonding occurs are developed using linear elastic fracture mechanics (LEFM). LEFM is assumed since concrete does not behave plastically under tensile situations. The horizontal slip resistance requires two experimentally determined parameters to model the interface slip, i.e. the coefficient of friction, \( \mu \), and the shear strength parameter, \( \tau \). The analysis is performed on a simply supported composite deck slab subjected to static point line loads. When subjected to loading, a discrete crack is assumed to form and propagate in the concrete under one of the loading points as shown in Figure 2.

The following assumptions will be made in the analysis:
1. The crack is assumed to propagate vertically upwards from the bottom.
2. Within the regime of LEFM, a linear elastic law will be assumed for concrete, and a rigid-plastic constitutive law for steel. This assumption can be considered valid since debonding normally occurs in the linear elastic regime.
3. The interface slip consists mainly of mechanical interlock and friction. The effect of the adhesion bond is assumed to be negligible.
4. Debonding of the steel/concrete interface occurs before the ultimate load.

In order to calculate the rotation in a cracked element, subjected to a bending moment and/or axial force, it is proposed to use the “Compliance Approach”, suggested by Okamura et al (1973, 1975) and Carpinteri (1984). For this approach to work, the composite cross-section is converted to an “equivalent” rectangular section by transforming the section so that the bending stiffness about the X-X axis remains the same, as shown in Figure 3. In the transformed section, the steel area is assumed to be concentrated at a height \( h \) from the bottom, and a crack of height \( a \) penetrating through the thickness of the model.

Consider a small element of the slab of length \( \Delta L_0 \) as shown in Figure 3, subjected to an opening external moment \( M \). Under the external moment, LEFM assumes that the force transmitted to the concrete increases with increasing moment until slippage or yielding of steel occurs. Due to the influence of the moment an eccentric axial force \( F_{st} \) will be induced in the steel deck. The force transmitted by the steel deck to the beam can be estimated by the principle of rotational congruence. This force is expected to increase linearly as the moment increases,
Figure 4 Equilibrium of forces

Table 1 Typical section/material properties of tested specimens

<table>
<thead>
<tr>
<th>Section/material properties</th>
<th>Bond-Lok</th>
<th>Bond-Dek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel depth, $h$</td>
<td>10 mm</td>
<td>36.38 mm</td>
</tr>
<tr>
<td>Width of slab, $b$</td>
<td>320 mm</td>
<td>940 mm</td>
</tr>
<tr>
<td>Depth of slab, $d$</td>
<td>150 mm</td>
<td>150 mm</td>
</tr>
<tr>
<td>Thickness of profile</td>
<td>0.92 mm</td>
<td>0.98 mm</td>
</tr>
<tr>
<td>Coefficient of friction, $\mu$</td>
<td>0.2</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Material properties

| Concrete strength, $f_{cm}$       | 40 MPa    | 40 MPa    |
| Concrete tensile strength        | 5.5 MPa   | 5.5 MPa   |
| Shear strength, $f_m$             | 0.095 MPa | 0.12 MPa  |
| Yield strength of profile        | 285 MPa   | 295 MPa   |
| Elastic modulus of profile       | 200 MPa   | 202 MPa   |

until debonding occurs. The stress intensity factor $K_1$ for a rectangular section (width $b$ and height $d$) that contains an edge crack of depth $a$, subjected to a bending moment $M_X$ about the centroid of the section has been shown by Okamura et al. (1973, 1975) to be given by:

$$K_1 = \frac{6M_X}{bd^2} g(\zeta)$$

(1)

where the function of the relative crack depth $\zeta$ is

$$g(\zeta) = 1.99\zeta^2 - 2.47\zeta^3 + 12.97\zeta^2 - 31.17\zeta^2 - 23.17\zeta^2 + 24.80\zeta^2$$

(2)

for $\zeta \leq 0.7$ and $\zeta = \frac{a}{d}$

Similarly, the stress intensity factor $K_2$ for a rectangular section subjected to an axial force $F_X$ acting at the level of the steel has been shown by Okamura et al. (1973, 1975) to be

$$K_2 = \frac{6F_X}{d^2} h(\zeta)$$

(3)

where

$$h(\zeta) = \frac{1}{9} - \frac{2}{9} \zeta^2 + \frac{8}{9} \zeta^2 - \frac{34}{9} \zeta^2 - \frac{50}{9} \zeta^2 + \frac{53.85}{9} \zeta^2$$

(4)

Whilst an uncracked section behaves as a perfectly fixed joint, a cracked section behaves as an elastic joint. Using a rotational congruence approach similar to that used by Okamura et al. (1973, 1975) and Carpinteri (1981) for reinforced concrete, the rotation $\phi$ of the crack due to an applied bending moment $M_X$ and an axial force $F_X$ can be evaluated by the principle of linear superposition as

$$\phi = \lambda_{MM} M_X - \lambda_{MF} F_X$$

(5)

Parameters $\lambda_{MM}$ and $\lambda_{MF}$ are the compliances of the member due to the existence of the crack and can be derived from energy methods by considering the moment $M_X$ and the axial force acting together to give

$$\lambda_{MM} = \frac{2}{d^2} \int_0^\zeta \overline{g}(\zeta) d(\zeta)$$

(6)

$$\lambda_{MF} = \frac{2}{d^2} \int_0^\zeta \overline{g}(\zeta) h(\zeta) d(\zeta)$$

(7)

At the point of slippage the net rotation $\phi = 0$ since the applied bending moment to open the crack and the axial force from the steel deck, to close the same crack, are of the same magnitudes. Setting $\phi = 0$ in Eq (5) yields the angular compatibility equation:

$$\phi = \lambda_{MM} M_X - \lambda_{MF} F_X = 0$$

(8)

From Figure 3, the applied bending moment $M_X = M - F_{st}(d/2 - h)$ and the axial force $F_X = -F_{st}$. Substituting $M_X$ and $F_X$ in Eq (8) gives

$$\lambda_{MM} [M - F_{st}(d/2 - h)] - \lambda_{MF} F_{st} = 0$$

(9)

Rearranging Eq 9 gives the indeterminate force $F_{st}$ as a function of the applied moment $M$, that is

$$F_{st} = \frac{M}{\frac{1}{2} - \frac{h}{d}} + r(\zeta)$$

(10)

where

$$r(\zeta) = \frac{\int_0^\zeta \overline{g}(\zeta) h(\zeta) d(\zeta)}{\int_0^\zeta \overline{g}(\zeta) d(\zeta)}$$

(11)

From Eq (11):

$$\zeta \rightarrow 0; \ r(\zeta) \rightarrow \frac{1}{6}$$

(12)

Substituting Eq (12) into Eq (10) results in

$$F_{st} = \frac{M}{\frac{1}{2} - \frac{n}{d}}$$

(13)

Now consider the equilibrium of horizontal forces, along the shear span as shown in Figure 4.

In the figure $L_s$ is the overhang length, $L_t$ is the shear span and $F_i$ is the interface force. If a load $P$ is applied to the composite system (see Figure 1), then the slab is subjected to a two-point line loading, with each point load of magnitude $P/2$. For horizontal equilibrium, the interface force is $F_i$ given by

$$F_i = F_c = F_{st}$$

(14)

where $F_c$ = Force in the concrete. The limiting force $F_i$ at the interface is due to the mechanical interlock and friction (Patrick & Bridge 1990; Schuster & Ling 1990) and is given by

$$F_i = r_m A_j + \mu N$$

(15)
where $V$ is the clamping force due the vertical reaction, $A_i$ is the interface area, $r_{si}$ is the mechanical shear strength parameter and $\mu$ is the coefficient of friction. As long as $F_i < F_1$ no slippage occurs. This implies that the limit force for slippage to occur is

$$ F_i = F_1 $$

(16)

It can be deduced from Eqs (14), (15) and (16) that debonding starts to occur when

$$ F_{st} = r_mA_i + \mu V $$

(17)

Substituting Eq (17) into Eq (13) yields

$$ r_mA_i + \mu V = \frac{M/d}{(3 - \frac{d}{b})} $$

(18)

where $A_i = (L_s + L_c)b$

Eq (18) can be solved to obtain the maximum loads that can be applied to the slab before debonding starts to occur.

**APPLICATION**

Fatigue tests were carried out by Mahachi (1997) in order to determine the response of composite Bond-Lok and Bond-Dek slabs to fatigue loading. Initial experimental tests involved the determination of the maximum ultimate static load, as well as the load at which debonding occurs. It was necessary to determine the load at which slip occurred, since applying a fatigue load above this resulted in immediate failure or a failure after a few hundred cycles. Typical section and material properties of the tested specimens are given in Table 1.

The shear bond parameters (mean shear stress per unit horizontal area ($r_m$) and coefficient of friction ($\mu$)) were established using the test method, developed by Patrick & Bridge (1990). The slabs were cast on the laboratory floor, and after 28 days the slabs were lifted onto the testing platform. In all tests, the composite slabs were one-way spanning and simply-supported on a span of 3.5 m. Tests were conducted using centre-line loading and two-point-line loading. The line loading was applied across the width of the specimen. For two-point-line loading, the shear span was maintained at 1.2 m. During testing, readings were taken of the vertical mid-span deflections and the horizontal differential movement or slip between the steel sheeting and the concrete slab using LVDTs and dial gauges.

Typical graphs of the load vs mid-span deflections and end-slip were taken of the vertical mid-span deflections and end-slip for the Bond-Lok and Bond-Dek are shown in Figure 5.

Similar load-deflection graphs were developed for composite Bond-Dek slabs, and were documented by Mahachi (1997).

**THEORETICAL CALCULATIONS**

Substituting $V$ and $M$ in Eq (18) for a slab subjected to a two-point loading, with a shear span $L_s$ yields the total load:

$$ P = \frac{2r_mb(1 + \frac{L_s}{L_c})bd}{(3 - \frac{d}{b})} $$

(19)

If, however, the coefficient of friction $\mu$ is neglected, then $\frac{\mu d}{r_m} = 0$ and Eq (19) reduces to:

$$ P = 2r_mb(1 + \frac{L_s}{L_c})d $$

(20)

From Eq (19) it can be observed that the effect of $\mu$ is to increase the load at which debonding occurs, particularly for shorter spans.

Eq (18) is valid for:

$$ \frac{1}{3} \frac{b}{d} \frac{d}{b} > \frac{\mu d}{r_m} \quad \text{i.e.} \quad L_s > \mu d(\frac{3}{2} - \frac{b}{d}) $$

(21)

Using a value of $\mu = 0.2$ and $r_m = 0.095 \text{ N/mm}^2$ (determined as suggested by Patrick & Bridge 1990), the shear span for the Bond-Lok can be shown to be greater than 19 mm ($L_s > 19$ mm). This implies that Eq (18) is valid for all practical span lengths. A plot of Eqs (18) and (19) is shown in Figure 6. From the figure, the debonding load increases with decreasing shear span, $L_s$. Also, the effect of friction $\mu$ is greater for shorter spans than longer spans. As the shear span increases, the debonding load approaches a constant value of 6 kN. Values of the shear span ($L_s$), overhang length ($L_c$), average depth of the specimens ($\bar{d}$) and the average position of the steel area ($\bar{h}$) are given in Table 2.

The equivalent sections of the Bond-Lok and Bond-Dek were calculated using the bending strength criteria. The equivalent widths ($b_e$) for the Bond-Lok and Bond-Dek were found to be 307 mm and 590 mm respectively. Using the shear values of $r_m = 0.095 \text{ N/mm}^2$ for the Bond-Lok and $r_m = 0.12 \text{ mm}^2$ for the Bond-Dek, the
theoretical loads \( (P_{td}) \) were calculated (see Table 2). It can be seen from this table that the theoretical debonding loads and the experimental loads \( (P_{ed}) \) are acceptably close. Also the effect of the coefficient of friction is to increase the debonding load by only 1% to 2%, which is significantly small.

**CONCLUSION**

The use of fracture mechanics enables the calculations of the debonding load \( P_d \) to be carried out without recourse to several experimental tests. This formulation requires only two experimentally determined input parameters, i.e. the coefficient of friction and the shear strength per unit area \( \tau_m \). However, the effect of the coefficient of friction is negligible compared to shear strength. It is anticipated that using non-linear elasto-plastic fracture mechanics will enable the determination of the ultimate load without embarking on several experimental tests.

**REFERENCES**


<table>
<thead>
<tr>
<th>Deck</th>
<th>( L_s ) (mm)</th>
<th>( L_c ) (mm)</th>
<th>( d_a ) (mm)</th>
<th>( b_a ) (mm)</th>
<th>( b_e ) (mm)</th>
<th>( P_{td} ) (kN)</th>
<th>( P_{td} (\mu=0) ) (kN)</th>
<th>( P_{ed} ) (kN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond-Lok</td>
<td>1750</td>
<td>105</td>
<td>155</td>
<td>10.85</td>
<td>307</td>
<td>5.78</td>
<td>5.72</td>
<td>6.20</td>
</tr>
<tr>
<td></td>
<td>1200</td>
<td>105</td>
<td>155</td>
<td>10.85</td>
<td>301</td>
<td>5.96</td>
<td>5.86</td>
<td>6.50</td>
</tr>
<tr>
<td>Bond-Dek</td>
<td>1200</td>
<td>100</td>
<td>150</td>
<td>36.38</td>
<td>590</td>
<td>9.86</td>
<td>9.76</td>
<td>10.50</td>
</tr>
</tbody>
</table>

**Figure 6** Effect of coefficient of friction \( \mu \)
Estimating soil plasticity properties from pedological data

G C Fanourakis

A number of pedological soil classification systems have been developed worldwide. These include an internationally accepted system and various national systems, some of which have been incorporated into databases which include maps. Such information is used primarily for agricultural purposes. Various physical and chemical soil properties are used for classifying soils according to these pedological systems. This paper proposes an approach, based on a research project, which may be used to statistically significantly determine the plasticity characteristics of soils from the physical and chemical properties that are used to pedologically classify soils by systems, such as the South African Binomial System. These plasticity characteristics may be used to establish the engineering soil classification groups which may, in turn, be used as a means of rapidly determining the general suitability of areas for proposed developments, particularly during the preliminary stages of transportation route locations and township developments, with a resultant saving of time and money.

INTRODUCTION

Pedology

Pedology is a branch of soil science which deals with the study of soils as natural phenomena, including their formation, morphological, physical, chemical, mineralogical and biological constitution.

Pedological classification systems

A number of pedological soil classification systems have been established worldwide and are used as a basis for mapping soils. Such systems include an internationally accepted system and many national systems.

The internationally accepted system, which is endorsed by the International Union of Soil Sciences (IUSS), is the World Reference Base for Soil Classification (IUSS-WRB 1998, 2006 and 2007). This system replaced the Food and Agriculture Organisation (FAO) system, which was originally developed as a legend to the Soil Map of the World in 1974 (FAO 1974, 1990) (Rossiter 2012).

National systems include the Australian (Isbell 1996), Canadian (SCWG 1998), English and Welsh (Avery 1980), French (Baize & Girard 1998, 2008), German (ABDBG 1998), Russian (Stolbovoy 2000), American (USDA 2010) and the South African Binomial Classification System (Mac Vicar et al 1977). These national systems are also utilised by private organisations.

According to Rossiter (2001), soils that are allocated a particular national class should ideally all be allocated the same WRB system class.

Published correlations of national classification systems to the WRB system include the South African Binomial System (Fey 2010) and the USDA system (Rossiter 2012). It appears that, in future, correlations will be established between other national pedological classification systems and the WRB system. The Canadian and Israeli classification systems have already been correlated to the FAO system, which was superseded by the WRB system (Rossiter 2012).

Pedological databases

Various international, regional and national databases comprising pedological, climatic, topographical and other information have been developed.

International databases include the SOTER (Soil and Terrain) world database. This digital database, which includes maps and soil information (in a standard format), has already been compiled for a number of countries (Isric 2012).

National databases include the following.
- Land Information System (LandIS) of England and Wales (Landis 2012)
- Australian Soil Resource Information System (Asris) of Australia (Asris 2012)
- National Soil Information System (Nasis) of the United States (Nasis 2012)
- Land Type Survey of South Africa (LTSS 2012)
- The soil data included in the Land Type Survey of South Africa (LTSS 2012) include

Keywords: plasticity, laboratory tests, chemical properties, classification
the pedological properties and pedological classification in accordance with the South African Binomial Classification System (after MacVicar et al. 1977).

From the above it is evident that a significant amount of pedological classification and mapping has been conducted worldwide. This data is used extensively and successfully primarily for the optimisation of the utilisation of agricultural land for crop production – so much so that pedology is often incorrectly regarded as part of agricultural science. It would be obviously beneficial if this data could benefit the engineering profession as well.

**Objectives of this paper**

This paper is based on a research project that was initiated by the premise that the physical and chemical properties of a soil, which are manifested in the pedological class of a soil, have a bearing on the engineering properties of that soil.

In this research, relationships were established for determining the engineering plasticity characteristics, namely liquid limit (LL), plasticity index (PI) and linear shrinkage (LS) of soils deriving from a range of pedological classes of the South African Binomial System (after MacVicar et al. 1977), from their physical and chemical properties which are determined for pedological classification purposes. These physico-chemical properties include the clay content, the percentage of the different exchangeable base cations and the cation exchange capacities (CEC), which are included in the data that is made available as part of the Land Type Survey of South Africa (LTSS 2012).

The estimated properties may be used to assess the suitability of soils for proposed engineering projects. Such data could also be used to supplement other available materials data such as that included in the defunct National Data Bank for Roads (NDBR), which is discussed in TRH 2 (1978).

**THE SOUTH AFRICAN BINOMIAL CLASSIFICATION SYSTEM**

**Structure**

The Binomial Classification System uses morphology (the science of form and structure) and composition of soils as criteria for differentiation in soil classification. It groups soils according to the similarity of the properties used in the identification. This comprehensive classification system employs two categories, namely an upper or general level consisting of soil forms, and a lower, more specific one, comprising soil series.

**Soil forms**

The upper level consists of 41 soil forms, each defined by a unique vertical sequence of diagnostic horizons or materials, not more than four in number, occurring in the uppermost 1.2 m of the soil profile.

**Soil series**

Soil forms are based on selected soil properties used to define diagnostic horizons with fairly wide permissible variations and, therefore, each soil is classified further in order to narrow down such wide variations that might exist. Hence, each soil form is sub-divisible into a number of soil series, varying from one to 36. A total of 504 soil series constitute the lower category of the classification system. Soil series have in common the properties of the relevant form (that is, the prescribed horizon sequence), but are...
differentiated within the form according to a variety of criteria.

The criteria used for series differentiation within each soil form include the clay content, sand grade (where the clay content is less than 15%), red and non-red colouring, the presence or absence of free calcium carbonate, the degree of base status, which is a measure of the degree to which the soil has been leached, pH, and in some cases, the presence of continuous clay skins (or cutans) on ped faces, or a crusting distinct from a self-mulching surface.

**Double designation**

Identification and communication are achieved by means of standard designations, for example Hutton Mangano, where Hutton is the form and Mangano is the series (abbreviated as Hu 33). The soil series designation is retained regardless of where the soil occurs (MacVicar et al 1977).

**Revisions**

A revised version of the South African Binomial Classification System, referred to as the "Taxonomic System for South Africa", was published by the Soil Classification Working Group (SCWG) in 1991. This system is a more specific extension of the South African Binomial Classification System. Many persons have found the newer system to be "unnecessarily complicated" and hence have not adopted it (Rossiter 2012).

Nevertheless, in this paper consideration is only given to the South African Binomial Classification System which has been solely adopted for soil classification in the Land Type Survey of South Africa (LTSS 2012), which was incorporated into the SOTER database, and which is still being used in South Africa.

**RESEARCH PROCEDURE**

**Research area**

The area studied, which is located approximately 150 kilometres northwest of the city of Rustenburg in South Africa, is approximately 4 200 km² in extent. The area was selected for its diversity of soil types.

A number of soil types occur in the area, differing from one another as a result of different parent materials and/or the different soil-forming processes to which they had been subjected. The research was confined to the dominant soil series which adequately cover the range of soil types occurring in the area.

Five soil forms were selected for the research. These are the Hutton, Shortlands, Valsrivier, Swartland and Arcadia forms, which are respectively shown in the typical photographs (with the associated horizon types) in Figures 1 to 5.

Note that A horizons overlie B horizons (where present) which, in turn, overlie C horizons (where present). Therefore, for example, soils of the Hutton form comprise an Orthic A horizon overlying a Red Aypedal B horizon.

The correlations of the soil series selected with the World Reference Base (WRB) Classification System are included in Table 1. These correlations are in agreement with the general correlations established by Fey (2010).

**Fieldwork**

The soils at 63 randomly selected sites within the study area were classified according

---

### Table 1 Soil forms selected for research purposes

<table>
<thead>
<tr>
<th>Soil form</th>
<th>Diagnostic horizons</th>
<th>Soil type</th>
<th>General correlation with World Reference Base (WRB) classification system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hutton</td>
<td>Orthic</td>
<td>Red Aypedal</td>
<td>Ferralsitic and Fersiallitic (not characterised by the dominance of smectitic clay minerals) Ferralsols and Arenosols</td>
</tr>
<tr>
<td>Shortlands</td>
<td>Orthic</td>
<td>Red structured</td>
<td>Fersiallitic (characterised by the dominance of smectitic clay minerals) Luvisols</td>
</tr>
<tr>
<td>Valsrivier</td>
<td>Orthic</td>
<td>Pedocutanic</td>
<td>Unconsolidated material Red series – Fersiallitic (characterised by the dominance of smectitic clay minerals) Non-red series – Planosolic</td>
</tr>
<tr>
<td>Swartland</td>
<td>Orthic</td>
<td>Pedocutanic</td>
<td>Saprolite Margallitic Vertisols</td>
</tr>
<tr>
<td>Arcadia</td>
<td>Vertic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A total of 111 disturbed soil samples, taken from the A and B (where present) horizons of these soil profiles, were tested to determine the physical and chemical properties which are used as criteria for pedological classification. In view of the fact that the physical and chemical properties of C horizons are not reflected in the pedological classifications of profiles which contain C horizons, C horizons were not sampled.

Details of the results, methods of analysis adopted, and references to the relevant methods, as well as brief discussions on the significance of each property determined, are given in the work of Fanourakis (1999).

The number of series encountered in the study area, the number of soil profiles classified and the number of samples tested for the determination of their plasticity properties for the soils of each form are shown in Table 2. The plasticity characteristics of the 99 samples were determined in accordance with TMH1 (1986). It should be noted that all the possible soil series in the forms considered could not be included in this research as they do not occur in the study area.

ANALYSIS

Justification for inclusion of linear shrinkage data
Generally, the plasticity index of a soil is approximately twice the magnitude of the linear shrinkage. For this reason, the possibility of excluding the linear shrinkage data from the analyses in this research was considered.

Figure 6 shows the relationship between plasticity index and linear shrinkage, for all the samples included in this research, which yielded a correlation coefficient of 0.953.

With reference to the regression equation pertaining to the line in Figure 6, the plasticity index is, on average, approximately twice the linear shrinkage. However, as indicated in Table 3, the actual (measured) plasticity index / linear shrinkage ratios for the soils of each form varied from 1.373 to 3.5. Hence, the actual linear shrinkage data was included in the relationships established.

Determining plasticity characteristics from the soil series
For each of the soil series included in the project, the liquid limits and plasticity indices of the samples taken from the diagnostic horizons (A and, where present, B) which displayed plasticity were plotted on a Casagrande plasticity chart (according to the USCS after USACE WES 1960).

The plasticity data pertaining to those soil series which constitute each soil form were plotted on separate plasticity charts. The purpose was to investigate whether the soils of a particular soil series would yield points which are located within a particular zone of the plasticity chart.

An examination of the five plasticity charts revealed that, in the case of the soil forms comprising B horizons, the points representing the B horizons were generally located to the right of the points representing the A horizons. This is because B horizons generally comprise a relatively higher clay content than the A horizons of their respective soil profiles. Furthermore, it is evident that the points representing soils of the same series were not located within a particular region of any of the charts. The reason for this is that series differentiation criteria are qualitative. Therefore, pedological information more specific than the soil series is required in order to determine the engineering properties of soils.

Chemical and mineralogical data as criteria for determining plasticity characteristics
The effect of quantitative chemical properties, such as the different exchangeable base cations (Ca, Mg, Na, K), S-value, base saturation, exchangeable sodium percentage (ESP), exchangeable magnesium percentage, exchangeable potassium percentage, and exchangeable calcium percentage on the plasticity characteristics of the soils constituting each soil form were evaluated with reference to liquid limit, plasticity index and linear shrinkage.

The combined effect of each of the above-mentioned chemical characteristics, in conjunction with physical characteristics such as the clay content, silt content, and percentage finer than 0.075 mm on the plasticity characteristics of the respective samples, was also evaluated.
In the evaluation of the above, the fact that the physico-chemical properties were determined using the fraction finer than 2.0 mm and the plasticity characteristics using the fraction finer than 0.425 mm was accounted for.

A comprehensive list of the possible relationships between the plasticity characteristics and the various chemical and physico-chemical properties of the soils of each form investigated is given in the work of Fanourakis (1999). The magnesium in these soils was found to be the only factor that correlated with the plasticity characteristics.

**Statistical methods**

None of the data was transformed and no multiple regression procedures were adopted. The regression equations and correlation coefficients that were not significant at the 5% level were rejected.

Lines were fitted to the data using the least squares method.

Significance tests were carried out for each relationship established using tables of Student’s t-distribution to determine the probability that the correlation coefficient could have arisen by chance in a sample of the size dealt with.

Fisher’s Z-transformation technique was used to combine independent correlation coefficients, which are significant at the 5% probability level, in order to obtain an estimate of the average (pooled) correlation coefficient. This procedure took into account each correlation coefficient and the corresponding number of sets of values on which it was based.

**RESULTS AND DISCUSSION**

**The effect of exchangeable magnesium**

In the case of the soils of each form studied, the magnitude of the liquid limit, plasticity index and linear shrinkage each correlated significantly, linearly and positively, with the amount of magnesium present in the clay-size portion of the fraction of the sample finer than 0.425 mm. This quantity which, in the succeeding discussion, will be referred to as "exchangeable magnesium" or "exchangeable magnesium percentage (EMgP)". The quantity referred to in Equation 1 was calculated by multiplying the value of the exchangeable magnesium (expressed in milli-equivalents per 100 g of soil finer than 2 mm, me %) by the ratio of the percentage by mass of the fraction of the total sample finer than 0.002 mm (P0.002 mm) to the percentage by mass of the fraction of the total sample finer than 0.425 mm (P0.425 mm).

The definition of *magnesium in the clay-size fraction* (Equation 1) is based on two premises, firstly, that the exchangeable magnesium occurs almost exclusively in the fraction finer than 0.425 mm, and secondly that the exchangeable magnesium is distributed uniformly within this fraction. These premises were regarded as applicable to all the soils investigated in this project, regardless of the degree of base saturation.

To test the validity of these premises, sets of relationships between exchangeable magnesium and plasticity characteristics were established, based on each of the following two assumptions:

Firstly, that the magnesium cations are uniformly distributed within the soil fraction finer than 2 mm, and secondly, that the magnesium cations are uniformly distributed within the minus 0.075 mm fraction.

The relationships established were not all significant and it can therefore be concluded that the upper size limit of the soil fraction in which the exchangeable magnesium predominates must be of the order of 0.425 mm. However, it must be emphasized that this is not because the engineering plasticity analyses are performed on the fraction finer than 0.425 mm of the soil.

The exchangeable cations in a soil are adsorbed onto the negatively charged clay particles (Brady 1974). In view of this fact, the relationships established between the plasticity characteristics and the *magnesium in the clay-size fraction* indicated that the soils researched must contain clay particles which are larger than 0.002 mm in size. Furthermore, these relationships indicated that only the exchangeable magnesium which is adsorbed onto the clay particles that are finer than 0.002 mm in size had an effect on the plasticity.

It is not unusual for soils to contain clay mineral particles which are larger than 0.002 mm in size. A study of the mineralogy of five weathering profiles developed from Archaean granite, conducted by Buhmann (1990), revealed that particles of clay minerals such as biotite occurred in the silt and sand size fractions of soils derived from Archaean granites. This was confirmed in a subsequent study of the compositional characteristics of the soils of 15 profiles developed from Archaean granite, which indicated that the proportions of certain weathering products of biotite (mica and mica/smectite interstratifications) were highest in the silt-sized fraction of the soil (Buhmann et al 1998).

The biotites have a platy structure and high charges. Therefore, exchangeable cations are adsorbed onto the surfaces of these mineral particles, regardless of their size. The relatively large size of these clay particles is thought to be as a result of the slow cooling of the molten magma (Buhmann 1990).

It is unlikely that the highly expansive soils of the Arcadia form derived from Archaean granite. In the study area these soils derived from the Bushveld Igneous Complex. Nevertheless, it has been shown that the swelling component of these soils can occur exclusively in the non-clay-sized fraction (Buhmann et al 1988).

The correlations determined for the soils of each soil form are discussed in the succeeding paragraphs.
Discussion of results

Soils of the Hutton form

The mean activity of the soils of the Hutton form was 0.47. Of the 48 soil samples of the Hutton form analysed, a total of 30 samples displayed plasticity. Examination of the grading and plasticity characteristics of these soils revealed that, on average, soils with a clay content of less than 12% did not display plasticity. This absence of plasticity is thought to be as a result of a low content of 1:1 clays in the soils.

The relationships established between liquid limit, plasticity index and linear shrinkage respectively with the magnesium in the clay-size fraction are shown in Figures 7 to 9. The regression equations and correlation coefficients are included in these figures.

The relationships established were all highly significant, being at the 0.001% level of probability. The correlation was positive – an increase in the magnesium in the clay-size fraction was associated with an increase in each of the three plasticity characteristics.

A statistical analysis of the correlation coefficients of the relationships referred to above, and the three pairs of values from which each relationship was determined, yielded a pooled correlation coefficient of 0.863. The mean of the probability that these relationships could have arisen by chance was 0.0002%.

Soils of the Arcadia form

The relationships established between liquid limit, plasticity index and linear shrinkage respectively with the magnesium in the clay-size fraction are shown in Figures 10 to 12. The regression equations and correlation coefficients are included in these figures.

The relationships established were all highly significant, being at the 0.5% level of probability. The correlation was positive – an increase in the magnesium in the clay-size fraction was associated with an increase in each of the three plasticity characteristics.

A statistical analysis of the correlation coefficients of the relationships referred to above, and the three pairs of values from which each relationship was determined, yielded a pooled correlation coefficient of 0.707. The mean of the probability that these relationships could have arisen by chance was 0.2%.

Soils of the Shortlands, Valsrivier and Swartland forms

The relationships between liquid limit, plasticity index and linear shrinkage respectively, and the magnesium in the clay-size fraction for the soils of these forms were analysed according to the six groups shown in Table 4, justified by the similarities described below.
The red structured B horizon of the Shortlands form may resemble the red pedocutanic B horizons of the Valsrivier and Swartland forms in physical appearance. When classifying a soil pedologically, differentiation between red structured B horizons and red pedocutanic B horizons is determined by certain differences in the physical characteristics (including the structure and texture) of the B horizons compared to the corresponding physical characteristics of the A horizons in that soil profile. For example, if the textural class of a B horizon is different from that of the A horizon in the same soil profile, then the B horizon would be a pedocutanic B horizon. If, on the other hand, the textural class of the B horizon is the same as that of the A horizon in that profile, then the B horizon would be a red structured B horizon (MacVicar et al. 1977).

In addition, the Valsrivier and Swartland forms consist of an Orthic A horizon underlain by a Pedocutanic B horizon, and identical series differentiation criteria are applied for both. The only difference between the two soil forms is that Valsrivier and Swartland B horizons are underlain by unconsolidated material and saprolite, respectively. Therefore, the soils of these two forms were analysed as a single group.

The pooled correlation coefficients, as well as the minimum, maximum and mean levels of significance for the six groups, are shown in Table 4. From the pooled correlation coefficients and mean probabilities shown in Table 4, it is evident that the best levels of significance were yielded when combining the data of the Shortlands, Valsrivier and Swartland forms.

Therefore, the engineering plasticity characteristics of soils of the Shortlands, Valsrivier and Swartland forms can be determined from the magnesium in the clay-size fraction of each soil sample using the relationships established when the test results for the soils of all three of the above forms were grouped. Furthermore, in the case of these soils (which had a mean activity of 0.52) qualitative physical characteristics such as uniform colour, predominant colour and the presence of cutans did not bear any relationship to the soil’s plasticity characteristics. This is true only for soils with physico-chemical properties within the ranges of the soils investigated, including a low exchangeable sodium percentage (ESP).

The relationships established between liquid limit, plasticity index and linear shrinkage respectively with the magnesium in the clay-size fraction are shown in Figures 13 to 15. The regression equations and correlation coefficients are included in these figures.

### Table 4 Pooled correlation coefficients and corresponding levels of significance

<table>
<thead>
<tr>
<th>Series</th>
<th>Pooled (r)</th>
<th>Levels of significance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Shortlands</td>
<td>0.956</td>
<td>3.3</td>
</tr>
<tr>
<td>Valsrivier and Swartland</td>
<td>0.890</td>
<td>1.5E-10</td>
</tr>
<tr>
<td>Red Valsrivier and red Swartland</td>
<td>0.836</td>
<td>2.8E-05</td>
</tr>
<tr>
<td>Shortlands, red Valsrivier and red Swartland</td>
<td>0.829</td>
<td>1.4E-06</td>
</tr>
<tr>
<td>Non-red Valsrivier and non-red Swartland</td>
<td>0.943</td>
<td>0.005</td>
</tr>
<tr>
<td>Shortlands, Valsrivier and Swartland (Luvisols)</td>
<td>0.886</td>
<td>3.7E-12</td>
</tr>
</tbody>
</table>

1 General equivalent according to the WRB Classification System (WRB 1998, 2006 and 2007).
The relationships established were all highly significant, being at the 5E-09% level of probability. The correlation was positive – an increase in the magnesium in the clay-size fraction was associated with an increase in each of the three plasticity characteristics.

A statistical analysis of the correlation coefficients of the relationships referred to above, and the three pairs of values from which each relationship was determined, yielded a pooled correlation coefficient of 0.886. The mean of the probability that these relationships could have arisen by chance was 2E-09%.

**CONCLUSIONS**

This research showed that the liquid limit, plasticity index and linear shrinkage of soils can be statistically significantly estimated from the physico-chemical data of soils, for a range of pedological groups.

All the relationships established for estimating the plasticity characteristics of the soils of the Hutton form (Ferralsols and Arenosols according to the WRB Soil Classification System), Arcadia form (Vertisols according to the WRB system) and Shortlands, Valsrivier and Swartland forms (Luvisols according to the WRB system) were highly significant (P < 1%).

With the aid of quantitative pedological data, the plasticity characteristics of soils, which are used to classify a soil for various intended engineering purposes, can be estimated with significant accuracy. Hence, this information may assist in the identification and avoidance of large areas of unfavourable soils in projects such as township development and transportation route alignment, resulting in a saving of time and money. Conversely, areas or routes covered by favourable soils can be located, hence providing the framework for more detailed testing required for the final engineering design.

The findings of this research are valid for soils of the pedological classifications investigated and their particular physical, chemical and mineralogical properties only, since the research project was intended to serve as a pilot study. The indications are that additional similar studies covering soils of the pedological classifications researched, but with different chemical properties, as well as soils with other pedological classifications with varying chemical properties, could undoubtedly lead to the extension of this work and improvement of its universality.

It is recommended that, in future research, the data of soils of forms of the same WRB class be grouped for relationship establishment purposes, as was done in the case of the Luvisolic soils in this research.

**Figure 13** Relationship between liquid limit and magnesium in the clay-size fraction for soils of the Shortlands, Valsrivier and Swartland forms

**Figure 14** Relationship between plasticity index and magnesium in the clay-size fraction for soils of the Shortlands, Valsrivier and Swartland forms

**Figure 15** Relationship between linear shrinkage and magnesium in the clay-size fraction for soils of the Shortlands, Valsrivier and Swartland forms
This would obviate the application of incorrect regression equations as a consequence of incorrect pedological classification of similar soils.

Finally, it is hoped that this research has succeeded in emphasising the inter-relationships between pedogenesis and the engineering behaviour of soils, and in suggesting an approach for the interpretation of pedological data for engineering purposes.

ACKNOWLEDGEMENT

The author thanks the Agricultural Research Council (ARC) for permitting the inclusion of the photographs of the soil forms in Figures 1–5.

REFERENCES


USACE WES (United States Army Corps of Engineers, Waterways Experiment Station) 1960. The Unified Soil Classification System. Appendix A – Characteristics of soil groups pertaining to embankments and foundations, 1953; Appendix B – Characteristics of soil groups pertaining to roads and airfields, Technical memorandum No 3-357, Vicksburg, Mississippi, US: USACE WES.

The importance of plane end-bearing surfaces when measuring the strengths of concrete core specimens

T. Bugai, D. Kruger, R.G.D. Rankine

INTRODUCTION

Incorrect measurement of the in situ strength of concrete by crushing extracted cores has serious economic and legal implications. Recent experience has shown an alarming spread of results when cores from a single structure were capped and tested by different laboratories using different procedures. In South Africa, the procedure for preparing and strength-testing cores is defined by SANS 5865 Concrete Tests – The drilling, preparation, and testing for compressive strength of cores taken from hardened concrete. This standard provides for two methods of preparing the load-bearing ends of the core specimens, namely the use of hot sulphur mortar and High Alumina Cement (HAC) mortar.

The SANS 5865 method of preparation using sulphur mortar entails mixing equal parts of molten sulphur (at a temperature of approximately 240°C) and sand with 2% – 3% carbon black. This molten mixture is then cast onto both ends of the core specimen using a collar (often a piston ring compressor) and a precision-made rigid metal plate that is plane to within 0.5 mm/m with a slightly bevelled rim to create the required plane surface.

The SANS 5865 method of preparation of end caps with HAC mortar is essentially the same, but a sheet of flat glass (of thickness at least 6 mm) is used (instead of metal) to form plane-bearing surfaces, since it does not need to withstand high temperatures.

A number of laboratories have elected to cap core specimens with the HAC mortar instead of hot sulphur for stated health and safety reasons – including risks of burns, fire and inhalation of sulphur fumes. Observations of HAC mortar-capped specimens confirmed that the end-bearing surfaces were often out-of-plane (lumpy), evidenced by rocking when placed on a plane surface and/or forming gaps between the cast ends and a plane surface when tested with a feeler gauge. Specimens with convex end-bearing surfaces often split in half vertically during testing – analogous to a cylinder splitting test. Further investigation by Bugai confirmed that a reason for the departure from planeness was due to the fact that laboratory workers trowel the end surfaces instead of casting them against flat glass plates. Figure 1 shows typical “footprints” of cores capped with and without plane casting surfaces. Caps formed with a steel trowel tend to make contact against a flat bearing surface (such as a precision-ground steel plate or a plate of float glass) at only three contact points. Two graphical techniques were found suitable to illustrate points of contact – transfer onto a plane surface using fingerprint paint, and transfer onto a plane surface using carbon paper and pressure. (Use of “engineers’ marking blue” proved too viscous and rapid drying.) The sulphur-capped specimen (Figure 1a) exhibits even distribution of colour, confirming a large area of contact. The steel-trowelled HAC mortar surface (Figure 1b) exhibits three pinpoint contact (ringed to highlight their positions).

EXPERIMENT

A simple experiment was conceived to explore the effect of these different casting techniques on measured compressive strength. In the first series of tests, a single homogeneous batch of concrete was cast. A slab of concrete (from which cores were later drilled and extracted) and a set of test cubes were cast using this concrete. As far as possible, variables that could influence the measured strength of these items were controlled (including compaction, wet curing conditions, temperature, etc). Twenty cores
Table 1 Strength results

<table>
<thead>
<tr>
<th>Specimen number</th>
<th>Series one tests</th>
<th>Series two tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cube strength (MPa)</td>
<td>Sulphur core strength (MPa)</td>
</tr>
<tr>
<td>1</td>
<td>28.0</td>
<td>27.6</td>
</tr>
<tr>
<td>2</td>
<td>28.2</td>
<td>27.3</td>
</tr>
<tr>
<td>3</td>
<td>28.2</td>
<td>26.9</td>
</tr>
<tr>
<td>4</td>
<td>27.7</td>
<td>26.1</td>
</tr>
<tr>
<td>5</td>
<td>26.8</td>
<td>23.5</td>
</tr>
<tr>
<td>6</td>
<td>28.5</td>
<td>26.0</td>
</tr>
<tr>
<td>7</td>
<td>21.8**</td>
<td>13.9</td>
</tr>
<tr>
<td>8</td>
<td>26.7</td>
<td>13.2</td>
</tr>
<tr>
<td>9</td>
<td>26.1</td>
<td>16.7</td>
</tr>
<tr>
<td>10</td>
<td>26.0</td>
<td>12.3</td>
</tr>
<tr>
<td>Average (MPa)</td>
<td>27.9</td>
<td>26.3</td>
</tr>
<tr>
<td>Deviation from ave cube strength (%)</td>
<td>5.9</td>
<td>46.6</td>
</tr>
<tr>
<td>Standard deviation (MPa)</td>
<td>0.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
<td>2.2</td>
<td>4.6</td>
</tr>
</tbody>
</table>

* Core 7 capped with sulphur mortar had a protrusion on the surface resulting in an outlier.
** One cap debonded. The core was therefore tested with only one side capped. It was therefore not compliant with SABS 865:1994. Consequently, this result was discarded.
Guidelines for the preparation of papers and technical notes

Authors should comply with the following guidelines when preparing papers for publication in the journal

SUBMISSION PROCEDURES AND REQUIRED FORMAT

- Hard copies: Hard copies should only be submitted if the author is unable to submit electronically, for whatever reason. In such a case only one unstapled copy, together with an electronic copy on CD, should be posted to the Chief Executive Officer of SAICE, Private Bag X2000, Halfway House, 1685, or delivered to Building 19, Thornhill Office Park, Bekker Street, Midrand. Hard copies should be printed on one side of A4 paper.

- Electronic submission: Manuscripts should be e-mailed to the editor at verelene@saice.org.za. File sizes should preferably not exceed 3 MB per e-mail – figures may for example be sent one by one or in groups not larger than 3 MB. Manuscripts should not be sent in PDF format as this precludes reviewing of papers per track changes.

- Format: Manuscripts should be prepared in MS Word and presented in double line spacing, single column layout with 25 mm wide margins. All pages should bear the authors’ names and be numbered at the bottom of the page. With the exception of tables and figures (see below) the document should be typed in Times New Roman 12 pt font. Contributions should be accompanied by an abstract of not more than 200 words and a profile not exceeding 100 words on each of the authors.

- First page: The first page of the manuscript should include the title of the paper, the initials and surnames of the authors, professional status (if applicable), SAICE affiliation (Member, Fellow, Visitor, etc), telephone and fax numbers, and e-mail and postal addresses. The name of the corresponding author should be underlined. Five keywords should be suggested. The copy must not contain instructions to the editor.

- Figures, tables, photos and illustrations: These should preferably be submitted in colour, as the journal is distributed internationally. Articles submitted in any of the other official South African languages should be accompanied by an expanded abstract in English.

- Discussion: A discussion, if applicable, is invited. This should be kept to a maximum of 1 500 words. Where appropriate, discussion contributions will be subject to the normal reviewing process and will be forwarded to the authors of the original article for reply.

POLICY REGARDING LANGUAGE AND ORIGINALITY OF SUBMITTED ARTICLES

- Language: Manuscripts should preferably be presented in English, as the journal is distributed internationally. Articles submitted in any of the other official South African languages should be accompanied by an expanded abstract in English.

- Original work: Papers and technical notes must be original contributions. Authors must affirm that submitted material has not been published previously, is not under consideration for publication elsewhere and will not be submitted elsewhere while under consideration by the SAICE Journal Editorial Panel. It is the responsibility of the authors to ensure that publication of any paper in the journal will not constitute a breach of any agreement or the transgression of any law. The corresponding author should confirm that all co-authors have read and approved the manuscript and accept these conditions. Authors are responsible for obtaining permission to publish experimental data and other information that may be confidential or sensitive. Authors are also responsible for obtaining permission from copyright owners when reproducing material that had been published elsewhere.

- Mathematical expressions: Equations should be typeset in a clear form which can be easily read by non-mathematicians. Symbols should be typed using the Times New Roman symbol set. Equations should be italics. Each equation should appear on a separate line and be numbered consecutively.

- Symbols should preferably be defined in the text, but if this is not feasible, a list of notations may be provided for inclusion at the end of the paper. Symbols should conform to accepted international usage. The SI system of units should be used. Decimal commas may be used, but decimal points are preferred.

- Since mathematical formulae are generally scanned in during the production process, they should be set out as they are to appear in the printed paper.

- Headings: Sections and paragraphs should not be numbered. The following hierarchy of headings should be followed:

  HEADING OF MAIN SECTION
  Heading of subsection
  Heading of sub-subsection

- References: References should follow the Harvard system. The format of text citations should be as follows: ‘Jones (1999) discovered that …’ or ‘recent results (Brown & Carter 1985; Green et al 1999) indicated that …’

- References cited in the text should be listed in alphabetical order at the end of the paper. References by the same author should be in chronological order. The following are examples of a journal article, a book and a conference paper:

- Footnotes, trade names, acronyms, abbreviations: These should be avoided. If acronyms are used, they should be defined when they first appear in the text. Do not use full stops after abbreviations or acronyms.

- Copyright: On acceptance of the paper or technical note, copyright must be transferred by the author to the South African Institution of Civil Engineering on the form that will be provided by the Institution.

- Photos of authors: The final corrected version of the paper should be accompanied by recent head and shoulders colour photographs of the authors.

- Proofs: First proofs of papers will be sent to authors in PDF format for verification before publication.