

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Estimating hill country rainfall without full data sets for the Manawatu River Catchment

**A thesis presented in partial fulfillment of
the requirements for the degree of
Master of Resource and Environmental Planning**

**At Massey University, Turitea,
New Zealand.**

Jorn Arjan Sijbertsma

2014

Abstract

Nowadays, people anticipate floods using flood warning systems, and building stop banks and flood ways in place that use flood models generated with hydrological information in their design. Nevertheless, various regions in the world are still hit by floods with catastrophic effects to urban areas, because of a lack of local hydrological knowledge, especially of upstream areas in their catchments. This lack of hydrological knowledge is a result of difficult accessible highly elevated upstream areas, which makes monitoring of hydrological variables difficult or impossible.

This thesis examines models for determining montane rainfall using spatial estimation methods and data sets. The distribution and quantity of montane rainfall were assessed by applying five appropriated spatial estimation methods, data of historical and current rain gauges, and a performance measurement.

The methodology applied to gain more knowledge about montane rainfall was established with the results of a literature analysis of 40 articles about montane rainfall. This literature analysis revealed that *ordinary kriging* is the most frequently applied spatial estimation method for montane rainfall, with *regression* and *regression kriging* completing the top three of the most applied methods. Also, two other spatial estimation methods, *empirical Bayesian kriging* and *geostatistical simulation*, performed well with rainfall data. The same literature analysis disclosed that the root mean square error was predominantly used as a performance measure of spatial estimation methods.

The literature analysis revealed a number of data gap-filling techniques, with *the inverse distance weighting* method and the *coefficient of correlation weighting* method as the two most suitable techniques. These techniques were applied to complete historical rainfall data sets and their performance was compared within this research. The result showed that the *coefficient of correlation weighting* method outperformed the *inverse distance weighting* method in 74% of all data gap estimations, and the *coefficient of correlation weighting* method was 22% more accurate (based on the overall performance) than the *inverse distance weighting* method. The most accurate data gap-filling technique, the *coefficient of correlation weighting* method, was used to complete the historical rain gauges data.

The overall ranking of the spatial estimation methods revealed that *Gaussian geostatistical simulation* performed the best. *Regression kriging* was the second best spatial estimation method, but there was no significant difference with *Gaussian geostatistical simulation*. At the same time, the results showed that the best performance of the spatial estimations was accomplished without the maximum number of rain gauges. However, better visual

representation of the distinct pattern of rainfall was generated with the historical rain gauges in the second and third experiment of the spatial estimations.

Finally, this research discussed the factors that can impact the performance of the spatial estimations. Two of these factors were the removal of “bad data” and the strategic placing of rain gauges. The results of this research clearly defined that the removal of “bad data” increased the accuracy of estimation, while a more even and strategic distribution of rain gauges was suggested to increase the accuracy of the spatial estimation of rainfall.

Acknowledgements

With sincere gratitude, I would like to thank a number of individuals and organisations. Firstly, I would like to thank my supervisors Dr. Jeff McNeill and Rachel Summers for their support, encouragement, and advice during my Thesis. Especially, Jeff's logic towards this research was particularly appreciated, as well as his ideas and feedback. Also, I acknowledged Rachel's GIS knowledge throughout this research.

I'm grateful to the regional council of the Manawatu-Wanganui region, also known as Horizons Regional Council, for some funding, providing data, and using of their resources. Particularly, I would like to acknowledge Sean Hodges, Jeff Watson, and Barry Gilliland for their thoughts and feedback.

The National Institute of Water and Atmospheric Research (NIWA) and Greater Wellington Regional Council provided data for no cost. Hereby, an appreciation to both organisations because their data sets played an important part in the performance measure of the spatial estimation and the estimation of the data gaps in the historical data sets.

Dr. Angela Feekery provided her expertise in writing a thesis, and feedback on my chapters. Thanks to her, I have improved my writing style.

Contents

Abstract	i
Acknowledgements	iii
Contents	iv
Tables	viii
Figures	x
Chapter 1: Montane Rain Quantity	1
1.1 Context	1
1.2 Scope and limitations.....	2
1.3 Overview of the study	3
Chapter 2: Literature Review	4
2.1 Introduction	4
2.2 Recent reviews.....	4
2.2.1 Bibliometric analyses	4
2.2.2 Applications of spatial estimation methods	5
2.2.2.1 <i>Spatial estimation methods overviews and decision trees</i>	6
2.2.3 Summary.....	8
2.3 Montane rainfall estimation	9
2.3.1 Selection of journal articles	9
2.3.2 Comparison of 40 studies	10
2.3.3 Challenges in reviewing the 40 studies	11
2.3.4 Results	11
2.4 Further investigation	13
2.4.1 Empirical Bayesian kriging	14
2.4.2 Geostatistical simulation	15
2.4.3 Data gap-filling techniques.....	16
2.4.3.1 <i>Gas concentrations</i>	16
2.4.3.2 <i>Climatology and hydrology data</i>	17
2.4.3.3 <i>Rainfall data</i>	18
2.4.4 Summary.....	19
2.5 Chapter summary	20
Chapter 3: Research Design	22
3.1 Introduction	22
3.2 Selection of methods, software, and data sets	22
3.2.1 Spatial estimation methods.....	22
3.2.2 Performance measure	24

3.2.3 Auxiliary data sets	25
3.2.4 Data gap-filling technique	25
3.3 Method for assessing the relative efficacy of estimation methods	26
3.3.1 Experiments	26
3.3.2 Comparison	27
3.4 Spatial estimation	28
3.5 Data sets	29
3.5.1 Primary rainfall data set and additional rainfall data points	29
3.5.2 Historical rain gauge values determined with data gap-filling techniques	31
3.5.3 Validation data set	31
3.6 Summary	31
Chapter 4: Data Sets	33
4.1 Introduction	33
4.2 Selection of the area	33
4.2.1 Geography characteristics	33
4.2.2 Rainfall characteristics	33
4.2.3 Rainfall data	35
4.2.3.1 <i>Rainfall data for spatial estimation</i>	35
4.2.3.2 <i>Rainfall data for the validation of the spatial estimation</i>	39
4.2.3.3 <i>Rainfall data used in the data gap-filling techniques</i>	39
4.3 Selection of the temporal data extent	41
4.4 Summary	42
Chapter 5: Results	44
5.1 Introduction	44
5.2 Data gap-filling	44
5.2.1 Challenges	44
5.2.2 Comparison	44
5.3 Principal components analysis	46
5.4 Spatial estimation methods	46
5.4.1 Performance comparison	47
5.4.2 The ranking assessment per temporal data set	47
5.4.3 The final ranking assessment	49
5.4.4 Rainfall distribution in the Manawatu River Catchment	50
5.4 Summary	50
Chapter 6: Analysis	51
6.1 Introduction	51

6.2 Validity of results.....	51
6.3 Comparison of the data gap filing techniques	52
6.4 Most appropriate spatial estimation method to manage sparse data.....	52
6.4.1 Empirical Bayesian kriging	53
6.4.2 Linear regression	54
6.4.3 Ordinary kriging.....	54
6.4.4 Regression kriging	56
6.4.5 Gaussian geostatistical simulation.....	57
6.4.6 Summary.....	58
6.5 Strategically placing limited rain gauges in montane country	59
6.6 Rainfall distribution in the Manawatu River Catchment	60
Chapter 7: Conclusion	62
7.1 Recommendations.....	63
References	65
Appendix 1: Comparison overview of SEMs by Burrough and McDonnell (1998)	72
Appendix 2: Comparison overview of SEMs by Johnston, et al. (2001)	75
Appendix 3: Comparison overview of SEMs by Basistha, et al. (2008)	76
Appendix 4: Decision tree for selecting Gaussian disjunctive kriging and indicator kriging by Lark and Ferguson (2004)	78
Appendix 5: Decision tree for selecting a SEM by Li and Heap (2008)	80
Appendix 6: The 40 selected studies	82
Appendix 7: The three databases with the number of identified publications per year	89
Appendix 8: Background information of the literature analysis	90
Appendix 9: The frequency and number times best performed of all SEMs in a review of 53 studies by Li and Heap (2008, 2011) and in this review of 40 studies	92
Appendix 10: Spatial estimation methods	95
Appendix 11: Data gap-filling techniques	100
Appendix 12: Root mean square error	101
Appendix 13: Primary rainfall data set	102
Appendix 14: Historical rainfall data set	105
Appendix 15: Rainfall data set covering rain gauges outside the catchment	108
Appendix 16: Validation rainfall data set	111

Appendix 17: NEMS' quality code schema for rainfall recording	114
Appendix 18: Rain gauges used to determine the historical data with data gap-filling techniques	115
Appendix 19: Generating slope data	117
Appendix 20: Ordinary kriging	119
Appendix 21: Linear regression	123
Appendix 22: Regression kriging	127
Appendix 23: Empirical Bayesian kriging	137
Appendix 24: Gaussian geostatistical simulation	142
Appendix 25: Principal components analysis	146
Appendix 26: Performance comparison of the spatial estimation methods applied to a data set per experiment	149
Appendix 27: The ranking assessment per temporal data set	158
Appendix 28: Maps of all experiments and spatial estimation methods that generated the lowest root mean square value	161
Appendix 29: Comparing the daily rainfall maps with the measured rainfall	188
Appendix 30: Annual rainfall maps by NIWA (2013) and by Burgess and New Zealand Meteorological Service (1988)	197
Appendix 31: Map outputs generated for all nine rainfall data sets with the most appropriate spatial estimation method Gaussian geostatistical simulation	199
Glossary	208

Tables

Table 2-1: The effectiveness of the top five of the most preferred estimation methods of the 40 studies	13
Table 2-2: The top three of each reported aspect	13
Table 2-3: The top 5 of the most applied SEMs in Li and Heap's review (2008, 2011) and this review	14
Table 2-4: Data gap-filling techniques that perform well	19
Table 3-1: The effectiveness of the top five of the most used SEMs of the literature analysis	23
Table 3-2: The selected SEMs	23
Table 3-3: The three predominant software products	24
Table 3-4: Software	24
Table 3-5: Top three of the most applied diagnostic statistics of the 40 readings	25
Table 3-6: Top three of the most incorporated auxiliary data sets of the 40 readings	25
Table 3-7: Experiments	26
Table 3-8: Performance comparison of the spatial estimation methods applied to a data set per experiment	27
Table 3-9: Example of the ranking assessment table for a temporal data set	28
Table 3-10: Example of the final ranking assessment table	28
Table 3-11: The main variables of the five SEMs	29
Table 4-1: Summary of all applied rainfall data sets	41
Table 4-2: The selection of temporal rainfall data sets	42
Table 5-1: The 2009 data gap (1 year data gap) filling results of the 18 historical rain gauges	45
Table 5-2: Overview of the number of times that each data gap technique performed better than the other in percentage for all data gap analysis	45
Table 5-3: The total increase in accuracy of the three runs of methods	46
Table 5-4: Minimum root mean square values for each experiment and SEM applied to a data set	48
Table 5-5: Summary of the best method for each data set, as determined by average ranking of each SEM per experiment	49

Table 5-6: The final ranking assessment	49
Table 6-1: Minimum root mean square values generated by that experiment per SEM per data set	51
Table 6-2: R-squared values of all rainfall data sets with rainfall and/or slope data	55
Table 6-3: Factors that may impact the performance of the SEM	59

Figures

Figure 2-1: A decision tree by Pebesma (2004)	7
Figure 2-2: A decision tree by Hengl (2009a)	9
Figure 2-3: The geographical locations of the 40 studies applied in the literature analyses	10
Figure 2-4: The frequency in percentage of all SEMs applied in the 40 articles	12
Figure 3-1: The three experiments	30
Figure 4-1: The Manawatu River Catchment	34
Figure 4-2: The rainfall distribution in the Manawatu River Catchment is influenced by weather systems	36
Figure 4-3: The geographical location of all rain gauges involved in the spatial estimation	38
Figure 4-4: The geographical location of all rain gauges involved in the data gap-filling techniques	40
Figure 4-5: Temporal extents of all applied rain gauges	43
Figure 6-1: The semivariogram of the 2009 data set (experiment 3) in ordinary kriging	56
Figure 6-2: The semivariogram of the June 2009 data set (experiment 1) in ordinary kriging	56