STATEMENT OF CORRECTIONS

Examiner 1

Understanding the nature and purpose of the investigation

The examiner's view is that I have introduced the aspect of stochastic modelling, highlighted the problems that exist as well as potential solutions that could solve the problems. I appreciate these compliments from the examiner.

Relevant literature

In response to the examiner's comments that a brief attempt was made to discuss the strengths and weaknesses of other approaches used for stochastic generation of data relevant for water resources applications, the following chapters in the Literature Review explain how this was done in the Research Report.

It is assumed that studies in EMD have evolved over time but generally being a relatively new data generation technique, most studies tabulated in Table 2.2 from Page 12-17 are for the purposes of breaking down a time series into components of IMFs and then evaluating the significance of those IMFs rather than modelling or generating new data. This finding is explained in Chapter 1, page 18. Thus as a data generation technique, EMD has not evolved much. I have rearranged Table 2.2 to detail a chronological update of the studies that have been carried out over time as proposed by the examiner.

The weaknesses and strengths of approaches used for stochastic generation of data relevant for water resources applications have been introduced in Chapter 1. In the first paragraph of page 2, I have explained the preferable choice of rainfall generators from streamflow generators in water resources assessment because it is more realistic and easier in catchment hydrological assessment. In page 2, paragraph 2, I have further explained in detail the choice of non-parametric data generation approaches due to their simplicity as opposed to parametric and linear approaches that are mostly characterized by large number of parameters .

I have also highlighted that water resources planners and managers prefer simpler models that are easier to understand in page 2. Further on, in page 3 paragraph 2, I have noted the preferable choice of monthly generators as compared with the daily because of simplicity since planning and management of water resources is most often done on a monthly time scale. The property of EMD to quantify proportions of hydro climatic time series variations at different time scales is important for sustainable management of water resources systems (Peel et al, 2005). This is detailed in page 4.

In Chapter 2, Table 2.1 page 7-8 , five non-parametric models; the wavelet (Bayazit et al, 2001; Ünal et al, 2004; Wang et al, 2011), the Reordering model (Clark et al., 2004a, b; Mehrotra and Sharma, 2009), Nearest Neighbour (Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Srikanthan et al, 2002; Yates et al, 2003; Mehrotra et al, 2006; Sharma and Mehrotra, 2006; Prairie et al, 2006), Kernel Density approach (Sharma et al, 2003; Srikanthan et al, 2005; Wang and Ding, 2007; Mehrotra and Sharma, 2007, 2009) and the VLB (Ndiritu, 2011a, b) were identified on the following strengths;

1. Ability to preserve historic characteristics
2. Ability to extrapolate beyond the range of historic data
3. Limitations of applicability
4. Possibility of generating negative values and
5. Ease of use

It is with these strengths and weaknesses that the VLB streamflow generator was the choice being adapted for rainfall data generation and to be coupled with EMD.

Understanding the science

The examiner appreciated that I understood the science relevant for the investigation as well as correctly interpreting the results as contained in Chapter 4.

Assessment of the significance of the results.

The generator's performance was assessed on the replication of generated statistics to the historic ones using a multi-site rainfall generation problem in which the closeness of the historic statistics to the generated are determined and plotted. The model's observed closeness demonstrates that the generator performed adequately and vice versa. The examiner appreciates that this was fairly done.

On the significance of the results to water resources planning and modelling the examiner is of the view that a discussion was not presented. However I have included this whereby the significance of Figures 4.6 -4.15(in Page 41- 45), I have explained in Chapter 4.1 paragraph 1 in page 36 that minimum run sums are essential to determine how the generator replicates long term dependence characteristics that is commonly used in water resources assessment. This is further discussed in page 64, Section 4.4 and Chapter 5 pages 65-66 in in the conclusion.

More significance of the results would entail the use of the model to simulate actual water systems so as to obtain a practical significance.

The examiner's view is that quantitative evaluations of the generator have not been presented. Additional quantitative evaluations are carried out for the yearly statistics in Tables 4.4-4.5 (page 51) and for the monthly in Tables 4.41 and 4.42 (page 62) in terms of how the means and medians of the simulated values differ from the historic.

The examiner indicates that the simulated monthly rainfall differ considerable from those of the observed. EMD-VLB generator obtains 80.2 % of the historic statistics falling within the inter-quartile range of the simulated statistics. In stochastic data generation, this signifies a relatively good generator that can be used for practical water systems assessment.

Further corrections

Referencing

I have corrected the references to reflect a chronological format where the references are separated by commas and more than one reference are enclosed by brackets. This has been carried out throughout the report.

Equation 2.1 in page 10 has been corrected to;

$$x(t) = \sum_{i=1}^{n} IMF\ (i) + \ e(n) \qquad 2.1$$

where, $x(t) -$ is the time series data being analyzed.

$IMF\ (i) -$ is the Intrinsic mode functions at time t.

$e(n) -$ is the Residual of the data set.

$i = 1, 2, 3, \ldots, n -$ is the number of extracted IMFs.

The meaning of notation i was not previously represented.

Figure 3.2.

On the x-axis, the months signify hydrological years (and not calendar years) that starts from July for the summer rainfall region and start from January for the winter rainfall region. All the stations except station 0020866 W are in the summer rainfall region. Thus plotting all of them on the same graph outlines better visibility than otherwise.

The examiner mentions two stations could be in the winter rainfall region but station 0052590 W is in the summer rainfall region. Its location is at the north-eastern tip of the Western Cape that borders both the Eastern Cape and the Western Cape.

Fig 3.3

The examiner indicates that a bar graph should be used to indicate the cross correlation coefficients for the rainfall stations. Representing a cross correlation in each station for all the stations would mean having 10 bar graphs each with 9 bars which comparatively, the line graph used in this case becomes better visible. Since the examiner stresses the need for a reduction of graphical illustrations, this would lead to an enormous increase in the number of graphs. It is thus my view that retaining the line graph as a means of showing the monthly cross correlation coefficient is the preferable option.

Figure 3.4

I have outlined that the X-axis indicates lag-1 monthly serial correlation for an easier understanding of the graph in the narration of the figure. This appears in Chapter 3, page 25.

Figure 3.5

This figure summarizes Figures 3.3 and figure 3.4. both being placed on the same graph. Average serial correlation for example in month 1 is calculated by obtaining the average for all the stations in that month. The same is carried out for the average cross correlation coefficients. An explanation to this effect has been added on page 25 below Figure 3.4.

Figure 3.7

The **No** and **Yes** from the diamond logical decision stages have been added in page 30.

Figure 4.6 -4.23

I have added a box plot legend to explain the percentiles used in the stochastic simulations as below in Page 40;

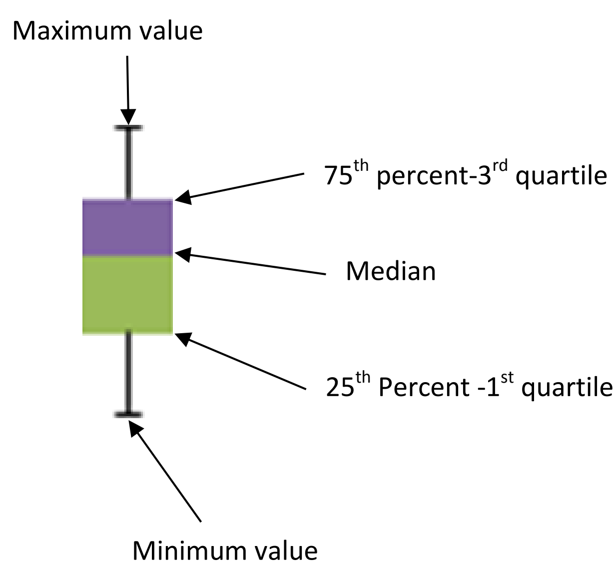

Figure 4.16-4.23

The line graphs connecting the stations are not meant to indicate variations in the historic statistics from one station to another but rather to enable easier visualization of those statistics.

Preserving the within year, and inter annual serial correlations

The examiner indicates that in Chapter 1, there is considerable discussion about the importance of preserving the correlations but this is mentioned in page 4 that one of the reasons behind the selection of the VLB generator was that it is able to preserve monthly serial correlation between end of one year and the beginning of the next year in streamflow modelling; a property that is a hindrance to many generators (Ndiritu 2011a). It is however seen that in Chapter 3, Table 3.2 and Figures 3.4 and 3.5 Annual Serial Correlation between stations are very low; averaging -0.03. There is no therefore no need for their preservation in rainfall generation. I have added this assessment below Figure 3.5 in page 26.

The examiner suggests that all the equations should be in brackets and I have effected this to all equations.

Equation A.1

x_1, x_2, x_3 are already defined as being in the range of $[x_1, x_n]\ for\ all\ i = 2, \ldots, n-1$ in page A77.

Appendix

I have included prefixes A and B in numbering both A and B appendices from page A77 to page B93.

Examiner 2

1. Introduction

Examiner 2 indicates that I should have used more than one catchment but the aim of the research was to investigate the performance of the model on the South African as a whole in an area that entails the multiple catchments represented by the whole country. This had been carried out by Ndiritu and Nyaga (2014) using VLB and the same catchment used to simulate by PEGRAIM (Pegram, 2011) after which the VLB was seen to perform relatively better than the PEGRAIM model. It was then considered that EMD as an improvement would be used and still undergo the same comparison to evaluate its performance. Pegraim-W (2011) is currently being used to simulate water resources systems to assess its practicability. Therefore using more than one catchment is goes beyond the scope of this Research Report. As a follow up in a separate research, I am investigating the performance of the EMD model on a smaller catchment within South Africa.

2. General comments on the thesis

1. I acknowledge that the examiner's appreciation that the Research Report is presented in good English and in a precise language making it easier to read and understand.
2. The examiner observes that it is overloaded with graphs.

The purpose of the Research was to develop a stochastic model and use the historic rainfall statistics to compare the generated ones. Quantitative graphical presentation was the preferred method of this comparison.

3. The examiner notes that formatting is not consistent by the inability to size Figures/Tables. Notably the examiner has marked page 24, 25, 41,43, 104 and 105 as the margins to be adjusted. I have adjusted the respective pages to fit on the margins.
6. Page numbers have been altered whereby chapter 1 starts at page 1 and end normal page numbering at the end of references IN PAGE 76. Appendixes have been altered to have the prefixes A starting from page A77 to page A82 in appendix A and prefix B from page B83 to page B93 in appendix B.

Detailed comments

Chapter 1

The examiner says that the background is too short to properly introduce the study as it should which contrasts examiner 1's comments. I in Chapter 1 page 1, paragraph 2, a further explanation of stochastic modelling.

Problem statement

I have added the problem statement in page 4 of the Research Report.

The objectives

The examiner notes that the objectives should be a bit more specific so that the conclusions have something specific to answer to. On page 5, I have outlined 2 objectives that are very specific namely;

1. To develop a simple, data-adaptive monthly stochastic rainfall generator by complementing the strengths of Empirical Mode Decomposition (EMD) and the Variable Length Bootstrap (VLB).

2. To assess the performance of the generator using representative rainfall data from South Africa.

These objectives have been answered to in the conclusions and recommendations in Chapter 5, page 65 and 66.

Arrangement of the dissertation

This is contained in Chapter 1 section 1.4 " Organization of the Research Report" .

The examiner indicates that there should be a conclusion of the literature Review pointing towards the methodology. I have included it as Section 2.4 in Pages 19-20 as Summary of the Literature Review.

Chapter 4

I have reduced the sizes of the graphs (that are as a result of statistical simulations in the main report) for four graphs to fit one page as opposed to two as was previously. I however cannot reduce the number as this forms part of the quantitative results of simulations that examiner 1 indicated that I should present.

Chapter 5

Conclusions and Recommendations.

It is not clear by the examiners statement ' Itemised References as they would be answering the Objectives would answer them better'

References 6

The references are matched.