Investigation of Enhancements to Two Fundamental Components of the Statistical Interpolation Method used by the Canadian Precipitation Analysis (CaPA)

by

Andrea Marie Evans

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Civil Engineering

University of Manitoba

Winnipeg

Abstract

The Canadian Precipitation Analysis (CaPA) generates gridded precipitation data outputs based on the assimilation of both observation and climate model data. CaPA outputs are highly valuable to modelling efforts dependent on precipitation inputs, and as such the quality of CaPA outputs is crucial. Two improvements to CaPA were investigated: reducing transformation bias though correction against moving-window averaged CaPA output that avoids transformation, and enhancing semivariograms through anisotropy and convection considerations. Accounting for convection in the semivariogram proved ineffectual, while the bias correction technique and anisotropic semivariograms both reduced bias and improved related metrics. No methods improved the Equitable Threat Score. If implemented separately, the bias correction or anisotropic semivariogram approaches will yield targeted benefits for CaPA users, particularly for applications focused on extreme precipitation values. Improvements were not so comprehensive as to warrant adoption in the operational CaPA configuration, although availability in experimental versions is recommended.

Acknowledgements

I would like to express my sincere gratitude to Dr. Peter Rasmussen, from the Department of Civil Engineering, for providing excellent guidance, advice, and technical expertise throughout my Master's degree. His supervision was invaluable to my efforts as a graduate student to absorb as much information as possible, and has provided me with an enduring appreciation for academic curiosity and the art of applying one's knowledge in intelligent ways.

My thanks goes out to my remaining committee members, Dr. Tricia Stadnyk from the Department of Civil Engineering, and Dr. Elif Acar from the Department of Statistics. Your thoughtful comments and guidance were greatly appreciated.

I wish to thank Environment Canada for accommodating my research needs in relation to CaPA, and for their generous permission to use the Canadian Lightning Detection Network data. Specifically, I would like to acknowledge Vincent Fortin and Franck Lespinas from the Meteorological Research Division at the Canadian Meteorological Centre, for sharing with me both the software for, and their knowledge of, the CaPA program.

My gratitude is also extended to Manitoba Hydro and the Natural Sciences and Engineering Research Council, for their support of this project. I would also like to acknowledge the financial support that I received from the following scholarships: the Duncan Norrie Memorial Scholarship, the Duncan Norrie Memorial Scholarship

in Water Resources, the Engineering Class of 1946 Graduate Scholarship, the UMA - John Allen Knowles Memorial Scholarship, and the Faculty of Graduate Studies Special Award.

I would like to thank my mother, Sandra Harding, for providing me with an inspiring and unfailing example of the merits of hard-work and determination, and the rest of my family, friends, and fellow colleagues for their encouragement. Finally, I would like to thank Shane Wruth for always providing sober second thought on my technical dilemmas and solutions, courageously offering advice and optimism during times of frustration, and bringing a bit more mirth to the graduate experience.

To the strongest people I know; my grandfather Marshall T. Harding, an engineer at heart, and my grandmother Doreen J. Deleu, whose hard-won wisdom is always so graciously shared.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

The creation of high-quality gridded precipitation data is an intricate endeavor. However, it is a task which can yield great benefit when completed successfully; accurate precipitation inputs can lend significant additional merit to any of a range of decision making, planning, and assessment tools which require precipitation inputs. This in turn translates into positive impacts for both private and public sectors, through applications such as: flood forecasting and modelling, agricultural and soil moisture management and planning, and hydrological and hydraulic modelling crucial to the planning of flow dependent activities, such as hydroelectric generation (*Mahfouf et al.*, 2007).

Currently, insufficient and inconsistent spatial and temporal coverage are key concerns for existing precipitation data records (*Mahfouf et al.*, 2007). Thus, to address the critical issue of gridded precipitation data quality, the Canadian Precipitation Analysis (CaPA) program has been created under the auspices of Environ-

ment Canada. CaPA addresses the quality problem through the use of numerical atmospheric modelling output, in combination with observation fields, to produce a temporally regular and spatially gridded precipitation product (*Mahfouf et al.*, 2007). This process is accomplished through the application of the geostatistical-type Statistical Interpolation (SI) methodology. As the driving force behind the combination of observation station and model data, the skill and accuracy of the SI technique lies at the heart of CaPA's performance capabilities.

The importance of producing high-quality gridded precipitation estimates is underscored by the benefits of applying such data in a hydrological context. Gridded precipitation estimates can offer a means of combating the lack of spatial variability information that is available when only point-form precipitation observations are used (*Schneider and Steinacker*, 2009). Enhancing the spatial details of precipitation is a significant accomplishment for hydrological applications. For example, a study by He et al. (2011) has shown that streamflow in small scale sub-catchments, with areas less than $400\,\mathrm{km}^2$, can exhibit sensitivity to the spatial variation of precipitation data. Gridded precipitation estimates are of particular use to distributed hydrological modelling applications (*Goudenhoofdt and Delobbe*, 2009).

Improved precipitation estimates can also enhance the calculation of extreme-value precipitation statistics, such as depth -duration-frequency or intensity-duration-frequency curves (*Overeem et al.*, 2007). Currently, the most commonly used data source for such statistics is station data, which brings into play the standard obser-

vation data problems related to low temporal and spatial resolution. This makes the robust calculation of such extreme-value statistics impossible in many regions (*Overeem et al.*, 2007). A study by Overeem et al. (2009) investigates the possibility of using radar data corrected with station data to generate such extreme-value statistics. While the relatively short radar record length and persistent error of the radar field hindered the extreme-value statistic's estimation somewhat, the approach shows promise. As the number of time periods for which CaPA analyses are available continues to grow, such an application for CaPA data will become possible, and may lead to more accurate extreme rainfall statistics, particularly for regions with sparse observation networks. Increased accuracy in extreme-value precipitation statistics can directly improve water management undertakings (*Overeem et al.*, 2007), including the design of urban drainage systems and flood infrastructure.

In the future, the Canadian Land Data Assimilation System (CaLDAS) can also become a direct recipient of gridded CaPA analyses. CaLDAS has been developed by the Meteorological Service of Canada, and is designed to provide land surface information to numerical weather prediction (NWP) models. The type of information offered by CaLDAS is key to understanding energy and water interactions at the land surface, where both can have a direct influence on the atmosphere. Beyond weather prediction, an NWP model enhanced by CaLDAS can be expected to supply information on projected crop yields, flood and drought potentials, and forest fire risks (*Balsamo et al.*, 2009).

CaPA has the potential to become an important player in the CaLDAS system. CaPA can provide precipitation information as a direct input into the CaLDAS model, where water storage and hydrology are directly considered through a Soil-Biosphere-Atmosphere modelling approach (*Balsamo et al.*, 2009). The role of CaPA in the CaLDAS system would be to provide background precipitation information, alongside the assimilation of observational and satellite data (*Balsamo et al.*, 2009). This contribution from CaPA would be particularly valuable, as inconsistencies in the land surface can be significant between gauge locations, making the prediction of soil moisture at unobserved locations difficult if only nearby values are to be relied on (*Balsamo et al.*, 2009).

Improved soil moisture and temperature estimates in the northern regions of Canada, obtained by fueling CaLDAS with CaPA outputs, may also improve the modelling of permafrost-type processes and snow melt (*Balsamo et al.*, 2009). These benefits can be expected despite the relatively low proportions of observations assimilated by CaPA in the winter months, due to CaPA's ability to improve upon precipitation estimates contributing to antecedent moisture conditions. Considering the importance that this can have not only on hydrological activities, but also on local infrastructure and communities, this potential increase in information in regions otherwise experiencing a dearth of observational data may prove to be indispensable.

Clearly, CaPA has much to offer users of precipitation data. The final value of what CaPA has to offer, however, is contingent upon the accuracy of its analyses. If

poor quality products are generated by CaPA, then their inclusion in any application will be of limited benefit. To ensure that the best possible analysis quality is attained, the very core of the program must be examined; the methodologies of CaPA's SI algorithm are placed under scrutiny in this thesis. Focus is honed in on two particular aspects of the SI method. First, the implications of data transformation on the final bias of analyses is reviewed and treated. Secondly, an upgrade from isotropic to anisotropic semivariograms to describe spatial relationships in the input data is evaluated.

Prior to delving into the two SI improvement topics presented in this thesis, a basic understanding of the study setup and data sources is required. Sections 1.1 to 1.3 supply this information, along with a brief outline of the CaPA program itself, a review of both the SI method and the verification scores used to assess the proposed modifications to CaPA, and an outline of the study objectives and thesis formatting.

1.1 Background

An appreciation of the CaPA program, with its current features, methodologies, and drawbacks, is required in order to understand the significance of the study presented here, which seeks to improve upon the foundational SI technique used by CaPA. In order to facilitate the growth of such an appreciation, a brief review of the CaPA program is provided subsequently, along with a summary of the SI method used by CaPA, the standard verification skill scores applied to assess CaPA products, and a

preliminary discussion regarding the basic attributes of the study.

1.1.1 The Study Domain, Time Period, and Data

Manitoba is a province intimately familiar with water. For example, hydroelectric power generation and flood forecasting are two activities that occur regularly in Manitoba, resulting in high demand for quality precipitation products. In recognition of this, and of the participation of Manitoba Hydro in the on-going CaPA research efforts, the Canadian prairies were selected as the domain of interest for this investigation. The specific delineation of the study domain was determined by the pre-defined prairie region grid available in CaPA, and is depicted in Figure 1.1.

As can be seen, this domain extends from the eastern half of British Columbia to the western half of Ontario, and from the northern United States to the southern portions of the territories. While this represents a considerably larger area than the province of Manitoba, it still contains this region of interest. Furthermore, it also contains a relatively diverse array of landscapes, including such regions as the prairies, the Canadian shield, the Boreal forest, and the Rocky Mountains. This helps to ensure that geographical or large scale weather influences on precipitation will be captured by the selected study region, and that the results of the improvements made will be applicable to a wide variety of applications throughout the full CaPA domain.

The depiction of the Global Environmental Multiscale (GEM) and observation

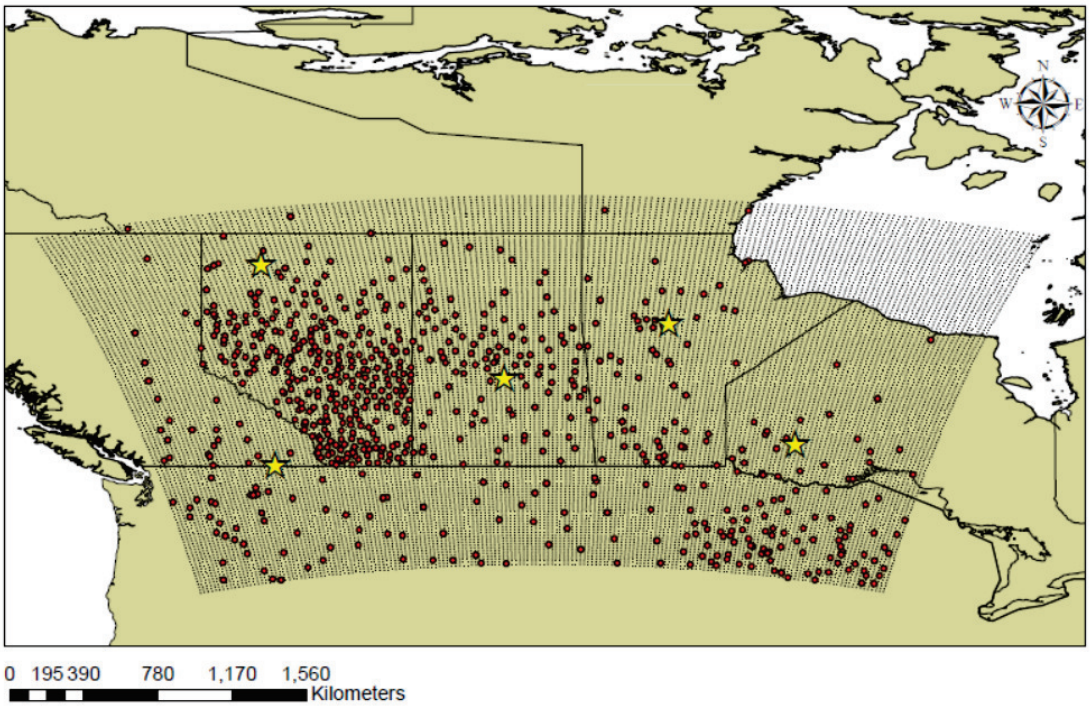

Figure 1.1: The study domain with observation and climate normal locations overlain. The CaPA prairie region grid and GEM grid are depicted by the small black dots, while the red dots indicate observation locations. The yellow stars mark the climate normal locations

data in Figure 1.1 clearly illustrates the desirable spatial characteristics of the GEM data as compared to observation data. The GEM data set consists of 22272 grid cells, at a spacing of 15 km. The GEM model was created by the Meteorological Research Branch and the Canadian Meteorological Centre of Environment Canada (*Côté et al.*, 1998), and was designed to be highly flexible, in order to meet a variety of needs ranging from weather forecasting and research to air quality and climate modelling.

GEM is a physically parametrized model (*Mailhot et al.*, 2006) capable of regional, medium-range, and global scale forecasting. An important feature of note for the GEM model is its ability to handle the large scale forcing needed for regional scale domains; because the model itself is capable of handling multiple scales, it can create a low resolution model environment surrounding a higher resolution regional domain in order to simulate large scale features and waves without the use of sharp boundaries (*Côté et al.*, 1998). Limitations exist, however, on the resolution of GEM outside of the North American region, where the minimum grid spacing is 300 km (*Mailhot et al.*, 2006).

GEM also boasts a multitude of other features designed to improve the forecasting capabilities of the model. Some of these features include (*Côté et al.*, 1998; *Mailhot et al.*, 2006):

- A rotated latitude-longitude mesh to promote regional modelling
- A 3D variational type data assimilation system

- Hydrostatic pressure vertical coordinates with 58 hybrid levels used for vertical resolution up to a maximum atmospheric height of 10 hPa. Focus is placed on the stratosphere, the boundary layer, and the upper-level jet
- Hydrostatic vertical momentum equations
- Orographic impacts such as drag
- Use of the Kain-Fritsch convection scheme, and the Kuo transient scheme for cases of shallow convection
- Distinction between land surface types such as sea ice, snow, open water, and land with vegetative cover, with land surface interaction physics specific to each surface type
- A variety of physics packages to handle cloud interactions, condensation, and vertical diffusion
- A 450 second model time step

Further discussion of the GEM model details listed above, as well as those not mentioned, is not within the scope of this thesis. However, more information can be found by referring to Mailhot et al. (2006). Note that the above description of the GEM model applies to the 15 km resolution version, which was used as the model data source for this investigation. A 10 km resolution version has been released and implemented as a data source for CaPA since the start of this project, but will not

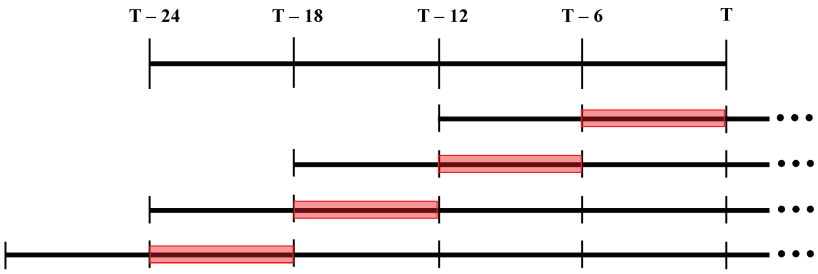

Figure 1.2: Assembly of the 24 hour GEM accumulations in CaPA. The top bar represents the 24 hour accumulation for valid time T, while all of the underlying bars represent 48 hour forecasts with start times indicated by their relative positions. The portion of each GEM forecast used to create the 24 hour accumulation value for CaPA is highlighted by a shaded box

be discussed here. It is not expected that the shift to the 10 km resolution GEM model will have a significant impact on the applicability of the results of this study.

GEM produces 48 hour forecasts, and are run at four time steps daily, at 06Z, 12Z, 18Z, and 00Z. In order to allow for an initial 6 hour model spin-up time, 24 hour accumulation periods are pieced together from four different forecasts. For a given time, the 24 hour accumulation value from GEM is derived as described in Figure 1.2. For 6 hourly data, the accumulation period in CaPA is always taken as the 6 to 12 hour period from a GEM forecast initialized 12 hours prior to the valid time in CaPA.

The observation data used in this study consists of a set of 781 stations. It is important to note, however, that not all of these stations report observations for every time step. The majority of the observation stations used in this study are taken from the Environment Canada data set, which is primarily comprised of records from the METAR, SYNOP, and SHEF networks (*Fortin*, 2012–2013). METAR data, or

Meteorological Aerodrome station data, is received from airports throughout North America, with Canadian airports providing precipitation information in the form of occurrence or non-occurrence data. SYNOP data is high quality precipitation data that is held to the standards of, and reported to, the World Meteorological Organization (WMO). SYNOP data can be derived from either automatic or manual gauging stations. Tipping bucket and weighing gauges are the most common types of automatic gauges used in Canada, while Fischer-Porter and Belfort Transmitting Precipitation gauges with electro-optical encoding are also used (*Goodison et al.*, 1998). At manual stations, Type B rain gauges are employed, which have a height of 36 cm, and 100 cm^2 orifices 40 cm off of the ground (*Goodison et al.*, 1998). Wind shielding is not used for tipping bucket gauges or manual stations (*Mahfouf et al.*, 2007). For the American portions of the study domain SHEF data is also available. SHEF is an American cooperative network based primarily on volunteer observations available for 24 hour accumulation periods. Finally, additional records of lower quality than the SYNOP data are also available from Environment Canada, and are thus used only for temperature and wind speed values as part of CaPA's quality control procedures (*Fortin*, 2012–2013). All of these data are received and input into CaPA in Binary Universal Report Protocol (BURP) format.

Additional observational data obtained from Manitoba Hydro was also incorporated into the input data set used for this investigation. This included observations from Alberta Agriculture and Rural Development, Alberta Sustainable Resource

Development, the Saskatchewan Forestry Association, and Manitoba Hydro gauges. However, only SYNOP data was used to verify the CaPA analyses produced through the methodologies proposed by this thesis. This distinction was made to guarantee that only the highest quality data is used in the verification process, and thus to ensure the accuracy of the verification statistics. Despite this restriction on the verification station selection, it is still possible to evaluate CaPA performances roughly throughout the domain, as evidenced by a 20% verification grid shown in Figure 2.4. Under operational verification conditions, the number of stations, and thus the corresponding coverage, is increased even further. This ensures that the performance of CaPA is verified under the variety of precipitation conditions that may exist at any one time within an analysis domain.

As previously stated, CaPA's data needs are not restricted to precipitation measurements. Temperature data is collected from GEM, while both temperature and wind data are retrieved from observation stations. This supplemental data is used in the quality control procedures in place in CaPA, primarily to handle quality concerns related to solid precipitation measurements.

The time period selected for this study ranges from $2010/05/01/06 : 00Z$ to $2010/08/31/18 : 00Z$. This limits the duration of the study to the months of May to August. This time frame was selected specifically to avoid known issues related to solid precipitation. Solid precipitation is challenging to observe, and results in specific biases manifesting in the CaPA results. Thus, the spring and summer months

were selected to avoid solid precipitation measurements, and to help focus the bias investigation on data transformation issues specifically; if the bias from solid precipitation was allowed to infiltrate the study results, it is possible that the significance of the data transformation bias corrections would be unfairly overshadowed by the remaining solid precipitation biases related to the habitual under-catch of this precipitation type. Further discussion on solid precipitation bias is provided in Section 2.1. The solid precipitation bias issue is currently being researched by other individuals associated with the CaPA project.

Only one year of spring and summer data was considered. It is felt that one year of data, analyzed at a 6 hour time step, is sufficient to capture the impact of the proposed statistical interpolation improvements on CaPA analyses. The chosen study time period results in 368 corrected analyses available for assessment under the proposed bias correction scheme, and 491 corrected analyses for the semivariogram improvement methodology. The selection of the study year, 2010, was based on the most recent year of data available at the start of this project.

In order to understand the implications of selecting 2010 as the focal year of the investigation, a comparison was made between the 2010 spring and summer precipitation records and the 1981 to 2010 climate normals for five stations scattered throughout the domain. The positions of these stations are illustrated in Figure 1.1. The stations were chosen so that they were loosely spread out across the full domain, met the WMO climate normal standards, and had a full data set for the 2010 study

time period. Figure 1.3 illustrates the 2010 data comparison to the climate normals.

A quick assessment of the results depicted in Figure 1.3 reveals that throughout the spring and summer of 2010, precipitation amounts at times exceed the climate normals, at other times approximately match them, and on occasion dip below the normal values. This confirms that a wide variety of precipitation regimes are captured by the 2010 spring and summer months, and that limiting the study time period to 2010 does not create an overt risk of verifying the proposed methods during only particularly wet or dry periods. It is possible that certain trends in precipitation may dominate in the domain during the study period, but deviations from such trends are likely in time and space, based on the results of Figure 1.3. Furthermore, it is evident from the daily precipitation curves that a wide spectrum of precipitation magnitudes is seen at all five stations. This supports the selection of May to August of 2010 as a representative time period for the domain.

Finally, lightning data was also used during this study. This data was obtained from Environment Canada, and taken from the Canadian Lightning Detection Network (CLDN). The CLDN, instituted in 1998, is operated by Vaisala Inc., along with the United States National Lightning Detection Network (*Burrows and Kochtubajda*, 2010). Together, the two networks produce a coverage area of 20 million km^2 (*Cummins and Murphy*, 2009). This data set explicitly distinguishes between cloud to ground and cloud to cloud lightning flashes. Multiple sensor types are used in the network, including Series 4 Lightning Positioning and Tracking Sensors (LPATS-IV),

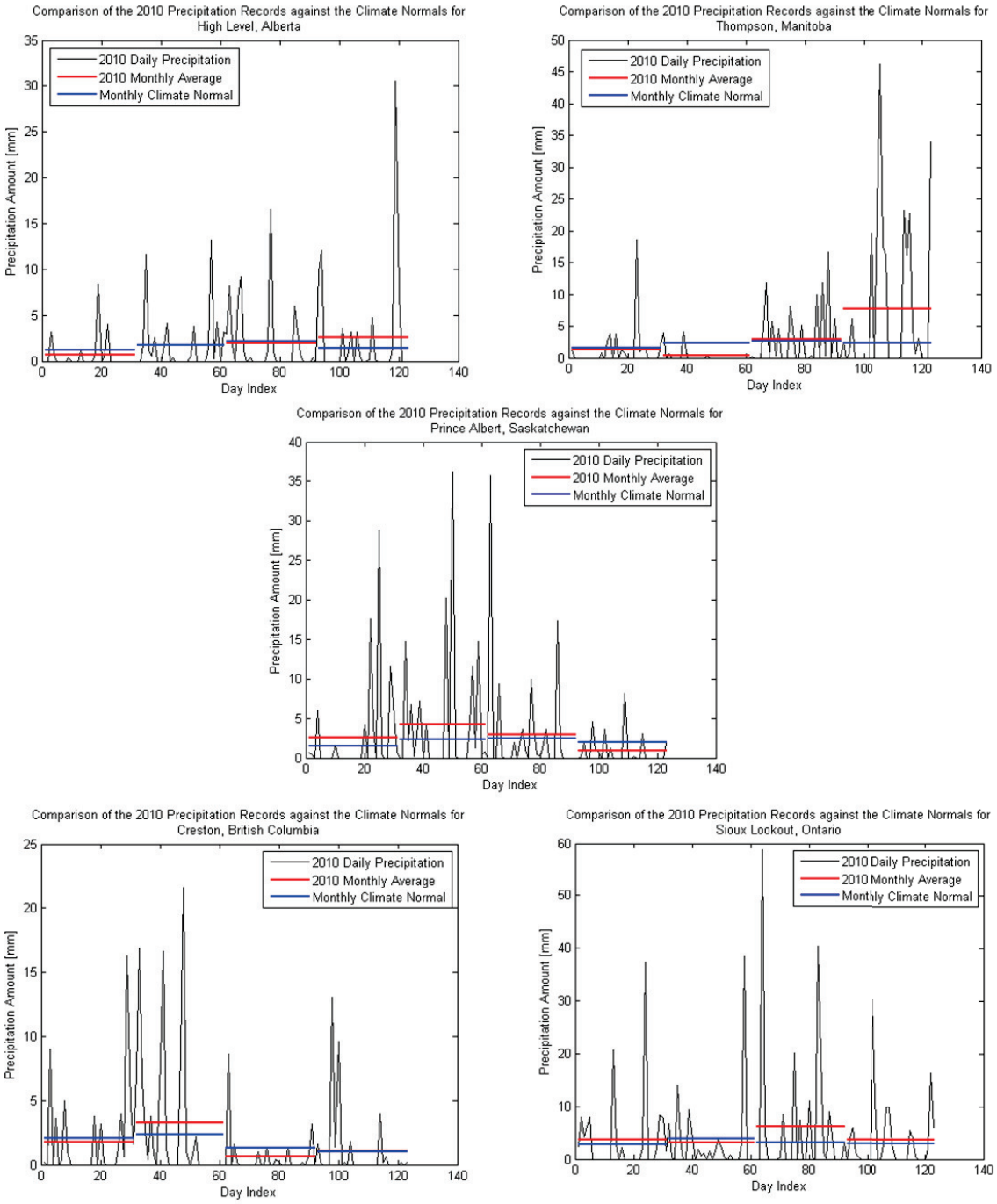

Figure 1.3: Comparison of 2010 precipitation data to climate normals for five stations in the study domain. Note that a day index of 1 refers to May 1st, 2010

Improved Accuracy from Combined Technology ES senors (IMPACT/ES) and Cloud to Ground Enhanced Lightning Sensors (LS7000) (*Burrows and Kochtubajda*, 2010). For the LPATS-IV case, three sensors are needed to detect the location of a lightning flash, while only two are needed for the other sensor types. Currently a total of 83 sensors exist across Canada, creating a network with an internal detection efficiency of 80 to 90%. At the outer boundaries of the Canadian network, this detection efficiency drops to 70%, and to 30% at distances of approximately 300 km outside the network. Detection efficiencies are notably less for cloud to cloud flashes, for example ranging from 1 to 4% when a 5 kiloAmpere threshold is implemented (*Burrows and Kochtubajda*, 2010). The location accuracy for the network is approximately 0.5 km (*Cummins and Murphy*, 2009).

1.1.2 Overview of the CaPA Program

The CaPA program is the culmination of many years of work on behalf of Environment Canada. The inception of CaPA experimentally dates back to approximately 2003, with CaPA becoming operational in 2011 (*Fortin*, 2012–2013). The program continues to undergo development today. CaPA is overseen jointly by the Meteorological Research and the Meteorological Services of Canada divions of Environment Canada.

The CaPA program is in essence a data assimilation system designed to combine observation and model data in as close to a statistically optimal way as possible,

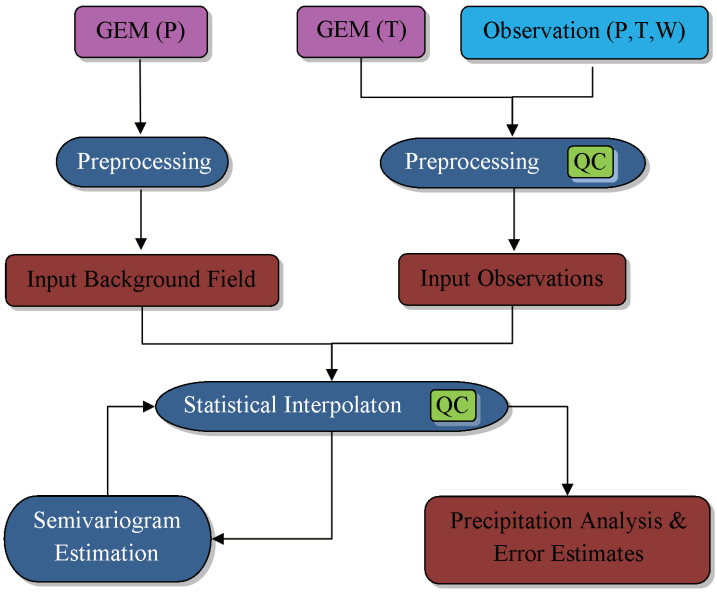

Figure 1.4: The CaPA work flow. The following abbreviations apply: P = Precipitation, T = Temperature, W = Wind, and QC = Quality Control. The ovals indicate CaPA processes, while the rounded rectangles represent data inputs and outputs

producing quantitative precipitation estimates. Running CaPA results in the production of a final analysis on the input model grid. To attain this, CaPA makes use of SI, as described in Section 1.1.3. More work is required by the program than just the execution of this geostatistical method, however. CaPA must orchestrate the input and processing of observation and model data, the execution of quality control processes, the application of the SI methodology, and the output of the final analysis along with a series of accompanying informative files. Figure 1.4 outlines the basic flow of data in CaPA, and the sequencing of some of the program's basic processes.

As Figure 1.4 illustrates, there are a series of steps and data sets involved in CaPA. First, GEM precipitation and temperature data is collected along with observed

precipitation, temperature, and wind. These inputs are fed into a pre-processing utility in CaPA, known as *obscapa*. The pre-processing converts the data inputs into forms compatible with the rest of the CaPA program, and also applies a quality control algorithm to the observation data, using the temperature and wind data. This step specifically addressess quality concerns associated with solid precipitation measurements, rejecting solid precipitation measurements at manual SYNOP gauges if wind speeds are greater than 2 m/s. To help combat wind-related measurement biases, Nipher and Geonor gauges with Alter shields are commonly used in Canada for solid precipitation, while shielded NWS8 gauges are more common in the United States (*Fortin et al.*, 2011). Belfort gauges with Nipher-type shields are also used (*Goodison et al.*, 1998). However, the shielding does not alleviate under-catch issues for wind speeds greater that 2 m/s. Automatic solid SYNOP precipitation records are also rejected.

After the pre-processing is complete, the outputted intermediary data is processed by CaPA's SI subroutines. This is handled by the *mist* component of CaPA. Additional quality control procedures occur during the SI process, including a leave-one-out type check. This check looks for internal consistency between observations by estimating the analysis value at an observation station when excluding that observation's value as input, and then comparing the difference between the resulting analysis value and the excluded observation to a pre-set tolerance (*Fortin*, 2012–2013). A temporal quality control check is also instituted, where differences between

analysis and observation values at the same locations are compared to a threshold, which is based on a multiple of the square root of the variance of the analysis and observation values at that location in time (*Fortin et al.*, 2011; *Fortin*, 2012–2013). The removal of an observation station due to quality control checks can be reversed, if subsequent elimination of stations warrants it (*Fortin*, 2012–2013).

The SI process also includes the estimation of the theoretical semivariogram. A semivariogram estimated for one time step is applied during the following time step. More details on the estimation of semivariograms can be found in Chapter 3. It is pertinent to note that precipitation, and precipitation errors, do not adhere to a normal distribution, but tend to approximate a log-normal distribution instead (*Mahfouf et al.*, 2007). However, the assumption of normality is required during the statistical interpolation methodology followed by CaPA, to ensure that semivariograms accurately represent the spatial structure of the precipitation inputs. Thus a Box-Cox transformation is applied to data inputs. Once the SI algorithm is completed, the inverse transformation is used, and the final precipitation analysis is outputted along with estimates of the analysis errors.

The entire CaPA process can be executed on either 6 or 24 hour accumulation periods. All aspects of this study use 6 hour accumulation periods, except where indicated during the bias correction methodology.

The CaPA program is continually being developed and updated by Environment Canada. For the purposes of this project, CaPA version 2.3b4 was used, and hence-

forth will be referred to as the current operational configuration. However, it is should be noted that newer versions exist experimentally and operationally, and version 2.4 is in the process of being released. Differences between the latest versions of CaPA and version 2.3b4 do not include any modifications to the statistical interpolation algorithm. Thus it is assumed that all findings related to version 2.3b4 are applicable to the latest CaPA versions as well.

The new features that have been incorporated into the latest CaPA versions include the following: the inclusion of radar observations experimentally; an increase in resolution from the 15 kilometer grid in version 2.3b4 to a 10 km grid operationally; the incorporation of additional station data from Canadian networks, as well as Mexico and China experimentally; and the increase of the possible CaPA domain to the global scale in the latest experimental version. Version 2.3b4 is limited to North America. The global version of CaPA is also able to assess analysis uncertainty through the use of a 24 member ensemble set (*Fortin*, 2012–2013).

Information regarding the CaPA program along with near real-time analysis data is publicly available at the following locations:

- http://loki.qc.ec.gc.ca/DAI/capa-e.html
- http://weather.gc.ca/analysis/

1.1.3 The Statistical Interpolation Method

SI is the underpinning concept by which CaPA operates. SI, as it is commonly applied to atmospheric scenarios, is a well developed and understood statistical tool. The fundamental principles of SI are not unique amongst statistical concepts; the roots of the method show significant similarities to kriging, a frequently used geostatistical methodology in geological endeavors (*Daley*, 1991). Much like kriging, the basic function of the SI method is the combining of a background data set with a series of point observations to create a final gridded output data set, or analysis. The combining of data sets using SI is based on the principle of minimizing the error variance of the final analysis. However, unlike traditional kriging, SI specifically operates on the basis of innovations, or the differences between the observation and background values at a given point.

In the SI paradigm, the background data set can be denoted as B_i, where i is the index value of the grid cell that the background value represents. Conversely, B_k refers to the background data interpolated to observation location k. Using the same indexing, the observation data is stated as O_k. Finally, A_i represents the analysis value produced as a result of the statistical interpolation technique at location i. Using this notation scheme, the fundamental equation of statistical interpolation can be stated (*Daley*, 1991; *Mahfouf et al.*, 2007):

$$A_i = B_i + \sum_{k=1}^{K} W_{ik} \left[O_k - B_k \right] \tag{1.1}$$

where W_{ik} denotes the set of weights that must be applied to the innovation values, $[O_k - B_k]$. As can be seen from the equation, the product of the innovation values and their appropriate weights is what is added to the background data value B_i, in order to produce the final analysis A_i.

Since in reality both the observed and background data contain error, it is necessary to denote the error of these terms as $O_k - T_k$ and $B_i - T_i$, where T_k and T_i are the true observed and background values at location k and i respectively. It should be noted that the observation error also contains errors of representativeness. Errors of representativeness result when weather features with smaller footprints than the average spacing of the observation stations are present, as they will either be missed entirely by the gauge network, or assumed to be of a larger scale than they truly are if captured (*Daley*, 1991).

Given the expressions for the error of the observation and background data, it is now necessary to make the following assumption regarding their expected value (*Daley*, 1991):

$$\langle B_i - T_i \rangle \;=\; \langle B_k - T_k \rangle \;=\; \langle O_k - T_k \rangle \;=\; 0 \tag{1.2}$$

where the <> brackets indicate an expected value. This assumption implies that both the background field and the observations are unbiased, and leads us to a confirmation of the unbiased nature of the analysis value A_i, since (*Daley*, 1991):

$$\langle O_k - B_k \rangle \;=\; \langle (O_k - T_k) - (B_k - T_k) \rangle = 0 \tag{1.3}$$

However, we know that the assumption that $\langle B_i \rangle = \langle T_i \rangle$ and $\langle O_k \rangle = \langle T_k \rangle$ is not true; the background field and observation data possess bias due to practical modelling and measurement limitations that prevent true precipitation amounts from being predicted or observed. Thus some bias does exist due to the observation and background data. Biased input data can be pre-treated or removed prior to use in the SI methodology to help alleviate this concern (*Daley*, 1991). In CaPA, this is accomplished in part through the quality control procedures contained in the obsCaPA subroutines, which prevent biased data, particularly untrusted solid precipitation observations, from being input to the SI component of the program. Moving forward in the development of the SI algorithm next requires that the true value at location i be subtracted from both sides of Equation 1.1, and the result squared (*Daley*, 1991):

$$A_i - T_i = B_i - T_i + \sum_{k=1}^{K} W_{ik} \left[O_k - B_k \right] \tag{1.4}$$

$$\left\langle (A_i - T_i)^2 \right\rangle = \left\langle (B_i - T_i)^2 \right\rangle + 2 \sum_{k=1}^{K} W_{ik} \left\langle (O_k - B_k) (B_i - T_i) \right\rangle + \sum_{k=1}^{K} \sum_{l=1}^{K} W_{ik} W_{il} \left\langle (O_k - B_k) (O_l - B_l) \right\rangle \tag{1.5}$$

The result in Equation 1.5 can be clarified by considering the meaning behind its component terms. Firstly, the term $\langle (O_k - B_k)(O_l - B_l) \rangle$ can be thought of as the covariance between innovation values at locations k and l. The full covariance term can be stated as $\langle (O_k - B_k)(O_l - B_l) \rangle - \langle O_k - B_k \rangle \times \langle O_l - B_l \rangle$, but given that Equation 1.3 defines both $\langle O_k - B_k \rangle$ and $\langle O_l - B_l \rangle$ as zero, they do not contribute to the covariance. Following the same logic reveals that the $\langle (O_k - B_k)(B_i - T_i) \rangle$ term also represents a covariance value, this time between the innovation value at location k and the background error value at the estimation point i (*Daley*, 1991). Finally, the $\left\langle (A_i - T_i)^2 \right\rangle$ and $\left\langle (B_i - T_i)^2 \right\rangle$ terms can be thought of as the expectations of the analysis and background error variances at location i (*Daley*, 1991).

Since the primary goal of the entire SI procedure is to produce an estimate at a location i in a way that minimizes the variance of that estimate, or $\left\langle (A_i - T_i)^2 \right\rangle$, the derivative of Equation 1.5 must be taken. This produces Equation 1.6, when the equation is differentiated with respect to the weights. Note that some rearrangement of terms is required to arrive at Equation 1.6, and the assumption that the observation and background error terms are not correlated is made (*Daley*, 1991):

$$\sum_{l=1}^{K} W_{il} \left[\langle (B_k - T_k)(B_l - T_l) \rangle + \langle (O_k - T_k)(O_l - T_l) \rangle \right] = 2 \langle (B_k - T_k)(B_i - T_i) \rangle \tag{1.6}$$

With Equations 1.1 and 1.6 now in place, it is possible to frame the SI formulation in terms of matrices. Let f_{A_i} and f_{B_i} be column vectors describing the analysis and

background values at the gridded analysis locations, while f_O and f_B are column vectors describing the observed and background values at the observation locations. The column vector of weights for grid location i, with a length equal to that of f_O and f_B, is denoted by W_i. Putting these newly defined vectors together allow for Equation 1.1 to be reproduced as follows (*Daley*, 1991):

$$f_{A_i} = f_{B_i} + W_i^T \left(f_O - f_B \right) \tag{1.7}$$

In order to restate Equation 1.6, a few more vector definitions are needed. First, let C_B and C_O represent the covariance matrices for the background and observation errors respectively. Next, C_{B0} is defined as the covariance vector describing the background error between all observation locations and the current analysis location i. Combining these terms yields the following (*Daley*, 1991; *Mahfouf et al.*, 2007):

$$W_i \left(C_B + C_O \right) = C_{B0} \tag{1.8}$$

It is important to remember that Equation 1.8 is used to define the set of weights that work to minimize the error variance of the analysis. However, since C_B, C_O, and C_{B0} require knowledge of the true background and observation values, which can never be known precisely, they are only estimates. Thus, the weights do not fully minimize the analysis error variance but only approximate the minimized result, which is what defines the above procedure as SI, rather than optimal interpolation (*Daley*, 1991).

While the above development of the SI methodology represents the theoretical approach, it is important to understand how SI is implemented in CaPA. A critical components of CaPA's execution of the SI technique is the use of a semivariogram, which is developed to describe the spatial relationships between innovation values for given separation distances. The semivariogram is used to replace the covariance matrices required in Equation 1.8.

To begin with, the term $C_B + C_O$ can be replaced with only one covariance matrix, C_I, which describes the covariance between innovation values, rather the the covariance between background and observation errors separately. Given this substitution, the following relationship can then be derived for a covariance value between locations i and j, denoted $C_{I_{i,j}}$, and the semivariogram value for the same locations, $\gamma_{I_{i,j}}$(*Isaaks and Srivastava*, 1989):

$$\gamma_{I_{i,j}} = \frac{1}{2}\left\langle (V_i - V_j)^2 \right\rangle \tag{1.9}$$

$$\gamma_{I_{i,j}} = \left\langle V^2 \right\rangle - \langle V_i \times V_j \rangle \tag{1.10}$$

$$\gamma_{I_{i,j}} = \left\langle V^2 \right\rangle - m^2 - \left(\langle V_i \times V_j \rangle - m^2 \right) \tag{1.11}$$

$$\gamma_{I_{i,j}} = \sigma_B^2 - C_{I_{i,j}} \tag{1.12}$$

where V_i and V_j are two innovation values at locations i and j, σ_B^2 is the error variance of the background, previously defined as $< (B_i - T_i)^2 >$, and m^2 is the assumed mean value of the innovations. Equations 1.9 to 1.12 are only valid if it is assumed that the innovation values, which are conceptualized as random variables, all have the same variance and mean (*Isaaks and Srivastava*, 1989).

In order to incorporate semivariograms in CaPA, a semivariogram must be estimated for each time step. CaPA begins this process by first determining the experimental semivariogram, and then using this as the basis for fitting an isotropic exponential theoretical semivariogram. The isotropic exponential model is somewhat beneficial for CaPA; because of its simplicity compared to other potential semivariogram models, semivariogram parameters can be predicted more confidently under conditions with data availability restrictions (*Erdin and Frei*, 2011). However, the use of an isotropic semivariogram is also concerning, as it is not uncommon for spatial covariance structures of precipitation to change with differing precipitation magnitudes (*Erdin and Frei*, 2011). Other geographic and atmospheric influences may also impart directionality into the semivariogram.

In CaPA, a semivariogram estimated during time step i is put to use during time step $i + 1$. The delay in the use of the semivariogram is deemed reasonable, as it is anticipated that the domain-wide semivariogram does not fluctuate wildly between adjacent time steps. Rather, seasonality is assumed to exist in the semivariogram, in the sense that fluctuations in the semivariogram are smoothed but still exist

throughout the course of a year, yet day-to-day changes are minimal. In order to accommodate this, smoothing of the semivariogram is practiced.

The semivariogram in CaPA is smoothed at the experimental stage. First, a weighting for the current time step's semivariogram is determined according to the following equation (*MathWorks*, 2013):

$$W_{ES} = \exp\left(\frac{-\delta t}{24 * 30}\right) \tag{1.13}$$

where W_{ES} is the weight to be applied in smoothing the current experimental semivariogram value, and δt is the time, in hours, since the last semivariogram was estimated. The denominator possesses a value of $24 * 30$ to represent 24 hours per day for 30 days, which is used to determine the dominant time period extending back in time that will influence the current smoothed semivariogram. A semivariogram calculated more than 30 days prior to the current time step may still be used, but will receive comparatively less weighting.

Using the value of W_{ES}, the experimental semivariogram is then smoothed. The experimental semivariogram exists as a series of semivariogram values estimated for precipitation ranges, or bins. The smoothing is done on a per bin basis as follows (*MathWorks*, 2013):

$$\gamma_{Smth_i,t} = \frac{NUM_{Exp_{i,t}} \times \gamma_{Exp_{i,t}} + W_{ES}\left(NUM_{Smth_{i,t-1}} \times \gamma_{Smth_{i,t-1}}\right)}{NUM_{Exp_{i,t}} + W_{ES}\left(NUM_{Smth_{i,t-1}}\right)} \tag{1.14}$$

where $\gamma_{Smth_i,t}$ is the smoothed experimental semivariogram value for time step t, $\gamma_{Exp_{i,t}}$ is the original semivariogram value for the same time step, and $\gamma_{Smth_{i,t-1}}$ is the smoothed semivariogram value for the previously time step. The bin number is denoted by i. Finally, $NUM_{Exp_{i,t}}$ and $NUM_{Smth_{i,t-1}}$ refer to the number of innovations used to calculate the semivariogram value for bin i for the current experimental semivariogram and the previous time step's smoothed semivariogram respectively. It is the values contained in γ_{Smth} that are used by CaPA during SI.

Once the smoothing is completed for the current time step, the $NUM_{Smth_{i,t-1}}$ value must be updated to reflect the inclusion of the current time step's innovation values in the estimation of the smoothed semivariogram. This is done by simply summing $NUM_{Exp_{i,t}}$ and $NUM_{Smth_{i,t-1}}$, to produce $NUM_{Smth_{i,t}}$.

Additionally, in CaPA a data transformation process is present. All of the input innovation data used in the SI scheme is treated with the Box-Cox transformation. As such, at the end of the SI procedure, the analysis values must all be treated with the inverse Box-Cox transformation, and a bias correction to address the use of the data transformation must be applied. This is discussed in greater detail in Chapter 2. Additionally, a brief review of some example alternatives to the SI method is provided in Appendix A.

1.1.4 Verification Statistics for CaPA Analyses

A standardized set of verification metrics has been established by the Comité des Passes Opérationnelles et Parallès of the Canadian Meteorological Centre of Environment Canada, to assist in evaluating the positive and negative aspects of modifications to the CaPA program. To maintain consistency with this practice, these metrics constitute the basic verification criteria used in this study. Two types of evaluators are used: continuous skill scores, and categorical skill scores. All scores are calculated on a per time step basis for the entire domain, and then averaged across the full time period of the study.

Continuous skill scores are generally built upon the idea of either using all of the available data in the domain, or else employing thresholds. When threshold are used, scores are calculated only for sub-sets of the domain where the precipitation amount is below a given threshold. This threshold is then continuously increased, until all values are included. The first score to be calculated in this way is the bias, although no thresholds are used in this case (*Environment Canada*, a).

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(O_i - A_i) \tag{1.15}$$

where O_i and A_i are the observed and analysis values at location i, and N is the total number of locations in the domain for the current time step. All analysis values are outputted from CaPA in a gridded form, which does not directly correlate to the observation locations. In order to bring the analysis values to the observation

locations, bi-linear interpolation is used.

The Departure from the Partial Mean (DPM) and Departure from the Partial Standard Deviation (DPS) scores incorporate the threshold concept. The threshold values used, in milimeters, are: 0.2, 1.0, 2.0, 5.0, 10.0, 25.0, and 50.0. For each threshold value, the DPM calculation adheres to the following equations (*Environment Canada*, a):

$$\bar{O}_q = \frac{1}{\xi\left(O_i < q\right)} \sum_{i=1}^{N} O_i \cdot I_{(O_i<q)}\left(O_i\right) \tag{1.16}$$

$$\bar{A}_q = \frac{1}{\xi\left(A_i < q\right)} \sum_{i=1}^{N} A_i \cdot I_{(A_i<q)}\left(A_i\right) \tag{1.17}$$

$$\mathrm{DPM}_q = \bar{A}_q - \bar{O}_q \tag{1.18}$$

where ξ denotes a function that returns the number of instances in the data set where the input conditions are met, q represents the current threshold value, and I signifies an indicator function which returns one if the input is less than threshold q and zero otherwise (*Environment Canada*, a). A similar set of equations are followed to calculate the DPS score (*Environment Canada*, a):

$$S_q^{(O)} = \sqrt{\frac{1}{\xi\left(O_i < q\right) - 1} \sum_{i=1}^{N} \left(O_i - \bar{O}_q\right)^2 \cdot I_{(O_i<q)}\left(O_i\right)} \tag{1.19}$$

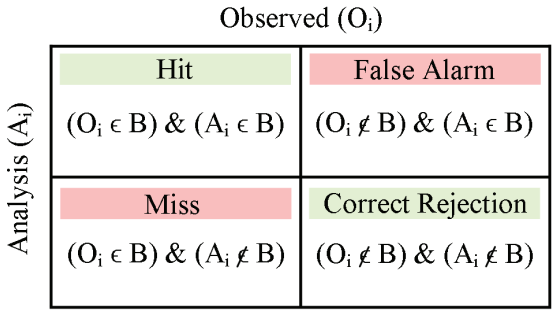

Figure 1.5: Contingency table for the calculation of categorical skill scores. The equation in each element of the table describes the requirements for classifying a pair of observed and analysis values as belonging to that category

$$S_q^{(A)} = \sqrt{\frac{1}{\xi\left(A_i < q\right) - 1} \sum_{i=1}^{N} \left(A_i - \bar{A}_q\right)^2 \cdot I_{(A_i<q)}\left(A_i\right)} \tag{1.20}$$

$$\mathrm{DPS}_q = S_q^{(A)} - S_q^{(O)} \tag{1.21}$$

In order to look only at specific ranges of values, or bins, categorical skill scores are used. However, the determination of such categorical scores first requires the construction of contingency tables for the domain. An example of such a table is provided in Figure 1.5, where the current bin is denoted by B, and the observation and analysis pair being considered are O_i and A_i. The symbol $\in$ indicates that the value belongs to the bin, and does not when the symbol is crossed out.

The values of a contingency table are determined by considering all pairs of observation and CaPA analysis values across the domain. CaPA analysis values are brought to the observation locations through bi-linear interpolation. Using these

pairs, four possible combinations are possible: a hit, where both the analysis and observations predict precipitation in the current bin; a false alarm, where the analysis predicts precipitation in the current bin but the observations do not; a miss, where the observations indicate precipitation has occurred in the current bin but it is not forecast by the analysis; and a correct rejection, where both the observations and the analysis agree that no precipitation occurred in the current bin. The number of occurrences of each combination are then calculated, and normalized by the total number of pairs in the domain. A contingency table must be calculated for every precipitation bin, and for every time step.

Once created, the contingency table can be used to calculate the Frequency Bias Index (FBI) categorical skill score as follows (*Environment Canada*, a):

$$FBI = \frac{h + f}{h + m} \tag{1.22}$$

where h, f, and m, are respectively the normalized hit, false alarm, and miss values from the contingency table. The FBI score considers the ratio of the total number of locations reporting precipitation amounts for the bin of interest from the analysis as compared to from the observations.

The second categorical skill score used by Environment Canada in assessing precipitation analysis performance is the Equitable Threat Score (ETS). The ETS score can be calculated according to the following equation (*Environment Canada*, a):

$$ETS = \frac{h - h_r}{h + f + m - h_r} \tag{1.23}$$

where h_r is further defined as follows (*Environment Canada*, a):

$$h_r = (h + f)(h + m) \tag{1.24}$$

and mathematically represents the number of hits that would be expected purely by chance (*Environment Canada*, a). This equation can be rewritten as follows (*Mesinger*, 2008):

$$ETS = \frac{h - FO}{F + O - h - FO} \tag{1.25}$$

where F and O are further defined as (*Environment Canada*, a):

$$F = h + f \tag{1.26}$$

$$O = h + m \tag{1.27}$$

The ETS score is not easily conceptualized, but can be loosely understood as a means of quantifying the analysis skill for specific precipitation bins, in terms of the placement of precipitation within the analysis domain (*Environment Canada*, a). The use of the ETS score when investigating precipitation products is commonplace,

with such organizations as the National Centers for Environmental Prediction relying heavily on its use (*Mesinger*, 2008). However, with a prevalence in the precipitation analysis community, the ETS has also been subjected to a notable amount of scrutiny. This attention has unearthed a basic deficiency associated with the skill metric; the ETS value is often found to be larger if the data possesses a greater amount of wet bias, as compared to a similar precipitation analysis with relatively less bias (*Mesinger*, 2008). This is particularly the case if the amount of bias being dealt with is greater than unity (*Mesinger*, 2008).

In order to combat this inadequacy of the ETS score, a method for adjusting the ETS value by modifying the number of hits has been proposed. This method requires the definition of a function $H(F)$, which allows for a h value corresponding to $F = O$ to be determined. This h value is labeled as h_a, and represents the h value associated with a no-bias case for the analysis. When h_a is plugged back into Equation 1.25, and F is replaced in the equation with O, the result is an ETS value effectively adjusted to remove the impact of bias (*Mesinger*, 2008). This value is known as the adjusted ETS, or aETS.

The equation used to determine h_a is derived based on a series of function requirements, which includes the following (*Mesinger*, 2008):

1. $h = 0$ when $F = 0$

2. for all known h and F pairs, $h(F)$ must equal the known h value for the given F values

3. $h(F) < O$ for all F, although as F increase, $h(F)$ must approach the value of O

4. $h(F) < F$ for all F

Based upon these requirements, the following equation for h_a is achieved (*Mesinger*, 2008):

$$h_a = O - \frac{F-h}{\ln\left(\frac{O}{O-h}\right)} \text{lambertw}\left(\frac{O}{F-h}\ln\left(\frac{O}{O-h}\right)\right) \tag{1.28}$$

where the lambertw function is defined as the inverse of the following (*Mesinger*, 2008):

$$y = x\exp(x) \tag{1.29}$$

However, it is important to note that this approach also makes the assumption that the observation and forecast areas with precipitation events for a given precipitation range can be idealized as circular areas (*Mesinger*, 2008).

The aETS values are not traditionally calculated when assessing CaPA analyses. However, all of the previous metrics are, and as such have been included in a verification code package available form Environment Canada, which automates their calculation. This coding package also makes it possible to estimate the amount of uncertainty associated with each of the scores through the use of a bootstrapping technique.

The bootstrapping method practiced in CaPA is based on the idea of drawing bootstrap samples of observation and analysis pairs from space and time. A date and location are first randomly selected from the study period and domain, then search radii in the space and time dimensions are generated randomly from independent exponential distributions. All pairs up to a threshold number are accepted into the bootstrap sample (*Fortin*, 2012–2013). This is done repeatedly until the number of pairs accepted into the bootstrap sample is equal to the original number of pairs in the domain. A total of 100 such samples are created, and percentile values of the metric results are used to determine the 10% and 90% confidence bounds, as well as the expected mean value (*Fortin*, 2012–2013). This combines with the original calculations to give a final picture of the skill score values and uncertainty.

1.2 Research Objectives

The focus of this thesis is placed on the improvement of the SI methodology used by CaPA. However, the SI framework includes many components, and not all are to be dealt with under the purview of this study. The specific objectives to be addressed by this investigation can be broken down into two main areas of interest as follows:

- *Correction of bias introduced through data transformation*: This will include the implementation of a moving-window averaging technique to create a new input data set for CaPA, and the correction of a standard CaPA analysis using the moving-window averaging based analysis

- *Improvement of the semivariogram*: Estimation and use of anisotropic semivariogram parameters will be investigated, along with the implementation of a semivariogram parameter filtering scheme to capture the influences of convective precipitation

1.3 Thesis Structure

The remaining portion of this thesis is divided into three chapters. Chapter 2 focuses explicitly on the bias correction objective. The chapter deals with the bias correction issue in its entirety, easing into the problem with a review of the relevant theory and concepts, and moving into the specific methodology followed. The results of implementing the methodology are presented, and a discussion and interpretation of these results follow.

Chapter 3 revolves around the issue of semivariogram improvement. A similar format to Chapter 2 is followed, with the necessary background information on the subject matter being presented first. This is followed by the methodology, results, and discussion portions of the chapter.

Finally, a summary of the work presented in this thesis is contained within Chapter 4. In addition to the study conclusions, this chapter also houses the recommendations that surfaced throughout the duration of this thesis.

Chapter 2

Evaluation and Correction of Bias Introduced through Data Transformation

The hydrology community uses precipitation data for a variety of modelling applications, but not without cognizance of the limitations that accompany it. In the case of CaPA precipitation analyses, these limitations include bias. The reduction of bias beyond what is currently achieved thus becomes an important goal, as it would serve to not only improve the quality of the analyses, but also to bolster the confidence of those who use them. However, bias correction is not a trivial problem, which is evident by its persistence in precipitation data. As such, only bias introduced through the use of data transformations will be considered henceforth for correction; other

forms of bias are outside the scope of this work and are touched upon only briefly in Section 2.1, although their importance should not be overlooked.

2.1 Background

The exercise of estimating or simulating precipitation values, whether it be through stochastic or deterministic means, invariably results in errors. It is the expected value of the error distribution that is commonly referred to as bias (*Isaaks and Srivastava*, 1989). CaPA is not immune to bias, as it not only inherits biases from its input data, but also generates new biases during the data assimilation procedure. During the discussion of bias presented in this chapter, the mathematical definition provided in Equation 1.15 will be used.

2.1.1 General Bias Sources

Biases stemming from the input data fed into CaPA can be partitioned into those that originate from the observation data, and those which are the progeny of the GEM background field. The biases present within the GEM precipitation data, and in climate model outputs in general, can in large part be attributed to erroneous assumptions made during model specification, including assumptions and simplifications required in numerical approximations and the handling of atmospheric physics. Additionally, climate models generally operate on a grid basis, requiring the discretization of climate processes as well as areal averaging, all of which may induce

biases (*Teutschbein and Seibert*, 2012). The summation of all bias sources in Global Circulation Models (GCMs) in general leads to the production of biases in the form of distribution errors, commonly translating into an over abundance of low intensity precipitation events (*Ines and Hansen*, 2006).

Observed precipitation biases are generally a direct result of the measurement tools and methods used to collect the precipitation information. Precipitation can be measured using a variety of instruments, such as non-recording collection vessels, tipping-bucket gauges, and weighing gauges (*Dingman*, 2008). Unfortunately, under-catch is a general concern. Many aspects of a gauge may contribute to the problem of under-catch, such as: gauge orifice size and orientation, gauge height, placement of the gauge in regards to nearby obstructions, and the presence or absence of wind shielding (*Dingman*, 2008). Generally, orifice openings should not be less than 30 mm, while the orientation of the gauge should be strictly horizontal unless persistent high winds require the leveling of the gauge to be parallel to the land surface (*Dingman*, 2008). General instrumentation errors, related to the mechanical properties and in-field operation of gauges, can also contribute to biases, particularly for recording gauge types. These types of biases may be systematic in nature or related to the precipitation event. For example, during extreme precipitation events, tipping-bucket gauges have been found to be increasingly prone to under-catch, while weighing gauges have shown reduced accuracy as well (*Dingman*, 2008).

Additional sources of bias include evaporative losses from gauges, water lost through splashing, and wetting losses, or adherence of a thin layer of water to the internal surfaces of the gauges (*Dingman*, 2008). Outside of the prairie region in higher elevation locales, occult precipitation, i.e. when precipitation is intercepted from clouds directly by vegetation in liquid or solid form, can also lead to under-catch problems (*Dingman*, 2008). The height of observation gauges is also an important consideration, as the intrusion of a gauge into the overlying atmosphere can cause localized turbulence which further hampers the catch capabilities of the gauge. This problem may be compounded by the presence of wind; for a gauge with an 8 inch height that does not incorporate wind shielding, catch errors of approximately 10% can be anticipated for liquid precipitation, with those errors growing to an expected value of approximately 50% for solid precipitation. Fortunately, correction factors can be applied to help account for biases related to various precipitation phase and shielding scenarios (*Dingman*, 2008).

Small precipitation amounts are also difficult to accurately quantify, resulting in the definition of trace precipitation thresholds to distinguish accumulations less than the resolution of the gauge. A standard trace threshold used in precipitation data processing is 0.2 mm (*World Meteorological Organization*, 2008), where any precipitation amount falling below this threshold is commonly treated as zero.

Finally, any manual component of the observation processes may also introduce bias, as the element for human error then exists. This can include such aspects

as the compiling and publishing of observation data, as well as the timing of readings for manual gauges which may lead to data gaps or misleading time stamps for precipitation events (*Dingman*, 2008).

In terms of bias, snowfall poses an even greater problem than liquid precipitation. The solid nature of snow introduces unwanted interactions with the gauge itself, such as accumulation of snow on the gauge rims (*Goodison et al.*, 1998), or the blowing off of snow originally adhered close to the gauge's orifice, all of which can lead to under-catch errors (*Dingman*, 2008). Many gauges in Canada designed for capturing solid precipitation are fitted with shields, which can cause additional problems such as increasing the instability of weighing gauges and enhancing the gauge catch of blowing snow (*Goodison et al.*, 1998). A delay in the recording of precipitation can also occur when dealing with snow, as accumulations on the gauge may actually be recorded during later time steps (*Goodison et al.*, 1998). Biases specifically related to snow events have already been recognized as being troublesome for CaPA, and independent research into the treatment of snow-related biases has been undertaken jointly between the University of Manitoba and Environment Canada.

While this summary encapsulates the main sources of bias that are expected to influence CaPA from an input perspective, other forms of bias do exist in the general hydrological context. Such biases can include, for example, those associated with temperature and wind data measurement, and bias introduced through the use of input data with spatial resolutions incongruous with the model (*Berg et al.*, 2012).

The identification of bias in any type of data used as input to a model is fundamental to understanding the limitations of that data, and how to properly interpret the significance of the model output. Understanding bias sources requires vigilant attention to the types of input data used and the means by which they were acquired, as well how the origin and characteristics of that data blend with the nature and setup of the model that intends to use it. To assist in the determination of the type of bias that is being dealt with, quantile-quantile plots can often be quite revealing: if a vertical shift in the plotted curve from the 45° line is observed, then a bias in the mean of the data should be suspected; if the slope of the curve varies throughout the plot, then bias in the variance of the data likely exists; and if curvature is present it may signify bias in the skewness (*Berg et al.*, 2012).

2.1.2 Data Transformation Bias

The primary concern for this investigation is the issue of data transformation bias, which unfortunately is not readily avoidable. The SI method employed by CaPA, like most statistical models, requires the input data to follow a normal distribution, which is a particularly tricky demand; precipitation amounts are well known to have highly asymmetrical distributions, with lower magnitude precipitation events occurring with relatively greater frequency than larger precipitation events. This non-Gaussian distribution also carries forward to the precipitation innovation values, which are the particular focus of the SI method.

The practical basis of this problem is exemplified by considering the calculation of statistics such as the semivariogram, which is based on the squared difference between innovation values at two different location (*Isaaks and Srivastava*, 1989), and is thus prone to over-emphasizing large difference values. Therefore, it would have negative consequences in terms of accuracy to use the raw input data. Data transformation is the tool applied to address this problem, by transforming the innovations such that they conform to the requisite normal distribution.

In CaPA, the data transformation method of choice is the BoxCox transformation (*Box and Cox*, 1964). This transformation can be expressed as follows:

$$y^{(\lambda)} = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} & (\lambda_1 \neq 0) \\ \log(y+\lambda_2) & (\lambda_1 = 0) \end{cases} \tag{2.1}$$

where the transformed data is denoted by $y^{(\lambda)}$ and the original data is given by y. The Box-Cox parameters are represented by λ_1 and λ_2. Additionally, if $\lambda_1 = 0$ and $y+\lambda_2 \leq 0$, or if $\lambda_1 \neq 1$ and $y+\lambda_2 < 0$, then $y^{(\lambda)}$ is set to zero (*Environment Canada*, b). Under the current operational configuration of CaPA, the λ_1 parameter is set equal to 0.3333, while λ_2 is set to a value of zero.

The selection of the λ_1-parameter value is not an inconsequential detail to the data transformation process. A λ_1 value close to 0.5 indicates approximately a square-root type of transformation, while a lower λ_1 value approaching 0.1 brings the transformation closer to a logarithm-type (*Erdin and Frei*, 2011). In the context

of geostatistical methods, once λ_1 starts to approximate a logarithm-type transformation, a skewness or extension of the upper-tail may be introduced in the probability distribution of precipitation estimates. This can translate into a positive bias in estimated precipitation quantities (*Erdin and Frei*, 2011). A λ_1 value of 0.3333, however, was found to be optimal for CaPA, and likely does not approximate a logarithm-type transformation closely enough for this potential bias to be of concern.

With the innovation data transformed according to Equation 2.1, the CaPA input becomes approximately normally distributed, and the resulting CaPA analysis can be expected to attain greater accuracy. However, once the analysis of transformed data has been completed, an inverse transformation must be applied, in order to regain the utility of CaPA analyses. For example, given the λ_1 and λ_2 values used currently by CaPA, the inverse function applied to the analysis data becomes (*Environment Canada*, b):

$$y = \begin{cases} \left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}} - \lambda_2 & (\lambda_1 \neq 0) \\ \exp\left(y^{(\lambda)}\right) - \lambda_2 & (\lambda_1 = 0) \end{cases} \tag{2.2}$$

where the $\lambda_1 \neq 0$ case applies specifically if $y^{(\lambda)}\lambda_1 + 1 \geq 0$. If $\lambda_1 \neq 0$ and $y^{(\lambda)}\lambda_1 + 1 < 0$, then $y = 0$ (*Environment Canada*, b).

It is the application of Equation 2.2 that creates trouble when considering the bias associated with the final CaPA analysis. Bias is introduced because an SI estimate produced in the transformed data space is essentially the mean value associated

with with a normal probability density function for the estimate, with a variance equal to the error variance that is also calculated by the SI model (*Erdin and Frei*, 2011). This mean, or expected value, however, is not necessarily consistent with the observed mean value if it is converted back to the original data space. This is due to the nature of the back-transformation, which ensures that the median quantile of the data's distribution in the transformed space becomes the median value of the distribution in the original data space (*Miller*, 1984). Since SI operates on the transformed, and thus approximately normally distributed data, the median in this case is approximately equal to the expected value of the SI analysis. However, it is the expected value of the estimate in the original data space, and not the median value, that is of interest (*Miller*, 1984), and the asymmetrical nature of the untransformed data means that the median and mean values are not equivalent. Thus, since the inverse transformed analysis estimates are not the true expected values, a bias is introduced.

In order for the CaPA analysis values to be of use, the final analysis values must be stated in the original data space, thus making the bias introduced during the inverse transformation essentially unavoidable. To ignore this bias is not a satisfactory solution. Instead, a bias correction factor (BCF), which estimates the amount of bias incurred throughout the data transformation process, is applied to the back-transformed data. In the operational version of CaPA, the BCF takes the following form (*Environment Canada*, b; *Fortin*, 2012–2013):

$$\text{BCF} = \left(\frac{1}{3}\right) \times \frac{v\,(v_p - v)}{v_p} \times \text{bc}\,(a)^{\frac{1}{3}} \tag{2.3}$$

where v represents the kriging variance of the analysis estimate, $\text{bc}\,()$ is used to denote the inverse Box-Cox function, and a is the calculated analysis value in the transformed space.

To understand how this equation is arrived at, the basics of the SI procedure must be considered. The two main data inputs into the SI formulation, stated in Equation 1.1, are the observation values, O, and the background values derived from GEM, denoted by B. These inputs are combined to produce an analysis value A. All of these quantities can also be considered in their transformed states, which are represented by o, b, and a respectively. Given this notation, a transformed analysis innovation can be described as $a - o$, and the distribution of the transformed innovations can be described by a normal distribution as follows (*Fortin*, 2012–2013):

$$(a - o) \sim \text{N}\,(0, v) \tag{2.4}$$

where v is the variance of the innovation. Using this variance, an unbiased estimate of the analysis value can theoretically be obtained. First, a multitude of samples can be drawn from the normal distribution of an analysis value in transformed space, $\text{N}\,(a, v)$. These samples can then be back-transformed by applying the inverse Box-Cox function. The mean value of the series of back-transformed samples then provides an estimate of the unbiased analysis value (*Fortin*, 2012–2013). The series

of samples is required to ensure that differences between the expected value of the transformed and un-transformed analysis is captured.

Computationally this is somewhat of a brute force approach, however, and thus the CaPA program instead back-transforms only the calculated analysis value a, and then applies the BCF value described above to account for the error that this choice incurs. The process of quantifying this error starts with considering a Taylor series approximation of the inverse Box-Cox function, $\mathrm{bc}\,()$, centered on the analysis value a (*Fortin*, 2012–2013; *Trim*, 2004):

$$\mathrm{bc}\,(y) \approx \mathrm{bc}\,(a) + \mathrm{bc}'\,(a)\,(y-a) + 0.5 \times \mathrm{bc}''\,(a)\,(y-a)^2 \tag{2.5}$$

Given the above Taylor series, and recalling that the analysis values in the transformed space are approximately normally distributed such that $\mathrm{E}\left[(y-a)\right] = 0$ and $\mathrm{E}\left[(y-a)^2\right] = v$, the expected value of Equation 2.5 can be written as follows (*Fortin*, 2012–2013):

$$\mathrm{E}\left[bc\,(y)\right] = \mathrm{bc}\,(a) + \mathrm{bc}'\,(a)\,\mathrm{E}\left[(y-a)\right] + 0.5 \times \mathrm{bc}''\,(a)\,\mathrm{E}\left[(y-a)^2\right] \tag{2.6}$$

$$\mathrm{E}\left[bc\,(y)\right] = bc\,(a) + 0.5 \times bc''\,(a) \times v \tag{2.7}$$

In order to substitute the full expression for $\mathrm{bc}''\,()$ into Equation 2.7, the expression for the derivative must first be obtained. The first four derivatives of the inverse

Box-Cox function can be found as follows, given that the value of λ_1 is set to be approximately $\frac{1}{3}$ (*Fortin*, 2012–2013):

$$\mathrm{bc}\,(y) = \left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}} - \lambda_2 \tag{2.8}$$

$$\mathrm{bc}'\,(y) = \left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}-1} \tag{2.9}$$

$$bc''\,(y) = (1-\lambda_1)\left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}-2} \tag{2.10}$$

$$bc'''\,(y) = (1-\lambda_1)(1-2\lambda_1)\left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}-3} \tag{2.11}$$

$$\mathrm{bc}''''\,(y) = (1-\lambda_1)(1-2\lambda_1)(1-3\lambda_1)\left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}-4} \tag{2.12}$$

It is important to note at this point that the term $(1-3\lambda_1)$ becomes zero when $\lambda_1 = \frac{1}{3}$ is plugged in. Since this term acts as a multiplier in the expression of the fourth derivative and onwards, all of these derivatives for CaPA's value of λ_1 are equal to zero. As the Taylor series approximation used here includes up to the second derivative, only the third derivative of bc () is ignored in the approximation, and thus becomes a source of error in the calculation of the BCF (*Fortin*, 2012–2013).

Since $\mathrm{bc}''\,()$ is of interest for Equation 2.7, it is worthwhile to rearrange the

equation of the derivative to use the value of A, which is a known value outputted by CaPA, as opposed to a. The requisite modifications are summarized by Equations 2.13 to 2.15, and give rise to Equation 2.16 for $\mathrm{bc}''()$, given that $\lambda_1 = \frac{1}{3}$ and $\lambda_2 = 0$ (*Fortin*, 2012–2013):

$$\mathrm{bc}''(y) = (1-\lambda_1)\left(y^{(\lambda)}\lambda_1 + 1\right)^{\left(\frac{1}{\lambda_1}\right)\times(1-2\lambda_1)} \tag{2.13}$$

$$\mathrm{bc}''(y) = (1-\lambda_1)\left[\left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}}\right]^{(1-2\lambda_1)} \tag{2.14}$$

$$\mathrm{bc}''(y) = (1-\lambda_1)\left[bc(x) + \lambda_2\right]^{(1-2\lambda_1)} \tag{2.15}$$

$$\mathrm{bc}''(y) = \left(\frac{2}{3}\right)[A]^{\frac{1}{3}} \tag{2.16}$$

If the expression for $\mathrm{bc}''(y)$ is now substituted into Equation 2.7, the following result is achieved (*Fortin*, 2012–2013):

$$\mathrm{E}\left[\mathrm{bc}(y)\right] = \mathrm{bc}(a) + 0.5\left(\frac{2}{3}\right)[A]^{\frac{1}{3}}v \tag{2.17}$$

and thus the difference between the expected value of the second order Taylor series approximation of $bc(a)$ and the result of simply taking the inverse of $\mathrm{bc}(a)$ is:

$$\mathrm{E}\left[\mathrm{bc}\left(y\right)\right] - \mathrm{bc}\left(a\right) = \left(\frac{1}{3}\right)\left[A\right]^{\frac{1}{3}} v \tag{2.18}$$

The quantified bias identified in Equation 2.18 must be further amended in order to reflect certain desired properties. The required attributes for the BCF include imposing no bias corrections under the following circumstances (*Fortin*, 2012–2013):

1. $v = 0$, or when a perfect analysis with zero kriging variance is achieved

2. $A = 0$, or when no precipitation is estimated by the analysis

3. $A = B$, or when no observation stations contribute

The first two cases are addressed naturally by the formulation of the BCF presented in Equation 2.18. The third requirement must be artificially inserted into the BCF, as Equation 2.18 will add precipitation to the analysis through the BCF so long as the variance v is greater than zero. This scenario is not ideal. Despite the fact that the GEM background field represents only one realization of the stochastic processes that contribute to the generation of the precipitation, for operational feasibility reasons it is assumed that the pre-processed GEM background field accepted into the SI subroutines is equivalent to the local expectation of the precipitation field; if no observation stations contribute to the analysis and only GEM values are taken into account, the assumed nature of the GEM field dictates that the resulting analysis values must be perceived as correct, and no bias correction should be applied from the data transformation perspective (*Fortin*, 2012–2013). This requirement is

accommodated by applying an additional BCF scaling factor that approaches zero as the kriging variance v becomes closer to the variance value of the GEM background, defined as v_p, and approaches one as v becomes smaller than v_p (*Fortin*, 2012–2013):

$$\text{BCF variance scaling factor} = \frac{v_p - v}{v_p} \tag{2.19}$$

Combining the variance scaling factor in Equation 2.19 with the BCF estimation of Equation 2.18 finally produces the operational BCF stated in Equation 2.3. Unfortunately, based on numerous trial runs incorporating Equation 2.3 into CaPA, the current BCF does show a tendency to underestimate the actual bias associated with data transformation. This is likely a direct result of the numerous simplifications and assumptions made during the derivation of the BCF expression (*Fortin*, 2012–2013). Thus, an as of yet un-answered need persists to address the remaining data transformation bias.

2.1.3 Alternative Bias Correction Approaches

In order to gain perspective on potential bias correction approaches that may be applied to CaPA analyses, a review of the methods currently used for general bias reduction, often in the context of bias related to the spatial resolution of climate variables, can be considered. Such overall biases have been treated in practice by an assortment of statistical prescriptions. This includes the delta change technique and other scaling methods which consider the ratio between observed and simu-

lated climate data means (*Graham et al.*, 2007; *Shabalova et al.*, 2003; *Gellens and Roulin*, 1998; *Teutschbein and Seibert*, 2012). Details pertaining to the delta change and linear scaling approaches are provided in Equations 2.20 and 2.21 as follows (*Teutschbein and Seibert*, 2012):

$$P_o^* = P_o \times \left[\frac{\mu_w\left(P_s\right)}{\mu_w\left(P_o\right)}\right] \tag{2.20}$$

$$P_s^* = P_s \times \left[\frac{\mu_w\left(P_o\right)}{\mu_w\left(P_s\right)}\right] \tag{2.21}$$

where in Equation 2.20, the original precipitation observation data is given by P_o, while the simulated precipitation data is represented by P_s. The corrected precipitation data set is denoted by P_o^*. The mean precipitation value for the observed reference data for a given time window, w, is denoted by $\mu_w\left(P_o\right)$, while the mean simulated precipitation value for the same time period is represented by the $\mu_w\left(P_s\right)$ term. These variable definitions also apply to Equation 2.21, and to all other subsequent appearances of the terms.

The difference between Equation 2.20 and 2.21 primarily comes down to which data set is being modified by the ratio of mean values; since the delta method modifies the observation data, it is not a candidate for application to the CaPA data set. The observation data does not lie on a grid, but rather is spatially sporadic with increasingly poor resolution in northern regions. The CaPA program was designed with the specific intent to address such quality concerns for observation data, and

thus the gridded and temporally regular CaPA analysis data must be the recipient of the bias correction.

Variations of the scaling method exist. Non-linear scaling can be used, which incorporates a non-linear transformation function into the scaling process, allowing for variations in the strength of the correction applied at different points of the simulated precipitation distribution and for alterations to the coefficient of variation (*Leander and Buishand*, 2006; *Teutschbein and Seibert*, 2012). An example of such a function is the power function, which can be applied as follows (*Leander and Buishand*, 2006):

$$P^* = aP^b \tag{2.22}$$

where P^* is the corrected precipitation amount, P is the original, uncorrected precipitation value, and a and b are parameters to be specified for the power function. The estimation of these parameters may be completed, for example, through any of the following approaches: assuming a distribution type for the precipitation data (*Shabalova et al.*, 2003); using the b parameter to ensure that the coefficient of variation of the observation data is matched by that of the adjusted analysis values, followed by designation of the a parameter to ensure agreement between the observation and adjusted analysis means (*Leander and Buishand*, 2006); or by solving a system of equations created by using the power function to equate an analysis quantile to an observation quantile, for two different quantile cases (*Leander and*

Buishand, 2006). The coefficient of variation method is summarized in Equations 2.23 to 2.24 (*Teutschbein and Seibert*, 2012):

$$\mathrm{f}\left(b_w\right) = 0 = \frac{\sigma_w\left(P_o\right)}{\mu_w\left(P_o\right)} - \frac{\sigma_w\left(P_s^{b_w}\right)}{\mu_w\left(P_s^{b_w}\right)} \tag{2.23}$$

$$P_s^* = P_s^{b_w} \times \left[\frac{\mu_w\left(P_o\right)}{\mu_w\left(P_s^{b_w}\right)}\right] \tag{2.24}$$

where in Equations 2.23 to 2.24 the coefficient of variation of the simulated data is made to match that of the observed data through the selection of the power transformation parameter b_w, and the subscript w indicates the window to be considered. The quantile system of equations method is recapped in Equations 2.25 to 2.26 (*Leander and Buishand*, 2006):

$$aQ_{p1,A}^b = Q_{p1,O} \quad and \quad aQ_{p2,A}^b = Q_{p2,O} \tag{2.25}$$

$$b = \frac{log\left(Q_{p2,O}/Q_{p1,O}\right)}{log\left(Q_{p2,A}/Q_{p1,A}\right)} \tag{2.26}$$

where $Q_{p1,O}$ and $Q_{p2,O}$ are the $p1$ and $p2$ quantiles of the observation precipitation data, and $Q_{p1,A}$ and $Q_{p1,A}$ are the $p1$ and $p2$ quantiles of the analysis data.

Another modification of the scaling method includes the pre-processing of the precipitation data to account for dry periods. This can be achieved by first sorting both the observation and climate model data, after which the observed intensity

ranking that corresponds to a pre-selected zero-precipitation threshold can be determined. This rank can then be used to isolate all climate model data with a lower ranking, which are then set to zero (*Berg et al.*, 2012; *Teutschbein and Seibert*, 2012). The inclusion of this type of wet and dry event frequency correction in the linear scaling methodology gives rise to what is termed the Local Intensity Scaling method (*Teutschbein and Seibert*, 2012). A summary of this bias correction approach is provided by Equations 2.27 to 2.29 (*Teutschbein and Seibert*, 2012).

$$P_s^1 = \begin{cases} 0, & \text{if } P_s < P_{st} \\ P_s, & \text{if } P_s \geq P_{st} \end{cases} \tag{2.27}$$

$$SF = \frac{\mu_w \left(P_o | P_o > 0\right)}{\mu_w \left(P_s | P_s > P_{st}\right) - P_{st}} \tag{2.28}$$

$$P_s^* = P_s^1 \times SF \tag{2.29}$$

where P_s^1 is an intermediate bias corrected precipitation data value, P_{st} is a precipitation threshold value selected to correct for the frequency of dry and wet events in the simulation data set, and SF is used to represent a scaling factor created by comparing the mean value of observation precipitation amounts greater than zero for a given window to the mean of the simulated precipitation values greater than the threshold value P_{st} for the same window.

Unlike the delta change method, it is possible for such correction methods as

linear scaling, local intensity scaling, and non-linear scaling to be used to treat CaPA analyses; the simulated values can be taken as CaPA analysis output values, and the various scaling factors and parameters required can be determined based on reference observation data assuming stationarity in time and space, unless spatially varying parameters are pursued. The methods can then be applied to analyses produced by CaPA to reduce overall biases. However, these approaches do nothing to further the understanding of the bias present in CaPA, and treat this bias generically rather than seeking methods to reduce this bias through a more causality-focused approach.

Additionally, because the linear scaling and local intensity scaling methods specifically concentrate on the mean of the precipitation amounts, this becomes the only statistical parameter improved through the application of these methods. All other moments of the precipitation data are ignored by these methods, which means that residual high-order biases may persist in the precipitation data set, remaining observable in the temporal aspects of the data (*Piani et al.*, 2010), and in the precipitation data's distribution. This is in part due to the fact that the extremities of the simulated precipitation distribution is subjected to the same scaling as all other portions of the distribution (*Leander and Buishand*, 2006). In essence, the same changes will be applied to both relatively frequent and extreme precipitation events, disallowing any alteration of the shape of the simulated precipitation distribution (*Teutschbein and Seibert*, 2012).

Furthermore, evidence has been found which suggests that the variability of cor-

rected precipitation values remains large when multiple Regional Climate Model (RCM) realizations are treated with linear and local intensity scaling. Ideally, a bias correction method should lead to the convergence of the treated RCM realizations on the observed values (*Teutschbein and Seibert*, 2012), and thus this result sheds doubt on the capabilities of such linear methods. However, non-linear bias correction has been found to have a comparatively more favourable impact on the variability of corrected results (*Teutschbein and Seibert*, 2012), and is also capable of correcting higher-order moments of the simulated data sets. Regrettably, issues regarding the assumption of stationarity and overly-demanding data requirements persist with this approach.

The shortcomings of simple scaling-of-the-mean-type bias correction approaches are in part addressed by histogram equalization methods, which can alternatively be referred to as statistical bias correction or quantile-quantile mapping (*Piani et al.*, 2010; *Teutschbein and Seibert*, 2012). Histogram equalization promotes the correction of the full precipitation distribution, and thus all moments of the data (*Hay et al.*, 2002). If a zero precipitation threshold correction is also applied, then histogram equalization, similar to the local intensity scaling method, also serves to correct the frequency with which zero precipitation time periods are observed (*Teutschbein and Seibert*, 2012). If errors in the simulated dry event frequency are left unchecked, the tendency for an over-prediction of wet days with small rainfall intensities by climate models may significantly alter the shaping of the fitted distribution and impact the

skill of this technique (*Teutschbein and Seibert*, 2012; *Berg et al.*, 2012).

Histogram equalization is generally achieved through the creation of quantile-quantile plots comparing the simulated climate model data and observation data records, which then support the development of a transfer function. This transfer function in turn allows for the conversion of a model climate data value associated with a given quantile to the amount associated with the same quantile from the observed record (*Berg et al.*, 2012). The fitting of the transfer function can raise important questions, however, as over-fitting may occur if complex forms are used, and yet if linear fitting is practiced, only the first and second moments can then be bias corrected (*Berg et al.*, 2012).

Unfortunately, drawbacks do exist for histogram equalization as well. One such downfall which cannot be ignored is the extensive amount of data that is required for the successful application of this method, as the accurate estimation of the distributions used to construct the transfer function is pivotal (*Berg et al.*, 2012). It is not unusual for data availability to be a limiting factor when intensive data requirements are in place. The histogram equalization method also relies on the the assumption of stationarity, requiring that the cumulative distribution transfer functions produced can be used either at different locations or for different time periods than those they were calibrated for, introducing the potential for further error (*Teutschbein and Seibert*, 2012).

It has also been shown in practice that while the histogram equalization method

allows for a treatment of higher-order moments in regards to bias, it does not necessarily surpass simpler multiplicative scaling methods when it comes down to correctly adjusting mean precipitation values (*Berg et al.*, 2012). Indeed, if a linear relationship is found between the climate model data to be corrected and the reference observation data, then scaling methods may perform equally well (*Berg et al.*, 2012).

As a general note, care must also be taken in the selection of the time periods used to construct reference observation data sets and to perform bias corrections; in the histogram equalization scenario, and in all bias correction cases, longer calibration time periods may not always imply greater accuracy at all time scales. If the selected time frame reaches an annual time scale, for instance, then it is not unlikely that smaller scale biases, such a monthly or daily biases, may become obscured. This may limit the overall bias correction potential of the method, particularly if the seasonal signal for the climate data is out of phase with the observational data (*Berg et al.*, 2012).

Finally, multiple bias correction methods that explicitly treat bias introduced through data transformation procedures exist. For example, this includes calculating bias adjustment factors through the fitting of inverse transformation functions based on sample data (*Miller*, 1984). Such an approach is similar in concept to the BCF bias correction methodology employed by the current operational version of CaPA, and separate research efforts are being focused on treating the data transformation bias issue by improving the BCF. Methods for calculating truly unbiased

and minimum variance estimates of inverse transformation functions have also been developed; unfortunately, these methods tend to be mathematically complex, and generally are not easily implemented in practice (*Miller*, 1984). The desire to avoid such complications was incorporated into the development of the current CaPA data transformation bias correction procedure, and remained a primary consideration in the development of the proposed methodology presented subsequently.

2.2 Correction Methodology for Bias Introduced by Data Transformation

In order to address the current underestimation of bias introduced due to internal data transformation processes in CaPA, a moving-window averaging (MWA) approach has been formulated. This methodology is designed to capture and correct the bias that is bred into a CaPA analysis through the Box-Cox transformation process without disallowing the use of this transformation method. The MWA technique achieves this by employing a MWA data set as input to CaPA, to produce a target analysis which is free of data transformation biases. This target analysis is then capable of correcting 6 or 24 hour analyses produced under basic operational conditions through a scaling technique. Analyses produced under the standard operational CaPA configuration will henceforth be termed *standard analyses*.

The keystone concept of the MWA method is the impact that moving-window

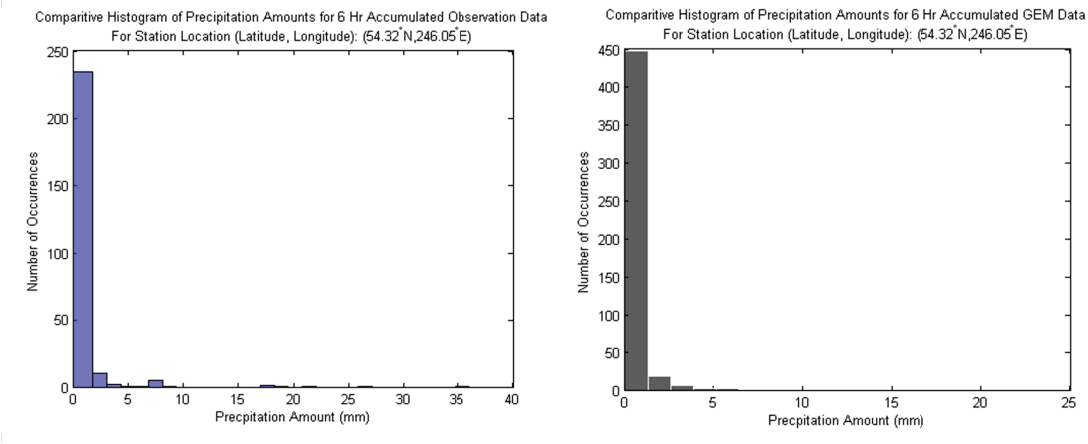

Figure 2.1: Histogram of 6 hour precipitation amounts illustrating the presence of asymmetry in precipitation distributions for the 2010 study period

averaging has on the distribution of the precipitation innovation inputs. Generally, smaller precipitation events occur with a greater frequency than larger, more extreme events. This produces a precipitation distribution that cannot be classified as Gaussian. This generalization also applies to the input innovation data used by CaPA. An example of this can be seen in Figure 2.1, which demonstrates the asymmetrical nature of precipitation distributions through the use of histogram plots. These plots are for a specific location in the study domain, and are presented for both observation and GEM data based on 6 hour accumulation periods.

As confirmed in Figure 2.1, the distribution of precipitation data tends towards a strong degree of asymmetry, with a significantly greater frequency of occurrence for relatively small precipitation amounts. While this asymmetric distribution my be approximated by a number of theoretical distribution types, the Central Limit Theorem dictates that, regardless of the true distribution of a given data set, if a

series of average values can be produced from the data then those average values can be asymptotically described by a normal distribution:

$$\left(X = \frac{1}{n} \sum_{i=1}^{n} x_i \right) \sim N \left(\mu_x, \frac{\sigma_x}{\sqrt{n}} \right) \tag{2.30}$$

where X is an averaged value, x_i denotes a value from the original data set of length n and upon which the averaging is performed, and μ_x and σ_x represent the mean and standard deviation of the original data set. If x_i is taken to be the precipitation input for CaPA, then average values derived from the input data should approximately follow a normal distribution and the effect of extreme precipitation events on the calculation of precipitation statistics such as the semivariogram should effectively be mitigated. Thus, by creating a CaPA analysis based upon monthly-averaged precipitation inputs, data transformation can be avoided, and so too can the associated biases; the effect of the averaging approximates the results of applying the Box-Cox transformation method and removes the need for its use.

The Central Limit Theorem requires that each element of the base data set is an independent random value; this poses a problem when applying this theorem to precipitation data, as each element of a precipitation series is not independent. Rather, each precipitation amount through time is linked through the presence of autocorrelation and persistence in precipitation occurrence inherited from the physical processes that dictate the evolution of precipitation events. Despite this, it is assumed that the averaging procedure should sufficiently normalize the data and

control the influence of extreme precipitation values to be effective in avoiding the need for data transformation, even though a truly normal distribution may not be attained.

With the creation of an approximately normally distributed input data set recognized as a priority, the creation of a MWA data set becomes the first step in the MWA approach. The MWA data set is constructed via the process illustrated in Figure 2.2. The MWA approach calls for the calculation of a MWA value for each time period throughout the entire analysis duration, for every location. However, because a window is taken to be 30 days, the first date of the analysis falls on the 30th day of the study time period, corresponding to the 12 hour accumulation interval for that day. The positioning of the corresponding analysis date within the averaging window is important; by placing the corresponding analysis date at the end of the window, as opposed to the beginning or middle, a CaPA analysis produced using the MWA methodology can still achieve only a modest analysis production delay, producing results in approximately real-time. This is an important feature in terms of the feasibility of implementing the MWA approach in the operational version of CaPA.

Using Figure 2.2 as a graphical reference for the MWA methodology, the calculation of the MWA value corresponding to window i can be expressed as follows:

$$MWA_i = \frac{\sum_{j=0}^{29} P_w}{n_i} \tag{2.31}$$

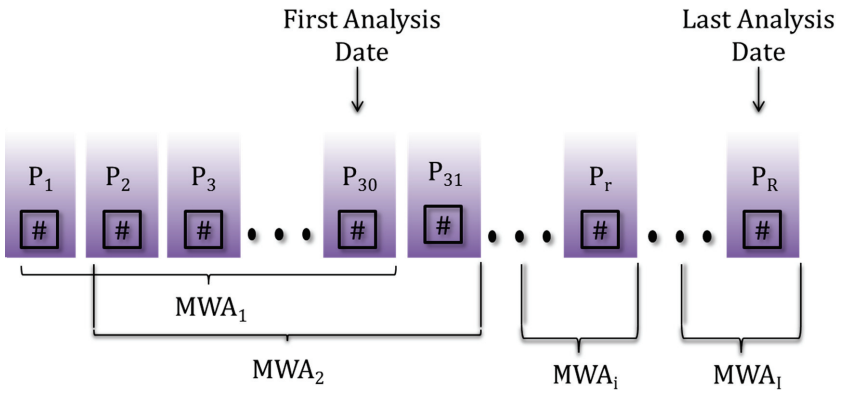

Figure 2.2: MWA scheme employed to create a CaPA input data set. The term P_r is used to denote the r^{th} precipitation value in the time series of observation amounts for a given station. Square brackets are used to group precipitation amounts belonging to a given window, MWA_i. The largest index values for r is R, and for i is I. The square boxes housing number signs denote individual precipitation amounts belonging to the time series

and

$$w = 29 + i - j \tag{2.32}$$

where MWA_i is the window-averaged precipitation value for the i^{th} window, P_r is a recorded precipitation amount contributing to the calculation of MWA_i, j is an index used to identify all time periods in the i^{th} 30-day window, and n_i is used to denote the total number of precipitation amounts actually recorded in the i^{th} window. Not all time periods in a window will necessarily have an associated observed value. Equation 2.31 assumes 24 hour accumulation periods for simplicity, although it can be easily adapted for 6 hour periods. All subsequent equations will be presented assuming 24 hour accumulation periods. It should also be noted that this procedure

is applied to both observed and GEM precipitation values, on a per observation or grid cell basis.

For a specific observation station location, the prevalence of missing observations in a given 30-day window was generally found to be considerable for the selected study time period. This reality was the primary driver for the selection of an averaging scheme over the use of a cumulative precipitation amount; the MWA scheme described accounts only for those days in the 30-day window that possess precipitation records. In this way the influence of missing observation records is minimized, and the MWA approach can proceed despite the temporal discontinuity in data records. By implementing a minimum record availability threshold of 20 out of 30 days, the representativeness of each MWA value for its windows is loosely controlled, and any bias introduced through inconsistencies in the placement of missing days of record across observation stations is minimized. Furthermore, the averaging scheme does not discriminate between occurrence and non-occurrence precipitation records; all recorded data values greater than or equal to zero are included in the averaging process.

Once the compilation of the MWA input data for each observation station and grid location is complete, the pre-processed MWA data are fed into CaPA to act as the observation and background inputs. A target analysis is then generated. Since the MWA data set precludes the use of the Box-Cox transformation, this aspect of the CaPA program is disabled when producing the target analysis. The Box-Cox

transformation is re-enabled for the production of the standard analyses, which can be based on either 6 or 24 hour accumulated precipitation inputs. The standard analyses become the focus of the scaling technique; each precipitation amount for each day within a standard analysis must be multiplied by its own unique scaling factor, in order to produce a corrected analysis vector. This procedure is summarized by the following equation:

$$A_r = \frac{\sum_{m=1}^{M} P_r \times SF_{r,m}}{M} \tag{2.33}$$

where A_r is the corrected standard analysis precipitation amount for the r^{th} time period, P_r is the original standard analysis amount for the same analysis time, and $SF_{r,m}$ is the m^{th} scaling factor associated with that time. The total number of scaling factors estimated for a time period, which indicates the number of times a time period is captured by the moving window, is given by M. The value of M is defined by the following function, given that i is set to I, or the total number of windows:

$$f_m(r,i) = \begin{cases} r, & \text{if } r < 30 \\ 30, & \text{if } 30 \leq r \leq i \\ R - r + 1, & \text{if } r > i \end{cases} \tag{2.34}$$

where R is the largest index value for r in the time period series. If $i < 30$, then the $30 \leq r \leq i$ case is removed from Equation 2.34.

Each of the scaling factors, $SF_{r,m}$, are determined in such a way that the bias introduced into the standard analysis through the data transformation process is accounted for and mitigated by the scaling procedure. When this is achieved, applying the same MWA scheme as previously described to the corrected analysis will result in the production of a corrected MWA data set that is consistent with the target analysis generated using the originally determined MWA input data. Thus, each $SF_{r,m}$ value must ensure consistency between the target analysis and the resulting corrected standard analysis. The scaling procedure is illustrated in Figure 2.3.

The method used to determine the $SF_{r,m}$ values applied during MWA bias correction possesses two primary manifestations. MWA scheme I represents the simplest scaling paradigm, generating one scaling factor applicable to all time periods for each window, and then averaging the resulting corrected analysis values for each time period as follows:

MWA scheme I

For a given window i, all analysis values for the grid cell of focus lying in the associated 30-day window are identified, and are denoted by the vector P_r. The vector of scaled analysis values, A_r, is determined as follows:

$$SF_{r,m}|_{r=i\rightarrow i+29} = \frac{(MWA_{C_i} \times 30)}{\sum_{j=0}^{29} P_w} \tag{2.35}$$

where MWA_{C_i} is the CaPA analysis value produced using MWA inputs for window i.

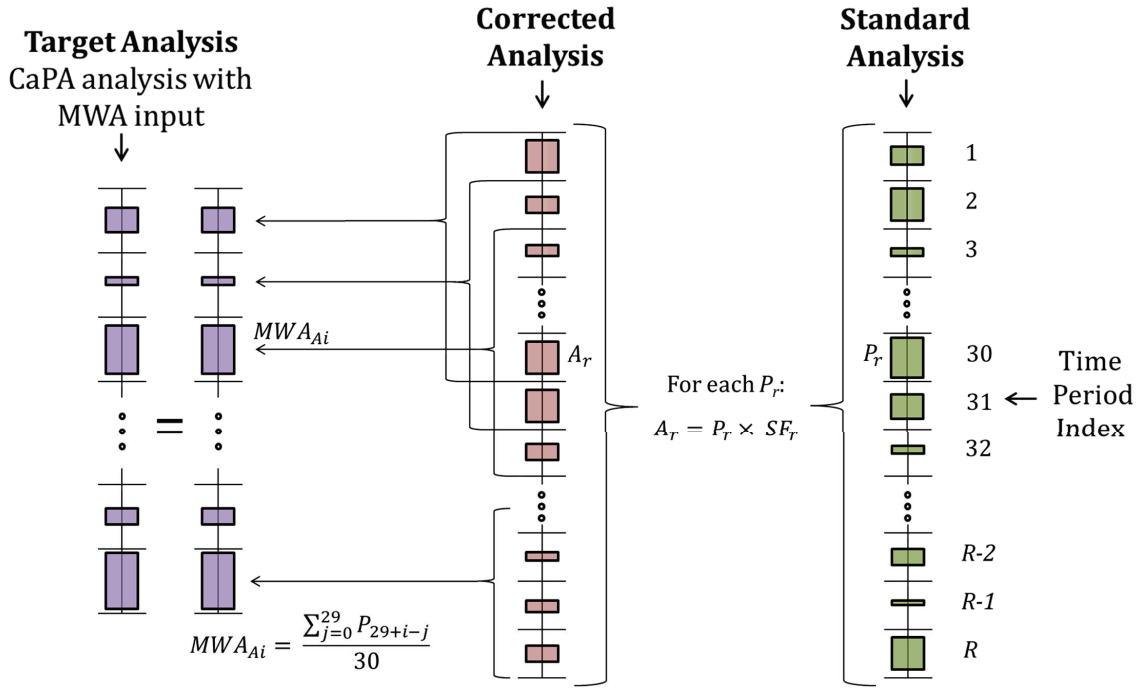

Figure 2.3: MWA bias correction process flow chart. For a given grid cell, each value in the standard analysis time series, denoted P_r, is treated with its own scaling factor, SF_r. This produces a series of corrected values, represented by A_r. When MWA values, given by MWA_{Ai}, are constructed from the corrected analysis, they are consistent with the target analysis, or the analysis values obtained by running CaPA with MWA inputs. The MWA_{Ai} value is divided by 30 as all 24 hour time periods are theoretically available from the CaPA analysis. This figure does not account for the over-lapping nature of the windows

The w index value is calculated according to Equation 2.32. The m index also varies depending on r, and can be determined by applying the $f_m(r, i)$ function provided in Equation 2.34. The value of $SF_{r,m}$ is then used in Equation 2.33 in combination with P_r to produce A_r.

The second scaling factor optimization approach is designed to take into consideration the analysis error values outputted along with an analysis by CaPA; if an analysis value possesses a relatively large analysis error, which is the square root of the kriging variance, then the amount of uncertainty that surrounds that value can be thought of as greater than that pertaining to a value with a lower analysis error. As such, there is merit in preferentially scaling values with larger analysis errors over values with lower errors. In order to take into account analysis errors during the scaling factor specification process, the following scheme is adhered to:

MWA scheme II

For a given window i, the vectors of corrected and original analysis values are once again defined as A_r and P_r, respectively. MWA scheme II is commenced through the following initialization:

$$A_r = P_r \tag{2.36}$$

Once the vector A_r has been duly initialized, scheme II progresses by attempting to minimize the following quantity through successive iterations for each window:

$$(MWA_{C_i} \times 30) - \sum_{j=0}^{29} A_w \tag{2.37}$$

where a tolerance threshold of 10^{-6} is used to define a satisfactory minimization. If the tolerance threshold is not reached in 10000 iterations, the minimization is ceased. Equation 2.32 once again applies to the w index. For each iteration, an updated form of A_r is produced according to the following,:

$$A_r = \begin{cases} A_r \times (1 + \eta_{r,m}), & MWA_{C_i} - \sum_{j=0}^{29} A_w > 0 \\ A_r \times (1 - \eta_{r,m}), & MWA_{C_i} - \sum_{j=0}^{29} A_w < 0 \end{cases} \tag{2.38}$$

where the scaling factor $SF_{r,m}$ is equivalent to either $(1 + \eta_{r,m})$ or $(1 - \eta_{r,m})$ depending on the case. The relationship between the r and m index values and the window value i is consistent with that described for scheme I. The term $\eta_{r,m}$ is calculated via the following for a given window i:

$$\eta_{r,m} = \frac{\sigma_r}{\sum_{j=0}^{29} \sigma_w} \tag{2.39}$$

where σ_r is used to denote the analysis error corresponding to P_r. Equation 2.33 then applies. Note that in applying the update in Equation 2.38, the correction may lead to the test criteria of Equation 2.37 transitioning from positive to negative, or vice-versa. When this occurs, the scaling factors applied during the correction are modified to produce a final result for Equation 2.37 equal to zero.

Obviously, the success of MWA scheme II hinges upon the accuracy of the analysis errors. If these values are not representative of the true uncertainty of each of the CaPA analysis estimates, then little value will be gained from incorporating them into the MWA correction methodology. MWA scheme I safeguards against the possibility that the reported CaPA analysis errors are not accurate depictions of the true estimate uncertainties.

The implementation of the MWA schemes is discussed in Appendix B. Multiple modifications to the CaPA code were required to accommodate the MWA methodologies. The following details regarding the use of the MWA schemes are also addressed in Appendix B: testing of CaPA's ability to estimate a semivariogram for the MWA input, selection of a minimum number of days of record required per window, and assessment of the normality of the MWA precipitation inputs.

2.2.1 Merits of the MWA Methodology

The MWA technique's theoretical avoidance of long-term data transformation bias in corrected analyses is demonstrated in Appendix C. Based upon this result, the MWA correction technique can be conceptualized as a candidate method for the production of a fully unbiased analysis in regards to data transformation. However, other promising bias correction methods also exist, as summarized in Section 2.1.3, which focus on data transformation specifically or on treating bias as a whole. With this in mind, it is important to note at this juncture that the MWA correction method

has a variety of advantages over existing bias correction practices.

The first advantageous attribute of the MWA schemes is the simplicity of implementing the technique in the CaPA environment. The MWA schemes primarily require three major new considerations for CaPA: the pre-processing of GEM and observation data, the ability to execute sets of runs to create the standard and MWA analyses in tandem, and finally the post-processing of analyses produced based upon the MWA scaling procedure. All of these requirements can be easily accommodated in a permanent fashion, although the work done for this investigation was external to the CaPA environment. All other modifications to the CaPA program require only mild coding adjustments and the re-specification of various CaPA configuration parameters.

While initial optimization of the technique may be required, the automation of the overall correction methodology allows for little user intervention once the MWA bias correction alterations are established. The calibrated parameters will likely require only minimal reassessment under changing climatic conditions in the future, whereas other methods may need to undergo more cumbersome re-calibrations; for many alternative bias correction approaches, the scaling factors themselves would need to be re-established, as opposed to modifying simple operational values such as the minimum days of record per window for the MWA case. Most importantly, the concepts driving the MWA method are simple compared to some of the more complex alternatives, making CaPA users more amenable to its integration now and

in the future.

Beyond the simplicity of the method, the MWA methodology also takes advantage of the analysis error estimates produce by CaPA, capitalizing on the full suite of information that the program offers. This allows for a probabilistic dimension to be embedded in the analysis, and allows the MWA method to become adaptive to the skill of the SI data assimilation procedure in ways that other bias correction approaches cannot. This may prove particularly useful in regions with spatially complex or highly graded precipitation fields, as kriging uncertainty estimates have been shown to be large under such conditions (*Erdin and Frei*, 2011).

Lastly, the MWA method requires relatively little data to operate effectively, as compared to other methods. Many bias correction practices involve estimating the scaling factor parameters based on long time series of data, often spanning decades when possible. Only 30 days of precipitation records are required per station and grid cell prior to the analysis start date under the MWA system. This reduces the impact of problems related to the temporal availability of quality data records felt in many regions of Canada, particularly in the northern portions of the country. Since the MWA method does not assume spatial stationarity, sensitivity to the spatial resolution of observations still exists, however, with the accuracy of corrections in data sparse regions being less accurate than those in data rich areas.

As a further means of supporting the two proposed MWA schemes, it should be noted that a similar approach involving the calculation of linear scaling factors

has been implemented in a study by Leander and Buishand (2006). These scaling factors were calculated for five day periods as the average precipitation values of the observations divided by the model data's average using a 65 day window centered on the five day period of interest. The performance of this correction method was found to be disappointing for the upper quantiles of precipitation (*Leander and Buishand*, 2006). Fortunately, the MWA scheme II proposed here includes the use of analysis errors, which diverges the method from true linear scaling. As such, it is anticipated that the downfalls of treating all precipitation amounts with the same scaling factor will be avoided. However, it is possible that MWA scheme I will encounter performance issues similar to those observed under strictly linear scaling conditions.

The proposed MWA schemes, similar to the Leander and Buishand (2006) methodology, have an additional advantage over many of the techniques currently used in practice; the MWA method is not dependent upon the assumption of stationarity. The use of a moving window in the MWA method allows for the scaling of standard analyses to take place on a per location and per time period basis, and thus does not require the assumption that any scaling factors remain applicable for locations or time periods for which they were not specifically calibrated. Many traditional bias correction methods, such as those discussed in Section 2.1.3, require such a stationarity assumption in order to reduce the number of scaling factor estimations or distribution parameterizations that are needed. As a result, bias correction factors

are often based on longer times series of data, in an attempt to make the bias corrections more robust when applied to future or spatially offset data. However, the stationarity assumption should be considered as a weakness for any bias correction method, as it assumes a constancy in the relationship between observed and simulated climate variables that cannot be confirmed, particularly in light of the effects of climate change (*Teutschbein and Seibert*, 2012).

2.3 Results

The performance of CaPA is routinely assessed through a supplementary program created and maintained by Environment Canada, which automatically calculates the verification statistics described previously in Section 1.1.4. This program uses the *statoma.obs* files produced by CaPA as input, which contain analysis results at all observation locations determined based on a leave-one-out approach; the leave-one-out process calculates an analysis value at an observation location without including that observation as input to CaPA. Verification scores can then be calculated based on all observation locations. However, common practice at Environment Canada is to verify analyses against only high quality synoptic station observations (*Fortin*, 2012–2013; *Lespinas*, 2012–2013). These synoptic stations include both manual and automatic types from the American and Canadian gauge networks (*Environment Canada*, 2013). Unfortunately, this approach to analysis verification is not applicable when using the MWA methodology, due to post-processing requirements.

During the post-processing component of the MWA method, a standard analysis must be scaled using a target analysis, in order to account for the data transformation bias. Thus, a target analysis must be generated independently of the standard analysis, by replacing all observation and GEM input into the CaPA program with MWA input for the entire time period. The requirement of two independent analyses renders the traditional leave-one-out verification approach cumbersome and time consuming.

The MWA approach inherently imposes a set of quality control restrictions on the data used to create the final MWA CaPA input. As a result, not all observation stations that possess values for a given time period necessarily have a corresponding MWA value. This means that a *statoma.obs* file produced for a target analysis will not contain leave-out-one analysis values for all observation stations that would be included in the corresponding standard analysis *statoma.obs* file, nor will the set of leave-one-out stations necessarily match those listed in other time period's target analysis *statoma.obs* files. This problem becomes amplified when the post-processing technique used in the MWA method is considered; a 30 day series of leave-one-out analysis values is required to be matched with each given target analysis value for the MWA processes to be executed for a given location. Thus, matching target and standard analysis values would be required not only for the current time period, but also for the proceeding 29 time periods. Given the well known issue of poor temporal resolution for many observation stations, this is not feasible.

If the traditional leave-one-out verification approach is to be pursued, three options for dealing with the problems outlined above can be identified: 1) The MWA technique can be employed with missing target analysis values within the sliding averaging windows during the scaling procedure, 2) Missing standard analysis values required for the MWA approach can be generated through interpolation between leave-one-out values that are available, 3) Only those cases where all required MWA analysis values are provided by the leave-one-out data set are considered. The first case is not desirable, because allowing for missing values only further imbeds uncertainty and possible error into the results of the approach, and may significantly undermine the ability of the method to properly correct bias. The second case should also be avoided, as it is not unlikely that no reasonably close stations are contained in the leave-one-out data set. Thus, the interpolation would not result in truly representative values for the required location, leading to artificially low verification scores. Finally, based on experience obtained in working with the study's data set, it was felt that the third option would result in too few qualified cases, and thus the verification scores calculated may not be statistically relevant.

Given that no course of action can be identified that would reasonably allow for the implementation of the traditional leave-one-out verification approach when dealing with the MWA technique, an alternative verification method was adopted. This method required the production of four different sets of CaPA analyses, each based on a different verification percentage or level; 10%, 20%, 30%, and 40% were selected.

For each case, a loosely approximated grid of verification stations was identified, representing the given percentage of eligible verification stations. In this case, eligible verification stations were defined by station ID code types 12 and 146, which correspond to manual and automatic synoptic gauges respectively (*Environment Canada*, 2005).

In order to identify the verification stations to be selected for each case, a generalization of the analysis domain to a rectangular area with a longitudinal width to latitudinal height ratio of 2:1 was made. This ratio assumption was felt to be sufficient for the purposes of extracting verification stations, despite the true domain being more adequately, albeit still roughly, represented by a rectangular region with a width to height ratio of 2.034, and defined by the following latitudinal and longitudinal coordinates: $(43.42°, 236.53°)$, $(62.12°, 236.53°)$, $(43.42°, 274.57°)$, $(62.12°, 274.57°)$. The act of identifying verification stations based on a grid framework is intended to assist in ensuring that the coverage of the domain captured by the verification stations is uniform, preventing specific regions which may posses unique precipitation characteristics, or regions with low gauge densities, from being ignored during the verification process.

Table 2.1 lists the number of verification stations selected for each verification case, and the corresponding true verification percentages. A graphical representation of where the selected verification station locations fall within the analysis domain is provided in Figure 2.4 for the 20% verification case. All other verification cases are

Table 2.1: Verification observation station extraction grids for all verification percentage cases. Grid dimensions in cells

$NominalVerif.\%$	$GridDimensions$	$No.ofStations$	$TrueVerif.\%$
$10\%Verif.$	$8x5$	40	10.72%
$20\%Verif.$	$12x6$	72	19.30%
$30\%Verif.$	$14x8$	112	30.03%
$40\%Verif.$	$17x9$	153	41.02%

presented in Appendix D.

Based on the assumed 2:1 ratio, the number of stations required in the longitudinal and latitudinal directions were determined such that the total number of stations required was found to be as close as possible to the total number required for the verification percentage case. Once the required number of stations in each direction was identified, the idealized spacing of these stations within a grid configuration and the resulting latitude and longitude coordinates of the locations were determined, assuming the approximate domain rectangle described previously. Note that along each dimension of the domain rectangle, a border of one spacing unit was assumed to apply. This was done to assist in preventing the selection of verification station locations with no observation stations in the immediate proximity, which was a concern given the preference for overestimating the domain extents rather than underestimating when defining the rough domain rectangle.

Once the idealized verification grid locations were pinpointed, the actual verification grid was determined by finding the closest observation station to each of the identified verification locations, and extracting these observation stations from the input data sets upon which the subsequent CaPA analyses were computed. The

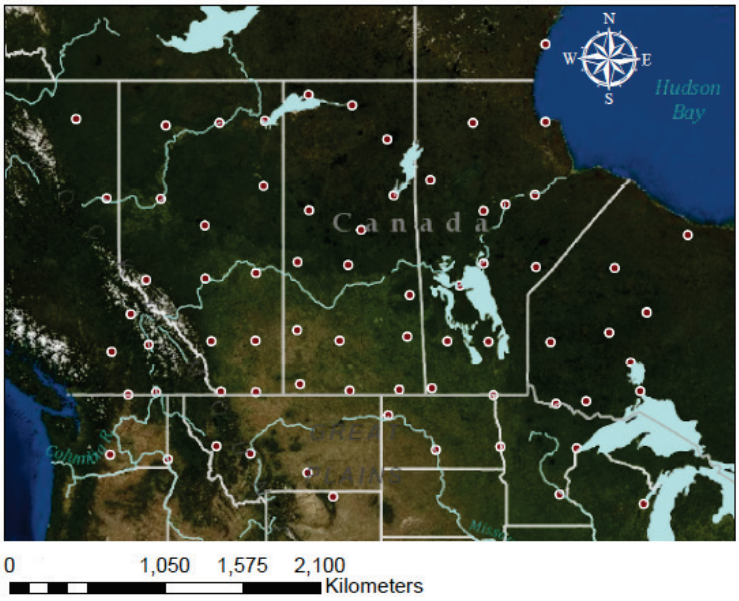

Figure 2.4: Verification station location map. The selected stations represent approximately 20% of the synoptic observation network within the domain, and are loosely distributed on a grid

CaPA analyses were then verified against these verification observation records by employing bi-linear interpolation to derive the analysis values that corresponds to each of the verification locations.

Given that the primary objective behind the implementation of the MWA scheme is the reduction of bias in the treated CaPA analyses, a cogent choice for commencing a review of the MWA method results is to examine the overall bias of all CaPA runs, for the verification cases as outlined above. These results are presented in Table 2.2, and represent the bias values for the analyses averaged across all time steps and all verification locations for the entire study domain and time period.

For all results presented in this chapter the following descriptions apply: SI de-

Table 2.2: Average bias results under simple interpolation conditions. SI refers to the no Box-Cox case, BCSI refers to the Box-Cox case and represents the current operational configuration, BCSIc denotes the corrected Box-Cox case using MWA scheme II, and BCSIcc refers to the corrected Box-Cox run using MWA scheme I.

Run	10% Verif.	20% Verif.	30% Verif.	40% Verif.
SI	0.26732	-0.00017	-0.01626	0.00314
BCSI	0.3434	0.07431	0.04854	0.06915
BCSIc	0.29454	-0.02481	-0.04541	-0.02303
BCSIcc	0.2897	-0.02804	-0.04977	-0.02724
GEM	0.2782	-0.00181	-0.00439	0.01671

notes a CaPA analysis based on simple interpolation; BCSI refers to an analysis based on simple interpolation and which includes the influences of a Box-Cox transformation, and thus is a standard analysis; BCSIc is a standard analysis corrected using MWA scheme II; and BCSIcc is a standard analysis corrected with MWA scheme I.

As can be seen from Table 2.2, both version of the MWA scheme produce absolute bias values that are lower than those obtained by the current operational version of CaPA. The only exception to this statement is the result obtained by MWA scheme I under the 30% verification scenario, at which point a 2.5% increase in the absolute bias value is observed.

The operational configuration produces positive biases in all cases. Conversely, the introduction of the MWA technique reduces this bias in the 20%, 30%, and 40% cases to result in negative final biases. The GEM background field and SI analysis also give rise to negative biases, but with smaller magnitudes than in all other cases. The SI analysis is a CaPA run with no data transformation and only the SI method enacted.

The results provided in Table 2.2 indicate that the reduction of bias accomplished through the use of the MWA method is an outcome that is consistent across all verification cases tested; because of this agreement, only the 20% verification case will be studied in greater detail, and the results presented for this case will be assumed to be representative for all verification cases. The details regarding the selection of the 20% verification case for continued focus, as well as further points of interest arising from the results presented in Table 2.2, are discussed in Section 2.4.

Figure 2.5 continues to focus on the average bias metric, providing a visual representation of the 20% verification results also reported in Table 2.2. This figure emphasizes the reduction and change in sign experienced by the average bias when the MWA method is executed, while also highlighting the notably lower bias amounts manifested when CaPA is run with no data transformation considerations, and when analyzing the GEM background field. Comparing the transformation-free results with the operational case highlights the fact that the majority of the operational bias in this study results from data transformation. This bias is also seen to dominate the GEM bias, which is derived from the GEM field's over-abundance of small precipitation events. This feature of GEM is confirmed in the following DPM and FBI scores as well.

Additional continuous analysis skill scores of interest are presented in Figures 2.6 and E.1. In Figure 2.6 the discrepancies in DPM values for varying thresholds can be seen across the run configurations tested. The current operational configuration,

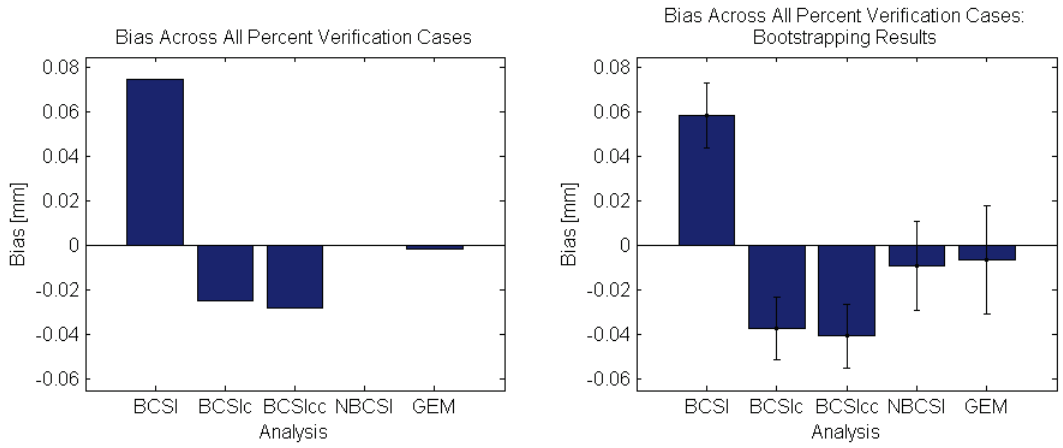

Figure 2.5: Bias results for multiple run scenarios averaged across all verification locations and all time periods in the study. Note that the results shown are for the 20% verification case

denoted by BCSI, performs well when small to moderate precipitation amounts are considered, but experiences a continual reduction in DPM values as threshold values increase, falling into negative values. All other CaPA run scenarios maintain positive DPM values. The SI case possesses the largest DPM values for small to moderate precipitation thresholds, and approximately converges with the GEM DPM results for threshold values greater than 25 mm. Both the SI and GEM cases produce a final DPM value that is better than that obtained by both MWA correction schemes. Scrutiny of the bottom panel of Figure 2.6 reveals the DPM results obtained through the bootstrapping procedure, and emphasizes the significance of the negative DPM results attributed to the BCSI case, as compared to all other scenarios which exhibit overlapping of their confidence intervals.

In Figure E.1, in Appendix E, the DPS continuous skill scores are depicted. As can be seen, the BCSI configuration still produces the most negative results

for larger thresholds, but in this case all scenarios terminate in negative values. Direct comparison of Figures 2.6 and E.1 highlights a slight worsening in the relative performance of the GEM background field for small to moderate threshold values, as well as a lower threshold associated with the transition from under to over-performing compared to the SI case. Apart from these differences, most of the basic trends and relative relationships witnessed in Figure 2.6 are upheld in Figure E.1, albeit with more pronounced differences between cases for larger precipitation thresholds. This increased spread in the DPS results for larger thresholds is preserved in the bootstrapping results presented in the bottom panel of Figure E.1, although all confidence intervals overlap with at least the adjacent intervals in the plot. In both the DPM and DPS figures, the MWA approaches outperform the current operational method for larger precipitation thresholds.

Categorical skill scores were also calculated to assess the MWA methodology's performance. FBI scores based on binned precipitation data are presented in Figure 2.7, where once again the top panel shows the results for the specific study data, while the bottom panel results are generated through the bootstrapping procedure. As the results presented are $FBI - 1$, zero represent a perfect score while a negative value indicates an underestimation of precipitation by CaPA, and vice versa.

The current operational case performs relatively well for precipitation amounts less than 2 mm, but beyond this value the performance deteriorates to the worst of all compared. In a physical context, CaPA's operational performance is strong when

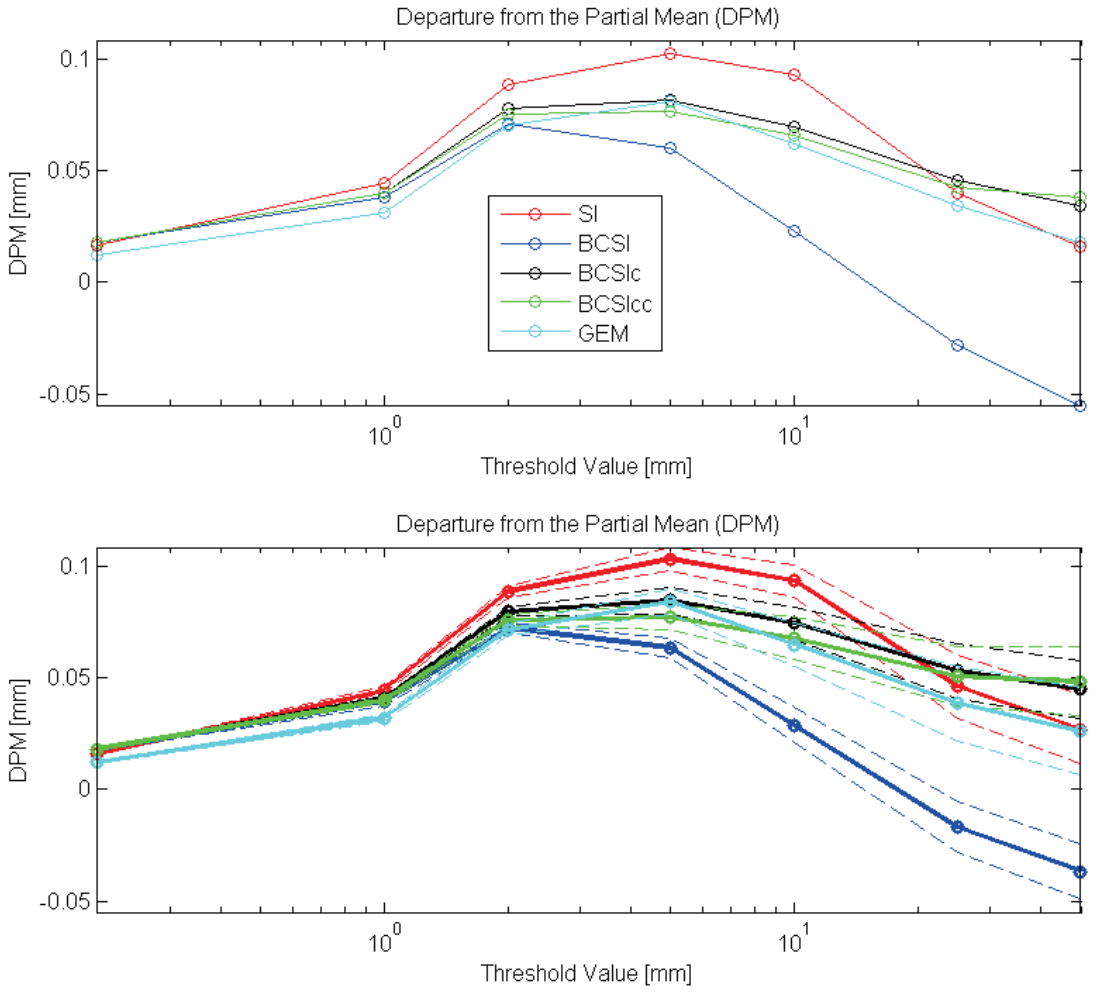

Figure 2.6: **Top:** DPM results averaged across all verification locations and all time periods in the study. **Bottom:** DPM bootstrapping results, with the bootstrapped mean DPM values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines. Note that the results shown are for the 20% verification case

dealing with drizzle-like precipitation, but suffers from a dry bias for more significant precipitant events. Both MWA approaches perform well for amounts greater than 2 mm, surpassing the current operational configuration. For precipitation amounts greater than and including 10 mm, the MWA approaches result in lower FBI scores than either the SI or GEM background cases as well. These results are echoed in the lower panel of the figure, although in many cases overlapping confidence intervals can be seen. For precipitation values in the 1 to 25 mm range, no overlap occurs when the current operational and MWA scheme II configurations are compared, implying that the differences are significant based on the bootstrapping assessment method.

A second set of categorical skill scores are summarized graphically in Figure 2.8, where larger ETS values are more desirable. This figure shows that the best performance, as gauged by the ETS metric, is consistently obtained by the current operational case, BCSI. The worst performance is put forth by the GEM background field. Except for bins within the range of 2 to 25 mm, relatively little difference is seen between the SI and MWA cases, which is supported by the bootstrapping results depicted in the lower panel. The SI analysis produced better ETS scores than the MWA analyses in the 2 to 25 mm range.

Adjusted ETS scores, denoted as aETS, are presented in Figure E.2 in Appendix E. Comparison of Figures E.2 and 2.8 reveals numerous deviations between the two sets of scores. When the impacts of bias are accounted for by the aETS, changes in the relative performance rankings can be seen in the 0 to 1 mm and 25 to 50 mm

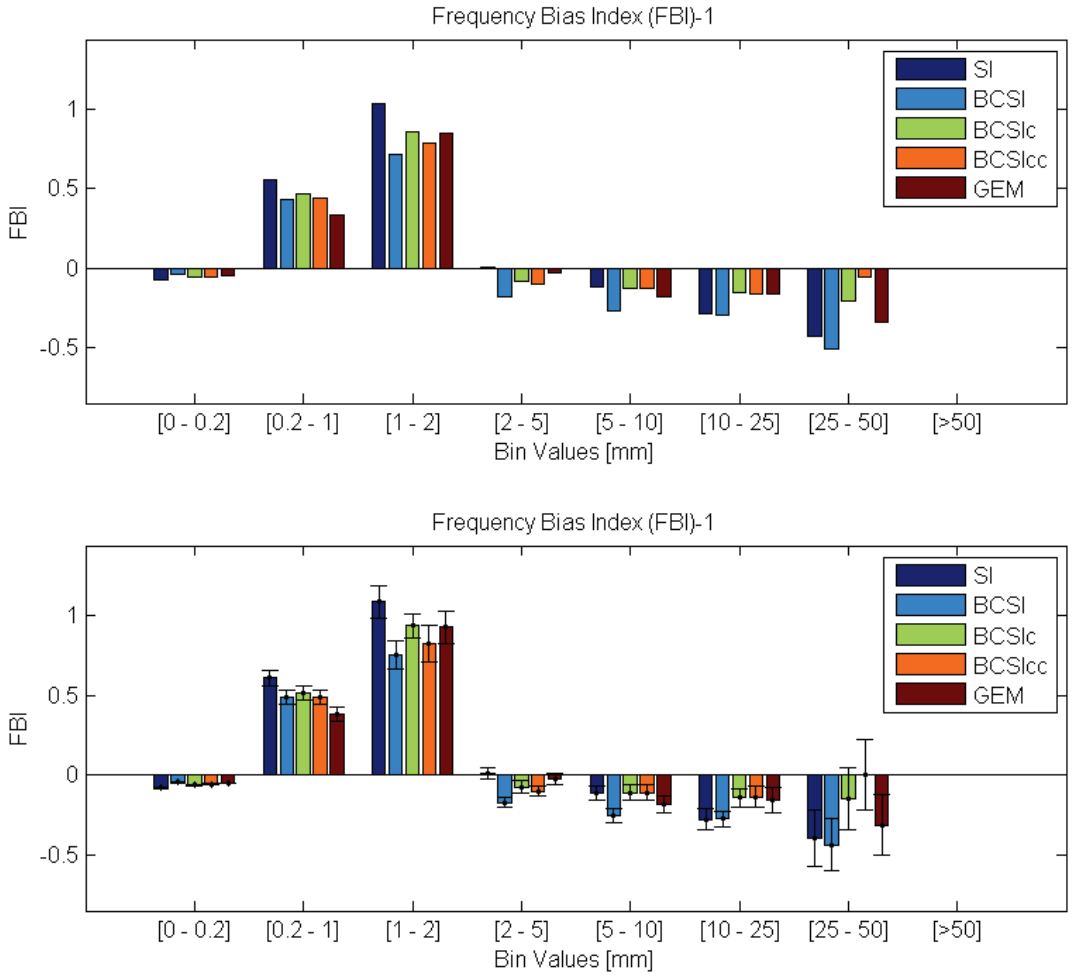

Figure 2.7: **Top:** FBI results averaged across all verification locations and all time periods in the study. **Bottom:** FBI bootstrapping results, with the bootstrapped mean FBI values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines. Note that the results shown are for the 20% verification case

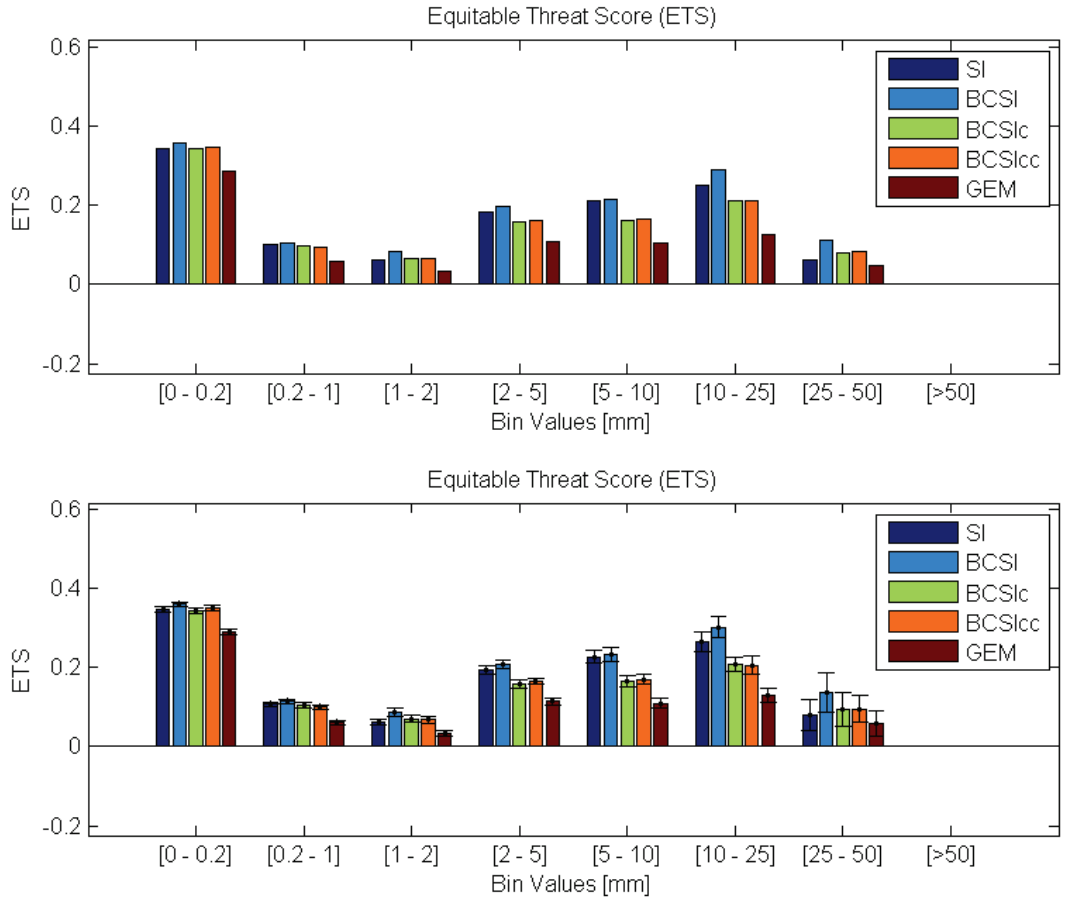

Figure 2.8: **Top:** ETS results averaged across all verification locations and all time periods in the study. **Bottom:** ETS bootstrapping results, with the bootstrapped mean ETS values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines. Note that the results shown are for the 20% verification case

ranges. However, the greatest relationship of interest is that between the performance of BCSI and BCSIc, or the operational and MWA scheme II runs; in both the ETS and aETS skill score sets, BCSI outperforms BCSIc. The same dominance of the BCSI analysis is also seen over BCSIcc, or the MWA scheme I data product, with the exception of the 0 to 0.2 mm bin values where BCSIcc produces a larger aETS value than BCSI. For the 1 mm to 25 mm range, the ranking of the assessed cases remains unchanged, although the magnitudes of the differences observed between cases in each bin appear more amplified in the aETS result set.

For the purposes of this investigation, the differences in performance between the current operational configuration of CaPA, and the MWA scheme II analysis is of the utmost interest. In order to better elucidate these differences, Figure E.3 has been provided in Appendix E. As can be seen, for all metrics the difference between the two cases of interest are found to be significant, except generally in specific regions where a reversal in the ordering of performance occurs.

Finally, Figure 2.9 provides a summary comparison between all runs conducted during the bias correction investigation. The figure clearly demonstrates a higher degree of agreement between the observation and BCSI, or operational run, quantiles for smaller precipitation amounts of up to approximately 1 mm, with the worst agreement generally seen for the SI, or un-treated, case. The results for the MWA schemes, as well as GEM, lie between these two run cases for these small precipitation amounts. For moderate precipitation quantities, the BCSI case degrades in

performance and exhibits the least skill in terms of quantile matching, while all other cases appear to perform relatively similarly. For precipitation events larger than approximately 6 mm, the MWA schemes are the most capable of matching the observation quantiles. The difference between MWA schemes I and II are minor in most cases, although for the largest quantile MWA scheme I does show a notably improved result over MWA scheme II.

2.4 Discussion

In order to extract something meaningful from the array of results presented in Section 2.3, a more detailed look into what is being communicated by these results, and what the implications of these results are for the project objectives, is imperative. In order to clarify the interpretation of these results, and to maintain an orderly and logical dissection of their components and implications, the discussion of the bias correction results is sub-sectioned based on the issue of focus, as presented below.

2.4.1 The Large Bias of the 10% Verification Case

To begin the inquest into the insights revealed by the bias correction results, attention can be focused on Table 2.2. One feature of this table stands out sharply; the reported biases across all tested CaPA setups are significantly greater for the 10% verification case, as compared to all other verification cases. In order to extract the reasoning behind this finding, Figure 2.10 was produced. This figure looks at the precipitation

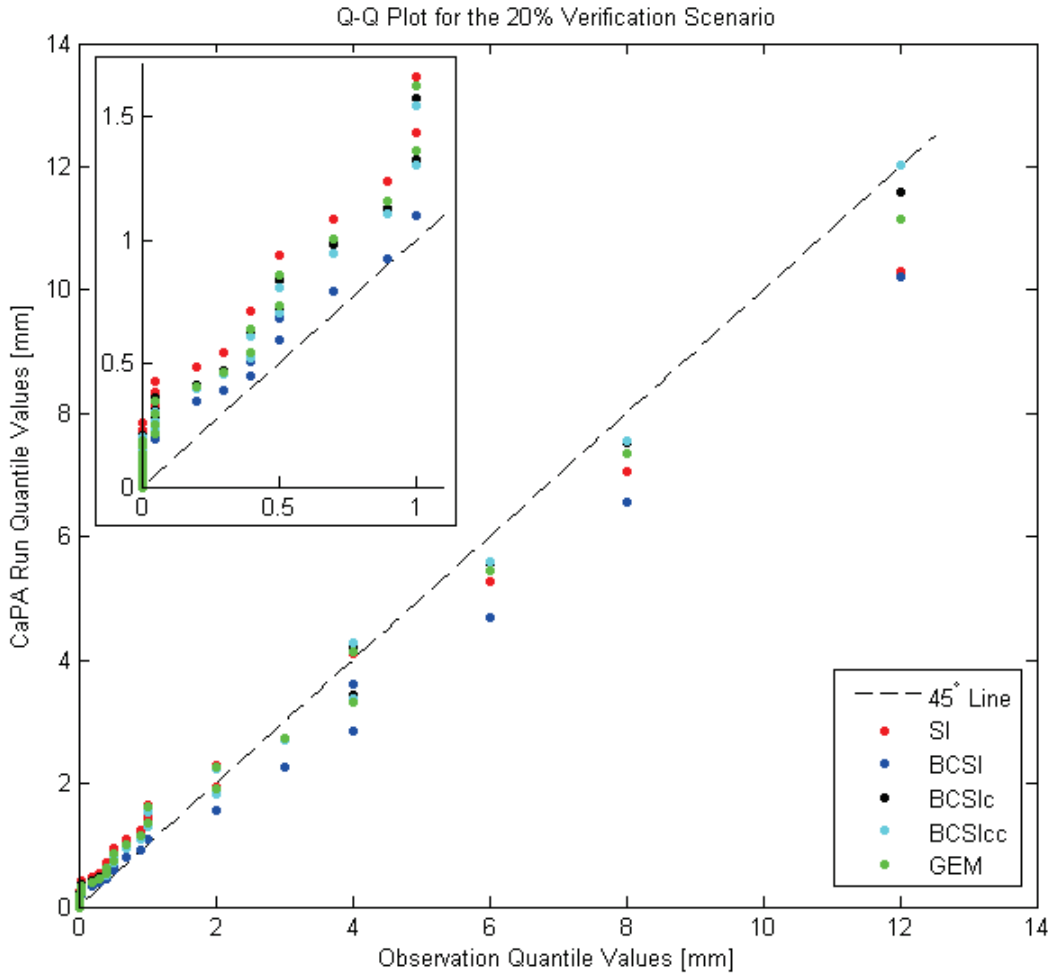

Figure 2.9: Quantile-quantile plot comparing the performance of all CaPA scenarios tested. The inset figure in the upper left corner is a zoomed-in view of the results for the smaller precipitation amounts

amounts that correspond with various quantile values for the BCSI, or operational CaPA analysis, as well as for the observed verification data, GEM data, and input innovation data.

Figure 2.10 pinpoints the scenario surrounding the 10% verification case as unique compared to the other verification scenarios. This is seen by the departure of the 10% verification empirical CDF curves from those associated with 20%, 30%, and 40% cases which otherwise display a strong visible similarity. However, this glaring difference in the empirical CDF curves between the 10% and all other verification cases only occurs when considering the CaPA analysis, GEM data, and innovation data. The plotted curves for all verification cases generally agree for the observed data. This presents a strong clue as to the source of the aberration for the 10% case.

In order to create a CaPA analysis, input data in the form of innovations must be fed into the CaPA program. This input data, or innovation data, can be decomposed into two base sources: observation data and GEM data. Since the observation data shows no indications of any notable distribution dissimilarities across the verification cases, GEM data becomes the likely culprit. This suspicion is confirmed through inspection of the bottom left panel of the figure, which does indeed show evidence of the 10% verification GEM data set possessing an empirical CDF that does not follow the trend set by all other verification cases. Once the GEM and observation data are combined to create the innovation data, the abnormal 10% empirical CDF understandably manifests, and likewise this irregular CDF is also translated in the

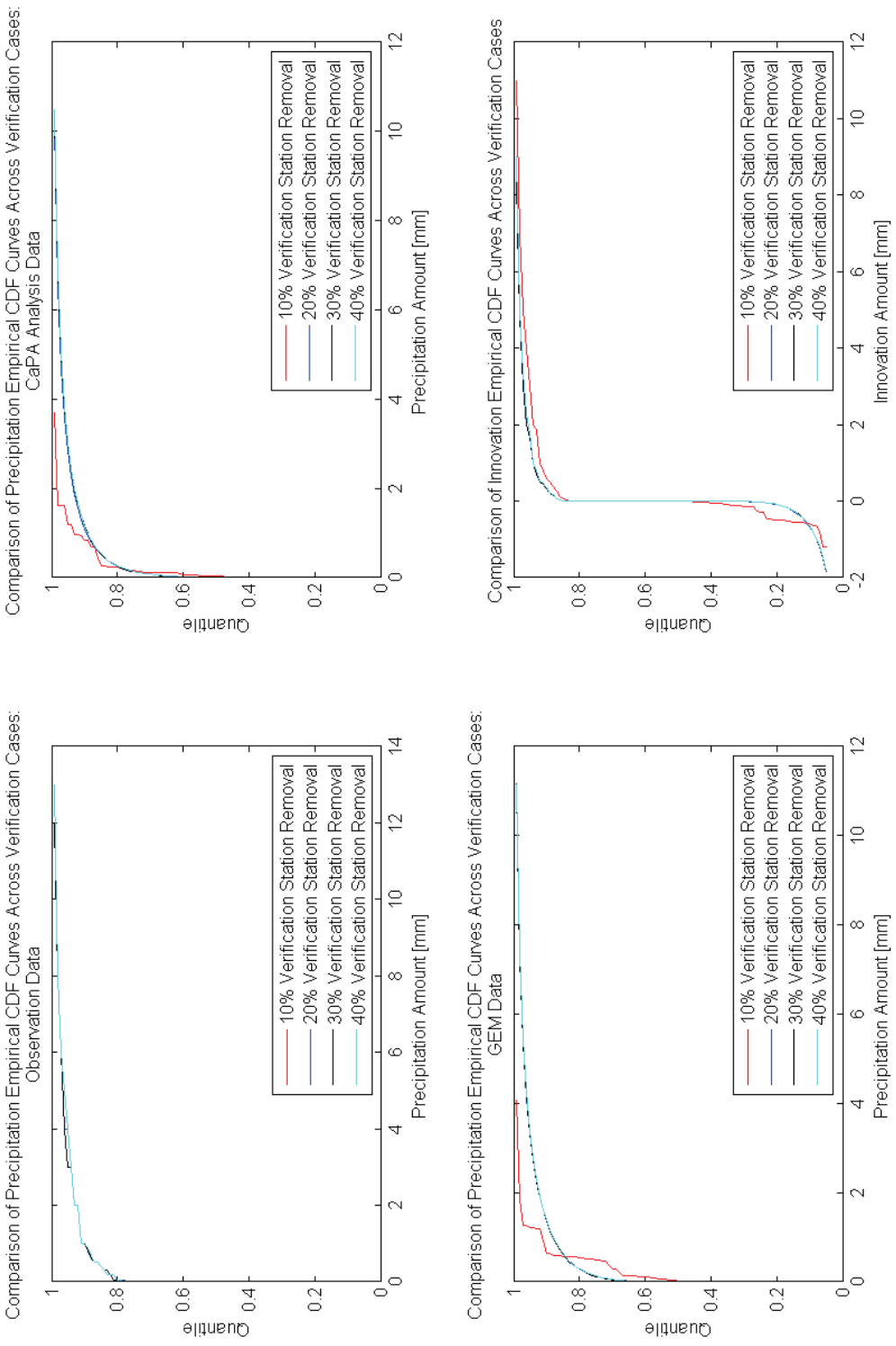

Figure 2.10: Precipitation quantiles associated with amount values for all verification cases. **Top Left:** Observation Verification Data **Top Right:** BCSI, or operational CaPA analysis data, **Bottom Left:** GEM data, **Bottom Right:** Innovation data.

CaPA analysis results, shown in the top right corner panel.

It is likely that the dissimilar empirical analysis CDF of the 10% verification case is the cause of the notably larger bias values witnessed in Table 2.2; it is not unreasonable to assume that the magnitude of the relative differences seen in the empirical analysis CDF of the 10% verification case is sufficient to produce the bias increases observed in Table 2.2. Furthermore, the deviation of the 10% verification case's CDF is such that an increase in the frequency of small to moderate precipitation amounts is experienced. It is plausible that having an increased amount of small to moderate precipitation values in the analysis data set can lead to increased positive bias results, as this would result in an overall under-prediction of total precipitation on behalf of CaPA.

Knowing that the large bias amounts reported for the 10% verification case are likely derived from idiosyncrasies associated with the distribution of its GEM input data provides only a partial solution to this problem. It is also important to investigate why this phenomenon is seen only at the 10% verification level, and yet not for the higher percentage cases. To do this, it is important to consider how the empirical CDF curves for the observation, GEM, and innovation data at the 10% verification level compare to the equivalent curves derived from the data used to generate the CaPA analyses, rather than the verification data. Figures 2.11, 2.12, and 2.13 depict empirical CDF curves for the data sets used to produce the BCSI analysis.

As can be seen from these three figures, in all cases, including for the GEM data,

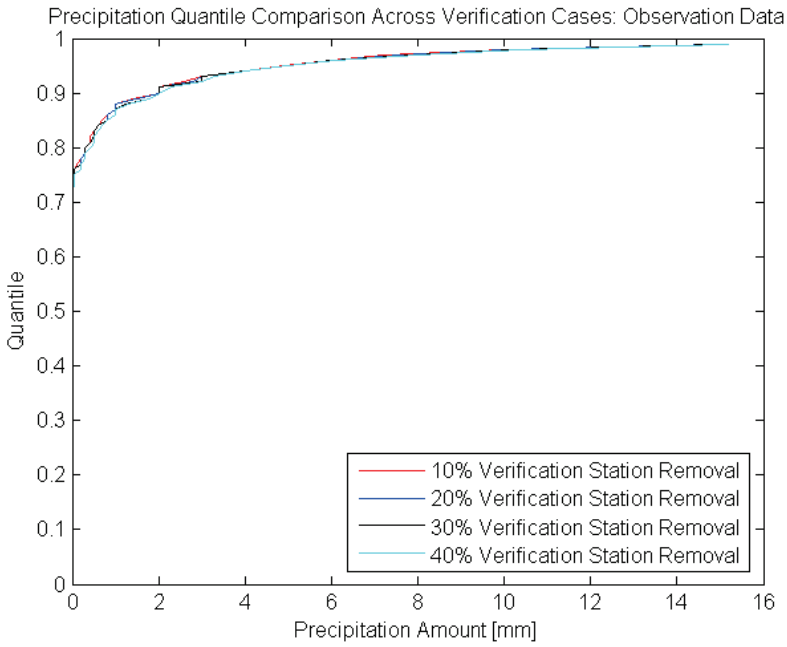

Figure 2.11: Precipitation quantiles associated with amount values for the observation data set used to create input for the production of CaPA analyses

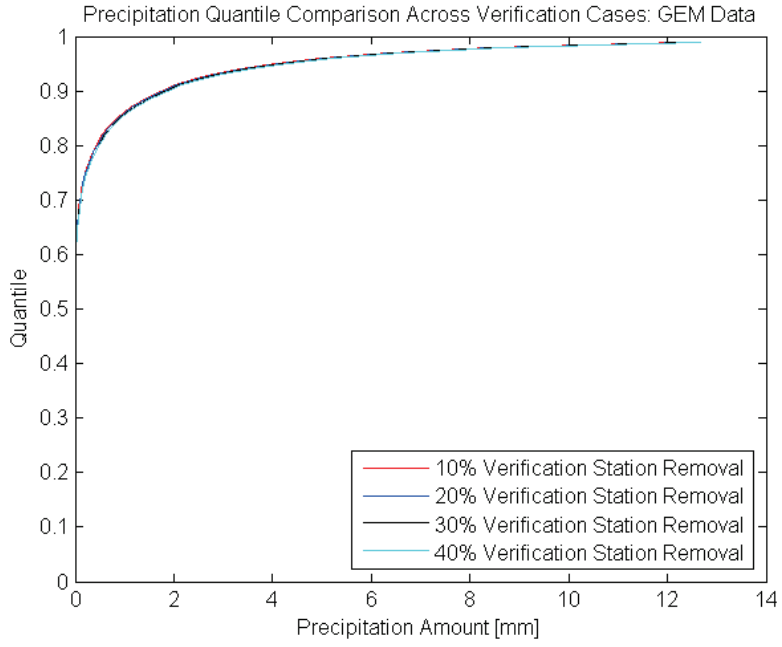

Figure 2.12: Precipitation quantiles associated with amount values for the GEM data set used to create input for the production of CaPA analyses

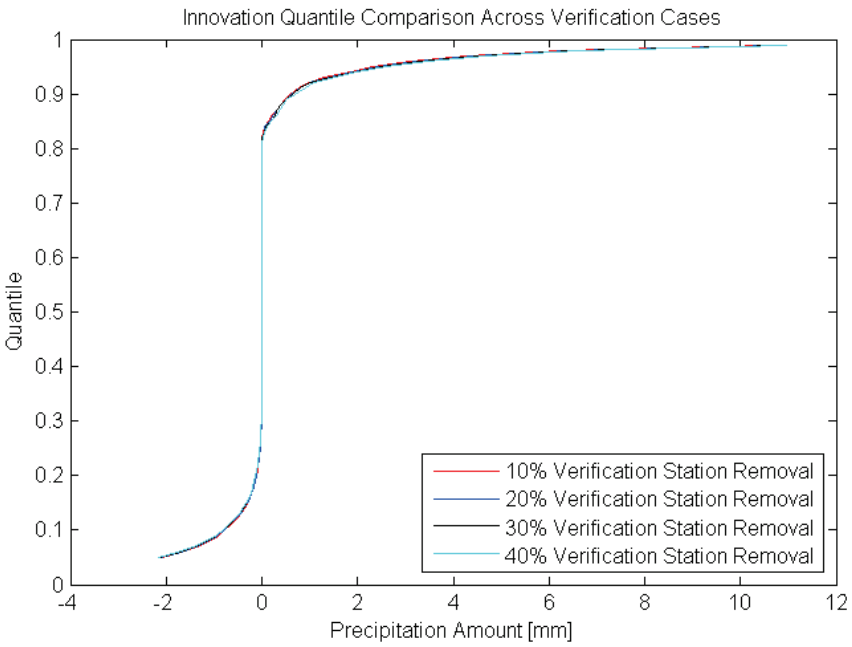

Figure 2.13: Precipitation quantiles associated with amount values for the innovation data set used to drive the production of CaPA analyses

the empirical CDFs are found to be in strong agreement across all verification cases. Since this agreement is seen in the GEM empirical CDFs for the run data sets as well as the 20%, 30%, and 40% verification cases, it is reasonable to deduce that the GEM CDF at the 10% level is indeed the irregularity. The cause of this anomalous 10% CDF can be most plausibly related to the verification sampling, with the selected station subset for the 10% level, purely by chance, capturing a sub-population of GEM data that does not adhere to the distribution seen when the entire sample population of the study is used.

This explanation is further supported by the fact that this CDF anomaly occurred at the 10% level, where the fewest number of stations are used for verification. This

makes it relatively more likely that the verification stations do not faithfully represent the distributional and statistical characteristics of the study's full sample population. This risk is reduced when a greater number of verification stations is used, and indeed this issue does not present itself for the larger verification cases.

As a general note for Figure 2.10, the overall shape of the innovation CDFs does not possess the characteristic curve witnessed in the other three panels. This is due to the nature of the innovation procedure, which generally results in an approximately more normally distributed data set.

It should also be noted that the above discussion focused on the BCSI CaPA analysis. As the distribution of the 10% GEM verification data was identified as the source of the increased bias, it can be safely stated that this reasoning applies to the BCSIc, BCSIcc, and SI cases as well.

Finally, it is also beneficial to provide a sense of context to the averaged bias values reported in Table 2.2. This can be accomplished by comparing the total precipitation experienced across the entire study domain and all time periods to the amount of bias incurred for the same spatial and temporal extents. This total amount of bias is calculated as the average reported bias multiplied by the total number of time periods from all verification stations. The results of performing this exercise are expressed in Table 2.3. The table clearly illustrates the substantial magnitude of the bias increase associated with the 10% verification case, as well as a general larger negative bias associated with the 30% case.

Table 2.3: Bias percentage of total precipitation within the study domain. SI refers to the no Box-Cox case, BCSI refers to the Box-Cox case and represents the current operational configuration, BCSIc denotes the corrected Box-Cox case using MWA scheme II, and BCSIcc refers to the corrected Box-Cox run using MWA scheme I

Run	10% Verif.	20% Verif.	30% Verif.	40% Verif.
SI	43.91%	-0.03%	-2.67%	0.52%
BC	56.41%	12.21%	7.97%	11.36%
BCc	48.38%	-4.08%	-7.46%	-3.78%
BCcc	47.59%	-4.61%	-8.18%	-4.47%
GEM	45.70%	-0.30%	-0.72%	2.74%

2.4.2 The Absence of Improvement in the Absolute Average Bias for the 30% Verification Case

A second feature of interest that necessitates further rumination is the comparatively minute improvement to the absolute value of the bias seen in the 30% verification case. In order to understand why such little improvement was seen as compared to the other percent verification cases, the issue of interest becomes the quality of the MWA data set produced under the 30% verification scenario. The previous investigation into the empirical CDF curves of the input data across all verification cases confirms that a quirk in the input data distributions is not the root of the poor improvement seen from the MWA method in the 30% case. Therefore, analyzing the quality of the MWA inputs specifically becomes a logical place to look for any departures from the other verification cases.

A basic proxy for the quality of a set of MWA inputs is the number of total stations used to create the MWA input data set initially, and derived from this is the number of total records across the full extents of the study both spatially and

temporally. The creation laws for the MWA data dictate that a MWA record is created only if all four 6 hour time periods required to make up one 24 hour daily period are present, and that at least 20 out of the 30 days possess values in a given window. This creates the potential for quality differentials to be seen between varying input data sets, depending on their temporal coverages. Unfortunately, the issue of CaPA input quality is not linked solely to the temporal coverage of the observation inputs; the often troublesomely low spatial distribution of observation stations in Canada, particularly in the northern portions of the country, is also a contributing factor in the MWA input quality question.

Both temporal and spatial input resolution unite to directly impact the caliber of the final CaPA analysis; if an MWA CaPA analysis is constructed based on inputs suffering from low resolution in either regard, then that analysis will have to be produced with a greater amount of reliance on the GEM background field, as opposed to the preferred option of a combination of GEM and observation data. Under such circumstances, the produced MWA analysis will not be able to perform optimally as a correction and scaling tool for a standard analysis, and may lead to issues such as the increased bias seen in the 30% verification case.

Figure 2.14 begins to explore the issue of the spatial distribution of the input MWA data by considering the number of observation stations contributing to the creation of MWA values for each day in the study's time period. As can be seen from the figure, as the percentage of stations extracted for use in the verification

Table 2.4: Total number of contributing stations for the entire study domain and time period for all verification cases

Verif. Case	Total Number of Contributing Stations
10% Verif.	27579
20% Verif.	24800
30% Verif.	21782
40% Verif.	18601

process increases, the number of stations contributing to a given MWA value calculation decreases. This trend holds without fail for the 30% case, indicating that no unexpected sampling consequences have manifested themselves to hinder the 30% verification case's MWA data fabrication.

A further confirmation that the number of contributing records did not likely play a role in reducing the efficacy of the MWA method in the 30% case is provided by considering the total number of data records contributing to the MWA data set's creation throughout the entire study. The total number of contributing records must be considered, as the number of records attributed to various observation stations fluctuates. Thus, even if the number of stations employed in the MWA data synthesis in the 30% case seems reasonable, that does not necessarily mean that these stations will possess lengthy records. The relevant data is provided in Table 2.4, and once again demonstrates that the number of contributing records for the 30% case is not sufficiently low so as to arouse suspicions. In fact, the number of contributing records for the 40% case is substantially lower.

Given that the amount of data being used to create the MWA input values does

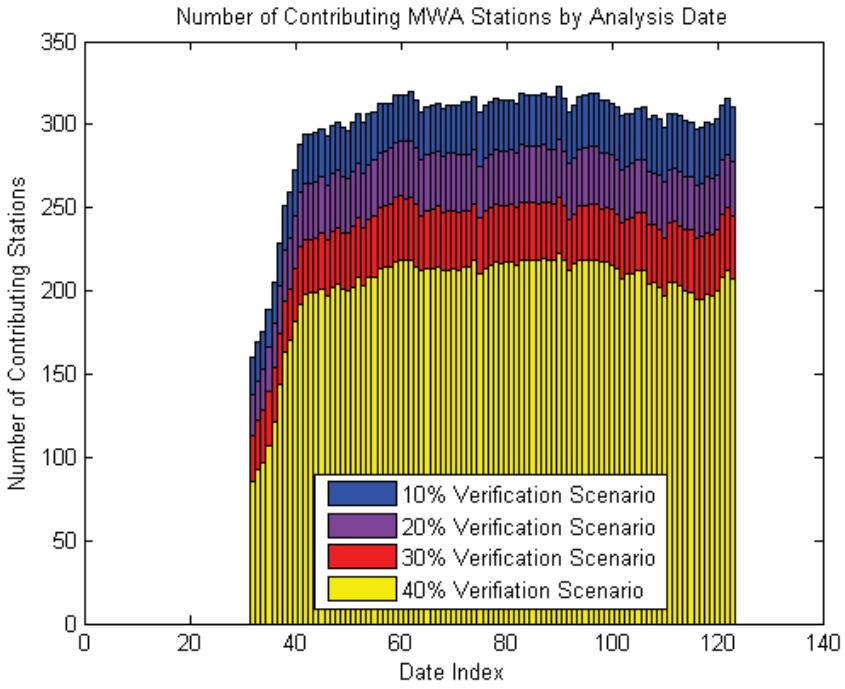

Figure 2.14: Impact of the verification scenario on the number of contributing MWA stations to each MWA CaPA analysis. Note that the Date Index refers to the index of each date in the study period, beginning May 1^{st}, 2010, and ending August 31^{st}, 2010

not appear to be unexpectedly low, the distances between the verification stations and the MWA input locations must also be considered. If the distances between the verification station and the MWA input stations is large, it is expected that use of the MWA analysis for scaling will have less positive benefits as opposed to if the verification station was in close proximity to a MWA location; closer proximity to an MWA input location reduces the influence of the GEM background in the scaling of the standard analysis and enhances the effectiveness of the MWA method. Table 2.5 presents minimum, maximum, and mean distances between verification station locations and the closest corresponding MWA input locations. Note that these table values were not constructed on a per time step basis, and look at the distances associated with the closest MWA station location for each verification location for all time steps.

As can be noted from the table, a distinct trend that pairs increasing the percentage of stations removed for verification with increased distance values can be seen. It can also be noted that the minimum and maximum distance values for the 30% and 40% cases are identical. This similarity may indicate that some type of a plateauing effect occurs, in regards to the distances that exist between verification and MWA stations, based on the specific layout of the stations in the study. Based on Table 2.5, however, there is no evidence that the distances between verification and MWA input locations for the 30% case are greater than should be expected given the established trend, and that these distances contributed to producing a less

Table 2.5: Minimum, maximum, and mean distances between verification observation stations and the closest lying MWA input station location

Verif. Case	Min Distance	Max Distance	Mean Distance
10% Verif.	0.02	2.54	0.98
20% Verif.	0.02	5.42	1.31
30% Verif.	0.01	7.24	1.57
40% Verif.	0.01	7.24	1.66

effective MWA scaling tool and ultimately a corrected analysis with more absolute bias.

The previous proxy measures for the MWA input quality failed to identify the cause of any potential input data quality shortfalls, but that is not to say that such shortfalls do not exist. Table 2.6 details the mean bias calculated by comparing the MWA CaPA analysis interpolated to the observation verification station locations, to MWA records of the verification observation data. This comparison is meant to highlight the base bias present in the MWA analysis used for scaling, which speaks to the influence of MWA input quality as all other bias incurring factors remain relatively constant across the verification cases. A quick review of the data listed in Table 2.6 confirms that the quality of the MWA inputs in the 30% case is the source of the increased bias.

Table 2.6 also finds the bias in the 10% verification case to be largest calculated. This supports the conclusion that the sampling of the 10% verification grid leads to irregular distributional characteristics in the GEM verification data which ultimately translates into increased bias; the increased bias in the 10% case is seen in the target analysis as well as the standard and corrected analyses, implying that it is not a

Table 2.6: Mean biases of the MWA observation data across all verification cases

Verif. Case	Mean Bias
10% Verif.	-0.61
20% Verif.	-0.49
30% Verif.	-0.58
40% Verif.	-0.47

run-specific phenomenon but rather verification related. This effect overrides the otherwise expected trend of an increasing negative bias in the target analysis as a greater number of stations are removed for verifications purposes.

Despite possessing the largest negative bias reported, the 10% case was the only scenario where scaling did not produce a negative final analysis bias. This is likely due to the much larger positive standard analysis bias that existed in the 10% case originally. However, the 30% case shows the next largest negative bias for the MWA analysis, while the 20% and 40% cases are relatively similar. This increased negative bias in the 30% case is likely the source of the relatively larger negative bias in its corrected analysis, resulting in essentially negligible absolute bias reduction.

A note of caution can be extracted from the investigation into both the 10% and 30% final bias anomalies. While the MWA scaling method produced the anticipated bias reduction results in the 20% and 40% verification cases, these two scenarios appear to represent more optimal conditions regarding the accuracy and spatial and temporal resolution of the input data. Thus, in order to ensure positive results from the implementation of the MWA method, care should be taken to ensure that good quality input data is being fed to the CaPA program; poor quality input data can

be expected to render any bias correction methodology ineffectual.

2.4.3 Choice of the 20% Verification Case for Detailed Analysis

Based on the discussion presented above regarding peculiarities associated with the 10% and 30% verification cases, these scenarios were not selected for a more detailed analysis of continuous and categorical skill scores. Both the 20% and 40% verification scenarios are comparatively more representative of the verification scores that would be seen if issues such as verification location sampling were not at play. However, the 40% verification case does require the removal of a greater number of potential input observation stations, which may still create a slight skew in the verification scores, even if it is not as blatant as the impacts seen in the 10% and 30% cases. Thus, the 20% case was selected for further investigation.

2.4.4 The Negative Bias of the MWA Corrected Analysis

While the discussion above addresses some of the unexpected characteristics of the bias results provided in Table 2.2, the overall trend of the corrections imparted by the MWA methodology should also be addressed. In all cases, applying the MWA method causes a reduction in average bias. In the 20%, 30%, and 40% cases, the initial operational bias value is small enough, and the reduction in the average bias is large enough, to yield a final bias value that is negative. This suggests that the

MWA method may be introducing negative bias sources, in addition to correcting the pre-existing data transformation bias.

This basic trend can be explained by looking primarily at the construction mechanism for the MWA input values. When the MWA inputs are created, only 20 out of the 30 window days are required to possess values, and the average of these value are then used to produce the MWA input. This was done in order to combat issues regarding the temporal resolution of the observation data, as strictly requiring 30 out of 30 days of record would simply leave too few qualifying stations per time period to produce a trustworthy MWA CaPA analysis. The reliance on the GEM background data would then override the influence of the MWA inputs at too many locations to be able to confidently use the MWA analysis as a correction tool.

Demanding only 20 out of 30 days of record to produce the averaged value for a given moving window requires the assumption that this averaged value is adequately representative of the 30-day window's true average. This may not be a valid assumption, as missing values from the record have the potential to lift up or drag down the average value. As can be seen from Figure B.9, the potential exists for either the under or over-prediction of a MWA value based upon the mandatory use of only 20 days of record per window; given that a tendency for a negative bias is witnessed in the results, it is plausible that in the specific case of this study, a tendency for over-predicting MWA values exists. This is supported by the results presented in Table 2.6, which clearly indicate a negative bias in the interpolated MWA analysis

values across all verification cases.

While Figure B.9 does not suggest that either over or under-prediction should dominate, it is important to note that this figure was created based on random sampling from those 30-day windows that were identified as having a complete set of daily records. For stations and time periods without full 30-day windows, it is possible that sampling biases that were not captured in the figure may come into play, producing a tendency for the use of the 20-day minimum to result in the overestimation of precipitation; this would be the case if lower precipitation values were not captured appropriately, or with the needed consistency. Given the comparatively lower frequency of extreme precipitation events, it is reasonable to expect missing days of records to correspond to zero or small precipitation events in many cases.

The lower spatial resolution of MWA inputs as compared to a standard analysis also contributes to the negative bias incurred by using the MWA scaling technique. A lower spatial resolution means that an increased reliance on the GEM background data is needed at more locations across the domain. As can be seen from Figure 2.6, GEM itself tends to over-predict the occurrence of smaller precipitation amounts, as can be verified by Figure 2.7. This leads to a slight negative average bias associated with the GEM data set that is incorporated into the MWA scaling analysis, and thereby also into the final corrected analysis. Good quality spatial and temporal input data is therefore key, and the availability of more complete records in time and

space for the observation data will likely assist in further improving the performance of the MWA scaling technique.

GEM's over-prediction of precipitation amounts, particularly those ranging from 0.2 to 2 mm, potentially plays a much larger role than simply in the context of an over-expressed background field in a observation-depleted CaPA environment. The tendency for GEM to produce too many smaller precipitation events creates a problem when calculating innovation values under MWA conditions. The GEM MWA values tends to be dominated by the more prevalent smaller precipitation amounts, generally producing MWA values that do not adequately account for those situations where GEM values are found to be larger than the corresponding observation values. This effect culminates in systematically positive innovation values that continually adjust the background field to increase the precipitation amounts in the analysis, providing little flexibility for instances where GEM may in fact predict larger precipitation values than the observation data set on a 6 hourly basis. This likely contributes to the artificial increase in the MWA analysis values, which in turn results in negative biases in the corrected analyses.

Support for the claim that the nature of the GEM data is the primary cause for the inflated MWA analysis precipitation amounts is supplied by Figures 2.15, 2.16, and 2.17. Figure 2.15 illustrates the difference in the empirical CDF curves for the MWA and 6 hourly innovation data sets, for the entire study domain and time period. For the majority of the CDF curve extents, until approximately the 0.95

quantile, the 6 hourly innovation CDF curve lies to the left of the MWA CDF curve, indicating that 6 hourly innovation data sets can be expected to possess smaller values than those belonging to the MWA innovation data set, except in cases where particularly large positive 6 hourly innovation values are encountered.

This concept is expressed graphically in Figures 2.16 and 2.17, where examples are shown of how 6 hourly innovation data compares to its corresponding MWA innovation value, for the 30 day period for which the MWA innovation amount was calculated. Figure 2.16 presents the window case for the smallest MWA value of the observation data, while Figure 2.17 corresponds to the window for the largest MWA value. These two figures were selected to give an idea of the influence that the magnitude of the MWA innovation has on the overestimation of innovation values as compared to the 6 hourly innovation results. It is clearly demonstrated by the figures that the MWA innovation value under-represents innovation cases where the GEM value is greater than the observed value, and that while this effect is magnified when the observation MWA value is itself large, it also persists in a notable fashion when the observation MWA value is small, despite producing an overall MWA innovation value that is slightly negative in this instance.

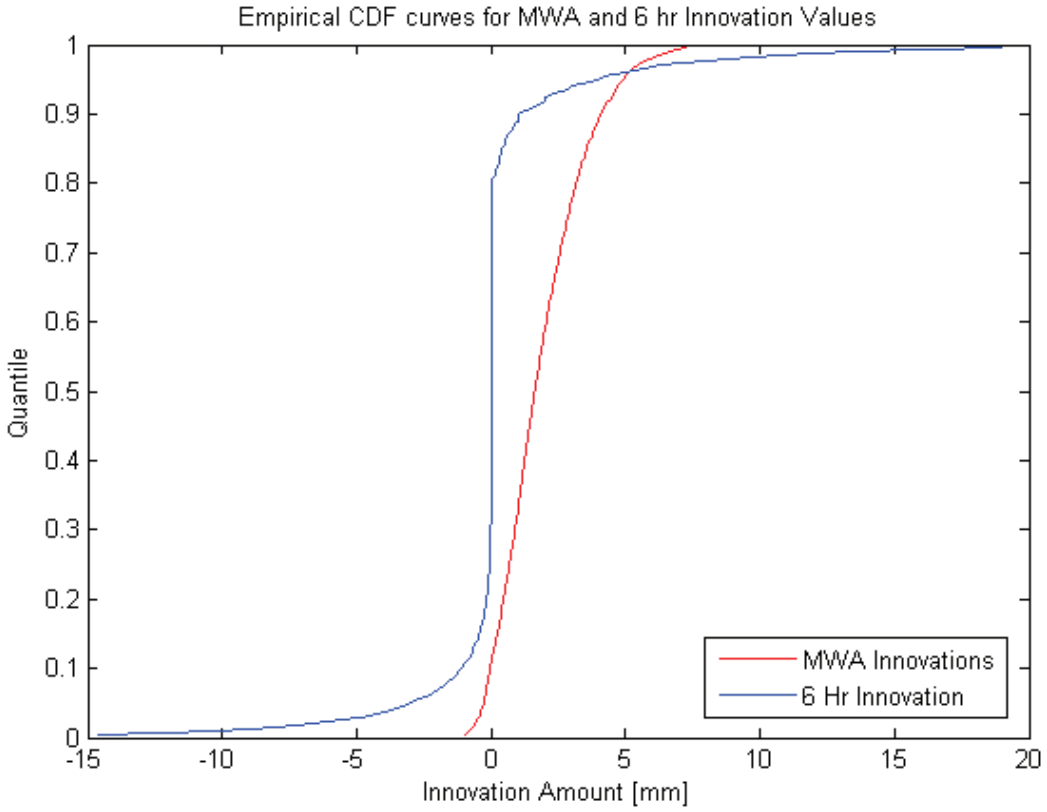

Figure 2.15: Empirical CDF curves for MWA and 6 hourly innovation data for the entire study domain and time period

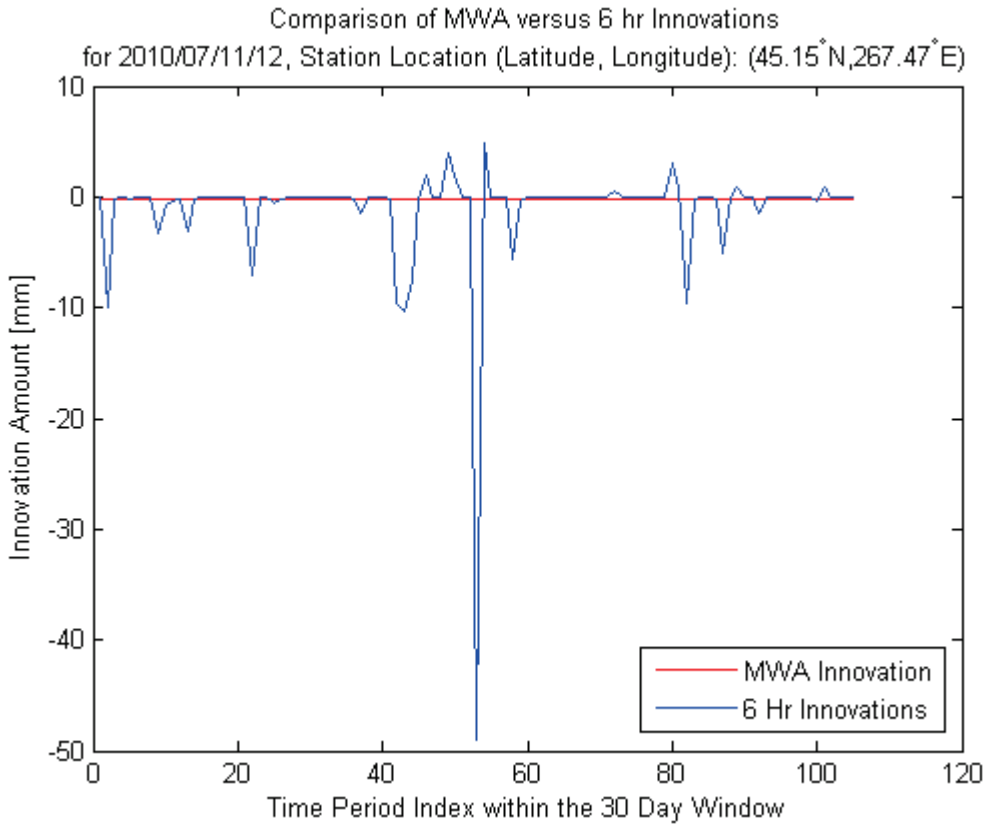

Figure 2.16: Station-specific comparison of the MWA innovation value to the 6 hourly innovation amounts that occur in the corresponding 30-day window. MWA data selection is based on the smallest observation MWA value for the chosen station.

Figure 2.17: Station-specific comparison of the MWA innovation value to the 6 hourly innovation amounts that occur in the corresponding 30-day window. MWA data selection is based on the largest observation MWA value for the chosen station.

2.4.5 Performance Comparisons and Identification of the Optimal CaPA Configuration

In terms of Table 2.2, MWA scheme II is shown to outperforms scheme I as well as the operation scenario in all cases, except under the 10% verification conditions. This is likely related to the previously discussed issue of the dissimilar distribution of the 10% verification GEM data set, which invites increased bias. If MWA scheme II is applied in this case, due to the differing distribution and overall increased positive biases, the use of the analysis errors to focus the scaling of the standard analysis may result in over-emphasizing the accuracy of low uncertainty analysis values. Under such conditions, MWA scheme I performs better, as it scales all analysis values without discriminating based on the analysis errors; MWA scheme I may serve as the better option in situations where the use of the analysis errors too narrowly limits the scaling focus based on issues related to the quality of the input data. However, it is important to note that even under circumstances where input data quality may not be ideal, the MWA methods still reduce bias to lower levels than the current operational configuration.

While Table 2.2 and Figure 2.5 represent a good means of communicating bias results for the entire set of approximately 35352 verification results within the study domain and throughout the study time period, averaging across all of these results can also be expected to mask more detailed patterns within the bias field. This issue is addressed in part in Figure 2.6.

The DPM results showcased in Figure 2.6 allow for a glimpse into how the average bias values depend on the magnitude of precipitation events. It is clear that while MWA scheme II produces a more desirable final average bias result than scheme I or the operational case, for lower to moderate precipitation amounts the operational case produces a DPM value closer to 0 than either MWA scheme. This suggest that the true prowess of the MWA schemes lies in their handling of larger precipitation amounts, as the operational analysis shows an overall under-prediction of large precipitation events. This may in part be due to the nature of the Box-Cox transformation, as it has been found that in a kriging with external drift setup, Box-Cox transformation with λ values that are not too low, for example greater than 0.1, result in precipitation estimates that are reduced in magnitude compared to the input data, though they may retain much of the more detailed structural characteristics (*Erdin and Frei*, 2011).

Figure 2.6 confirms the the operational case struggles to correctly handle large precipitation amounts, showing that the operational analysis results in too few occurrences of precipitation events greater than 2 mm in magnitude. While both MWA schemes also suffer the same affliction, the degree to which they under-predict the occurrence of events within this magnitude range is less severe. Generally the two schemes perform very similarly in this regard. For the 25 to 50 mm range, MWA scheme I produces results that appear distinctly better than those for scheme II in the plot, but referring to the bootstrapping results in the lower panel of the figure

confirms that the statistical difference between the performance of the two schemes even in this precipitation magnitude range is not significant.

An examination of the DPS scores in Figure E.1 also supports the choice of MWA scheme II over the current operational configuration of CaPA, for at least large precipitation events. Here, the standard deviation of the precipitation data is more faithfully represented by both of the MWA schemes than it is by the operational configuration of CaPA, with a statistically significant improvement suggested for moderate to large precipitation thresholds. For the largest precipitation threshold only, the MWA schemes also significantly outperform the SI case. However, it is not possible to statistically distinguish between the performance of the two MWA cases in regards to the DPS scores calculated, based on the bootstrapping results presented. It is important to note that DPS is a key score to consider, as the variability of precipitation data is crucial for truly capturing the statistical character of the precipitation data, and for producing output that can be reliably used in any form of modelling sensitive to hydrological inputs. For example, accurately capturing precipitation variability can positively impact the estimation of hydrological responses, such as extreme flood magnitude quantiles (*Leander and Buishand*, 2006).

The two MWA schemes also perform in essence identically when considering the ETS and aETS scores. On both cases, the operational configuration of CaPA produces better scores than do either of the MWA schemes, with this dominance becoming more pronounced for the moderate to large precipitation bins. This result

is somewhat damaging for the MWA schemes, as the ETS score is weighted heavily in terms of evaluating overall performance. For the moderate to large precipitation bins, the SI case also results in higher ETS and aETS scores than do the MWA corrected analyses.

An examination of the the suite of results presented in Section 2.3 confirms that in terms of bias, DPM, DPS, and the upper bin ranges of the FBI scores, the SI CaPA setup results in a more desirable analysis than the BCSI, or operational CaPA case. This is important to note, as the MWA technique is intended to essentially create a hybrid between the SI and BCSI analyses, bringing the positive characteristics of the SI configuration into the results of a MWA corrected analysis, while attempting to still retain the ETS improvements associated with the BCSI case. Unfortunately, the ETS and aETS scores of the MWA approaches for most bins prove to be either statistically equivalent to or worse than even those for the SI case.

Overall, the SI case outperformed the MWA schemes except in regard to the DPS and FBI scores for certain bins and threshold, and in the aETS for the 0.2 to 1 mm, 1 to 2 mm, and 25 to 50 mm bins, although perhaps not significantly. However, the MWA schemes showed improvements over the operational configuration for moderate to large precipitation thresholds for most metrics, except for the ETS and aETS. When these results are weighed in combination with the strong quantile-quantile performance of the MWA schemes for larger precipitation amounts, and the overall bias reduction capabilities of MWA scheme II, it becomes clear that MWA scheme

II is viable for future use in an experimental version of CaPA. Appropriate applications for MWA scheme II may include any activities where dealing with hydrological extremes are a priority, such as in flood forecasting or hydroelectric utility operation planning.

Chapter 3

Improvement of Semivariogram Estimation and Representation within CaPA

Semivariogram estimation is a critical component of the SI methodology used by CaPA, and the quality of the semivariogram model directly impacts how successful a CaPA analysis will be in accurately portraying the precipitation field. Yet the semivariogram has a somewhat daunting task. Precipitation is one of the trickiest variables in the hydrological world to characterize in terms of its spatial and temporal structure, due to the significant amount of variability precipitation shows in both of these dimensions (*Yavuz and Erdogan*, 2012). Regardless, an attempt is made in this investigation to improve the semivariogram specification in CaPA. After a

presentation of the relevant background information, the methods developed to accomplish this task are presented. These methods include the use of anisotropy and accommodation of convection. A review of the results obtained through this study, and a discussion on the implications of these results, are also presented.

3.1 Background

One of the primary draws for CaPA analyses is the presentation of results on a spatially regular grid. The gridded nature of a CaPA analysis is the key to its functionality as a source of precipitation input data for hydrological and other model types. CaPA owes its gridded output to the SI process, which uses GEM data to create the basic gridded framework. The SI method, in turn, also relies heavily on certain tools to make the regular spacing of CaPA output possible; without a semivariogram to describe spatial correlations in the innovation data set, the prized spatial structure of CaPA analyses would not materialize.

The semivariogram, therefore, is an indispensable contributor to the SI procedure. As such, its accuracy is an issue of great concern. A poorly estimated semivariogram function may still allow the SI procedure to execute smoothly, but the inaccuracies in the semivariogram can be expected to transfer directly into errors in the final analysis results. A semivariogram which reliably reflects the true spatial structure of the innovation data will instead enhance CaPA analyses. Therefore, as a basic step in pursuing an improved CaPA model, the best possible semivariogram specification

technique must be sought.

Improvements to the semivariogram are not necessarily anchored solely in the mathematical characteristics of the semivariogram model selected, however. It is also important to show due deference to how the mathematical attributes of the semivariogram interact and capture the physical nature of the data set being predicted. Given that precipitation data is the focus in CaPA, the impact of different types of weather systems on the spatial structure of the innovation data must also be contemplated.

3.1.1 Semivariogram Estimation and Types

The ability to accurately specify a semivariogram model must be cultivated. This includes the estimation of an experimental semivariogram. After this is accomplished, the selection of an appropriate theoretical semivariogram model must follow. Regardless of the type of theoretical model selected, a decision between an isotropic or an anisotropic configuration is then required. It is this basic process that is outlined below, along with a brief summary of alternate improvement techniques and a reiteration of the importance of enhancing semivariogram performance in CaPA.

Estimation of the Experimental Semivariogram

Semivariograms are flexible tools with utility beyond precipitation interpolation applications. Semivariograms and geostatistical analysis in general have their roots in geological applications, but beyond this semivariograms have also been used in biol-

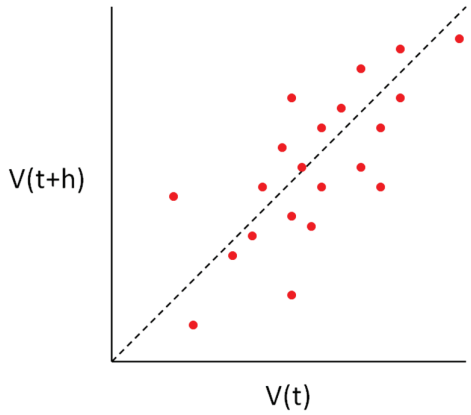

Figure 3.1: Hypothetical example of a h-scatterplot. $V(t)$ represents a data value at t, while $V(t+h)$ denotes a data value at a distance h away. The dashed line is the 45° line

ogy, ecology, economics, the atmospheric sciences (*Yu et al.*, 2007), and in alternative hydrological applications, such as to quantify land surface heterogeneity from remote sensing observations in the red and near infrared spectrum (*Garrigues et al.*, 2008). In general, the semivariogram is used to describe the spatial relationship between values in a data set. In order to understand what the values of a semivariogram mean, a h-scatterplot can first be considered. A h-scatterplot is created by plotting all data pairs from a data set that can be obtained by taking an initial data value, $V(t)$, and a value that exists at a distance h away, $V(t+h)$. Thus each h-scatterplot is specific to a distance, h. An example of a h-scatterplot is shown in Figure 3.1.

If a series of h-scatterplots are created for different values of h, then for each plot a moment of inertia value can be calculated. This moment of inertia value is taken with respect to a 45° line drawn through the h-scatterplot, and is given by the

following equation (*Isaaks and Srivastava*, 1989):

$$MOI_h = \frac{1}{2n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3.1}$$

where MOI_h is the moment of inertia value for the distance h, n is the number of points plotted on the h-scatterplot, and x_i and y_i are the x and y coordinates, respectively, of the point i. This function between h and the moment of inertia value is known as the semivariogram. If correlation coefficients or covariances are calculated instead, then the result is a correlogram or covariance function, respectively.

Speed and efficiency in data processing is always desirable. Thus it is important to find a means of approximating the experimental semivariogram directly from a data set. This is done by using the following version of the semivariogram equation (*Isaaks and Srivastava*, 1989):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij}=h} (v_i - v_j)^2 \tag{3.2}$$

where $\gamma(h)$ denotes the semivariogram value for the distance h, v_i and v_j represents two data points that form a pair at the requisite separation distance, and $N(h)$ is the number of pairs that meet the separation distance requirement. A concern related to the value of $N(h)$ quickly becomes apparent; in an average data set, it is not likely that the data points will be regularly placed in space, and thus the number of pairs likely to be exactly a distance h away from each other can be expected to be very small, if not zero for many distances. Therefore, the concept of tolerance must be

adopted, and h must become an approximate distance. The tolerance on the h value should be selected based on the spatial characteristics of the data set (*Isaaks and Srivastava*, 1989).

The tolerance on h is often selected in conjunction with the lag increment, or the spacing of h values for which the experimental semivariogram is determined. Both parameters can be adjusted to ensure that a clear semivariogram structure is achieved. The lag increment is often selected as either the value of the grid spacing, or the approximate average distance between data measurements (*Isaaks and Srivastava*, 1989). Once the lag spacing is determined, the tolerance on h is commonly set to half of the value of the lag (*Isaaks and Srivastava*, 1989).

In order to understand the anatomy of an estimated semivariogram function, Figure 3.2 has been provided with all of the features of interest clearly labeled. This example also serves as a guideline for what a clear semivariogram structure is. The *range* refers to the separation distance at which a leveling-off of the semivariogram values occurs, indicating roughly the distance beyond which no significant correlation between values exists. The semivariogram value associated with the range distance is known as the *sill*. Finally, the *nugget* is used to account for semivariogram values at extremely small separation distances. Theoretically, the semivariogram value at a distance of zero must itself be zero, but minuscule separation distances can result in a non-zero value estimated for the y-intercept of the semivariogram plot. Under such conditions, the semivariogram value at a separation distance of zero is known

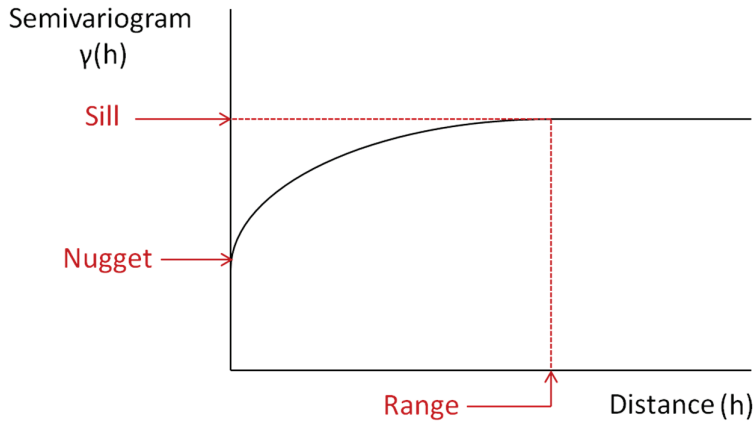

Figure 3.2: Anatomy of a semivariogram function, with the sill, nugget, and range values highlighted

as a nugget. A nugget may arise due to the variability of the data set that naturally occurs at small scales, or due to measurement errors (*Isaaks and Srivastava*, 1989). A simple test to determine if a nugget is required is provided in Equation 3.3, where if the inequality described is true as $h \to 0$, then a nugget exists with a value of c_o (*Yu et al.*, 2007):

$$\gamma(h) \to c_o > 0 \tag{3.3}$$

Introduction of Anisotropy into the Semivariogram

The difference between isotropic and anisotropic semivariograms during the experimental calculation phase can be parred down to the h variable. If h is considered to be a distance, with no angularity, then semivariogram values originating from all directions around the point of interest are pooled to form the isotropic estimate. If

h is considered to be a vector, then the direction in which the scalar component of h extends becomes useful as a means of pooling only semivariogram estimates with similar angles (*Garrigues et al.*, 2008); instead of including all those data pairs v_i and v_j that are a distance h apart, the semivariogram value $\gamma(h)$ becomes direction specific, and limits the pairs further to those separated by distance h in a given direction (*Isaaks and Srivastava*, 1989). For example, one anisotropic semivariogram value may be based on data pairs separated by 5 kilometers in a NE direction.

Under an anisotropic configuration, a tolerance range is also required for the search angle. To determine the tolerance on the search angle, caution must be exercised to ensure that the tolerance is not so restrictive that too few data pairs qualify and the resulting semivariogram structure is unclear. However, if too large a tolerance is selected, the anisotropy may be blurred (*Isaaks and Srivastava*, 1989). Thus a guess-and-check approach is often best (*Isaaks and Srivastava*, 1989). The h tolerance and lag increment may also be found to vary with direction (*Isaaks and Srivastava*, 1989).

It is not sufficient merely to elect to incorporate anisotropy in the semivariogram. The process must be taken a step further, to include a decision on the type of anisotropy to allow in the semivariogram model. Two basic forms of anisotropy can exist. If the sill parameter remains constant under directional analyses, and only the range value fluctuates, then the anisotropy type is said to be geometric. Conversely, if it is the range that is found to be invariant to the search direction

of the semivariogram, while the sill parameter shows directional dependence, then the anisotropy type is zonal (*Isaaks and Srivastava*, 1989; *Mateu et al.*, 2008). It is geometric anisotropy that is considered in this investigation.

When undertaking semivariogram estimation manually, a quick means of assessing the presence of anisotropy is by looking first to an isotropic semivariogram. If the isotropic semivariogram, which contains the maximum amount of data pairs per separation distance, appears to give a clear semivariogram structure, then it is logical to move towards anisotropic semivariograms. If the result is muddled and no clear relationship between distance and the semivariogram value is evident, then it is unlikely that allowing for anisotropy will provide an improvement (*Isaaks and Srivastava*, 1989). Once anisotropic semivariograms are created, a quick way to determine the nature of the anisotropy is by creating semivariogram contour maps, which may assist in the identification of the primary and secondary axes of anisotropy, under elliptical conditions (*Isaaks and Srivastava*, 1989).

Theoretical Semivariograms

While an experimental semivariogram is relatively simple to estimate, its usefulness is restricted by the limited set of distances for which semivariogram values are determined. To surmount this obstacle, a theoretical semivariogram can be fit to the experimental semivariogram values. In order to select an appropriate type of theoretical semivariogram, the statistical requirements of the candidate models must first be understood. For any ordinary kriging applications, it is required that the

covariance matrix derived from the semivariogram be positive definite. A positive definite covariance matrix results in only one unique solution to the kriging problem. The positive definite trait is attained if the following condition is met, where w_i and w_j denote arbitrary weights in a vector of length n assigned to data values i and j, and $C_{i,j}$ are the covariance values from the covariance matrix (*Isaaks and Srivastava*, 1989):

$$\sum_{i=0}^{n}\sum_{j=0}^{n} w_i w_j C_{i,j} > 0 \tag{3.4}$$

An alternative interpretation of this condition is that any random variable that is created from a weighted linear combination of other random variables must have a positive variance.

Using the positive definiteness criterion as a basis, a series of basic functions can be identified that are themselves positive definite, and which can be used as basic building blocks for theoretical semivariograms. Such functions can either be used independently, or in a weighted linear combination. Four basic positive definite semivariogram functions are (*Isaaks and Srivastava*, 1989):

The Nugget Effect Model

$$\gamma_0(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1 & \text{otherwise} \end{cases} \tag{3.5}$$

The Exponential Model

$$\gamma(h) = 1 - \exp\left(-\frac{3h}{a}\right) \tag{3.6}$$

where a is a rough approximation of the true range, defined as the distance value corresponding to 95% of the sill value (*Isaaks and Srivastava*, 1989).

The Spherical Model

$$\gamma(h) = \begin{cases} 1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 & \text{if} h \leq a \\ 1 \quad \text{otherwise} \end{cases} \tag{3.7}$$

where a denotes the range of the semivariogram (*Isaaks and Srivastava*, 1989).

The Gaussian Model

$$\gamma(h) = 1 - \exp\left(-\frac{3h^2}{a^2}\right) \tag{3.8}$$

where a is defined the same as in the exponential model case (*Isaaks and Srivastava*, 1989).

In order to accommodate anisotropy in the theoretical semivariogram, an elliptical model can be used. Such a model allows the range of the semivariogram to vary, depending on the direction. It is important to note that assuming elliptical geometric anisotropy is a hefty assumption. However, if anisotropy exists, it is anticipated that it will be detected by an elliptical model, even if that model is not the best depiction of the true anisotropy. Thus an elliptical anisotropic assumption provides a reasonable starting point to investigate the influence of anisotropy on the

semivariogram. Equations 3.9 to 3.12 define the elliptical semivariogram model as follows (*Trim*, 2004):

$$1 = \frac{x^2}{a^2} + \frac{y^2}{b^2} \tag{3.9}$$

where x and y represent Cartesian coordinates associated with a position on the ellipse, and a and b are parameters used to specify the major and minor axes of the ellipse (*Trim*, 2004). The anisotropic semivariogram information collected in this investigation involves only the direction angle, major and minor range, nugget, and sill values. Given that this is the only information available, the following adjustments can be made to develop an equation that calculates the range value using the ellipse, based on the direction angle:

$$x = r \times \sin\theta \tag{3.10}$$

$$y = r \times \cos\theta \tag{3.11}$$

$$r = \sqrt{\frac{1}{\left(\frac{\sin^2\theta}{a^2} + \frac{\cos^2\theta}{b^2}\right)}} \tag{3.12}$$

where r is used to denote the range value determined from the elliptical model of the semivariogram's anisotropy, and θ is the angle used to determine r.

Performance Evaluation of the Current Operational Semivariogram

In order to gauge the quality of the semivariograms currently operational in CaPA, a plot was generated comparing the true relative locations of a selected sub-set of observation stations against the locations of the same stations estimated using a semivariogram generated by CaPA. This was accomplished using a multi-dimensional scaling technique automated in Matlab by the built-in *cmdscale* function.

First, the sub-set of points was selected such that the largest number of locations that could be reasonably and clearly plotted in one window were captured. Once all of the locations were identified, the distance between each pair of locations was calculated.

Next, an experimental semivariogram was determined for the entire domain using a bin width of 10 km. The approximated semivariogram values associated with each distance in the station location sub-set were then pulled out. This collection of semivariogram values was then converted back into distance values by applying the inverse of the exponential semivariogram using the CaPA-determined parameters. Thus two distance vectors were created; one for the true distances between the sub-set of observation locations, and one for the semivariogram estimated distances. The *cmdscale* function was then used to transform each of these distance vectors into maps of the corresponding spatial arrangement of the observation locations. The results of this exercise are shown in Figure 3.3. This procedure was completed on a per time step basis.

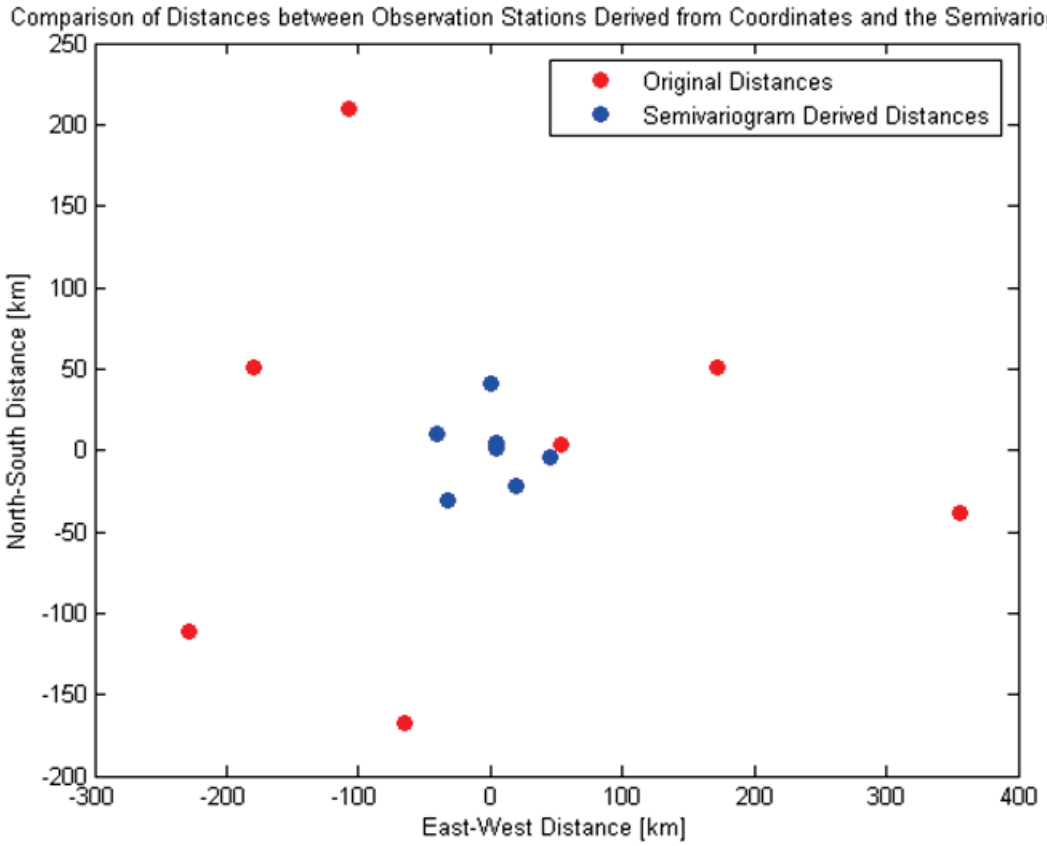

Figure 3.3: Visual evaluation of the current semivariogram performance in CaPA based on multi-dimensional scaling. The example shown uses the semivariogram estimated for the 2010/06/01/00:00Z time step

The advantage of using this approach, and comparing distance vectors rather than correlation values directly, is that an explicit visualization of the semivariogram performance is provided. A similar analysis could be done by directly comparing correlation values, but Figure 3.3 provides a performance gauge in terms of distance measurements, which can be more meaningfully interpreted by individuals not accustom to assessing statistical metrics.

By directly comparing the plotted spatial arrangements, an estimate of the performance of the theoretical semivariogram in CaPA emerges. If the theoretical semivariogram performed perfectly, the two sets of points would overlap. Differences in the plotted points reflect inadequacies in the current theoretical semivariogram. As can clearly be seen, significant shortcomings exist for the current exponential semivariograms being fit by CaPA, and an investigation into possible improvement strategies is well warranted. The most likely source of the semivariogram errors seen in Figure 3.3 is over-simplification in the specification of the theoretical semivariogram, which is a single, isotropic exponential semivariogram applied for all regions of the domain for each time step.

Alternate Semivariogram Modelling Approaches

Thinking beyond the initial question of anisotropy during semivariogram estimation brings up the possibility of other forms of semivariogram improvement. The best place to begin such a discussion is at the experimental estimation stage, as improving the experimental semivariogram estimates can only serve to further improve the quality of the theoretical semivariograms that are subsequently fit.

The traditional approach to estimating an experimental semivariogram involves employing the method-of-moments technique (*Yu et al.*, 2007). This is not the only tried and tested option available for defining a semivariogram, however. Robust estimators have an undeniable appeal, as they are well suited to handling data sets that do not conform to a normal distribution (*Cressie and Hawkins*, 1980). Robust

estimators for the semivariogram can be obtained through a variety of methods, whose details vary only in the practicalities of how the data is to be treated. Such approaches include the following (*Cressie and Hawkins*, 1980):

- Allowing the difference between data pairs, $(V(t+h) - V(t))^2$, to be represented as a single random variable, and conceptualizing the semivariogram as the expected value of that variable, which posses a chi-squared distribution with one degree of freedom

- Assuming stationarity of the mean in the data set and estimating the semivariogram through the variance of the $(V(t+h) - V(t))^2$ data set, which must possess a symmetrical distribution

- Assigning an autocorrelation function value for each value of h for which the semivariogram must be specified, and assessing the semivariogram estimate via the variance of the data set $V(t)$ multiplied by one minus the autocorrelation

The above approaches can be combined with the concepts of robust statistics to produce the final required values. For example, this can be accomplished through the use of trimmed means, which exclude extreme-valued order statistics, or M-estimators, depending on the approach followed (*Cressie and Hawkins*, 1980).

Another possible source of improvement that relates to the generation of the experimental semivariogram is a method that replaces the standard concept of distance in the semivariogram formulation with a nearest-neighbour based distance measure.

This method assumes stationarity in the semivariogram. The method requires the specification of a nearest-neighbour parameter, δ, which defines the acceptable range. The δ value is used in combination with an indicator function to limit the basis of the semivariogram value at h to only those data pairs where the difference $h- \parallel s_i - s_j \parallel$ is in the range $-\delta$ to δ, and s_i and s_j represent the data pair in question (*Yu et al.*, 2007). Improved estimation performance can then be obtained by building upon this nearest-neighbour concept, to allow the δ value to vary throughout the domain. Such a modification provides the needed flexibility to account for heteroskedastic conditions, and to improve estimation by allowing for adaptive estimation that can address the following potential localized problems: data availability and continuity with distance, non-monotonically increasing behaviour, and consistency in the smoothness of the semivariogram (*Yu et al.*, 2007). The variability of the δ parameter can be instituted in the semivariogram estimation framework through a function defining the relationship between δ and $\parallel s_i - s_j \parallel$, in combination with a non-parametric kernel function (*Yu et al.*, 2007).

Focusing now on the model setup for the semivariogram, another possibility for improvement stems from the concept of nonstationarity in the semivariogram. Nonstationarity may arise due to a number of localized factors that directly influence the generation of precipitation (*Chang et al.*, 2010). For example, in the study domain considered, this may include topographic influences in the West due to the Rocky Mountains. If nonstationarity is present and not accounted for, it is likely

that the performance of the semivariogram as it is applied across the domain will suffer (*Chang et al.*, 2010).

A solution to this problem is to employ nonstationary semivariogram functions. Such functions can be developed by similar means as nonstationary spatial covariance functions, which enjoy a variety of potential estimation methods which generally either require a series of realizations of the data set, or only a single realization given a sufficiently large data set describing the spatial relationships under review (*Chang et al.*, 2010). These methods range in their mathematical foundations from Empirical Orthogonal Functions and kernel-type methods to the strategic combinations of stationary processes and random coefficients with local basis function constituents (*Chang et al.*, 2010).

Another example of how CaPA may benefit from the use of nonstationarity involves the spatio-temporal semivariogram. A spatio-temporal semivariogram allows for spatial covariances to be linked with temporal covariances (*Mateu et al.*, 2008); given the evolution of weather patterns and precipitation that occur through time, considering the linked spatial and temporal covariance of innovation data may improve the performance of the resulting mathematical characterization.

Such nonstationary techniques could be applied to the CaPA program, to allow the semivariogram to be modelled differently at different locations throughout the domain. However, the limited availability of innovation data in many regions of the domain would likely require a moving-window approach to be adopted, with

each window extending back in time for each time step in order to develop the nonstationary spatial models. Thus, the inclusion of nonstationarity in CaPA would necessitate the addition of another modelling component to the program, in order to produce the desired series of geographically specific semivariograms. This represents a considerable increase in the amount of computations and data management that would be required by CaPA, and most importantly understood by the users of CaPA analyses. Therefore, jumping into the realm of nonstationary semivariograms at the current juncture is not a reasonable undertaking, without first at least attempting to improve the performance of the simpler, stationary semivariogram estimation method.

All of the above potential semivariogram improvements have been described only briefly, and are mentioned solely with the intent of highlighting the many options that exist. They represent only a sampling of what advancements in the field of variography might be able to offer to CaPA, although some techniques may be more feasible than others. The important thing to note, however, is that the method proposed in the following section is a simple, and logical first step in discovering the optimal semivariogram setup. Considering only anisotropy and convection also allows the investigation to avoid deviating too far from the operational setup, so as to confirm first and foremost that modifying the semivariogram will translate into detectable changes in the CaPA analysis and is worthwhile, and to provide a foundation for future investigations.

3.1.2 Convective Precipitation

The need to properly capture the spatial extents of convective systems in CaPA is a direct result of the importance that such systems can have for society. When a convective system develops, the speed of development and the precipitation intensities experienced can result in hydrological responses such as flash flooding and erosion problems, as well as agricultural impacts through damage to crops and potentially livestock. In the most severe cases, convective systems can even pose a threat to public safety (*Tadesse and Anagnostou*, 2009). Given these potentially severe consequences, and knowing that convective activity tends to be the most prevalent during the summer and fall (*Ruiz-Leo et al.*, 2013), it is imperative that proper consideration be given to the distinct physical nature of convective precipitation in CaPA during the warm season.

Under the current operational configuration of CaPA, no special consideration is given to convective systems. This makes it possible for stratiform precipitation to dominate the specification of the semivariogram, due to its larger scale and less variable nature. Within the SI methodology, this may lead to too much weight being given to innovation values located outside of convection zones for convective locations; the result would be the incorporation of inappropriate innovation values in the estimate of a convective innovation amount. By developing an understanding of how convective and stratiform precipitation differ from each other, and how they can be partitioned, it becomes possible to modify CaPA to directly consider convective

processes. If CaPA can handle convective systems realistically, it is hoped that the final CaPA analyses will show an increase in accuracy in regions experiencing convection.

The Basics of Stratiform and Convective Precipitation

Precipitation formation, whether it be through stratiform or convective processes, always entails the same basic series of events; rising air must be cooled adiabatically until the dew-point temperature is reached and precipitation can form liquid droplets or solid ice crystal through condensation. The droplets or crystals then continue to form and grow so long as vertical mass transport can supply the required moisture (*Dingman*, 2008). The more specific details surrounding these basic processes define what can be considered as stratiform precipitation, and what is convective.

Stratiform precipitation is commonly observed in mid-latitudes (*Houze*, 1997), and is related closely to atmospheric phenomena such as frontogenesis and the convective movement of air resulting from orographic uplift (*Ruiz-Leo et al.*, 2013). Stratiform precipitation can occur whenever saturated air is forced upwards into the atmosphere, forcing the generation of precipitation (*Houze*, 1997). When frontal convergence produces precipitation through such features as extra-tropical cyclones, the result is a system that can be up to 1500 km in diameter, thus covering huge swaths of land that are then showered with low to moderate precipitation amounts, relative to convective precipitation. Stratiform precipitation quantities are generally lower than convective amounts due to the weak nature of the uplift experienced (*Ding-*

man, 2008). An important characteristic to note is that the production of stratiform precipitation is a thermodynamically stable event. This stability is not echoed in convective precipitation formation, where local vertical atmospheric instability is a driving factor (*Ruiz-Leo et al.*, 2013).

Convection is triggered when the rate of heating of a surface or near-surface air mass out-paces the capability of diffusive action to re-establish the density balance between the surface air mass and the air above it. This results in the lower, less dense air mass becoming buoyant, and vertical uplift and overturning then start to occur in conjunction with downdrafts (*Houze*, 1997), fueling the generation of intense precipitation as water vapour condenses with uplift. The main updraft zone can be thought of as the core of the convective system; the majority of convective precipitation falls to the Earth's surface within only a few kilometers of the system's updraft location (*Houze*, 1997). Aside from the low-level air convergence at the heart of convective systems, upper-level divergence zones are also created, due to the release of latent heat when precipitation forms and falls in convective cells. This latent heat release results in the creation of buoyancy in the upper regions of the system that pushes the surrounding air masses outwards and down through the action of gravity waves (*Houze*, 1997).

These processes describe the average setup of one convective cell, but individual convective systems can also merge into a single cirriform-type mesoscale convective system (*Tadesse and Anagnostou*, 2009). However, the average coverage of a convec-

tive system is commonly limited to only a few square kilometers (*Dingman*, 2008). Overall, convective storms are generally both small scale, and produce large precipitation intensities. This means that precipitation fields containing greater proportions of convective systems exhibit larger standard deviation values in regards to rain rate, as compared to stratiform-dominated fields (*Varma and Liu*, 2004).

Convective systems also undergo developmental life stages that impact the characteristics of the systems. The amount of lightning in a convective system is one such attribute, with the greatest amount of lightning flashes detected during the initial system growth phase, and less lightning occurring as the system ages. The areal coverage also varies with the convective system's development stage. The coverage area grows as a system matures, peaking during the middle to latter portion of the life cycle, and then dropping in size towards the dissipation point (*Tadesse and Anagnostou*, 2009). The rain volumes of convective storms follow a similar trend as the system area, peaking during maturity, or the mid life-cycle stages (*Tadesse and Anagnostou*, 2009).

It is clear from the above descriptions that stratiform and convective precipitation are different down to the very nature of their evolution. However, these differences can be highlighted even further by paying specific attention to the vertical motions associated with each. In the convective case, the vertical movement of air can reach speeds on the order of 10 to 30 meters per second (*Dingman*, 2008), which can match the falling speeds of ice particles. In the stratiform case, falling precipitation

meteors have speeds that dominate the vertical air speed (*Ruiz-Leo et al.*, 2013), which can reach up to approximately 1 meter per second (*Dingman*, 2008). Due to this difference in the vertical winds, convective precipitation ice crystals tend to grow due to accretion mechanisms such as ice-crystal riming and liquid water coalesence (*Houze*, 1997). Convective precipitation droplets and crystals spend longer, and become larger, in the cloud environment than stratiform precipitation (*Ruiz-Leo et al.*, 2013). Stratiform precipitation, which is capable of falling as it is generated, evolves primarily due to vapour condensation (*Houze*, 1997). Further differences can be found by looking beyond vertical updraft speeds, to the resulting cloud formations associated with each precipitation type. Convective precipitation is frequently paired with cumulus and cumulonimbus clouds, while stratiform precipitation is linked with nimbostratus clouds (*Ruiz-Leo et al.*, 2013).

Despite the clear contrasts between stratiform and convective precipitation, it is not always simple to distinguish between the two types, even from a theoretical perspective. For example, the distinction between convective and stratiform precipitation is complicated by the possibility of both types of precipitation occurring jointly in the same system (*Houze*, 1997). This type of simultaneous occurrence can be witnessed in well developed convective environments, such as in cumulonimbus clouds, where both new and relatively older regions of convection can exist. Even though the driving factors instigating the weather event may be held constant, the younger, new convection zone alone results in what is traditionally thought of as

convective precipitation, with the characteristic high intensity. Older convection regions become weaker, with tamer vertical wind speeds that allow for precipitation to fall in a manner more in line with stratiform precipitation (*Houze*, 1997).

The difference between young and old convective regions is related to atmospheric layering, and to interactions with external air masses that occur in the different parts of the system; younger convective regions frequently display a two layered system with divergence and convergence of air occurring at upper and lower levels respectively, while for the older convective regions convergence of air into the convection zone is seen in the middle, with divergence occurring at upper and lower levels. This effect is particularly well noted in the tropics, although it can also be observed in mid-latitudes (*Houze*, 1997).

Precipitation Partitioning Approaches

The practical partitioning of precipitation as either convective or stratiform is an issue that has received considerable attention. The ability to designate convective and stratiform classifications has been found to be useful in many applications, ranging from the interpretation of remote sensing data (*Lang et al.*, 2003), to enhancing the validation processes of numerical weather prediction models (*Houze*, 1997). The division of convective and stratiform events can generally be based on such characteristics as rain rates, cloud water, updraft velocities, precipitation fall velocities, and assumed spatial relationships and convective extents. Such methods are commonly applied to either surface precipitation data or radar data (*Lang et al.*, 2003).

The performance of such methods can vary when used in an atmospheric modelling context, with different approaches being capable of improving different aspects of convective system attributes such as spatial characteristics, microphysics, and precipitation production. For example, it is the methods that employ a convective rain rate threshold that tend to result is the largest classifications of stratiform processes (*Lang et al.*, 2003). However, not all of these methods are useful to interpreters of observational data outside of the atmospheric sciences, as they are often forced to work with limited information.

Conveniently, the classification of precipitation events as either stratiform or convective is supported not only by theoretical differences in the genesis and evolution of the events, but also by the statistical characteristics of the observed precipitation. For example, even if identical average precipitation for a relatively small scale region of interest is observed, such as for a 30 by 30 km^2 domain, statistical differences in such attributes as the fractional rain cover and the conditional PDF of the instantaneous rain rates can be noted; convective events will often possess a conditional PDF with a greater amount of spread, and a lower fractional rain cover (*Varma and Liu*, 2004).

However, it is important to recognize that internal sources of variability also exist in the statistical characterization of both stratiform and convective rain. This is due to factors such as the ocean or land surface type underlying precipitation events, and the latitude of occurrence of events. Surface type has the greatest influence

on convective events, while latitude impacts stratiform events the most (*Varma and Liu*, 2004). Therefore, variability in statistical characteristics of stratiform and convective precipitation event types is an important consideration when developing a classification system dealing with diverse geography in a single domain, as it possess the potential to blur classification lines.

Many methods for diagnosing precipitation as convective or stratiform also exist which are dependent on radar data. As an example, such methods can allows for the application of texture differentiation methods that commonly concern themselves with radar reflectivity measures and their local maxima and averages (*Tremblay*, 2005). More generally, radar-based classification often depends on the rule that stratiform precipitation roughly appears to be uniform in the horizontal, with a band of high reflectivity associated with the transition from solid to liquid precipitation as falling stratiform ice crystals melt during their descent. Convective precipitation possesses different characteristics, with localized high reflectivity that cuts through the vertical cross section and shows up as a cell in plan view (*Houze*, 1997; *Varma and Liu*, 2004). Additionally, radar reflectivity data can be used to classify precipitation by assigning a convective reflectivity threshold value, or by detecting reflectivity pixels that are stronger than neighbouring pixels, for example in a 10 km radius, by a pre-selected factor. If a convective centre is detected in such a way, the spatial extent of the convective precipitation can be estimated by assuming a 4 to 5 kilometer radius of connectivity extending out from the identified centre (*Houze*, 1997).

Lightning is another means by which precipitation classification can occur. Lightning occurs when a charge difference builds up between clouds and the surrounding environment, and is used to regain a balance by dissipating the excess charge (*Burrows and Kochtubajda*, 2010). Lightning can occur within a cloud, between clouds, between the cloud and the air, and between clouds and the ground. The direction of movement of lightning is often from the negatively charged area to a positive one (*Cummins and Murphy*, 2009). Cloud-to-ground lightning is often of the most interest, in regards to precipitation classification. This type of lightning is usually discretized by referring to flashes, where each flash can be comprised of multiple strokes, often 2 to 3. Each stroke usually lasts for tens of microseconds, has current values on the order of a few to 200 kiloAmperes (*Cummins and Murphy*, 2009), and are separated from other strokes in time by about 20 to 100 milliseconds (*Cummins and Murphy*, 2009). For the CLDN data set, the location of a lightning flash is designated as the location of the first stroke in that flash (*Burrows and Kochtubajda*, 2010). Often strokes will hit the ground at the same point as the first, but in approximately 30 to 50% of cases the stroke may hit at a different location. This has lead to the popular flash definition as encompassing all strokes inside of a 5 km radius that occur in one second (*Cummins and Murphy*, 2009).

Lightning is normally tightly associated with convective, or thunderstorm-type weather events. The more intensive movement of air and condensation of water vapour that occurs in convective systems generally promotes cloud electrification

(*Mattos and Machado*, 2011). Convective systems do not always display all types of this convenient identification feature, however. It is possible for a convective system to experience no cloud-to-ground lightning strikes (*Tadesse and Anagnostou*, 2009; *Mattos and Machado*, 2011). When cloud-to-ground flashes do occur, certain storm characteristics are usually seen.

A study conducted over a US domain during the summer period has highlighted that convective systems possessing cloud-to-ground lightning flashes generally last longer, reaching durations up to 15 hours, while convective systems that do not produce this type of lightning generally last only up to 7 hours. Assessments of electrical activity growth during the formation period of convective systems can provide a hint as to the anticipated intensity and duration of the system (*Mattos and Machado*, 2011). Furthermore, systems with cloud-to-ground flashes also exhibit a greater restriction in their time of development, showing increased formation after 15:00 UTC, due to the increased surface heating needed to drive the strong updrafts that sustain ice particle supplies, and which allow for the development of lightning (*Tadesse and Anagnostou*, 2009). Storms that do not produce cloud-to-ground lightning do not show such a strong relationship between development time and the time of day (*Tadesse and Anagnostou*, 2009).

Unfortunately for classification, convective systems do not hold the exclusive ability to produce lightning. Cold fronts in such systems as extra-tropical cyclones can instigate more intensive, stratiform-type weather at somewhat smaller spatial

scales than is normally seen. If the uplift achieved at such a front is sufficiently intense, thunderstorms with lightning production can result (*Dingman*, 2008).

Introducing Convection into CaPA

While all of the above listed methods of precipitation classification have merit, their use in this investigation is limited by the restriction of data to CaPA inputs and NLDN data. Thus only those methods that solely require rain rates or lightning data are applicable, and it is specifically the methods proposed by Ruiz-Leo et al. (2013) and Tremblay (2005) that are used in this investigation. These methods focus on the idea that when total cumulative precipitation is plotted against rain rate, the stratiform component can be modeled by an exponential curve. The curve is fit such that all deviations from the exponential lie above the curve, and are then assumed to have convective origins (*Ruiz-Leo et al.*, 2013; *Tremblay*, 2005). This method of first specifying the stratiform component of precipitation, and then defining convective precipitation contributions afterward, is concerning when it comes to storm systems that give rise to both precipitation types. Any intermediate precipitation, which may have kinematics distinct from stratiform and convective types, may be lumped in with truly convective precipitation despite being more similar to the stratiform type (*Houze*, 1997). This potential complication was not addressed in this investigation. However, the rain rate approach was also not the sole criterion for deciding upon the classification of precipitation in this study. The standard deviation of the precipitation field, as well as lightning coverage, were also considered.

It is anticipated that classifying precipitation as stratiform or convective in CaPA will allow for the generation of semivariograms tailored to each to type. Such semivariograms, which by definition capture spatial correlation, would then be able to properly reflect the unique spatial characteristics of each precipitation type, generated not only by the scales of their extents but also by the gradients of the precipitation intensities across those extents. By properly representing the spatial characteristics of each precipitation type, the SI combination of observation and GEM data can be expected to improve, specifically for areas of convection that were previously overshadowed by larger scale stratiform influences. It is anticipated that this will allow CaPA analyses to perform better under convective conditions, where dramatic precipitation rates can have significant implications on flood forecast modeling, infrastructure risk assessments, and agricultural land management modeling (*Tadesse and Anagnostou*, 2009).

3.2 Semivariogram Estimation Methodology

Semivariograms play an integral role in the SI process, directly influencing the way in which innovations are combined to produce an innovation estimate at a required location. Needless to say, the way in which semivariograms are estimated and modelled is intrinsically linked to the quality of innovations at estimation locations. As a result, improvement to the semivariogram estimation technique is a natural point of focus for efforts to improve CaPA analyses. The semivariogram improvements

proposed here can be divided into two types: improvement through the introduction of anisotropic semivariogram parameters, and enhancement by accommodating convective storm activity.

3.2.1 Estimation and Inclusion of Anisotropic Semivariogram Parameters

Semivariogram estimation can be a tricky process, and when done manually can require lengthy data investigations and critical judgments on when a theoretical semivariogram should be deemed sufficient. This means that the process invites subjectivity, which can hinder the performance consistency of applications dependent on the semivariogram. To avoid this issue, and to stay in line with the current semivariogram estimation approach practiced by CaPA, an automated semivariogram estimation approach was undertaken using ArcGIS.

To estimate the semivariogram parameters, ArcMap 10.1 and ArcCatalog 10.1 were used. Both programs are part of the ArcGIS 10.1 suite, a geographic information system (GIS) software package created by ERSI. The ArcMap program was used for visualization of the innovation data and the results of the semivariogram estimation processes during the initial investigations, while ArcCatalog provided a graphic-free and more efficient means of estimating the required semivariogram parameters once the setup of the estimation process was established.

Within ArcGIS, multiple options exist regarding the type of theoretical semivari-

ogram to be fit to the input data. In order to narrow down the selection of theoretical semivariogram types, a preliminary visual inspection was conducted, during which the fit of each possible theoretical type was assessed for a single sample time step of innovation data. During this appraisal, the fit to the experimental semivariogram was not the only consideration. The range of possible curve shapes potentially captured by each candidate type also played an important part in the selection process. This lead to the selection of the following semivariograms for further testing: exponential, spherical, Gaussian, and a linear combination of all of the above. Note that the operational semivariogram was also replicated in ArcGIS, using an isotropic exponential semivariogram with innovations based only on non-zero precipitation amounts. The ArcGIS version of the operational semivariogram is what all other configurations are compared against.

Using the ArcGIS Geostatistical Wizard, contained in the Geostatistical Analyst toolbox, each of the four selected semivariogram types were specified in detail, creating model definition files. For each model type, an isotropic and anisotropic version was created, with the nugget parameter enabled. For all cases, the maximum number of neighbours considered was set to be 32, while the minimum number was designated as 2. Similarly, a lag size of 10 km for 50 lags was also defined. These values are in agreement with the default CaPA semivariogram parameters, and were adopted for consistency.

Once the semivariogram model files were established, the *auto* field for all values

expected to vary with differing input data sets had to be modified using the *arcpy.GASetModelParameter_ga* python command. By re-setting the *auto* field to true, the corresponding value becomes up-datable when the model file is used to create a new geostatistical layer in ArcGIS. This is the process by which the semivariogram parameters are automatically estimated, with the newly determined parameters being accessible through the *arcpy.GAGetModelParameter_ga* command.

Upon completion of the model file setup, the actual work of going through the innovation data set for each time step and coming up with the required semivariogram parameter sets had to be undertaken. This process was eased through the production of a python script, which was created to automate the remaining extent of the estimation procedure. A summary of the steps taken during the execution of this script is provided in Appendix F.

It is important to note that all four semivariogram types were estimated by the python script, including the current operational case which is captured by the isotropic exponential semivariogram estimated by innovations that require both the observation and GEM components to be non-zero. Estimating the operational version of the semivariogram in ArcGIS ensures that the CaPA results obtained from each set of semivariogram parameters are directly comparable both with one another, and with the operational case.

The intention behind estimating the full series of theoretical semivariograms was never to run all four types through a full CaPA-based investigation, but rather to

preliminarily assess their potential usefulness in capturing the spatial relationships of the innovation data. The geostatistical analyst toolbox was used to create a simple kriging surface for each of the estimated semivariograms, from which a variety of prediction errors were determined based on a cross-validation analysis. Figure 3.4 shows the prediction errors for the unrestricted innovation case, where zero precipitation values are allowed. Clearly all semivariogram types perform relatively similarly, with the optimal type fluctuating based on the type of prediction error. Similar results were seen for all other innovation restriction cases. Therefore, only the exponential semivariogram was considered for further investigation; it performs reasonably well in comparison to the other candidate semivariograms, and is already used by CaPA. This choice allows for a more direct comparison between the operational case and the modified semivariograms, allowing for greater clarity in attributing the investigation results to anisotropy or convection filtering changes.

To further ensure that anisotropy is reasonable for the input innovation data, and that the proposed anisotropic semivariogram methodology is justified, a quick visual assessment of the semivariogram rose diagrams outputted by ArcGIS was conducted. An example of one such diagram, for the 2010/06/01/06:00Z time step, and assuming an exponential model, is provided in Figure 3.5. As can be seen, clear indications of anisotropy exist, based on the elliptical patterning of the diagram. This supported the continuation of the investigation.

The impacts of anisotropy on the semivariogram are also depicted in Figure 3.6,

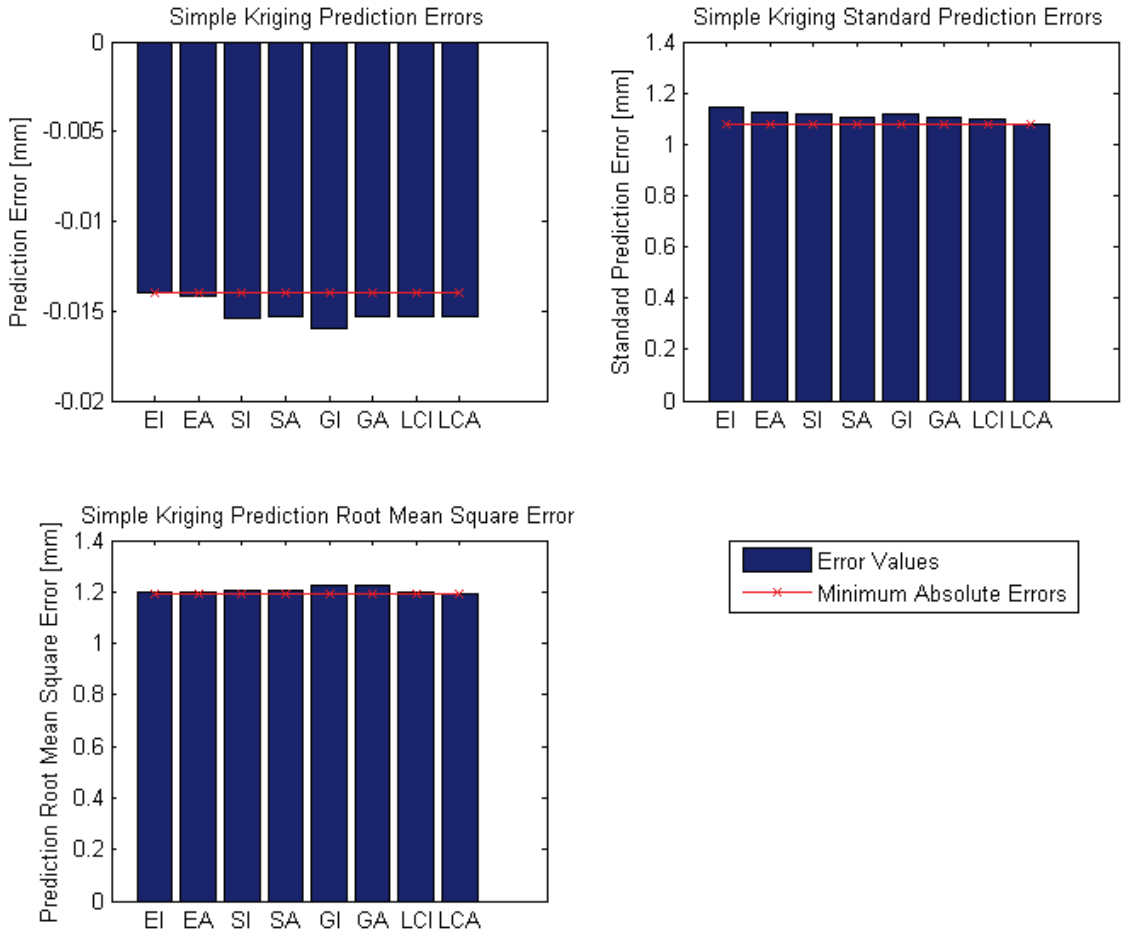

Figure 3.4: Prediction errors for various semivariogram types. Input innovations include zero precipitation amounts from both the observations and GEM. The following codes apply: E = exponential, S = spherical, G = Gaussian, LC = linear combination, I = isotropic, and A = anisotropic

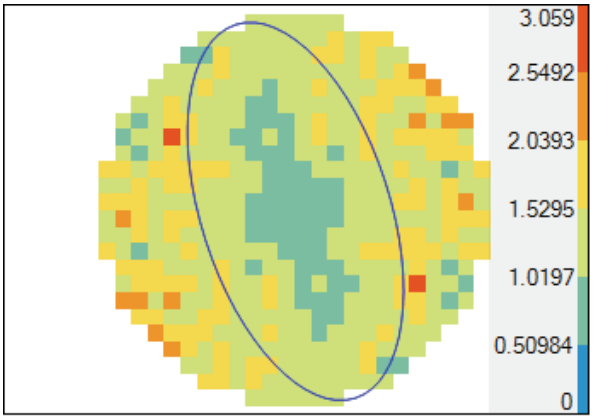

Figure 3.5: Exponential semivariogram rose diagram, depicting semivariogram values in space

which plots a series of directional exponential semivariograms against the binned and bin-averaged experimental data. The thickness of the modeled semivariogram line is due to the differences in the multiple, overlapping directional semivariograms plotted. It should be noted while the exponential semivariogram was selected for further investigation, it is possible that the degree of impact anisotropy might have on other model types may vary. This is highlighted in Figure 3.7, where the differences in the modeled directional semivariograms are more distinct than those seen in Figure 3.6.

In order to make use of the estimated semivariogram parameters, modification of the ArcGIS output files was also required. A Matlab script was created to transform the ArcGIS results into a format usable by CaPA. Further details regarding the Matlab script can be found in Appendix F. During this process, the correct translation of the ArcGIS values to the CaPA variables is critical, and can be understood by relating back to Figure 3.2. Based on this figure, the following links can be made:

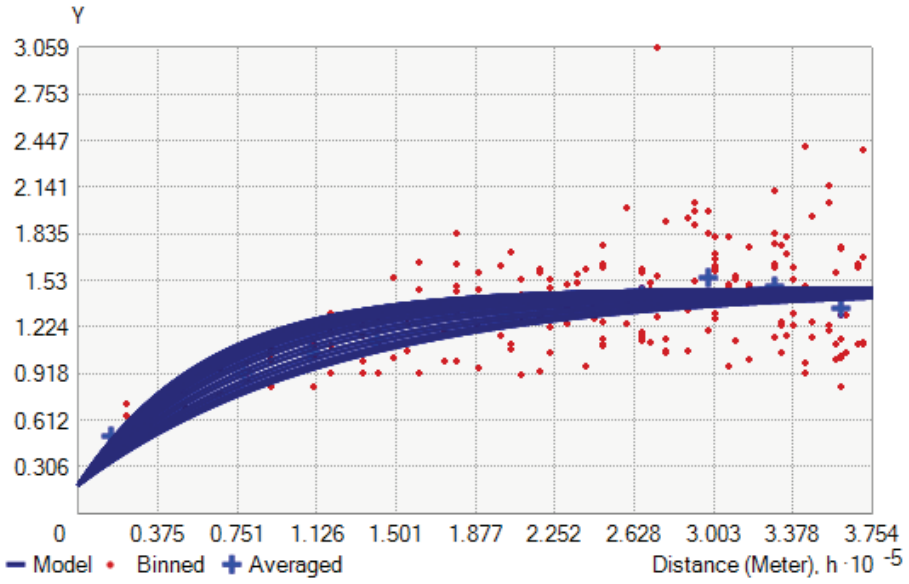

Figure 3.6: Exponential semivariogram model fit, with anisotropic model results depicted for the the 2010/06/01/06:00Z time step

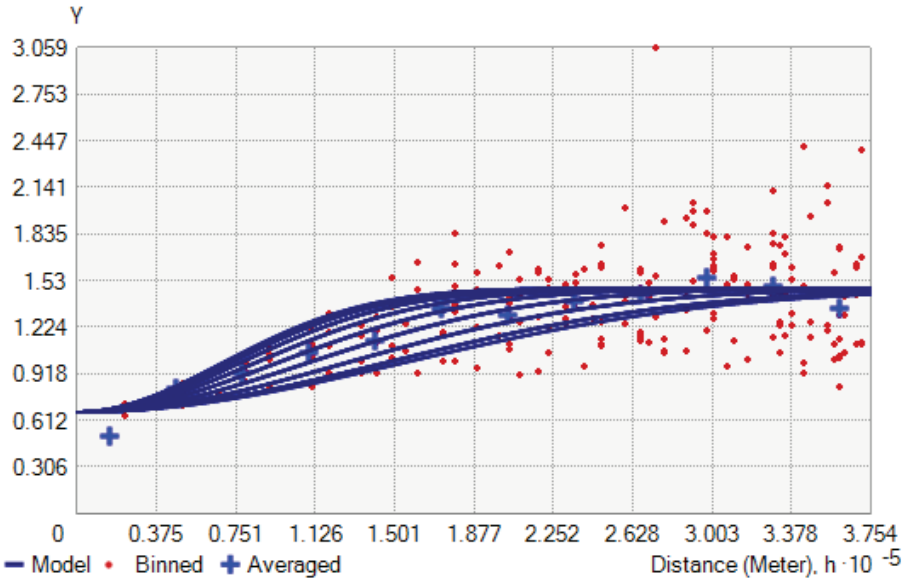

Figure 3.7: Gaussian semivariogram model fit, with anisotropic model results depicted for the the 2010/06/01/06:00Z time step

$$\text{Observation Error} = \text{nugget} \tag{3.13}$$

$$\text{Background Error} = \text{sill} \tag{3.14}$$

$$\text{Correlation Length} = \text{range} \tag{3.15}$$

where all of the above quantities describe basic attributes of interest during the SI methodology. Note that the correlation length is a measure of the distance between innovations at which a correlation is still detectable.

Beyond the above equalities, the distance value between innovations also warranted special attention. Under the traditional scheme, the distance between two innovation points was only required to be a scalar quantity. With anisotropy introduced into the CaPA code, this distance measure had to be transitioned into a vector with the requisite angle also being calculated and stored in a new variable. It is this angle that allows for anisotropy to be realized, as the angle is used in conjunction with the elliptical semivariogram equation and the major and minor semi-axis quantities determined by ArcGIS to calculate the appropriate correlation length. The updated correlation length variable can then be applied in the traditional *covariogramme_xyz* Fortran function to generate SI weights as needed.

Furthermore, a semivariogram filtering scheme had to be applied, to replicate

the theoretical semivariogram estimation procedure followed in the operational version of CaPA. This scheme follows the basic methodology outlined in Section 1.1.3. The total number of innovations used to generate each parameter set are used in place of $Num_{Exp_{i,t}}$ and $Num_{Smth_{i,t-1}}$ in Equation 1.14, for the current time step's estimated parameters and the previously smoothed parameters respectively. Similarly, the smoothing filter is applied to the parameter values, rather than the binned experimental semivariogram values.

The obtained and refined semivariogram parameter sets for the study period, whether isotropic or anisotropic, only provide value if they are put to work in CaPA. The implementation process followed required multiple alterations to the CaPA code, in order to accommodate the new information describing semivariogram anisotropy. The modifications applied to the CaPA program are detailed in Appendix F.

3.2.2 Convection Filtering Methodology

Convective precipitation has spatial characteristics that distinguish it from larger-scale stratiform precipitation. As such, it is a logical point of departure from which to develop a means of improving semivariogram performance in CaPA. However, an initial conundrum must first be resolved: What constitutes convective activity in the context of CaPA's semivariogram estimation procedure? To answer this question, a series of convection metrics were developed and applied to help identify the most convective days in the study period.

Standard Deviation of Precipitation

The standard deviation of precipitation fields was assessed for both the observation data and the GEM background field. Standard deviation values are expected to increase with the presence of convective systems (*Varma and Liu*, 2004). The standard deviation values were calculated based on the logarithm of the rain rates (*Varma and Liu*, 2004), either reported by observation stations or modelled by GEM. This was done across the entire domain, and for each time step as follows:

$$\sigma_O = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(r_{O_i} - \bar{r_O}\right)^2} \tag{3.16}$$

$$\sigma_G = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(r_{G_i} - \bar{r_G}\right)^2} \tag{3.17}$$

where σ_O and σ_G are the standard deviations calculated for the observed and GEM data sets respectively, n is the total number of rain rates available in a data set for the time step being considered, and r_O and r_G are the rain rates for the observed and GEM data sources. The rain rates are calculated on a per station or grid cell basis, and are converted to millimeters per hour.

Lightning Occurrence

The occurrence of lightning is also a signifier of convective activity. Using lightning data obtained from Environment Canada for the study domain and time period, the

percentage of lightning coverage was determined for each time step. These values were calculated by first isolating all lightning occurrences during the time interval spanning from the previous time stamp to the current time of interest, for all time steps throughout the study duration. Once all of the lightning data was attributed to the proper time period, spatial allocation was required. This was done on a per time step basis, and required the assignment of all lightning occurrences during a given time to a GEM grid cell, by finding the minimum distance between the latitude and longitude coordinates of a flash and the centers of the GEM cells. The percentage of lightning coverage for a time step was then determined according to the following:

$$\text{Coverage} = \frac{\sum_{i=1}^{n} L_i}{n} \tag{3.18}$$

where n is the total number of GEM grid cells within the study domain, and L denotes a binary vector indicating the occurrence or non-occurrence of lightning strikes in a grid cell by ones and zeros respectively.

It is important to note that the percentage of lightning coverage is required as opposed to a simple binary indicator of lightning occurrence for the domain, as lightning occurs for almost every time period of the study. This is emphasized in Figure 3.8. The figure also clearly demonstrates that the number of lightning occurrences, and thus the percentage of lightning occurrences, varies through time and may be useful in identifying the most strongly convective days during the study period.

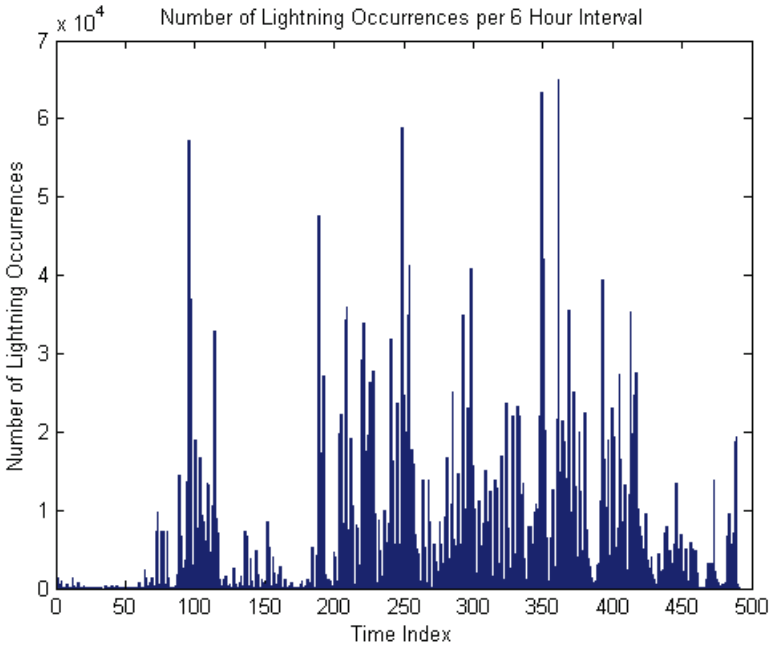

Figure 3.8: Lightning occurrence counts across the study domain, per 6 hour interval for all study time steps

Convective Precipitation Coverage

A methodology outlined by Ruiz-Leo et al. (2013) allows for the direct calculation of the amount of convective precipitation that falls in a given region for a single realization of a precipitation field. This technique was applied with some modifications to the study domain, and was used to determine the amount of convective precipitation per time step based on the GEM precipitation field. The GEM data was used in lieu of the observation data due to the improved resolution of data available per time step. Use of the observation data may have otherwise undermined the quality of the quantitative estimates by missing significant smaller scale precipitation features, and thus failing to properly capture the distribution of the precipitation data (*Tremblay*,

2005). GEM is known to over-estimate small precipitation amount occurrence, and thus to have its own distributional errors, but the advantage of GEM's spatial resolution overcomes this potential drawback. Errors associated with the identification of a convective precipitation threshold due to the use of the GEM field are also mitigated by the use of multiple convection metrics.

In order to determine the amount of convective rain coverage that occurs for a time step, the critical rain intensity for that time step must first be identified. The critical intensity is the tool that is used to classify the precipitation type, which can be considered as stratiform or convective. If precipitation intensity is plotted against the total cumulative precipitation that falls at each intensity level, the result can be approximated by an exponential curve fit to the minimum total cumulative precipitation amounts. This is done in such a way that any observed data point on the plot then falls above the exponential curve. The result can be approximated by the following function, which is decomposable into stratiform and convective components (*Ruiz-Leo et al.*, 2013; *Tremblay*, 2005):

$$P(R) = A(t) \exp(-B(t) R) + p'' \tag{3.19}$$

where $P(R)$ is the total cumulative precipitation associated with the rain rate R, $A(t)$ and $B(t)$ are parameters of the exponential distribution that change with time step t, and p'' is used to accommodate positive anomalies falling above the exponential curve. These positive anomalies can be thought of as convective contribu-

tions combined with any statistical errors (*Tremblay*, 2005), while the smoother, underlying precipitation distribution, described by $A(t)\exp(-B(t)R)$, arises due to stratiform precipitation (*Ruiz-Leo et al.*, 2013).

While this method of describing the rain intensity versus total cumulative precipitation plot is both convenient and physically based, it did not suit the study data set. As can be seen from Figure 3.9, when all of the data from all time steps is pooled, the result can easily be fit to an exponential distribution with a high degree of confidence. However, at such time scales, convective influences cannot be resolved. It is the 6-hour time step that is of interest for this investigation. At the 6-hour time scale, the plot should theoretically still develop a similar basic exponential base curve, but with the addition of convective spikes, as convective systems typically take at least 6 hours to materialize (*Ruiz-Leo et al.*, 2013). Furthermore, the standard analysis time-step used in this investigation is 6 hours.

When the data is plotted for 6 hour time steps, an underlying exponential distribution cannot reliably be identified. This may be due to the use of GEM model data rather than data from a dense observation network, as biases related to specific precipitation magnitudes in GEM may conspire to muddy the underlying distribution to a degree where a exponential approximation is inappropriate (*Tremblay*, 2005). This can be seen in Figure 3.10, which shows attempts at fitting an exponential curve by considering all of the data for the time step, as in the *Full Exponential Curve* case, or by considering only the data from the peak onwards for the time step, as in the

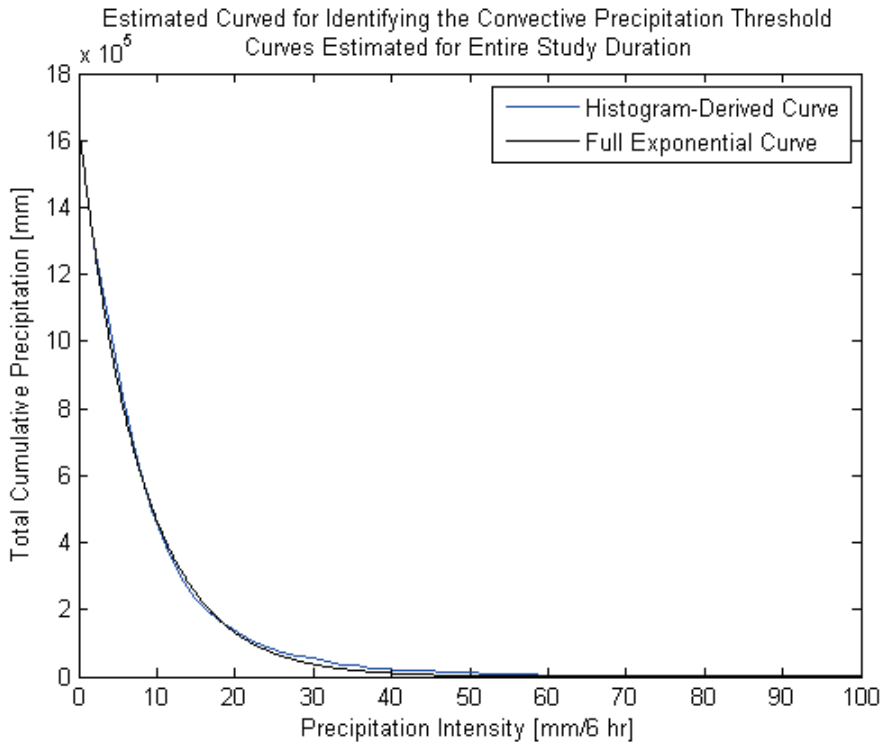

Figure 3.9: Exponential stratiform curve fit to the full study record

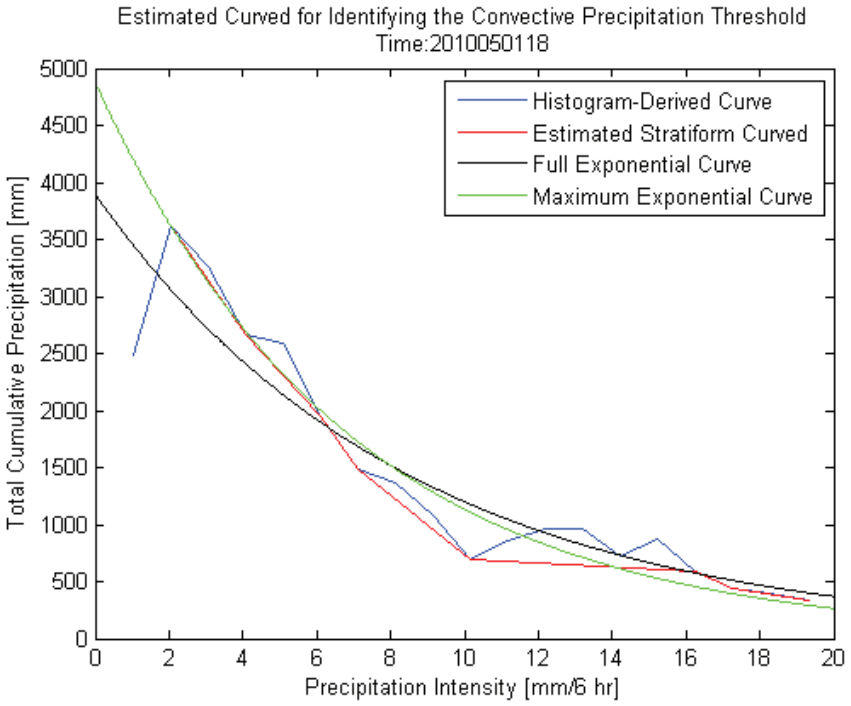

Figure 3.10: Stratiform curves poorly fit to the 2010/05/01/18:00Z time step

Maximum Exponential Curve case. Both attempts at fitting the exponential curve produce poor results.

Some instances do occur during the study where an exponential curve can be fit to the stratiform component of the intensity versus total cumulative precipitation plot reasonably well. This is highlighted by Figure 3.11 based on visual inspection. However, because of the inconsistency in the exponential curve's appropriateness, an alternative means of isolating stratiform and convective precipitation components was required.

The concept adopted by this study to distinguish between stratiform and convective precipitation is founded on principles similar to those of the exponential

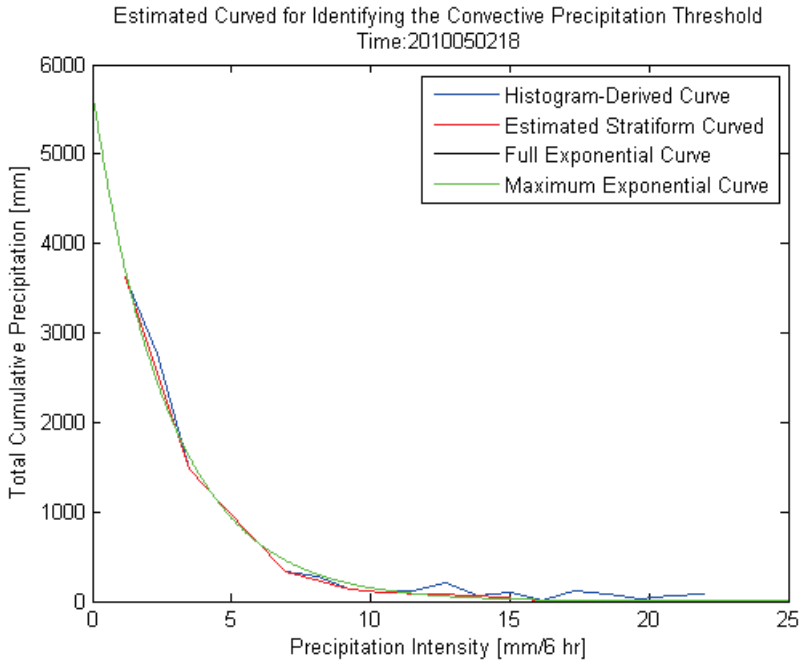

Figure 3.11: Stratiform curves reasonably well fit to the 2010/05/02/18:00Z time step

distribution approach; the stratiform precipitation should comprise the underlying, smoother trend in the plot, while the anomalous spikes seen in the data are characterized as convective. Based on this idea, the stratiform curve was developed for each 6 hour period by linearly connecting binned rain rate data points such that none of the original binned plot values fall below the estimated stratiform curve. This in effect roughly creates a somewhat smooth base stratiform curve. Examples of such estimated stratiform curves can also be seen in Figures 3.10 and 3.11, and details related to the creation of the stratiform curve are provided in Appendix F.

Under the original methodology proposed by Ruiz-Leo et al. (2013), the critical intensity, denoted R_c, is set as the rain rate that partitions the data into 60% con-

vective and 40% stratiform in origin. This simple means of specifying R_c, however, is based on generalizations regarding the average amount of total precipitation normally attributable to convective systems. Rather than depending on this assumption to apply consistently across all study time steps, the concept proposed by Tremblay (2005) was implemented. The Tremblay (2005) method specifies R_c as the intensity below which the total cumulative precipitation is dominated by stratiform-type events, while above R_c convective precipitation prevails.

In order to implement this method, the differences between the empirical plot points and the estimated stratiform curve are first found. Note that this is done after a preliminary correction of the stratiform curve, to change all negative total cumulative precipitation amounts to zero. Using the obtained difference values, the R_c value is set as the intensity for which all larger bins consistently have empirical total precipitation amounts greater than the total precipitation associated with the stratiform curve for each bin, indicating a transition into convection-dominated precipitation events.

With the R_c value established, all GEM precipitation amounts across the domain were then classified as either stratiform or convective in nature. This led to an overall quantification of the amount of convective precipitation coverage across the domain for each time step, by calculating the percentage of GEM grid cells reporting convective precipitation. The idea of determining whether or not an entire area is convective based on the percentage of coverage of identified convective precipitation

is supported by a similar approach used in classifying radar data, where satellite pixels are classified in this manner (*Varma and Liu*, 2004).

Finally, it is important to note that no attempt is made to ensure that the spatial extents and coherence of stratiform and convective features are realistic. While it is possible that undertaking such an endeavor may increase the amount of convective precipitation identified for a given time step (*Lang et al.*, 2003), all time steps are treated similarly as thus no significant relative difference should be seen when comparing the convective coverage of different time steps.

Metric Combination and Filtering Modification

While all three of the above convection metrics are useful, it is a combination of these metrics that is finally employed to determine if a given day in the study period is considered to be convective or not. This designation must be handled with care. If a day is incorrectly classified as being convective, then the results of the convection filtering technique in CaPA focused on the convective day subset will likely be misleading, and may erroneously suggest that the convection filtering method is ineffective. As such, the precaution was taken to identify the days that are not simply convective, but rather are considered to be the most convective.

In order to determine if a day should be classified as being one of the most convective during the study period, the 0.85 quantile was used. For the standard deviations of the observation and GEM precipitation fields, all those days with standard deviation values greater than the 0.85 quantile value were identified, while a similar

process was used to pinpoint the days whose percentage of lightning strike coverage and percentage of convective precipitation coverage were also greater than the 0.85 quantile value. The indexes of the qualifying days were then compared across the four metrics, and those days where the convective precipitation coverage and at least one other metric agreed were taken as the subset of the most convective time steps during the study.

Note that the convective precipitation coverage was used as the primary metric of concern, as it was felt that it was the most direct indicator of true convective precipitation. For example, the standard deviation of the precipitation field may still be relatively high if moderate regions of stratiform precipitation are present within an otherwise low rainfall rate system, as compared to truly intense convective precipitation in a very small area that is otherwise dominated by very uniform precipitation rates, particularly if the uniform rate is moderate in intensity. In terms of the lightning data, the possibility for lightning flashes to be misleadingly associated with extreme stratiform events exists. More importantly, the metric developed for the lightning data does not directly account for the amount of convective precipitation that occurs. It may be possible for two time steps with the same percentage of lightning coverage to experience different amounts of convective precipitation. In such a scenario, it is the day that experiences the most convective precipitation that is of primary interest, as such a day is more likely to highlight the impacts of applying the convection filtering methodology which is already anticipated to be diluted

by the use of one semivariogram for the entire study domain.

Once the subset of convective days was determined, these days were treated with a modified semivariogram parameter filtering regime, to reflect the significant presence of convective weather in the domain during those days. Convection is a small scale phenomena, and the estimated semivariogram for days dominated by convective activities should capture this effect on the spatial correlations of innovation values. However, filtering the semivariogram parameters allows for previously estimated parameters, for instance from stratiform-dominated time steps, to have an influence which may hamper how well the semivariogram truly reflects the convective nature of the current time step. Thus convection filtering was implemented.

The convection filtering method is twofold. The first part of the method allows for the semivariogram parameters of convective time periods to by-pass any filtering, remaining true to the ArcGIS estimates. The second part of the method produces another version of the convective semivariogram parameters, where the influence of the filtering is limited such that at least 50% of the final parameter values must be derived from that time step's ArcGIS estimated values. The two sets of convective semivariograms were both tested independently in CaPA, with the verification statistics being calculated in both cases only for the determined convective time periods.

3.3 Results

The semivariogram verification results presented in this section are not determined using the modified verification method detailed in Section 2.3, but rather the standard verification method commonly used when assessing CaPA analyses, as programmed by Environment Canada. Due to the lack of any substantial post-processing activities required for either the anisotropic or convection filtering methodologies, no limitations on applying the standard verification methodology existed. Thus all verification scores reported in regards to semivariogram improvements are based on the leave-one-out analysis results outputted by CaPA, as opposed to the use of verification cases. This is preferable, as it presents the verification scores without the hindrance of withheld observation stations.

For the purposes of comparison, all semivariogram results have been calculated using the verification method of Section 2.3 as well, although only for the 20% verification case. These results were found to be in good agreement with the leave-one-out verifications. This confirms that the verification approach used for the bias correction methodology is a realistic approximation of the leave-one-out approach standard at Environment Canada.

The final set of results in this section, which combine MWA scheme II and an anisotropic semivariogram, are presented in accordance with the verification methodology used in Section 2.3, using a 20% verification case. This is due to the limitations in the verification score calculation procedure imparted by MWA scheme II.

3.3.1 Anisotropy Implementation Results

The CaPA analysis scenarios used for the calculation of verification skill scores in this section encompass six different configurations, each posing an alternative means of estimating the semivariogram. These CaPA runs are defined as follows:

- 'noInc': CaPA setup with semivariogram estimation based on innovation data that includes zero or no-precipitation values
- 'noInc_A': CaPA setup which includes anisotropy in the 'noInc' configuration
- 'all': CaPA setup equivalent to the current operational version of CaPA, with semivariogram estimation based only on innovations calculated when all of the contributing sources, being both observation and GEM inputs, report precipitation values greater than zero
- 'all_A': CaPA setup includes anisotropy in the 'all' configuration
- 'one': CaPA setup with semivariogram estimation based only on innovations calculated when at least one of either the observation or GEM sources report precipitation values greater than zero
- 'one_A': CaPA setup includes anisotropy in the 'one' configuration

The average bias scores for the CaPA analyses serve as overall indicators of the success of each semivariogram modelling technique. These scores are summarized in Figure 3.12, and clearly indicate that the least amount of absolute bias is incurred

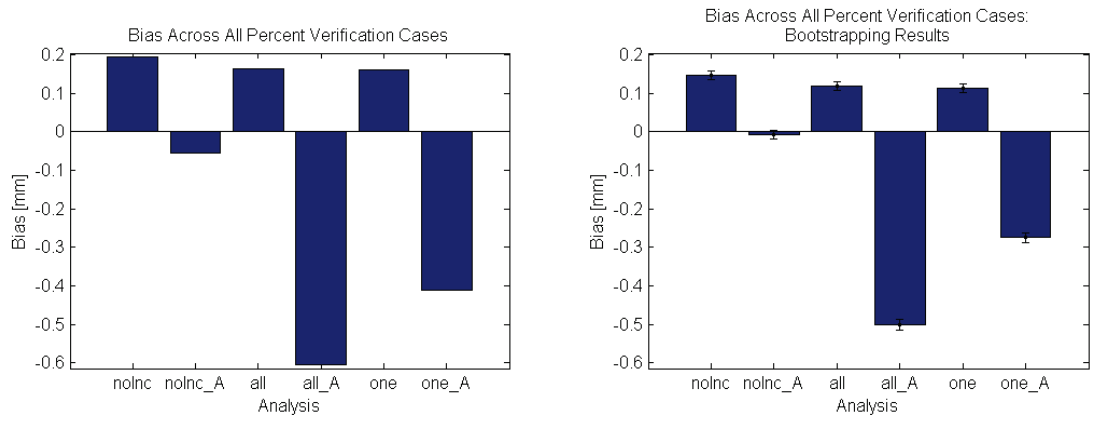

Figure 3.12: Anisotropy investigation bias results for multiple run scenarios averaged across all verification locations and all time periods in the study

when anisotropy is combined with the inclusion of zero precipitation values in the semivariogram estimation process to produce the *noInc_A* case. However, the incorporation of anisotropy does not always prove valuable to the improvement of the bias score. The biases calculated for analyses that include anisotropy show a notable increase when innovations are required to have at least one non-zero precipitation value contributed from either the observation or GEM data. Bias is increased even further when both observation and GEM data must be non-zero to create an innovation. When anisotropy is not considered, the same progression of change in the estimation of the semivariogram has the opposite effect, resulting in successive decreases in the average bias as innovation constraints slacken.

A more insightful glimpse into the development of the bias depicted in Figure 3.12 is provided by the DPM continuous skill scores, which are plotted in Figure 3.13. The DPM scores confirm the strong performance of the *noInc_A* scenario, which shows

consistently low DPM scores for all thresholds, and can be seen to perform the best of all cases analyzed for larger precipitation thresholds specifically. The DPM scores associated with all of the isotropic semivariogram estimations perform relatively similarly, with the *one* configuration generally resulting in the lowest absolute DPM scores. However, the bootstrapping results in the lower panel of the figure suggest that this dominance of the DPM scores is not statistically significant as compared to the *all* case. A similar ranking of the *all* and *one* runs can be seen when reviewing the cases that possess anisotropy, although a shift from negative to positive DPM scores occurs along with an increased spread between the two analyses. In all cases, an increase in DPM values is seen with larger precipitation thresholds, which correlates to increased bias. Similar trends and relationships are seen in the DPS scores as well, which are portrayed in Figure G.1, in Appendix G.

The FBI plot displayed in Figure 3.14 can also be used to help gleam a better understanding of how analysis bias breaks down across different precipitation ranges. From the figure, it is easy to discern that the *all_A* and *one_A* cases behave similarly to each other, and uniquely compared to the remaining cases. For precipitation amounts less than 2 mm, the *all_A* and *one_A* cases exhibit the worst performance as assessed by the FBI categorical skill score, with the *all_A* run generally having the largest FBI scores. However, bootstrapping results indicate that the *all_A* run is statistically indistinguishable from the *one* case for the 0.2 to 1 mm bin. For precipitation amounts greater than 2 mm, all other cases possess negative FBI scores,

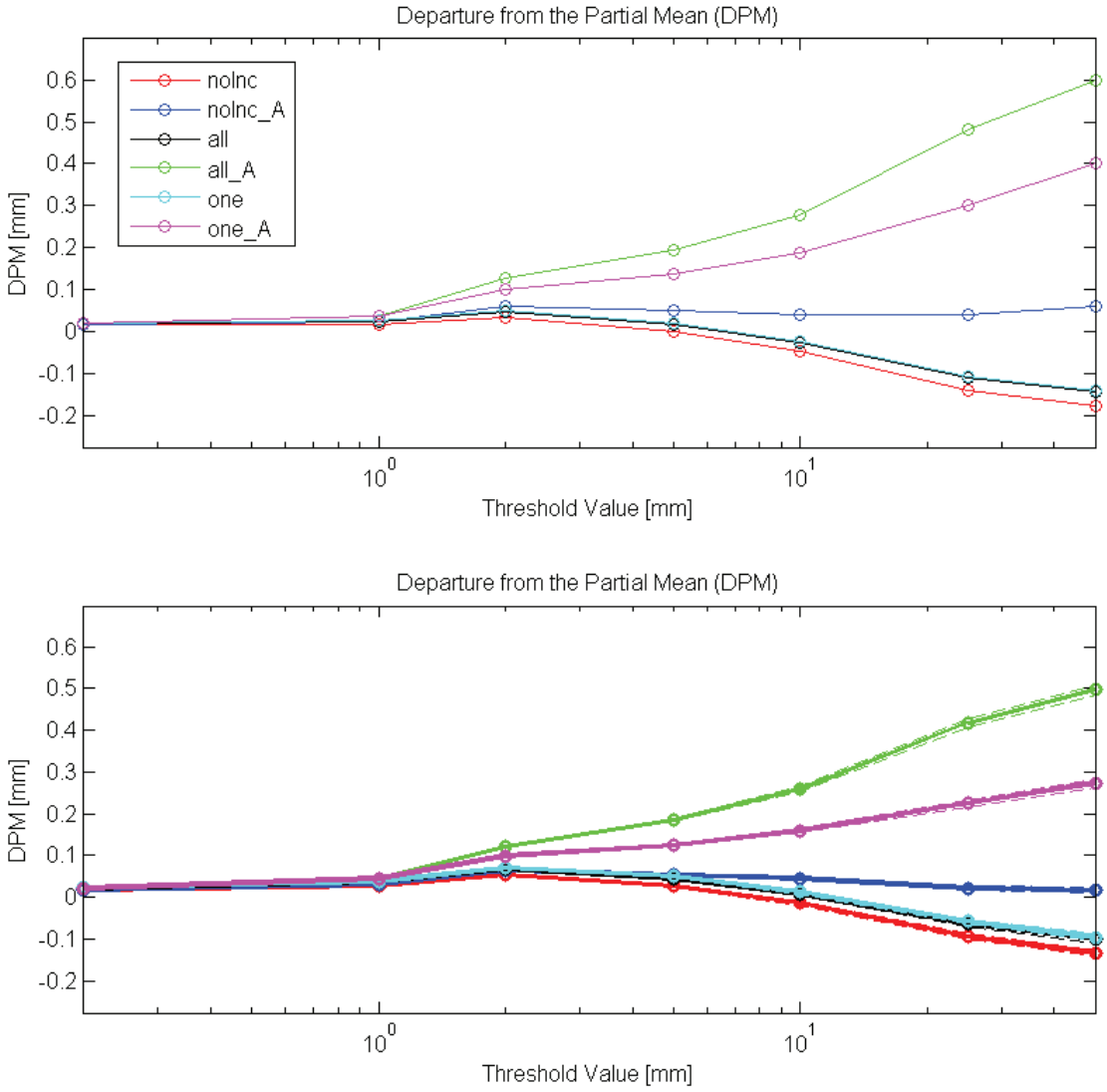

Figure 3.13: Anisotropy investigation DPM results. **Top:** DPM results averaged across all verification locations and all time periods in the study. **Bottom:** DPM bootstrapping results, with the bootstrapped mean DPM values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

but the *all_A* and *one_A* cases continue to have positive FBI results. As precipitation amounts reach magnitudes of 10 mm and larger, the *all_A* case results in absolute FBI values that are notably larger than all other cases.

For all bins, the *noInc_A* configuration performs better than all other anisotropic cases. For precipitation amounts larger than 2 mm, *noInc_A* performs better than the isotropic cases as well. The isotropic cases tend to showcase a similar performance amongst themselves, although for precipitation amounts between 0.2 and 2 mm a slight improvement can be seen in the *noInc* run.

An assessment of the ETS scores presented in Figure 3.15 provides another means of looking into the categorical skill of each proposed semivariogram configuration. All three isotropic cases show similar ETS results, and generally outperform the anisotropic cases across all bin ranges, although the skill improvements over the anisotropic cases for the 25 to 50 mm bin are not statistically significant. The three anisotropic scenarios tend to adhere to a specific ranking in their performance, with the strongest ETS scores attributed to the *noInc_A* case, and the poorest scores belonging to the *all_A* case. In addition to the ETS scores, aETS values are also presented in Figure G.2 in Appendix G. Little difference can be seen in terms of the relative performances of the configurations tested when comparing the ETS and aETS values, although the aETS indicates that the isotropic cases significantly outperform the anisotropic cases even for the 25 to 50 mm bin.

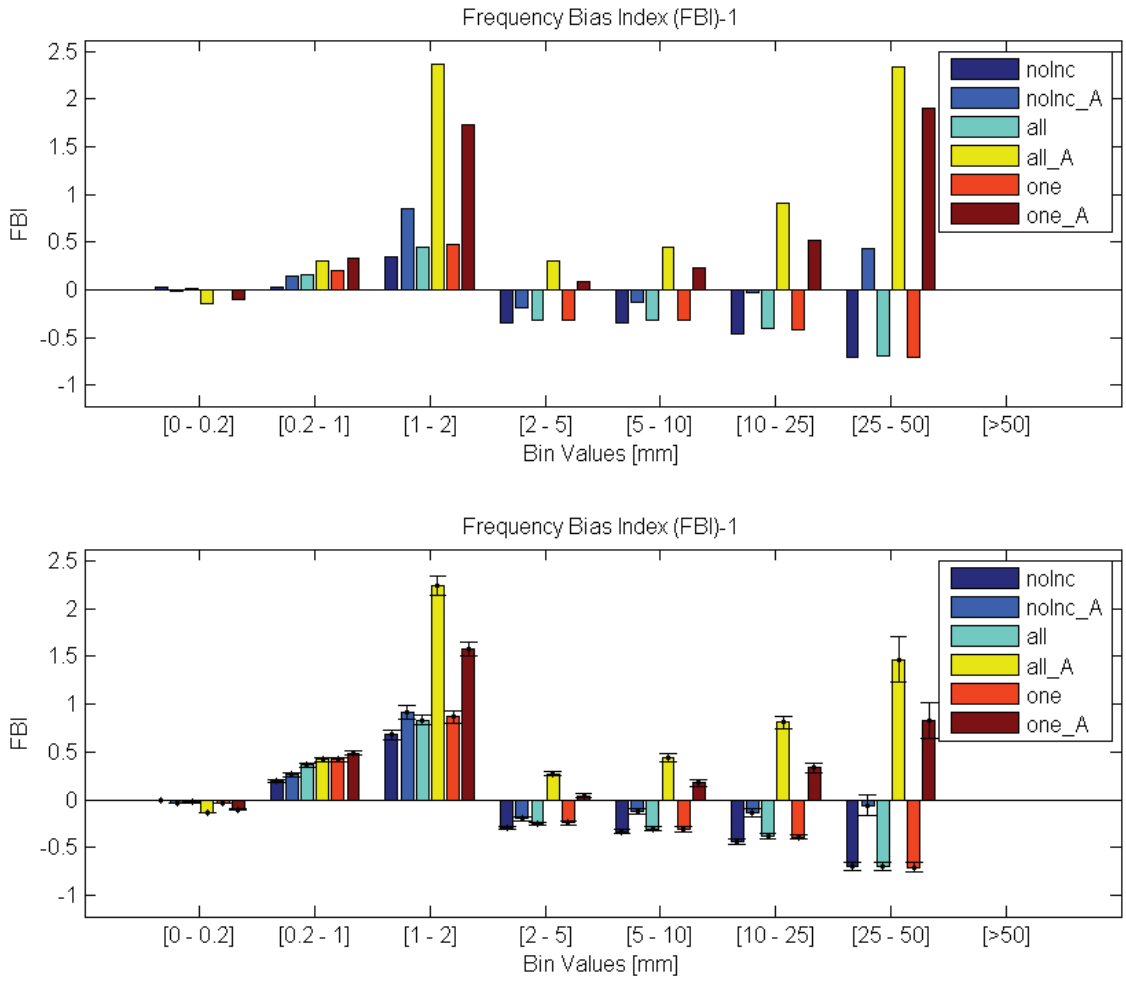

Figure 3.14: Anisotropy investigation FBI results. **Top:** FBI results averaged across all verification locations and all time periods in the study. **Bottom:** FBI bootstrapping results, with the bootstrapped mean FBI values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines

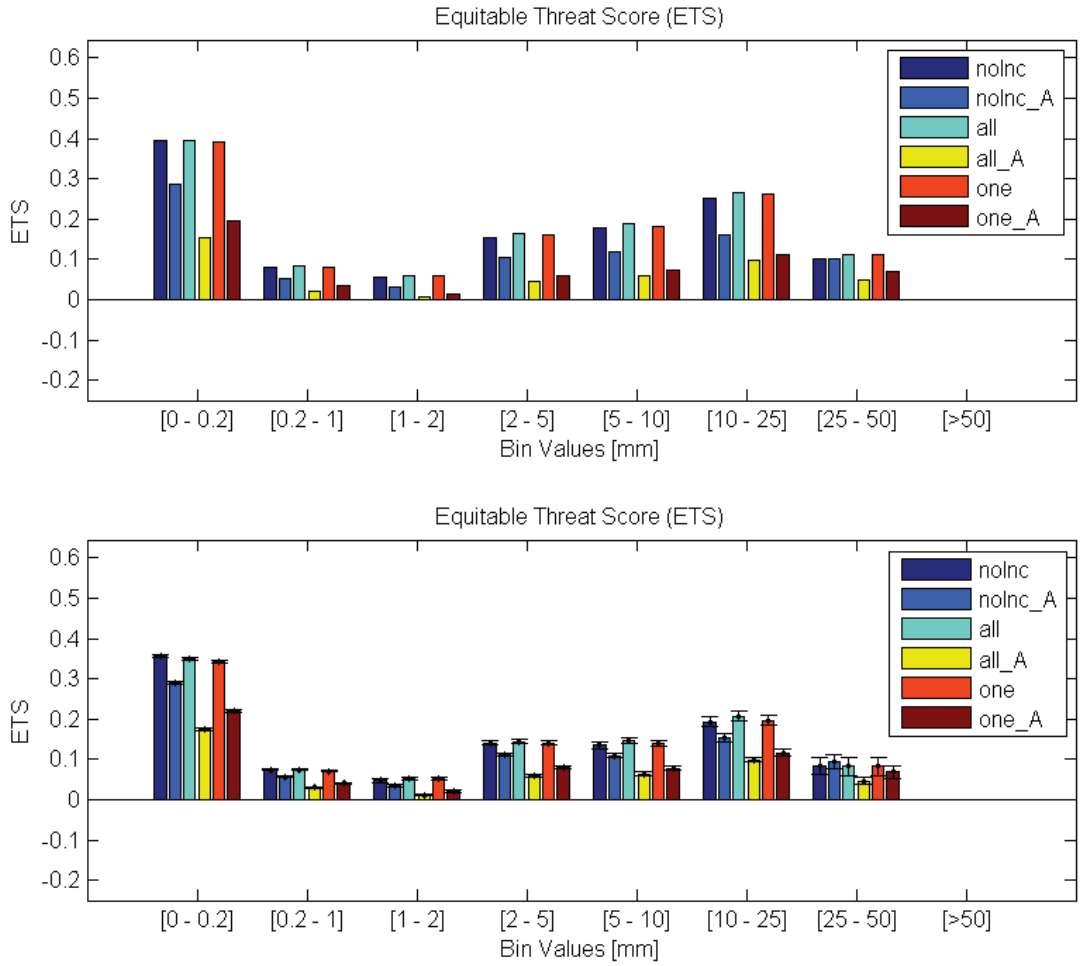

Figure 3.15: Anisotropy investigation ETS results. **Top:** ETS results averaged across all verification locations and all time periods in the study. **Bottom:** ETS bootstrapping results, with the bootstrapped mean ETS values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines

The bootstrapping results quantifying the difference, along with 10% and 90% confidence bounds, between each isotropic and anisotropic configuration pair are presented in Figures G.3 to G.5, provided in Appendix G. The bootstrapping results suggest that the particular results obtained from this investigation may not always be representative of the expected results for the verification metrics, although the basic trends and patterns in the difference are replicated closely. In almost all cases, the differences between the metric values of the pairs are significant, with the confidence lines not encompassing zero except where a transition from a positive to negative difference occurs, or vice-versa. However, in a few cases it can be seen that the differences reported for the largest bin values are not significantly different from zero.

Finally, a quantile-quantile plot is presented in Figure 3.16, which demonstrates the capability of the semivariogram configurations to properly capture the distributional characteristics of the observed precipitation. Generally, the *all_A* and *one_A* scenarios consistently over-predict precipitation amounts, while the *noInc_A* case lies closes to the 45° line and may over-predict smaller precipitation amounts and under-predict large amounts. The isotropic cases all under-predict precipitation consistently, except for amounts approximately less than 1 mm when relatively good agreement with the observed precipitation is seen. It should be noted that the black points do not appear on the plot, because they are overlapped by the green markers.

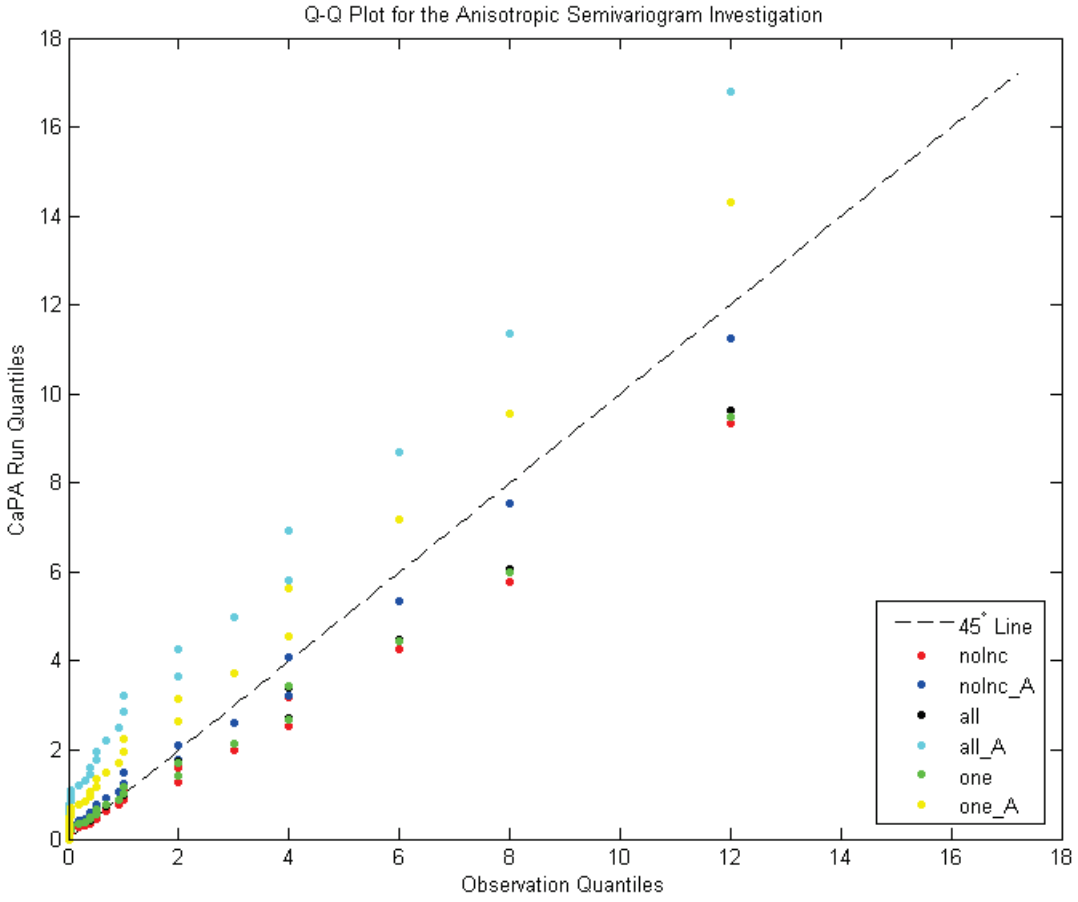

Figure 3.16: Quantile-quantile plot comparing the performance of all CaPA scenarios tested during the semivariogram anisotropy investigation

3.3.2 Convection Filtering Results

The results presented here for the convection filtering technique are based only on those days that have been identified as being the most convective during the study period. The run scenarios considered for the calculation of the verification skill scores are defined as follows:

- 'nc': CaPA setup equivalent to the operational version of CaPA, with a smoothing filtered applied to all semivariogram values; this is the no convection filtering case
- 'nc_A': CaPA setup which includes anisotropy in the 'nc' configuration
- 'c': CaPA setup with a semivariogram smoothing filter applied only to semivariogram values for time periods not classified as convective; this is the convection filtering case
- 'c_A': CaPA setup includes anisotropy in the 'c' configuration
- 'c50': CaPA setup with a standard semivariogram smoothing filter applied to non-convective days, and a modified smoothing filter applied to convective time periods; this is the 50% convection filtering case
- 'c50_A': CaPA setup includes anisotropy in the 'c50' configuration

The average bias results for the convection filtering runs are provided in Figure 3.17. Since only those days identified as being the most convective during the

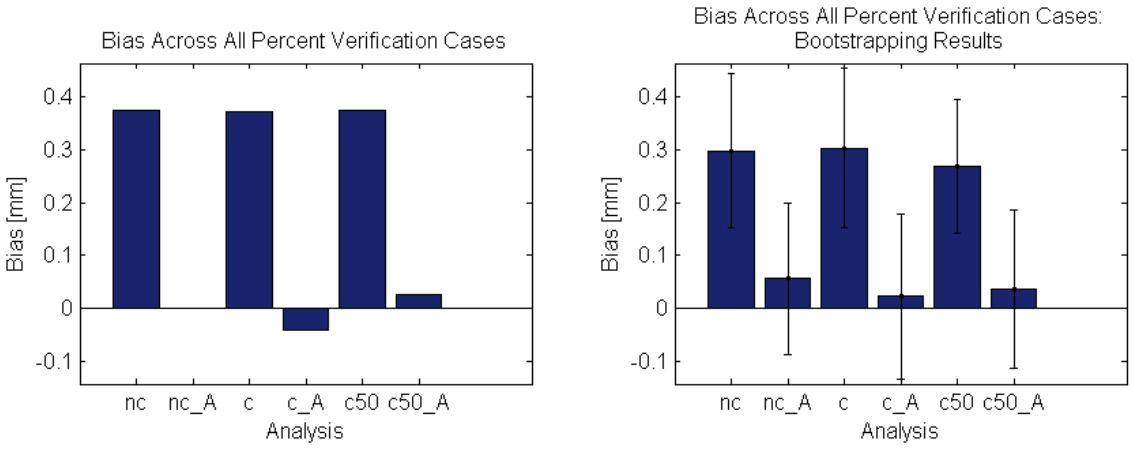

Figure 3.17: Convection filtering bias results for multiple run scenarios averaged across all verification locations and all time periods in the study

study period are considered, these results are for only a subset of the full study time period. This explains how the *nc_A* run can achieve an average bias value approximately equal to zero, despite bias being present when the equivalent *noInc_A* run was analyzed in Section 3.3.1. Based on the bias results represented by Figure 3.17, it is clear that accounting for convection through modified semivariogram parameter filtering does not result in a reduction of the average bias, in either the isotropic or anisotropic cases. This is confirmed by the bootstrapping results in the right panel of the figure.

A summary of the DPM skill scores calculated for the convection filtering runs is depicted by Figure 3.18. The high degree of agreement between the isotropic cases that was seen in Figure 3.17 is echoed, with all three cases plotting almost identically and with no significant difference in their results. All isotropic cases tend towards increasingly negative DPM values as the precipitation threshold increases. This is

the opposite of the trend that is displayed for the anisotropic cases, which generally tend to become increasingly positive in their DPM scores as the threshold becomes larger. However, in the anisotropic case overall values closer to zero are maintained. While a greater amount of distinction between the anisotropic cases can be seen in the DPM results, all differences are once again shown to be insignificant based on the bootstrapping analysis. Similar outcomes are seen for the DPS results, summarized in Figure G.6 in Appendix G.

The strong similarity between the isotropic configurations is once again seen in the FBI categorical skill scores, displayed in Figure 3.19. To a lesser extent, this consistency is also seen amongst the anisotropic results except for precipitation amounts of 5 mm or greater. Generally, the isotropic cases produce smaller FBI scores than do any of the anisotropic runs for precipitation amounts less than 2 mm, while the anisotropic cases produce lower FBI values for the 2 to 5 mm and the 10 to 25 mm bins. Only the *nc_A* case results in a lower FBI score than the isotropic scenarios for the 5 to 10 mm bin. Based on these findings, and the results reported by the bootstrapping bar plot in the lower panel of the figure, it is evident that no significant difference exists in the FBI scores based on the time steps treated with the convection filtering schemes.

The ETS scores for the tested convection cases offer the same story as presented by all previous results; the anisotropic-based configurations all perform relatively equivalently, with only a slight staggering of the values for the 10 to 25 mm bin

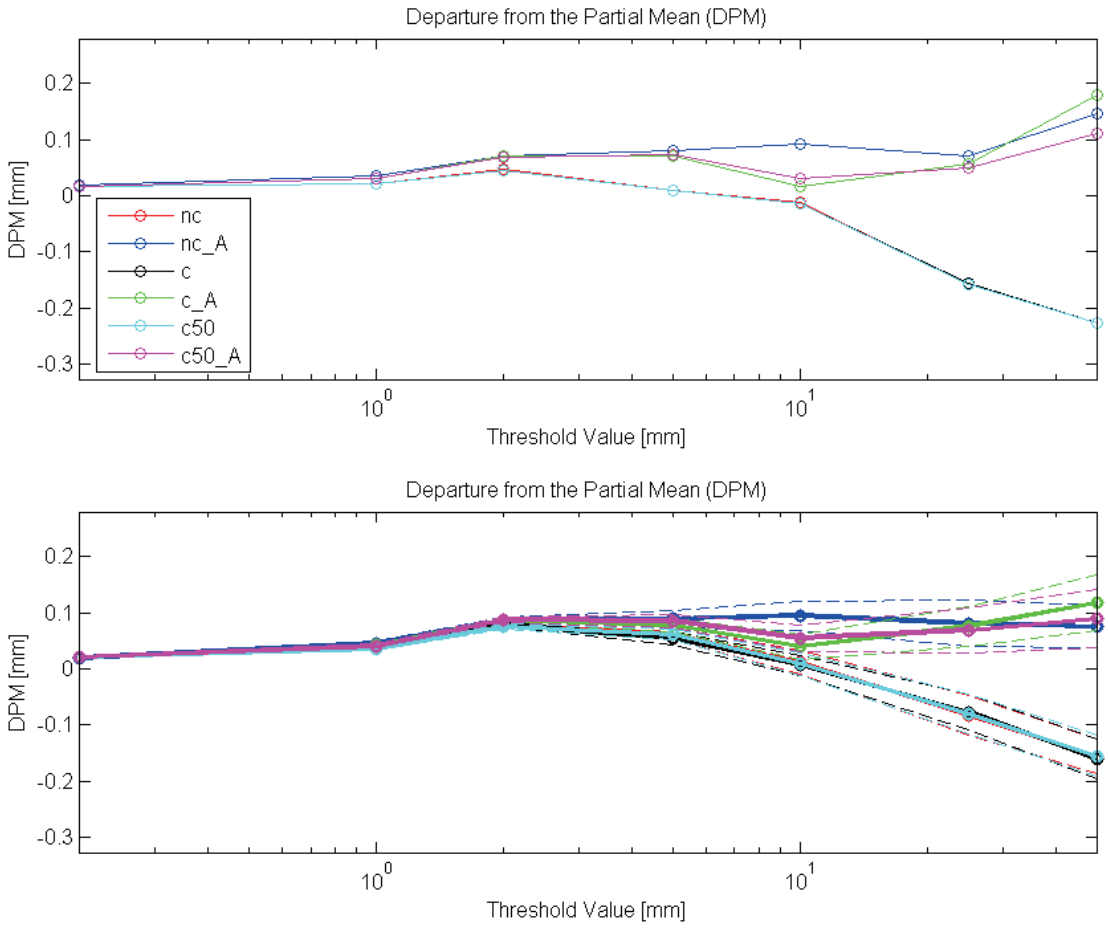

Figure 3.18: Convection filtering DPM results. **Top:** DPM results averaged across all verification locations and all time periods in the study. **Bottom:** DPM bootstrapping results, with the bootstrapped mean DPM values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

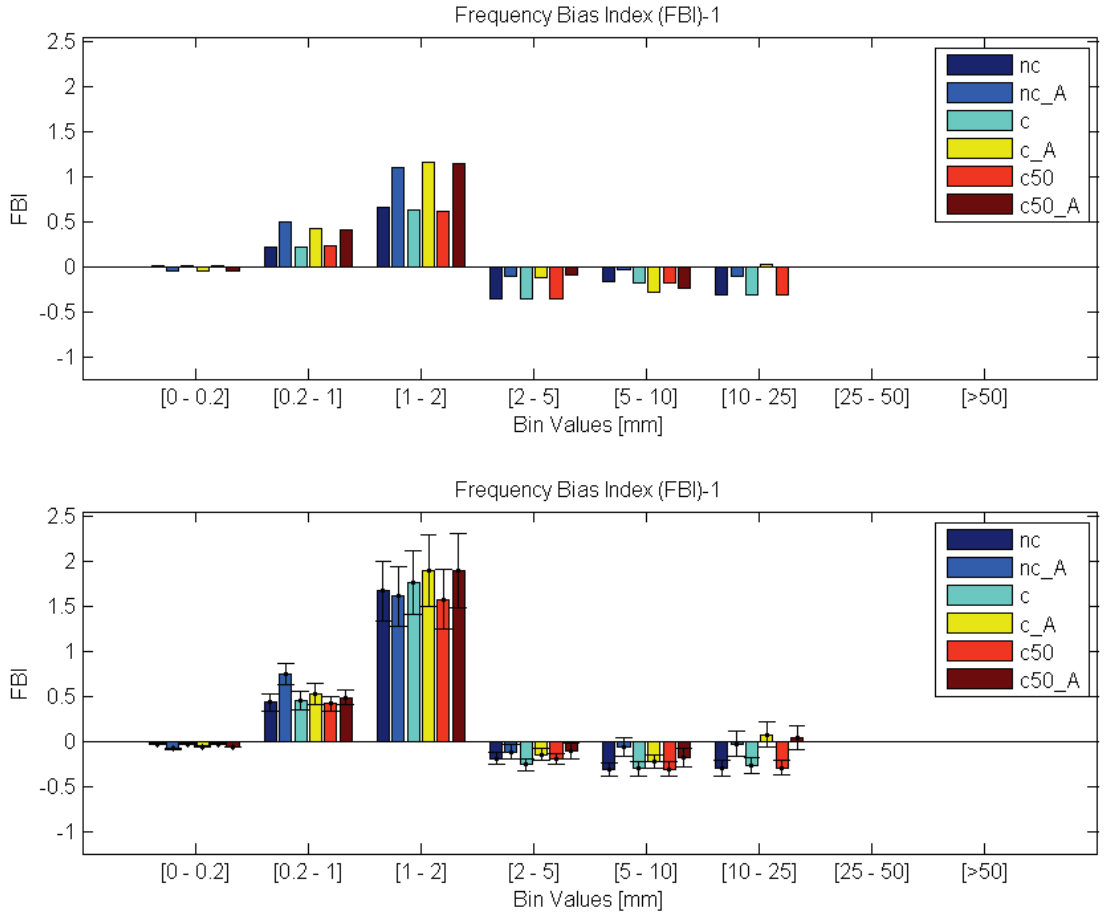

Figure 3.19: Convection filtering FBI results. **Top:** FBI results averaged across all verification locations and all time periods in the study. **Bottom:** FBI bootstrapping results, with the bootstrapped mean FBI values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines. Note that c50_A does not appear for the 10 - 25 mm bin, due to a calculated value visually indistinguishable form zero

according to the following ascending order: *nc_A*, *c_A*, and *c50_A*. The isotropic runs are also almost indistinguishable from each other, and produce the largest ETS values across all precipitation ranges without fail. In both the isotropic and anisotropic cases, bootstrapping results confirm that no statistically significant difference is seen when convection filtering techniques are implemented. The results reported by the aETS scores replicate the same basic relationships, with a stronger dominance of the isotropic cases over the anisotropic runs for larger precipitation amounts. The ETS scores are presented in Figure 3.20, while the aETS scores are depicted in Figure G.7 in Appendix G.

Figures G.8, G.9, and G.10 are presented in Appendix G, and show differencing results between isotropic and anisotropic cases in order to gain a stronger understanding of how anisotropy impacts the convection filtering technique. All of the difference figures show roughly the same trends, which indicates that anisotropy does not consistently produce a significant difference from isotropy for a given convective configuration. The range of precipitation amounts for which no significant difference is seen, based on the bootstrapping results, varies depending on the metric under review. It was also found that the study results did not always lie within the confidence bounds produced through bootstrapping, although in most cases the same basic trends in the curves are seen.

Finally, Figure 3.21 summarizes how well the observed precipitation distribution is captured by the various convection filtering runs. The worst results are obtained

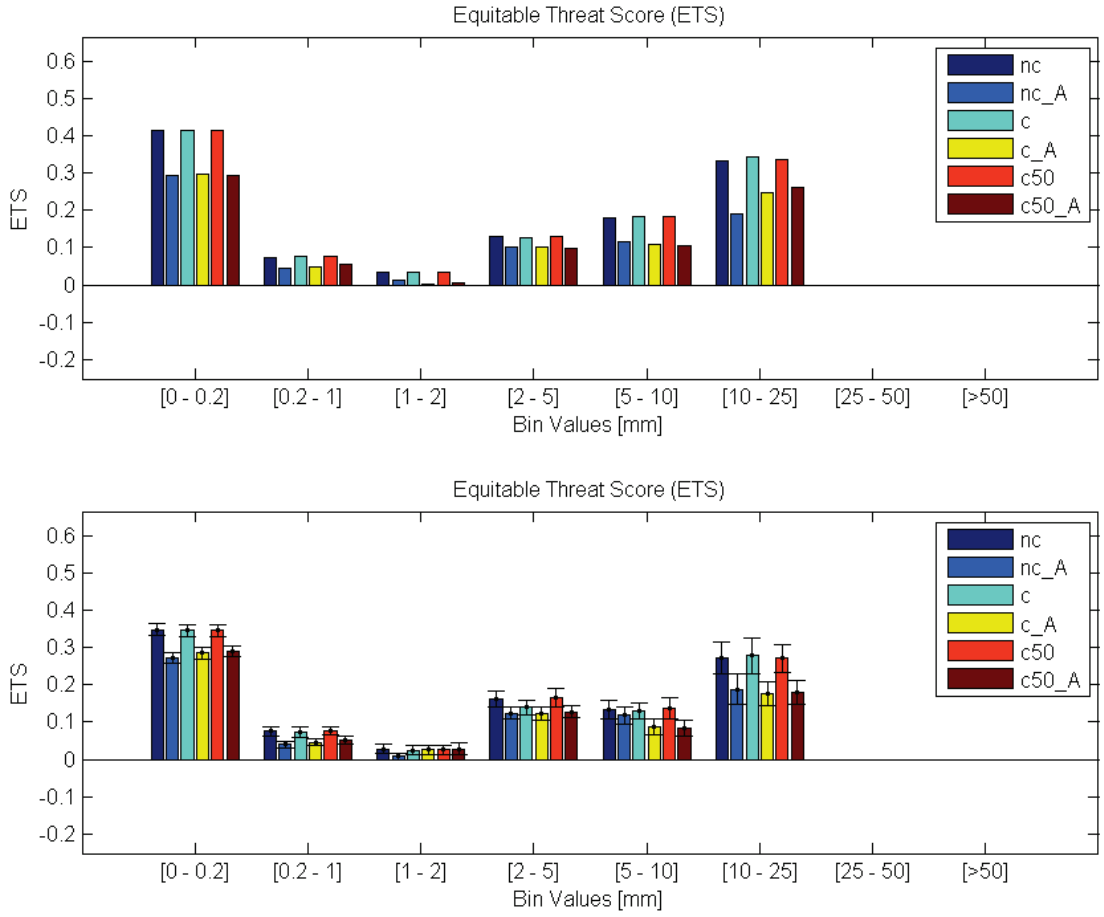

Figure 3.20: Convection filtering ETS results. **Top:** ETS results averaged across all verification locations and all time periods in the study. **Bottom:** ETS bootstrapping results, with the bootstrapped mean ETS values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines

by the isotropic cases, with the under-prediction of precipitation amounts dominating except for amounts less than 1 mm. Within the isotropic context, the case where no filtering modifications are made appears to perform the best, with these points falling closest to the 45° line. When considering the anisotropic results, as precipitation amounts begin to exceed 1 mm, *c_A* and *c50_A* begin to under-predict precipitation overall, while the *nc_A* case continues to over-predict. For precipitation amounts less than 6 mm, the *c_A* and *c50_A* cases outperform *nc_A*, but as larger precipitation events are considered the tendencies to under- and over-predict precipitation become approximately equal in magnitude across all three scenarios. Based on visual inspection of the entire range of values plotted, *c_A* and *c50_A* appear to be the most preferable of all cases.

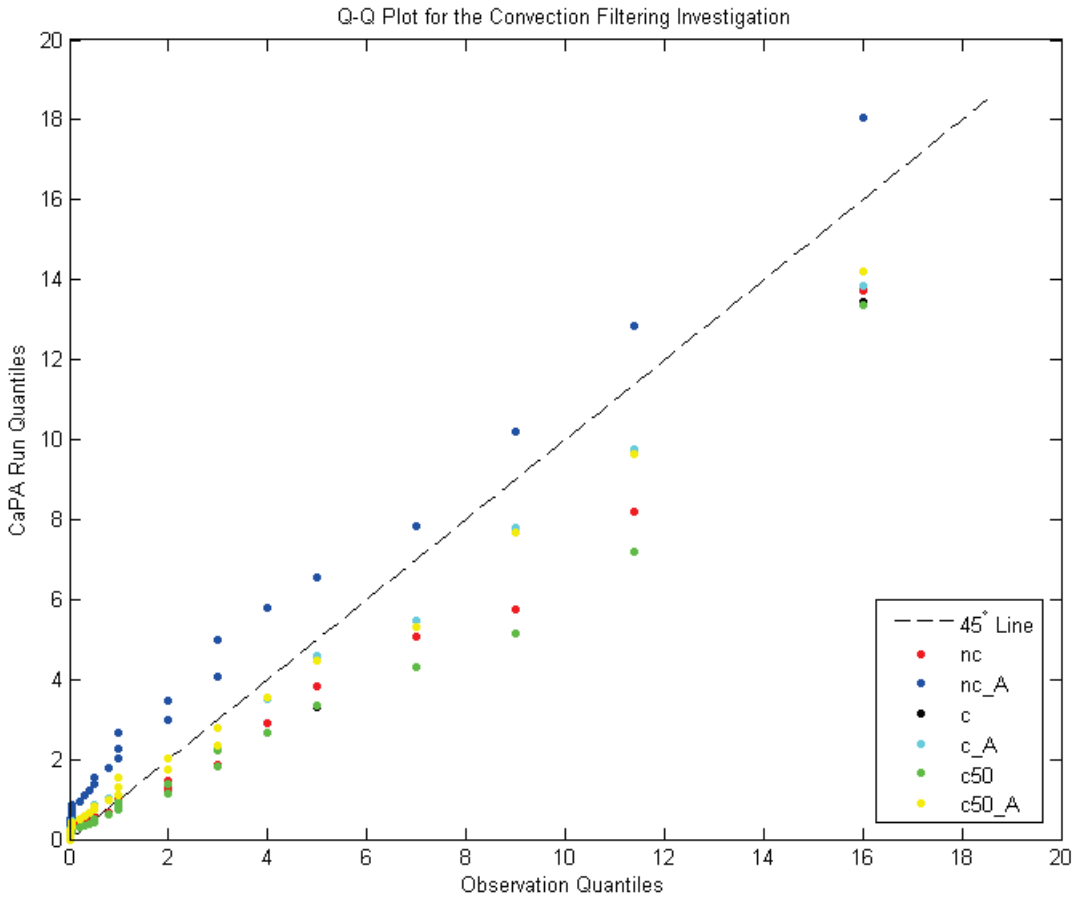

Figure 3.21: Quantile-quantile plot comparing the performance of all CaPA scenarios tested during the convection filtering investigation. The inset figure in the upper left corner is a zoomed-in view of the results for smaller precipitation amounts

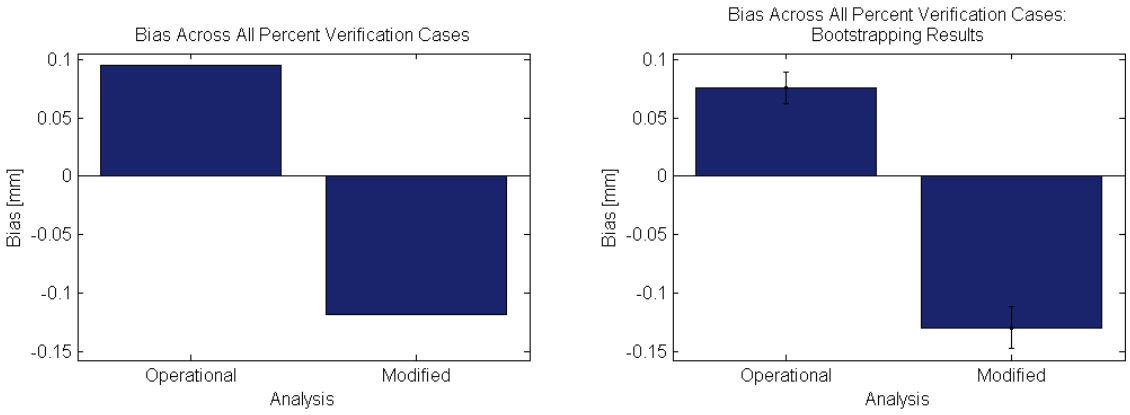

Figure 3.22: Average bias results under the combined influence of MWA scheme II and semivariogram anisotropy averaged across all verification locations and all time periods in the study

3.3.3 Combined Bias Correction and Anisotropic Semivariogram Results

A final appraisal of MWA scheme II and the use of anisotropic semivariograms is provided by the results reported here, for a CaPA configuration combining both of these elements and henceforth referred to as the modified case. The average bias results comparing the operational CaPA configuration to the modified case are depicted in Figure 3.22. It is evident that the combination of the two proposed enhancements to the SI method merge to produce an overall increase in the absolute bias, shifting the bias type from dry to wet when comparing the CaPA analysis to the observation verification data.

The DPM results for the operational and modified cases further confirm the sub-optimal performance of the modified configuration. The DPM scores of the modified

run steadily increase as the precipitation threshold increases. For the operational case, the DPM scores peak around 2 mm, after which they commence a falling trend that leads to final DPM results for the largest threshold values that are negative, and yet possess an absolute value less than the corresponding modified case DPM values. Figure 3.23 illustrates the DPM scores calculated. The DPS scores are consistent with the DPM results, except for a slight lowering of DPS values from 2 to 10 mm, prior to their final increase for the largest thresholds. The DPS scores are shown in Figure G.11, in Appendix G.

Looking at the FBI scores reported in Figure 3.24 reveals that the the modified case may have some bias-reduction merit in comparison to the operational base case; the operational case outperforms the modified scenario for the smaller precipitation events, but for bins encompassing values greater than 2 mm the modified case produces lower FBI scores, with a switch of the sign of the FBI score also seen for precipitation amounts greater than 10 mm.

A review of the ETS scores presented in Figure 3.25 reestablishes the operational case as the pre-eminent setup choice, as ETS values are seen to be larger than those for the modified case for all precipitation magnitude ranges. The aETS results captured in Figure G.12 support the ETS results, and are presented in Appendix G.

Figure G.13, also housed in Appendix G, can be used to check that the differences between the operational and modified cases seen in the previous figures are statistically significant. For the DPS values corresponding roughly to moderate and large

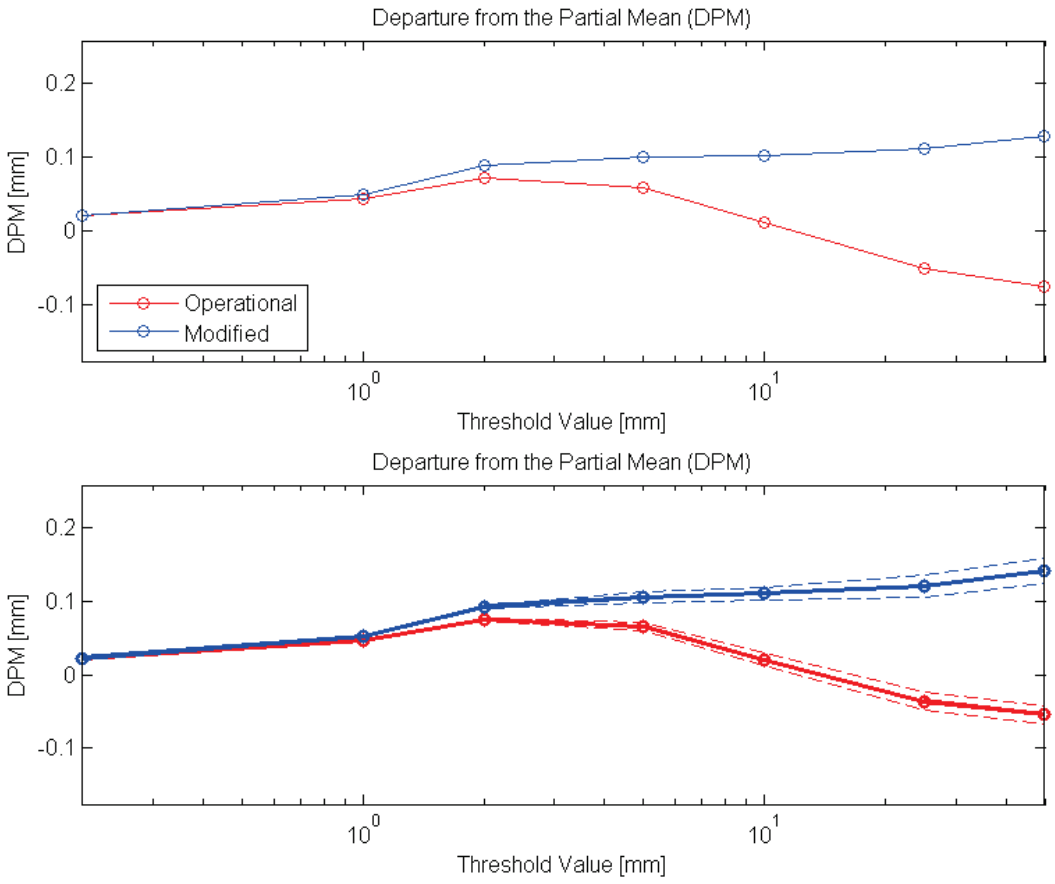

Figure 3.23: DPM results for the combined MWA scheme II and anisotropic semi-variogram case. **Top:** DPM results averaged across all verification locations and all time periods in the study. **Bottom:** DPM bootstrapping results, with the bootstrapped mean DPM values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

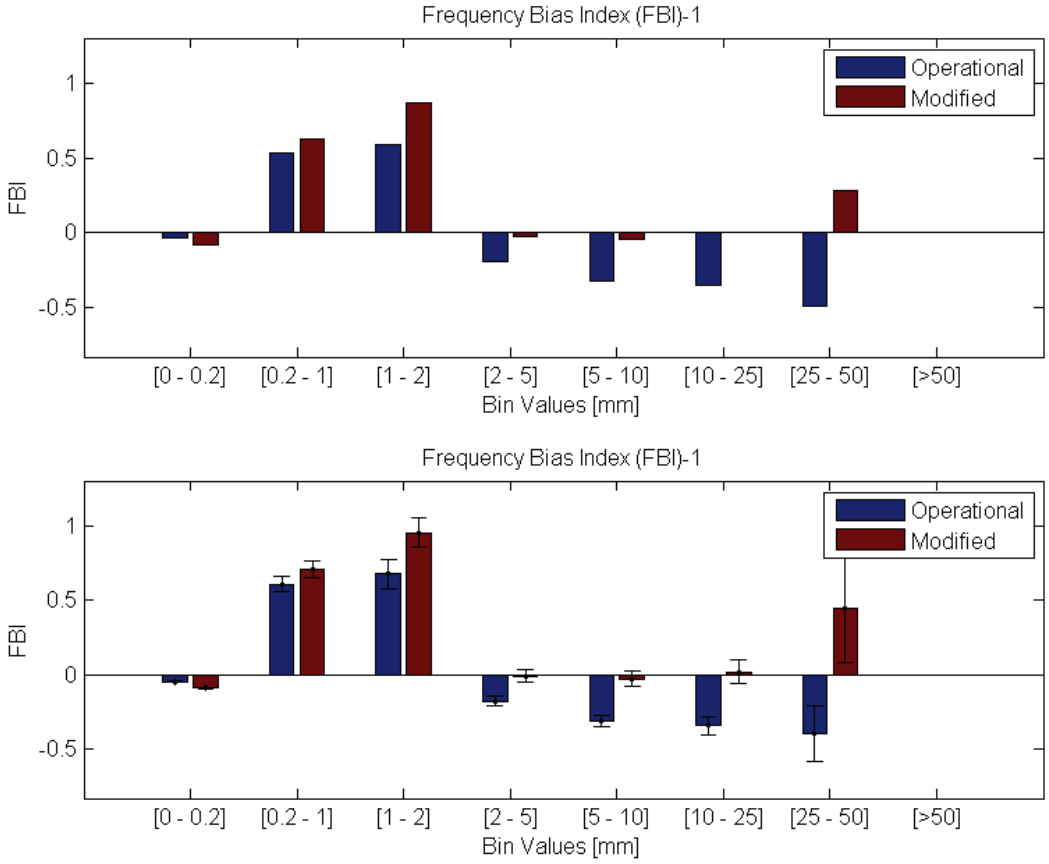

Figure 3.24: FBI results for the combined MWA scheme II and anisotropic semivariogram case. **Top:** FBI results averaged across all verification locations and all time periods in the study. **Bottom:** FBI bootstrapping results, with the bootstrapped mean FBI values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines

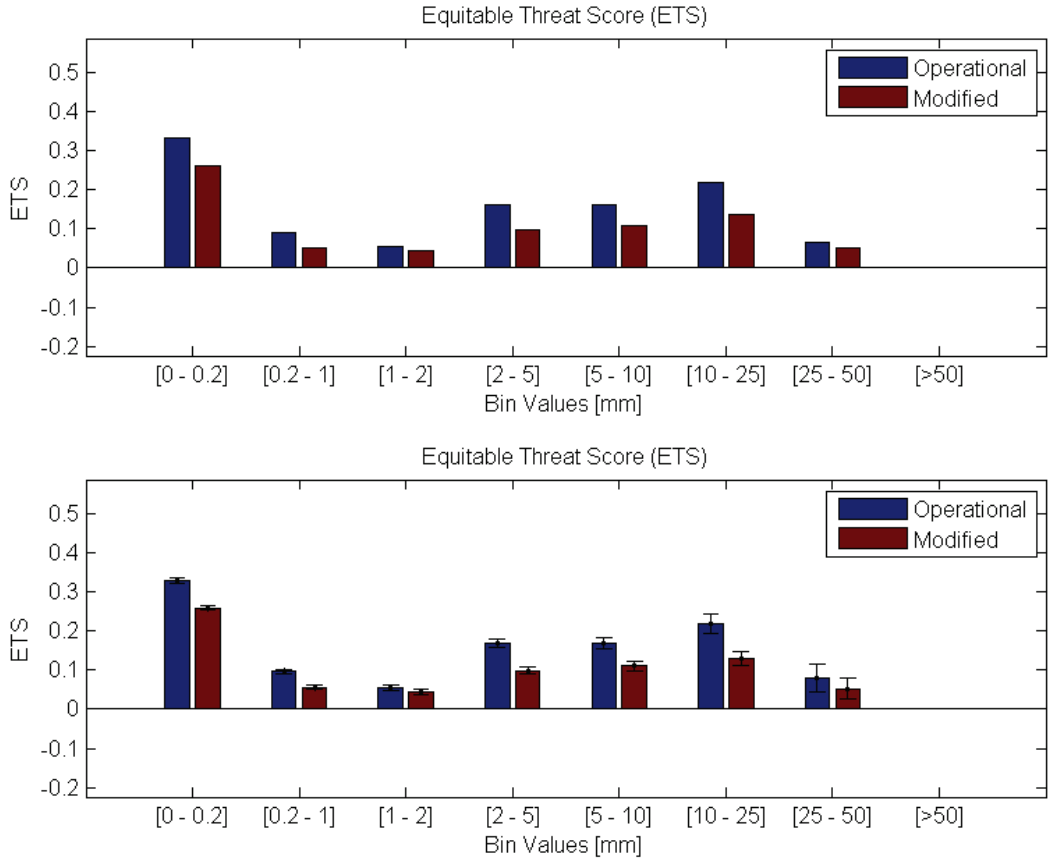

Figure 3.25: ETS results for the combined MWA scheme II and anisotropic semivariogram case. **Top:** ETS results averaged across all verification locations and all time periods in the study. **Bottom:** ETS bootstrapping results, with the bootstrapped mean ETS values shown by the solid coloured bars, and the 10% and 90% confidence limits shown by the thin black lines

precipitation thresholds, and FBI scores for the large amounts, differences between the two cases of interest are not found to be statistically different from zero.

Finally, a direct comparison of the observation and CaPA quantile values is afforded by the plot presented in Figure 3.26. The under-prediction of precipitation by the operational run and the over-prediction by the modified case are both highlighted in the plot. A slightly tighter agreement with the 45° line for the operational case can be identified visually for amounts approximately less than 2 mm, while the modified case shows closer agreement form amounts larger than 2 mm.

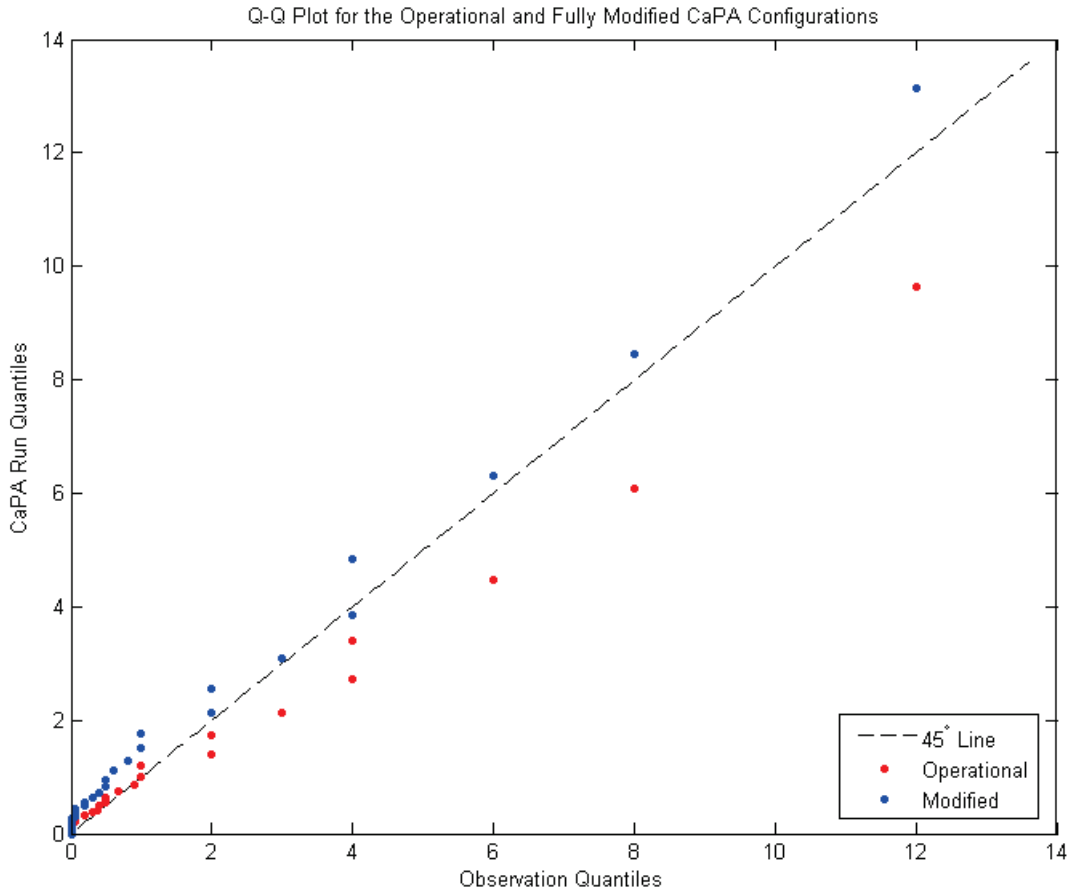

Figure 3.26: Quantile-quantile plot comparing the performance of the operational and modified CaPA configurations. The modified CaPA setup includes the use of MWA scheme II as well as semivariogram anisotropy

3.4 Discussion

The plethora of results provided in Sections 3.3.1, 3.3.2, and 3.3.3 summarily and repeatedly reject the hypothesis that modification of the semivariogram within CaPA, be it through the introduction of anisotropy or convection filtering, can improve the performance of the CaPA program in all aspects. However, this failure of the semivariogram improvement methods to impart widespread benefits is not a total condemnation of their use. Improvements due to the semivariogram modifications were seen under specific conditions, and warrant a deeper look into the results presented.

3.4.1 Anisotropy Implementation

The first point of interest arising from the results of the anisotropy investigation is the notable bias reduction associated with the *noInc_A* case. When comparing the remaining two anisotropic configurations, *all_A* and *one_A*, against the three isotropic runs, it becomes clear that the three isotropic and the two anisotropic cases possess distinct and different bias characteristics; the isotropic runs have lower absolute bias values, and are positive in nature compared to the negative average bias values of the anisotropic runs. While *noInc_A* shares the negative bias value trait of the other two anisotropic runs, it has the lowest absolute average bias value of all runs tested, contrary to the suggestion of the *all_A* and *one_A* runs that including anisotropy may lead to an increase in absolute bias.

The significance of these bias results is confirmed by looking at the percentage of the total precipitation that the bias represents. For the *noInc*, *all*, and *one* isotropic cases, the percentages of bias are 24.39%, 20.54%, and 20.17% respectively. For the anisotropic scenarios, *noInc_A*, *all_A*, and *one_A*, the bias percentages are -6.78%, -75.77%, and -51.48%. Note that the amount of bias for the *all* case, which is the operational configuration, is not consistent with that presented in the bias correction component of this study. This is due to the use of the arcGIS appointed semivariogram. It is evident from these results that the *noInc_A* case has a notably lower bias than all other configurations, and that CaPA analyses in general show sensitivity to the specification of the semivariogram.

The most likely basis for the improved performance of the *noInc_A* case is the lack of limitations that this configuration places on input innovations used to define the semivariogram parameters. Unlike the other two anisotropic cases, the *noInc_A* run defines its semivariogram based on all possible innovation values. The *one_A* case requires that either the observation or GEM value, or both, contributing to a innovation value must be greater than zero. The *all_A* scenario strictly requires that both the observation and GEM values are greater than zero in order for an innovation value to be deemed acceptable.

The benefits of requiring that the input data used to formulate an innovation value be non-zero relates back to the basic concept of the semivariogram. Firstly, precipitation can be divided into two components: occurrence and amount. Under

this system, the forecast first needs to accurately estimate precipitation occurrence for a given location, and if precipitation is predicted to occur, the precipitation amount must then be accurately estimated. It is reasonable to assume that the greatest amount of agreement between observations and GEM is witnessed at locations where precipitation does not occur, as the extra requirement of quantifying precipitation amounts is alleviated in such regions. Thus, under such circumstances an innovation value of zero would be incurred. This is confirmed when considering the study data set, as zero innovation values were produced only when both the observations and GEM reported dry conditions. Such zero innovation values constituted 47.2% of all innovation values for all locations and time steps.

It is generally likely that a significant portion of the analysis domain, for any given time stamp, does not experience precipitation. Furthermore, based on the spatial coherence of precipitation events, it is also likely that large sub-regions within the domain may experience dry conditions, even if precipitation is falling elsewhere in the domain. Because of this, and the zero innovation values that are likely to accompany dry conditions, semivariogram values estimated for locations within a dry zone will likely overestimate the range of the semivariogram; because locations relatively far away may also be dry, the semivariogram estimated for the domain will over-predict the distances at which stations with reasonably well-correlated precipitation amounts can be found. This is in comparison to semivariograms estimated in wet zones, where precipitation and innovation amounts can be expected to vary more significantly in

space and produce semivariograms with a smaller range value. Note that it is also possible for zero innovation values that are sensed at large distances to muddy the semivariogram, as a sudden increase in the correlation with greater distance might then occur.

Since the performance of CaPA is measured in part on its ability to accurately capture precipitation amounts greater than zero, the development of a semivariogram that performs sub-optimally for non-zero precipitation amounts is counterproductive. Thus, excluding cases with zero innovation values assist in rectifying this problem, by disallowing such innovations to negatively influence the semivariogram estimation. This is the logistical premise behind the requirement that either the observation or GEM values possess a non-zero precipitation amount. By still allowing for one of the two contributors to possess a zero value, a slightly larger data set can be used to develop the semivariogram, and the model capabilities at the edges of precipitation events are enhanced. This is an important benefit, as such areas are still zones of interest and should not contribute to the overestimation of the range. This likely explains the strong performance of the *one* scenario, as compared to the *noInc* run.

The *all* case's bias falls between the other two isotropic bias values, likely due to the influence of the increased innovation data restrictions, which require the semivariogram to be estimated based on fewer innovation data points. This likely reduces the probability that the final semivariogram is truly applicable at all locations throughout the study domain. This effect may be slight, but then so too is the difference

in performance seen between the *all* and *one* cases. The percentage of all potential innovation amounts that qualify under the *all* regulations is 20.8%, while the *one* requirements passes 51.5% of all possible innovations.

Despite the described effect of zero innovation values on the estimation of the semivariogram, including zero innovations under anisotropic conditions results in the best performance. This is a quandary best addressed by first considering the data requirements for semivariogram parameter estimation. By introducing anisotropy, two additional parameters are required: the minor range value, and the direction angle. Thus a total of five, rather than three, parameters must be estimated under anisotropic conditions. If the number of parameters being estimated increases, so too does the amount of data required for an accurate estimate. Thus, since the *noInc_A* setup results in the largest amount of data being used during the semivariogram estimation, it is possible that the results are more accurate at the domain scale; fewer total eligible innovations may still allow for accurate localized anisotropic semivariograms, depending on the density of the innovations in that particular area, but are more likely to provide insufficient spatial characterization to capture the generic trends of the entire domain in a meaningful way.

Additionally, the presence of the zero innovation values may prove valuable to the anisotropic semivariogram parameter estimation process in ways that were not applicable in the isotropic case. If the only innovations used in determining the semivariogram are those that do not include zero precipitation values, then zero

precipitation borders which define the spatial extents of precipitation events will not be sensed by the semivariogram. Under such circumstances, it is possible that any directionality in the precipitation borders will be missed, as emphasis is placed firmly on spatial patterns internal to the precipitation weather features. Directionality may still be present in such cases, but it may also be under-represented. By including innovations where at least one contributing zero precipitation value is allowed, the borders of precipitation features may be captured, and anisotropic properties of the precipitation field may be more strongly reflected. However, by allowing all innovation values to contribute, including those based fully on zero precipitation values, all aspects of the spatial configuration of the precipitation can be clearly defined.

This is demonstrated by Figure 3.27, which provides an example of the spatial organization of precipitation across the study area for one time step. As can be seen, directionality clearly exists, with vaguely linear precipitation features oriented loosely in an ENE direction. Most of the innovation locations on the map fall in the dark blue zones, and would therefore likely not be included if only values with no contributing zero precipitation amounts are considered. This would lead to significantly less innovations to base the semivariogram on, and the resolution of the qualified innovations may not be sufficient to estimate the semivariogram values adequately. This is particularly the case if innovations in close proximity to the zero precipitation borders are also discounted due to either GEM or the observation station predicting

a zero amount.

Clearly, if an anisotropic semivariogram is used, it does not appear likely that the correct semivariogram parameters can be identified by focusing only on innovations based on non-zero precipitation amounts. In fact, a potentially more inappropriate semivariogram may result than if a simple spherical model was assumed, which did not attempt to apply incorrect anisotropic spatial characteristics. Thus it appears that the ability to successfully estimate the anisotropic semivariogram is contingent upon both the number of innovations used in the approximation, and the ability of the innovations to reflect the structure of the precipitation events present during the time step. The importance of these two factors appears to outweigh the negative impacts of including zero innovation values.

The bias results presented in Section 3.3.1 also indicate that the incorporation of anisotropy in the semivariogram used by CaPA creates a shift from a general under-estimation of precipitation to an over-estimation by the CaPA analyses. This effect becomes increasingly more pronounced as first innovations are limited to only those where at least one of the GEM or observation contributions are non-zero, and then as all innovations must be based purely on non-zero precipitation amounts. This reversal in the sign of the average bias is likely attributable to the influence of the anisotropic semivariogram on CaPA's ability to resolve precipitation. The anisotropic semivariogram applies the elliptical semivariogram to enhance the influence of innovations lying at greater distances in the major axis directions relative to

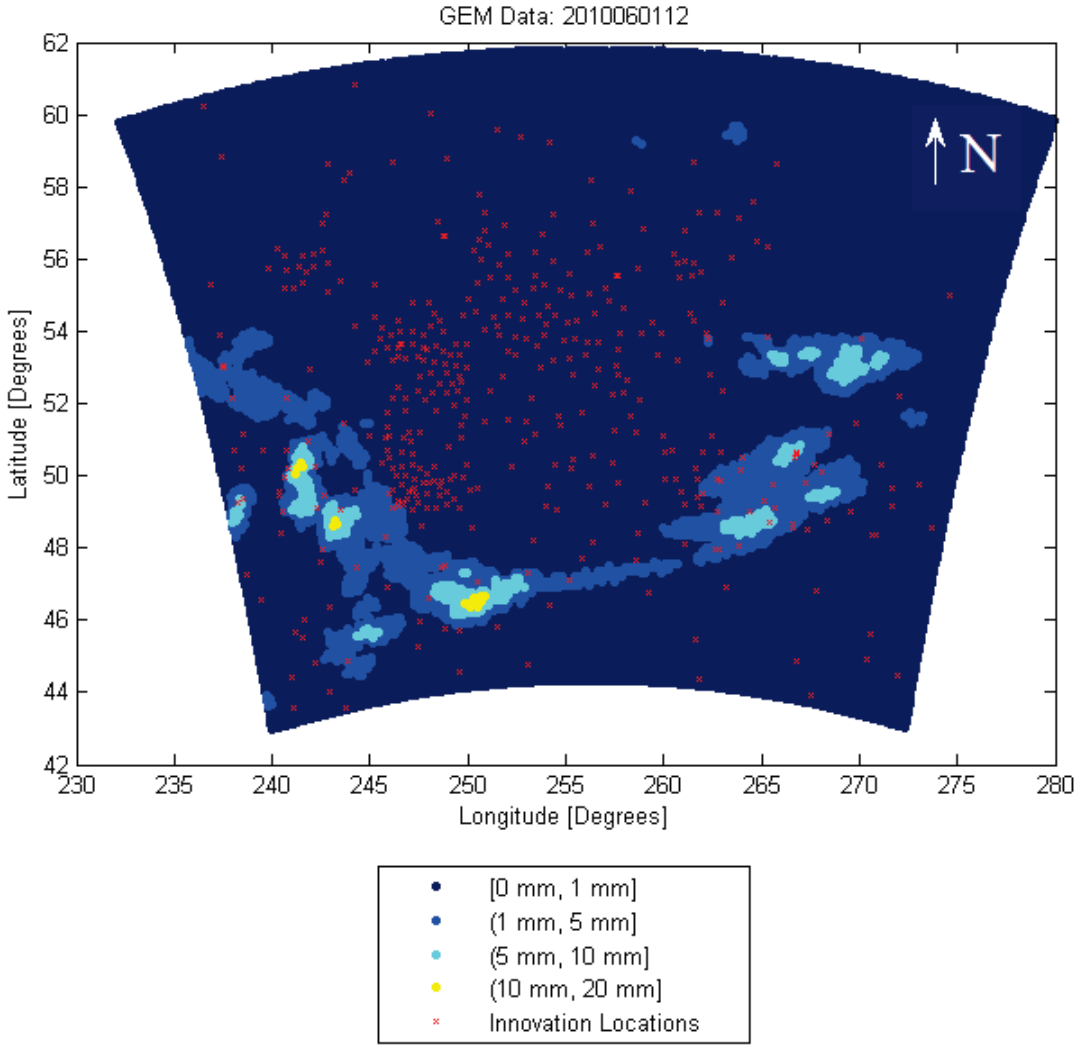

Figure 3.27: Precipitation map of the GEM background field for 2010/06/01/12:00Z

a point of interest, beyond the level of influence that would be experienced under spherical conditions. This can potentially allow for large precipitation amounts to be carried outwards from the main event region.

The problem of the anisotropic configuration propagating larger precipitation events along its major axis is outlined in Figures 3.28 and 3.29, which provide examples from the 2010/06/01/12:00Z time stamp of how the *all* and *noInc_A* cases compare. The *all* case represents the operational case. The *noInc_A* scenario depicts a greater amount of larger precipitation events, as well as more small scale events isolated from the main weather system in areas where in the operational analysis only 0 to 1 mm of rain was forecast. It is also clear that the rough alignment of the sporadic precipitation anomalies present in Figure 3.29 loosely follow the anisotropic directionality that can be seen in the dominant, more clearly defined precipitation systems.

The same effect as was described for the *noInc_A* case can also be seen for the *one_A* and *all_A* runs, but with increasing severity. The culprit behind the degradation of the spatial organization, and general realism, is likely the inability of the anisotropic semivariogram under the restricted innovation eligibility criteria to generate reasonable semivariograms that are applicable for the precipitation patterns seen throughout the entire domain for the entire time step. Figures 3.30 and 3.31 display the *one_A* and *all_A* analysis results. It can also be seen in these figures that vague directionality in the scattered precipitation is present, and approximately

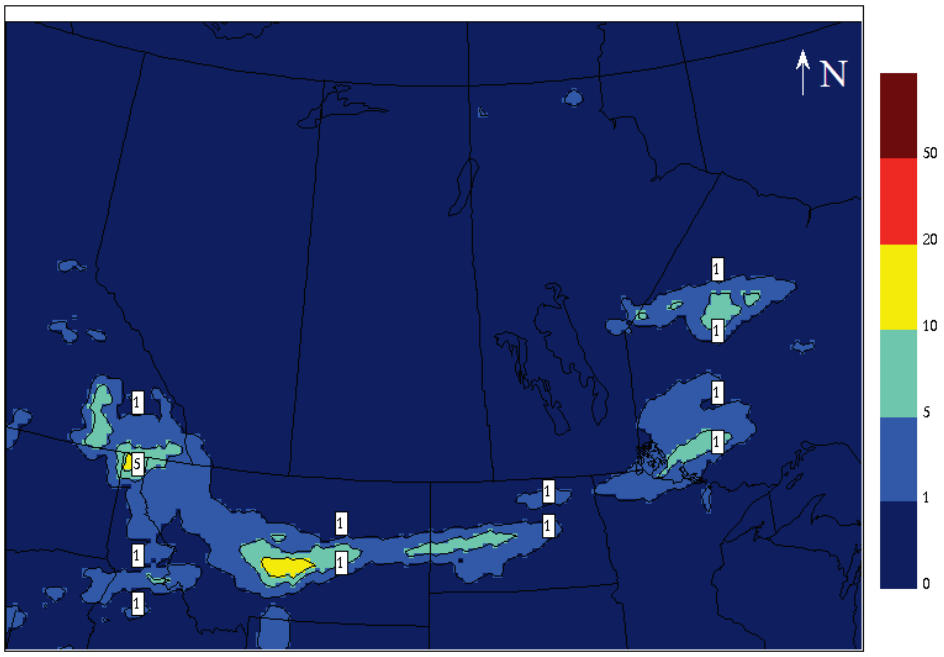

Figure 3.28: Precipitation map of the *all* CaPA analysis field for 2010/06/01/12:00Z in the CaPA projection. The colour legend is in units of mm

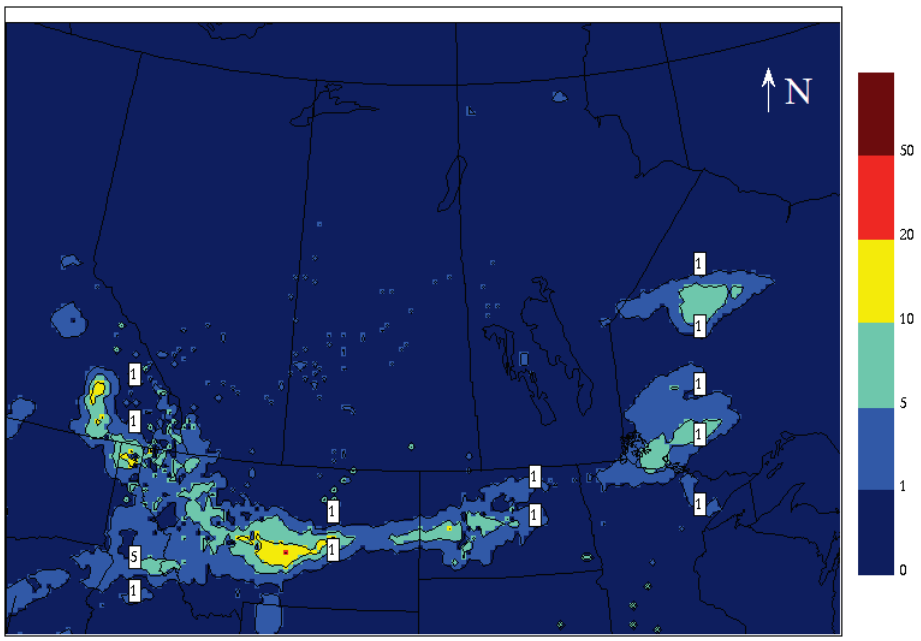

Figure 3.29: Precipitation map of the *noInc_A* CaPA analysis field for 2010/06/01/12:00Z in the CaPA projection. The colour legend is in units of mm

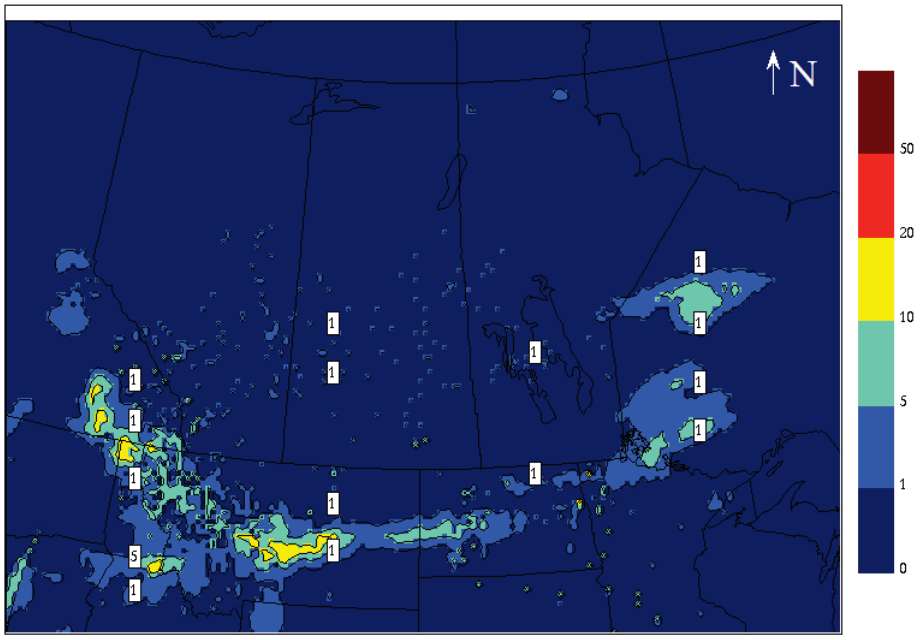

Figure 3.30: Precipitation map of the *one_A* CaPA analysis field for 2010/06/01/12:00Z in the CaPA projection. The colour legend is in units of mm

based on the dominant precipitation structure, but displays more spread than is seen in Figure 3.29.

All of the bias features that have been noted, and the possible explanations that have been attributed to them, can be applied to the DPM and FBI results as well. The DPM and FBI metrics reflect much of the same type of information about the analyses as the average bias. While the DPS score more precisely focuses on the standard deviation statistic, this can also be considered as a reflection of how well the spatial structure of the precipitation events is captured. This comparison is reinforced by the agreement between the performance rankings offered by the DPS, and those derived from the DPM and FBI scores.

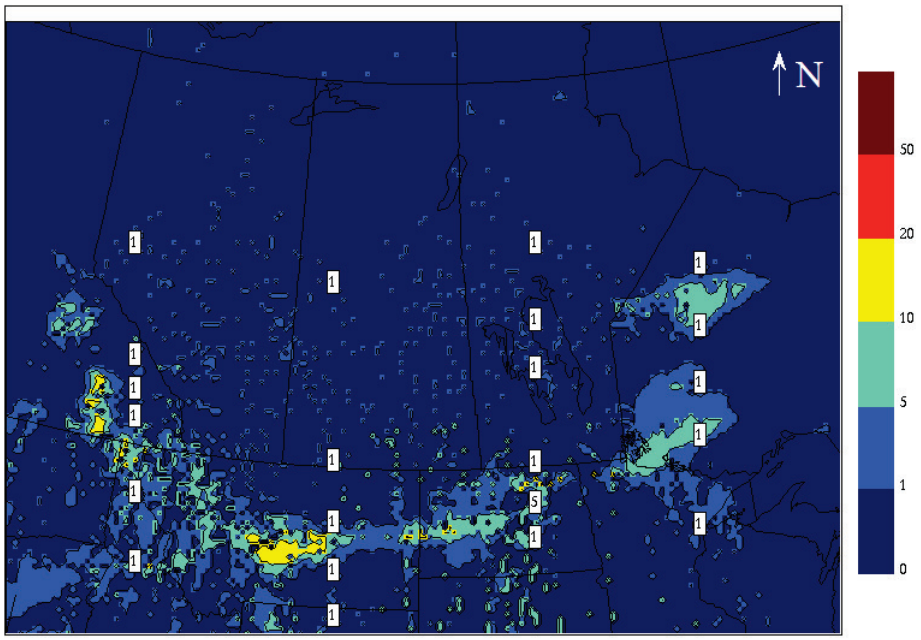

Figure 3.31: Precipitation map of the *all_A* CaPA analysis field for 2010/06/01/12:00Z in the CaPA projection. The colour legend is in units of mm

A quick perusal of the ETS results clearly indicates that introducing anisotropy does not have a positive impact on this categorical skill score. The ETS score is a means of assessing the ability of the CaPA runs to skillfully predict the occurrence of precipitation in a certain range of magnitudes, by looking at the number of hits for a certain precipitation bin as compared to the total number of predicted precipitation occurrences for the same bin from both the CaPA and observations sources, with both quantities undergoing an adjustment for the probability of randomly occurring hits. When the aETS is used, and the impact of bias is removed, the interpretation of the results transforms into a specific estimation of the placement skill of the CaPA analysis (*Mesinger*, 2008). With this understanding of what the ETS and aETS

measure, it becomes clear as to why introducing anisotropy can lead to a reduction in these scores despite improving the bias at least in the *noInc_A* case.

The anisotropic configurations are distinct from the isotropic cases because of their use of an elliptical semivariogram which attempts to capture the dominant two-dimensional structure of spatial correlation in the study domain. The isotropic semivariogram is spherical in nature, and thus is limited to assuming that the correlation of innovation values is equivalent in all directions. While this makes the isotropic semivariogram ignorant of many realistic atmospheric, and potentially geographical, factors that can instigate directional correlation patterns, it may serve to better capture the correlation structure at a specific point than the anisotropic semivariogram.

By assigning an elliptical semivariogram to be applied at all points in the domain, it is likely that at those points where such an elliptical structure is observed the performance of the CaPA analysis will be strong. However, a given elliptical structure will not be adhered to at all locations, as the elliptical parameters are based on the estimated semivariogram and thus on averaged, binned values. For such locations, a spherical semivariogram may avoid incorrectly emphasizing innovation values at large distances along the direction of the major axis; even if the spherical assumption is itself inadequate for that location, it may result in a more reasonable approximation than an elliptical semivariogram with incorrect direction and range properties by more strictly limiting the range.

This anisotropy effect can influence the ETS and aETS scores by creating problems with the placement of precipitation within the analysis domain, even if on average the distribution of the precipitation is more accurately defined across the study region. Thus the importance placed on the ETS and aETS scores by users of CaPA data should vary in accordance with their particular applications, as distributional characteristics may be of greater importance for some modelling applications than others.

An improvement over the isotropic case in the overall distribution of precipitation can still be expected, however, when anisotropy is implemented. This is because the elliptical semivariogram, like its isotropic counter-part, is based on parameter estimation for the entire domain, but with the added benefit of capturing any directional influences. If an elliptical semivariogram is not found to be appropriate on average, then the anisotropic semivariogram can still adopt a spherical shape. This serves as a possible explanation for the poor anisotropic ETS and aETS scores, while the scores that are not placement-sensitive, such as the average bias, DPM, DPS, and FBI scores for amounts greater than 2 mm, show improvement in the *noInc_A* case.

Overall, significant benefits are seen as a result of implementing an anisotropic semivariogram. When comparing to the other cases tested, these advantages include a significant reduction in the absolute average bias, and notably improved performance in the DPM, DPS, and FBI scores, and in the quantile-quantile plot for moderate to large precipitation events. However, the inclusion of anisotropy is detri-

mental to the ETS scores, implying that issues with the placement of precipitation are augmented overall. These results are specific to the *noInc_A* case, however; the *one_A* and *all_A* runs both show no improvement over the current operational setup. Therefore, anisotropy shows promise in the CaPA context only when no restriction on the type of innovations used to estimate the semivariogram are in place, and when the applications of the final analysis require reductions in overall bias or improved predictions of larger precipitation events specifically.

3.4.2 Convection Filtering

The results of the convection filtering tests generally do not support the continued use of this technique in CaPA under the current setup. These results were provided in Section 3.3.2, and generally indicate that little impact occurs when convection filtering is implemented. This is particularly true for the isotropic runs analyzed, all of which were based off of the *noInc* case. This case was selected as it was found to have similar skill scores and attributes as the operational case, *all*, but was the isotropic counter point for the *noInc_A* scenario which was selected as the base case of interest due to its performance in the anisotropy implementation testing. The percentage of total precipitation that is represented by each isotropic bias amount is approximately 27% for the *nc*, *c*, and *c50* cases. This confirms the minimal influence of convection filtering on the average bias scores. The only visually noticeable difference between the isotropic cases arises when considering the quantile-quantile

plot in Figure 3.21, where the *nc* case slightly outperforms the other two isotropic configurations that possess a convective filtering variant. This difference is strongest for the moderate precipitation range, from approximately 5 to 12 mm.

For the anisotropic cases, a marginal increase in sensitivity of the introduction of convection filtering is seen. Almost no average bias exists across the subset of convective time periods selected for the study for the operational configuration, resulting in an increase in bias when the convection filtering is applied. Full filtering was found to produce a slight negative bias, while moderate filtering resulted in a positive bias. The *nc_A*, *c_A*, and *c50_A* cases have bias percentages of the total precipitation of 0%, -3%, and 2% respectively. Bootstrapping results suggest that these differences are not significant. A similar assessment can be made for the DPM, DPS, FBI, and ETS scores; the differences seen are not significant based on the bootstrapping results, but can be identified visually for larger precipitation thresholds and bins. However, the *nc_A* case did show a general tendency to over-estimate precipitation as compared to the *c_A* and *c50_A* cases, which under-estimate precipitation consistently for amounts greater than approximately 3 mm.

It is also interesting to note that, when contemplating the results of Figures G.8, G.9, and G.10, evidence of wider confidence intervals for lower threshold values can be seen as compared to the results for the anisotropic semivariogram study. This is particularly the case for the continuous skill scores. The expectation for the behaviour of the confidence intervals is a widening for larger precipitation amounts

or thresholds, due to the asymmetrical nature of characteristic precipitation distributions which carries through to the results of the SI method (*Erdin and Frei*, 2011). The appearance of wide intervals for smaller thresholds may indicate that the number of stations analyzed for the convection filtering was not large enough to produce robust bootstrapping results. This possibility is not sufficient to discredit the bootstrapping results, which still show a moderate level of agreement with the actual analysis realization assessed for many metrics. It does serve as a reminder, however, that the bootstrapping results are only an approximation of the uncertainty associated with the difference estimates presented, and should be viewed as guidelines.

There are hints in the DPM, DPS, FBI, and quantile-quantile plots that employing convection filtering may be useful for improving specific metrics over specific precipitation ranges. Unfortunately, the potential for anything useful to be taken from the convection filtering technique is negated by the highly targeted nature of these potential improvements, combined with the bootstrapping results which indicate that they are not significant. This conclusion is not altogether surprising. The concept of the convection filtering is based on the distinct spatial characteristics of convective precipitation; convective storms are small scale events, and thus semivariograms derived through a filtering process that includes the influence of time periods dominated by stratiform weather will likely over-estimate the range of a suitable semivariogram. The problem with the convection filtering method is that the impor-

tance of the small scale nature of convective storms likely remains under-estimated, as the method still attempts to fit one catch-all semivariogram for the entire analysis domain.

A more promising environment in which to test the convection filtering method would be one which estimates the semivariogram on a sub-region basis. This would mean calculating multiple semivariograms per time step, based on boundaries derived from climatological patterns, geography, ecological units, or other similar distinguishing traits. Since the semivariograms would be determined specifically for regions with common attributes, the likelihood that they would perform better at all locations within that sub-region increases. Under such conditions, the impact of convective activity on the semivariogram is likely to be more tangible. Depending on the size of the sub-region, locations within a sub-region may experience convection more consistently than in the full domain, leading to a stronger influence on the semivariogram and greater benefits from the convection filtering methodology. Such an approach to semivariogram estimation can be anticipated to have a similar positive impact on the incorporation of anisotropy as well.

3.4.3 Combined Bias Correction and Anisotropic Semivariogram

Both MWA scheme II and the anisotropic semivariogram show improvements over the current operational configuration of CaPA for moderate to large precipitation

amounts, for example over 2 mm, for most metrics except the ETS and aETS. However, it is not prudent to simply accept these improvements blindly as being compatible and additive. As such, the consequences of combining the bias and semivariogram improvement strategies was assessed, with the results presented in Section 3.3.3. The results quickly confirm that the benefits proffered by the two improvement techniques are not complimentary, but rather act in harmony to over-correct the standard analysis in a common way, resulting in the ultimate reduction of the skill of the corrected analysis.

The bias results reported in Figure 3.22 exemplify this over-adjustment issue. As can be seen, the bias of the modified case, which combines MWA scheme II and a semivariogram estimated with anisotropy and zero-precipitation innovations, results in a shift from positive bias to negative. The absolute value of this bias is greater than that produced by the current operational version of CaPA. Given that both MWA scheme II and the use of anisotropic semivariograms individually lead to negative bias, it is not a surprising revelation that when employed in tandem their effect is to produce an even larger negative bias than either experienced individually.

The amplified tendency for the fully modified CaPA analysis to over-predict precipitation is also observable in the DPM skill score set, where it becomes clear that the impact of the fully modified configuration manifests to the greatest extent for moderate to large precipitation values, or amounts greater than approximately 2 mm. However, for precipitation thresholds and bins greater than 10 mm and 2 mm, im-

provements can be identified in the DPS and FBI scores respectively. This suggests that small precipitation amounts, approximately less than 2 mm, are specifically problematic for the modified analysis. When the cumulative bias that arises from the smaller precipitation ranges is combined with that of the larger precipitation bins, the net effect is to produce greater overall biases than those seen when the modified analysis methodology is broken down into its component techniques. The *operational* and *modified* scenarios result in bias percentages of total precipitation of 12% and -15% respectively.

Given that the overall bias, FBI, and continuous skill scores of the *modified* case were all found to be in general agreement with the results seen when the MWA scheme II and anisotropic semivariogram methods were evaluated independently, it was anticipated that no improvements would be seen in the ETS and aETS scores for the *modified* analysis. This hypothesis was upheld by the results depicted in Figures 3.25 and G.12. Both the MWA and anisotropic improvement techniques improve the average accuracy of CaPA analyses, but are inadequate at the local scale with the domain, making the *modified* ETS and aETS results expected. This failure of the *modified* analysis to improve the ETS scores is alleviated slightly, however, by the improvement of the distributional characteristics of its analysis field for amounts greater than 2 mm, which is depicted by the quantile-quantile plot shown in Figure 3.26.

Chasing grander improvements in CaPA analyses by combining MWA scheme II

and anisotropic semivariograms does not prove to be a fruitful endeavour. While FBI scores are lowered for precipitation values greater than 2 mm, and DPS values reflect a more realistic snapshot of precipitation variance in the modified analysis for thresholds greater than 10 mm, similar benefits are also felt when the techniques are used separately, but with the added benefit of reduced absolute bias. Thus, it is not desirable to merge the two methods until further optimization of the techniques occur, but rather to select the method with the characteristics, physical foundations, and final impacts most desirable for the end use of the CaPA analyses treated; if distinct, large scale directionality exists in the observed precipitation field, then anisotropic semivariograms should be selected over the MWA technique.

Chapter 4

Conclusions

This thesis delves into the realm of gridded precipitation quality, through an examination of the CaPA program. CaPA is a data assimilation system that depends upon a SI methodology to combine gridded GEM background data with spatially and temporally irregular observation data, in a way that seeks to minimize the final analysis error variance. The result is a CaPA analysis with potential applications in a range of hydrological, agricultural, weather forecasting, and infrastructure management activities. However, the usefulness of CaPA data products hinges upon the accuracy of its final analyses, which in turn is dependent on the skill of the SI methodology.

In order to tackle the prospect of improving the SI methodology, the current areas of weakness in the interpolation scheme were dichotomized into the following research areas: the introduction of bias as a result of employing a Box-Cox data transformation procedure, and the enhancement of the semivariogram beyond the

current isotropic form. These topics were treated independently on a common domain, consisting of the Canadian prairies. Both investigations focused on the May through August 2010 time period.

For the bias study, two basic methodologies were proposed that built off of the concept of constructing MWA data sets. Using a 30-day window, input MWA data was created and used to fabricate a MWA target analyses, without the need for data transformation. Standard 6 hour analyses were then corrected based on the target analysis, using either MWA scheme I or II. Scheme I corrects a standard analysis through the use of a simple scaling factor applied uniformly across all time periods. Scheme II adopts an iterative correction approach, and adapts the adjustment factor used during each iteration based on the current updated analysis values. In scheme II, each adjustment factor for each time period is also made to be proportional to the analysis error value outputted by CaPA, to reflect the uncertainty surrounding that estimate. In either case, the presence of data transformation bias is reduced by correcting a standard analysis against a target analysis baseline that is exempt from data transformation, although the procedure still attempts to inherit the improved skill historically achieved by implementing the Box-Cox transformation in the first place.

In order to standardize the evaluation procedure, verification statistics predefined by Environment Canada were used to quantify the impact of MWA scheme I and II. A modification to the verification procedure was required, in order to accommodate

the post-processing necessitated by the MWA schemes. This led to the production of four verification cases, each with a unique percentage of synoptic stations removed from the input data set and used in the subsequent verification processes. The 20% verification case was selected for in-depth analysis, due to issues with representative sampling of the GEM data for the 10% case, poor quality MWA input data resulting from the 30% case, and a preference for a smaller total number of observations stations removed for verification.

The results of the 20% verification indicate a decrease in the amount of average bias when MWA schemes I and II are used, as compared to the current operational configuration. The bias that remains is also seen to shift from positive to negative, likely due to both a potential negative bias introduced into the MWA input data through the 20 out of 30 days of record requirement, and the increased dependence on GEM data for the MWA CaPA runs. The greater emphasis on the GEM background field promotes a negative bias due to the tendency for GEM to over-simulate small precipitation events. This is a characteristic which also creates a negative bias due to its influence on MWA innovations, which fail to capture the frequency with which GEM values may actually be larger than the observation values. MWA scheme II outperforms scheme I in overall bias reduction.

Both MWA schemes performed similarly for the remainder of the verification metrics. This includes outperforming the current operational configuration in regards to DPM, DPS, FBI, and in a quantile-quantile plot comparison, particularly for

moderate to large precipitation ranges. However, both schemes failed to improve upon the ETS scores of the operational CaPA program. As such, MWA scheme II is a promising technique for CaPA when user needs are focused around moderate to large precipitation magnitudes, such as amounts greater than 2 mm, and when average bias in the data is of particular concern. This may include hydrological modelling ventures focused on flood forecasting, or those which are conducted over long time scales, such as on an annual basis, given that solid precipitation under-catch biases that are not addressed by the MWA technique are adequately accounted for. Under such circumstance, the total bias can become significant, and total rainfall may be of primary concern, as opposed to the specific timing and placement of precipitation in the basin.

The attempt to enhance semivariogram estimation in CaPA was focused around the assumption of isotropy. Flexibility was introduced into the theoretical semivariogram model, in order to allow for anisotropy. The estimation of anisotropic semivariogram parameters was facilitated by ArcGIS, external to the CaPA program. A smoothing filter was applied to anisotropic parameters for consistency with the CaPA approach. This filter was modified, however, in order to account for the impact of convective precipitation on the spatial relationships of the innovation data. Time periods identified as possessing convection were either not filtered, or treated to only partial smoothing. The designation of a convective time period was based on a combination of assessments that explored the standard deviations of the obser-

vation and GEM precipitation fields, the coverage of lightning, and the coverage of convective precipitation on a per time step basis.

CaPA analyses synthesized with the modified semivariogram parameters were evaluated using the standard verification techniques without the need for verification cases. The results of the verification tests indicate that the use of anisotropy improves CaPA's performance in regards to all metrics for moderate to large precipitation magnitudes roughly greater than 2 mm, except in the case of the ETS. This is only seen, however, when precipitation occurrence limitations on the innovations are removed, and all innovation values are used during parameter estimation regardless if zero-precipitation values occur. The lack of limitations on the innovation data during estimation is likely the key to the strong performance of the anisotropy case; allowing all innovation values to contribute increases the number of data records that can be used to estimate anisotropic semivariogram parameters, and helps to more clearly define all aspects of spatial precipitation patterns, including in regions bordering and external to precipitation zones in the domain. The transition from under to over-estimating precipitation in CaPA was also identified as a consequence of including anisotropy. It is suspected that this is due to the increased ability of an anisotropic semivariogram to propagate innovation values characteristic of heavy precipitation outwards into low to zero precipitation regions. This also serves as an explanation of the poor placement skill of the anisotropic semivariogram, which is captured by the ETS score.

The use of convection filtering on the semivariogram parameters did not yield promising results. Little to no significant difference was seen in the verification scores of convective days, based on whether or not convection filtering was employed. When anisotropic semivariograms were tested with convection filtering, a slightly greater degree of sensitivity to the method was witnessed, although the differences between scenarios were still not significant. The domain-wide application of a single semivariogram per time step is thought to be the cause of CaPA's indifference to convection filtering. The estimation and application of multiple semivariograms on reduced spatial domains, such as sub-regions within the full CaPA domain, is expected to provide a more suitable environment in which to test the benefits of considering convection during semivariogram specification.

Combination of the bias correction and semivariogram techniques also failed to improve upon the current operational version of CaPA. This was likely due to similarities in how each technique modifies CaPA output, which results in over-correction for many precipitation magnitudes. Thus, independent application of the techniques is advisable.

Much like MWA scheme II, the anisotropic semivariogram enhancement is particularly adept at improving upon skill scores for moderate to large precipitation magnitudes, or generally magnitudes larger than 2 mm, with the ETS scores once again being the exception. Thus, the set of applications for which the bias correction methodology is beneficial is in essence the same as the applications for which

an anisotropic semivariogram can be recommended. In order to distinguish between the two alternatives, the prevalence of anisotropy in the input data of a user should be assessed and then handled through the appropriate technique. In either case, two methods for improving the performance of the existing SI methodology have been developed and outlined, and offer a means of increasing CaPA data accuracy and reliability for users intent upon obtaining high quality estimates of moderate to extreme precipitation amounts, and who are not primarily concerned with the placement skill of CaPA inside the analysis domain.

Recommendations

The procedures and findings discussed in this thesis open the door to further research pursuits. The following recommendations reflect the areas of further study that have been identified as the most promising throughout the course of this investigation:

1. *Implementation of the MWA and anisotropic semivariogram analyses in a hydrological model*: The importance of viewing the impact of the MWA bias correction methodology and the semivariogram improvement technique in the context of hydrological modelling is unquestionable. A primary driver for the creation of CaPA analyses is their use in modelling applications; directly testing the consequences of the two streams of improvement investigated here will more clearly illuminate the true effects of the proposed modifications for end-users dependent on precipitation data in the modelling community. Previous

studies have already shown that hydrological models can display sensitivity to the quality of their inputs, even if the shifts in quality sensed by the model are purely attributable to bias correction techniques. For example, evidence of this sensitivity materializing in both the calculated errors and the variability of model results in an ensemble scenario has been noted (*Teutschbein and Seibert*, 2012). Therefore, it is anticipated that model-based testing of MWA scheme II and anisotropic semivariograms will yield valuable insights into their practical importance. Special attention should be paid to the impact on moderate to extreme events, as these magnitudes benefited the most from the proposed improvement techniques.

2. *Continued Investigation into Bias Correction Approaches*: The bias correction methodology explored in this study considered only data transformation bias. Based on the results obtained, it is clear that other sources of bias persist in CaPA. As such, the following general bias treatment options should be entertained: the use of a holistic bias treatment method, explicit treatment of specific bias sources outside of data transformation, and potential modifications to the MWA technique to enhance its performance. Work is currently being done regarding the improvement of the current bias correction factor scheme, as well addressing solid precipitation biases. The optimal use of a combination of bias correction methods should also be explored.

   An alternative method which may prove valuable for the bias correction of

CaPA analyses is a non-linear approach to MWA scheme II. Currently, iterative corrections to a standard CaPA analysis under MWA scheme II are based on relative adjustments directly proportional to analysis errors. A deeper look into optimizing the relationship type between adjustment factors and analysis errors at a grid point may prove fruitful. An additional amendment to MWA scheme II which should be evaluated is the use of scaling limits, in conjunction with adjustments proportional to the analysis errors. These scaling limits would ensure that during every correction iteration, the magnitude of the correction does not reclassify the precipitation amount, for example in regards to the bin delineations used during the verification processes. This would help ensure that no particular estimates are over-corrected, and may also assist in preserving the ETS skill achieved by the current operational configuration of CaPA. Similarly, only applying MWA corrections to precipitation amounts for which benefits have been demonstrated is also a promising approach.

3. *Investigation into alternate semivariogram types*: Consideration should also be given to alternative forms of the theoretical semivariogram. Currently, an exponential semivariogram is used. However, the preliminary investigation into the analytical model choice for the semivariogram revealed little cross-validation difference between models. Despite this, visual assessment of the anisotropic semivariogram maps in ArcGIS hinted at differing levels of sensitivity of the models types to potential anisotropy. As such, an investigation into the optimal

type of semivariogram model to be used in CaPA is recommended.

Another possibility that deserves evaluation is the use of nonparametric semivariogram specification procedures. For example, the experimental semivariogram data can be calculated from the innovation data set, and then using a Fast Fourier Transform (FFT) technique, this information can be converted from its original spatial domain into a frequency domain (*Velasco-Forero et al.*, 2009). At this point, smoothing of the covariance matrix can be done to ensure that the rules for a proper probability density function are followed, which require that the sum of the values equate to one and are greater than zero. The inverse of the FFT can then be applied to regain the spatial domain. This ensures that a positive definite covariance matrix is achieved, without the use of any analytical model. Such a technique benefits from avoiding the issue of selecting an appropriate theoretical model, and automatically incorporates anisotropy (*Velasco-Forero et al.*, 2009).

Finally, zonal anisotropic semivariograms should also be evaluated for their potential benefits, as only geometric anisotropy was considered here.

4. *Estimation of semivariograms on a sub-region basis*: The addition of anisotropic capabilities, or the use of convection filtering, is likely limited due to the specification and use of only one semivariogram domain-wide. The benefits of estimating the semivariogram for sub-regions, or in a more localized context should be studied. Sub-regionalization of the semivariogram may be done based on a

number of attributes, such as geography, atmospheric patterns, or eco zones. This may allow anisotropic semivariograms in CaPA to become more powerful in their ability to characterize spatial relationships between the innovation data, and is anticipated to yield more meaningful semivariograms in the convection context as well. The small scale nature of convection lends itself to small scale estimation of spatial relationships. Attention should also be given to the quality control procedures at work in CaPA, to ensure that large, valid observations resulting from convective events are not being rejected by the program (*Fortin*, 2012–2013), as this may influence the perceived efficacy of convection filtering even if conducted on semivariograms for smaller regions.

5. *Use of anisotropic semivariograms only under anisotropic conditions*: Due to the large degree of agreement that exists in how the bias correction and anisotropic semivariogram methods proposed by this thesis improve the skill of CaPA, a decision on the precise type of modification to implement should be made based on the specific needs of the user. If the user possesses a data sat that displays strongly anisotropic spatial correlations in the innovation data, then the anisotropic semivariogram should be put to use. If strong anisotropy does not appear, use of anisotropic semivariograms may not be warranted, and the bias correction methodology may better serve the user. It is important to note that no direct comparison was made between the two improvement techniques, due to the use of different semivariogram estimation methods. It

is recommended that such a comparison be undertaken.

6. *Testing of improvement methods under improved and varied input data scenarios*: The study time period and domain for this investigation were restricted to the months of May to August, 2010, for the Canadian prairies. The performance of the suggested methods should be tested for alternate time periods and study domains, to ensure that any specific hydrological regimes or geographical influences which may have been under-represented in the selected study setup are assessed. It particular, regions lying farther north or in more arid climates may be of interest. While the 2010 data year experienced both wetter and drier than normal conditions for various months and locations throughout the domain, excessively wet or drought-type conditions should be tested. Application of the methodologies to winter months should also be considered, with due acknowledgment of the pre-existing bias issues of solid precipitation measurement.

   The performance of both methods should also be tested under improved input data resolution conditions. This is specifically in reference to the observation data, which can limit the number of innovations used to estimate semivariograms and the quality of MWA inputs during the data transformation bias correction. Improved observation density may be obtainable by shifting the study domain or time period, or by continuing to acquire additional gauging sites and observation networks beyond those currently contributing to

the CaPA input data. The incorporation of radar data, which has now been instituted in experimental versions of CaPA, may assist with this issue. Regardless of the means through which the observational resolution is enhanced, improvements in occurrence accuracy can also be expected, as a clearer picture of the extents of precipitation events will result.

Finally, from a more practical stand point, the implementation of the MWA and anisotropic semivariogram methodologies in an experimental version of CaPA is also recommended. The methodologies proved to be the most beneficial for larger precipitation magnitudes, and specifically showed improved bias characteristics. This supports the implementation of the two modifications in CaPA, so that users with interests targeted to these attributes of precipitation data quality may benefit. However, due to the inability of the two methods to improve the ETS skill scores, it is not recommended that these modifications make their way into the operational version of CaPA.

The above recommendations, in conjunction with the conclusions drawn from the work presented, provide the context in which the importance of this thesis can be viewed. The community of gridded precipitation data users can be expected to benefit from the methodologies proposed herein, insofar as their needs align appropriately with the strengths of the methods. Following through on the outlined recommendations will serve to reinforce the practical advantages of implementing both the bias correction and anisotropic semivariogram methods.

Bibliography

Balsamo, G., J. Mahfouf, A. Bélair, and G. Deblonde (2009), A land data assimilation system for soil moisture and temperature: An information content study, *Journal of Hydrometeorology*, *45*.

Berg, P., H. Feldmann, and H. J. Panitz (2012), Bias correction of high resolution regional climate model data, *Journal of Hydrology*, *448*, 80–92.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *Journal of the Royal Statistical Society*, *26*, 211–252.

Brandes, E. A. (1975), Optimizing rainfall estimates with the aid of radar, *Journal of Applied Meteorology*, pp. 1339–1345.

Burrows, W. R., and B. Kochtubajda (2010), A decade of cloud-to-ground lightning in Canada: 1999-2008. part 1: Flash density and occurrence, *Atmosphere-Ocean*, *48*, 177–194.

Chang, Y., N. Hsu, and H. Huang (2010), Semiparametric estimation and selec-

tion for nonstationary spatial covariance functions, *Journal of Computational and Graphical Statistics*, *19*, 117–139.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth (1998), The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulations, *Monthly Weather Review*, *126*, 1373–1395.

Cressie, N., and D. M. Hawkins (1980), Robust estimation of the variogram: I, *Mathematical Geology*, *12*, 115–125.

Cummins, K. L., and M. J. Murphy (2009), An overview of lightning location systems, history, techniques, and data uses, with an in-depth look at the U.S. NLDN, *IEEE Transactions on Electromagnetic Compatibility*, *51*, 499–518.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press.

Dingman, S. L. (2008), *Physical Hydrology*, Waveland Press, Inc.

Environment Canada (a), *CPOP Standard for the Precipitation Analysis (CaPA)*, provided by the Canadian Meteorological Centre of Environment Canada.

Environment Canada (b), capa.f fortran source code, provided by the Canadian Meteorological Centre of Environment Canada.

Environment Canada (c), *capa.m Mist Source Code*, provided by the Canadian Meteorological Centre of Environment Canada.

Environment Canada (2005), *Le Format BURP*, provided by the Canadian Meteorological Centre of Environment Canada.

Environment Canada (2013), filtoma.v23.py help documentation, provided by the Canadian Meteorological Centre of Environment Canada.

Erdin, R., and C. Frei (2011), Data transformation and uncertainty in geostatistical combination of radar and rain gauges, *Journal of Hydrometeorology*, *13*, 1332–1346.

Fortin, V. (2012–2013), Personal communication, scientific researcher at the Canadian Meteorological Centre of Environment Canada.

Fortin, V., G. Roy, and A. Mahidjiba (2011), L'analyse de preécipitation CaPA: état des lieux, presentation.

García-Pintado, J., G. G. Barberá, M. Erena, and V. M. Castillo (2009), Rainfall estimation by rain gauge-radar combination: A concurrent multiplicative-additive approach, *Water Resources Research*, *45*.

Garrigues, S., D. Allard, F. Baret, and J. Morisette (2008), Multivariate quantification of landscape spatial heterogeneity using variogram models, *Remote Sensing of Environment*, *112*, 216–230.

Gellens, D., and E. Roulin (1998), Streamflow response of Belgian catchments to IPCC climate change scenarios, *Journal of Hydrology*, *210*, 242–258.

Goodison, B. E., P.Y.T.Louie, and D. Yang (1998), WMO solid precipitation measurement intercomparison final report, Instruments and Observing Methods Report No. 67.

Goudenhoofdt, E., and L. Delobbe (2009), Evaluation of radar-gauge merging methods for quantitative precipitation estimates, *Hydrology & Earth System Sciences*, *13*, 195–203.

Graham, L. P., J. Andreasson, and B. Carlsson (2007), Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales, and linking methods - a case study on the Lule river basin, *Climatic Change*, *81*, 293–307.

Hay, L. E., M. P. Clark, R. L. Wilby, W. J. Gutowski, J. G. H. Leavesley, Z. Pan, R. W. Arritt, and S. Takle (2002), Use of regional climate model output for hydrologic simulations, *Journal of Hydrometeorology*, *3*, 571–590.

He, X., F. Vejen, S. Stisen, T. O. Sonnenborg, and K. H. Jensen (2011), An operational weather radar-based quantitative precipitation estimation and its application in catchment water resource modelling, *Vadose Zone Journal*, *10*, 8–24.

Houze, R. A. (1997), Stratiform precipitation in regions of convection: A meteorological paradox?, *Bulletin of the American Meteorological Society*, *78*, 2179–2196.

Ines, A. V. M., and J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies, *Agricultural and Forest Meteorology*, *138*, 44–53.

Isaaks, E. H., and R. M. Srivastava (1989), *Applied Geostatistics*, Oxford University Press, Inc.

Lang, S., W.-K. Tao, J. Simpson, and B. Ferrier (2003), Modeling of convective-stratiform precipitation processes: Sensitivity to partitioning methods, *American Meteorological Society*, *42*, 505–527.

Leander, R., and T. A. Buishand (2006), Resampling of regional climate model output for the simulation of extreme river flows, *Journal of Hydrology*, *332*, 487–496.

Lespinas, F. (2012–2013), Personal communication, scientific researcher at the Canadian Meteorological Centre of Environment Canada.

Mahfouf, J., B. Brasnett, and S. Gagnon (2007), A Canadian precipitation analysis (CaPA) project: Description and preliminary results, *Atmosphere-Ocean*, *45*, 1–17.

Mailhot, J., S. Bélair, L. Lefaivre, B. Bilodeau, M. Desgagné, C. Girard, A. Glazer, A. Leduc, A. Méthot, A. Patoine, A. Plante, A. Rahill, T. Robinson, D. T. nad A. Tremblay, P. Vaillancourt, A. Zadra, and A. Qaddouri (2006), The 15-km version of the Canadian regional forecast system, *Atmosphere-Ocean*, *126*, 133–149.

Mateu, J., E. Porcu, and P. Gregori (2008), Recent advances to model anisotropic space-time data, *Statistical Methods and Applications*, *17*, 209–223.

MathWorks (2013), Matlab help documentation.

Mattos, E. V., and L. A. T. Machado (2011), Cloud-to-ground lightning and mesoscale convective systems, *Atmospheric Research*, *99*, 377–390.

Mesinger, F. (2008), Bias adjusted precipitation threat scores, *Advances in Geosciences*, *16*, 137–142.

Miller, D. M. (1984), Reducing transformation bias in curve fitting, *The American Statistician*, *38*, 124–126.

Overeem, A., T. A. Buishand, and I. Holleman (2007), Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar, *Water Resources Research*, *8*, 1225–1242.

Piani, C., J. O. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theoretical and Applied Climatology*, *99*, 187–192.

Ruiz-Leo, A. M., E. Hernandez, S. Queralt, and G. Maqueda (2013), Convective and stratiform precipitation trends in the Spanish Mediterranean coast, *Atmospheric Research*, *119*, 46–55.

Schneider, S., and R. Steinacker (2009), Utilization of radar information to refine precipitation fields by a variational approach, *Meteorology and Atmospheric Physics*, *103*, 137–144.

Shabalova, M. V., W. P. A. van Deursen, and T. A. Buishand (2003), Assessing

future discharge of the river Rhine using regional climate model integrations and a hydrological model, *Climate Research*, *23*, 233–246.

Tadesse, A., and E. N. Anagnostou (2009), Characterization of warm season convective systems over U.S. in terms of cloud to ground lightning, cloud kinematics, and precipitation, *American Meteorological Society*, *91*, 36–46.

Teutschbein, C., and J. Seibert (2012), Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *Journal of Hydrology*, *456–457*, 12–29.

Todini, E. (2001), A bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements, *Hydrology & Earth System Sciences*, *5*, 187–199.

Tremblay, A. (2005), The stratiform and convective components of surface precipitation, *Atmospheric Science*, *62*, 1513–1528.

Trim, D. (2004), *Calculus for Engineers*, Pearson Education Canada, Inc.

Varma, A. K., and G. Liu (2004), A near-global survey of the horizontal variability of rainfall, Department of Meteorology, Florida State University.

Velasco-Forero, C. A., D. Sempere-Torres, E. F. Cassiraga, and J. J. Gómez-Hernández (2009), A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data, *Advances in Water Resources*, *32*, 986–1002.

World Meteorological Organization (2008), Guide to meteorological instruments and methods of observation, WMO-No.8, Seventh Edition.

Yavuz, H., and S. Erdogan (2012), Spatial analysis of monthly and annual precipitation trends in Turkey, *Water Resources Management*, *26*, 609–621.

Yu, K., J. Mateu, and E. Porcu (2007), A kernel-based method for nonparametric estimation of variograms, *Statistica Neerlandica*, *61*, 173–197.

Appendix A

Alternatives to the SI Method

It should be noted that although CaPA employs SI as its underlying geostatistical method, many alternatives do exist in terms of combining gridded, or spatially regular precipitation data, with point-form observation data. These methods can vary from simplistic large scale adjustments intended to combat bias, to more sophisticated statistical combination techniques (*Erdin and Frei*, 2011). Examples of such methods include the following:

- *Objective Analysis*: In order to take intermittent values in space and place them on a user defined grid, objective analysis proposes the following equation (*Brandes*, 1975):

$$WT_i = \exp\left(\frac{-d^2}{EP}\right) \tag{A.1}$$

where WT_i is the weight that is applied to the i^{th} value from the original data

set, d is the distance from that value to the current grid point at which an estimate is required, and EP is a control on the smoothing of the values. In the objective analysis scheme, EP can be selected as required by the user, and may be based on the spacing between the original data points, the level of variation of the original data that is to be preserved, or the amount of error associated with the original observation value (*Brandes*, 1975). Once calculated, the weights are applied as follows (*Brandes*, 1975):

$$F = \frac{\sum_{i=1}^{N} WT_i V_i}{\sum_{i=1}^{N} WT_i} \tag{A.2}$$

where F is the final objective analysis value at the grid point location, V_i is an observation value contributing to the estimation of F, and N is the total number of observations relevant to the determination of F.

This objective analysis set up is traditionally used in a radar context, to create a grid of correction factors for radar derived accumulated precipitation. These correction factors are often multiplicative, and based on comparison with available gauge observations. The same approach can be used with the observation gauge data itself, allowing for the final combination of the gridded radar and gauge data (*Brandes*, 1975). This method has been applied with a variety of modifications, including the use of both multiplicative and additive optimized bias correction factors simultaneously, to tackle biases in both radar-based precipitation magnitudes and fractional occurrence (*García-Pintado et al.*, 2009).

The above objective analysis approach is essentially a modification of the Successive Corrections Method, which in its original form replaces V_i with observation increments over co-located background values, and F with analysis increments over the background field. Additionally, the denominator of Equation A.2 also would include the ratio of the observation error variance to the background error variance (*Daley*, 1991).

- *Deterministic Weights*: Deterministic weighting schemes follow the same basic principal of using weights to combine various data sources into one final gridded output. However, in the deterministic case, these weights are based purely on the characteristics of the data sets being considered. An example of this technique can once again be pulled from the problem of combining radar and gauge precipitation data.

  Using a deterministic weighting approach, a polynomial function can be derived which describes the relationship between the distance from a grid point to a gauge location, and the bias of the radar at that location. This is used to formulate a distance-dependent adjustment field, denoted $F_{r(i,j)}$ (*He et al.*, 2011). To create a spatial adjustment field , $F_{s(i,j)}$, weights can be derived for precipitation gauges in a manner similar to objective analysis, and then combined with radar bias estimates at each location. This results in a first-guess field, $F_{(i,j)}$, that can then be combined with a similar field, created by substituting radar bias estimates at each location with the average of the four

closest bias values, denoted $\hat{F}_{(i,j)}$. Weights can then be derived from the spatial correlation of the radar bias data set, producing $w_{s(i,j)}$ (*He et al.*, 2011). This produces the final adjustment field, $\bar{F}_{(i,j)}$, used to treat the radar data and produce a final gridded output, as follows (*He et al.*, 2011):

$$\bar{F}_{(i,j)} = F_{r(i,j)} + w_{s(i,j)} \left(F_{s(i,j)} - F_{r(i,j)} \right) \tag{A.3}$$

and,

$$F_{s(i,j)} = \hat{F}_{(i,j)} + F_{(i,j)} \tag{A.4}$$

- *Splines*: Spline functions can be fit to data to allow for the interpolation of that data to a spatial grid. Spline functions can generally be fit through a minimization function, which both quantifies the goodness of fit of the spline approximations to observation values, and penalizes the over-smoothing of the data (*Daley*, 1991). This penalization is based on the cubic-spline property that the following curvature quantity, expressed here in one dimension, is minimized (*Daley*, 1991):

$$\int_{x_a}^{x_b} \left[f'' (x) \right]^2 dx \tag{A.5}$$

where x_a and x_b are the limits defining the range over which the spline function

is fit, and $f''(x)$ denotes the second derivative of the values to be interpolated (*Daley*, 1991).

An example of an application of splines in the production of gridded precipitation estimates is provided by Schneider and Steinacker (2009). In this study, radar and gauge data are transferred to a grid by breaking down each data source into waves, which are then combined to produce quantitative estimates at the grid locations. These waves are split into two categories: fingerprint waves that define known patterns in time and space in the data, and a second class of waves which simply characterizes aspects of the precipitation that are not encompassed by the fingerprint waves. These two types of waves are combined at an estimate location by adding the non-fingerprint wave value to a weighted linear combination of the values of the fingerprint waves. The weights are determined based on the allowable degree of smoothing. This concept derives itself from the theory that radar data is particularly adept at capturing spatial patterns in precipitation, if not accurate magnitudes (*Schneider and Steinacker*, 2009).

- *Bayesian Conditioning*: Bayesian conditioning provides gridded combined estimates through a minimum variance approach. This can be done by specifying the gridded model or radar data as the prior information, and the gauge observations as the measurement data, or the likelihood information. These two sources can then be combined by applying Bayesian probability laws to pro-

duce the posterior precipitation estimates on the original radar grid (*Todini*, 2001). In the implementation of this technique outlined by Todini (2001), block kriging is used to create a gridded observation field to serve in the Bayesian estimation procedure, which also relies on the application of a Kalman filter in its setup. This method allows for an estimate of the error covariance matrix to be produced.

The result of applying the techniques outlined above is generally an improved picture of precipitation in space, as compared to using only gauge-type observations (*Brandes*, 1975; *He et al.*, 2011; *Schneider and Steinacker*, 2009; *Todini*, 2001). While all of the method examples discussed above dealt with radar data, the basic concepts are transferable to gridded model data as well. Despite this, SI is the method of choice for CaPA. SI is a technique grounded on the minimization of estimate error variance, which is desirable in regards to limiting estimation uncertainties. Furthermore, it provides quantification of those estimate error variances, which can be potentially useful in determining how to apply estimates after they are determined (*Erdin and Frei*, 2011). SI also has the benefit of determining its semivariogram directly from the input data, reducing the need to select or calibrate weighting parameters as may be the case for other methods (*Erdin and Frei*, 2011). Thus the SI methodology used by CaPA does possess a series of advantages that support its use in the program.

Finally, a study by Yavuz and Erdoğan (2012) shows that ordinary kriging, a close relative of SI, outperforms alternatives such as a spline-based technique and a basic inverse-distance weighting method as a means of interpolation. A study conducted in Belgium over a four-year verification period by Goudenhoofdt and Delobbe (2009) also examines methods for combining radar and observation data, and confirms that geostatistical kriging methods, the category into which SI falls, perform better than other methods tested. The alternatives compared included a mean field bias adjustment method, a form of deterministic weighting, and a form of objective analysis similar to that proposed by Brandes (1975). It was also found that the geostatistical methods performed better than the alternatives under a variety of gauge densities, but that the degree of their relative improvement increased with increasing density (*Goudenhoofdt and Delobbe*, 2009). These results lend further credence to the use of SI in CaPA.

Appendix B

MWA Bias Correction Methodology Implementation Details and Procedures

The functionality of the MWA method in an operational sense is dependent on the ability of CaPA to be re-configured from a standard operational setup to a MWA setup. This is achieved through a series of small configuration modifications to CaPA, including the following:

1. Implementation of a modified *preparerchampessai* subroutine, which is housed within the *libcapa.py* python code, and is called during the obsCaPA portion of the CaPA program. This modified subroutine is designed to create MWA GEM inputs for latter use within the CaPA Mist program component.

2. Replacement of the observation input data generated by *obsCaPA* in the 24 hour "entrees" data folder in the CaPA directory, with the MWA observation files produced externally through pre-processing accomplished via Matlab. Note that these files must be created to exactly replicate the formatting of the *obsCaPA* produced files.

3. Modification of the CaPA *mist.c* code to output enhanced statoma-type files with all relevant data, including analysis, analysis error, and GEM values for each grid cell

4. Deactivation of the Box-Cox transformation, by setting the *LAMBDA_BOXCOX* and *OFFSET_BOXCOX* parameters to values of 1 in the *mist.cfg* configuration file.

While the above modifications encapsulate all required configuration changes, these changes are only useful if it assumed that, given the MWA input, CaPA will produce an analysis with equivalent performance capabilities as those that exist under standard operational conditions. A key consideration when weighing the validity of this assumption is CaPA's estimation of a semivariogram to fit the input data. If the exponential semivariogram type used in CaPA is not appropriate for the MWA input, then the possibility of producing a poor precipitation analysis is heightened, and the use of such an analysis as a correction tool for standard analyses becomes a dubious practice. Thus, the quality of the semivariogram estimation procedure in CaPA, when applied to MWA input, was tested to confirm adequate semivariogram

Statistic	Value
Mean R^2	0.9683
Min R^2	0.9249
Max R^2	0.9963

Table B.1: Curve fit statistics for MWA semivariograms

fitting. The results of the exercise can be reviewed in Table B.1, which provides a summary of the R^2 values associated with the estimated semivariograms for each time period in the study period when MWA inputs are considered.

As is evident in Table B.1, high R^2 values are observed, with a mean R^2 value across all time periods of 0.97. Even the lowest R^2 values were found to be above 0.92. Based on this simple investigation, it was determined that the fitting of semivariograms to the MWA input is accomplished with sufficient skill so as not to hinder the production of an accurate precipitation analysis. Visual examples of the fit of the theoretical exponential semivariogram to the experimental semivariogram data for 6 hour and MWA inputs are shown in Figures B.1 and B.2, for the specific 2010/06/01 12:00 UTC time stamp. While the 6 hour experimental semivariogram data appears to be more tightly dispersed around the theoretical semivariogram, the MWA theoretical semivariogram captures the basic information contained in the MWA experimental semivariogram data.

It is also important to consider the validity of the assumption that the MWA approach does create approximately normally distributed input data, as this is the fundamental premise driving the use of the MWA method. This assumption was first assessed through the use of the built-in Matlab function "lillietest". This test

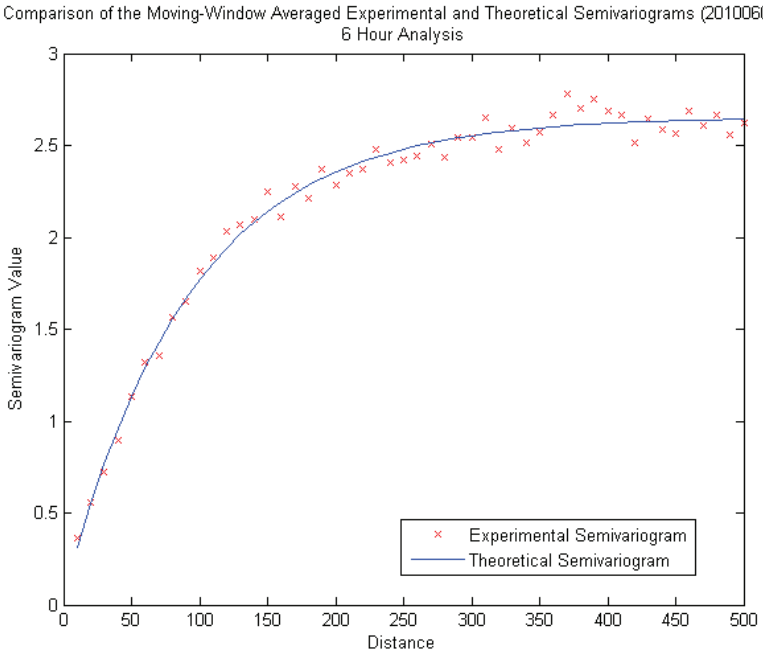

Figure B.1: Theoretical semivariogram fit for 6 hour experimental semivariogram data.

operates at a 5% significance level, and evaluates the null hypothesis that the data provided to the function is normally distributed (*Environment Canada*, c). Acceptance percentages of this null hypothesis were calculated based on a survey of the function results for all observation locations. These results are presented in Figure B.3. Zero precipitation values have been included in this testing, as they are included in the MWA technique once applied within CaPA. Given this inclusion, the low null hypothesis acceptance rate for the untreated and Box-Cox transformed inputs are not surprising. Clearly, the MWA technique shows substantial improvement on a comparative basis, although the acceptance rate still varies depending on the minimum number of required days of record per window, and never registers above

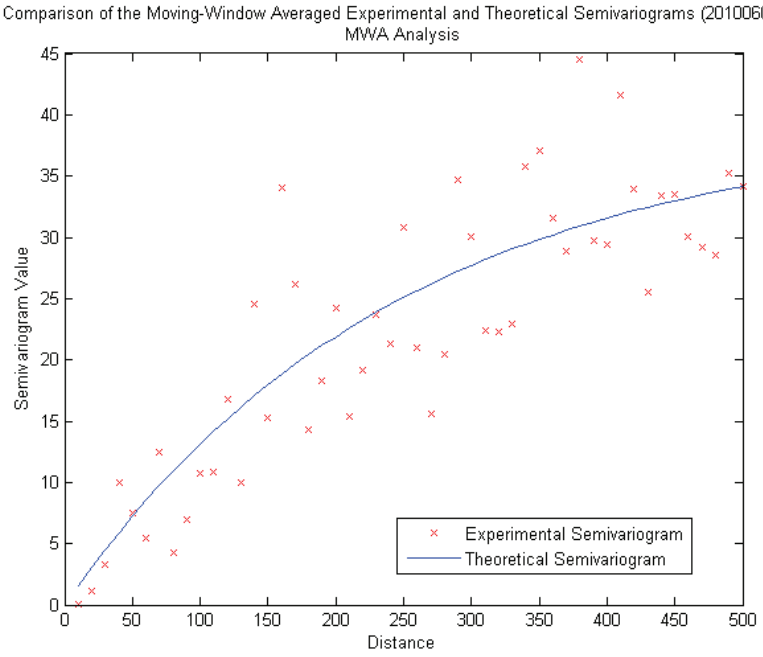

Figure B.2: Theoretical semivariogram fit for MWA experimental semivariogram data.

15%.

The effect that removing all zero-precipitation amounts has on the hypothesis acceptance rates was also considered, and the results are presented in Figure B.4. Here it can be seen that applying the Box-Cox transformation leads to a notable improvement of the null hypothesis acceptance rate over the untreated case, and even surpasses the performance of the MWA method. However, when a minimum of 20 days of record are required per window, the two treatment options show little difference in their abilities to create normally distributed input data, with the acceptance rate in both cases being close to 15%. This finding supports the use of a minimum of 20 days of record required during the MWA analysis. Furthermore, the results also

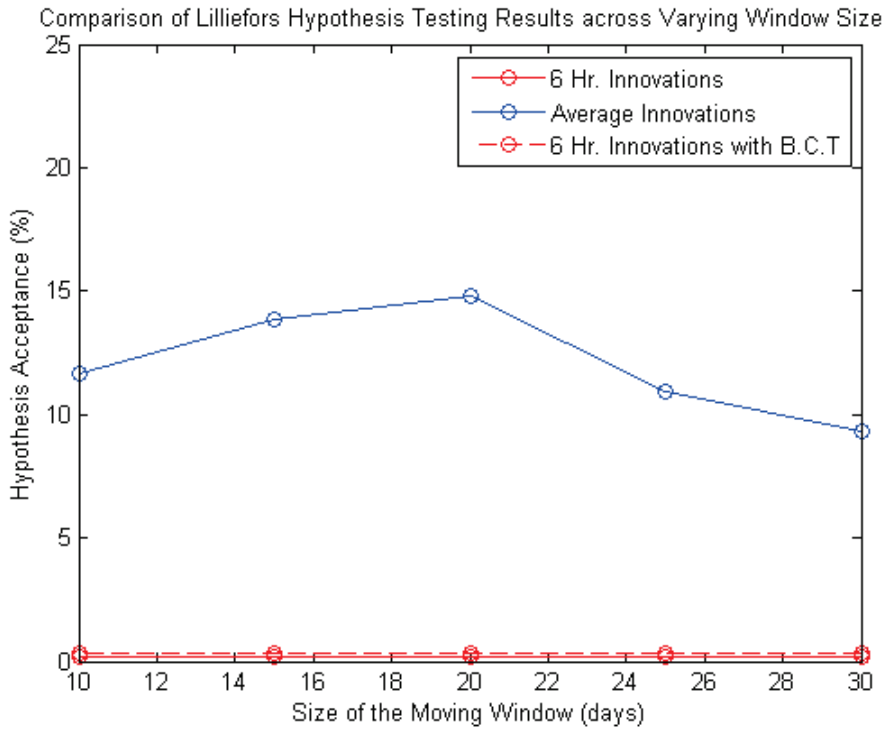

Figure B.3: Comparison of Lilliefors hypothesis testing results across varying input data treatment schemes. The Box-Cox transformation method is denoted as B.C.T.

indicate that the hypothesis acceptance rates demonstrated by the MWA technique are reasonable, despite reaching a maximum of only approximately 15%, as a similar performance was seen for the Box-Cox transformation case which has already been adopted operationally.

As an additional means of assessing the performance of the MWA methodology, variance, coefficient of variation, skewness, and kurtosis statistics for the untreated, Box-Cox transformed, and MWA treated cases were assessed. Figure B.5 summarizes the results found, and clearly indicates that the use of the MWA method reduces the variance and coefficient of variation for the input innovation data. This is a

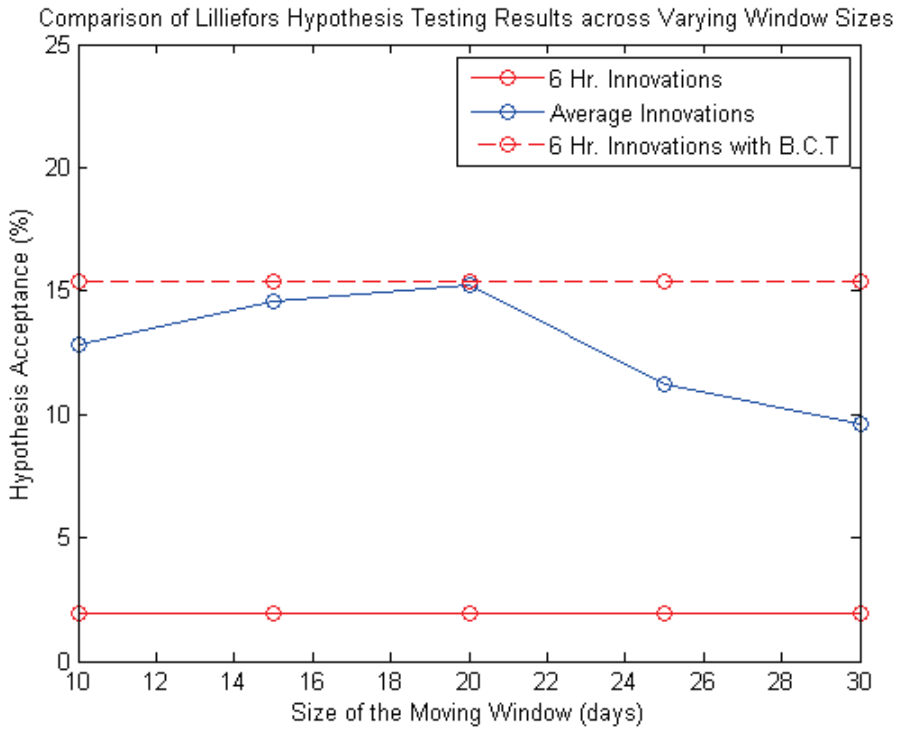

Figure B.4: Comparison of Lilliefors hypothesis testing results across varying input data treatment schemes for non-zero precipitation amounts only. The Box-Cox transformation method is denoted as B.C.T.

desirable feature, as the variance and coefficient of variation terms both depend in part on $(x_i - \overline{x})^2$, where x_i is an input value and $\overline{x}$ is the mean of the input data set. Due to the squaring of this term, the variance and coefficient of variation are both sensitive to large differences between extreme precipitation events and the mean precipitation amount within a data set; this is the exact feature that plagues the semivariogram estimation and can lead to poor SI performance. Thus, the variance and coefficient of variation act as proxy indicators for the MWA's ability to reduce this phenomenon.

The skewness and kurtosis of the input were also considered. These statistics are indicators of the shape of the distribution of the input data. A normal distribution possesses a value of zero for each of these statistics, and using this as a reference we can see that while the Box-Cox treated input data performs better than the MWA method when considering skewness or asymmetry, the reverse is true when looking at kurtosis or peakedness. Both treatments represent obvious improvements over the untreated input case, so long as a minimum of 10 days of record are required for the MWA windows. Based on these results, the MWA method is found to perform reasonably well in terms of resulting in approximately normally distributed input data in comparison to the Box-Cox method.

The results in Figure B.5 include the influence of zero precipitation amounts on the input data set. Figure B.6 presents the outcome of the same statistical tests with zero precipitation amounts removed from the input data. As can be seen, the

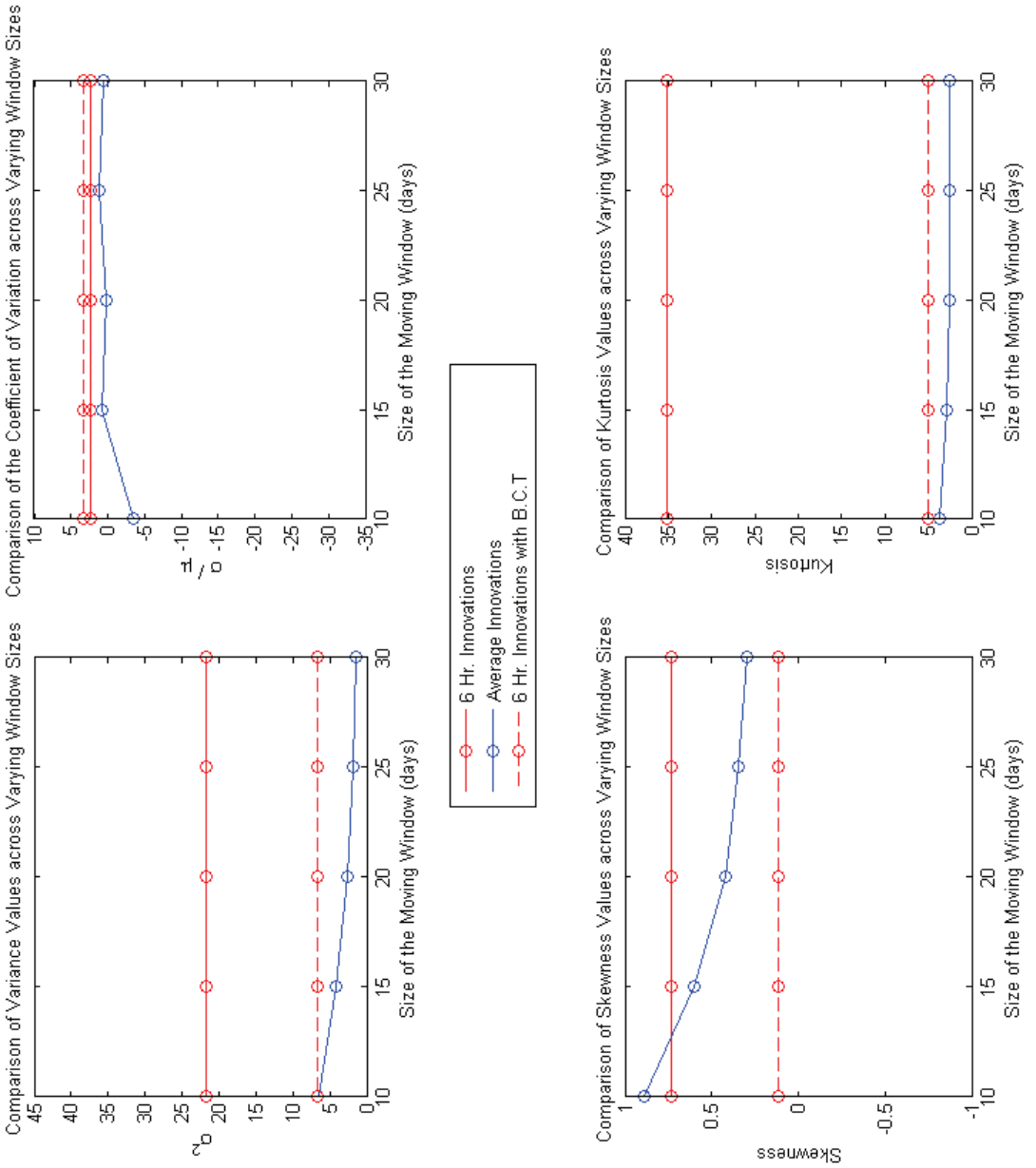

Figure B.5: Comparison of variance, coefficient of variation, skewness, and kurtosis statistics across varying input data treatment schemes. The Box-Cox transformation method is denoted as B.C.T.

variance is reduced through the use of the MWA method, although the same cannot be said for the coefficient of variation. However, the performance of all three cases is still relatively similar for the coefficient of variation, with the MWA treatment being the only case exhibiting positive values. Once again the Box-Cox treatment performs best in regard to the skewness statistic, although the MWA method still improves upon the untreated case in this regard, given a minimum of 10 required days of record per window. In terms of skewness, the Box-Cox and MWA treatments show no considerable differences in the magnitude of their improvements over the untreated kurtosis values. Based upon these results, the MWA method can still be considered as justifiable for implementation in CaPA and for further testing in this thesis.

A visual confirmation of this judgment is provided by Figure B.7, which clearly demonstrates in a graphical sense the normalizing capabilities of the MWA method, and the ability of the averaging technique to reduce extreme input values as compared to the Box-Cox treated and untreated 6 hour precipitation data. Such extreme values may otherwise wreak havoc on the calculation of the semivariogram. However, it is important to note that this figure does not represent the innovation amounts, but rather reflects only the improvement of the MWA technique on the precipitation observation data in this case.

Finally, the choice of a minimum of 20 days of record in each 30-day window was selected based on the above figures. In all cases, using 20 days of record as a

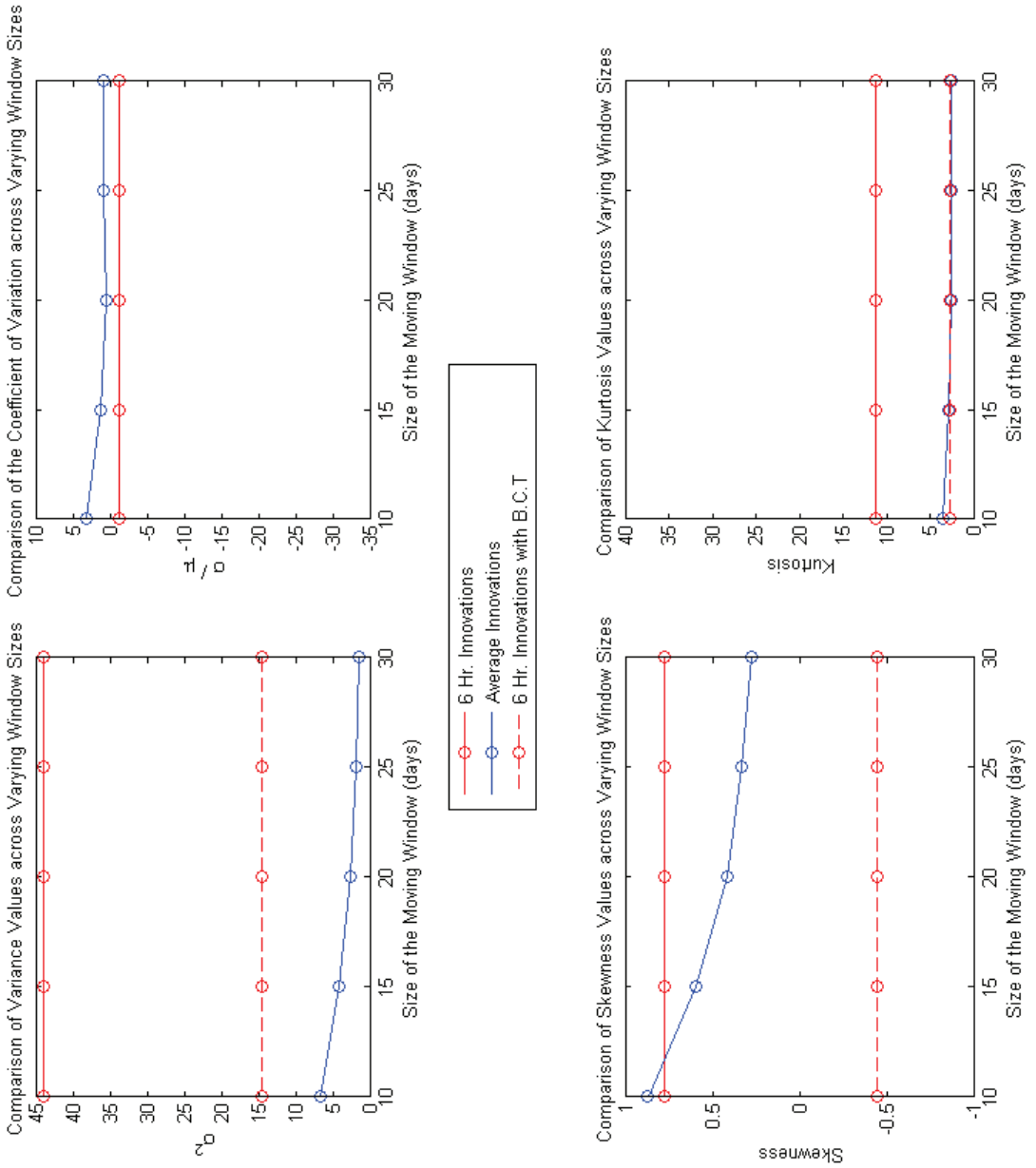

Figure B.6: Comparison of variance, coefficient of variation, skewness, and kurtosis statistics across varying input data treatment schemes when excluding zero precipitation events. The Box-Cox transformation method is denoted as B.C.T.

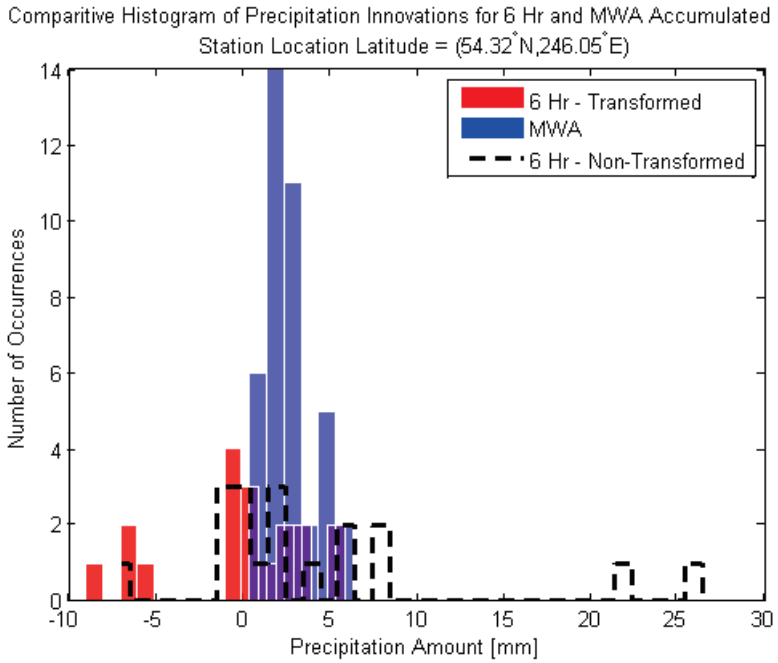

Figure B.7: Moving-window monthly averaging scheme employed to create a MWA data set

minimum yields statistical characteristics in the treated data that are approximately similar, if not improved in the Lilliefors testing case, to those seen when a full 30 out of 30 days are demanded. Thus the amount of performance loss experienced by lowering the minimum requirement for days of record to 20 for any given window was deemed to be acceptable, and was found to represent a balance with the need to prevent an overly strict minimum record requirement. If the minimum days of record required is too strict, this could potentially yield too few MWA observation records for input into the CaPA analysis or any given time period.

To augment the validity of the 20 -day minimum threshold choice, Figures B.8 and B.9 show the percentage of absolute difference and the difference, respectively,

between MWA observed precipitation estimates based on a given minimum number of days of record required, and the MWA value that would result if all 30 days of record were used. This was accomplished by identifying all windows, across all times periods and observation locations, that possessed 30 out of 30 days. Using these predetermined windows, a sample of 5 to 30 precipitation amounts were randomly drawn without replacement from each window to calculate MWA values, for comparison with the true MWA amounts. This process was iterated 1000 times.

Ideally no difference between the two types of MWA values would be noted in the two figures when looking at 20 days of record per window being required. This unfortunately is not the case in Figure B.8, although when the average difference in the values across all samples in Figure B.9 is considered, the difference is approximately zero. However, many individual simulations do report differences, both positive and negative. Because of this, the results in Figure B.8, which focus on the percentage of absolute difference, are not surprising. Once again, a trade-off must be made in terms of requiring more days of record per window to gain a more accurate MWA estimate, and producing enough MWA observation values per time period to have confidence that the CaPA analysis produced is truly improved over the background field. Therefore, the decision to use a 20 day minimum requirement was upheld.

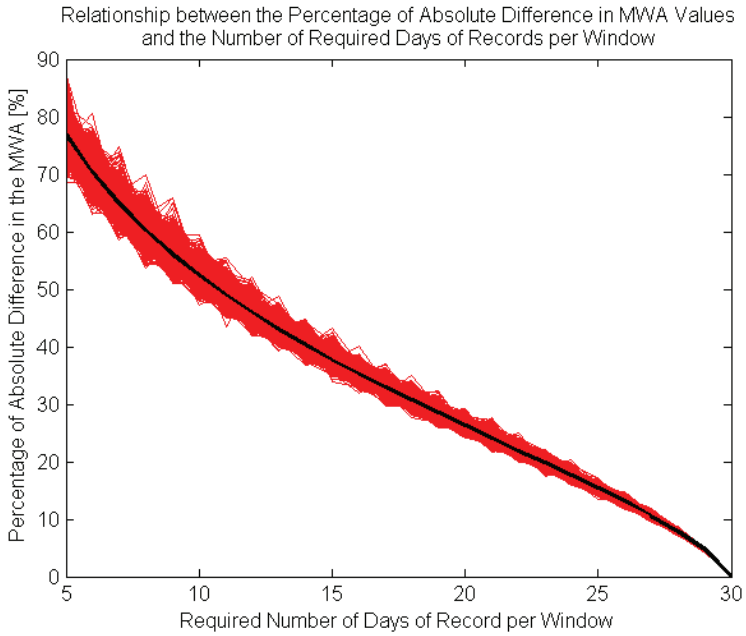

Figure B.8: Assessment of the impact of the minimum number of days of record required per window on the accuracy of the estimated MWA values, based on the percentage of absolute difference.

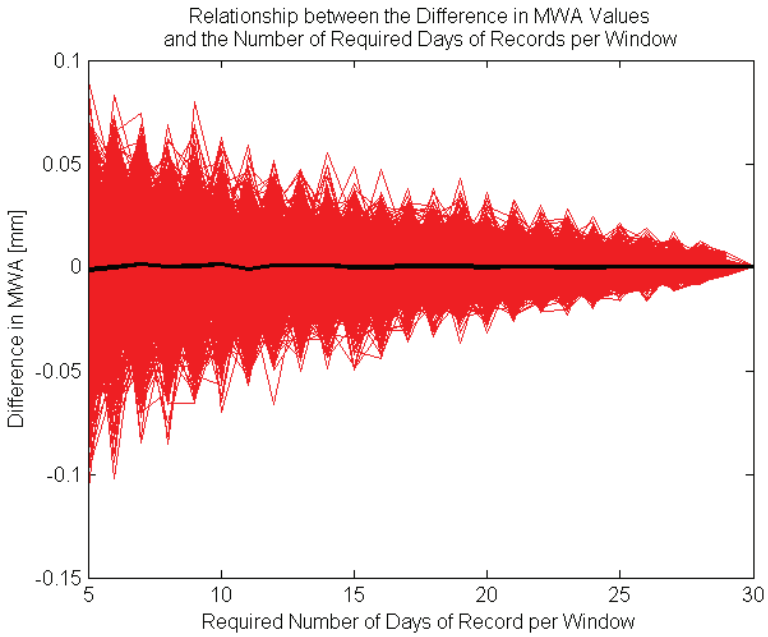

Figure B.9: Assessment of the impact of the minimum number of days of record required per window on the accuracy of the estimated MWA values.

Appendix C

Long-Term Bias of the MWA Schemes

It is important to demonstrate that the MWA methodology is truly beneficial to the data transformation bias correction problem. If a correction method itself instills its own, new bias into the analyses it is intended to correct, then it may be possible for the net resulting bias to be greater than in the initial case.

For the purposes of this long-term bias exploration, the basic question that must be addressed is whether or not simply averaging the series of corrected values for a given time period and location produces any new form of bias. As the moving window shifts through time, the corrected analysis value for a given time period can vary; confirmation is required that taking an average of all calculated corrected values for a time period is the optimal way to address this issue.

To begin, it will be assumed for simplicity that it is the scaling factors, and not the corrected analysis values, that are averaged. This assumption allows for greater clarity in the required mathematical expressions, and represents an equivalent approach to that taken in this thesis.

For a single window, i, the difference in total cumulative precipitation between using averaged and non-averaged scaling factors can be expressed as follows:

$$d_i = \sum_{j=0}^{29} P_r SFA_r - \sum_{j=0}^{29} P_r SF_{r,f_m(r,i)} \tag{C.1}$$

where r is calculated via the following:

$$r = 29 + i - j \tag{C.2}$$

and where P_r is the time series vector of analysis precipitation values for the entire study period for a given grid cell location, SFA_r is a corresponding vector of averaged scaling factors, and $SF_{r,f_m(r,i)}$ denotes the scaling factor calculated based on the precipitation analysis index r and the window i. The function $f_m\,(r, i)$ is described in Equation 2.34. The difference, or bias, that is described in Equation C.1, is on a per-window basis. Such a bias can be acceptable, if over the course of the study it cancels itself out. To appraise the value of d_i throughout the study, Equation C.1 can be updated as follows:

$$\sum_{i=1}^{I} d_i = \sum_{i=1}^{I} \sum_{j=0}^{29} P_r SFA_r - \sum_{i=1}^{I} \sum_{j=0}^{29} P_r SF_{r,f_m(r,i)} \tag{C.3}$$

where the desired value of $\sum_{i=1}^{I} d_i$ is zero. The first step in proving that no cumulative bias is introduced through the study period is acknowledging the following equality:

$$\sum_{r=1}^{R} SFA_r P_r \times f_m(r, R) = \sum_{i=1}^{I} \sum_{j=0}^{29} P_r SF_{r,f_m(r,i)} \tag{C.4}$$

where R is the largest index value of r. The left hand side of Equation C.4 represents analysis amounts corrected using average scaling factors for each time period multiplied by the total number of windows that encompassed that time period, and then summed across all time periods in the study. The right hand side of the equation represents the sum of the analysis amounts corrected using un-averaged scaling factors across all windows. This equality is founded on the basic principal that the sum of n data elements is equal to the average of those data elements multiplied by n.

An alternative way to phrase the left hand side of Equation C.4 is to sum the analysis amounts corrected using average scaling factors on a per-window basis, and then again across all windows. This accomplishes the same task, as summing across all windows is equivalent to summing across each time period only once, and then multiplying by the total number of times that that time period was captured by a moving window. Thus the following statement can be made:

$$\sum_{i=1}^{I}\sum_{j=0}^{29} P_r SFA_r = \sum_{r=1}^{R} SFA_r P_r \times f_m\left(r, R\right) \tag{C.5}$$

and by substituting this result into Equation C.6, it is shown that no bias is introduced when the full study time frame is considered as follows:

$$\sum_{i=1}^{I} d_i = \sum_{i=1}^{I}\sum_{j=0}^{29} P_r SF_{r,f_m(r,i)} - \sum_{i=1}^{I}\sum_{j=0}^{29} P_r SF_{r,f_m(r,i)} = 0 \tag{C.6}$$

Appendix D

Additional Verification Station Maps

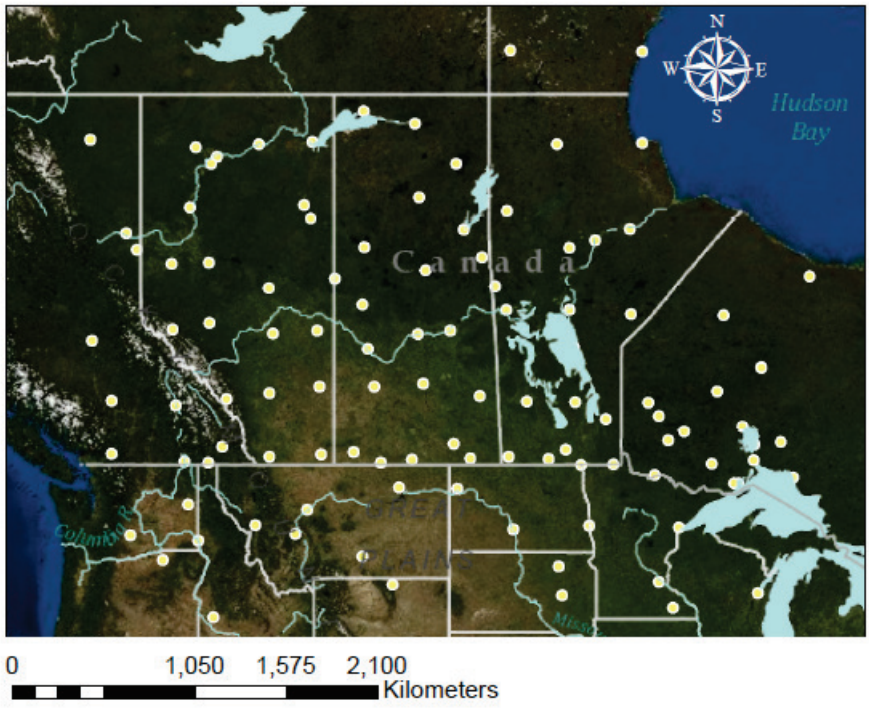

Figure D.1: Verification station location map. The selected stations represents approximately 30% of the synoptic observation network within the domain, approximately distributed on a grid.

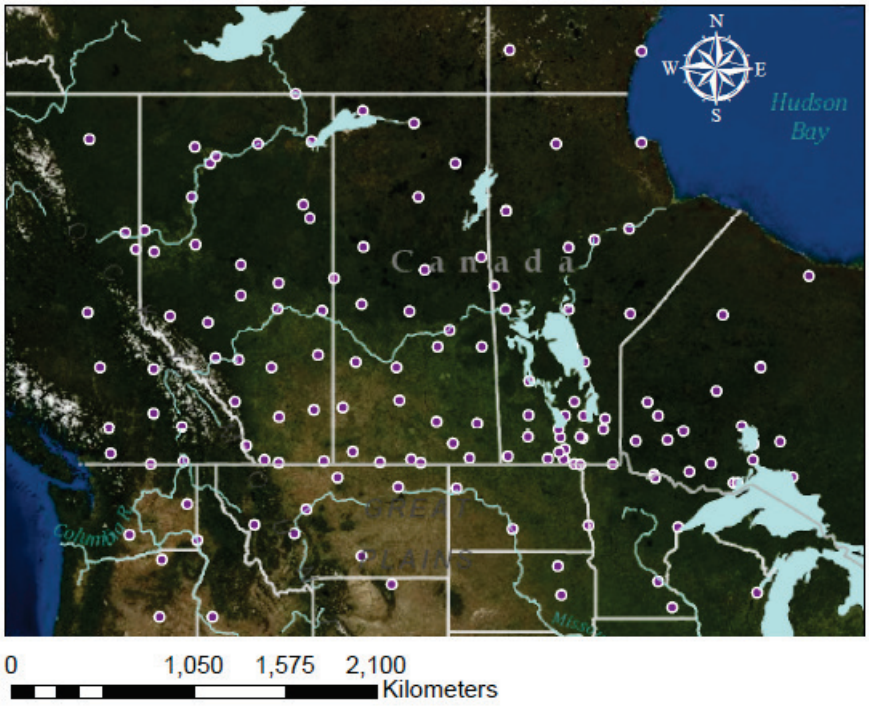

Figure D.2: Verification station location map. The selected stations represents approximately 40% of the synoptic observation network within the domain, approximately distributed on a grid.

Appendix E

Additional Verification Results: Data Transformation Bias Correction

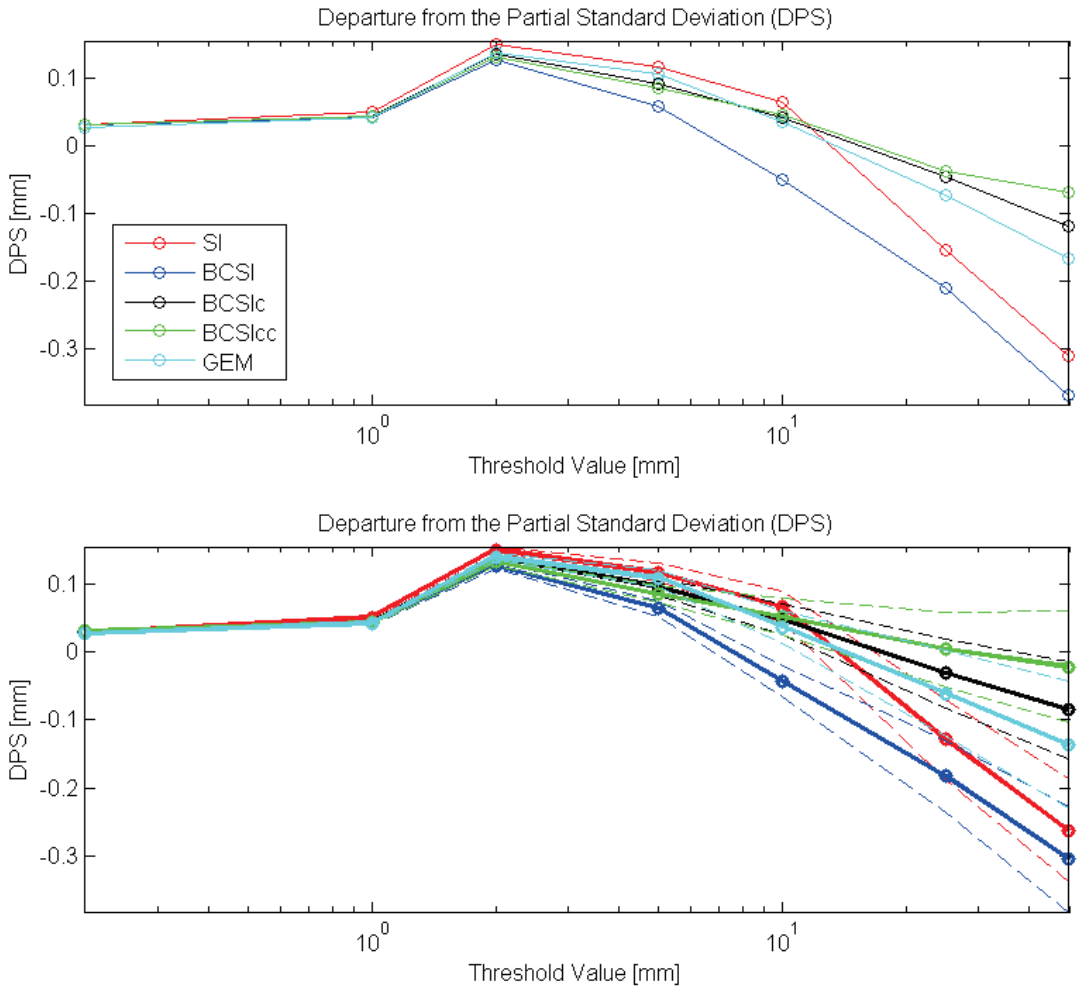

Figure E.1: **Top:** DPS results averaged across all verification locations and all time periods in the study. **Bottom:** DPS bootstrapping results, with the bootstrapped mean DPS values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines. Note that the results shown are for the 20% verification case

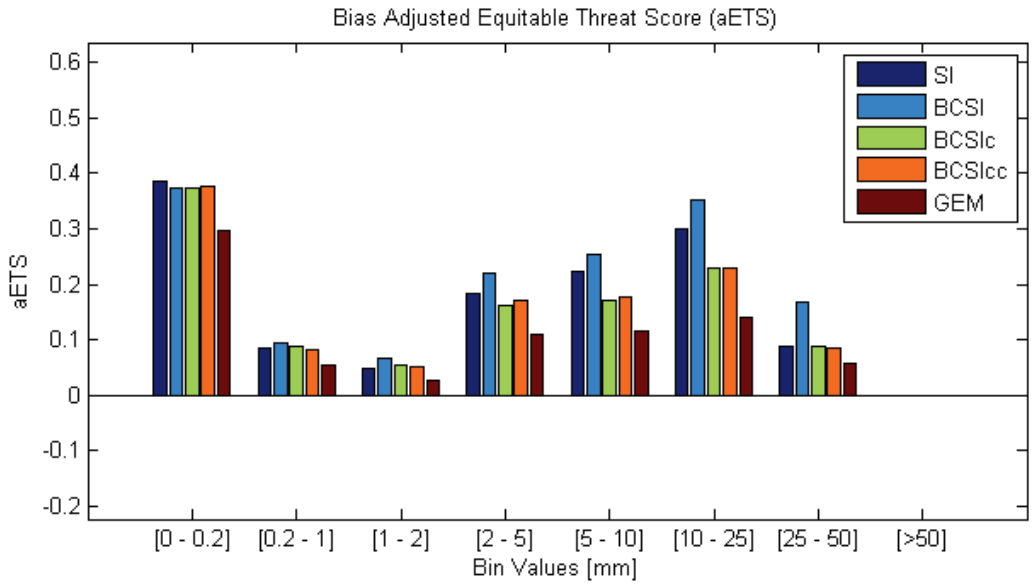

Figure E.2: aETS results averaged across all verification locations and all time periods in the study.

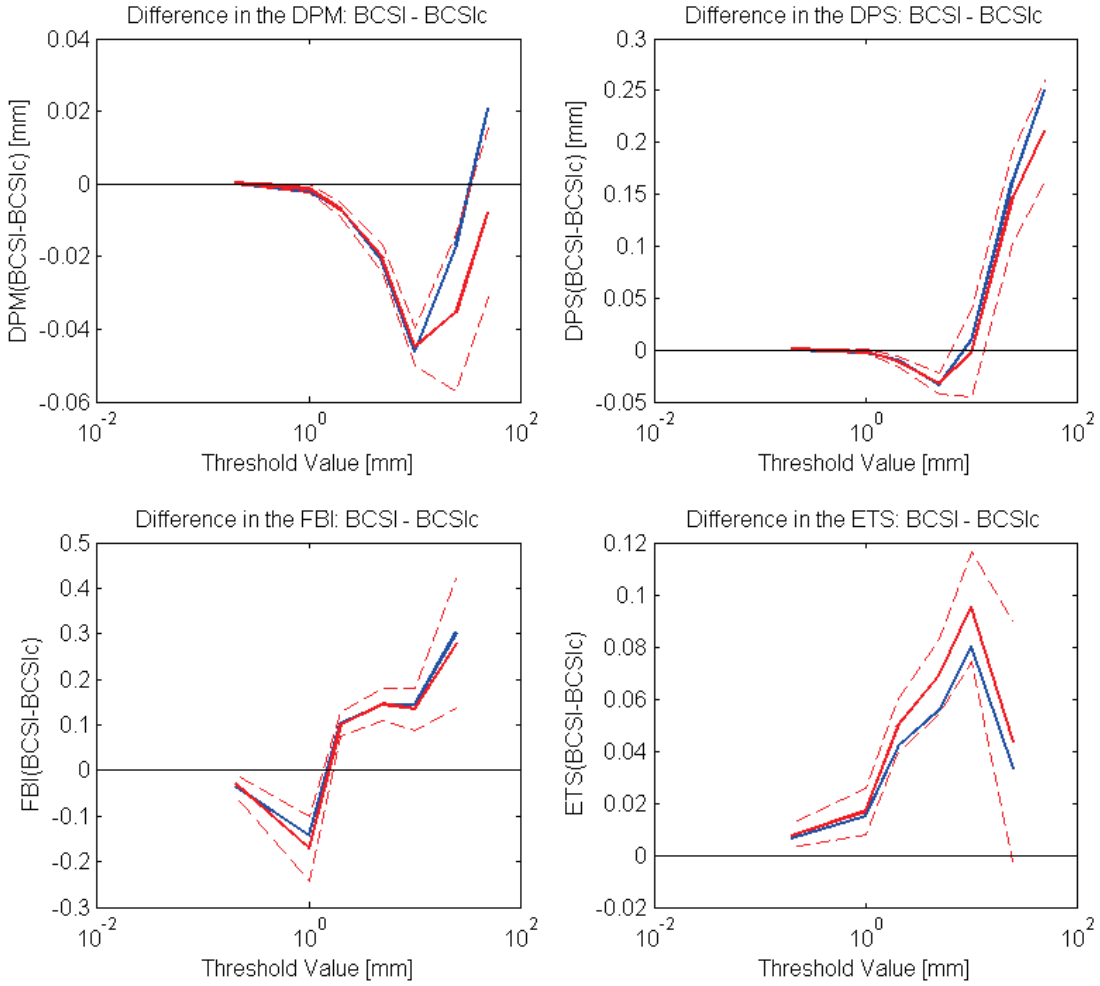

Figure E.3: Difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the operational configuration, BCSI, and the BCSIc MWA analysis (scheme I). The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

Appendix F

Semivariogram Methodology Implementation Details and Procedures

Automatic Estimation of the Semivariogram Parameters

The automation of the semivariogram estimation procedure in ArcGIS was achieved through a python script. The basic series of events triggered by the script is as follows:

- definition of the data projection as WGS84
- reading and conversion of the input data
- transformation of the input data to the Universal Transverse Mercator (UTM)

coordinate system in order to obtain meaningful metric distance measures

- creation of a new geostatistical layer based on previously defined model files
- retrieval of new semivariogram parameter values with updates to a master parameter output file
- creation of a prediction surface using simple kriging to produce and output cross-validation error statistics

Once the input data has been read into ArcGIS and duly manipulated, all remaining processes related to the generation of new semivariogram parameters and error estimates must be done for all semivariogram configurations. This entailed eight sets of semivariograms. The entire semivariogram estimation process was also executed for each time step, with only the data projection information carrying forward through each iteration. Beyond this, the process was also repeated for all three innovation limitation types: innovations where any observation and GEM combination is allowed, innovations where at least one of the observation or GEM contributions have to be non-zero, and innovations limited to cases where both the observations and GEM experience precipitation amounts greater than zero.

It is important to note that a WGS84 datum defined for the arcGIS workspace is not equivalent to the rotated latitude-longitude system used by CaPA. However, all cases dealt with use the same projection in the ArcGIS environment, ensuring that all final CaPA results are directly comparable.

In order for the ArcGIS generated theoretical parameter values to be compatible with the *mist.c* program responsible for handling the semivariogram, a rearrangement of the file formatting from the original output was necessary. This was accomplished through the use of a Matlab script, which performs the following tasks:

- Reads in the cumulative parameter files generated by ArcGIS
- Breaks down the contents into time step dependent series
- Applies a 30 day smoothing filter equivalent to that used by CaPA
- Writes the results to an output file consistent with the formatting of the the *prmsemivartheo* files that are created during the execution of *mist.c*

Finally, to make use of the anisotropic semivariogram parameters estimated for CaPA, the following modifications to the program had to be made:

- Alteration of the *creerVecteurVoisin* and *creerMatriceVoisinStation* subroutines in the *mist.c* file, to use the anisotropic semivariogram parameters to supply the appropriate correlation length values to the covariogramme_xyz subroutine in the *capa.f* file
- Amending of the *extraireChampConfig* and *imprimerConfiguration* subroutines in the *mist.c* file, and the *mist.h* file to properly read in and output the anisotropic parameters

- Redirection of the internal CaPA reference to the theoretical semivariogram parameters, to read in the newly generated files capable of characterizing anisotropy.

- Addition of the LONGUEUR_CORR_PREV_MNR and DIR_ANGL anisotropic parameters to the *mist.cfg* configuration file

Estimation of the Stratiform Curve for the Rain Rate versus Total Cumulative Precipitation Plots

To form the base stratiform curve for the rain rate versus total cumulative precipitation plot, the rain rate data must first be binned. In order to bin the precipitation data, intensity bins, given by $[R, R + dR]$, had to be created for each time step. In each case, the bins were created to span the entire intensity range. Using the *linspace* Matlab function, 20 evenly spaced division were created from 0 to the maximum intensity, resulting in 19 bins per time step. All non-zero GEM precipitation rates were then classified into the appropriate bins. The total amount of precipitation allocated to each bin was then determined for the time step, resulting in total cumulative amounts in millimeters. This formed the empirical data set to be plotted and fit with a base stratiform curve.

In order to form the necessary stratiform curve, the following sequence of assessment were made to determine which points to include in the estimated stratiform curve:

- The first point to be accepted into the stratiform curve corresponds to the maximum total cumulative precipitation amount
- The slopes between all consecutive empirical points are calculated
- Moving towards greater precipitation intensities, the next consecutive empirical point is considered if the slope to that point is both negative and less steep than the preceding slope evaluated
- If the point remains a candidate, it is accepted into the stratiform curve series if the total cumulative precipitation amount associated with that point is less than the amount corresponding to the previously accepted point

Once the stratiform curve data is established, linear linkages between the points are used to fill out the data set. This is done such that each bin possesses a stratiform curve value, starting from the bin corresponding to the maximum total cumulative precipitation amount. This is accomplished via the following sequence of steps which is repeated for each bin value, with increasingly large bins being considered:

- Check if the current bin already has a point associated with it that has been accepted into the stratiform curve, if so then move to the next bin
- If the bin does not already have a stratiform curve point, find the stratiform curve point that has the closest, yet still larger, bin value as compared to the current bin. Determine the slope between this value and the immediately previ-

ous stratiform curve point, and use this slope to determine the total cumulative precipitation for the current bin:

$$SC_i = \left(\frac{b_{lrg} - b_{lrg-1}}{SC_{lrg} - SC_{lrg-1}} \times (b_i - b_{i-1}) \right) + SC_{i-1} \tag{F.1}$$

where SC refers to the total cumulative precipitation amounts of the stratiform curve, b refers to the set of bins for the current time step, the index i denotes the current bin, and the index lrg denotes the index of the already accepted stratiform curve point with the closest, yet larger, bin value to the current bin.

- For the last bin in the series, or if the bin value is greater than the bin of the last accepted stratiform curve point, the slope used to determine the current total cumulative precipitation amount for the stratiform curve is taken as the slope between the last two accepted points.

Appendix G

Additional Verification Results:

Semivariogram Improvement

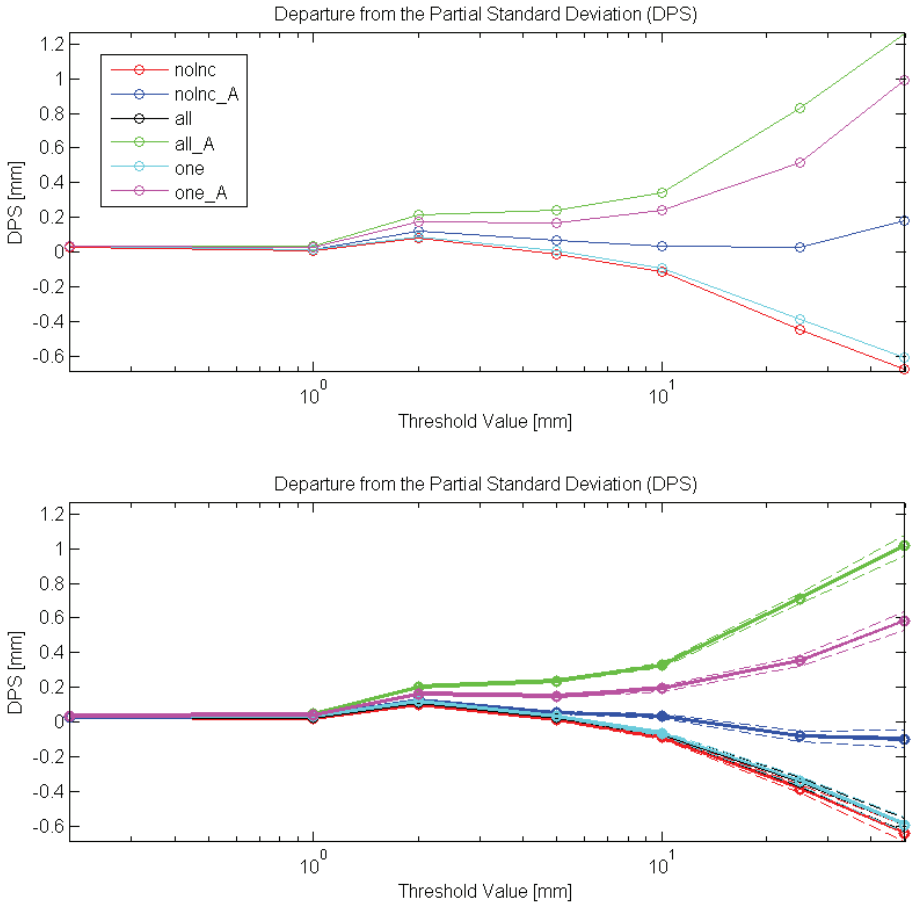

Figure G.1: Anisotropy investigation DPS results. **Top:** DPS results averaged across all verification locations and all time periods in the study. **Bottom:** DPS bootstrapping results, with the bootstrapped mean DPS values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

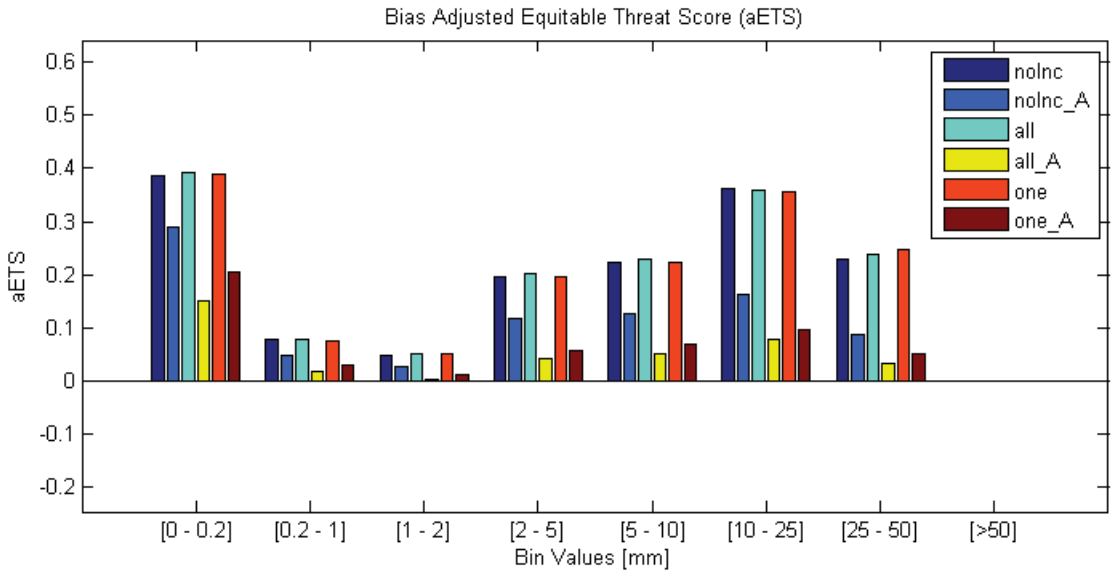

Figure G.2: Anisotropy investigation aETS results averaged across all verification locations and all time periods in the study

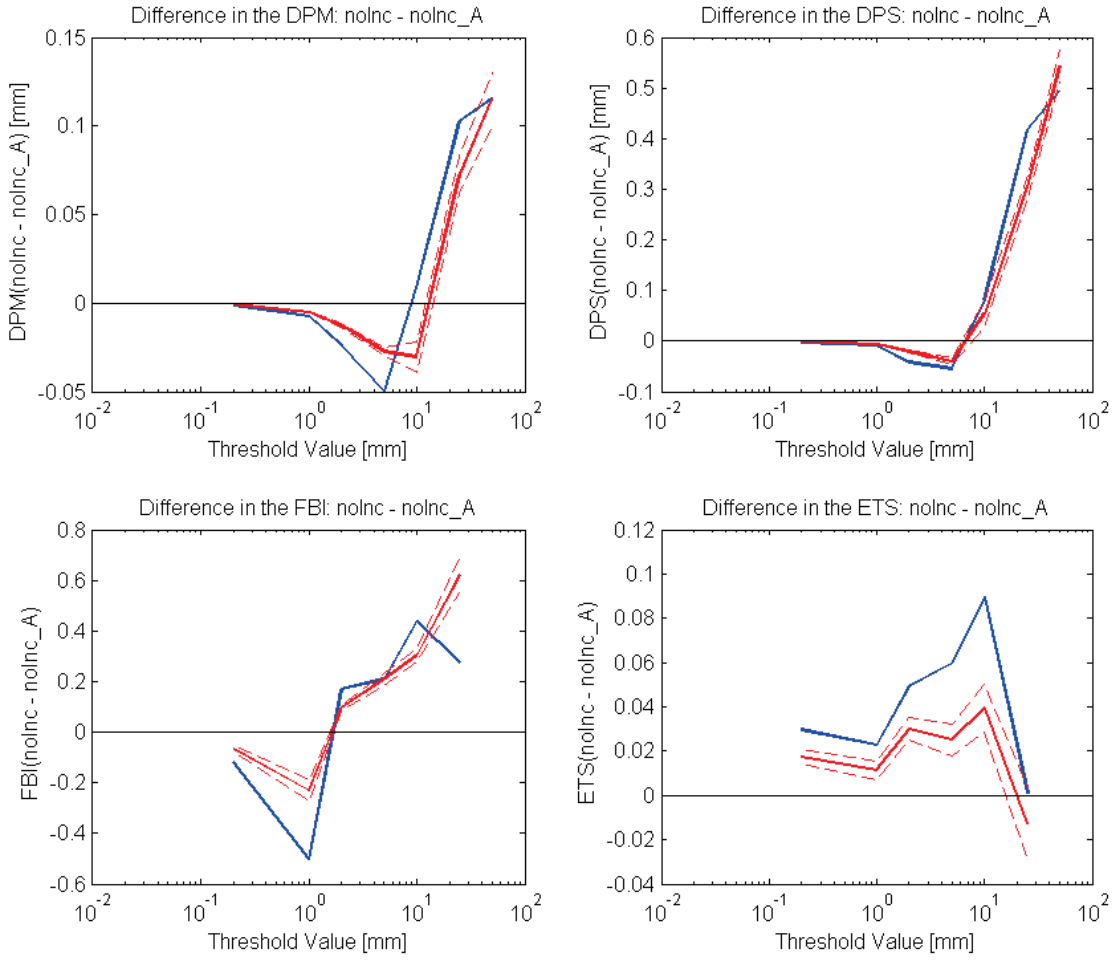

Figure G.3: Anisotropy investigation difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *noInc* and *noInc_A* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

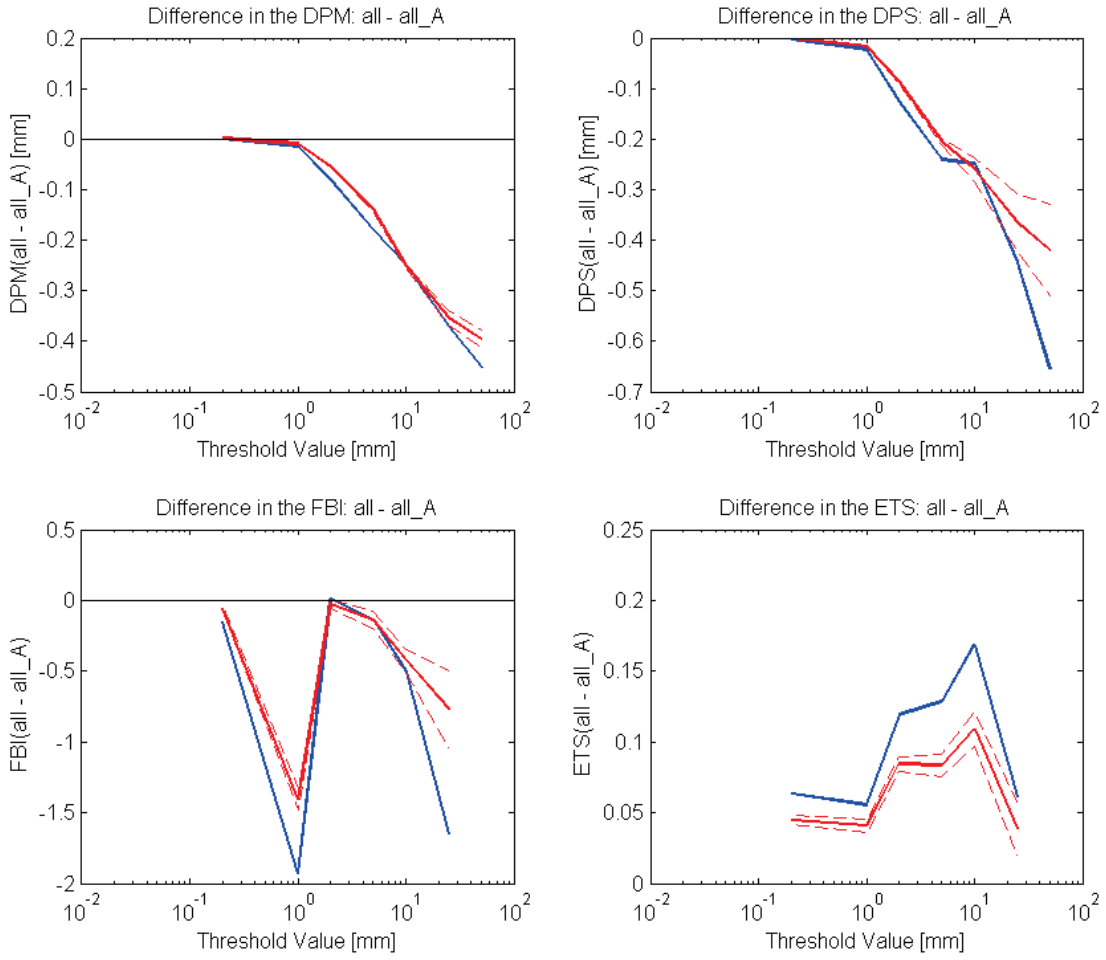

Figure G.4: Anisotropy investigation difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *all* and *all_A* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

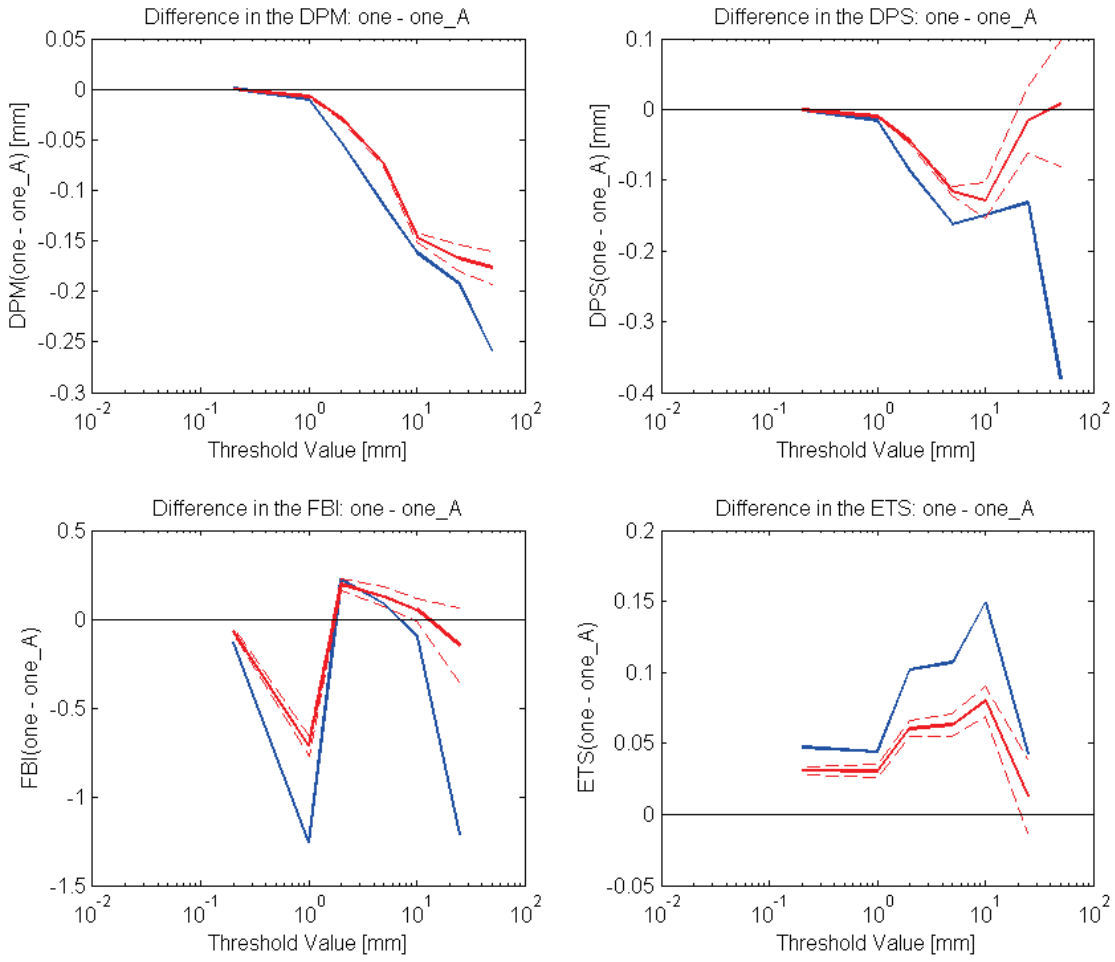

Figure G.5: Anisotropy investigation difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *one* and *one_A* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

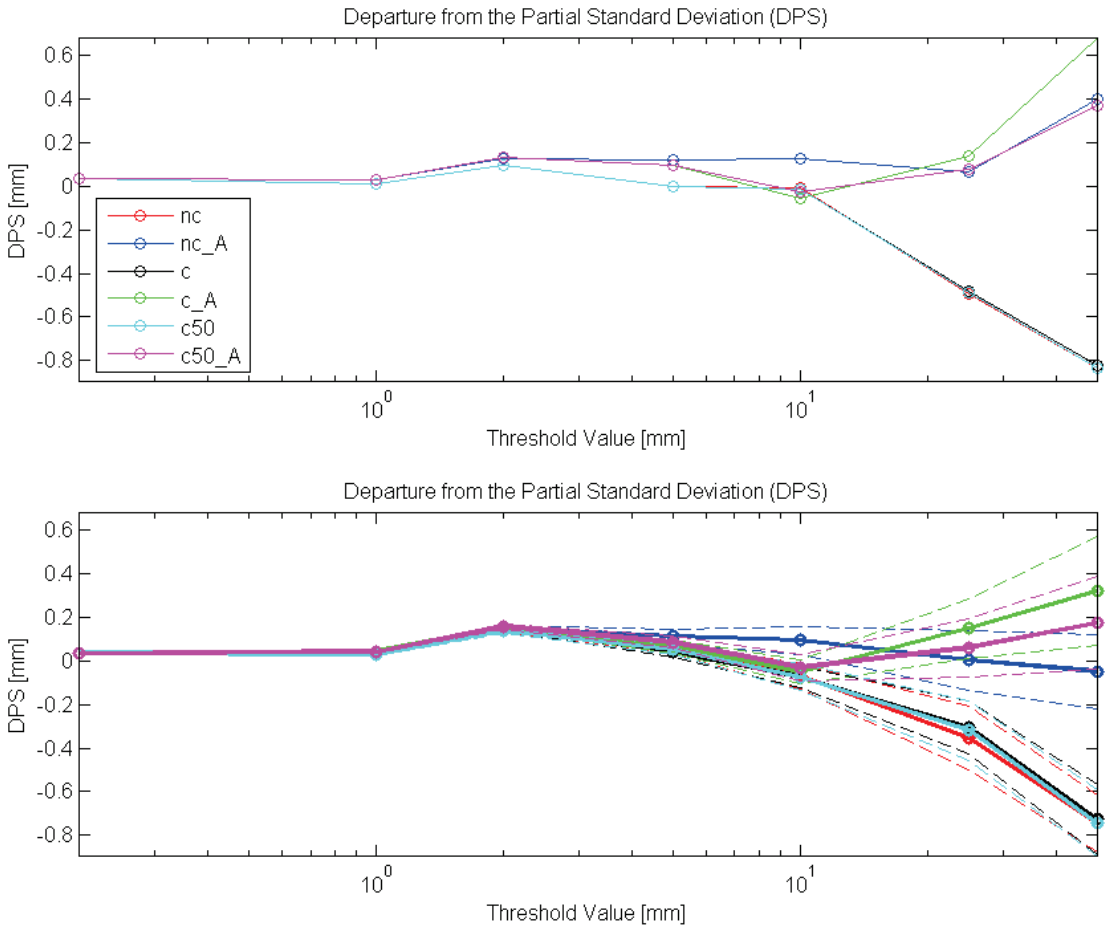

Figure G.6: Convection filtering DPS results. **Top:** DPS results averaged across all verification locations and all time periods in the study. **Bottom:** DPS bootstrapping results, with the bootstrapped mean DPS values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

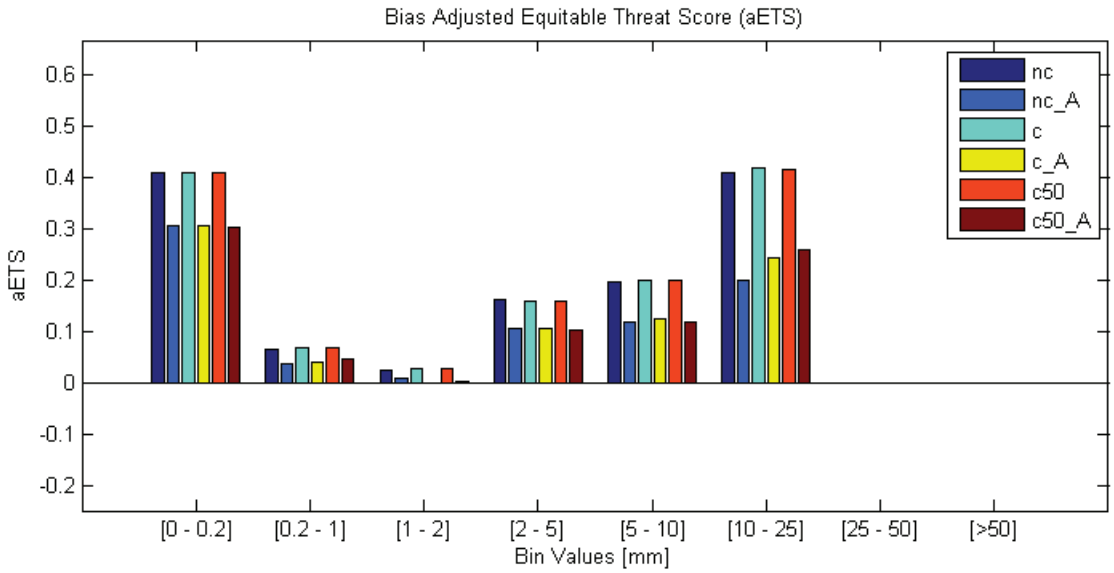

Figure G.7: Convection filtering aETS results averaged across all verification locations and all time periods in the study

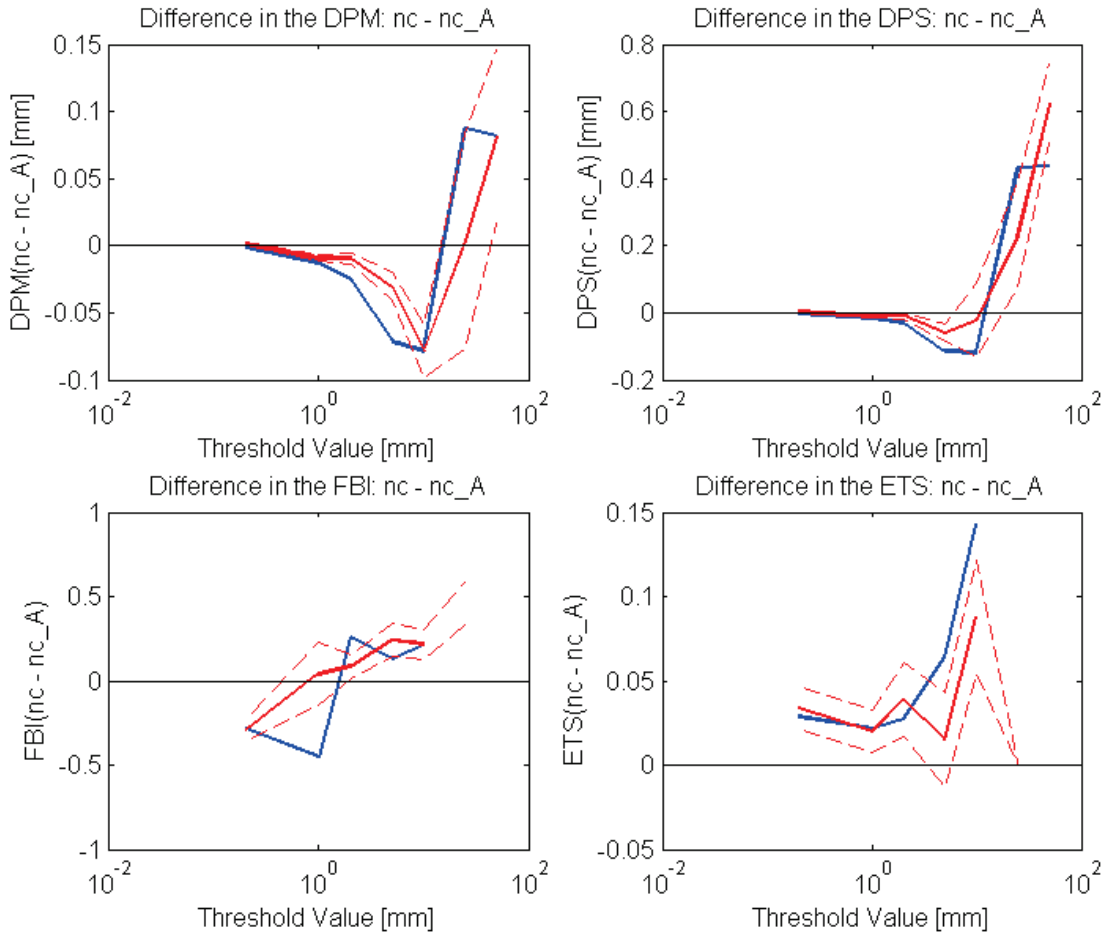

Figure G.8: Convection filtering difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *nc* and *nc_A* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

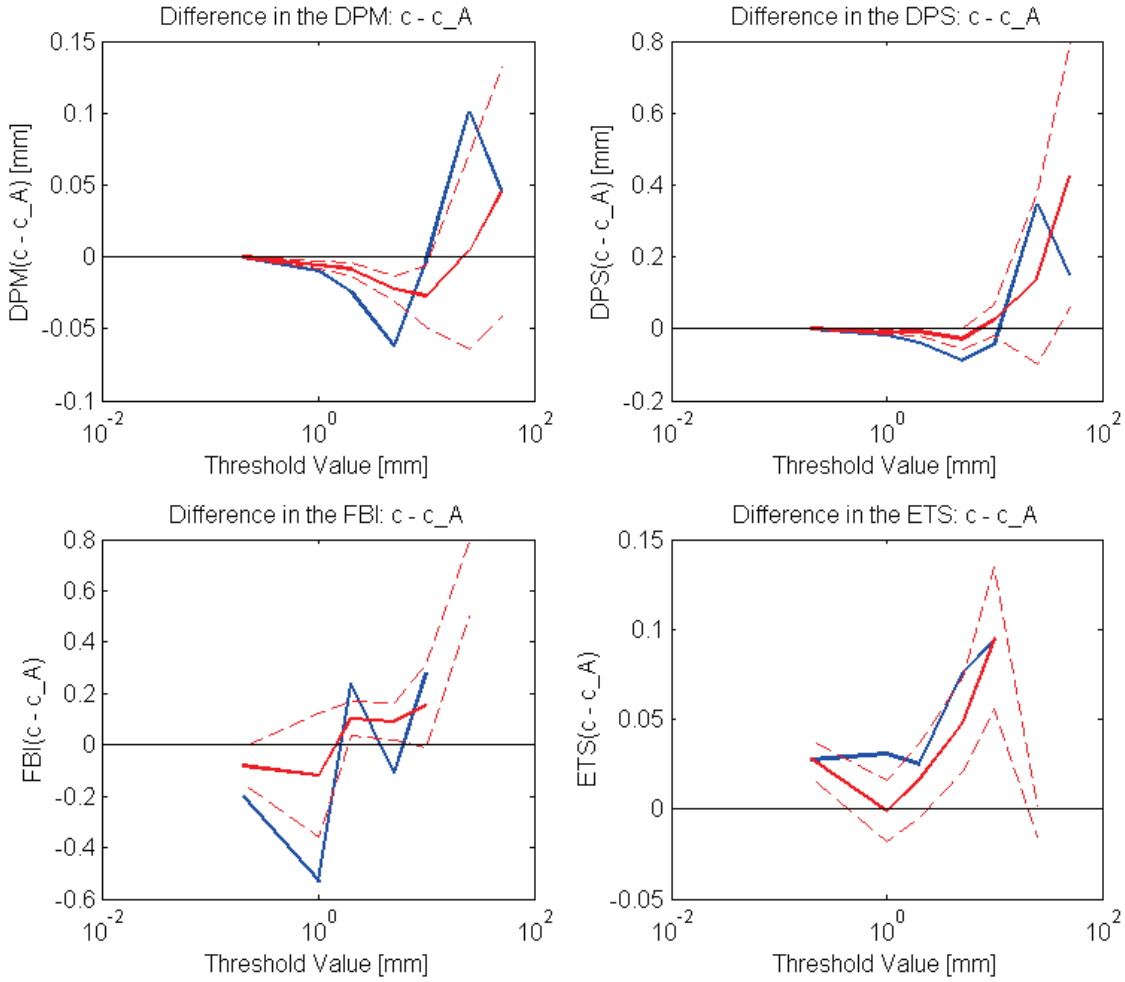

Figure G.9: Convection filtering difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the c and c_A analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

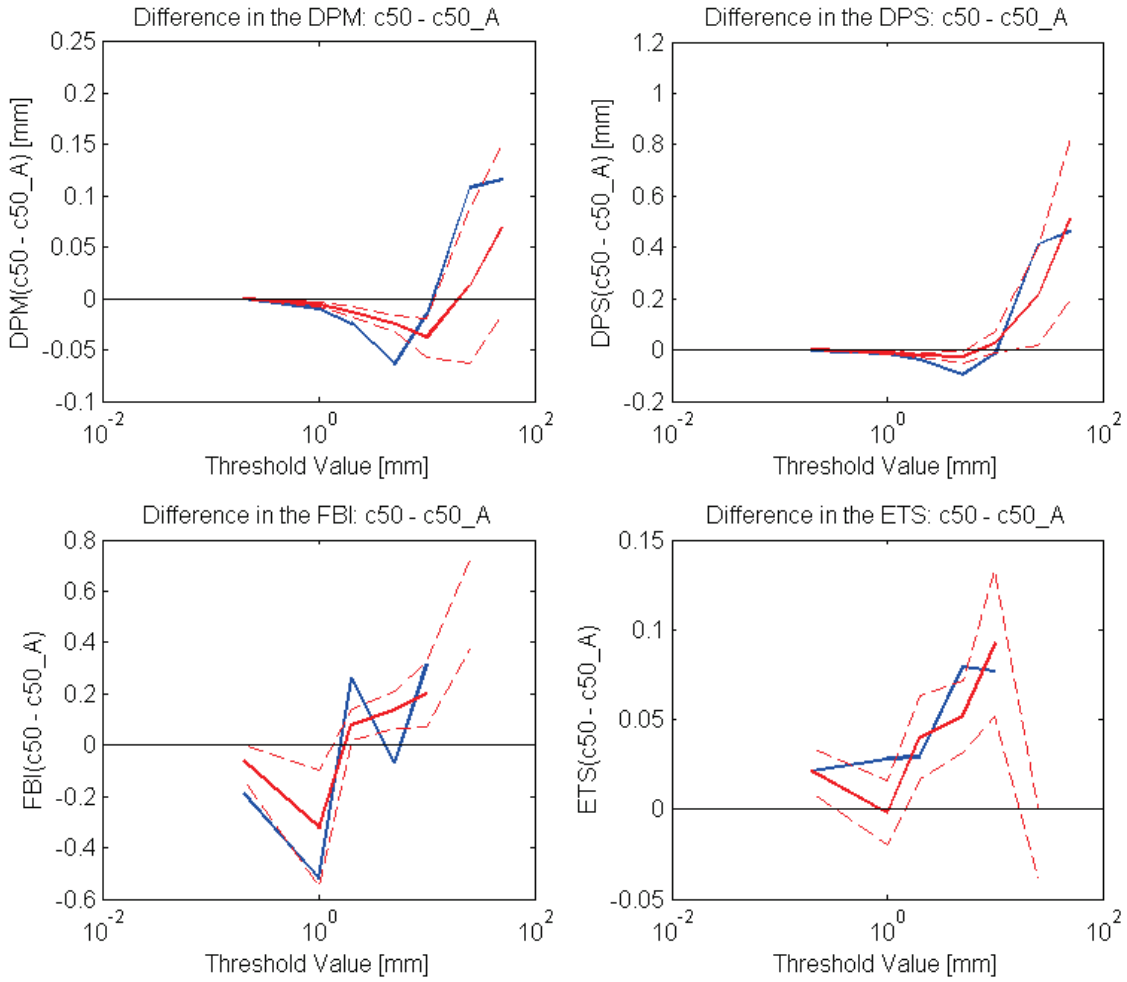

Figure G.10: Convection filtering difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *c50* and *c50_A* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences

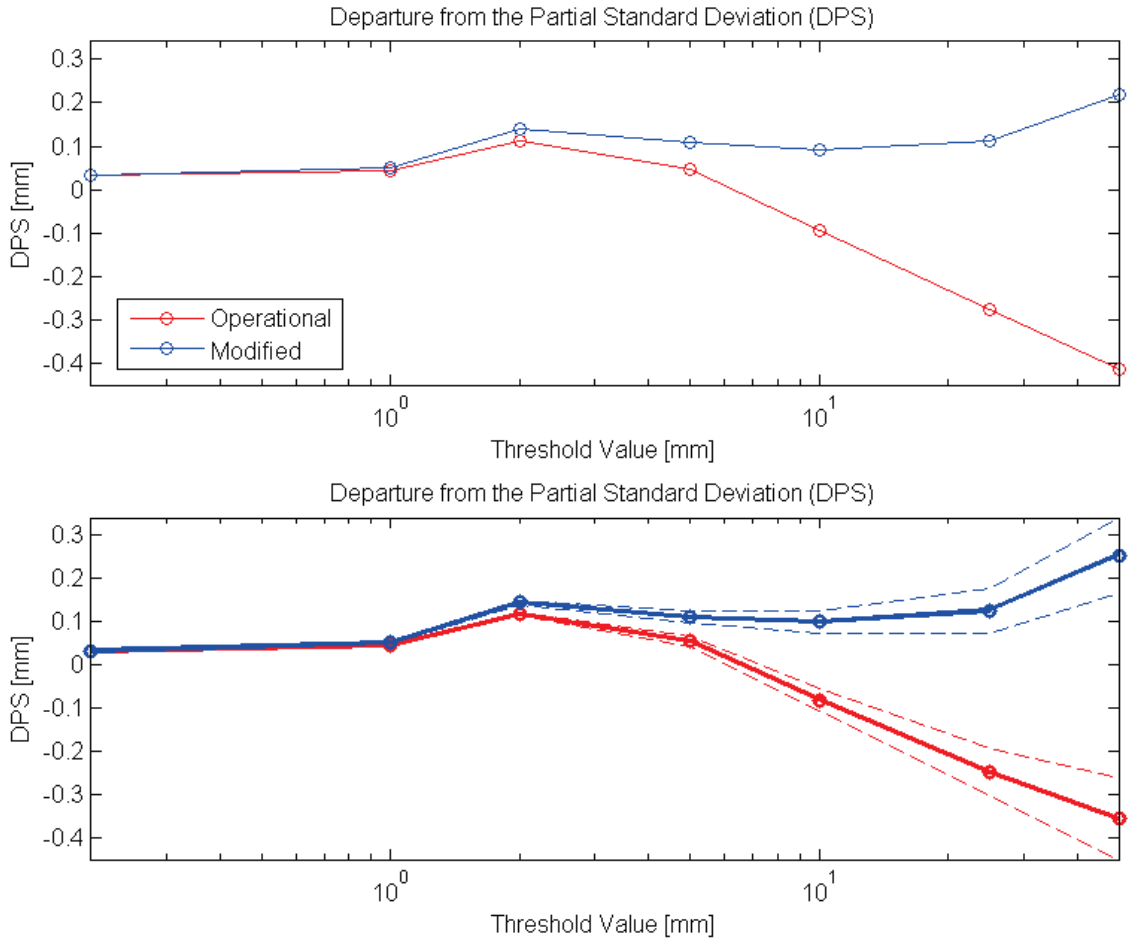

Figure G.11: DPS results for the *operational* and *modified* cases. **Top:** DPS results averaged across all verification locations and all time periods in the study. **Bottom:** DPS bootstrapping results, with the bootstrapped mean DPS values shown in bold, and the 10% and 90% confidence limits shown by the dashed lines

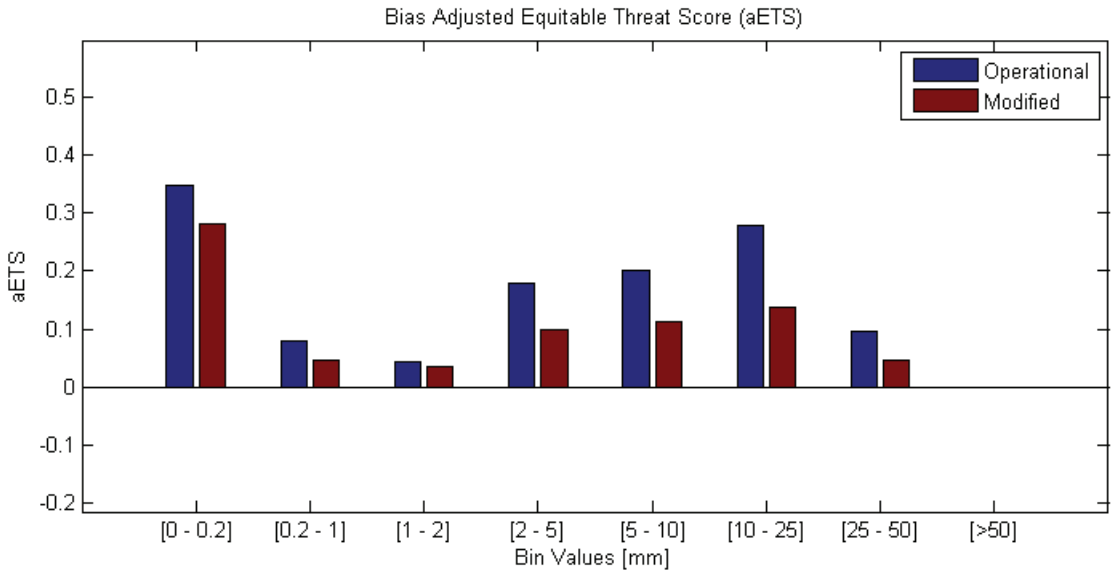

Figure G.12: aETS results averaged across all verification locations and all time periods in the study, for the *operational* and *modified* cases

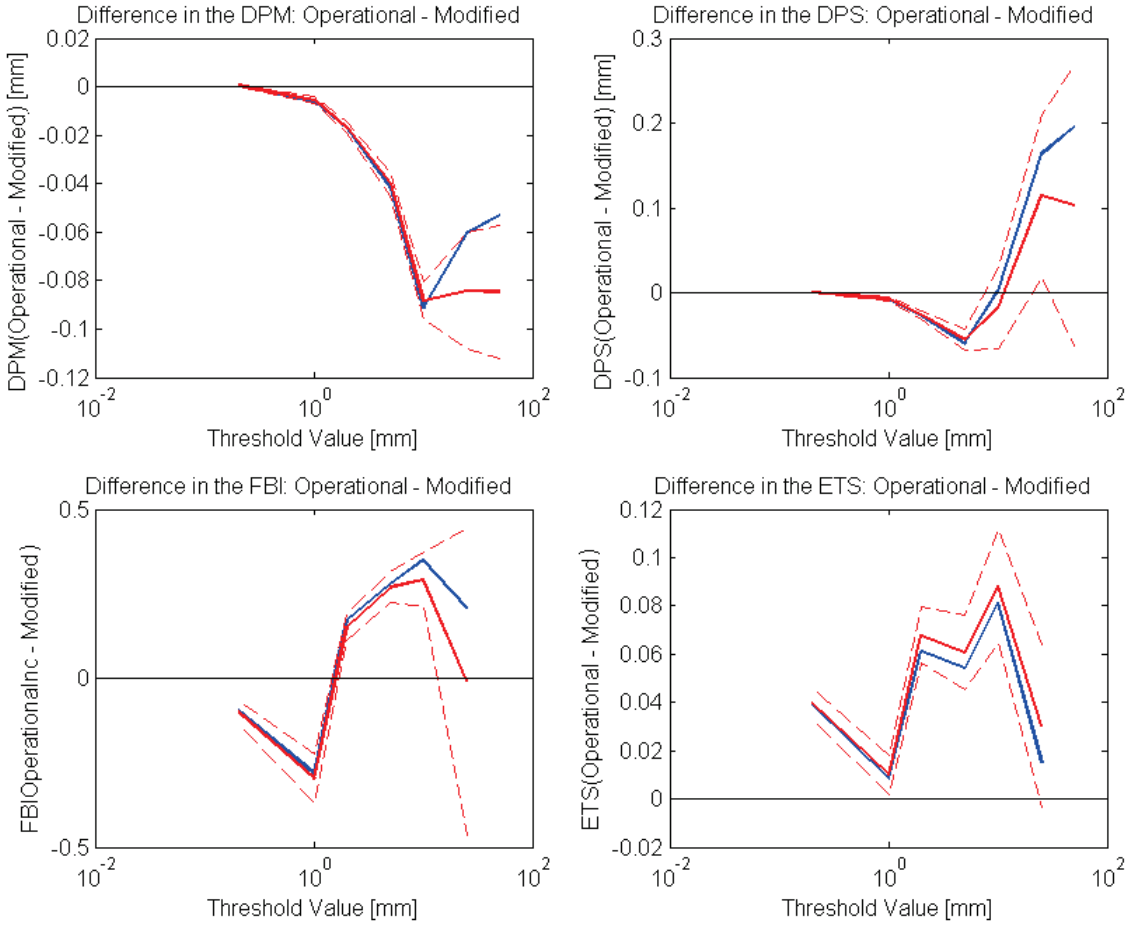

Figure G.13: Difference in the DPM (top left), DPS (top right), FBI (bottom left) and ETS (bottom right) skill score metrics between the *operational* and the *modified* analyses. The blue line represents the difference between the actual analysis skill scores, while the bold and dashed red lines represents the mean and 10% and 90% confidence limits of the bootstrapped differences