Evaluation of Reanalysis Precipitation Estimates in the Canadian Precipitation Analysis (CaPA)

by

Hyaesun Choi

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

Canadian Precipitation Analysis (CaPA) has been developed by Environment Canada to produce the most accurate near-real-time gridded precipitation estimates. It uses the Global Environmental Multiscale model (GEM) as a background and assimilates the synoptic network of weather stations through Optimal Interpolation. Accurate estimation of gridded precipitation is useful for hydrological modeling, stream flow forecasting, and climate change studies. However, the calibration and validation of hydrologic models requires long temporal coverage of data for a better performance. Since GEM/CaPA data are available only for the recent past (2002-present), the development of historical data sets starting earlier than 2002 becomes important. Using alternative models for producing the atmospheric gridded background is one solution to overcome the short temporal coverage of archived GEM data. This thesis evaluates and analyzes two candidate data sets. ERA-Interim and NARR were selected as potential alternatives to GEM background. The general conclusion of the study is that the use of ERA-Interim and NARR as background fields leads to performance results that are not significantly inferior to GEM after assimilation with stations in the CaPA framework. While result with the GEM background remains the best, one can cautiously conclude that for most practical applications, ERA-Interim and/or NARR may be used for the period that predates archived GEM data. The thesis

presents a more detailed evaluation of ERA-Interim and NARR for different seasons and different regions of Canada.

Acknowledgements

First and foremost, I would like to give thanks to the Almighty God for getting me through this long journey.

I would like to express my deepest gratitude to my advisor, Dr. Peter Rasmussen, from the Department of Civil Engineering, for providing not only the great guidance for my Master's study but also the encouragement during the difficult moments in my personal life.

My thanks are extended to the committee members, Dr. Tricia Stadnyk from the Department of Civil Engineering, and Dr. Ronald Steward from the Department of Environment and Geography for providing valuable comments.

I would like to thank Manitoba Hydro and the National Sciences and Engineering Research Council (NSERC) for their support of this project. I would also like to thank Dr. Vincent Fortin from the Meteorological Research Division at the Canadian Meteorological Centre, Environment Canada, for sharing knowledge of the CaPA system, and Post doctoral fellow, Alaba Boluwade, for the support and encouragement toward the completion of my study.

Last but not the least, I would like to thank my family for their love and patience during my study. This completion would not be possible without their support and encouragement. Thank you all.

To my family

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Water resources is one of main natural resources in Canada. There are over one million lakes and massive ice fields in Canada which represent approximately nine percent of the world's supply of fresh water. Water resources in Canada are used in various fields such as water supply, irrigation, and hydroelectric-power generation. Despite the importance of water in Canada, Canada does not have a high-density weather station network due to the environment and low population over the region. To manage the water resources in Canada more effectively, many long-term studies of climate change and hydrological modelling have been undertaken.

There are a lot of studies contributing to determining long-term trends of climate change and the impact of hydrological cycles in Canada. Zhang et al. (2000) studied the trends of temperature and precipitation in Canada during the 20th century and found that the Canadian climate has become wetter and warmer. Fernandes et al. (2007) related such trends of temperature and precipitation to the trend of evapo-transpiration. Furthermore, Latifovic and Pouliot (2007) analyzed climate change impacts on lake processes in Canada and highlighted the strong relationship between lake processes and climate variability, which has important implications for hydrological systems. They also mentioned the significant decrease in the number of observations after 1985. Many studies require long-term historical climate data. Canada is a large country and the population density in many areas is very low, making it difficult to maintain a high-density network of weather stations, which can significantly affect the quality of many studies using climate information.

When observations are missing, reanalysis data sets can be a useful alternative due to their well-distributed spatial and temporal coverage. Reanalysis data are generated from numerical weather prediction (NWP) models and provide a reasonable alternative to observations of precipitation in many studies. Reanalysis data may be used for various purposes such as hydrological modelling calibration, streamflow forecasting, and climate change studies. Those three aspects of reanalysis data use are interrelated. Streamflow forecasting has been investigated using reanalysis data in many studies (Hay and Clark, 2003; Choi et al., 2009a; Dibike and Coulibaly, 2005) which show the usefulness of reanalysis data input when weather stations are sparse. A common tool to forecast streamflow is a hydrological model. Most hydrological models require distributed forcing data which are usually generated by traditional interpolation methods such as Inverse Distance Weighting, Thiessen polygons, Normal-Ratio, and Kriging. There are also more sophisticated methods such as geostatistical and optimal interpolation methods. However, there are some risks to use the methods when the network of data is very sparse; in such cases, reanalysis data can possibly provide practical input. In addition, reanalysis data are useful for climate change studies. Reanalysis data can be used as input to future climate models, provide possible climate patterns, and further simulate possible future streamflow scenarios (Giorgi et al., 1994; Fernandes et al., 2007; Choi et al., 2009b). While reanalysis data are only an approximation of the truth, they may still be useful in areas having few available observations. Reanalysis data sets provide gridded climate information, which can be used for studies requiring spatially well-distributed climate information. Since many observed records of precipitation in Canada suffer from spatially and temporally missing information, useful data to fill those gaps are necessary.

A common caution one has to be aware of with the use of reanalysis products is the existence of model bias and bias in data assimilated into the model. Bias must be estimated using a common reference and corrected using a proper method. There are many studies dealing with bias correction of reanalysis products. Gutjahr and Heinemann (2013) and Teutschbein and Seibert (2012) evaluated different bias correction methods to a regional climate model, applied them to raw model data, and compared performances. Teutschbein and Seibert (2012) especially focused on the performance of different methods for hydrological streamflow simulations. Ines and Hansen (2006) introduced a simple method to remove bias of daily rainfall intensity in a climate model, called a multiplicative shift method. It uses a scaling factor derived from mean monthly climate information for each calendar month. This thesis used the multiplicative shift method.

However, when the density of observations is sparse, the method is not easy to implement. Scaling factors can only be calculated at the locations of observations, but must be applied to the entire grid to determine scaling factors for all grid points. This study employed an interpolation method inspired by the Australian National University Spline (ANUSPLIN) model. ANUSPLIN is based on the thin-plate spline interpolation algorithms developed by Hutchinson (1995). ANUSPLIN is a good example of a smooth interpolation method based on thin plate splines that has been used to produce gridded surface data in Canada. The advantage of the thin plate spline method is that it provides spatially continuous surface without manual tuning using closed-form solution. Therefore, this study used the thin plate spline method to interpolate scaling factors to correct bias of reanalysis products on each grid using observation data set.

1.1.1 Description of Canadian Precipitation Analysis (CaPA)

Environment Canada has developed a precipitation data assimilation system called the Canadian Precipitation Analysis (CaPA) that aims to provide 6-hourly or 24-

hourly gridded precipitation products over Canada. The project was initiated in November 2003 by the Meteorological Research Division (MRD) and the Meteorological Services of Canada (MSC) of Environment Canada and the operational version is available at the Canadian Meteorological Centre (CMC) starting April 2011. CaPA assimilates and quality-controls information of observations from various sources such as SYNOP, METAR and RMCQ (Réseau Météorologique Coopératif du Québec) into a model using enhanced statistical interpolation (SI) techniques (Mahfouf et al., 2007). CaPA data are particularly useful in areas where the network of observations is sparse. CaPA provides enhanced weather data set for studies of hydrological model calibration, streamflow forecasting, and climate change downscaling specially for Canada. In addition, CaPA data are used as input to the Canadian Land Data Assimilation System (CaLDAS) which provides information such as soil moisture, soil temperature and snowpack across Canada. It uses output from the Global Environmental Multiscale (GEM) model as a first guess and then assimilates information from the synoptic network of weather stations to produce a refined estimate of gridded precipitation.

Although CaPA provides relatively fine grid resolution (15km, now available in 10km) which is an improvement over existing reanalysis data sets, it is available only for the recent past (2002 to present). This poses an issue when using CaPA for long-term studies where longer historical data series are preferred. One of the difficulties of calibrating hydrological models of large Canadian watersheds is the temporal limitations of climate input. CaPA, for example, is limited to climate data after 2002. This temporal limitation is from the lack of availability of GEM prior to 2002. Therefore, the motivation of this study is to overcome this limitation of CaPA. Several alternative climate data sets will be evaluated and assessed in terms of extending the time series of CaPA.

1.2 Objectives

Objectives of this study are described below:

- Evaluate reanalysis products and select the best candidates to use in Canada;
- Develop a proper methodology to create bias-corrected gridded precipitation surface from selected reanalysis products;
- Develop alternative CaPA analyses from the selected reanalysis products using the CaPA framework

1.3 Thesis Structure

An outline of the thesis structure is explained in this section. ERA-Interim and NARR reanalysis data sets were used in this study as potential alternative background of CaPA (GEM). The two data sets were subjected to pre-processing in the form of bias correction before the CaPA framework application. Analysis of the data sets was conducted before and after pre-processing. The CaPA framework was then applied to both original and pre-processed data sets. The results were analyzed over different regions and seasons. Various analysis methods were performed but particularly focused on two categorical scores, the Equitable Threat Score (ETS) and the Frequency Bias Index (FBI), to estimate the forecasting skills. The analysis was divided into four cases, as shown in Figure 3.1. Each case was analyzed over different regions and seasons.

This study is divided into three major steps. As a first step, evaluation and comparison of reanalysis candidates (ERA-Interim and NARR) with observations (CDCD) and GEM was carried out. The study analyzed daily and monthly precipitation from ERA-Interim and NARR using observed data as a reference with the goal to find a proper bias correction method. Before the CaPA assimilation method is applied using ERA-Interim and NARR for the common period (2002-2011), bias correction was implemented using the simple multiplicative shift scheme and statistical interpolation. This step is described in Chapter 3. Once biases are corrected from ERA-Interim and NARR, observations will be assimilated with ERA-Interim and NARR using the CaPA framework. This process requires a variogram analysis. Approval of access to Pegasus was required for the process and then running the CaPA framework using original and corrected ERA-Interim and NARR was done. The analysis and evaluation of the CaPA analyses in five cases (original ERA-Interim, original NARR, corrected ERA-Interim, corrected NARR, and GEM) will provide the foundation for selecting the best candidate to run CaPA back in time to 1979.

Chapter 2

Study domain, Time period, and Data

2.1 Study domain and time period

The present investigation considers the performance of CaPA across Canada. Since CaPA data are available from 2002 and the latest year of availability of the data set was 2011, the 10-year period from 2002 to 2011 was selected for the study. The data sets in daily (24h) time steps were chosen instead of 6-hours time steps due to the long study period (10 years).

For the purpose of this study, Canada was divided into four regions based on characteristics of Canadian climatology and topography as shown in Figure 2.1. Western, Prairie, Central, and Atlantic regions are descried as red, green, blue, and purple box, respectively, in Figure 2.1. The Western region is located along the

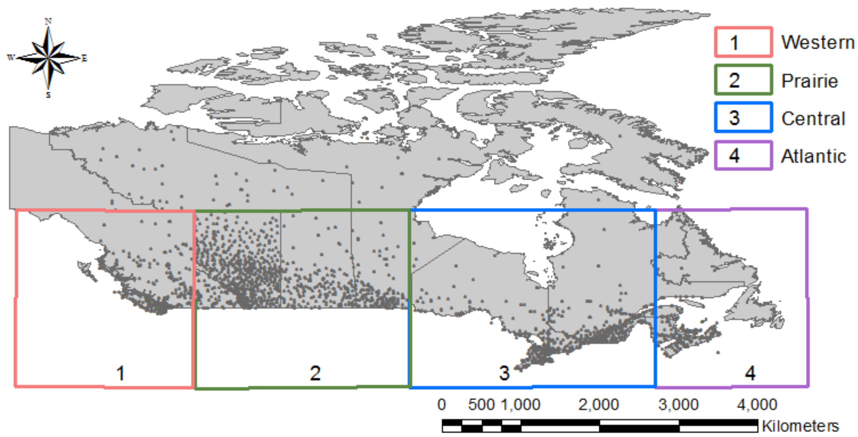


Figure 2.1: Division of regions over Canada

Pacific Ocean and has a humid mild climate. Humid air from the Pacific causes vast amounts of orographic rainfall in mountainous areas. Summer precipitation is less than winter precipitation in the region because low pressure systems moves northward in the summer.

The Prairie region consists mainly of the interior plains. The region has a continental climate with cold winters, hot summers, and relatively low precipitation since the west-coast mountains blocks marine air moving from west to east. Precipitation in spring and summer is higher than in winter. In summer, convection is the main source of precipitation.

Upper side of the Prairie region is heavily influenced by the Arctic Ocean. The region has very short and cool summers and winters lasting up to ten months. The amount of precipitation is low and evaporation is limited since the Arctic Ocean is frozen most of year. The number of stations near the Arctic Ocean is few and the stations contain a large amount of missing information. The main analysis of the study does not consider the region above 60N latitude as seen in Figure 2.1.

The Central region shows both continental and maritime characteristics. The region have more rainfall than the Prairie region due to the water vapour from the Great Lakes, Hudson Bay, the Atlantic Ocean, and the Gulf of Mexico. The amount of snowfall in winter is similar to summer precipitation since winter is not as cold as the Prairie region, which creates less dry climate in winter.

In the Atlantic region, maritime characteristics are dominant since the Atlantic Ocean creates a cool and humid climate. Most of the precipitation in the area is cyclonic and evenly distributed over the year. The region is characterized by few thunderstorms, a little orographic rainfall due to low Appalachian Mountains and generally less rainfall than at the west coast due to dominant offshore wind.

2.2 CaPA and GEM

As described in Chapter 1, CaPA and GEM were developed by the Meteorological Research Branch (MRB) in partnership with the Canadian Meteorological Centre (CMC) of Environment Canada to generate better weather products for Canada.

To compensate for the short period of time coverage, this study intends to use alternative backgrounds for CaPA analysis, which later on will be compared to the original CaPA background (GEM) and CaPA analysis. The CaPA framework can be executed in either 6-hour or 24-hour accumulation time frames. This study uses

24-hour accumulations of GEM and CaPA. The version of CaPA used is 2.4b8 (currently, the 2.5 version of CaPA is available) along with the use of 'capaverif-2.0.0' for the verification process. In this thesis, 'Forecast', 'CaPA analysis' will be referred to a background model and output of CaPA framework, respectively. Côté et al. (1998b), Côté et al. (1998a), Mailhot et al. (1997), Mailhot et al. (2006) describe the development of Canadian forecasting system by CMC including the increase of both the horizontal and vertical resolutions and the improvements of the physics.

Currently, radar data assimilation has been undertaken in the operational version of CaPA. Fortin et al. (2014) shows the positive improvement after the assimilation of radar data which is not surprising as the performance of atmospheric models heavily depends on how much data are assimilated. There are some challenges with satellite data due to the difficulties to detect frozen precipitation in winter which is an important part of the hydrological cycle. Since Canada has long winters, there are some limitations to use the data as most of results related to satellite studies focus on seasons which have only liquid precipitation. The analyses performed in the thesis do not consider radar and satellite data.

2.3 Observation data

Three observation data sets were initially considered for this study: Canadian Daily Climate Data (CDCD), Adjusted and Homogenized Canadian Climate Data (AHCCD), and ANUSPLIN.

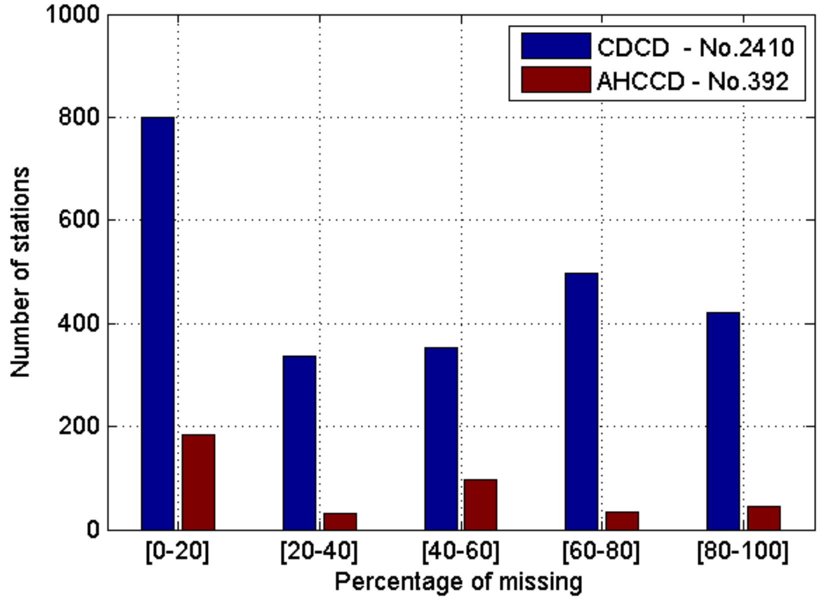


Figure 2.2: Histogram of percentage of missing records

Historical weather and climate data over Canada are available from Environment Canada's website and is commonly referred to as the Canadian Daily Climate Data (CDCD) (Environment Canada, 2013). The earliest year of any station in the CDCD archive is 1840 and it includes more than 7,000 Canadian climate stations. Each station contains the World Meteorological Organization (WMO) identifier as Canadian stations are assigned to standards of WMO. It provides daily temperature, rainfall, snowfall, total precipitation and depth of snow on the ground. CDCD is Canada's official weather and climate data and is used as the primary data source for many climate studies in Canada. The Adjusted and Homogenized Canadian Climate Data (AHCCD) was developed for the studies of climate trends in Canada (Mekis and Hogg, 1999). The amount of daily precipitation is adjusted for almost 500 stations. A number of techniques are applied to handle measurement issues such as under-catch due to wind, evaporation and wetting losses for each type of rain gauge, snow water equivalent from ruler measurements, trace observations and accumulated amounts from several days. Observations nearby are sometimes combined in order to create longer time series. All snow ruler measurements are adjusted by density corrections based on coincident ruler and Nipher measurements (Mekis and Brown, 2010). Accumulated amounts of precipitation are distributed over the affected days in order to prevent misleading extreme values. However, AHCCD data are not in situ station records of the official Meteorological Service of Canada and should be considered with awareness of the adjustment techniques used. From 2002 to 2011, the available number of AHCCD stations over Canada is 392 which is significantly lower than the number of station in CDCD. In addition, the available number becomes even lower if one considers missing records as shown in Figure 2.2. Therefore, AHCCD data was not considered in this study.

ANUSPLIN is a gridded observation data set which is based on a multi-variate non-parametric surface fitting technique for interpolation of various climate parameters such as daily maximum and minimum temperature, precipitation and solar radiation (McKenney et al., 2011). Natural Resources Canada's Canadian Forest Service (CFS), the Australian National University (ANU) and Environment Canada (EC) have collaborated to develop spatially continuous climate data over Canada using the ANUSPLIN model. The ANUSPLIN model uses a trivariate thin-plate smoothing spline approach which permits spatially varying dependence on ground

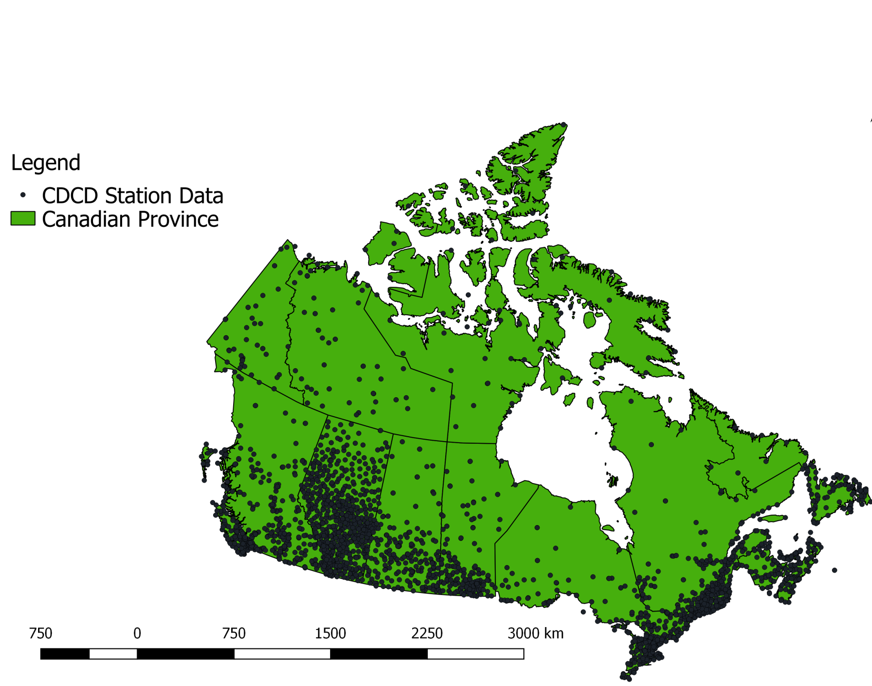


Figure 2.3: Location of Canadian Daily Climate Data (CDCD) Stations over Canada

elevation and optimizes the climate data by minimizing errors using generalized cross validation (GCV) (Hutchinson et al., 2009). The ANUSPLIN model is a useful tool for the Canadian region where large variation in the density of stations exists. However, Hutchinson et al. (2009) found that there are larger errors in the result of the model for northern Canada which has a sparse network of stations. Since this study does not employ processed observation data, it will not use the ANUSPLIN model data as ground truth.

CDCD is selected among the three datasets as a reference due to coverage, time span, and no uncertainty due to data processing. This study focuses on daily (24h) based analysis so the CDCD data set is adequate to use. There are a total of 2610 stations over Canada from 2002 to 2011. The locations of these stations are shown in Figure 2.3. Missing records are not accounted for in the analysis in this study.

In-situ observations are valuable for quantifying variables such as precipitation generated by atmospheric models; it is also useful for adjustment, validation, and verification of the atmospheric models (Rudolf et al., 1996). It is however important to obtain information on sources of errors of these observations and consider them as factors affecting the results of the study. Systematic errors are caused from unfavorable conditions of measurement such as drift of rain or snow particles due to wind under-catch and due to evaporation losses (Rudolf and Schneider, 2005). In addition, records of trace amount of precipitation also introduce some errors. Last but not the least, data transmission and processing will introduce errors to the observation as well (Rudolf and Schneider, 2005).

2.4 Description of reanalysis data sets

Reanalysis data sets are widely used to study weather systems and climate variability as they provide a valuable information from combination of background model and observations (Bosilovich et al., 2008). Background models (also called Numerical Weather Prediction models) offer forecast information which is based on numerical calculation. It then assimilates observation using its own routines. Numerical weather prediction (NWP) models and observed data have their own distinctive advantages and disadvantages. Due to its distributed nature, NWP models are able to provide meaningful information at locations where observations do not exist. However, observed values are considerably more accurate but also more sparse. Reanalysis data utilize the advantages of each of the types of information. They use the NWP data as a background and assimilates the observed data in order to adjust the background value.

Often the main use of reanalysis data is to assimilate observations during the stage of initialization of NWP models. Its application is expanded to other processes of NWP simulations and many improvements have been made to the quality of background models, input variables, and assimilation system for each reanalysis model. Figure 2.4 shows different types of assimilation systems as an example. Reanalysis model performs differently because of the combination of the quality of the components and the method of assimilation. Many studies have been evaluated the performance of different reanalysis data sets for specific purposes (Decker et al., 2012; Bao and Zhang, 2013; Bosilovich et al., 2008; Zhao and Fu, 2006; Ward et al., 2011).

Reanalysis models use observations from a wide variety of sources such as automatic climate stations, aircraft survey, satellite data, and radiosonde measurements. They assimilate key parameters such as vorticity, divergence, temperature, humidity, ozone and derives variables such as precipitation and radiation (Newman et al., 2000). Reanalysis models take gridded fields of an NWP model and adjust them through the processes of data assimilation and bias correction.

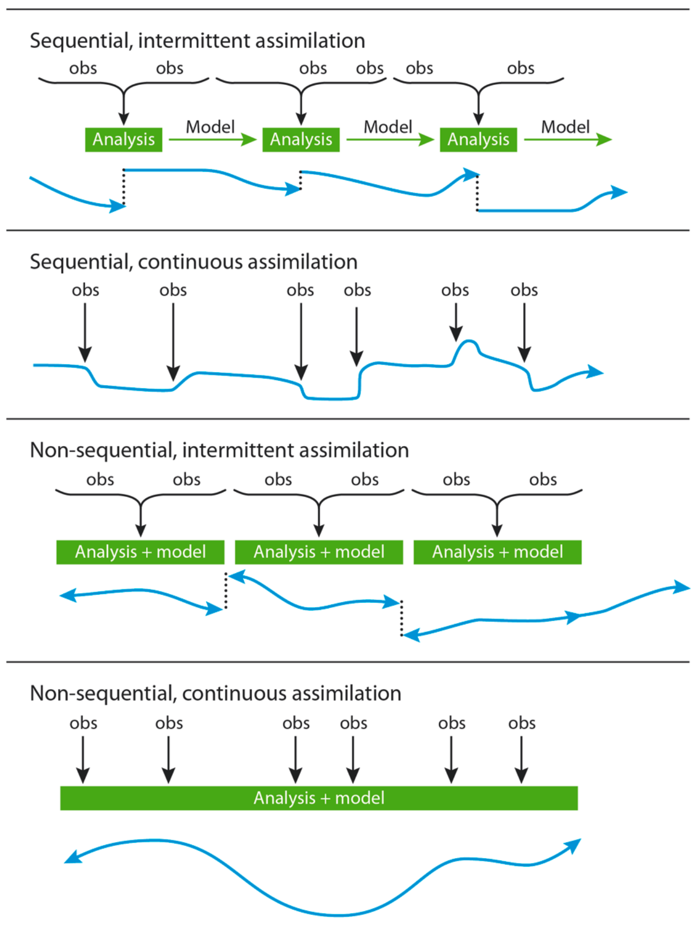


Figure 2.4: Data assimilation process of Reanalysis data (National Center for Atmospheric Research, 2013)

NWP models initialize their simulation at the end of each time step. This process introduces potential inconsistencies to the model. On the other hand, reanalysis models produce data that are consistent over space and over time with the physical processes of the model within the boundaries of the NWP model. This consistency is achieved by conducting Quality Control (QC) and Error Analysis procedures of all observation types. The biggest difference between NWP and reanalysis data is that reanalysis data is consistent with observation time series over longer time periods. During the process of initialization and assimilation, reanalysis models consider not only the consistency of near time but also the flow over longer periods of time (Mesinger et al., 2006). Therefore, such output from reanalysis models is useful.

On the NOAA Precipitation Datasets Attributes website, several gridded reanalysis data sets are available as shown in Table 2.1. However, only those providing precipitation variables at a time step of 24-hrs or less can be used in this study. In addition, some data sets are available free of charge while others, such as the high-resolution ERA-40 or ERA-Interim, have a cost. For this study, only data sets which can be downloadable with no cost were used. More specifically, ERA-Interim in 0.75 degree resolution and NARR in 32km resolution were used in this study. These two datasets have relatively higher resolution than other existing data sets and they provide surface precipitation information at a 24-hr time step. There are a great number of reanalysis data sets produced by NCEP. However, the average resolution of the data sets is 2.5 degree by 2.5 degree Gaussian which is too coarse to use in regional studies. In addition, there are a large number of data sets providing fine resolution as well. However, the given time resolution is larger than 24-hr and/or spatial coverage is limited to the Unites States. There are also temporal coverage limitation. Some data sets cover longer periods of time, but were not considered because they were discontinued after 2002. Therefore, taking into account those six criteria (spatial coverage, spatial resolution, temporal coverage, temporal resolution, availability of surface precipitation information, and cost), the three data sets shown in Table 2.2 were selected.

Since CaPA/GEM is available from 2002, the time period for evaluation starts from 2002 for comparison with NARR and ERA-Interim to CaPA/GEM. The NARR dataset and ERA-Interim dataset were downloaded from NOAA's website and ECMWF's website, respectively. Both ERA-Interim and NARR datasets are provided in NetCDF format. NetCDF defines latitude, longitude, and precipitation values within the defined spatial and temporal coverage. Each grid value is represented by a point at the centre of the grid.

Table 2.1: List of data sets on the NOAA Precipitation Datasets Attributes website

Datasets	Areal Coverage	Grid Size	Time Step	Period of Record
NCEP/NCAR Reanalysis 1	Global	2.5 deg	4× Daily	1948-present
NCEP/NCAR Reanalysis 2	Global	2.5 deg	4× Daily	1979-present
CPC .25 × .25 Daily US Unified Precipitation	U.S.	0.25 deg	Daily	1948-2006
CPC Hourly Precipitation	U.S.	2.0 deg ×2.5 deg	Hourly	1948-2002
U. of Delaware Precipitation and Air Temperature	Global Land	0.5 deg	Monthly	1901-2014
NARR	Northern Hemisphere	32 km	8× Daily	1979-2012
ERA-Interim	Global	0.75 deg	2× Daily	1979-present
CaPA/GEM	Canada	15 km	4× Daily	2002-present

Table 2.2: List of data sets used

Datasets	Areal Coverage	Grid Size	Time Step	Period of Record
NARR	Northern Hemisphere	32km	8× Daily	1979-2012
ERA-Interim	Global	0.75 deg	2× Daily	1979-present
CaPA/GEM	Canada	15km grids	4× Daily	2002-present

Prior to the use of the reanalysis products, assessment of the quality of the data sets is necessary. Data quality is a function of uncertainties contained in numerical weather prediction models (NWP), input data, and the process of data assimilation (Hodges et al., 2011). Daily precipitation from the two candidates were evaluated and compared to observations and GEM. This process includes estimating forecasting skills which will be explained in detail in Chapter 3. The goal of the evaluation is to determine which dataset is the best candidate to be used as an alternative to GEM.

Many reanalysis data sets have been developed for the region of North America. However, most reanalysis projects have focused on the United States and not extended to Canada.

2.4.1 NARR

The North American Regional Reanalysis (NARR) is produced by the National Center for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR). NARR uses the ETA model and the ETA model's 3D-VAR data assimilation system (EDAS). The NARR data set covers the period of 1979 to present and is being continued in real time a temporal resolution of 3 hours and a spatial resolution of 32km/45 layer. This study uses daily accumulated total precipitation

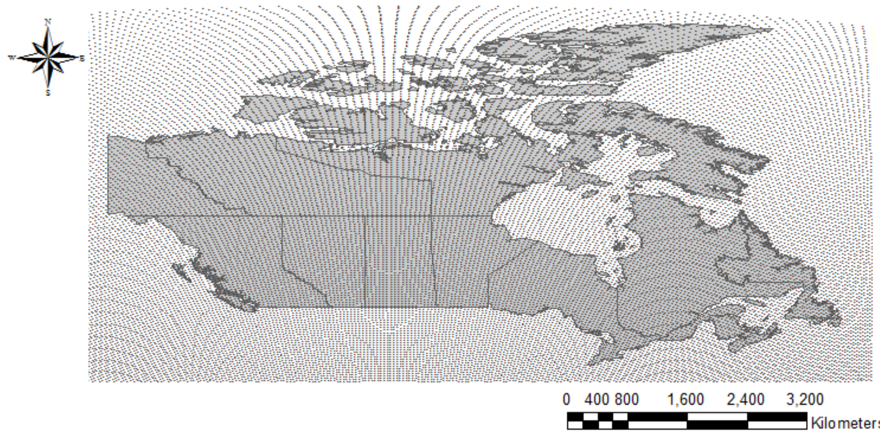


Figure 2.5: NARR Resolution over Canada

at the surface with a horizontal resolution of 32km for the North American domain. The regional reanalysis is based on the NCEP-NCAR Global Reanalysis. NARR uses precipitation observations for assimilation into the atmospheric analysis.

The production of NARR data involves: the lateral boundaries from the NCEP-DOE Global Reanalysis, the NCEP Eta Model and its Data Assimilation System, a recent version of the Noah land-surface model, and the use of a number of additional data sets. In addition, the land-surface model used in NARR provide enhanced representation of the interaction between land hydrology and land-atmosphere. NARR also provides improved atmospheric circulation throughout the troposphere. Particularly, NARR introduces high-quality and detailed precipitation observations into the atmospheric analysis, which promotes hydrological analysis through more accurate forcing to the land-surface model component (Mesinger et al., 2006).

The North American Regional Reanalysis (NARR) take a great amount of observed information into account. In addition to data sets used for NCEP-DOE Global

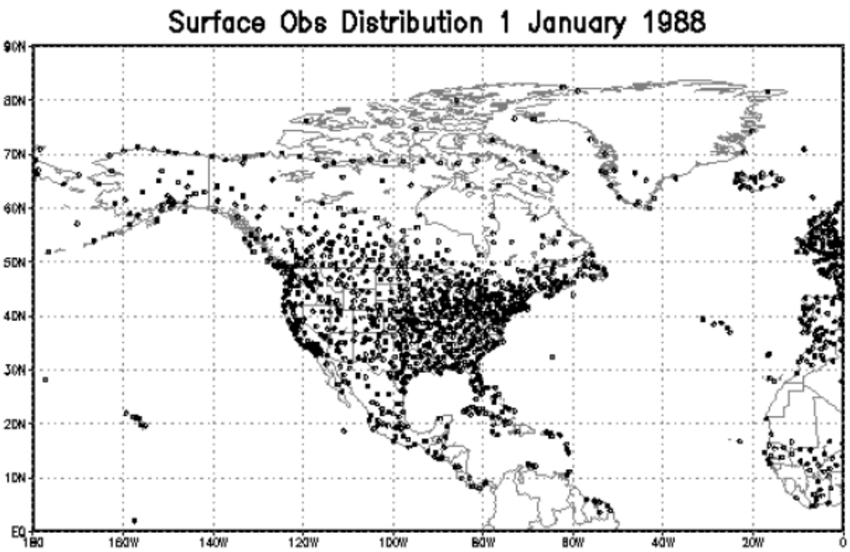


Figure 2.6: Distribution of surface observations assimilated in the NARR 1 January 1988 (Source: Mesinger et al. (2006))

reanalysis, such as temperature, wind, moisture, pressure from rawinsondes, dropsondes, pibals, aircraft, surface, and geostationary satellites, NARR adds data sets from TOVS-lb radiances, COADS (Ship and buoy data), Snow depth by Air Force, Sea-surface temperatures (SST), and Sea ice. It contains SST of the Great Lakes as well as Canadian lake ice from the Canadian Ice Center.

Perhaps the most successful improvement in NARR is the use of observed precipitation information obtained from various sources. Over the region of the United States, it uses a 1/8-degree gauge dataset which is analyzed using PRISM and a least-squares distance weighting algorithm, while 1-degree gauge datasets are used over Canada and Mexico. For the rest of the NARR domain (Ocean), it uses CMAP information which is combined with satellite and gauge precipitation (Mesinger et al., 2006).

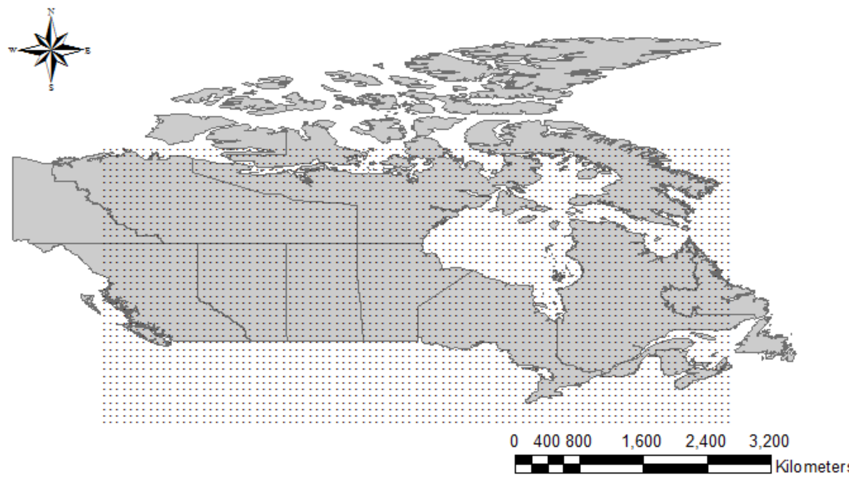


Figure 2.7: ERA-Interim Resolution and Coverage over Canada

The distribution of surface observations assimilated into NARR is shown in Figure 2.6. The precipitation information used, converted into latent heat in the model, results in precipitation patterns similar to observed ones in summer and winter as well as improves the representation of the hydrological cycle. However, the use of gauge precipitation over Canada was discontinued at the end of 2002 (Mesinger et al., 2006). Therefore, precipitation over Canada is not assimilated since then.

2.4.2 ERA-Interim

ERA-Interim is the most recent global atmospheric reanalysis product developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA-Interim uses a 4D-VAR system for the atmospheric analysis and a 6 hourly 3D-VAR system for data assimilation with different resolutions. The ERA-Interim project was initially implemented to overcome some problems of ERA-40 and is currently completed for the period back to 1979, with plans to extend the analysis back to the early twentieth century. In addition, the ERA-Interim project aims to test different technical aspects such as data selection, quality control (QC), bias correction, and performance monitoring as well as how each of these components impacts the quality of reanalyses (Dee et al., 2011). ERA-Interim improves several aspects ERA-40 has some difficulty with: the representation of the hydrological cycle, the quality of the stratospheric circulation, and the consistency in time of the reanalyzed geophysical fields (Dee et al., 2011). The most successful contributions for the improvements are the development of data assimilation system from 3D-VAR to 4D-VAR as well as improved variational bias correction of the satellite observation (Betts et al., 2009). ERA-Interim covers the period from January 1979 to present. Each analysis cycle starts at 00:00 UTC and 12:00 UTC with 3-h intervals. ERA-Interim uses 12-hourly analysis cycles for sequential data assimilation. Each cycle assimilates available observations into the forecast model information. The information is calculated from the combination of upper level atmospheric fields, near-surface fields, soil moisture and soil temperature, snow, and ocean waves. It then provides the initial state of a short-range model forecast which is used for the next analysis cycle.

Many studies show the use of the ERA-Interim data set over various regions. Szczypta et al. (2011) verified the daily precipitation estimates and the Incoming Solar Radiation (ISR) from ERA-Interim with the SAFRAN reanalysis data set and the Brion reference ISR products. Szczypta et al. (2011) found a good agreement over France in ERA-Interim. Rubel and Rudolf (2001) evaluated two different global gridded estimates of daily precipitation: the Global Precipitation Climatology Projects (GPCP) and ECMWF model predictions which is used for ERA-Interim as a background, over the European Alpine region. Rubel and Rudolf (2001) concluded that the ECMWF prediction shows high performance of forecasting in terms of the spatial distribution of precipitation. In addition, Decker et al. (2012) evaluated various reanalysis products using flux tower observations. Their study indicated that ERA-Interim shows the best result in terms of air temperature, 6-hourly wind speed, latent flux at monthly averaged diurnal time scales, incoming shortwave radiation, among other data sets such as MERRA, ERA-40, GLDAS, NCEP, and CFSR. The result interestingly shows that ERA-Interim performs better than MERRA in some aspects even thought MERRA assimilates satellite rain rates and uses a smaller grid than ERA-Interim.

This thesis uses ERA-Interim precipitation (parameter: total precipitation) with $0.75^{\circ} \times 0.75^{\circ}$ resolution (finest) (shown in Figure 2.7) and 12-h intervals starting at 00:00 UTC and 12:00 UTC.

ERA-Interim uses a 12-hourly 4D-VAR data assimilation system at the upper-air atmospheric state to overcome several problems of ERA-40. In contrast to the 3D-VAR assimilation system, 4D-VAR takes into account tendencies from the mass field, which allows the system to use observations more efficiently (Rabier et al., 2000). Several studies show that 4D-VAR systems outperform 3D-VAR systems in data-sparse cases and are able to produce more accurate output of the large-scale tropospheric circulation using only observations of surface pressure.

Currently, advanced reanalysis data sets from ECMWF are available such as ERA-20C Land, ERA-SAT, CERA-20C, and CERA-SAT in improved resolution and longer time coverage. They are suggested to be used in future studies, as discussed in Chapter 7.

Chapter 3

Methodology

An overview of the work flow in this study is shown in Figure 3.1. To assess the reliability of ERA-Interim and NARR as an alternative to GEM, analysis and evaluation were conducted following three steps. Anomalies, RMSE, and correlation for daily precipitation are examined based on different regions and seasons. In addition, Q-Q plot will be presented for different regions and different seasons. Q-Q plots are used to compare probability distributions of two different samples. The results will be presented in Chapter 4, 5, 6. Through the analysis, it will become clear which dataset is the most reliable to use as background field.

The grid points of the three data set (GEM, NARR, and ERA-Interim) selected for comparison to observations are shown in Figure 3.2. All distances between a model point and an observed point are less than 30 km (ERA-Interim), and 10 km (NARR and GEM).

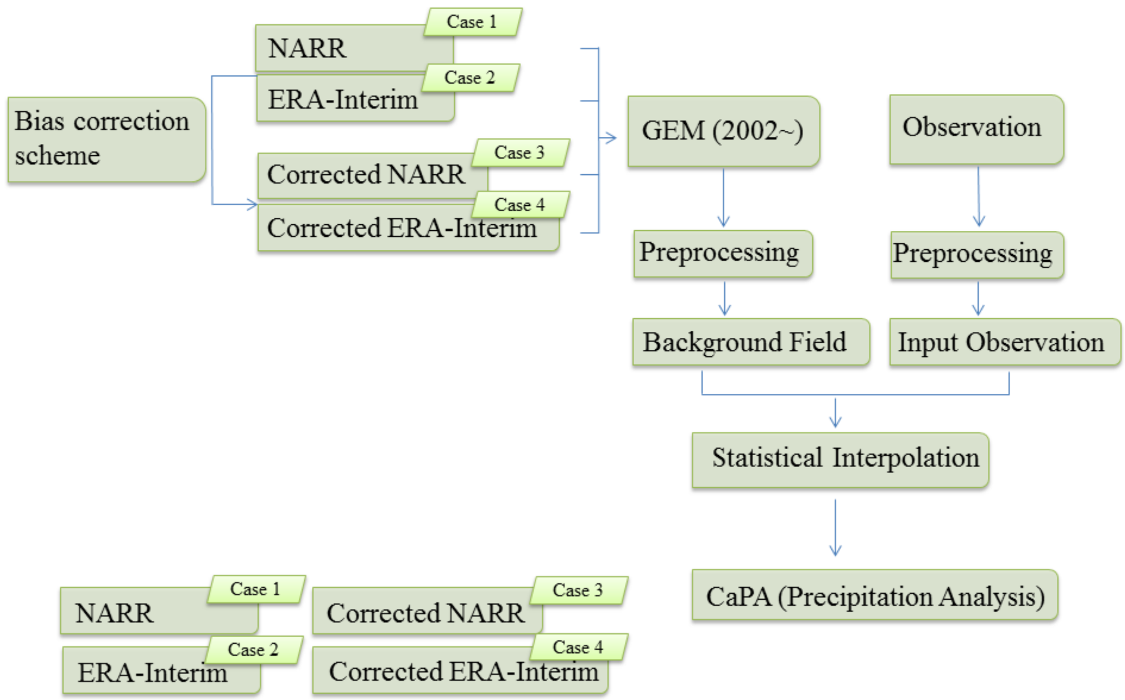


Figure 3.1: Chart of CaPA framework

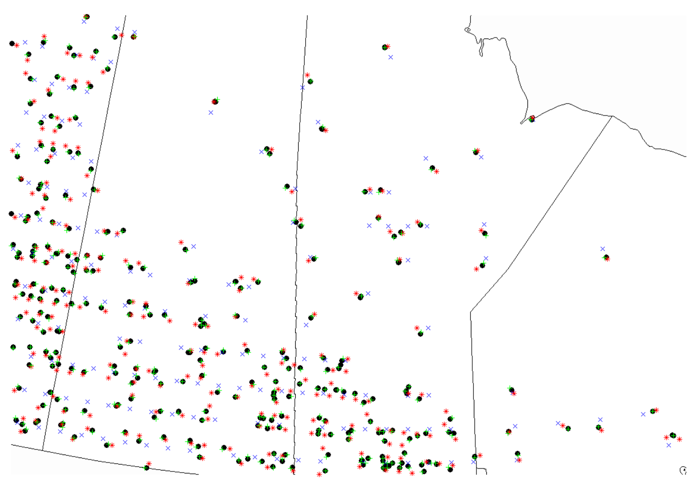


Figure 3.2: Locations of precipitation values. o,x,*,and + represent the location of Observation, ERA-Interim, NARR, and GEM respectively

3.1 Evaluation of reanalysis data sets

The original reanalysis data sets before pre-processing were evaluated in terms of reliability. The bias-corrected reanalysis data sets and CaPA analyses were then evaluated. The process is implemented at different regional and seasonal scales. Seasonal validation is made over four season: Spring (March, April, and May), Summer (June, July, and August), Fall (September, October, and November), and Winter (December, January, and February). The four regions considered for this study (Western, Prairie, Central, and Eastern) were described in Chapter 2 (see Figure 2.1).

3.1.1 Data preparation

The traditional CaPA background, GEM, was provided by Environment Canada. The 24-hr GEM dataset was used to allow for comparison with daily observations. The original format of GEM dataset is binary ('.fst') which had to be converted to ascii format ('.r2c'). The Pegasus server is exclusively used in this study for the extraction of information from the binary files and application of the CaPA framework. One example of the Pegasus is shown in Figure 3.3. The details of the process to execute the CaPA framework is explained in Section 3.3.

NARR and ERA-Interim datasets were downloaded in NetCDF format from NOAA and ECMWF websites, respectively. The NetCDF file consists of three components: latitude, longitude, and variable. The location of each grid is defined by

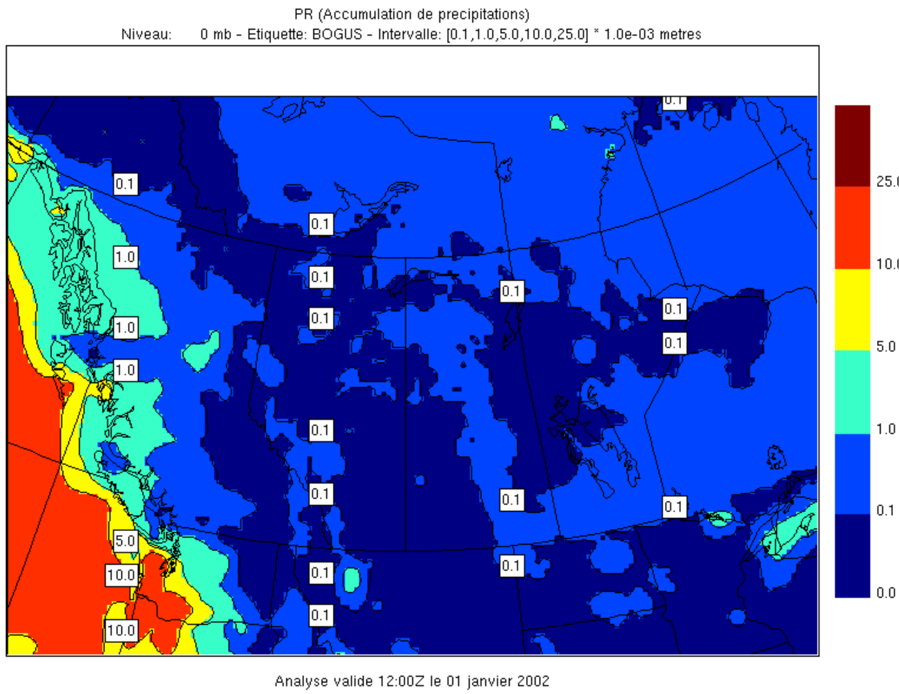


Figure 3.3: Visualization of NARR precipitation in 2002 January 1st using Pegasus software

latitude and longitude. Each grid contains various variables such as temperature, precipitation, humidity and pressure at different vertical levels. The variable 'precipitation on the surface' is the only one of interest in this study. MATLAB is used to extract information from the NetCDF files and format them into 'r2c'-files.

The NARR data set is provided in Lambert Conical projection system and must be converted to a regular latitude-longitude grid as required by the CaPA framework. Bi-linear interpolation is used for the regridding process. Therefore, the evaluation of CaPA analyses of the precipitation products was all based on the Gaussian grid system.

Table 3.1: Contingency Table

	Observed(YES)	Observed(NO)
Forecast(YES)	Hits	False Alarms
Forecast(NO)	Misses	Correct Negatives

3.1.2 Verification of Forecasting Skills: Categorical Scores

Categorical scores are widely used to estimate forecasting skills. This verification method is also used for the CaPA project, see Mahfouf et al. (2007) and the detailed description in Mason (2003). There are several useful skill scores for categorical data: Equitable Threat Score (ETS), Frequency Bias Index (FBI), Probability of Detection (POD), False Alarm Rate (FAR), and Yule's Q (YQ). ETS and FBI are mostly used to verify forecast skill. Each of skill scores are calculated by the number of events falling in each component as defined in the contingency table in Table 3.1. ETS is traditionally defined as follows:

$$ETS = \frac{H - R(H)}{H + F + M - R(H)} \tag{3.1}$$

where

$$R(H) = \frac{(H + F)(H + M)}{H + F + M + C} \tag{3.2}$$

with denotations of the correct forecast area (Hits) H, false forecast area (False Alarms) F, forecast missing area (Misses) M, and correct missing area (Correct Negatives) C following Table 3.1. H is related to the number of occurrences in a forecast (H+F) and the number of real occurrences (H+M). H is zero when there is no forecast

(H+F) and increases as the number of forecast increases but remains less than the number of observation (H+M). ETS represents correct forecasting skill when events actually occur. To detect the number of occurrence in Table 3.1, different thresholds are set up, precipitation amount in this study. There are two types to use the thresholds 1) detect the number of forecast falling between lower and upper thresholds (each bin) of observed precipitation 2) detect the number of forecast when it is larger than each threshold of observed precipitation. This study chose second type for both of ETS and FBI. ETS indicates a value of one for a perfect forecast as both of F and M are zero in a perfect forecast, a value of zero for a random forecast. If ETS is less than zero, the forecast is worse than random. Mesinger (2008) addresses possibility of high threat scores by increasing bias above unity or/and the wetter forecast. To reduce the problems Mesinger (2008) introduces the bias adjusted value of H shown below:

$$H_a = O - \frac{F - H}{\ln(\frac{O}{O-H})} lambertw(\frac{O}{F - H} \ln(\frac{O}{O - H})) \tag{3.3}$$

where lambertw is the inverse function of

$$z = \omega e^{\omega} \tag{3.4}$$

Therefore,

$$lambertw(z) = \omega \tag{3.5}$$

with denotations of the forecast area F, correct forecast area (Hits) H, and the observed area O. Using Table 3.1, the forecast area F and the observed area O are equal to H (Hits) + F (False Alarms) and H (Hits) + M (Misses), respectively. As a result, it is converted as follows:

$$H_a = H + M - \frac{F}{\ln\left(\frac{H+M}{M}\right)} lambertw\left(\frac{H+M}{F}\ln\left(\frac{H+M}{M}\right)\right) \tag{3.6}$$

The study initially considered both ETS using H and ETS using H_a, but no significant difference between the two ETS measures. Therefore, ETS using H is adopted for the verification of reanalysis data sets in this study.

The Frequency Bias Index (FBI) is defined as:

$$FBI = \frac{H+F}{H+M} \tag{3.7}$$

FBI is the ratio of the number of occurrences in a forecast to the number of real occurrences. In a perfect forecasting systems, H is equal to 1, and F and M are equal to 0, which results in FBI=1. A value of FBI above 1 indicates that events are more frequently forecasted than observed. A value of FBI under 1 indicates that a given event magnitudes is forecasted less frequently than actually observed. In this study, FBI-1 is used, which indicates 0 for the perfect forecasting as F and M indicate zero, above 0 for overestimation, and under 0 for underestimation.

3.2 Application of bias correction

Bias correction is necessary to minimize systematic errors from observations and models. This study uses the simple multiplicative shift scheme and the thin plate spline scheme.

3.2.1 Simple multiplicative shift scheme

Due to the large temporal and spatial coverage, the data sets in the study are fairly large. Here, the simple multiplicative shift is chosen for bias correction. The method was introduced by Ines and Hansen (2006) and is used to correct the bias of NARR and ERA-Interim. Ines and Hansen (2006) applied several bias correction schemes to correct General circulation models (GCM) which tends to simulate too many events of low intensity rain. Ines and Hansen (2006) indicates that a simple multiplication shift method performs as well as other methods to correct monthly and seasonal total rainfall. This method only corrects rainfall intensity, so any systematic error in frequency is not corrected (Ines and Hansen, 2006). The simple multiplication shift method uses the ratio of mean monthly precipitation of observation and simulation to correct observed daily precipitation:

$$x_i' = x_i \frac{\bar{X}_{obs}}{\bar{X}_{sim}} \tag{3.8}$$

where x_i and x'_i refer to raw and corrected reanalysis rainfall on day i, and $\bar{X}_{obs}$ and $\bar{X}_{sim}$ are long-term monthly mean rainfall from the reanalysis data and observations for a given month. To obtain observed mean monthly precipitation of observations, the Canadian Daily Climate Data (CDCD) is used. Gridded monthly observation data sets can be an other option as discussed in Chapter 7. However, this study does not use data sets already processed due to the inherent uncertainties in such data. The bias correction ratio $\frac{\bar{X}_{obs}}{\bar{X}_{sim}}$ for NARR and ERA-Interim is presented in Appendix D.

3.2.2 Thin plate spline scheme

Interpolation methods are used to provide spatially continuous information from discrete information. In hydrological studies, interpolation methods are generally used to distribute observed rainfall information to fill in missing data over a given area. Rainfall data are generally available as geographical point information. There are several interpolation methods available for the distribution of point information in space such as IDW (Inverse distance weight), kriging, splines, etc. In this study, interpolation is applied to spatially distribute the correction factors of the simple multiplicative shift method, which can be calculated only at station locations, over all the grid points of the reanalysis data sets.

The Thin plate spline method is one of several useful tools to interpolate and smooth a given dataset. The Thin plate spline method is a spline-based technique, applying a penalty involving the smoothness of the fitted surface. It comes from the idea of bending of a thin sheet of metal. The thin metal sheet has smoothness but also rigidity controlling the smoothness. The following equation shows simple form of energy function:

$$E_{tps}(f) = \sum_{i=1}^{n}(v_i - f(x_i, y_i))^2 \tag{3.9}$$

where v_i denotes the target function values at locations (x_i,y_i) in the plain. A mapping function $f(x_i, y_i)$ minimizes the energy function gives a sets of control points, n, with measure of goodness of fit where deflection occurs in z direction based on x and y coordinates. The energy function considering smoothness is described as below:

$$E_{tps,smooth}(f) = \sum_{i=1}^{n}(v_i - f(x_i, y_i))^2 + \lambda \int\int[(\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2]dxdy \tag{3.10}$$

The tuning parameter λ controls and balances the deformation versus the goodness-of-fit. It plays a role to reduce noise in the specific values v_i by relaxing exact interpolation and controlling the amount of smoothing. Limiting λ to zero causes the exact interpolation. More detail can be found in Donato and Belongie (2002).

In this study, correction factors are distributed and filled into each grid of the reanalysis data sets using the thin plate spline method as described in Figure 3.4. It allows each gridded reanalysis data sets to be bias-corrected in a smooth manner. Thin plate interpolation provides more advanced results of interpolation than other interpolation methods.

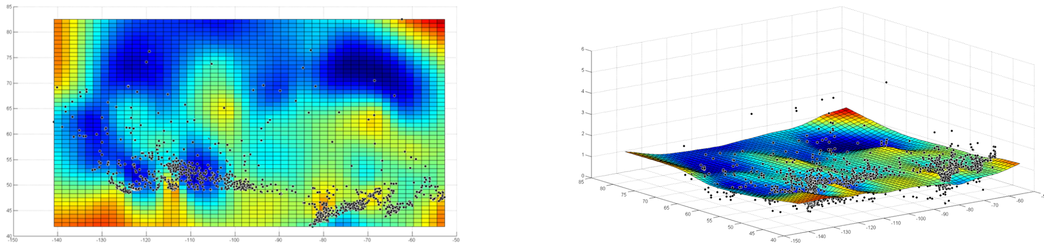


Figure 3.4: Application of Thin Plate Interpolation to bias correction factors: visualized in 2-D (left) and 3-D (right)

3.3 Optimal Interpolation process in the CaPA framework

A Statistical Interpolation (SI) technique (also known as Optimum Interpolation) is a core component of the CaPA framework. It is used for spatial interpolation (Mahfouf et al., 2007). SI uses innovations (differences between an observation and the corresponding background value), and error statistics to perform the analysis. The equation of the SI technique is shown below (Mahfouf et al., 2007; Daley, 1993).

$$x_a^j = x_b^j + \sum_{i=1}^{N} W^{ij}(y_o^i - y_b^i) \tag{3.11}$$

An analysis value, x_a^j, is updated from x_b^j by the statistical interpolation. x_b^j represents the initial guess which is a background field value at model grid point j. SI performs the analysis on innovations using y_o^i and y_b^i which represent neighboring observations and the spatially interpolated background field to the observation points i, respectively. W^{ij} are weights calculated by minimizing the variance of the analysis

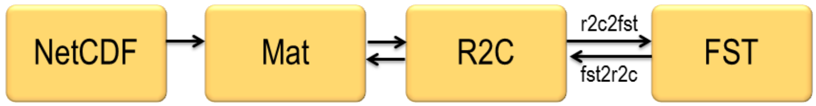


Figure 3.5: Data conversion procedure

error given a set of N. The equation can be simplified as:

$$\delta_x^j = \sum_{i=1}^{N} W^{ij} \delta_y^j \tag{3.12}$$

where δ_x^j (analysis increment) states the correction at a given grid point which is a linear combination of the innovations, δ_y^j. The SI technique is most commonly used for meteorological analysis. This technique is similar method to Kriging, but requires prior information to perform the analysis on innovations. It provides an estimate of the analysis error (Lorenc, 1986). In a NWP system, prior information is generally obtained from continuous model background and combined with discrete observations. Therefore, the NWP ensures that the final analysis is available at each grid point in the model domain.

As described previously, the purpose of this study is to replace the CaPA background, GEM, by NARR and ERA-Interim and to determine their usefulness for the purpose of historical development of CaPA analysis. Original and corrected NARR and ERA-Interim are plugged into the CaPA framework through the Pegasus software and compared with the traditional CaPA analysis that uses GEM. To execute the CaPA framework process using the Pegasus software, several procedures need to be followed. Data conversion process is necessary, as briefly described in Figure 3.5. All of the CaPA procedures work on files in binary format, '.fst'. Since the only conversion tools installed in the Pegasus are 'fst2r2c' and/or 'r2c2fst', it is necessary to prepare input files in ASCII format, 'r2c', and convert these to '.fst' files. Matlab is used to create '.r2c' files. Matlab is used not only for reading, converting, and formatting data sets but also to analyze the data sets. In Matlab, the format, '.mat', is mostly used which requires additional conversion steps. Most of reanalysis data sets including NARR and ERA-Interim are formatted in 'NetCDF' and are also processed in Matlab.

Chapter 4

Evaluation/Assessment of reanalysis data sets

Gridded precipitation data sets are useful due to their well-distributed spatial coverage which can enhance many studies. There are two general approaches to generate gridded precipitation: (1) Interpolation of observed data and (2) Simulation of meteorological processes using atmospheric models. The first approach uses station information only and applies quality control (QC) and various interpolation techniques. However, when there is a lack of observations and the existing observations are further apart, the first approach is not particularly useful. There is a gridded precipitation data set over Canada that uses the Australian National University Spline (ANUSPLIN) model which is based on the thin-plate spline interpolation algorithms developed by Hutchinson (1995). ANUSPLIN was initially developed for forest applications but is now also used for many studies requiring continuous spatial precipitation information and has been applied successfully in other fields (Hopkinson et al., 2012; Benyahya et al., 2014; Murdock et al., 2013). However, as mentioned earlier, the data is only available below latitude 60N due to significant missing information in Northern Canada.

Even when sophisticated interpolation methods and techniques are used to generate gridded information, there are always some challenges in remote areas where large gaps exist in the network of stations. This can significantly limit the usefulness of data obtained using the first approach. The second approach, on the other hand, uses numerical weather models, which do not rely on observed data, and provides information where observations are not available. Therefore, the second approach may overcome part of the aforementioned limitations imposed by sparse observation networks. Data obtained using the second approach are often called reanalysis data. Reanalysis data uses a numerical forecast model as background and assimilates available observations to generate gridded output. The details are explained in Chapter 2. As a result, the reanalysis data fields are often more seamless and smoother than that of interpolated fields (Bosilovich et al., 2008).

Precipitation is an essential component of the energy balance in the water cycles and the dynamical circulation in climate models (Bosilovich et al., 2008). Despite its biases introduced by various sources, station data are often used as a reference and considered as the ground truth (Keenan et al., 1988; Sugita and Brutsaert, 1993).

The Canadian Daily Climate Data (CDCD) set is used as a reference in this study to assess the two selected reanalysis data sets (ERA-Interim and NARR). The CDCD data set has many stations with long periods of record and has relatively good spatial coverage. Precipitation over Canada varies widely in time and space. Precipitation from the reanalysis data sets will be compared to CDCD precipitation at different spatial and temporal scales in this chapter. Since this thesis aims to use reanalysis data sets as alternatives to GEM, reanalysis data sets will also be compared to GEM.

4.1 Results and discussions

Figure 4.1 shows Equitable Threat Score (ETS) scores of GEM, NARR, and ERA-Interim over all regions and seasons (2002-2011). ETS is used to test the forecasting ability of the model associated with different thresholds of precipitation. As explained in Section 3.1, the higher the ETS value, the better forecasting skill of the model. GEM is superior in terms of ETS score across all precipitation thresholds. ERA-Interim shows slightly better performance than NARR in skill. Regional and seasonal ETS are discussed later in this chapter.

Figure 4.2 shows the Frequency Bias Index, FBI-1, of GEM, NARR, and ERA-Interim over all seasons and regions. GEM is also superior to the other data sets in terms of the FBI-1 index. The value of FBI-1 in GEM is close to 0 at all thresholds while for ERA-Interim and NARR, it varies depending on the threshold. It means GEM has the least bias comparing to ERA-Interim and NARR. ERA-Interim and

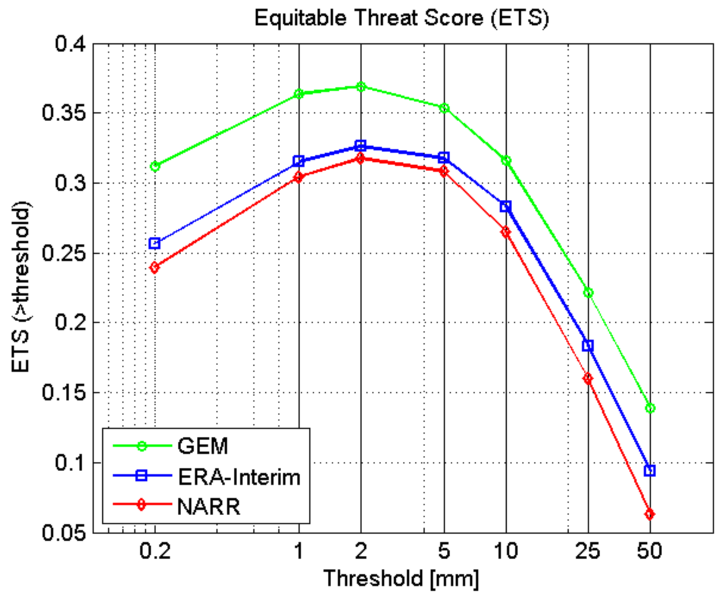


Figure 4.1: Equitable Threat Score (ETS) results of reanalyses over all seasons and all regions in this study

NARR show similar tendencies in FBI-1 at all thresholds. In terms of the FBI-1, NARR shows less bias than ERA-Interim up to 5mm threshold while the opposite is true when thresholds are more than 10mm. At relatively extreme thresholds (25mm or more), NARR and ERA-Interim show more underestimation than GEM. In addition, when focusing on the thresholds between 5mm and 10mm, NARR and ERA-Interim indicate less bias than GEM. ERA-Interim shows higher score in ETS and less bias in FBI-1 at large thresholds which implies ERA-Interim performs better than NARR at extreme precipitation. Therefore, ERA-Interim may be recommended for extreme precipitation.

The overall results of bin type ETS and FBI are presented in Appendix A. The regional and seasonal results of bin type are presented in Appendix B and C, respectively.

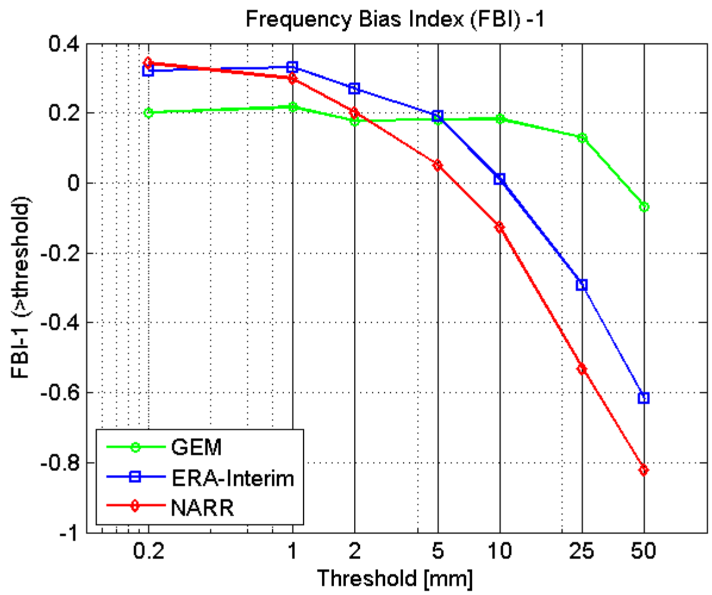
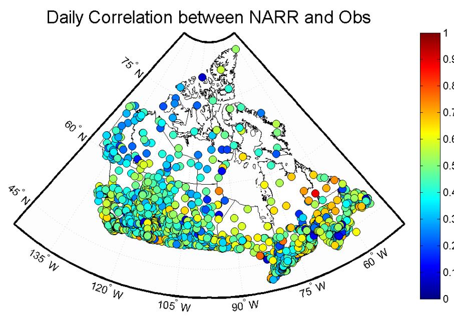


Figure 4.2: Frequency Bias Index (FBI) results of reanalyses over all seasons and all regions in this study

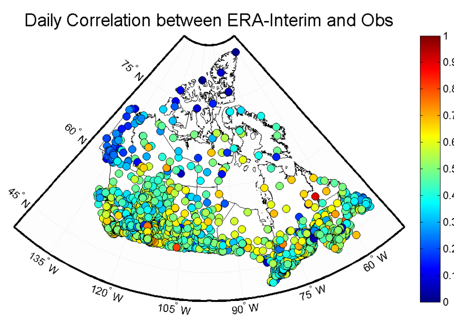
4.1.1 Regional Evaluation

Analysis of daily precipitation from 2002 to 2011 was performed overall regions as shown in Table 4.1. GEM shows smaller RMSE and higher correlation to observations than NARR and ERA-Interim. NARR and ERA-Interim show similar values in RMSE and correlation, although ERA-Interim indicates slightly better correlation than NARR. On the other hand, NARR has less RMSE than ERA-Interim. The regional daily correlation is visualized in Figure 4.3, which helps to see the distribution of correlation values in space. Although RMSE and correlations for NARR and ERA-Interim are less good than for GEM, they still appear to be reasonable alternative backgrounds for use in the CaPA framework.

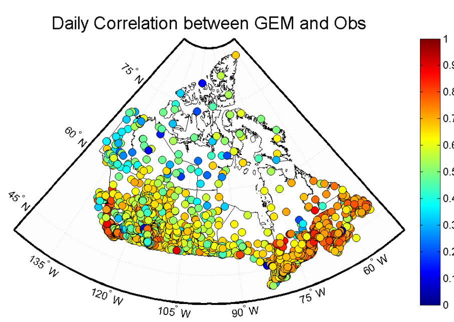
Q-Q plot of GEM and reanalyses against observations over regions are shown



(a)



(b)



(c)

Figure 4.3: Daily correlation between reanalyses and CDCD over Canada: (a) NARR (b) ERA-Interim (c) GEM

Table 4.1: Comparison of daily precipitation Average[mm], RMSE[mm], and Correlation(R) over regions

Region	Obs	GEM			NARR			ERA-Interim		
		Avg	RMSE	R	Avg	RMSE	R	Avg	RMSE	R
Western region	3.9	5.5	8.7	0.69	4.0	8.7	0.51	4.6	9.0	0.53
Prairie region	1.3	1.7	3.8	0.62	1.6	4.1	0.46	1.5	4.1	0.49
Central region	2.7	3.1	5.0	0.69	2.2	5.7	0.48	2.7	6.0	0.48
Atlantic region	3.3	3.8	5.7	0.72	2.1	7.1	0.45	3.1	7.1	0.52
Average	2.8	3.5	5.8	0.68	2.5	6.4	0.48	3.0	6.6	0.50

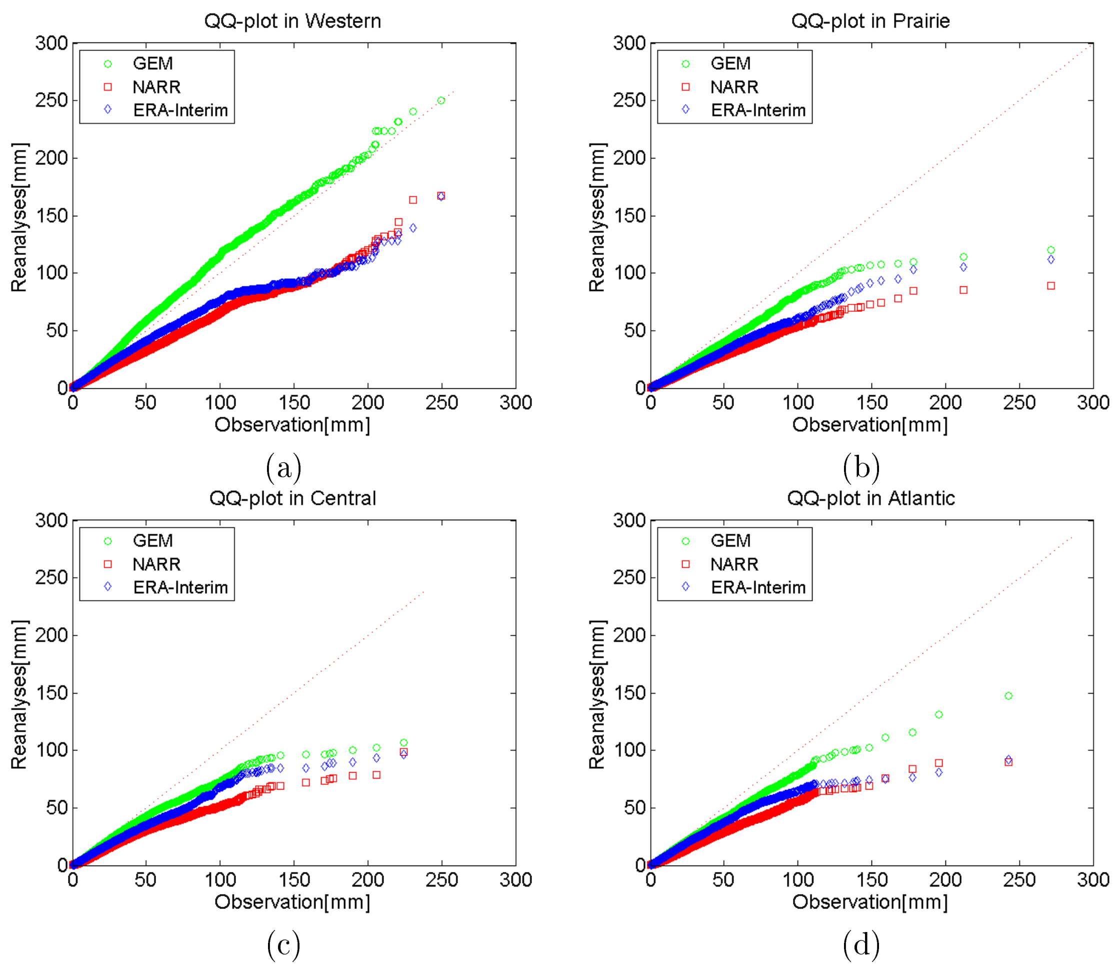


Figure 4.4: QQ plot of reanalyses against observations over regions: (a) Western (b) Prairie (c) Central (d) Atlantic

in Figure 4.4. GEM is superior among all data sets over all regions, especially in the Western region. Although GEM slightly overestimates in the Western region, its distribution is close to the distribution of observations, even at extreme precipitation. Distributions of NARR and ERA-Interim are similar in the Western and Atlantic regions, although NARR is slightly closer to GEM and observation at extreme precipitation. In the Prairie and Central regions, the distribution of NARR and ERA-Interim become slightly separated as precipitation becomes larger. Here, ERA-Interim is closer to observations than NARR in terms of the distribution of extreme precipitation.

Figure 4.5 shows regional boxplot analysis of differences between the reanalysis data sets and observations. The previously shown Q-Q plots do not take timing into account. It only compares distribution of two data sets. The boxplot analysis, on the other hand, looks at the differences between reanalyses data sets and observations on the same day. Using this presentation, it will be easy to compare the consistency of GEM, NARR and ERA-Interim with observations.

The differences between GEM and observations cover the widest range in the Western region. Except for the Central region, the distribution of GEM-Observation has a wider range than ERA-Interim-Observation and NARR-Observation. Among ERA-Interim and NARR, NARR-Observation has a wider range than ERA-Interim-Observation in the Western and Central regions, while ERA-Interim-Observation has a wider range than NARR-Observation in the Atlantic region. Overall, the GEM-

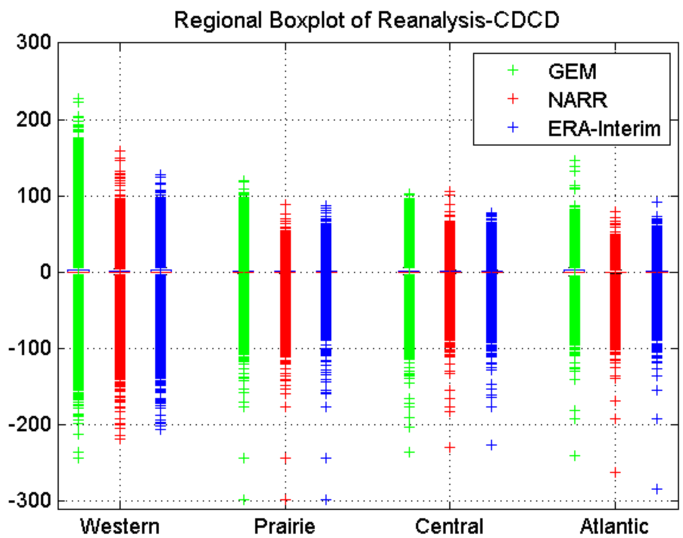


Figure 4.5: Boxplot of differences between reanalyses and CDCD over regions

Observation seems denser above the zero line than ERA-Interim-Observation and NARR-Observation. The result might be expected, since as explained previously, NWP tends to generate more frequent precipitation than observations. NARR and ERA-Interim are reanalysis products which means that they are already adjusted through assimilation of observations.

Figure 4.6 shows ETS of forecasts over regions. As expected, the ETS skill score of GEM is superior over all regions. ERA-Interim and NARR show similar performance in the Western, Prairie, and Central regions, while ERA-Interim shows noticeably higher scores than NARR and is closer to GEM in the Atlantic region. When the precipitation threshold is less than 2mm in the Western and Prairie regions, ERA-Interim is slightly more skillful than NARR. In the Central region, ERA-Interim shows higher ETS score than NARR when precipitation is more than 5mm, especially

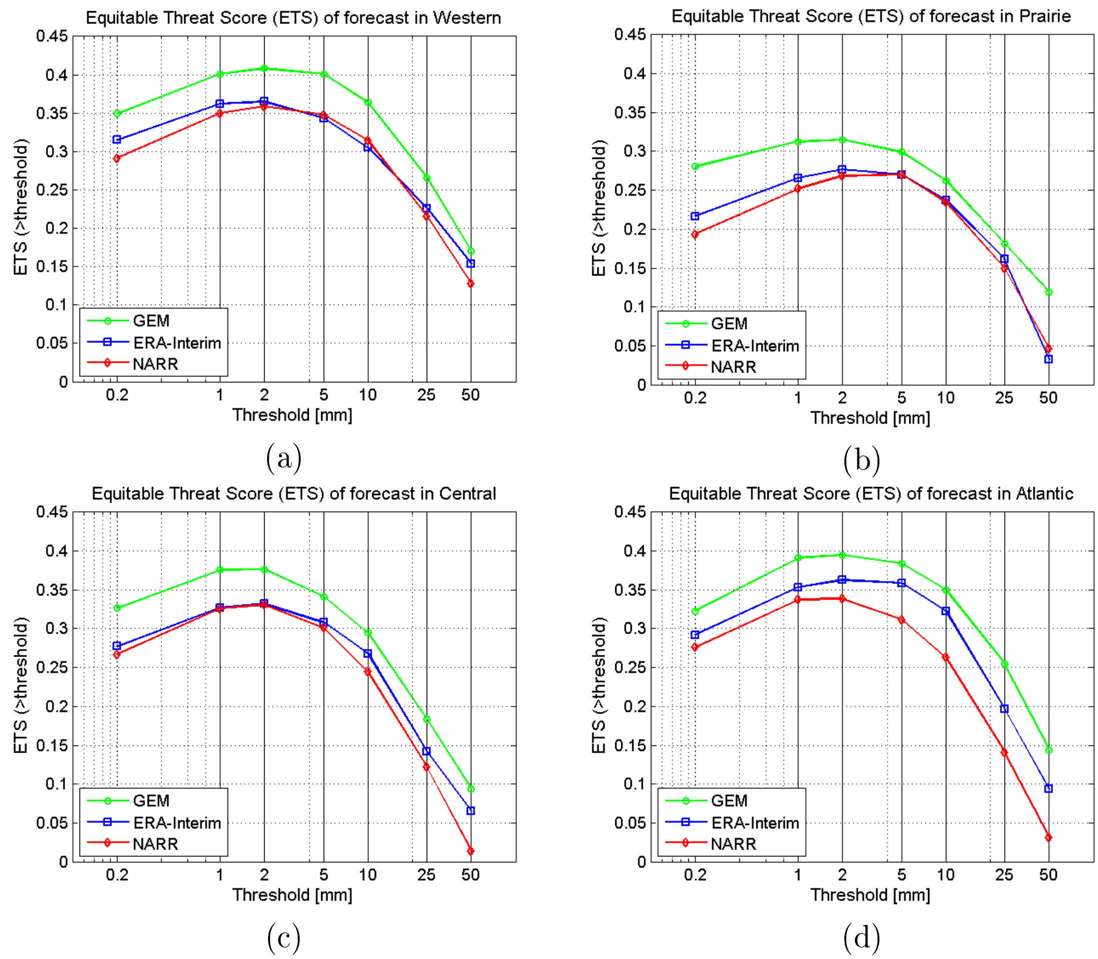


Figure 4.6: Equitable Threat Score (ETS) of forecasts over Canadian regions: (a) Western (b) Prairie (c) Central (d) Atlantic

at extreme precipitation (50mm). In the Atlantic region, the difference between ERA-Interim and NARR is small when the threshold is less than 1mm but becomes bigger at thresholds between 2 and 10mm.

Figure 4.7 shows the bias (FBI-1) of forecasts over regions. GEM is closer to zero than ERA-Interim and NARR over all thresholds in most of the regions. The biases of GEM, ERA-Interim, and NARR are close at thresholds up to 2mm in the Western, Central, and Atlantic regions. The biases of ERA-Interim and NARR are

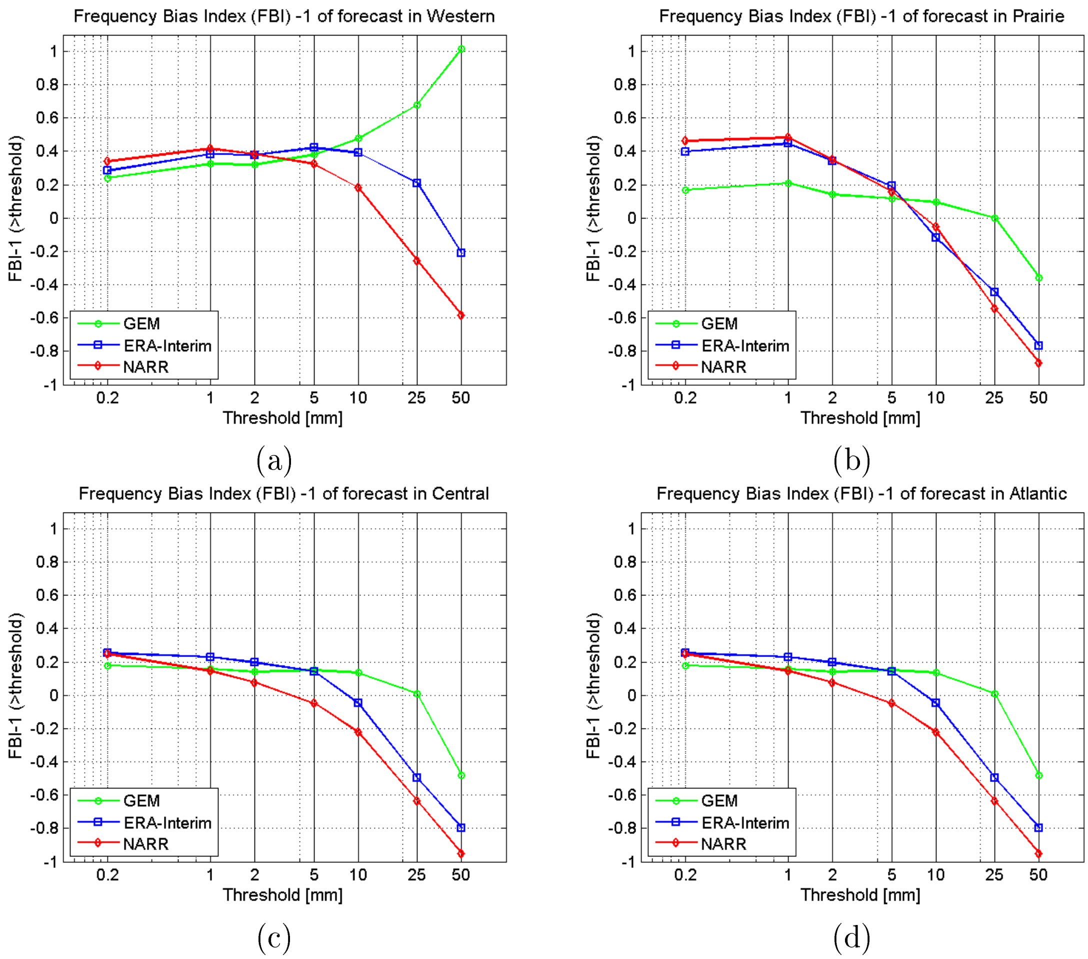


Figure 4.7: Frequency Bias Index (FBI) of forecasts over regions: (a) Western (b) Prairie (c) Central (d) Atlantic

close to each other over all thresholds in the Prairie region, but both overestimate more than GEM when thresholds are less than 5mm, and underestimate more than GEM when thresholds are over 5mm. The tendency of ERA-Interim and NARR to underestimate more than GEM at more than 5mm thresholds shows in the Western, Central and Atlantic regions as well. However, ERA-Interim seems to have less bias than NARR in the Central and Atlantic regions, and to be closer to GEM in the Western, Central, and Atlantic regions.

4.1.2 Seasonal Evaluation

Comparison of daily precipitation bias, RMSE, and correlation was performed over seasons as shown in Table 4.2. One of several noticeable points is that the correlation of GEM is exceptional regardless of region and season. GEM also shows the least RMSE over all regions and seasons, except in Spring and Winter of the Western region. Comparing NARR and ERA-Interim in terms of correlation, ERA-Interim shows slightly better correlation than NARR in all seasons in the Prairie and Atlantic regions, NARR shows better correlation in all seasons in Central region. In the Western region, NARR shows better correlation in Summer while ERA-Interim is better correlated in Fall and Winter. In terms of RMSE, NARR and ERA-Interim show very similar performance in all seasons in the Prairie region. RMSE of NARR is less than the RMSE of ERA-Interim in the Western (except Summer season) and Central regions while ERA-Interim shows less RMSE than NARR in the Atlantic region except Winter season.

Each region of boxplot is divided into four seasons as shown in Figure 4.8. As previously mentioned, the range of GEM-Observation in the Western region is wider than in other regions, and it is clear from Figure 4.8 that it is mostly due to Spring, Fall, and Winter seasons in the region. The differences between GEM and observations in Summer in the Western region covers a significantly smaller range compared to other seasons.

In the Atlantic region, overestimation is found more in GEM than in reanalysis datasets. From the seasonal analysis, it is seen that overestimation in GEM is mainly from Winter.

Table 4.2: Comparison of daily precipitation Average[mm], RMSE[mm], and Correlation(R) over seasons

Region	Season	Obs	GEM			NARR			ERA-Interim		
			Avg	RMSE	R	Avg	RMSE	R	Avg	RMSE	R
Western	Spring(MAM)	3.4	5.1	7.7	0.67	3.8	7.2	0.52	4.2	7.5	0.52
	Summer(JJA)	1.7	2.3	4.4	0.68	2.1	5.0	0.47	1.9	5.0	0.44
	Fall(SON)	5.2	7.0	10.3	0.70	4.8	10.7	0.50	6.0	11.2	0.53
	Winter(DJF)	5.4	7.7	11.2	0.67	5.2	10.5	0.50	6.4	11.0	0.51
Prairie	Spring(MAM)	1.2	1.7	3.4	0.64	1.6	3.6	0.50	1.5	3.6	0.53
	Summer(JJA)	2.3	2.5	5.4	0.59	2.3	5.7	0.42	2.4	5.7	0.44
	Fall(SON)	1.1	1.4	3.2	0.65	1.3	3.4	0.49	1.3	3.4	0.56
	Winter(DJF)	0.8	1.0	2.1	0.58	1.1	2.4	0.44	0.9	2.4	0.55
Central	Spring(MAM)	2.5	3.0	4.2	0.73	2.0	4.9	0.52	2.6	5.3	0.50
	Summer(JJA)	3.1	3.4	7.0	0.53	2.7	7.2	0.40	3.2	7.3	0.40
	Fall(SON)	3.0	3.5	4.9	0.76	2.4	6.0	0.54	2.9	6.3	0.53
	Winter(DJF)	2.3	2.7	3.5	0.74	1.9	4.2	0.51	2.1	4.5	0.50
Atlantic	Spring(MAM)	2.8	3.4	4.6	0.76	1.9	6.4	0.42	2.8	6.1	0.53
	Summer(JJA)	3.1	3.4	6.5	0.60	2.0	7.2	0.42	2.8	7.0	0.47
	Fall(SON)	3.9	4.4	6.6	0.75	2.3	8.4	0.50	3.6	8.3	0.56
	Winter(DJF)	3.1	4.0	5.2	0.73	2.3	6.4	0.46	3.2	6.6	0.51
Average		2.8	3.5	5.8	0.68	2.5	6.4	0.48	3.0	5.3	0.50

The seasonal characteristics of models are captured through the seasonal Frequency Bias Index (FBI)-1 as shown in Figures 4.9 - 4.12. The first, second, and third columns represent FBI-1 of GEM, ERA-Interim, and NARR, respectively. Over most regions, FBI-1 of GEM is closer to the zero line than ERA-Interim and NARR over all seasons. Especially at extreme precipitation thresholds ($\geq 25mm$), FBI-1 of NARR and ERA-Interim show relative underestimation compared to GEM.

Over the Western region in Figure 4.9, the overestimation seen in Figure 4.7 of GEM in FBI-1 is also noticeable. From the seasonal analysis, it is found that the FBI-1 overestimation of GEM in Figure 4.7 is mainly from Spring and Winter seasons. As a result, ERA-Interim and NARR performs better than GEM in terms of FBI-1 at extreme precipitation in Spring and Winter, and GEM performs better at extreme precipitation in Summer.

However, in terms of FBI-1, GEM performs better than ERA-Interim and NARR in Spring and Winter over Prairie region. Over the Prairie region, GEM is actually superior among all data sets.

Over the Central region, the FBI-1 of ERA-Interim shows the biggest bias in Summer up to thresholds of 5mm and then dramatically decreases till the threshold of 25mm. In addition, the FBI-1 of ERA-Interim in Spring and Summer indicates the biggest overestimation under thresholds of 10mm, and the biggest underestimation over thresholds of 10mm.

Over the Atlantic region, like in other seasons, the FBI-1 of GEM outperforms

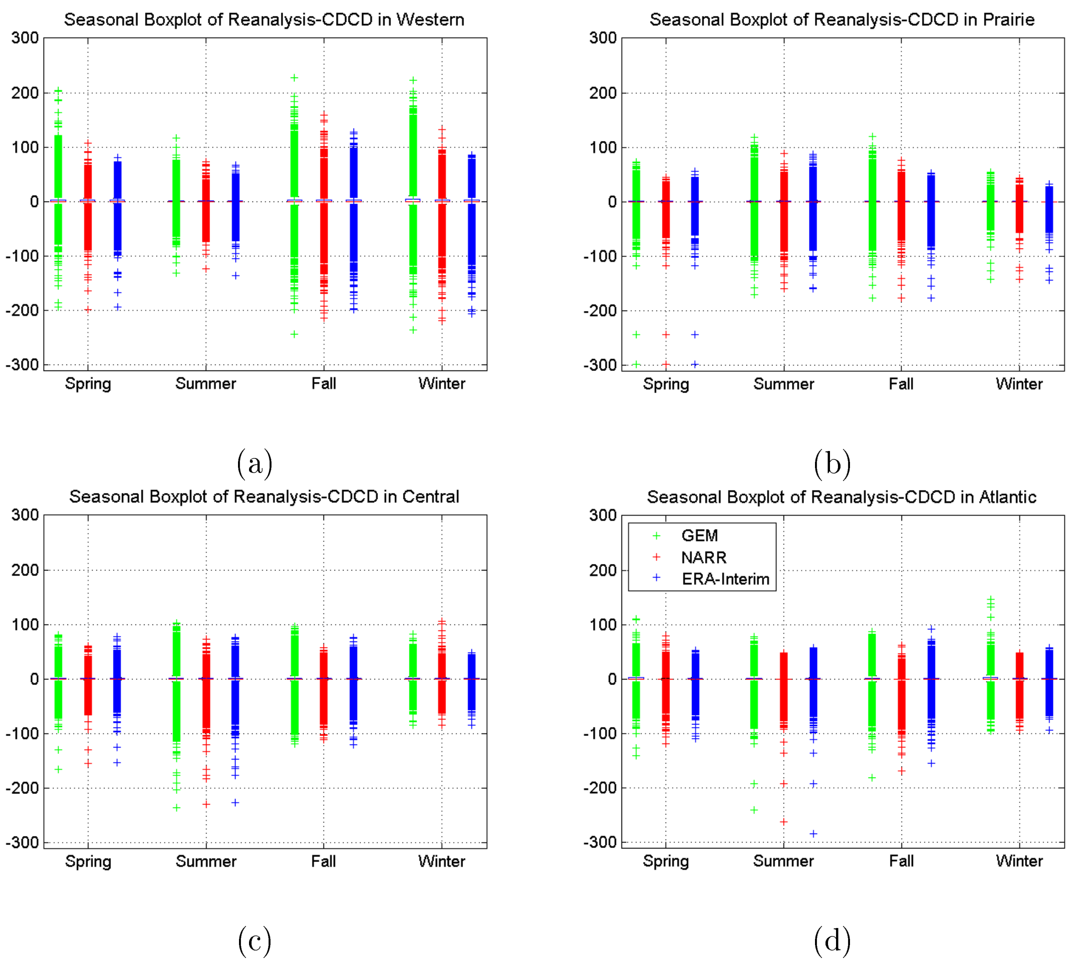
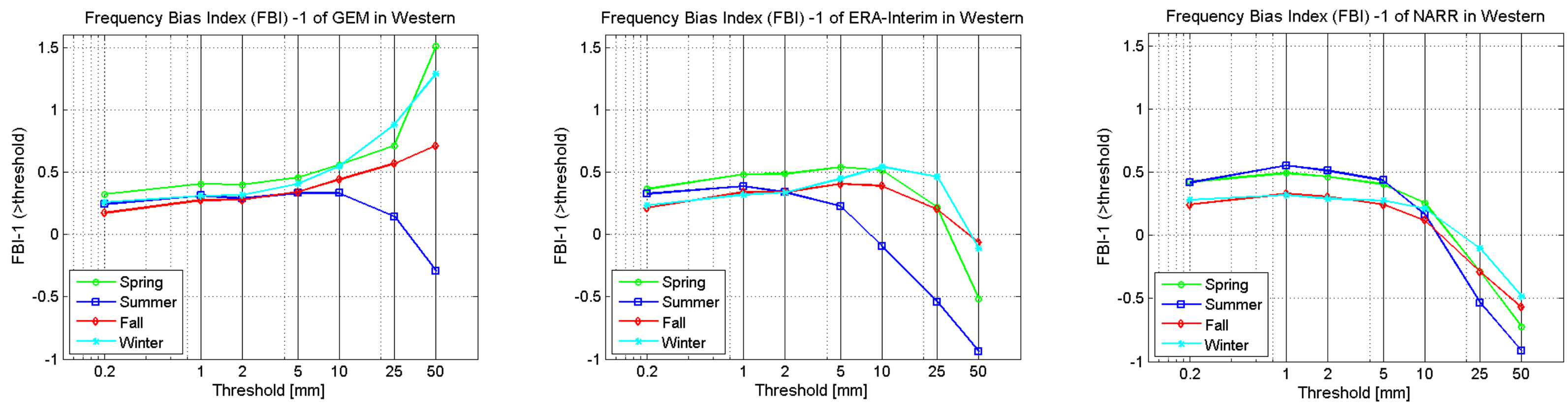


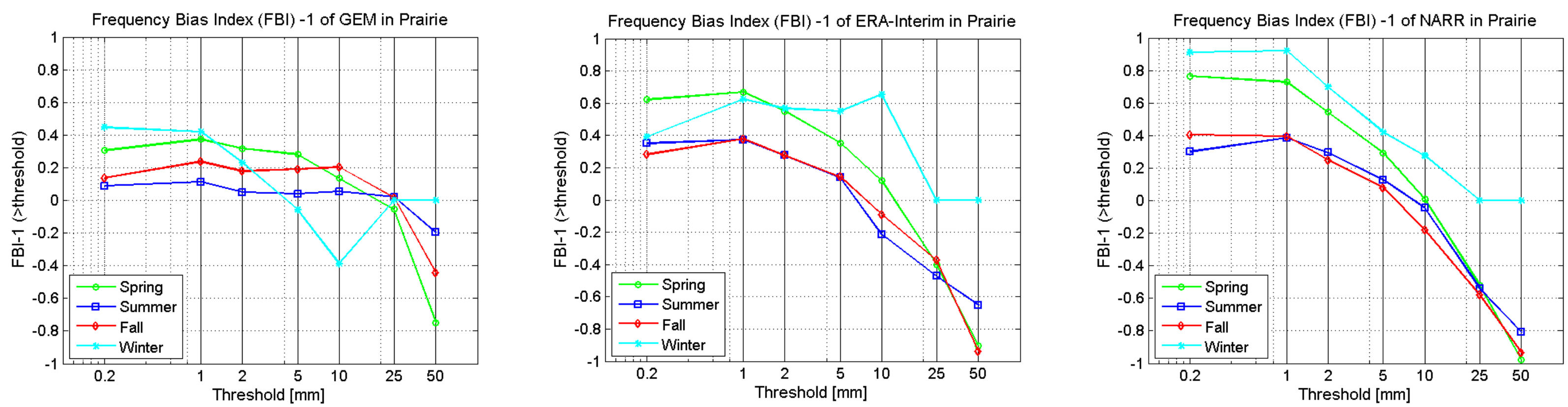
Figure 4.8: Boxplot of differences between reanalyses and CDCD over seasons in each region: (a) Western (b) Prairie (c) Central (d) Atlantic

at extreme precipitation thresholds ($\geq 25mm$), especially in Winter.



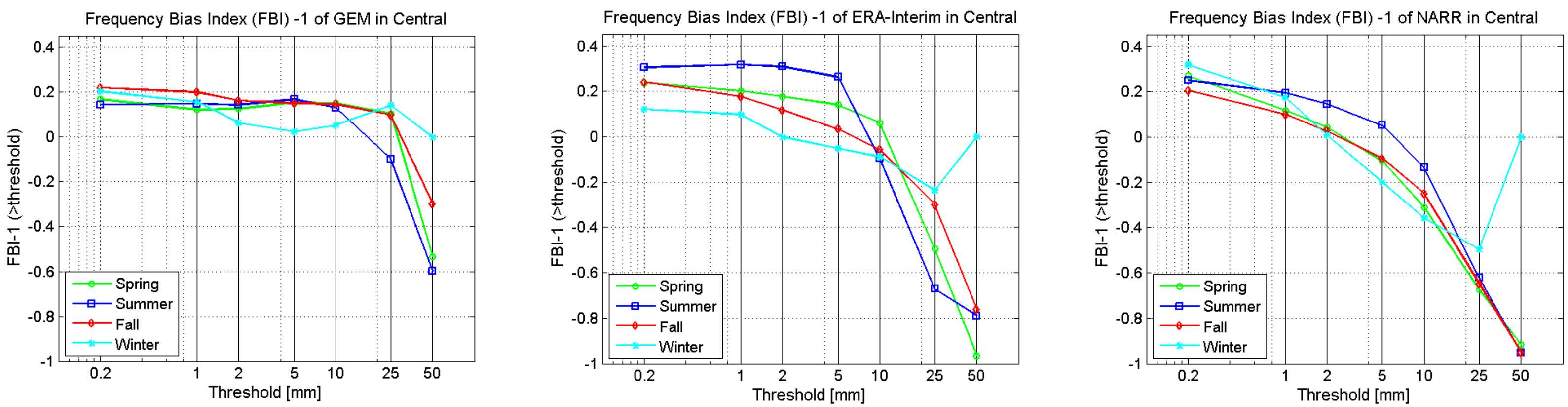
(a) (b) (c)

Figure 4.9: Frequency Bias Index (FBI)-1 of forecasts over seasons in Western region: (a) GEM (b) ERA-Interim (c) NARR



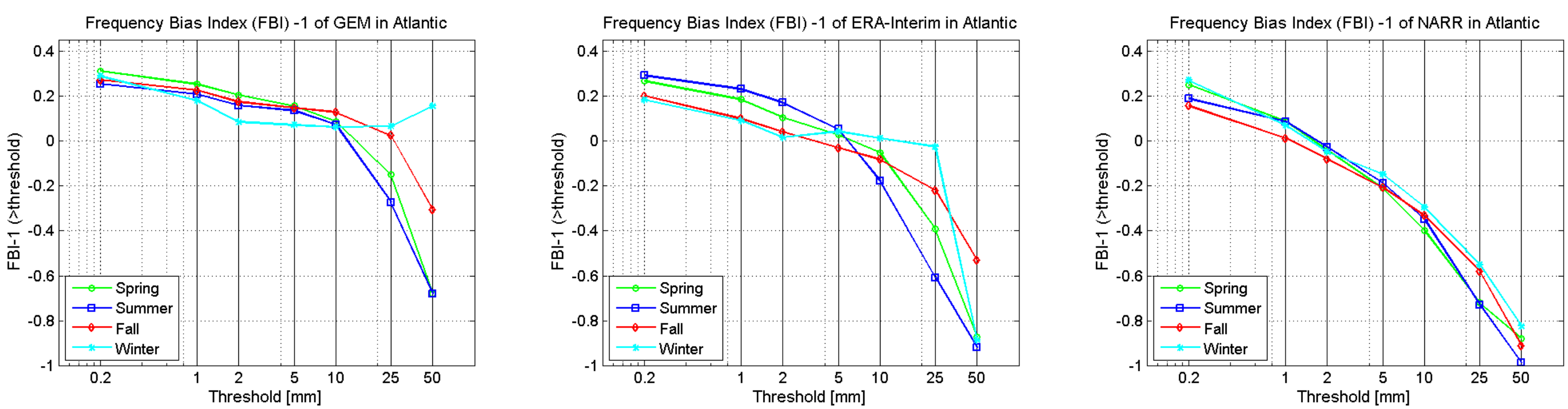
(a) (b) (c)

Figure 4.10: Frequency Bias Index (FBI)-1 of forecasts over seasons in Prairie region: (a) GEM (b) ERA-Interim (c) NARR



(a) (b) (c)

Figure 4.11: Frequency Bias Index (FBI)-1 of forecasts over seasons in Central region: (a) GEM (b) ERA-Interim (c) NARR



(a) (b) (c)

Figure 4.12: Frequency Bias Index (FBI)-1 of forecasts over seasons in Atlantic region: (a) GEM (b) ERA-Interim (c) NARR

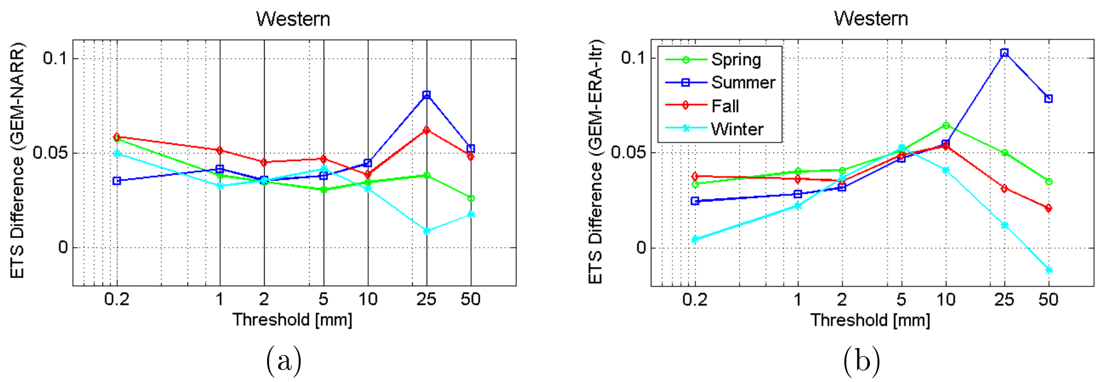
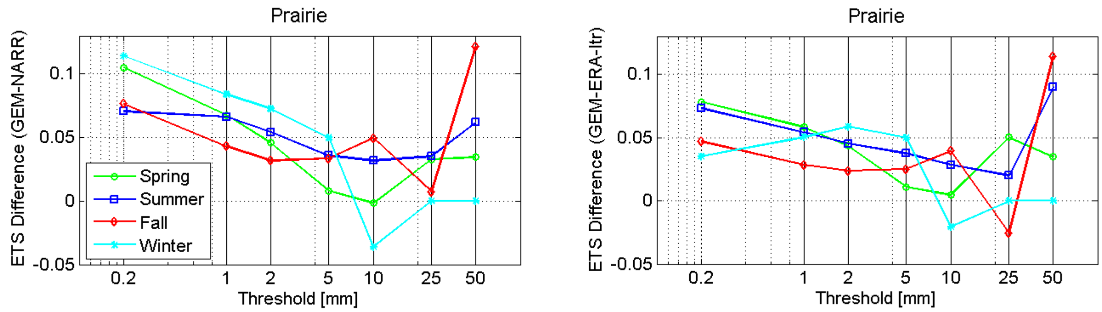


Figure 4.13: Equitable Threat Score (ETS) differences between GEM and reanalysis forecasts in Western region: (a) GEM minus NARR (b) GEM minus ERA-Interim

4.1.3 Differences in skill between GEM and reanalysis data sets

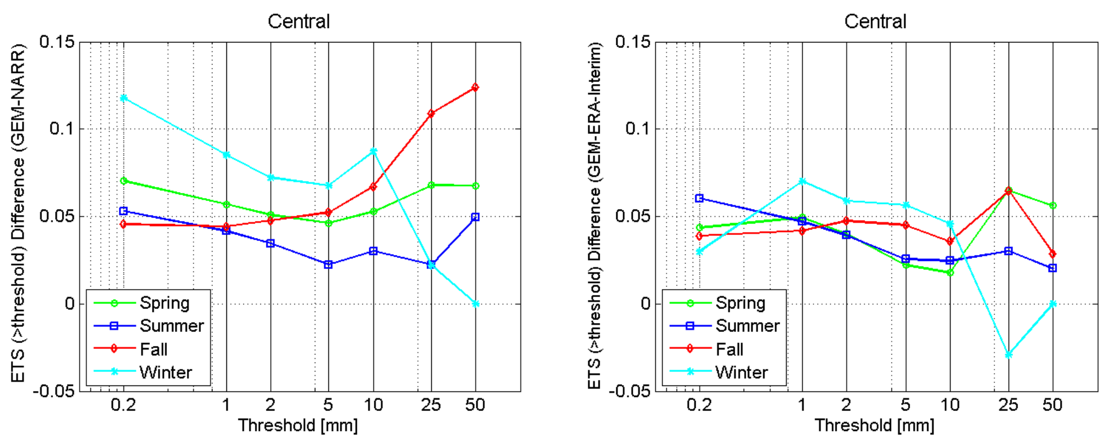
Seasonal ETS differences over each regions between GEM and reanalysis data sets are shown in Figures 4.13 - 4.16. The purpose of this study is to find a model as an alternative to GEM; therefore, the ETS score differences between GEM and reanalysis data sets are presented instead of ETS scores themselves. A difference close to 0 indicates that the reanalysis data sets is similar to GEM, in terms of skill. Left columns represent the difference of ETS between GEM and NARR; right columns represent the difference of ETS between GEM and ERA-Interim.

Over the Western region (Figure 4.13), the biggest difference of GEM-NARR and GEM-ERA-Interim shows in Summer and Winter. The difference of ETS score of GEM-NARR in all seasons seems relatively steady up to threshold 10mm, becomes larger in Summer and smaller in Winter as the threshold reaches 25mm. The difference of ETS of GEM-ERA-Interim shows a spike at threshold 25mm and decreases



(a) (b)

Figure 4.14: Equitable Threat Score (ETS) differences between GEM and reanalysis forecasts in Prairie region: (a) GEM minus NARR (b) GEM minus ERA-Interim



(a) (b)

Figure 4.15: Equitable Threat Score (ETS) differences between GEM and reanalysis forecasts in Central region: (a) GEM minus NARR (b) GEM minus ERA-Interim

a bit at threshold 50mm in Summer. As a result, NARR seems better in Summer for high thresholds relative to GEM over the Western region.

Over the Prairie region (Figure 4.14), the ETS score differences of GEM-NARR and GEM-ERA-Interim are similar, except in Winter. In Winter, the ETS difference of GEM-NARR is larger than GEM-ERA-Interim at threshold of 0.2mm, and gradually decreases and becomes similar to GEM-ERA-Interim at a threshold of 5mm.

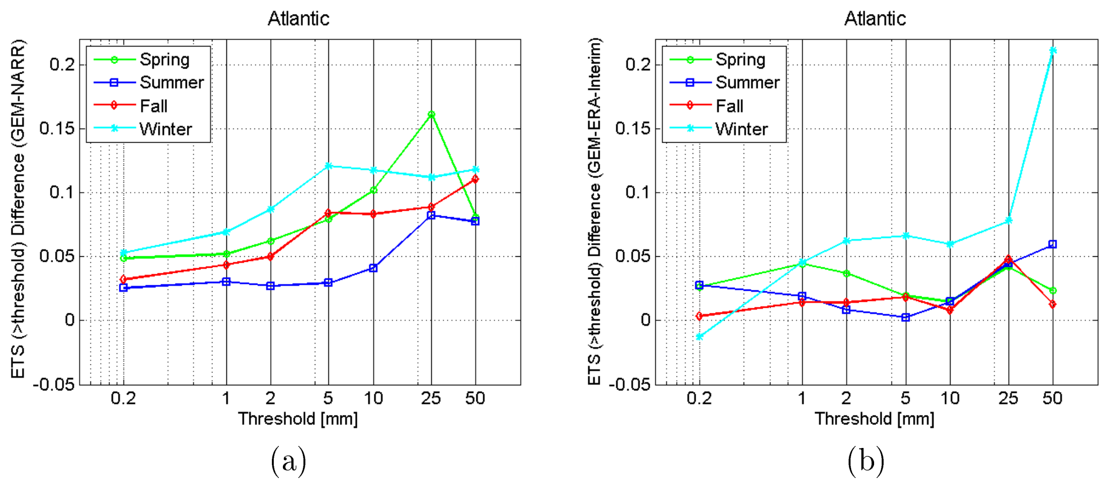


Figure 4.16: Equitable Threat Score (ETS) differences between GEM and reanalysis forecasts in Atlantic region: (a) GEM minus NARR (b) GEM minus ERA-Interim

Over the Central region (Figure 4.15), the ETS score difference of GEM-ERA-Interim is closer to zero than the score difference of GEM-NARR over all seasons. Particularly in Fall and Winter, the ETS differences of GEM-NARR and GEM-ERA-Interim seem noticeably different compared to other seasons. While the ETS score difference of GEM-NARR increases with threshold in Fall, the ETS score difference of GEM-ERA-Interim does not increase as much. In Winter, the ETS difference of GEM-NARR is significantly larger than GEM-ERA-Interim at a threshold of 0.2mm, and decreases gradually but remains larger than GEM-ERA-Interim up to 10mm. Overall, the forecasting ability of ERA-Interim is closer to GEM in all seasons in the Central region.

Over the Atlantic region (Figure 4.16), the ETS difference of GEM-ERA-Interim is closer to zero than GEM-NARR overall except in Winter. In Winter, the ETS difference of GEM-ERA-Interim is less than GEM-NARR up to a threshold of 25mm, but there is a spike at 50mm in GEM-ERA-Interim which is larger than GEM-NARR.

4.2 Summary

In this chapter, a pre-assessment of reanalysis data were conducted. As expected, GEM is overall superior among the candidate reanalysis data sets, but there are slight differences depending on each region and season.

- ERA-Interim is slightly better than NARR in terms of correlation with stations over all regions.

- In the regional analysis of QQ-plots in Western region, GEM shows a slight overestimation but has distributions most similar to the distribution of observations.

- In the regional analysis of QQ-plot, ERA-Interim is closer to observations than NARR in the Prairie and Central regions.

- In the regional analysis of ETS, ERA-Interim and NARR show fairly similar skill but ERA-Interim has higher skill than NARR in the Atlantic region.

- In analysis of FBI-1, GEM has the least bias over all regions. However, GEM shows a spike of overestimation at extreme thresholds in the Western region. Therefore, reanalysis datasets are better than GEM in terms of FBI-1 of the Western region at high thresholds.

- In the seasonal analysis of FBI-1 in the Prairie region, both ERA-Interim and NARR are far from GEM in Spring and Winter.

- In the seasonal analysis of FBI-1 in Central region, ERA-Interim is closer to GEM than NARR in Winter.

- In the analysis of FBI-1 in the Atlantic region, NARR and ERA-Interim show significant underestimation compared to GEM.

- In the seasonal analysis of ETS differences, ERA-Interim skill is closer to GEM than NARR in most of seasons and regions.

- From the seasonal analysis of ETS differences in the Atlantic region, ERA-Interim would be more useful than NARR as alternative to GEM, particularly to forecast larger amounts of daily precipitation ($\geq 25mm$) in Fall and Winter.

Generally, ERA-Interim shows better performance over most of seasons and regions. Especially in Winter, ERA-Interim is more similar to GEM than NARR.

Chapter 5

Evaluation/Assessment of Bias Corrected Reanalysis Datasets

Biases are introduced into models for several reasons. One of the benefits of reanalysis data is that it does not solely use the numerical model output but assimilates observed information. However, some bias contained in observations may be transferred into reanalysis data (Adler et al., 2001) and add to the the model bias.

A problem usually found in numerical weather prediction models is that they have difficulty representing small scale meteorological changes which can have a significant impact on weather processes (Wilks, 2006). Since observed and simulated atmospheric circulation are never in perfect agreement, the selection and application of proper bias correction is necessary prior to use of the numerical weather models. Ines and Hansen (2006) points out that NWP models tend to predict more frequent low-intensity events which could affect the simulation of crop growth and yield in their study.

To make forcing data from NWP such as precipitation and temperature useful, some bias correction procedures are necessary. Several studies have employed different bias correction methods on reanalysis data and analyzed the results for hydrological and climate change studies. For example, Zhang et al. (2004) performed bias correction of daily precipitation from Tretyakov gauges and investigated the values of the correction factor (CF) seasonally and annually over the Mongolia region; Gutjahr and Heinemann (2013) compared three different bias correction methods using the regional climate model COSMO-CLM (consortium for small-scale modelling - climate limited area modelling), and Terink et al. (2010) evaluated bias-corrected ERA-15 (ECMWF-reanalysis data) precipitation and temperature in the Rhine basin. The methods of bias correction used in this study were descried in Chapter 3.

5.1 Results and discussions

The bias correction over Canada is generally successful since the differences between corrected reanalysis datasets and observations and RMSE have become smaller as shown in Table 5.1. The values of correlation stay relatively unchanged as expected.

The overall ETS score after bias correction is shown in Figure 5.1. Both ERA-Interim and NARR are slightly improved in terms of ETS after bias correction but not significantly.

Table 5.1: Comparison of daily precipitation Bias (Reanalysis-Observation) [mm] and RMSE [mm] over regions after bias correction ('Corrected ERA' denotes 'Corrected ERA-Interim')

Region	NARR		Corrected NARR		ERA-Interim		Corrected ERA	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Western region	0.05	8.7	-0.53	8.7	0.70	9.0	-0.39	8.4
Prairie region	0.24	4.1	-0.05	3.8	0.16	4.1	0.01	4.0
Central region	-0.48	5.7	-0.17	5.0	-0.05	6.0	0.05	6.0
Atlantic region	-1.13	7.1	-0.44	5.7	-0.14	7.1	0.12	6.0
Average	-0.33	6.4	-0.30	5.8	0.17	6.6	-0.05	6.4

The ETS score of NARR is improved at all thresholds, while the ETS of ERA-Interim is improved only at thresholds under 5mm. The ETS of ERA-Interim at extreme precipitation ($\geq 25mm$) is slightly decreased after bias correction, however, the difference is small. Before bias correction, the ETS of ERA-Interim performs better than NARR, but after bias correction they are rather similar in terms of ETS.

Figure 5.2 shows FBI-1 including data before and after bias correction. It seems that the impact of bias correction on ERA-Interim is bigger than on NARR. For NARR, the bias correction reduces bias from overestimation at thresholds less than 5mm and from underestimation at thresholds of more than 5mm. However, it seems to move FBI-1 of ERA-Interim to overall underestimation, which ends up causing more underestimation at thresholds over 10mm although it reduces bias at thresholds under 5mm.

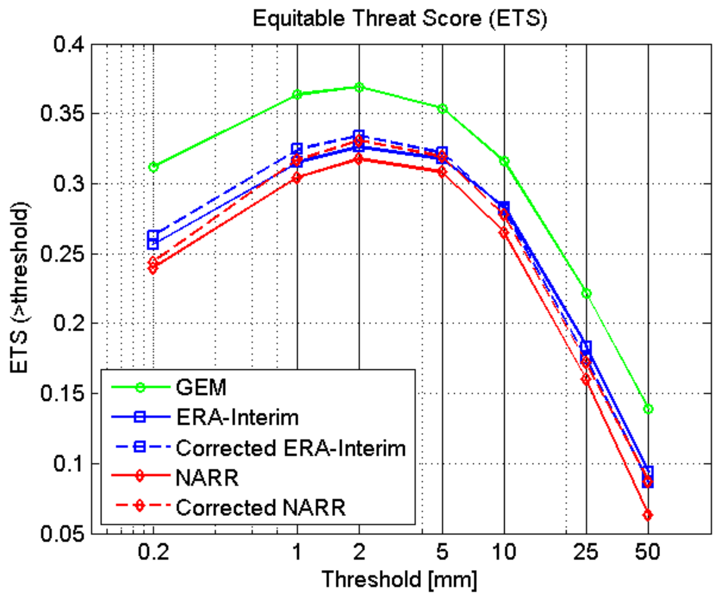


Figure 5.1: Equitable Threat Score (ETS) result of reanalysis datasets over all seasons and all study regions after bias correction: solid and dashed lines represent original and bias corrected data set, respectively

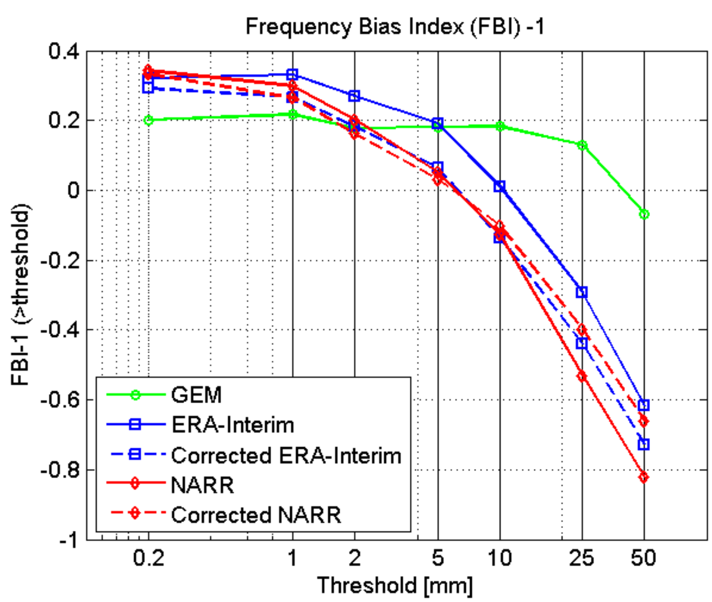


Figure 5.2: Frequency Bias Index (FBI)-1 result of corrected reanalyses over all seasons and all study regions: solid and dashed lines represent original and bias corrected data set, respectively

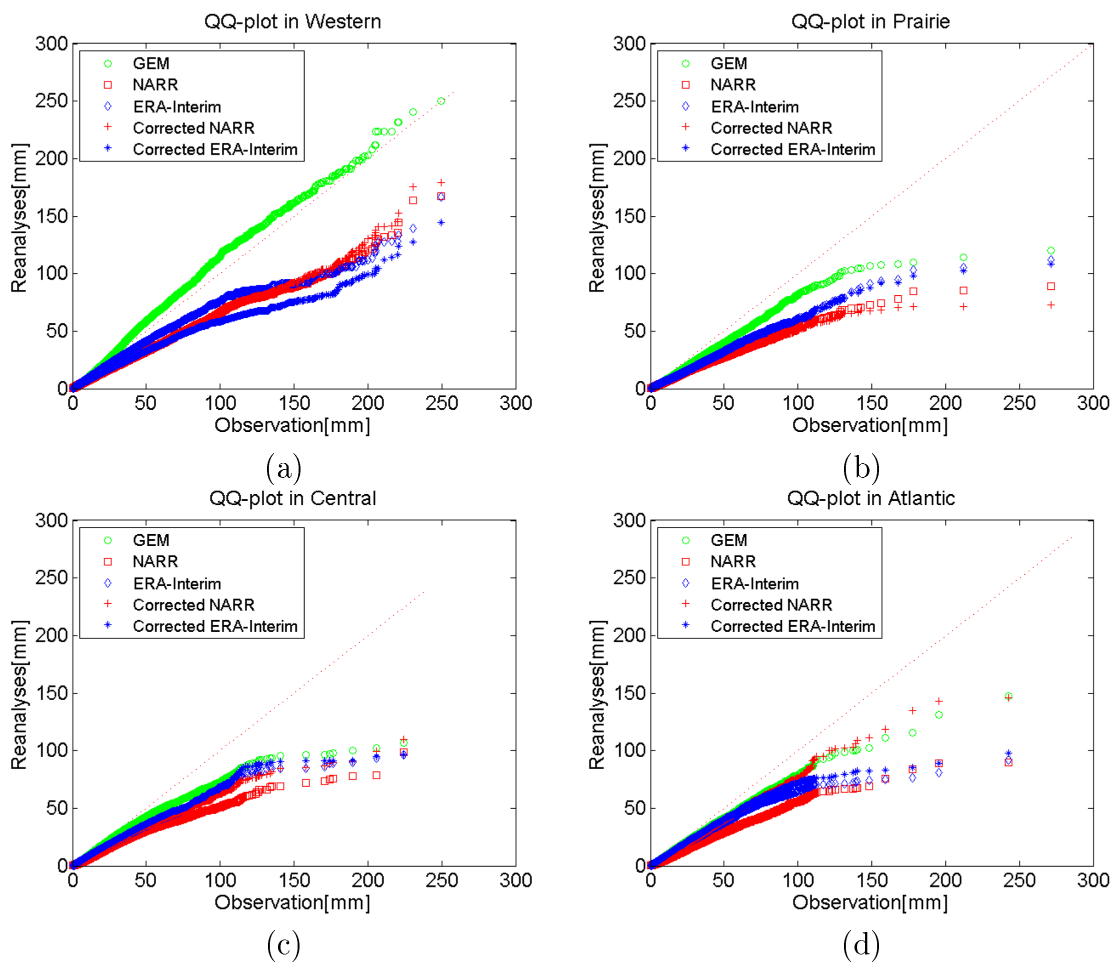


Figure 5.3: QQ plot of reanalysis datasets over regions after bias correction: (a) Western (b) Prairie (c) Central (d) Atlantic

5.1.1 Regional Evaluation of Bias-corrected Reanalysis Datasets

Figure 5.3 shows regional Q-Q plots which compares distributions of ERA-Interim and NARR before and after bias correction and shows how improved the data sets are compared to GEM. It seems of degree to which ERA-Interim and NARR are affected by the bias correction depends on the region. In the Western region, NARR becomes slightly improved after bias correction as the distribution is closer to the 1:1 line against observations. However, the ERA-Interim distribution is further away from the 1:1 line, showing more underestimation after bias correction. In the Prairie region, ERA-Interim stays almost the same while NARR shows more underestimation after bias correction. In the Central region, ERA-Interim also stays almost the same while NARR becomes improved and more similar to ERA-Interim after bias correction. In the Atlantic region, NARR shows significant improvement after bias correction while ERA-Interim shows just slight improvement.

In terms of Q-Q plots, the bias correction seems to work well on NARR in the Western, Central, and Atlantic regions, but not in Prairie region. Especially in Atlantic region, significant improvement is seen in NARR, even better than GEM after bias correction.

Compared to NARR, the bias correction seems less successful for ERA-Interim. There are improvements in Central and Atlantic regions but the improvements are small. In the Western region, ERA-Interim before bias correction seems better to use than the corrected ERA-Interim.

GEM generally is superior to even the bias-corrected reanalysis data sets. After bias correction, biases in the reanalysis data sets are reduced, and RMSE and correlation are unchanged. This is from the smoothing effect from spline interpolation method and reduction of some systematic errors that existed in the observation data. Through the application of bias correction, the distribution of NARR against observations becomes closer to the distribution of GEM against observations over most regions while ERA-Interim is slightly more off from the distribution of GEM. However, NARR shows slightly more underestimation against observation in the Central region after application of bias correction.

Figure 5.4 shows ETS changes after bias correction in different regions. There are improvements at thresholds under 5mm in the Western and Prairie regions, however, improvements are not noticeable at thresholds over 5mm in the regions. The ETS of NARR and ERA-interim are very similar before and after bias correction in the Central region. The ETS improvement of ERA-Interim is small as well in the Atlantic region, but NARR seems significantly improved at thresholds over 2mm, and especially for extreme precipitation ($\geq 25mm$).

Figure 5.5 shows FBI-1 after bias correction. The changes after bias correction seems bigger in FBI-1 than in ETS, but varies depending on the region. Biases are reduced at relatively lower thresholds ($\leq 5mm$) while models become more underestimated at the relatively higher thresholds ($\geq 10mm$) in the Western, Prairie and Central regions. In the Western, Prairie and Central regions, the FBI-1 of NARR

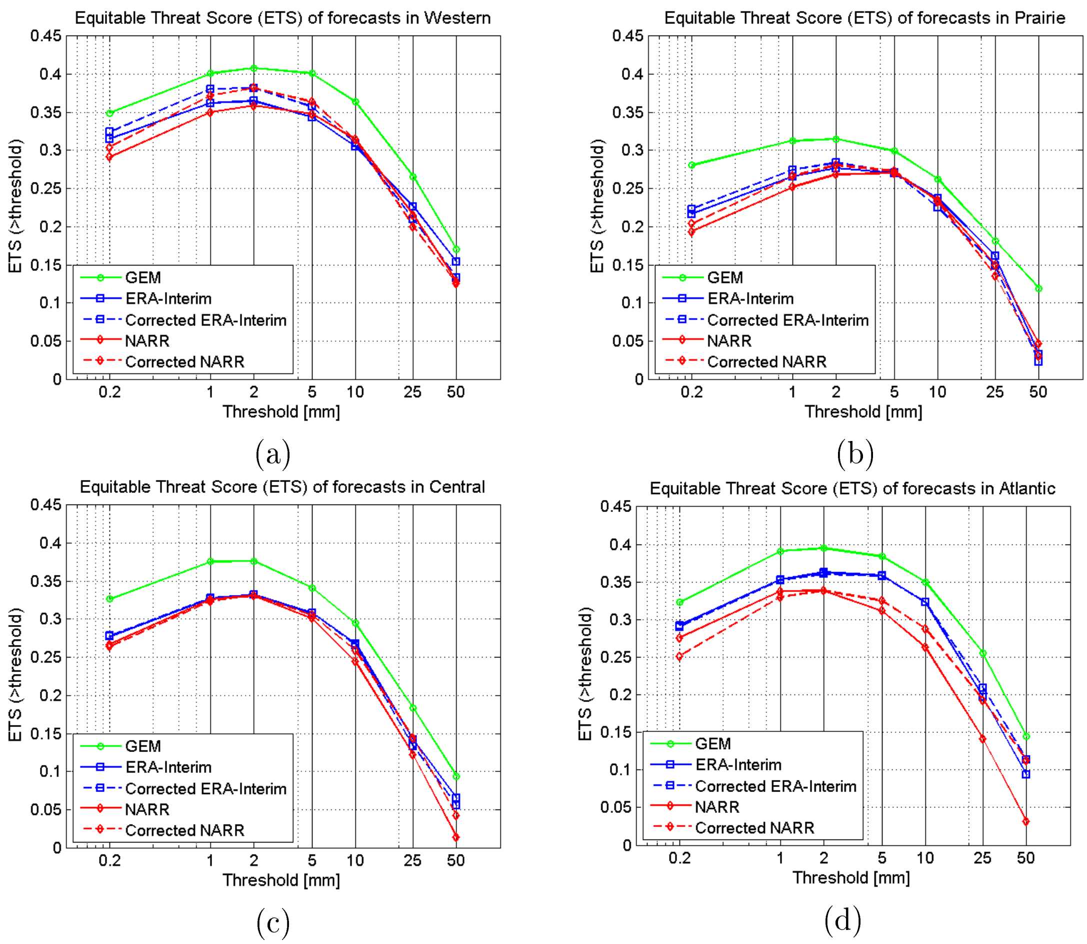


Figure 5.4: Equitable Threat Score (ETS) of forecasts over regions. Solid and dashed lines represent original and corrected forecasts, respectively: (a) Western (b) Prairie (c) Central (d) Atlantic

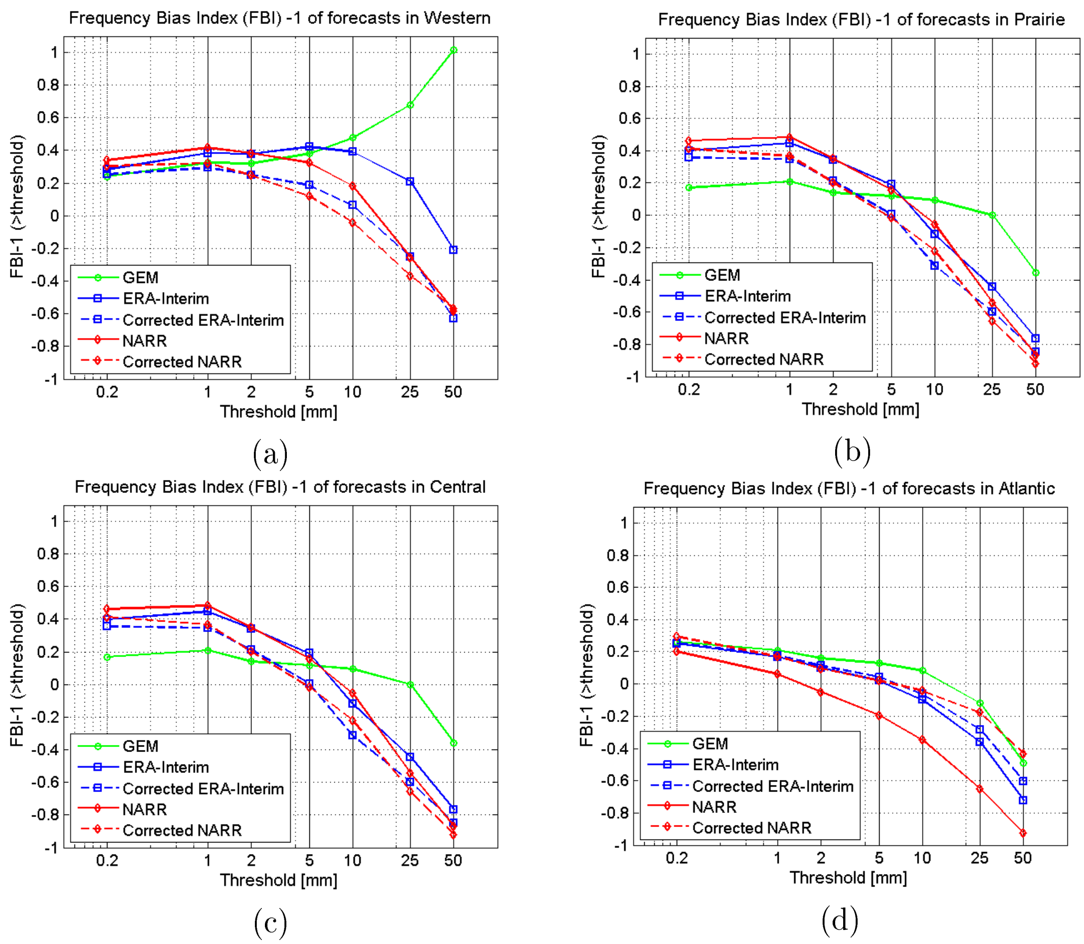


Figure 5.5: Frequency Bias Index (FBI) of forecasts over regions. Solid and dashed lines represent original and corrected forecasts, respectively: (a) Western (b) Prairie (c) Central (d) Atlantic

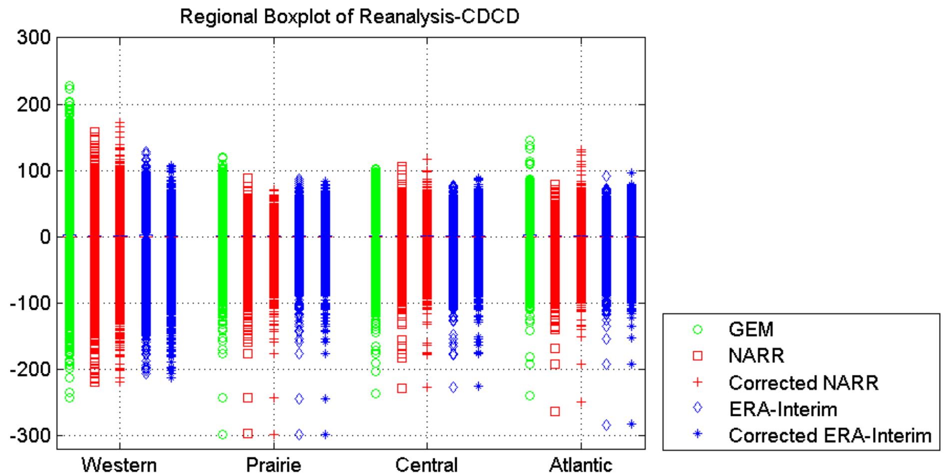


Figure 5.6: Boxplot of differences between corrected reanalyses and CDCD over regions

and ERA-Interim is more similar to GEM at lower thresholds ($\leq 5mm$) after bias correction, but not at higher thresholds ($\geq 10mm$).

However, in the Atlantic region, the FBI-1 becomes significantly improved for NARR and slightly improved for ERA-Interim at all thresholds, being very similar to GEM after bias correction.

Figure 5.6 shows regional boxplot of the differences between reanalysis data sets and CDCD. The impact of bias correction on overestimation is bigger than on underestimation.

In the Western region, the corrected NARR ranges more toward overestimation. Some positive biases of NARR become larger after bias correction. However, biases in ERA-Interim on the overestimation side become smaller.

In the Prairie region, bias values of the corrected NARR moves closer to zero after bias correction while ERA-Interim remains relatively unchanged.

In the Central region, both NARR and ERA-Interim show slightly increased bias values after bias correction.

In the Atlantic region, bias values in overestimation increase in NARR while ERA-Interim remains similar after bias correction. Although the distribution of NARR moves upward a bit, it ends up making NARR close to GEM.

Through regional FBI-1 and boxplot analysis, it seems the bias correction method used works best for NARR in the Atlantic region.

5.1.2 Seasonal Evaluation of Bias-corrected Reanalysis Datasets

Figure 5.7 shows seasonal boxplot over the four regions after bias correction. The impact of bias correction is various depending on seasons and regions, but seems more effective over the Western and Prairie regions than the Central and Atlantic regions.

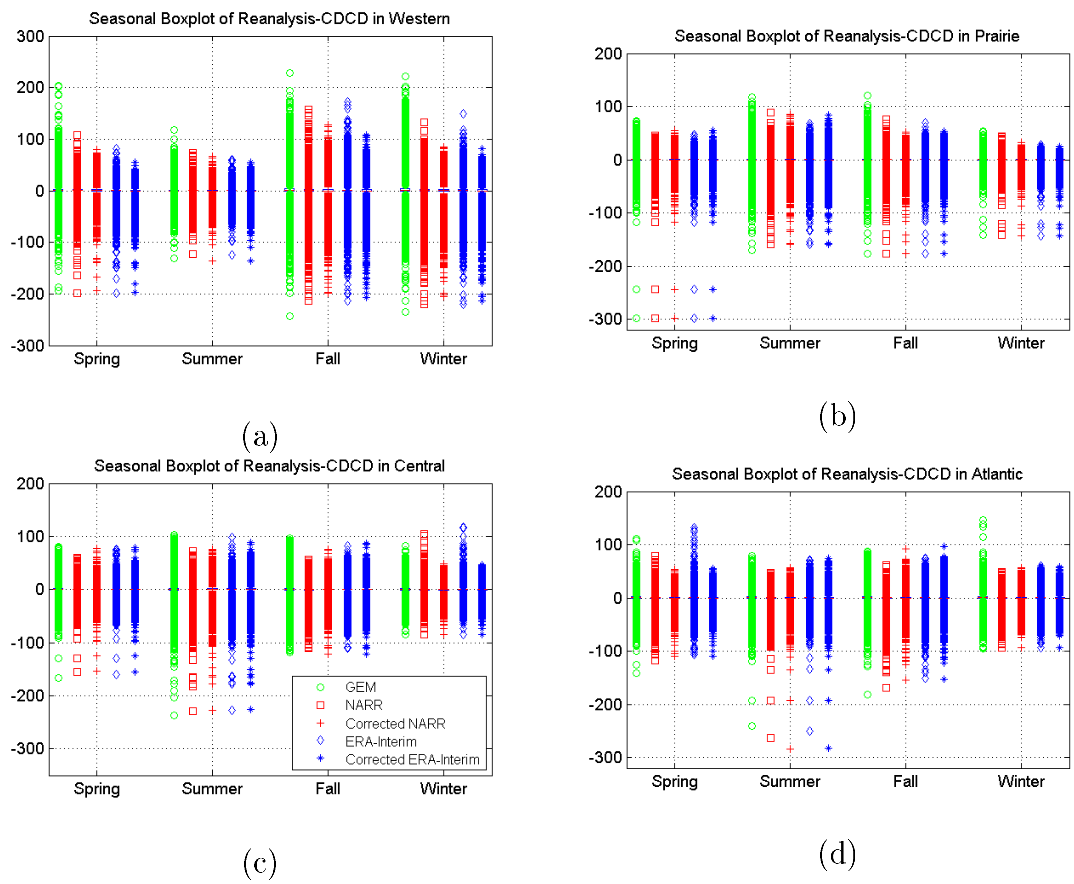
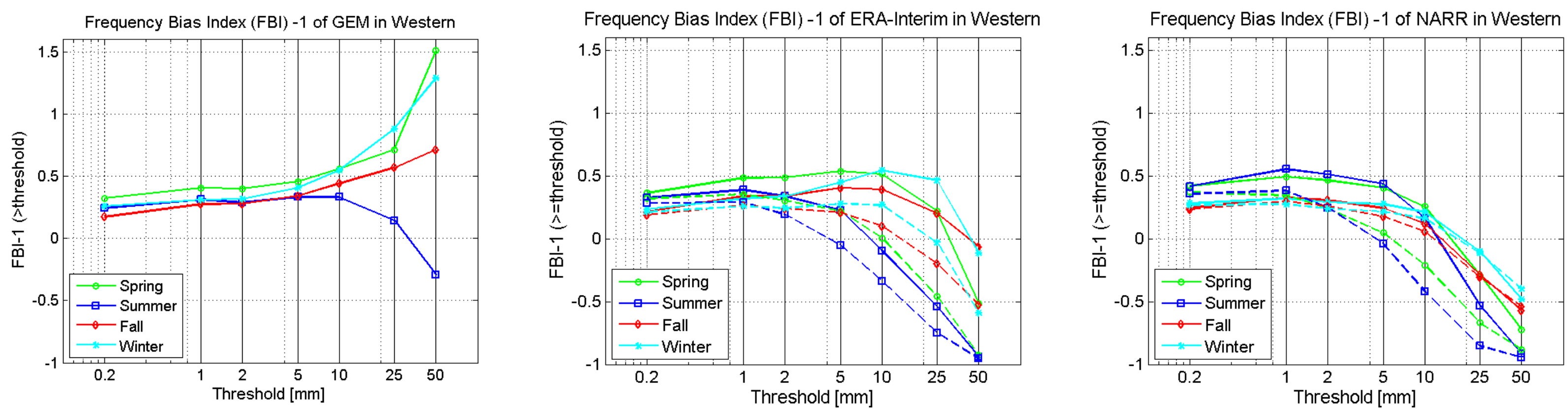
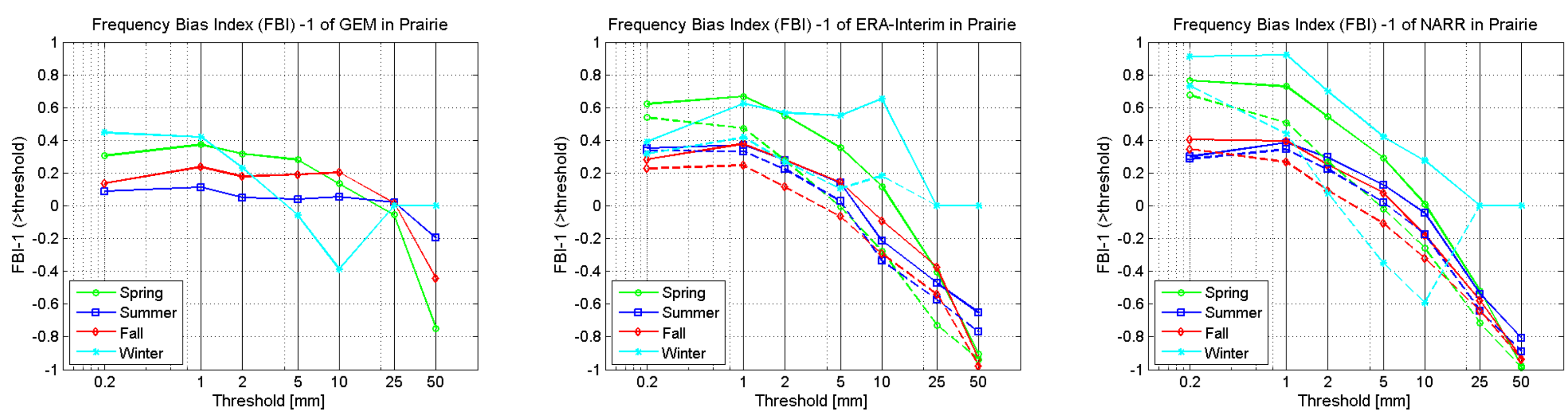


Figure 5.7: Seasonal boxplot of differences between corrected reanalyses and CDCD in different regions: (a) Western (b) Prairie (c) Central (d) Atlantic



(a) (b) (c)

Figure 5.8: Frequency Bias Index (FBI)-1 of corrected forecasts over seasons in the Western region. Solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM (b) ERA-Interim (c) NARR



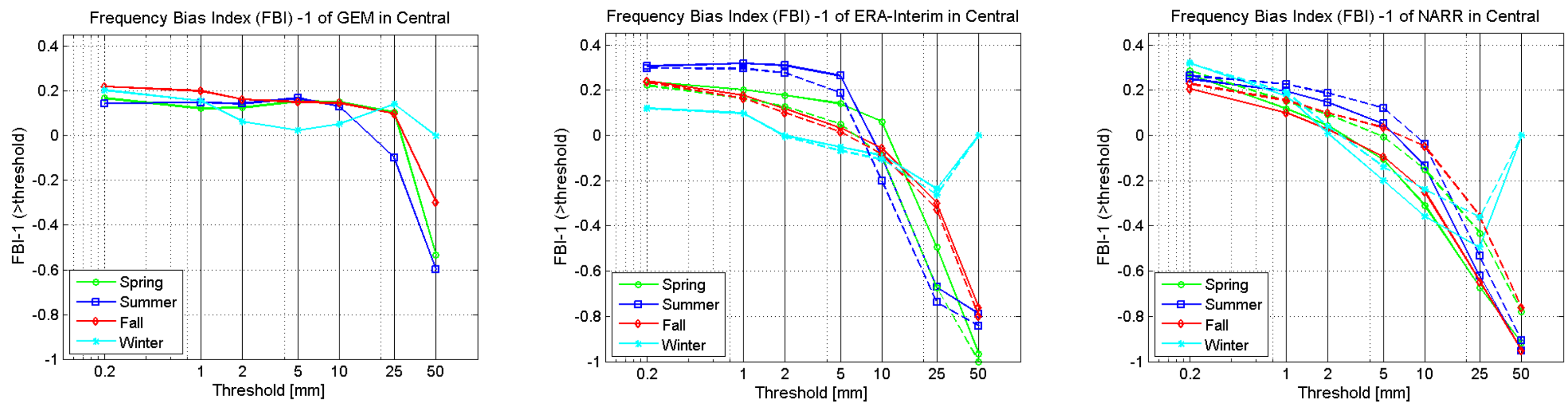
(a) (b) (c)

Figure 5.9: Frequency Bias Index (FBI)-1 of corrected forecasts over seasons in the Prairie region. Solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM (b) ERA-Interim (c) NARR

Figure 5.8 - 5.11 show seasonal FBI-1 before and after bias correction in Western, Prairie, Central, and Atlantic, respectively. The first, second, and third columns represent FBI-1 of GEM, ERA-Interim, and NARR, respectively.

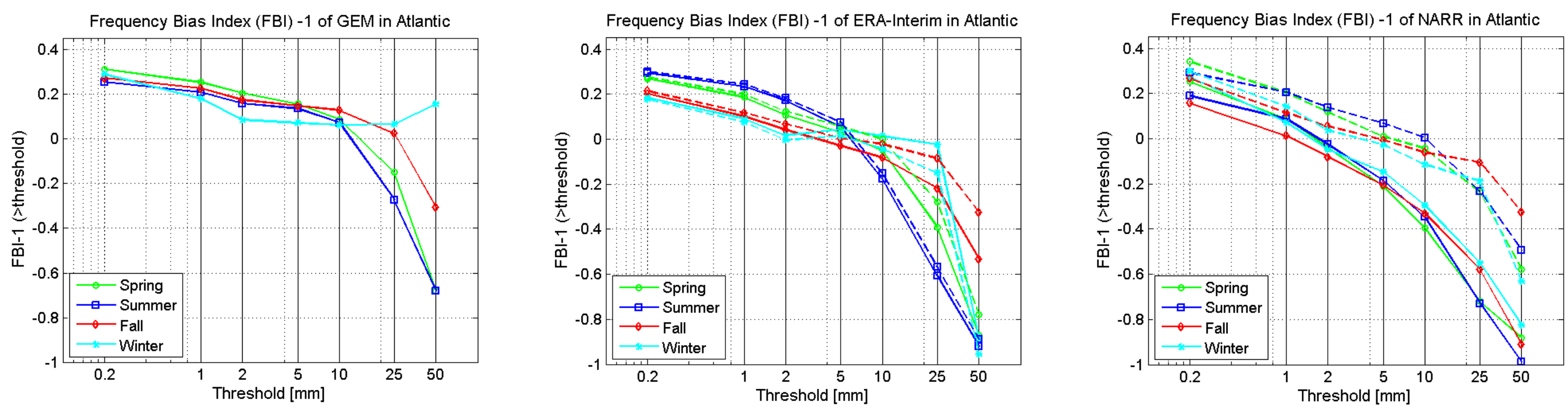
Over the Western region in Figure 5.8, overall FBI-1 of ERA-Interim shows relatively big changes compared to other regions after bias correction. Over all thresholds, the FBI-1 is moved to lower value after bias correction. It ends up causing the reduction of overestimation in thresholds equal or less than 10mm while increase of underestimation is caused in thresholds above 10mm. Similar to FBI-1 of ERA-Interim, FBI-1 of NARR also shows lower values after bias correction in all seasons. However, the changes are more noticeable in Spring and Summer, and significantly slight in Fall and Winter. As found in FBI-1 of ERA-Interim, overestimation is reduced in lower thresholds ($\leq 5mm$) but underestimation is increased in larger thresholds ($\leq 10mm$) in Spring and Summer in the FBI-1 of NARR.

Over the Prairie region in Figure 5.9, the changes of FBI-1 of ERA-Interim and NARR are larger in Spring and Winter than in Summer and Fall after bias correction. In Spring and Winter, reduction of overestimation is noticeable over all thresholds after bias correction. However, there are more underestimation found at threshold 10mm in Winter after bias correction. Overall FBI-1 of ERA-Interim and NARR becomes close to zero and FBI-1 of GEM after bias correction.



(a) (b) (c)

Figure 5.10: Frequency Bias Index (FBI)-1 of corrected forecasts over seasons in the Central region. Solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM (b) ERA-Interim (c) NARR



(a) (b) (c)

Figure 5.11: Frequency Bias Index (FBI)-1 of corrected forecasts over seasons in the Atlantic region. Solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM (b) ERA-Interim (c) NARR

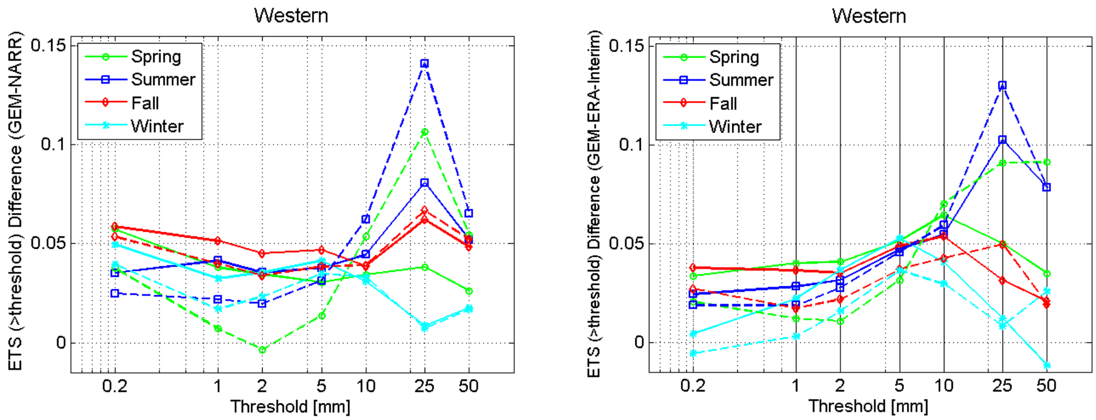
Over the Central region in Figure 5.10, FBI-1 of ERA-Interim and NARR are not much affected from bias correction compared to in the Western and Prairie regions. Even their changes are slight, the FBI-1 of both ERA-Interim and NARR becomes closed to zero and the FBI-1 of GEM after bias correction.

Over the Atlantic region in Figure 5.11, relatively large improvement is found in NARR after bias correction at large thresholds ($\geq 25mm$) which also shown previously in Q-Q plot and ETS. FBI-1 of ERA-Interim shows slight improvement after bias correction. Both ERA-Interim and NARR become closed to GEM after bias correction. Therefore, the bias correction seems contributing in NARR in the Atlantic region, especially at thresholds of extreme precipitation ($\geq 25mm$).

In summary, more positive impact of bias correction is found in Prairie and Atlantic.

5.1.3 Differences in skill between GEM and reanalysis datasets

Figure 5.12 - 5.15 show differences of ETS between datasets before and after bias correction. Solid lines represent 'before bias correction' and dashed lines 'after bias correction'. Values close to zero means that the given dataset, either NARR or ERA-Interim, has similar forecasting skill score as GEM. The bias correction used in this study does affect the forecasting skills of each dataset, however, the average of ETS score changes is less than 0.05. The relatively biggest changes in the analysis is seen in GEM-NARR over large precipitation thresholds ($\geq 25mm$) in the Central and



(a) (b)

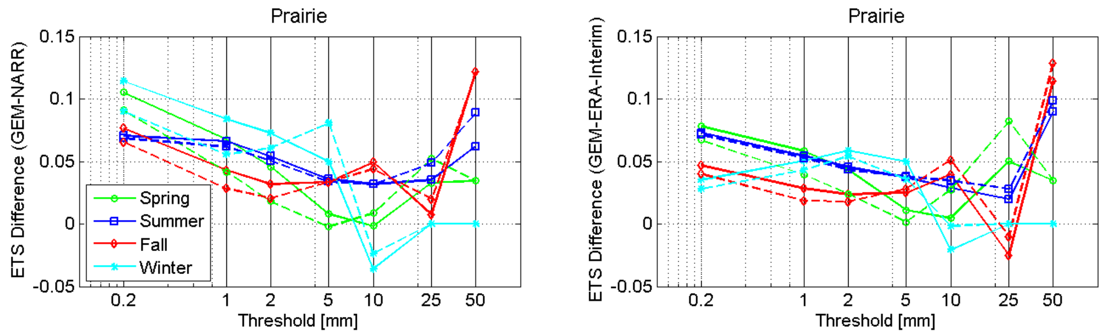
Figure 5.12: Equitable Threat Score (ETS) differences between GEM and corrected reanalysis forecasts in Western region, solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM minus NARR (b) GEM minus ERA-Interim

Atlantic regions in Fall and over small amount of precipitation thresholds ($\leq 2mm$) in the Western region in Spring.

Both GEM-NARR and GEM-ERA-Interim increase after bias correction at thresholds of extreme precipitation ($\geq 25mm$) in the Western region. This implies that it is not recommended to use the corrected NARR and ERA-Interim at any thresholds in Western region in any season.

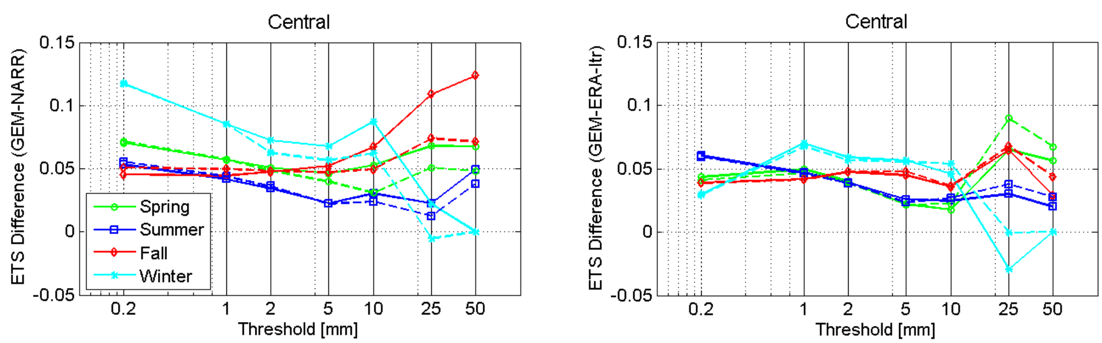
5.2 Summary

The bias correction used in this study was relatively more effective for NARR than for ERA-Interim. As mentioned earlier, part of the reason is the impact of the smoothing process and resolution of the datasets. Relatively simple bias correction methods were used in this study due to the size of the datasets. Modification of the



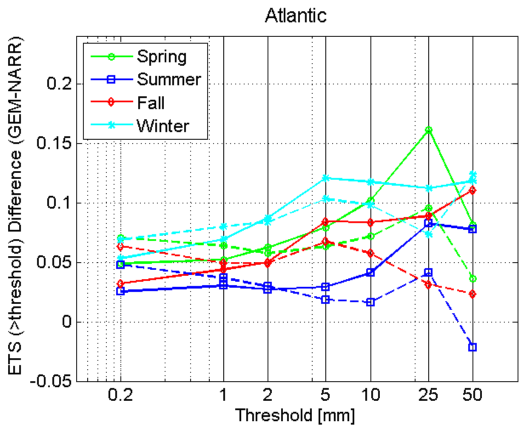
(a) (b)

Figure 5.13: Equitable Threat Score (ETS) differences between GEM and corrected reanalysis forecasts in Prairie region, solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM minus NARR (b) GEM minus ERA-Interim

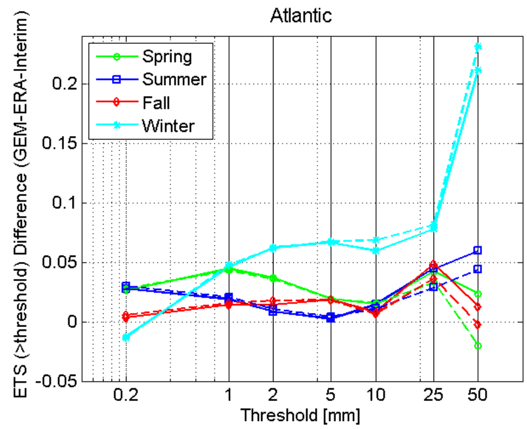


(a) (b)

Figure 5.14: Equitable Threat Score (ETS) differences between GEM and corrected reanalysis forecasts in Central region, solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM minus NARR (b) GEM minus ERA-Interim



(a)



(b)

Figure 5.15: Equitable Threat Score (ETS) differences between GEM and corrected reanalysis forecasts in Atlantic region, solid and dashed lines represent original and corrected forecasts, respectively: (a) GEM minus NARR (b) GEM minus ERA-Interim

method is suggested for future studies for more improvement in the quality of the datasets.

- ERA-Interim shows improvements in mid-range of precipitation thresholds (2-25mm) after correction, however, it seems to no improvements on extreme precipitation (more than 50mm) especially in the Western region.

- In the Q-Q plots, the bias correction seems to work well on NARR in all regions, except for the Prairie region. Especially in the Atlantic region, significant improvement is seen in NARR, even better than GEM after bias correction.

- NARR improves after bias correction over all regions in the Q-Q plots, boxplots, ETS and FBI-1.

- ERA-Interim is not improved as much as NARR after bias correction. Sometimes, the original ERA-Interim is better than the corrected ERA-Interim, for example, in the Western region.

- In the seasonal analysis of FBI-1 in the Prairie region, overestimation in both of NARR and ERA-INTERIM is significantly reduced after bias correction in Spring and Winter.
- In the seasonal analysis of FBI-1 in the Central and Atlantic regions, underestimation in NARR and ERA-INTERIM is reduced after bias correction. However, the changes are noticeable only in NARR in the Atlantic region.

Chapter 6

Evaluation/Assessment of Reanalysis Datasets after CaPA Assimilation Process

After bias correction, the gridded corrected reanalysis data sets were subjected to the CaPA assimilation process. As mentioned in Section 2.2, CaPA also has its inherent bias correction and interpolation processes, and therefore, this study uses the original reanalysis datasets for CaPA process along with the corrected datasets. The CaPA process is more effective at the 6-hour time frame than at the 24-hour time frame due to the way it is set up for the operation purpose. This study used the 24-hour accumulated precipitation time step due to the long study time period (10 years). Accordingly, bias correction schemes were needed prior to the CaPA assimilation

Table 6.1: Comparison of daily precipitation Average[mm], RMSE[mm], and Correlation(R) over regions after CaPA assimilation process

Region	Obs	GEM			NARR			ERA-Interim		
		Avg	RMSE	R	Avg	RMSE	R	Avg	RMSE	R
Western region	3.9	3.5	6.3	0.76	3.0	6.7	0.72	3.1	6.7	0.72
Prairie region	1.3	1.2	3.4	0.66	1.1	3.4	0.64	1.1	3.4	0.64
Central region	2.7	2.6	4.6	0.70	2.3	4.9	0.65	2.3	4.8	0.66
Atlantic region	3.3	3.3	6.0	0.69	2.7	6.6	0.58	3.0	6.4	0.62
Average	2.8	2.7	5.1	0.70	2.3	5.4	0.65	2.4	5.3	0.66

process. Both the original and bias-corrected reanalysis datasets were applied in the CaPA assimilation process. Through the step, one is able to see not only how the CaPA process impacts on datasets but also how reanalysis datasets are positively affected as comparing to GEM.

6.1 Results and discussions

Table 6.1 shows the average, RMSE, and correlation of datasets over different regions after the CaPA assimilation process. After CaPA process, original and bias-corrected NARR and ERA-Interim are improved as comparing to CDCD observation. GEM is still superior, having the least RMSE and the highest correlation, but the differences between GEM and the reanalysis datasets are significantly reduced after the CaPA assimilation process.

Figure 6.1 shows ETS after the CaPA assimilation process over all regions and seasons. The figure clearly shows that the ETS of the reanalysis datasets becomes almost identical to GEM over all regions and seasons. This could be a good sign

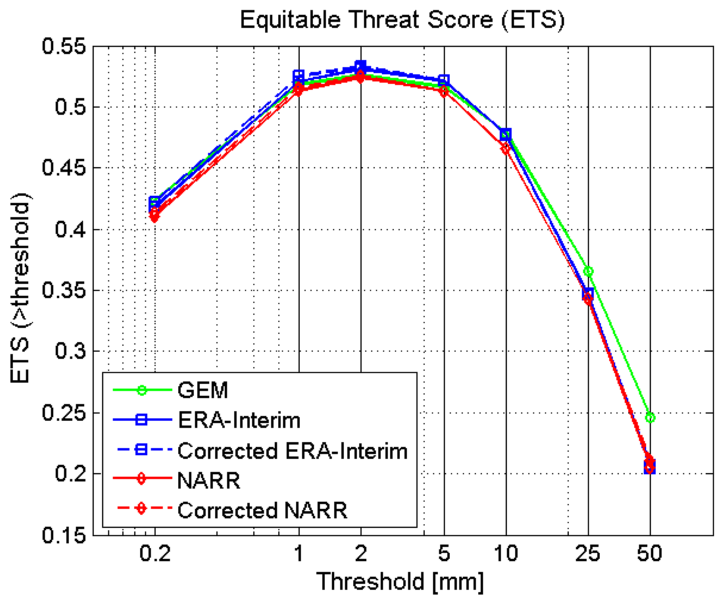


Figure 6.1: Equitable Threat Score (ETS) results of GEM, NARR, ERA-Interim after CaPA assimilation process

since these reanalysis datasets have now shown to be good alternatives to GEM to generate CaPA back in time.

Figure 6.2 shows the FBI-1 after the CaPA assimilation process over all regions and seasons. Although the FBI-1 of reanalysis datasets is not as close to zero as GEM at large thresholds ($\geq 25mm$), improvement is shown as the scores become close to GEM at all thresholds. In addition, corrected reanalysis datasets after CaPA assimilation process do not show much difference from the datasets before CaPA assimilation process in both of ETS and FBI-1.

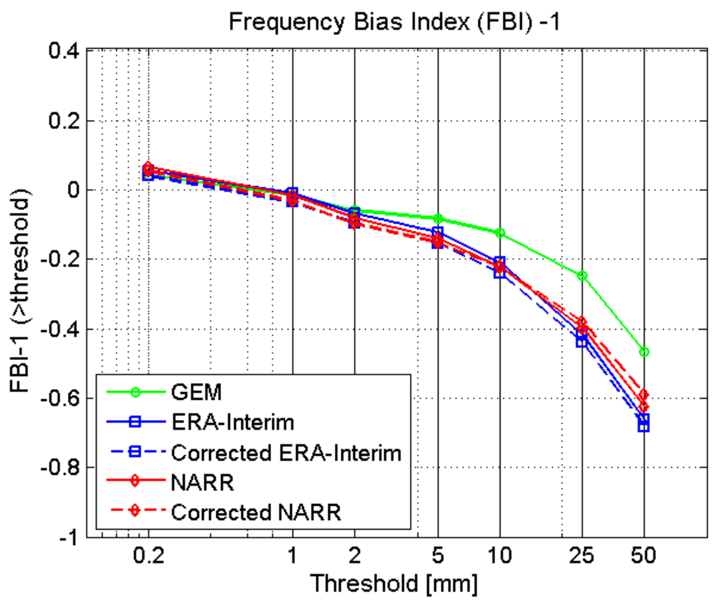


Figure 6.2: Frequency Bias Index (FBI)-1 results of GEM, NARR, ERA-Interim after CaPA assimilation process

6.1.1 Regional Evaluation of Reanalysis Datasets after CaPA Assimilation Process

QQ-plots are also used to evaluate the performance of the reanaylsis datasets after the CaPA process. Regional QQ-plots after the CaPA assimilation process are shown in Figure 6.3. Over all regions, the CaPA process transforms distributions of all datasets in such a way that they become close to each other. Through the CaPA assimilation process, the distributions of NARR and ERA-Interim become closer to GEM. It further demonstrates the possibility of using reanalysis datasets as alternative background for the CaPA process, even if the time frame of analysis (24-h) is longer than the operation one (6-h). There are some differences between datasets in the Western region and at extreme precipitation over all regions, but the difference

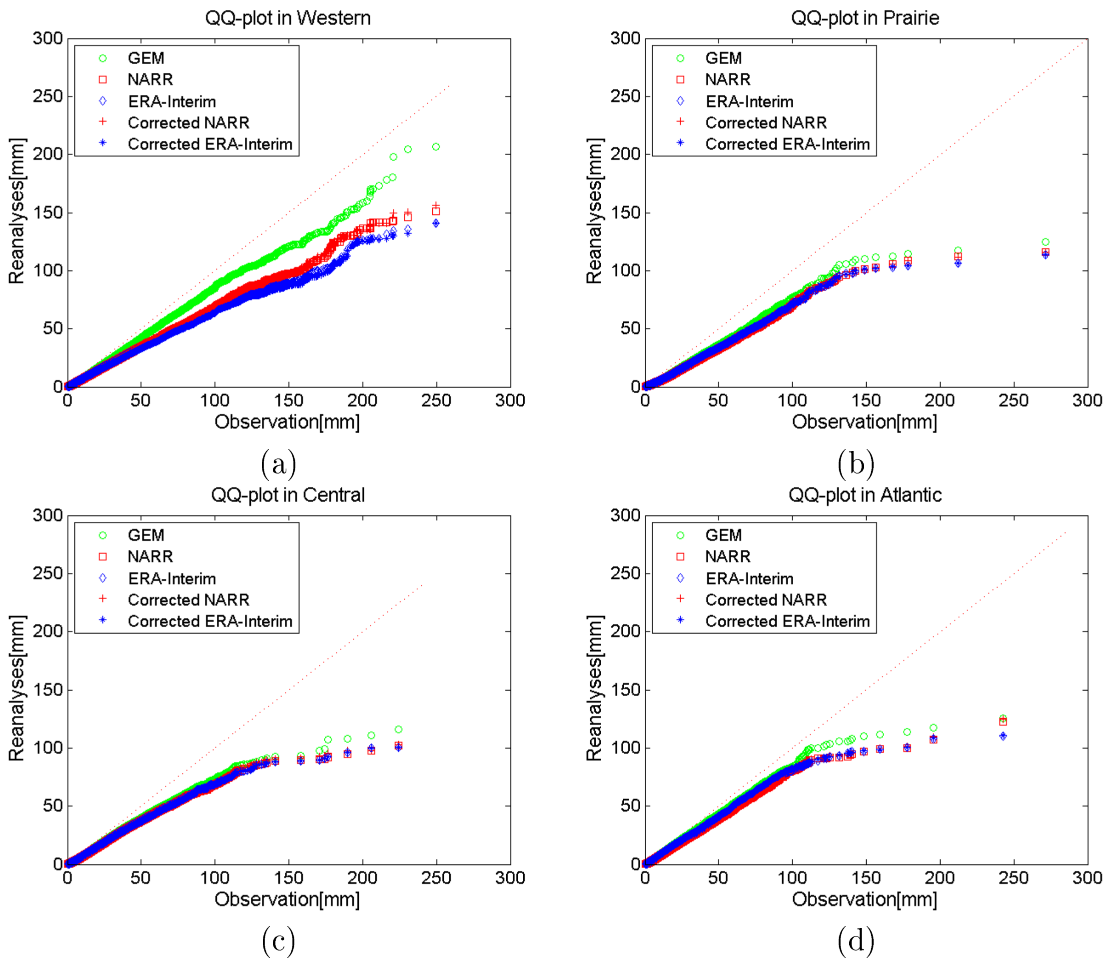


Figure 6.3: QQ plot of reanalysis datasets against observations over regions after CaPA assimilation: (a) Western (b) Prairie (c) Central (d) Atlantic

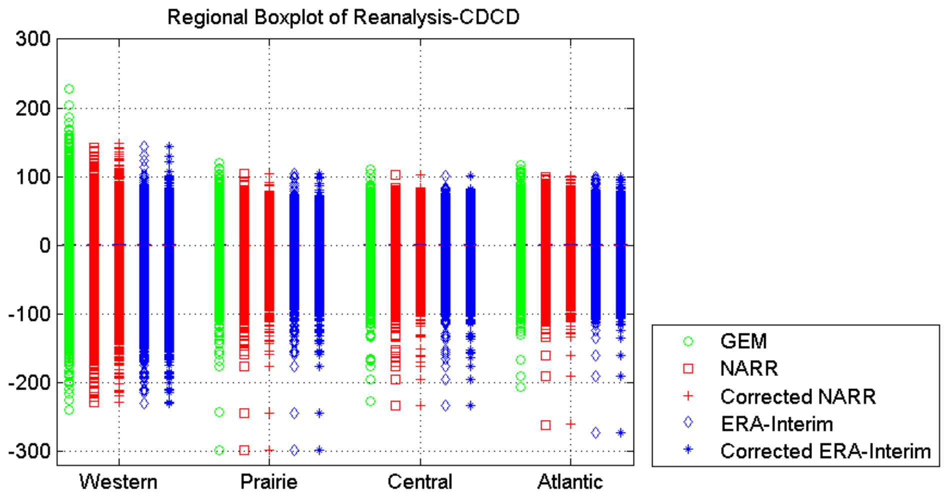


Figure 6.4: Boxplot of differences between reanalysis datasets and CDCD over regions after CaPA assimilation process

becomes significantly reduced after the CaPA process.

Figure 6.4 shows boxplots of differences between reanalysis datasets after the CaPA process and observations over regions. Similar to previous analyses, the datasets become closer to GEM after the CaPA process, except for the Western region. The differences from observations in GEM range wider than for NARR and ERA-Interim in the Western region while is narrower than NARR and ERA-Interim in the Atlantic region.

Figure 6.5 shows ETS over regions after the CaPA assimilation process. As in Figure 6.1, datasets become close to GEM after the CaPA process no matter whether the datasets are corrected or not. However, in the Atlantic region, ERA-Interim is closer to GEM, which implies ERA-Interim is recommended as alternative background to GEM in that region.

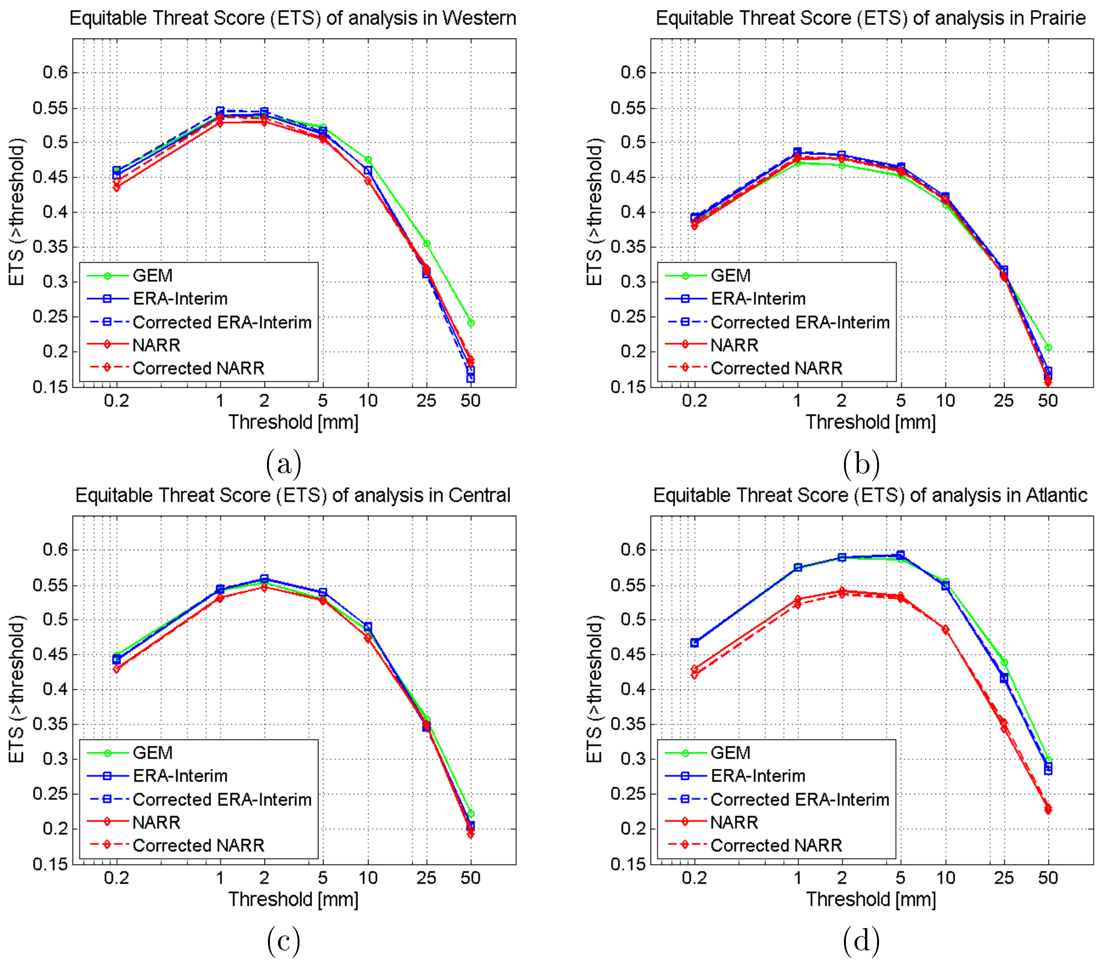


Figure 6.5: Equitable Threat Score (ETS) of reanalyses against observation over regions after CaPA assimilation process: (a) Western (b) Prairie (c) Central (d) Atlantic

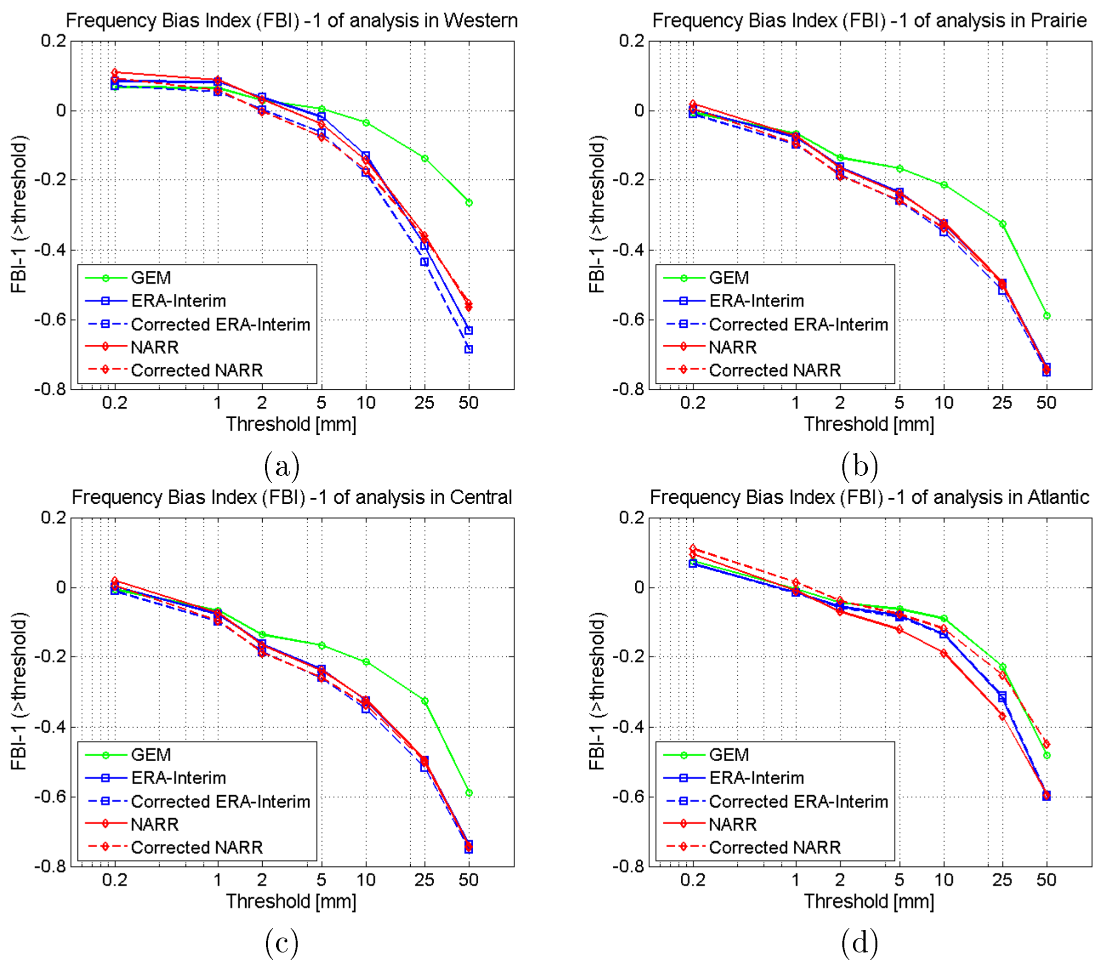


Figure 6.6: Frequency Bias Index (FBI)-1 of reanalysis datasets over regions after CaPA assimilation process: (a) Western (b) Prairie (c) Central (d) Atlantic

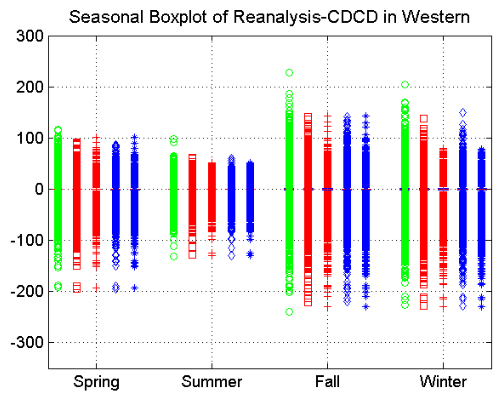
Figure 6.6 shows the FBI-1 of reanalysis datasets after the CaPA assimilation process over regions. In terms of FBI-1, the reanalysis datasets are to GEM over all regions. The improvement can be found more at lower thresholds. GEM still shows less bias than NARR and ERA-Interim at extreme thresholds over all regions.

6.1.2 Seasonal Evaluation of Reanalysis Datasets after CaPA Assimilation Process

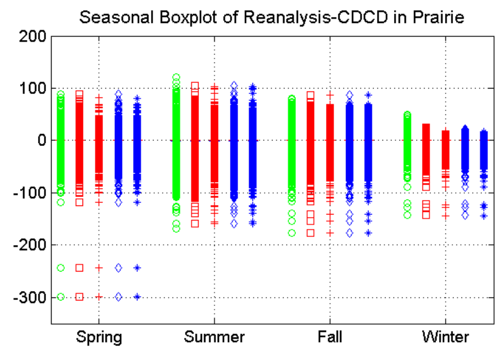
Figure 6.7 shows seasonal boxplot over the four regions. The wide range of differences between GEM and observations in the Western region observed in Figure 6.4 are mainly due to the Fall and Winter seasons, while Summer has a narrower range. However, overall ranges of GEM-observation, NARR-obervation, and ERA-Interim-observation become significantly closed each other after the CaPA assimilation process regardless of bias correction.

Figures 6.8 and 6.9 show FBI-1 of GEM, NARR, and ERA-Interim in the Western and Prairie regions. In the Western region, the FBI-1 of NARR and ERA-Interim are similar. Both the bias-corrected NARR and the bias-corrected ERA-Interim have slightly smaller values than the original NARR and ERA-Interim. NARR and ERA-Interim produce more underestimation than GEM over all seasons as the threshold becomes larger.

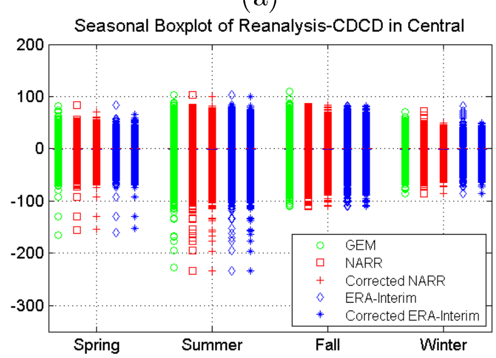
In the Prairie region, the FBI-1 of GEM, NARR, and ERA-Interim are almost identical except in Winter. In Winter, the bias-corrected ERA-Interim shows closer



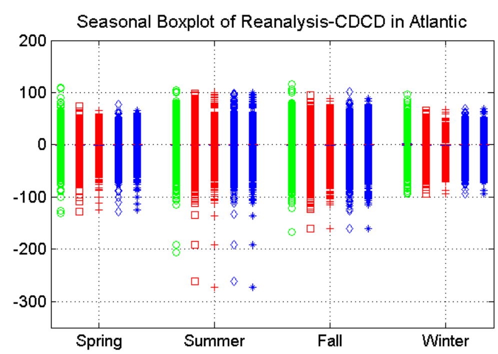
(a)



(b)



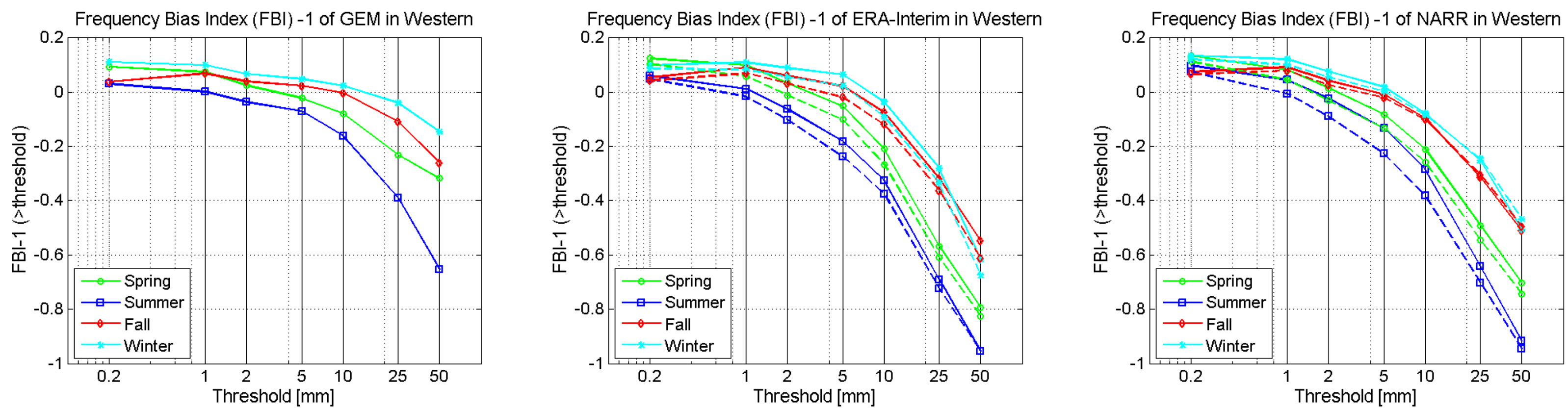
(c)



(d)

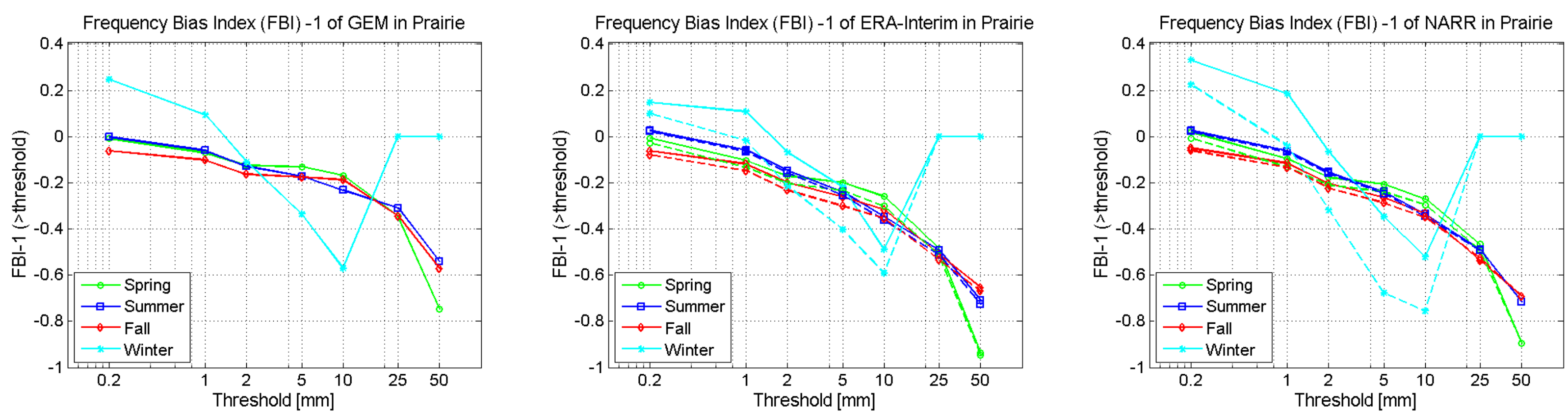
Figure 6.7: Seasonal boxplot of differences between reanalysis datasets and CDCD over regions after CaPA assimilation process: (a) Western (b) Prairie (c) Central (d) Atlantic

to GEM than ERA-Interim at thresholds more than 2mm while the corrected NARR is closer to GEM than NARR at thresholds less than 2mm.



(a) (b) (c)

Figure 6.8: Frequency Bias Index (FBI)-1 of reanalysis datasets after CaPA process over seasons in Western region: (a) GEM (b) ERA-Interim (c) NARR



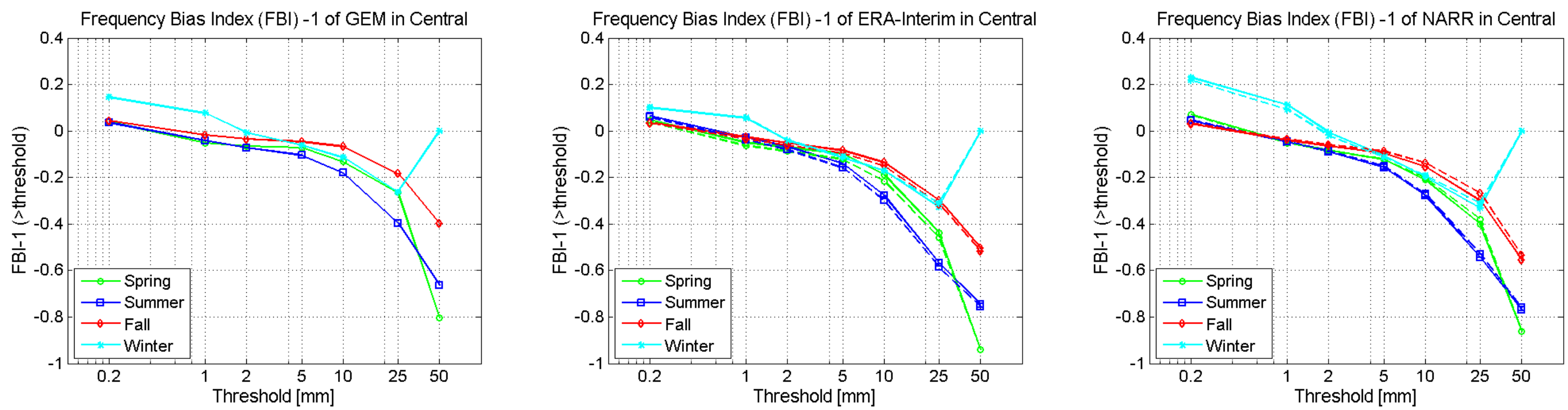
(a) (b) (c)

Figure 6.9: Frequency Bias Index (FBI)-1 of reanalysis datasets after CaPA process over seasons in Prairie region: (a) GEM (b) ERA-Interim (c) NARR

Figures 6.10 and 6.11 show the FBI-1 of GEM, NARR, and ERA-Interim in Central and Atlantic regions after application of CaPA-assimilation framework.

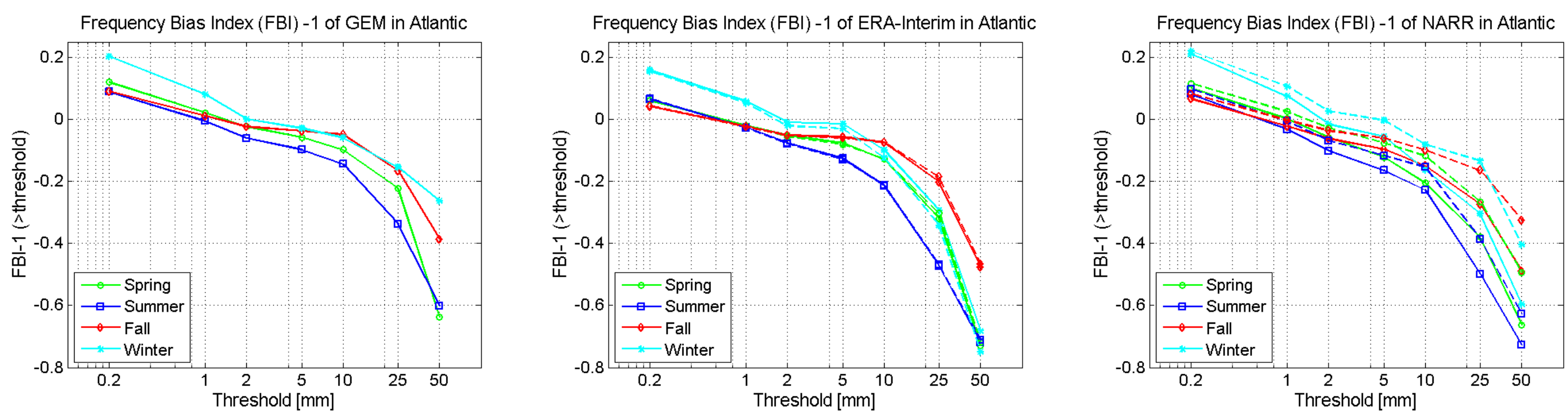
In the Central region, FBI-1 of NARR and ERA-Interim are almost the same as for GEM. There is a slight difference at lower thresholds ($\leq 1mm$) but the difference is small.

Overall the FBI-1 of GEM, NARR, and ERA-Interim are similar in the Atlantic region. The only noticeable point is that the FBI-1 between NARR and the bias-corrected NARR shows some difference. The corrected NARR shows less underestimation than NARR which results in it being closer to GEM over all seasons. In contrast, ERA-Interim and the bias-corrected ERA-Interim are almost the same in terms of FBI-1 after the CaPA process even though they had differences before the CaPA process.



(a) (b) (c)

Figure 6.10: Frequency Bias Index (FBI)-1 of reanalysis datasets after CaPA process over seasons in Central region: (a) GEM (b) ERA-Interim (c) NARR



(a) (b) (c)

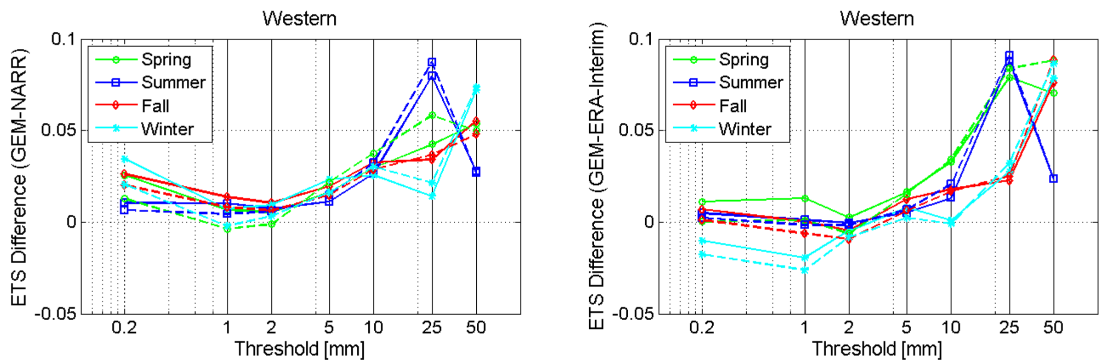
Figure 6.11: Frequency Bias Index (FBI)-1 of reanalysis datasets after CaPA process over seasons in Atlantic region: (a) GEM (b) ERA-Interim (c) NARR

6.1.3 Differences in Skill between GEM and Reanalysis Datasets

Figures 6.12 to 6.15 show differences of ETS scores between GEM and reanalysis datasets, considering both original and bias-corrected datasets after the CaPA assimilation process. We choose to show the differences with GEM rather than the actual value of ETS to better highlight the "cost" of using ERA-Interim or NARR instead of GEM. Interestingly, once datasets are through the CaPA process, the results become almost identical between original and bias-corrected datasets. This would seem to indicate that bias correction is not a critical or even necessary step when it is used in CaPA.

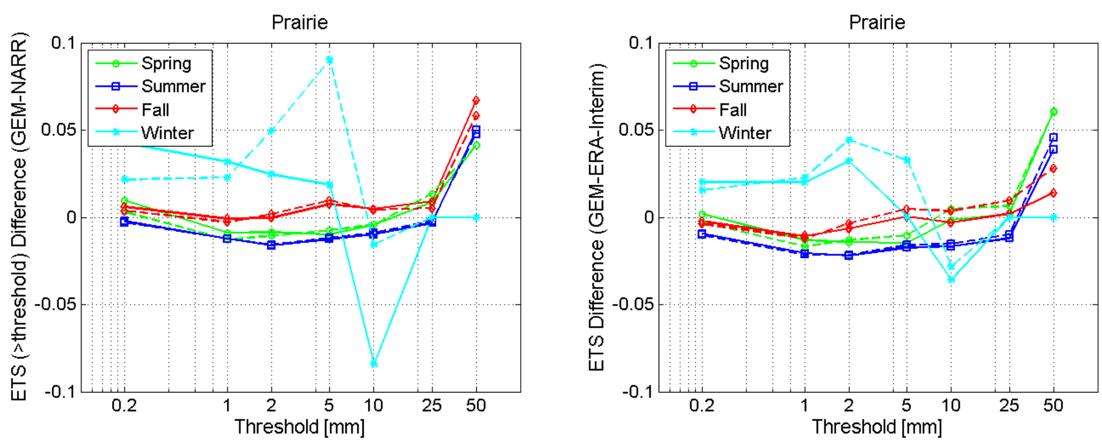
In the Western region, NARR and ERA-Interim have similar ETS over all thresholds. Both of them have ETS close to GEM at lower thresholds ($\leq 10mm$) and the differences becomes bigger as the threshold gets larger. At extreme precipitation thresholds ($\geq 25mm$), the differences between GEM and the reanalysis datasets reach a peak except in Summer. In Summer, the difference is the largest at the threshold of 25mm, and become lower at the thresholds of 50mm. In addition, ERA-Interim has slightly better ETS than GEM at thresholds lower than 2mm in Winter.

In the Prairie region, the ETS of NARR and ERA-Interim are quite close to GEM as they stays along the zero line, except in Winter. There are some differences between NARR and the corrected NARR in Winter in ETS. The ETS of NARR is closer to GEM than the corrected NARR at thresholds between 2mm and 10mm while corrected NARR is closer to GEM than NARR at thresholds over 10mm.



(a) (b)

Figure 6.12: Equitable Threat Score (ETS) differences between GEM and reanalysis datasets after CaPA process in Western region: (a) GEM minus NARR (b) GEM minus ERA-Interim



(a) (b)

Figure 6.13: Equitable Threat Score (ETS) differences between GEM and reanalysis datasets after CaPA process in Prairie region: (a) GEM minus NARR (b) GEM minus ERA-Interim

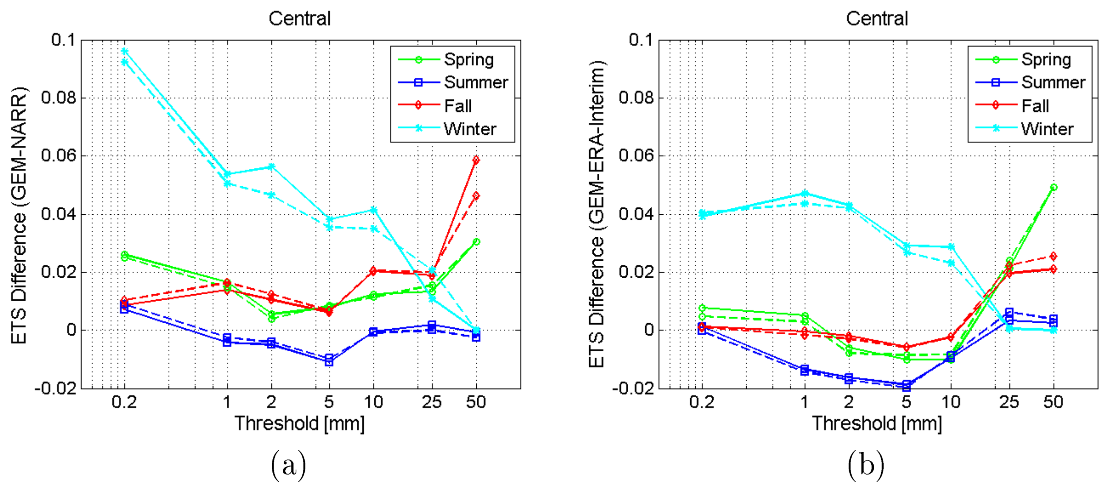


Figure 6.14: Equitable Threat Score (ETS) differences between GEM and reanalysis datasets after CaPA process in Central region: (a) GEM minus NARR (b) GEM minus ERA-Interim

In the Central region, the ETS difference of GEM-NARR and GEM-ERA-Interim is similar except in Fall and Winter. In Fall, the ETS difference of GEM-NARR is slightly bigger than GEM-ERA-Interim at the 50mm threshold. In Winter, significant differences are visible between GEM-NARR and GEM-ERA-Interim, especially at low thresholds ($\leq 1mm$), which shows that ERA-Interim is closer to GEM than NARR.

In the Atlantic region, the ETS difference of GEM-NARR sits more above zero while the ETS difference of GEM-ERA-Interim mostly sits under zero. This implies that the skill of GEM is better than NARR while the skill of ERA-Interim is better than GEM overall. NARR is the closest to GEM in Summer and ERA-Interim is the closest to GEM in Winter, except for thresholds more than 25mm.

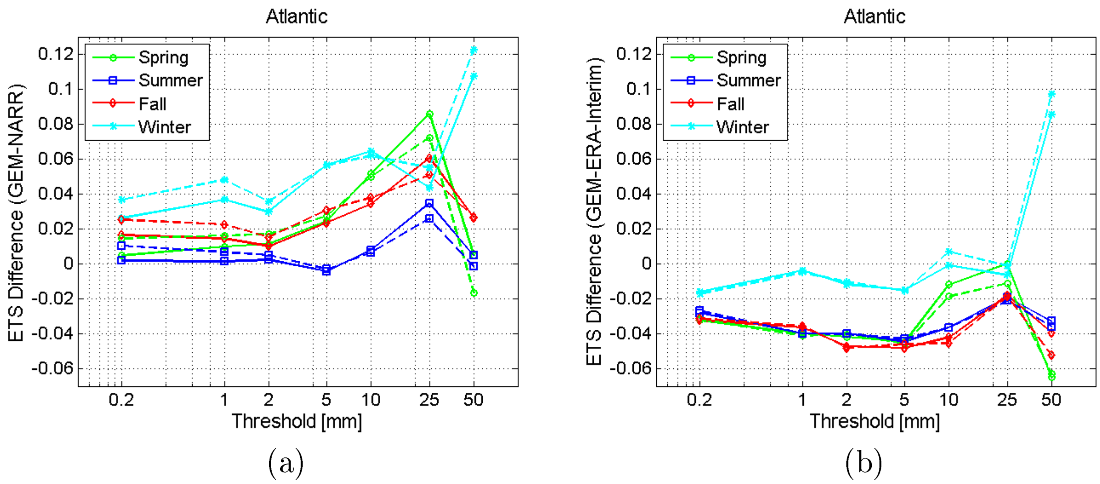


Figure 6.15: Equitable Threat Score (ETS) differences between GEM and reanalysis datasets after CaPA process in Atlantic region: (a) GEM minus NARR (b) GEM minus ERA-Interim

6.2 Summary

The CaPA framework does its job through statistical interpolation, quality control, assimilation of observations to generate improved analysis of the background field. It seems to work well on alternative background fields, as one can observe by generally higher scores of ETS and FBI-1 values closer to zero after the CaPA process. Overall, the improvement after the CaPA process is founded on mid-range thresholds of precipitation (2mm-25mm). Below is a summary of the results obtained in this chapter.

- After application of the CaPA process to the reanalysis datasets, overall ETS and FBI-1 improve and are closed to those of GEM.

- In the regional analysis of Q-Q plots, NARR and ERA-Interim become improved and close to GEM after the CaPA process. However, GEM is still superior at extreme precipitation. In the Western region, the distribution of NARR is slightly better than ERA-Interim and closer to GEM.

- In the regional analysis of Boxplots, overestimation of observations is found more in GEM than in the reanalysis datasets in the Western region. In the Atlantic region, overestimation are similar for GEM and reanalysis datasets but underestimation is pronounced less in GEM than in the reanalysis datasets.

- In the regional analysis of ETS, ERA-Interim shows higher skill than NARR in the Atlantic region. Higher skill in ERA-Interim is also shown over all seasons in seasonal analysis of ETS differences.

- In the regional analysis of FBI-1, GEM has the least bias at thresholds more than 10mm over all regions. In the Atlantic region, the corrected NARR is close to GEM.

- In the seasonal analysis of FBI-1 in the Prairie region, the corrected ERA-Interim is closer to GEM than ERA-Interim at thresholds more than 2mm in Winter. However, the corrected NARR is closer to GEM than NARR at thresholds less than 2mm in Winter.

- In the seasonal analysis of FBI-1 in the Atlantic region, the corrected NARR is noticeably closer to GEM than NARR after the CaPA process over all seasons. However, the corrected ERA-Interim and ERA-Interim are almost the same after the CaPA process even though they show differences before the CaPA process.

- In the analysis of ETS differences in the Western region, ETS of NARR and ERA-Interim are almost the same as GEM when thresholds are lower than 5mm.
- In the analysis of ETS differences in the Prairie region, the corrected NARR is not recommended at thresholds between 1mm and 5mm.
- In the analysis of ETS differences in the Central region, ERA-Interim is closer to GEM than NARR at lower thresholds ($\leq 5mm$) in Winter.

After application of the CaPA assimilation process, NARR and ERA-Interim not only show improvements but also become close to each other, compared to before the CaPA process. Original and corrected reanalysis datasets have only small differences after the CaPA process. Accordingly, original and corrected datasets are not really separated in the analysis description in this chapter. Both NARR and ERA-Interim become similar after the CaPA process, therefore, both of them are equally recommended as alternatives to GEM. However, there are two points that need to be emphasized: higher skill in ERA-Interim than NARR in Atlantic region, and recommendation not to use corrected NARR at thresholds between 1mm and 5mm in the Prairie region due to the relatively large skill gap with GEM.

Chapter 7

Conclusions and Recommendations

The Canadian Precipitation Analysis (CaPA) is considered the most accurate gridded data in Canada. However, the short temporal coverage limits its use in studies such as hydrological modeling and climate change assessment. Reanalysis datasets, specifically NARR and ERA-interim, were used in this study in place of GEM as background field in the CaPA framework. The results show that prior bias correction improves the quality of the datasets overall, but the difference between the original and the bias-corrected datasets becomes notably smaller after the CaPA assimilation process. Results in this study suggest that comprehensive bias correction of the background field is not a strict requirement to obtain useful results, at least if biases are of the order of magnitude found in NARR and ERA-Interim. Summarized conclusions and suggestions for future work are presented below.

- Based on the pre-assessment of reanalysis datasets (i.e. prior to CaPA assimilation), ERA-Interim shows slightly better performance than NARR overall. This study is aiming to find alternative datasets of GEM for the extension of the CaPA product and found that GEM is better than NARR and ERA-Interim. Therefore, this study relate the performance of the reanalysis datasets to GEM. Both ERA-Interim and NARR are similar in terms of forecasting skill, but ERA-Interim is closer to GEM than NARR, especially in the Atlantic region. In addition, both NARR and ERA-Interim show some gaps in FBI-1 from GEM in Spring and Winter.

Other datasets such as JRA-25, JRA-55, ERA-20C Land, ERA-SAT, CERA-20C, and CERA-SAT are suggested to use in future studies. Those are the latest models available with high resolution as well as longer period coverage than NARR and ERA-Interim. Their resolution ranges from 25km up to 40km and the time frame is back to 1900.

- The bias-correction method used in this study is a simple multiplicative shift scheme combined with thin plate spline interpolation for full spatial coverage. The bias correction affects NARR more than ERA-Interim. ERA-Interim remains similar after bias correction, but sometimes actually shows slightly less good performance. Corrected NARR datasets are recommended in the Atlantic region since the level of improvement toward GEM is the highest here. However, the improvement in NARR from bias-correction is founded particularly in the Prairie region. In addition, the gaps between NARR/ERA-Interim and GEM are reduced after bias correction in Spring and Winter over Prairie region.

When bias correcting using the multiplicative shift method, using long term monthly means from other datasets is recommended for calculation of factors. This study calculates monthly means based on the study period (10yr) but it would be more accurate if the period were longer. There are datasets such as CANGRID, CRU (Climate Research Unit), and GPCC(Global Precipitation Climatology Centre) providing monthly means based on reference periods longer than ten years.

- After the CaPA assimilation process, both the original and bias-corrected reanalysis datasets are improved and close to each other. As a result, both of them are recommended as alternatives to CaPA. However, there is higher skill in ERA-Interim than in NARR in the Atlantic region. In addition, it is not recommended to use the corrected NARR after CaPA assimilation at thresholds between 1mm and 5mm in the Prairie region due to the relatively large skill gap with GEM.

  Studies of hydrological modelling using the reanalysis datasets after CaPA assimilation is recommended for the future. Through the analysis, it would be more clear how useful alternative datasets are for practical applications in water resources.

- The original objective of this study was to extend the CaPA product back in time. This study chose the way using alternative background datasets in order to extend the temporal coverage. The results of this study show various performance depending on regions and seasons, however it overall has several advantages. Based on the results in this study for the 10 years, extension of the CaPA product back in time (up to 1980) is suggested for future work. The results can also be used as a reference when the interest is in extending CaPA prediction back in time, especially for specific region and season.

Bibliography

Adler, R. F., Kidd, C., Petty, G., Morissey, M., and Goodman, H. M. (2001). Intercomparison of global precipitation products: The third precipitation intercomparison project (pip-3). *Bulletin of the American Meteorological Society*, 82(7):1377–1396.

Bao, X. and Zhang, F. (2013). Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the tibetan plateau. *Journal of Climate*, 26(1):206–214.

Benyahya, L., Gachon, P., St-Hilaire, A., and Laprise, R. (2014). Frequency analysis of seasonal extreme precipitation in southern Quebec (Canada): an evaluation of regional climate model simulation with respect to two gridded datasets. *Hydrology Research*, 45(1):115–133.

Betts, A. K., Köhler, M., and Zhang, Y. (2009). Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *Journal of Geophysical Research: Atmospheres*, 114(D2).

Bosilovich, M. G., Chen, J., Robertson, F. R., and Adler, R. F. (2008). Evaluation of global precipitation in reanalyses. *Journal of Applied Meteorology and Climatology*, 47(9):2279–2299.

Choi, W., Kim, S. J., Rasmussen, P. F., and Moore, A. R. (2009a). Use of the North American Regional Reanalysis for hydrological modelling in Manitoba. *Canadian Water Resources Journal*, 34(1):17–36.

Choi, W., Rasmussen, P. F., Moore, A. R., SungJoon, K., et al. (2009b). Simulating streamflow response to climate scenarios in central Canada using a simple statistical downscaling method. *Climate Research*, 40(1):89–102.

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A. (1998a). The operational CMC-MRB global environmental multiscale (GEM) model. Part ii: Results. *Monthly Weather Review*, 126(6):1397–1418.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A. (1998b). The operational CMC-MRB global environmental multiscale (GEM) model. Part i: Design considerations and formulation. *Monthly Weather Review*, 126(6):1373–1395.

Daley, R. (1993). *Atmospheric data analysis*. Cambridge university press.

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., and Bosilovich, M. G. (2012). Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal of Climate*, 25(6):1916–1944.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Dibike, Y. B. and Coulibaly, P. (2005). Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, 307(1):145–163.

Donato, G. and Belongie, S. (2002). Approximate thin plate spline mappings. In *European conference on computer vision*, pages 21–31. Springer.

Environment Canada (2013). Historical climate data. Retrieved 2013, from `http://climate.weather.gc.ca/`.

Fernandes, R., Korolevych, V., and Wang, S. (2007). Trends in land evapotranspiration over Canada for the period 1960-2000 based on in situ climate observations and a land surface model. *Journal of Hydrometeorology*, 8(5):1016–1030.

Fortin, V., Roy, G., and Donaldson, N. (2014). Assimilation of radar QPE in the Canadian Precipitation Analysis (CaPA). In *ASCE International Symposium on Weather Radar and Hydrology*.

Giorgi, F., Shields Brodeur, C., and Bates, G. T. (1994). Regional climate change scenarios over the United States produced with a nested regional climate model. *Journal of Climate*, 7(3):375–399.

Gutjahr, O. and Heinemann, G. (2013). Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology*, 114(3-4):511–529.

Hay, L. E. and Clark, M. (2003). Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology*, 282(1):56–75.

Hodges, K., Lee, R., and Bengtsson, L. (2011). A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *Journal of Climate*, 24(18):4888–4906.

Hopkinson, R. F., Hutchinson, M. F., McKenney, D. W., Milewska, E. J., and Papadopol, P. (2012). Optimizing input data for gridding climate normals for Canada. *Journal of Applied Meteorology and Climatology*, 51(8):1508–1518.

Hutchinson, M. (1995). Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems*, 9(4):385–403.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P. (2009). Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961-2003. *Journal of Applied Meteorology and Climatology*, 48(4):725–741.

Ines, A. V. and Hansen, J. W. (2006). Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology*, 138(1):44–53.

Keenan, T., Holland, G., Mantón, M., and Simpson, J. (1988). TRMM ground truth in a monsoon environment-Darwin, Australia.[tropical rainfall measuring mission. *Australian Meteorological Magazine*.

Latifovic, R. and Pouliot, D. (2007). Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record. *Remote Sensing of Environment*, 106(4):492–507.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Royal Meteorological Society, Quarterly Journal*, 112:1177–1194.

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

Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, C., Glazer, A., Leduc, A.-M., Méthot, A., Patoine, A., et al. (2006). The 15-km version of the Canadian regional forecast system. *Atmosphere-Ocean*, 44(2):133–149.

Mailhot, J., Sarrazin, R., Bilodeau, B., Brunet, N., and Pellerin, G. (1997). Development of the 35-km version of the Canadian regional forecast system. *Atmosphere-Ocean*, 35(1):1–28.

Mason, I. B. (2003). Binary events. *Forecast Verification: A Practitioners Guide in Atmospheric Science*, pages 37–76.

McKenney, D. W., Hutchinson, M. F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R. F., Price, D., and Owen, T. (2011). Customized spatial climate models for North America. *Bulletin of the American Meteorological Society*, 92(12):1611–1622.

Mekis, É. and Brown, R. (2010). Derivation of an adjustment factor map for the estimation of the water equivalent of snowfall from ruler measurements in Canada. *Atmosphere-Ocean*, 48(4):284–293.

Mekis, E. and Hogg, W. D. (1999). Rehabilitation and analysis of Canadian daily precipitation time series. *Atmosphere-Ocean*, 37(1):53–85.

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

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E. H., et al. (2006). North American regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360.

Murdock, T., Cannon, A., and Sobie, S. (2013). New statistical downscaling for Canada. In *AGU Fall Meeting Abstracts*, volume 1, page 1069.

National Center for Atmospheric Research (2013). The climate data guide: Simplistic overview of reanalysis data assimilation methods. Retrieved from https://climatedataguide.ucar.edu/climate-data/simplistic-overview-reanalysis-data-assimilation-methods.

Newman, M., Sardeshmukh, P. D., and Bergman, J. W. (2000). An assessment of the NCEP, NASA, and ECMWF reanalyses over the tropical west Pacific warm pool. *Bulletin of the American Meteorological Society*, 81(1):41.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1143–1170.

Rubel, F. and Rudolf, B. (2001). Global daily precipitation estimates proved over the European Alps. *Meteorologische Zeitschrift*, 10(5):407–418.

Rudolf, B., Hauschild, H., Rüth, W., and Schneider, U. (1996). Comparison of rain-gauge analyses, satellite-based precipitation estimates and forecast model results. *Advances in Space Research*, 18(7):53–62.

Rudolf, B. and Schneider, U. (2005). Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. In *Proc. Second Workshop of the Int. Precipitation Working Group*, pages 231–247.

Sugita, M. and Brutsaert, W. (1993). Comparison of land surface temperatures derived from satellite observations with ground truth during FIFE. *International Journal of Remote Sensing*, 14(9):1659–1676.

Szczypta, C., Calvet, J.-C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., and Meurey, C. (2011). Verification of the new ECMWF ERA-Interim reanalysis over France. *Hydrology and Earth System Sciences*, 15:647–666.

Terink, W., Hurkmans, R., Torfs, P., and Uijlenhoet, R. (2010). Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin. *Hydrology and Earth System Sciences Discussions*, 7(1):221–267.

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

Ward, E., Buytaert, W., Peaver, L., and Wheater, H. (2011). Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Advances in Water Resources*, 34(10):1222–1231.

Wilks, D. (2006). *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series. Academic Press.

Zhang, X., Vincent, L. A., Hogg, W., and Niitsoo, A. (2000). Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean*, 38(3):395–429.

Zhang, Y., Ohata, T., Yang, D., and Davaa, G. (2004). Bias correction of daily precipitation measurements for Mongolia. *Hydrological Processes*, 18(16):2991–3005.

Zhao, T. and Fu, C. (2006). Comparison of products from ERA-40, NCEP-2, and CRU with station data for summer precipitation over china. *Advances in Atmospheric sciences*, 23:593–604.

Appendix A

ETS and FBI (bin type)

There are two ways to calculate ETS and FBI-1. First way is 'bin type' which is based on model precipitation falling into each bin. Each bin is between two specific thresholds of precipitation. Second way is based on model precipitation larger than specific threshold of precipitation, which is used in this study. Overall ETS and FBI-1 of bin type are presented in this Appendix.

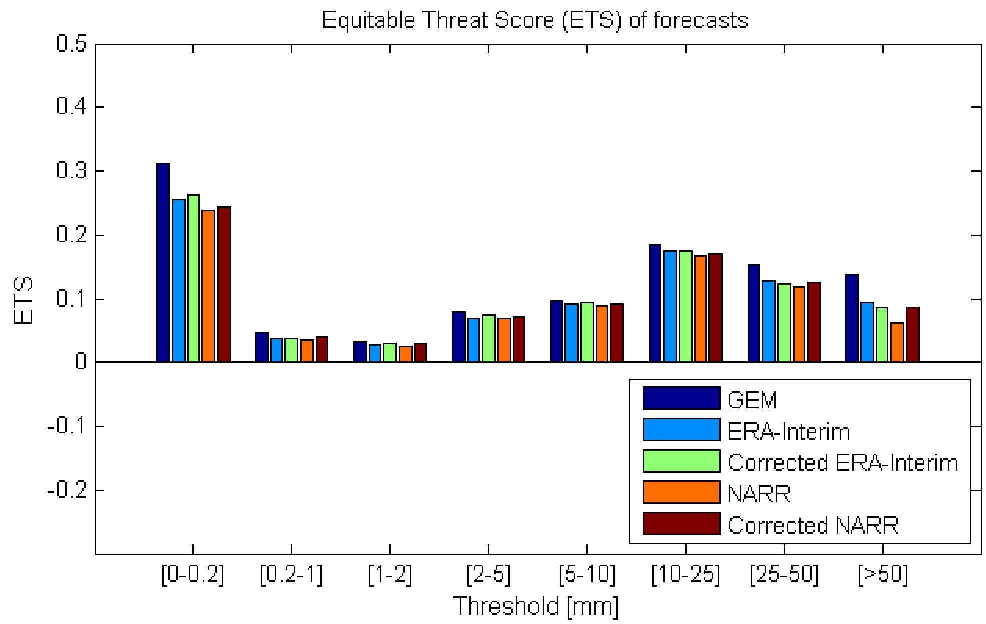


Figure A.1: Equitable Threat Score (ETS) results of reanalysis data sets

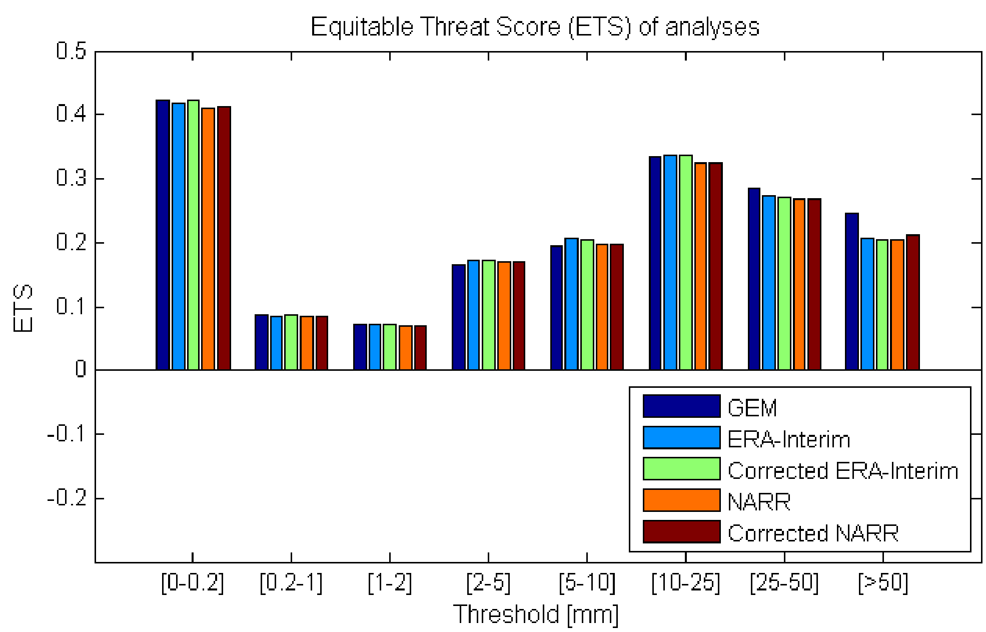


Figure A.2: Equitable Threat Score (ETS) results of reanalysis data sets after CaPA assimilation process

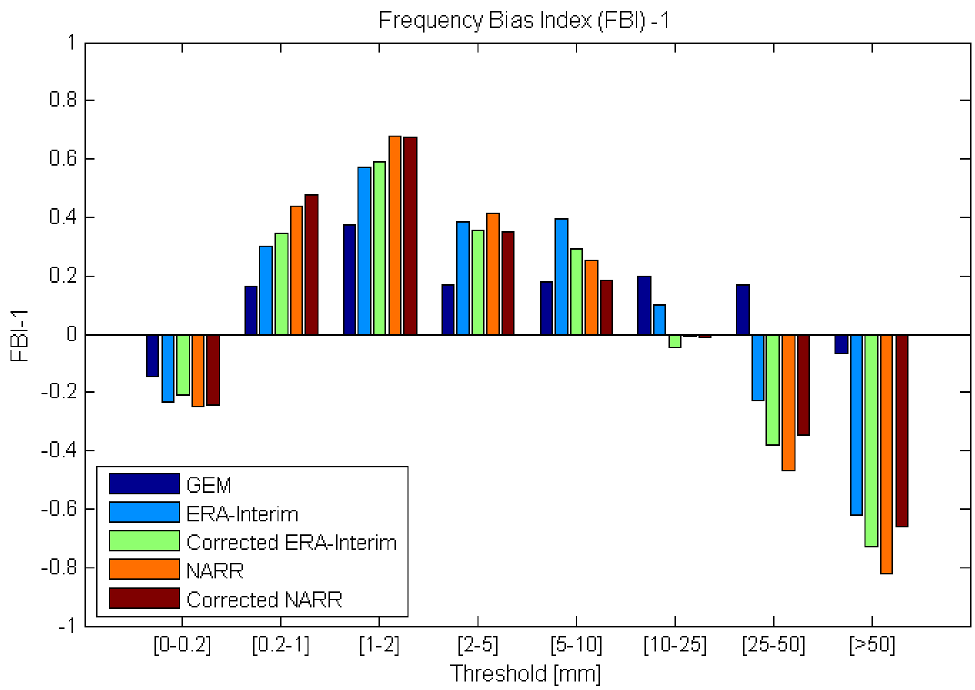


Figure A.3: Frequency Bias Index (FBI)-1 results of reanalysis data sets

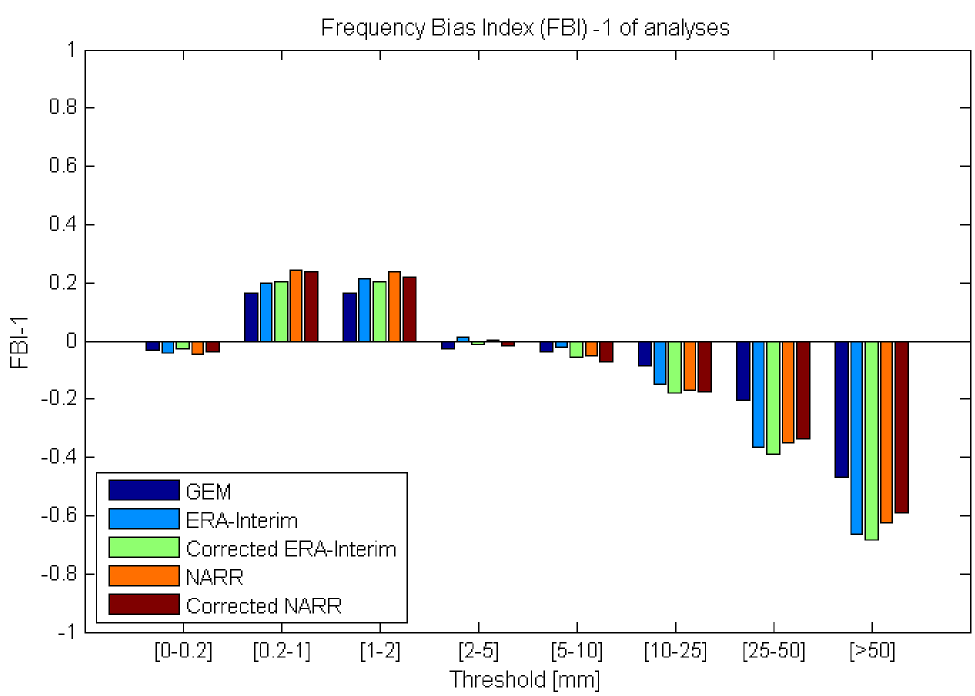


Figure A.4: Frequency Bias Index (FBI)-1 results of reanalysis data sets after CaPA assimilation process

Appendix B

Seasonal comparison between forecast and analysis of GEM, ERA-Interim, and NARR (from Pegasus)

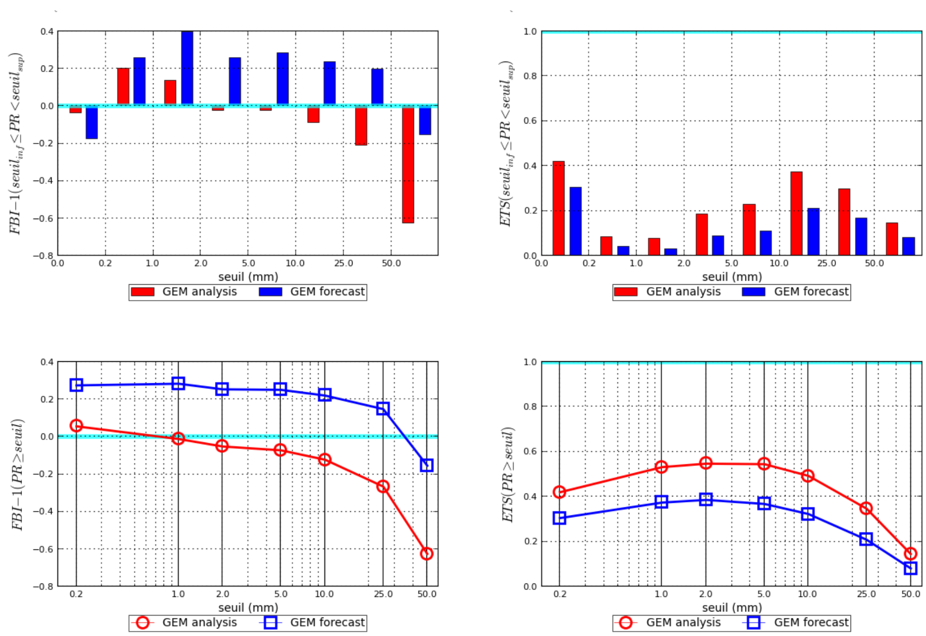


Figure B.1: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Spring

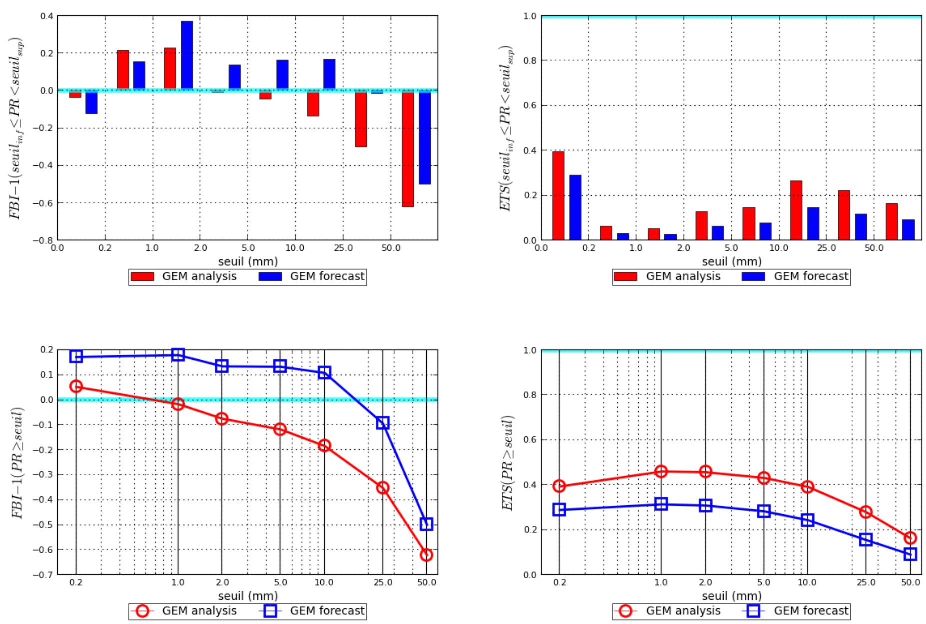


Figure B.2: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Summer

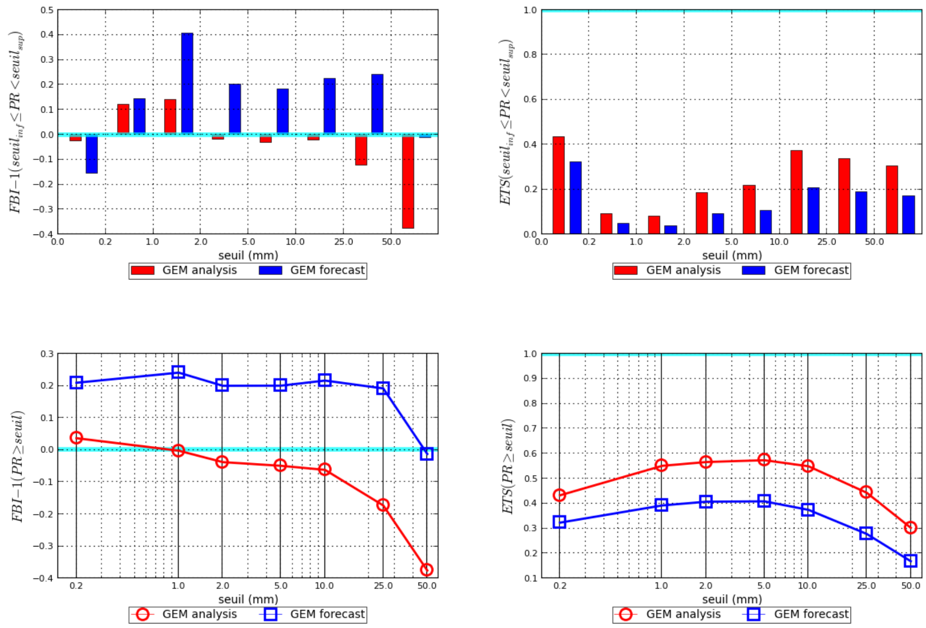


Figure B.3: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Fall

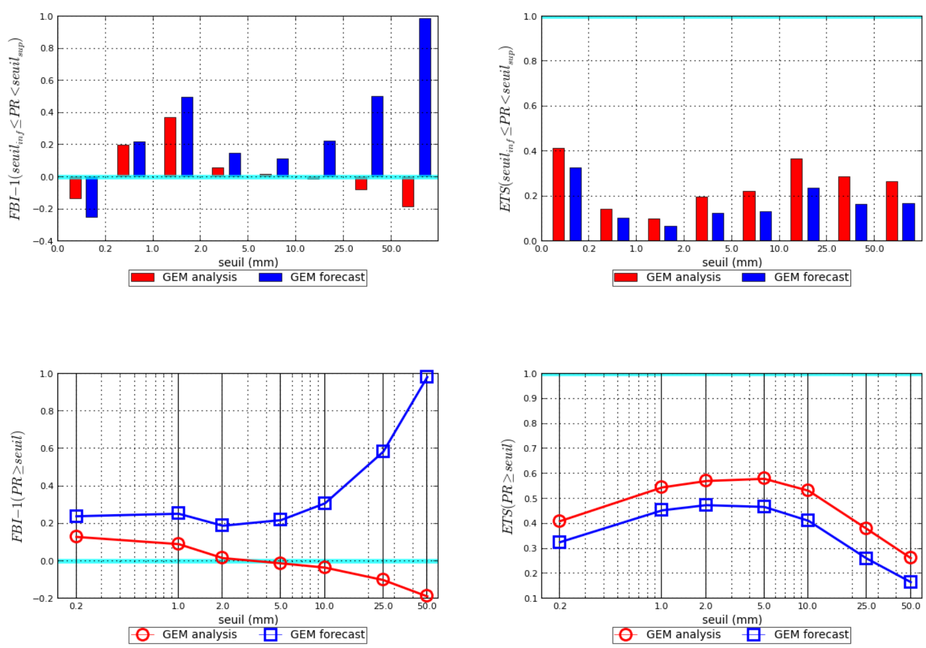


Figure B.4: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Winter

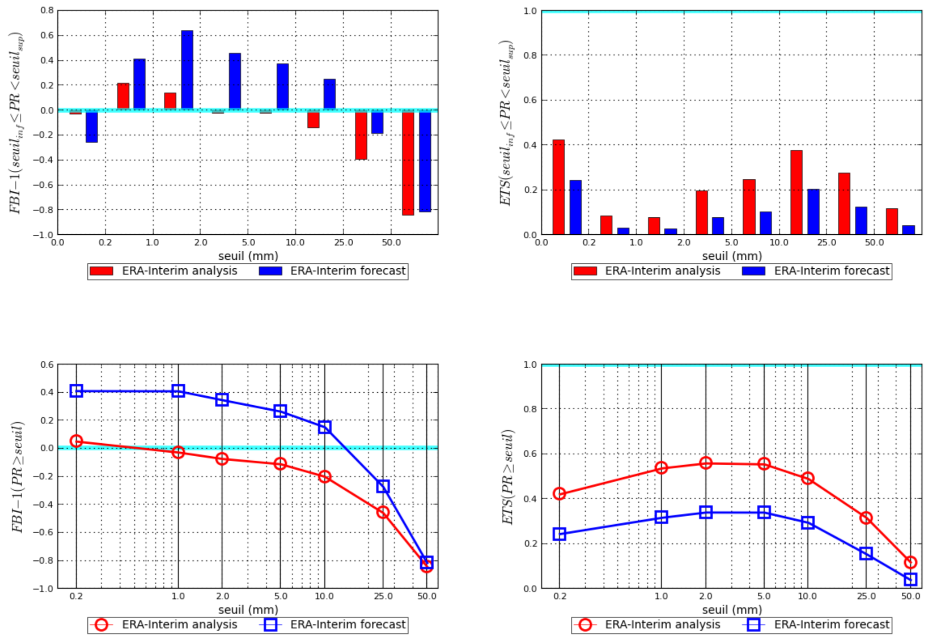


Figure B.5: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Spring

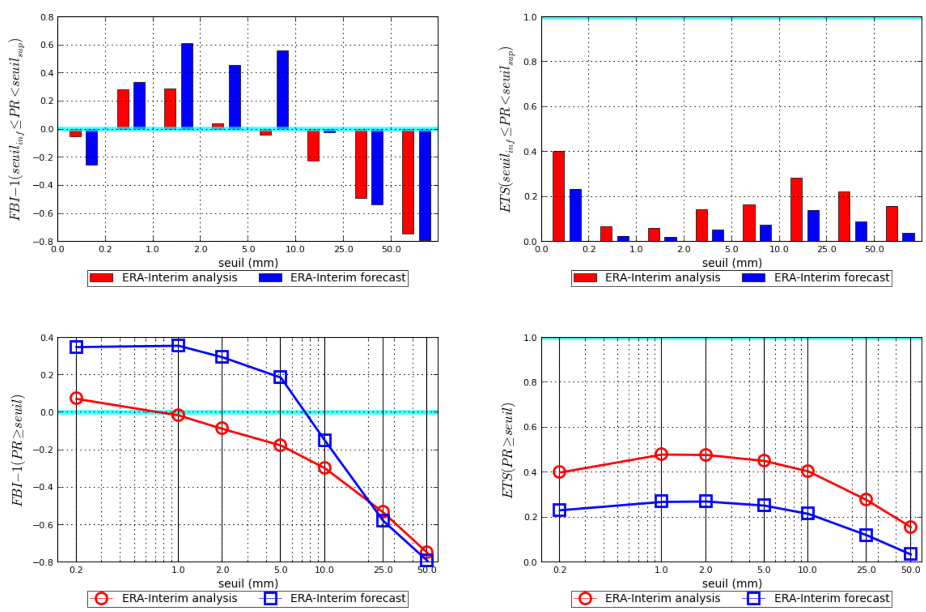


Figure B.6: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Summer

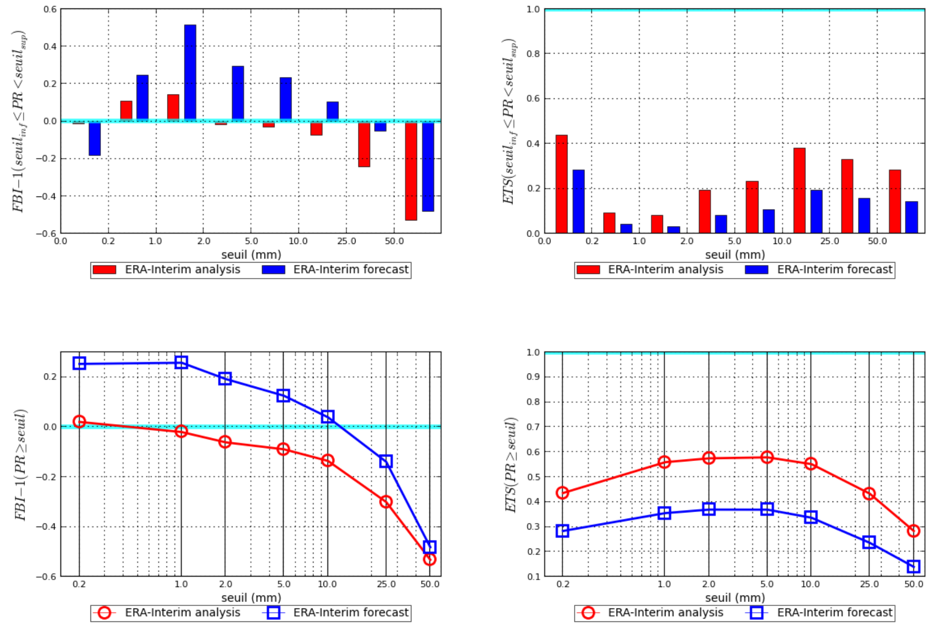


Figure B.7: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Fall

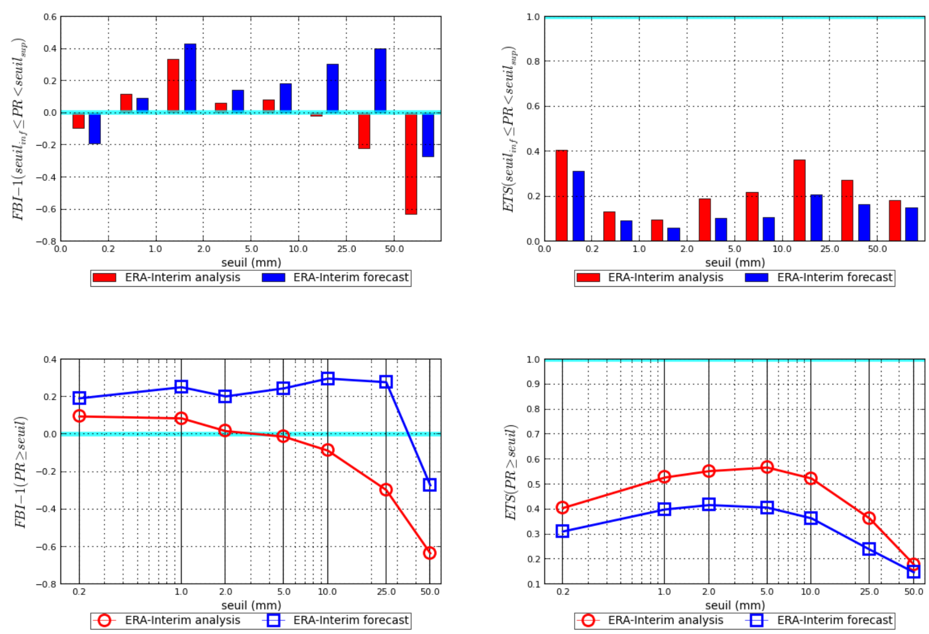


Figure B.8: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Winter

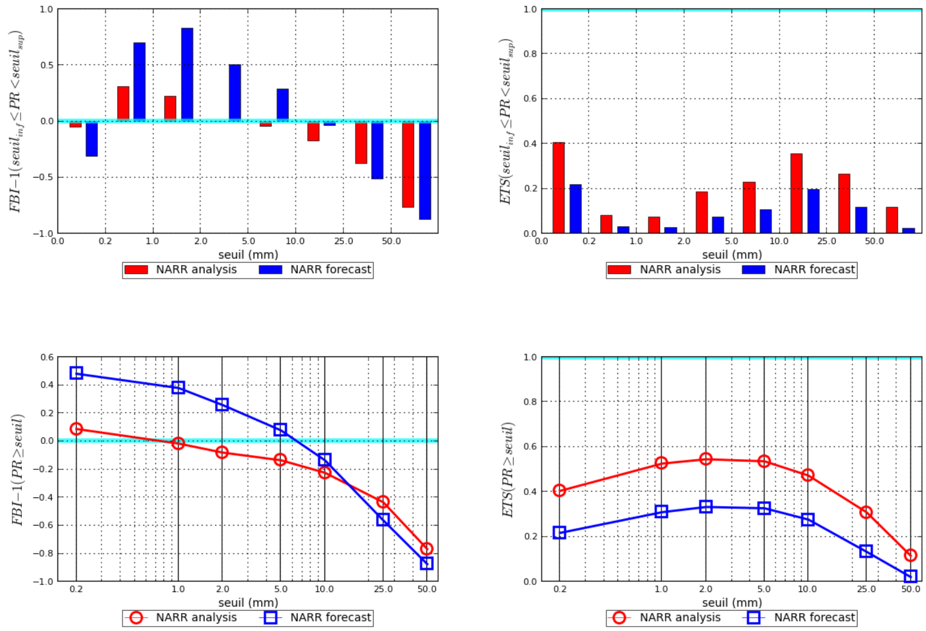


Figure B.9: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Spring

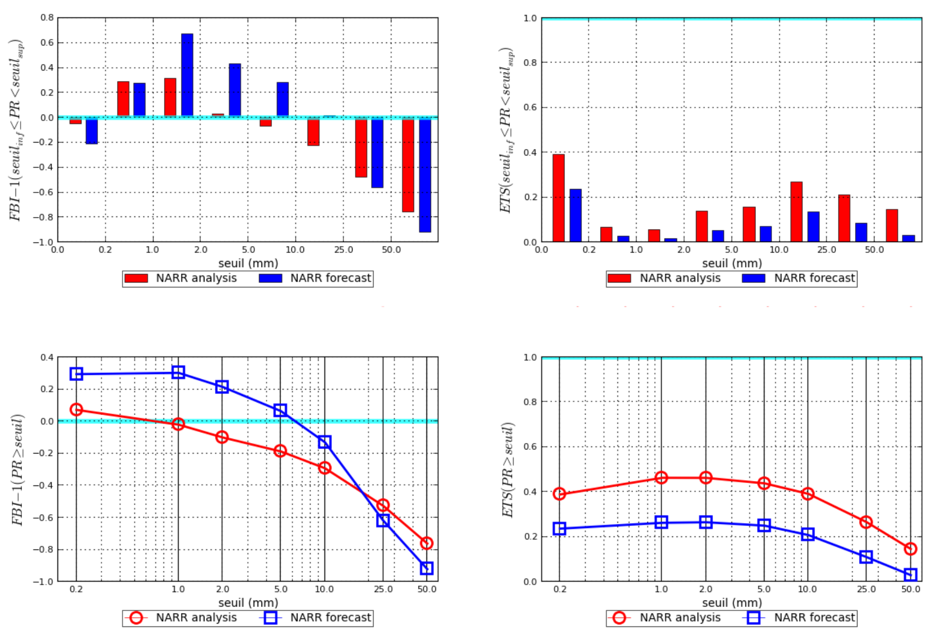


Figure B.10: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Summer

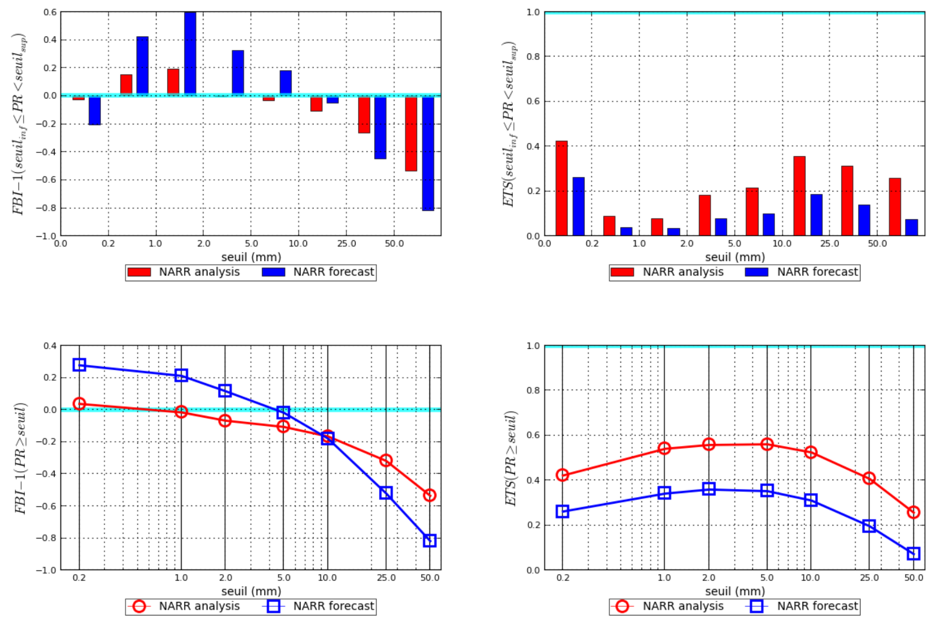


Figure B.11: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Fall

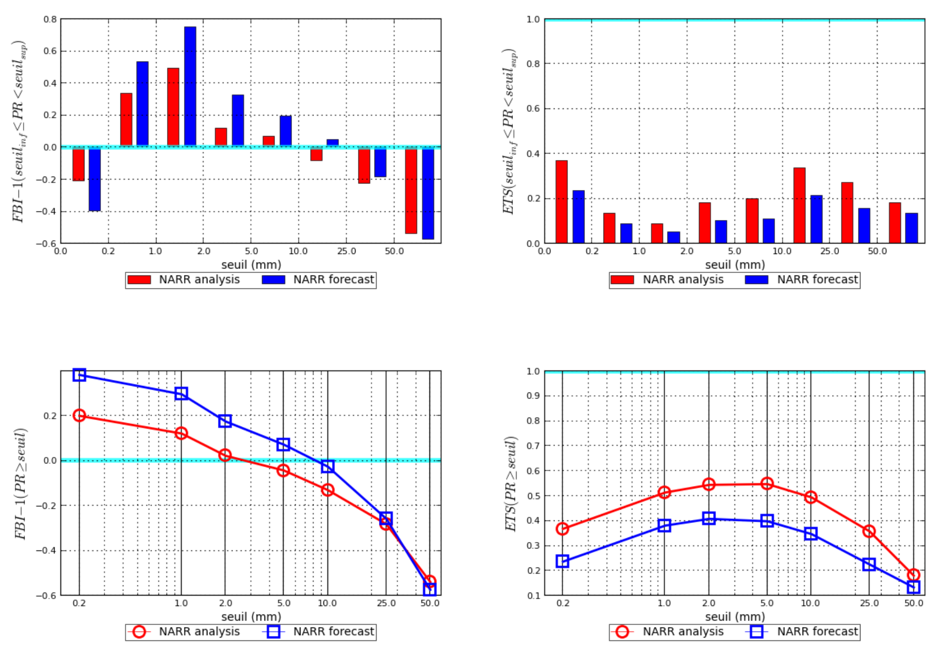


Figure B.12: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Winter

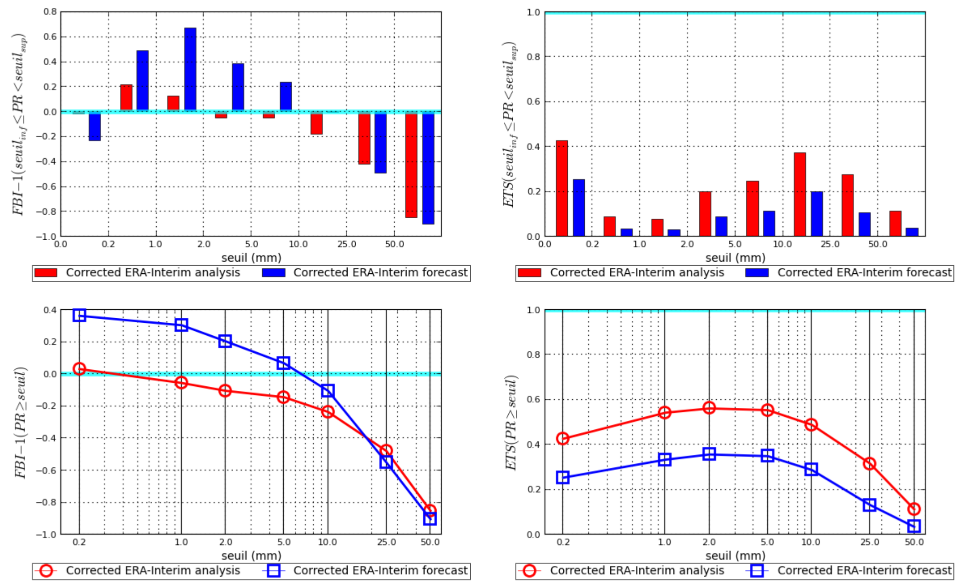


Figure B.13: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Spring

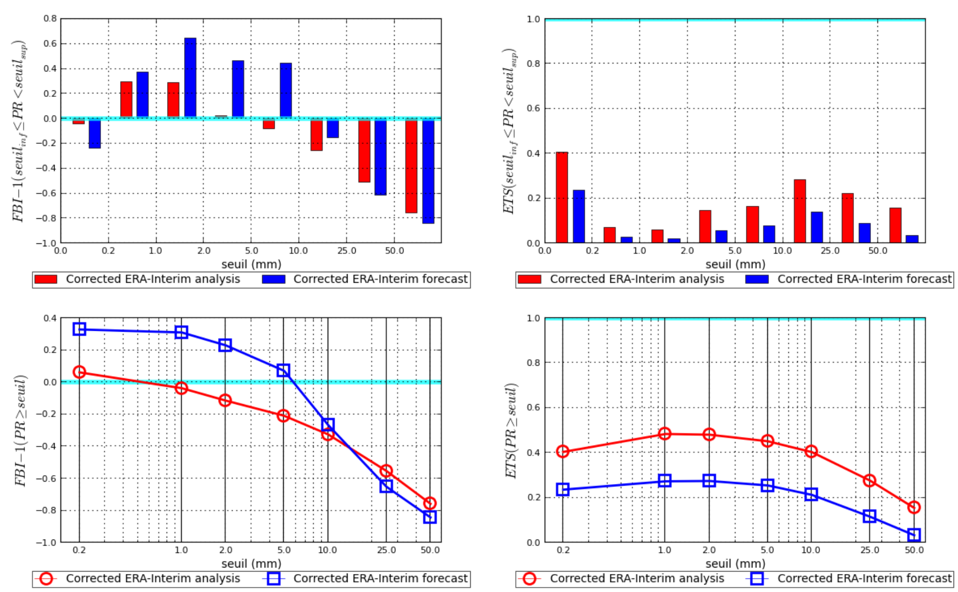


Figure B.14: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Summer

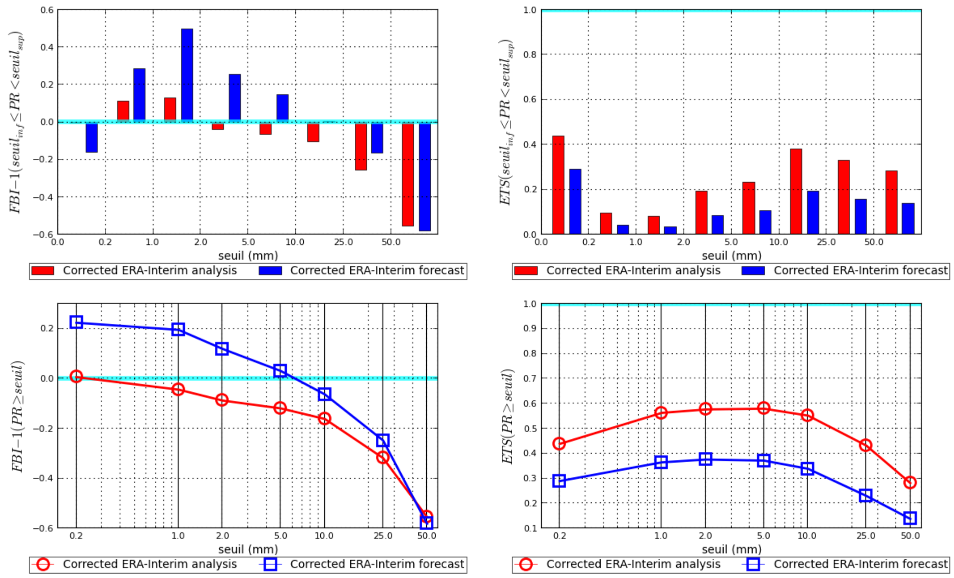


Figure B.15: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Fall

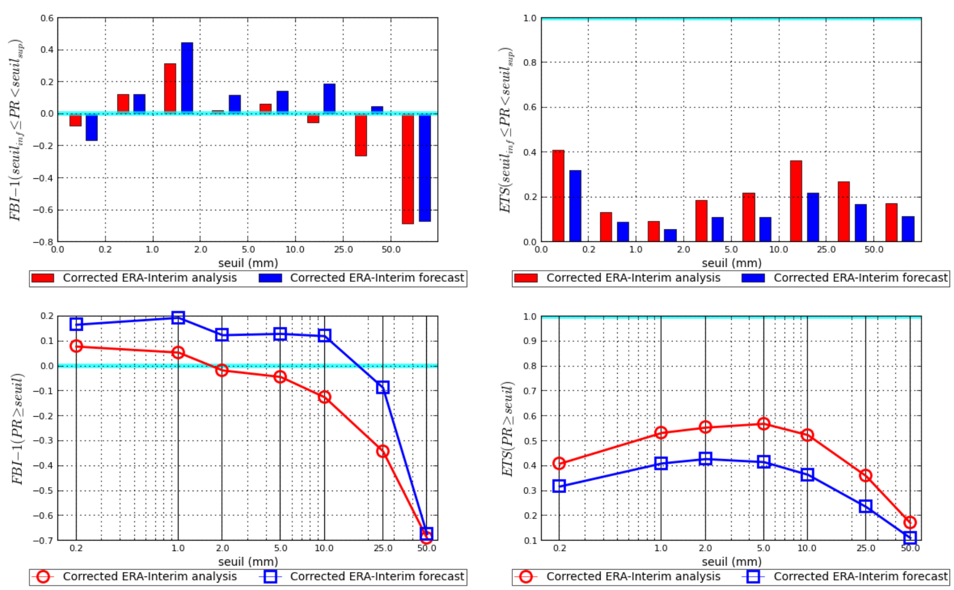


Figure B.16: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Winter

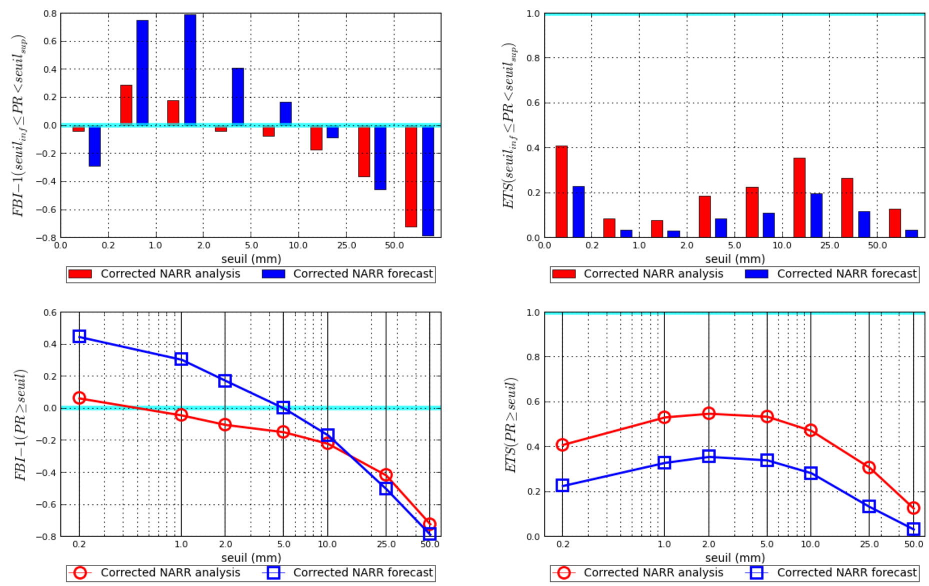


Figure B.17: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Spring

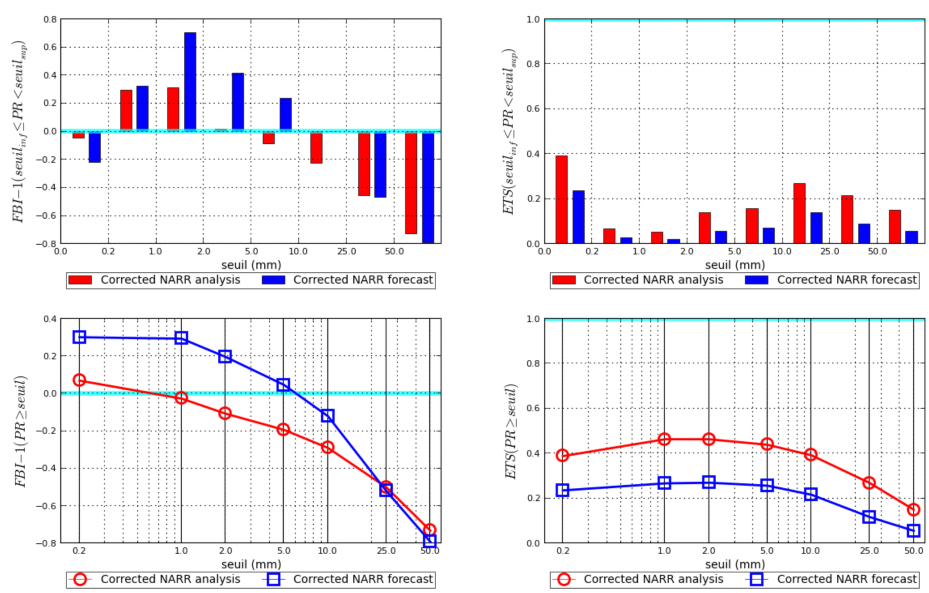


Figure B.18: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Summer

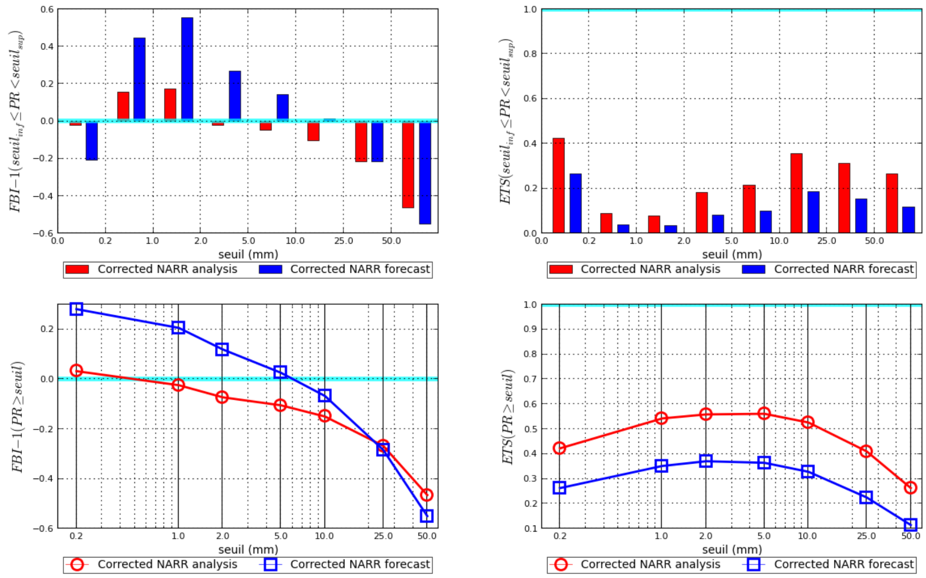


Figure B.19: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Fall

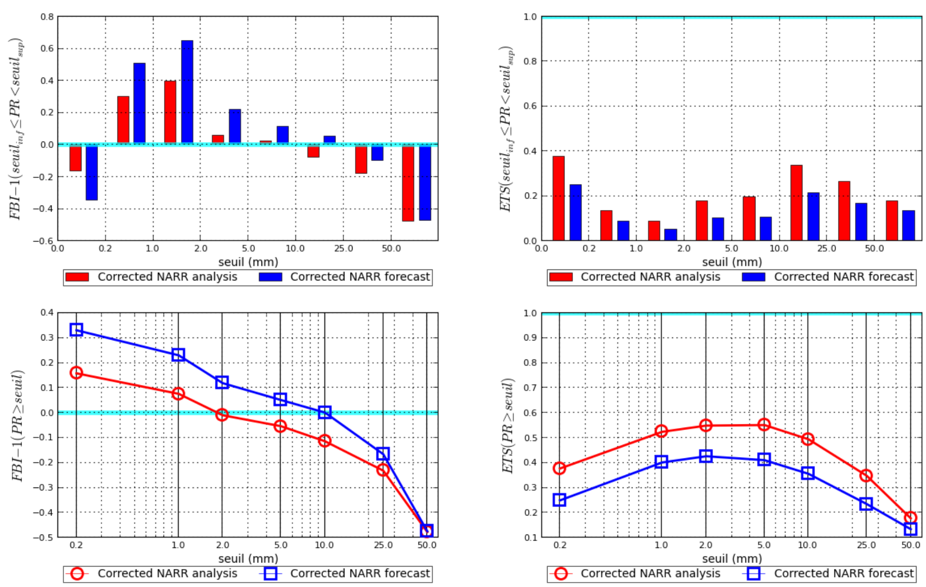


Figure B.20: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Winter

Appendix C

Regional comparison between forecast and analysis of GEM, ERA-Interim, and NARR (from Pegasus)

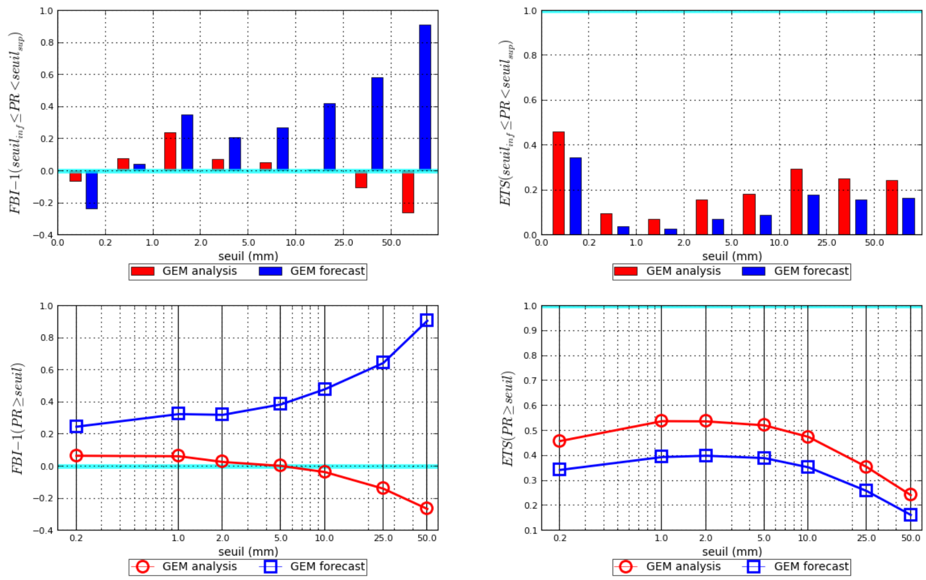


Figure C.1: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Western region

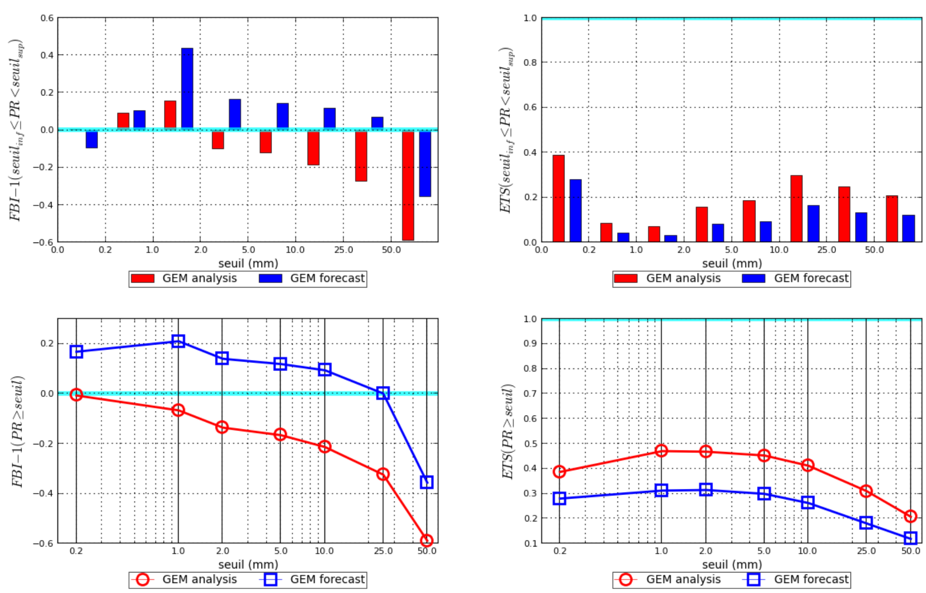


Figure C.2: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Prairie region

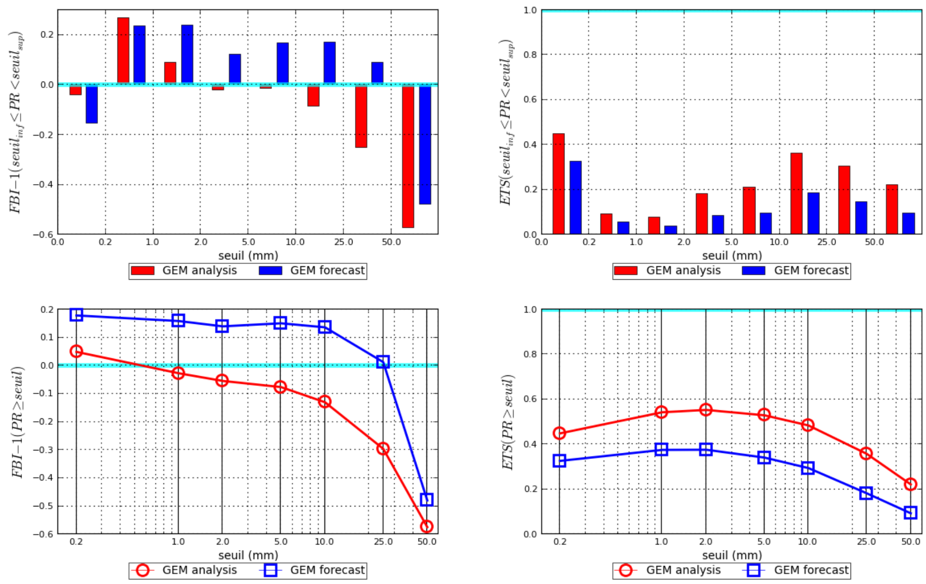


Figure C.3: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Central region

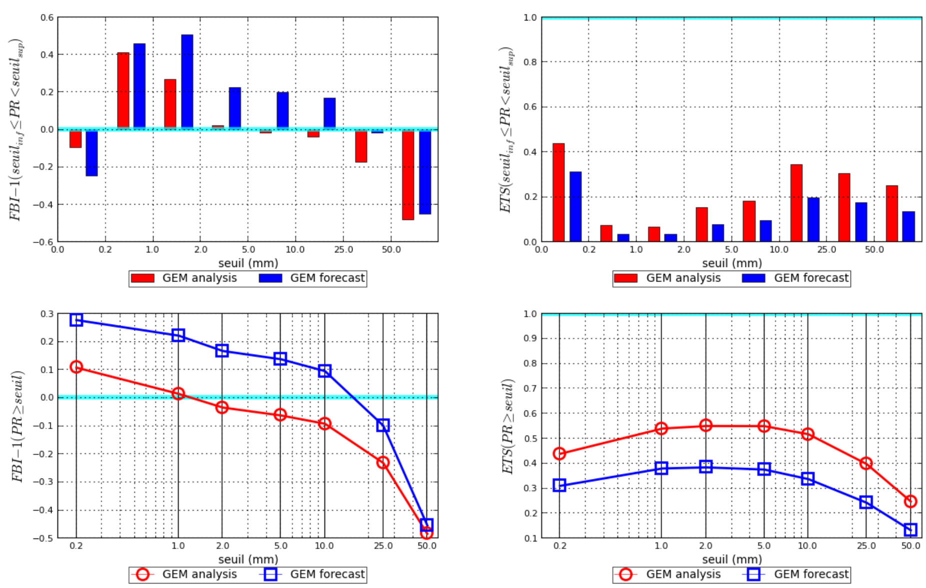


Figure C.4: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of GEM in Atlantic region

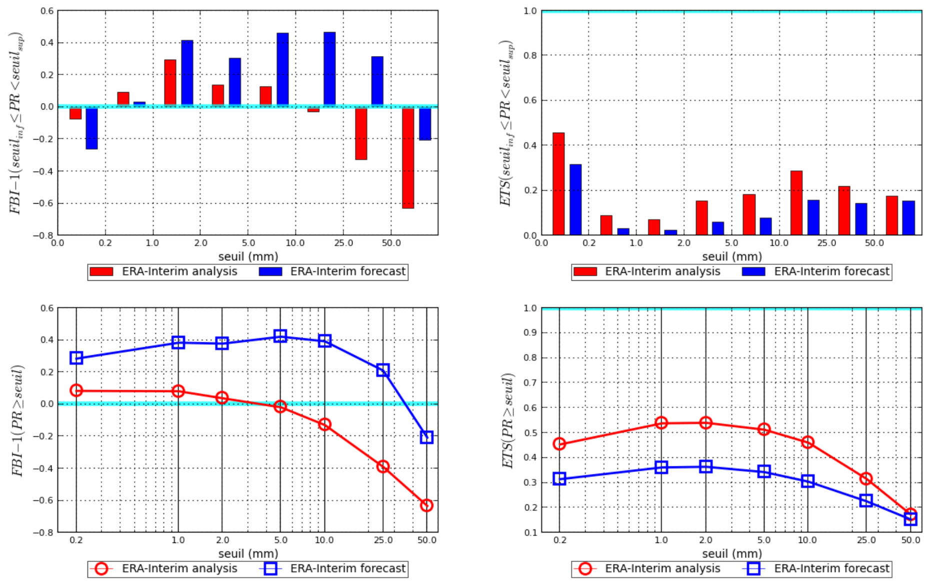


Figure C.5: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Western region

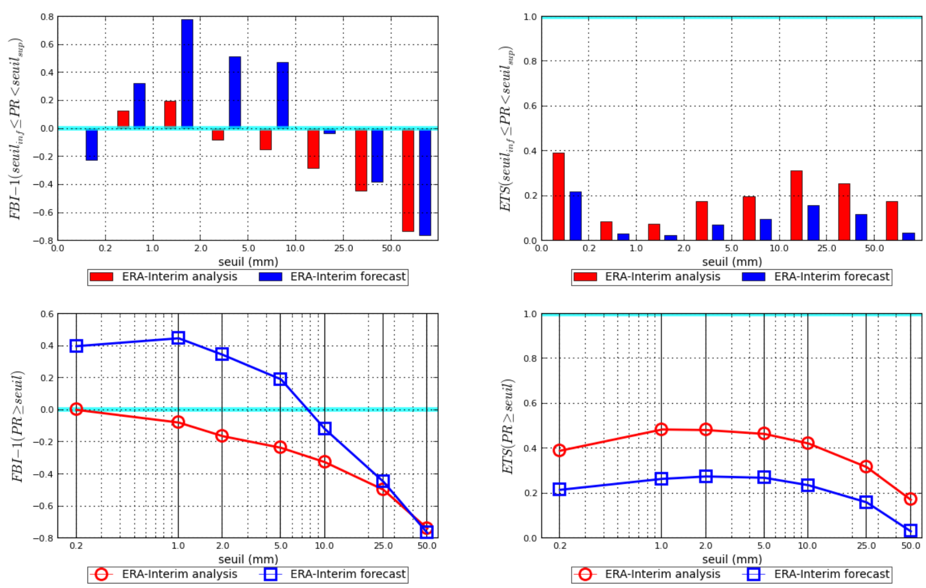


Figure C.6: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Prairie region

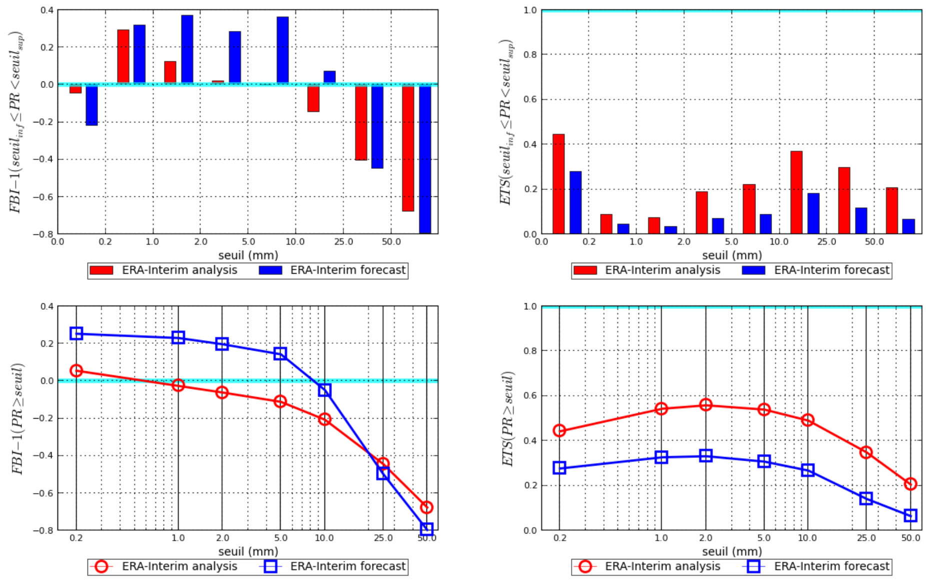


Figure C.7: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Central region

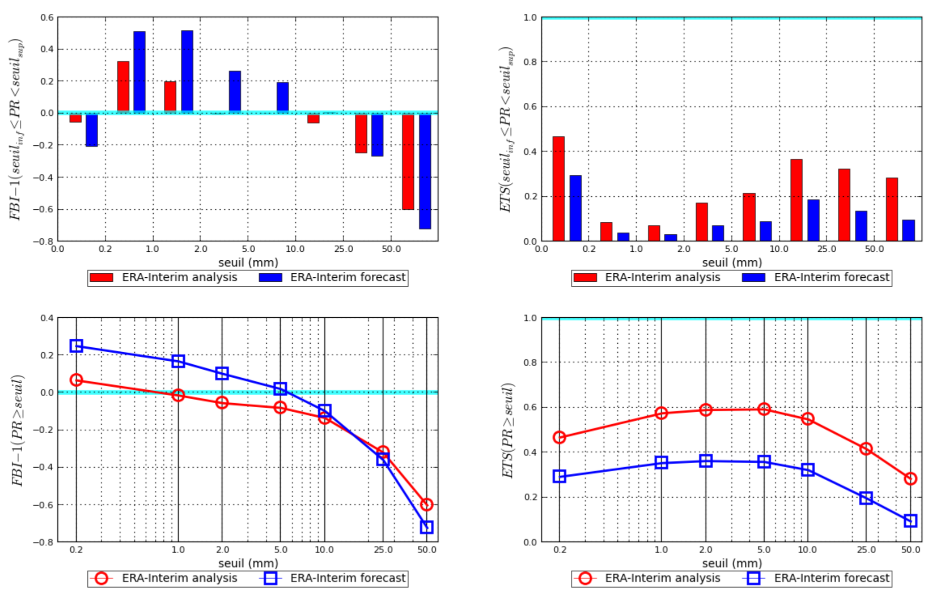


Figure C.8: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of ERA-Interim in Atlantic region

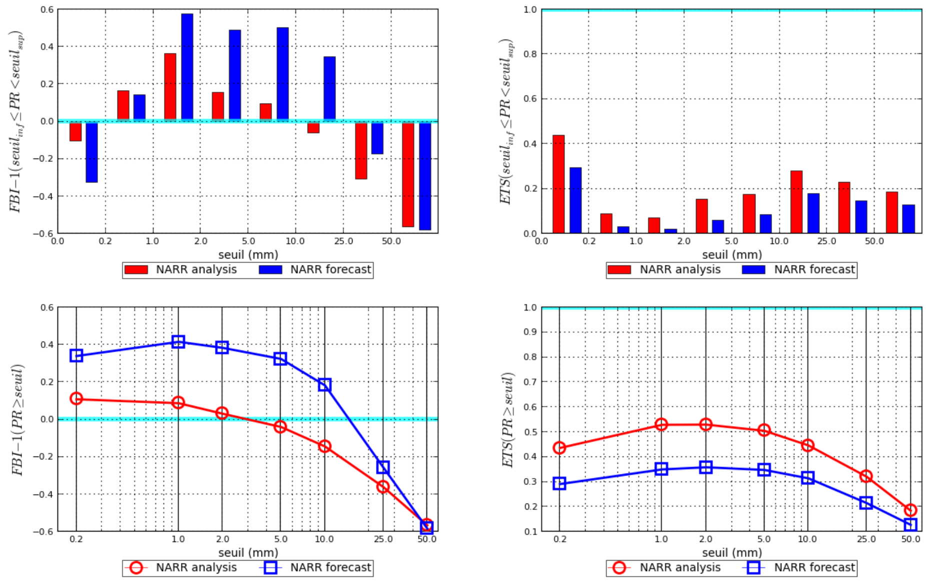


Figure C.9: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Western region

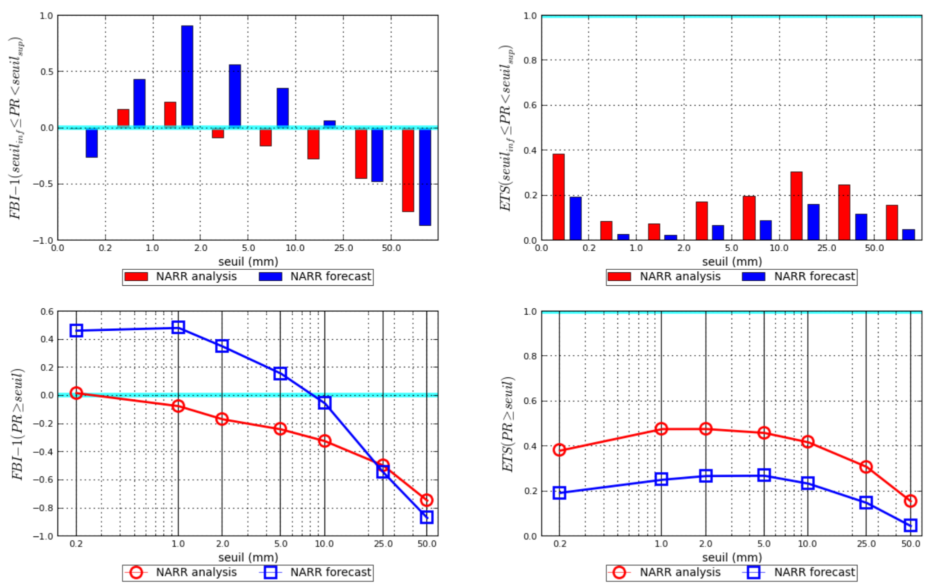


Figure C.10: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Prairie region

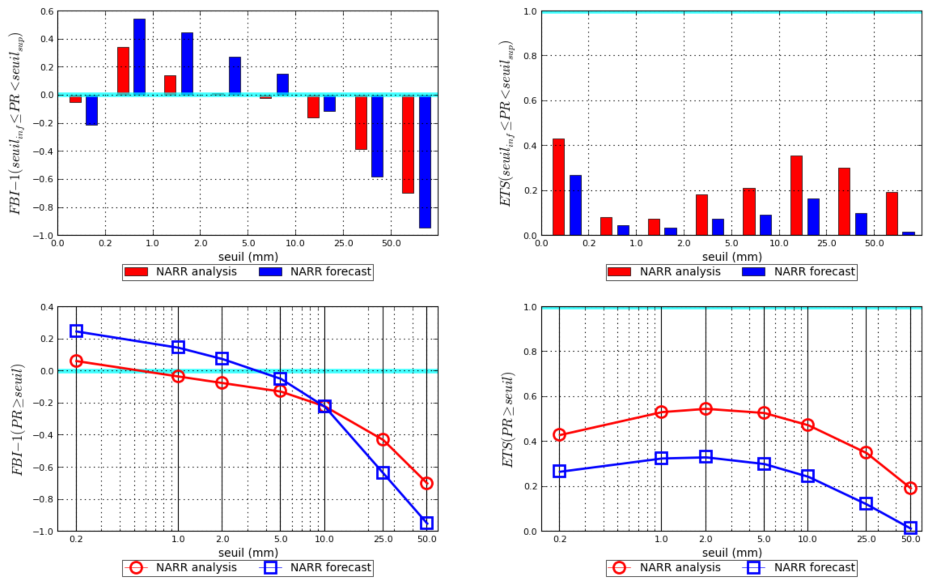


Figure C.11: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Central region

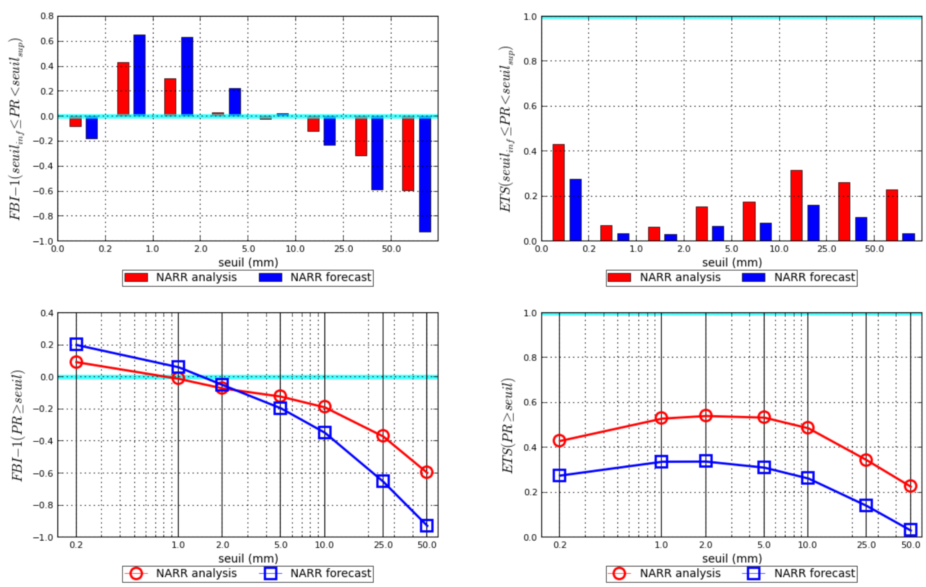


Figure C.12: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of NARR in Atlantic region

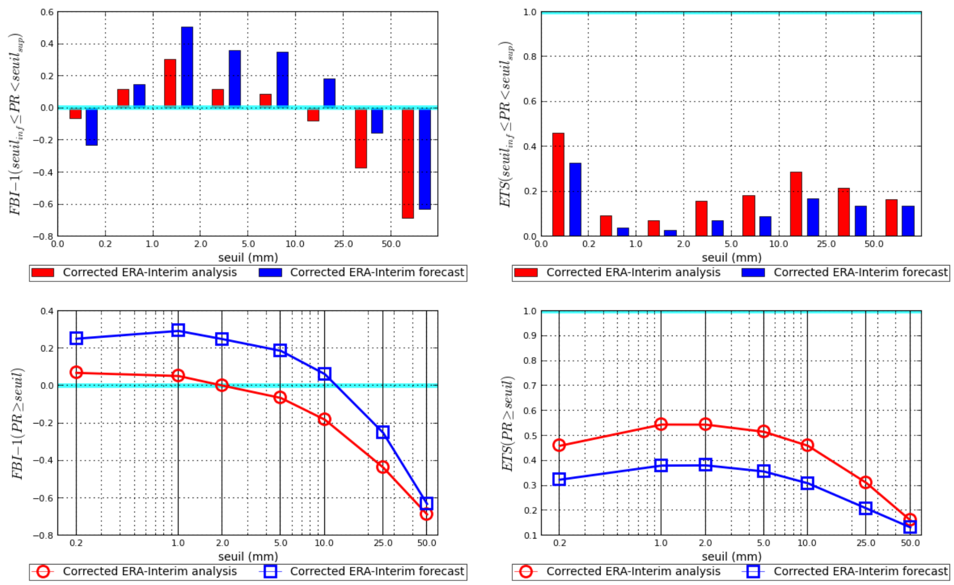


Figure C.13: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Western region

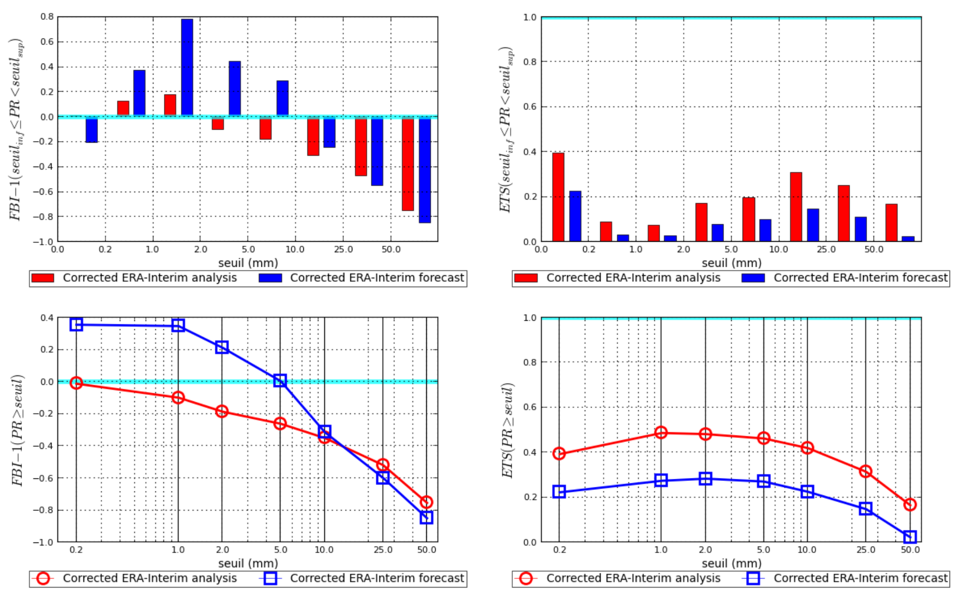


Figure C.14: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Prairie region

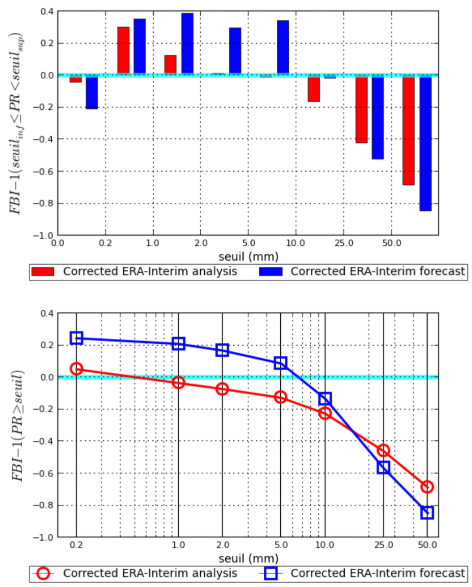


Figure C.15: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Central region

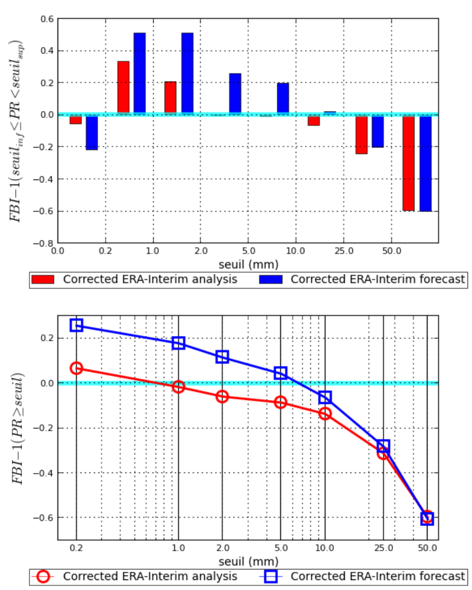


Figure C.16: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected ERA-Interim in Atlantic region

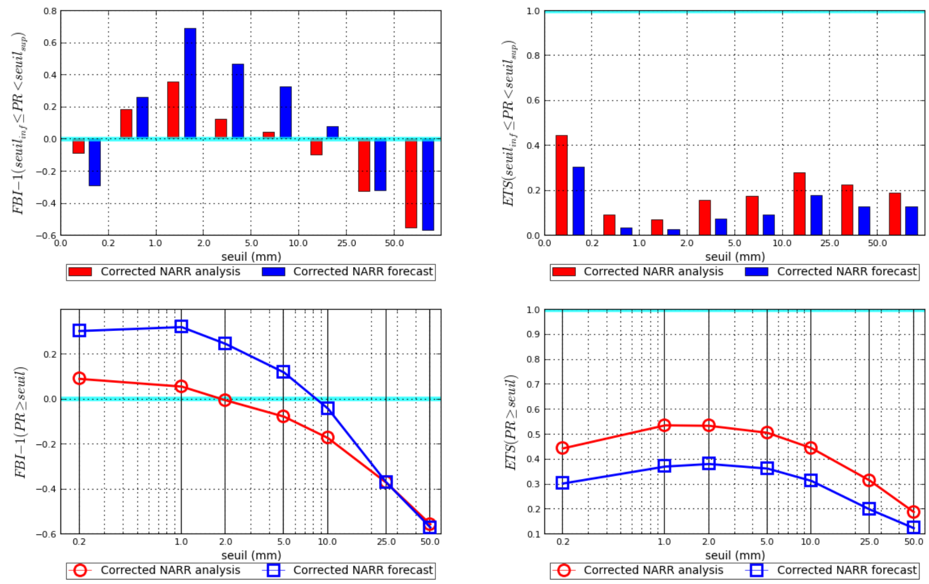


Figure C.17: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Western region

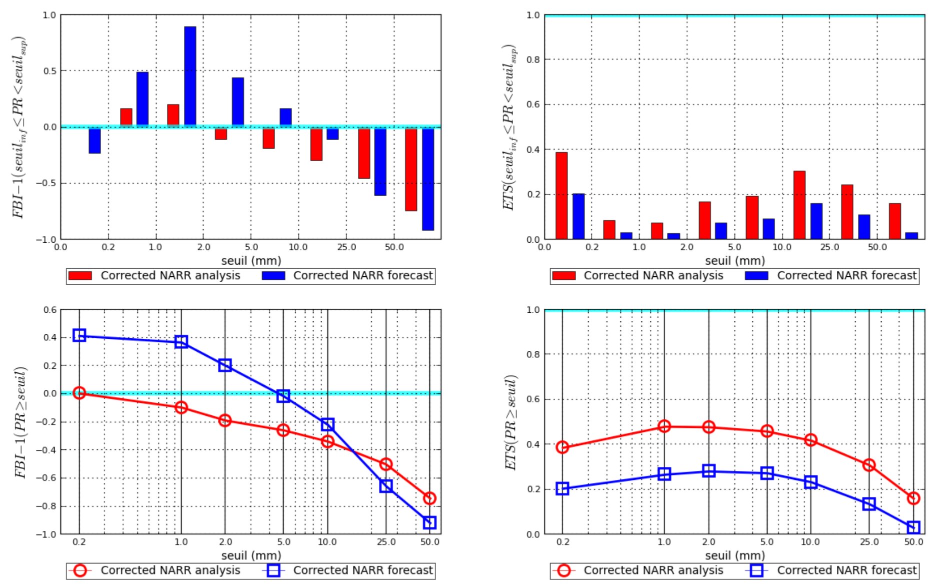


Figure C.18: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Prairie region

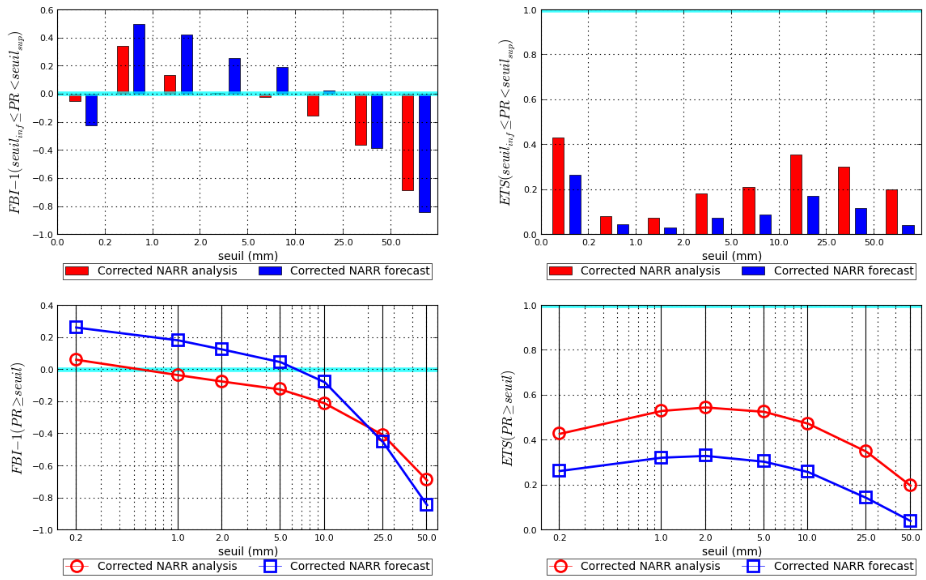


Figure C.19: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Central region

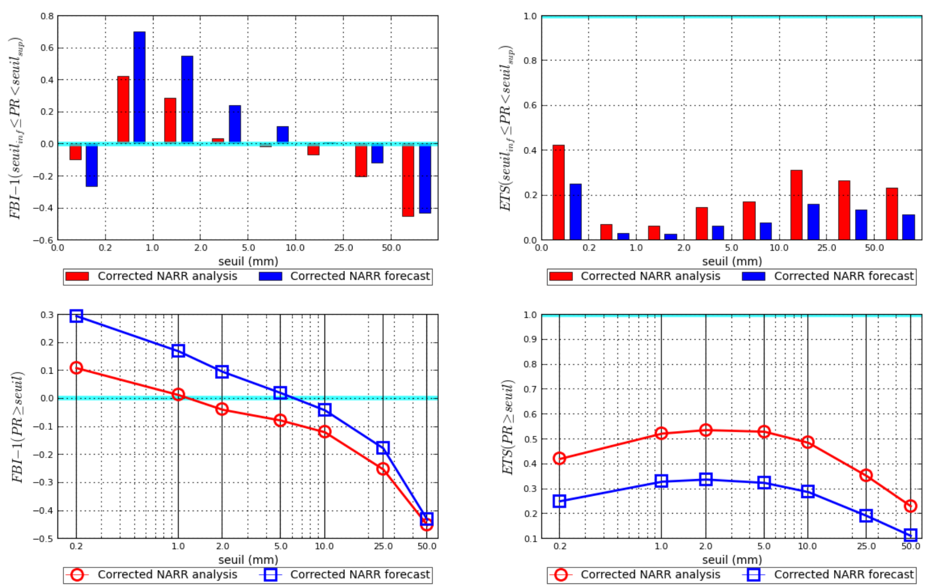


Figure C.20: Comparison of Equitable Threat Score (ETS) and Frequency Bias Index (FBI)-1 against Synoptic between analysis and forecast of Corrected NARR in Atlantic region

Appendix D

Bias correction ratio for each month

Bias correction ratio of each month in NARR and ERA-Interim is presented. Bias correction ratio is calculated using mean monthly precipitation of reanalysis data set and CDCD at each location of stations and used for bias correction. The detail is described in Section 3.2.

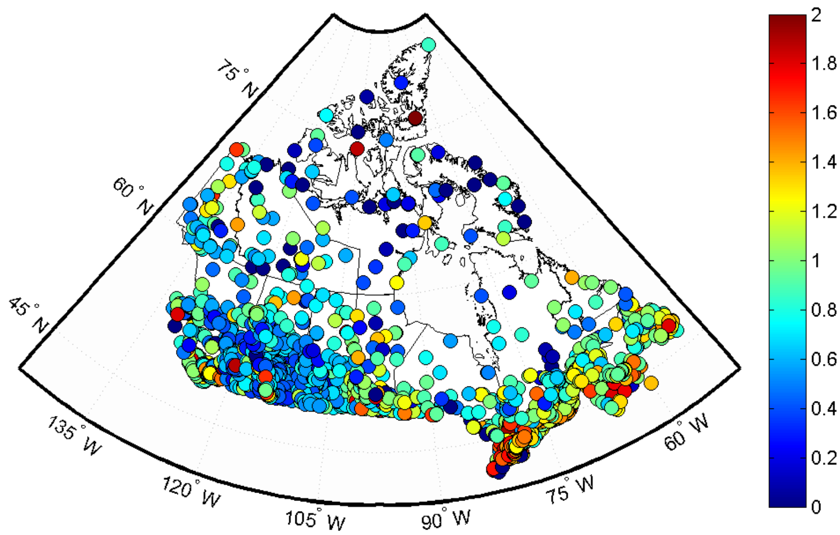


Figure D.1: Bias correction ratio of NARR in January

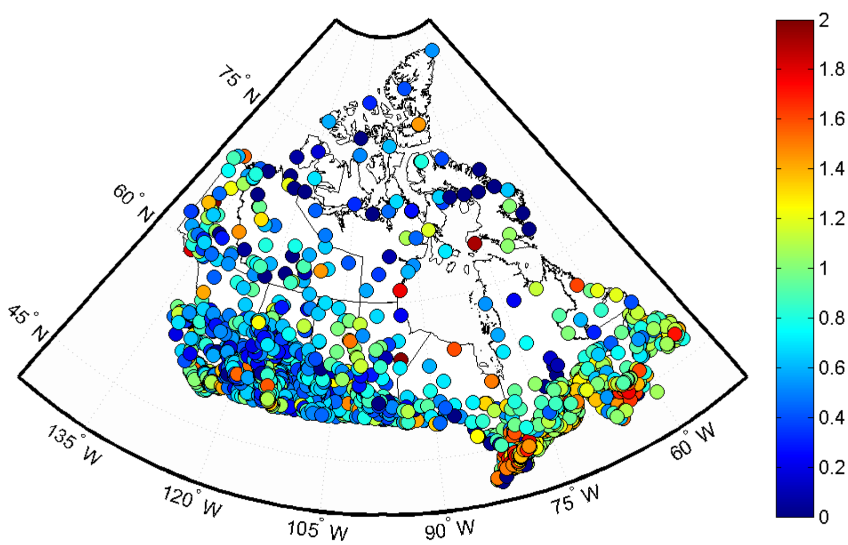


Figure D.2: Bias correction ratio of NARR in February

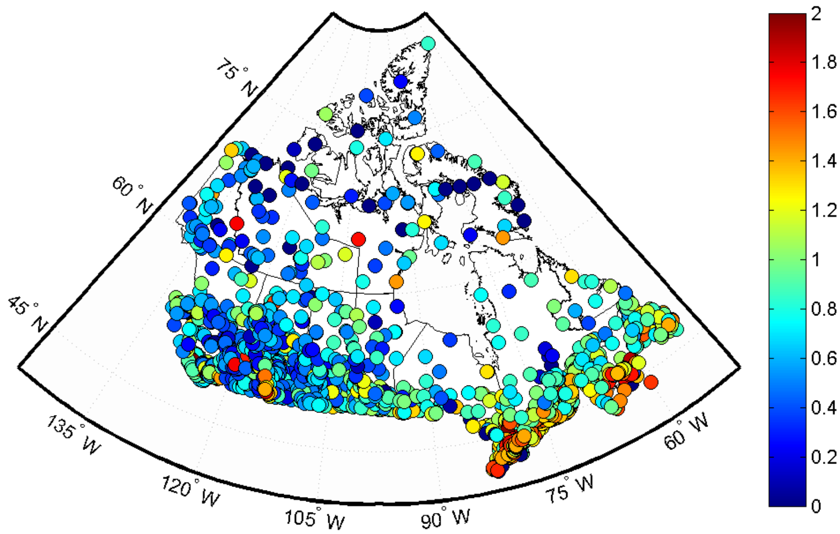


Figure D.3: Bias correction ratio of NARR in March

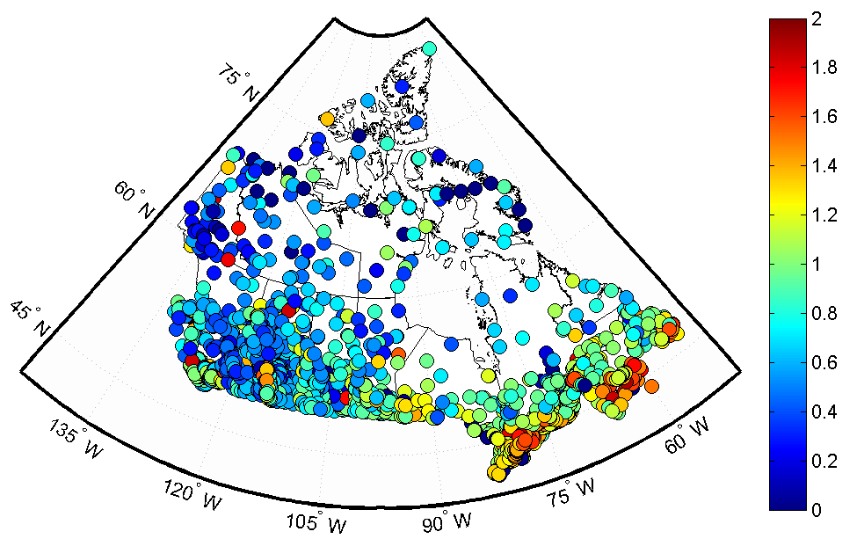


Figure D.4: Bias correction ratio of NARR in April

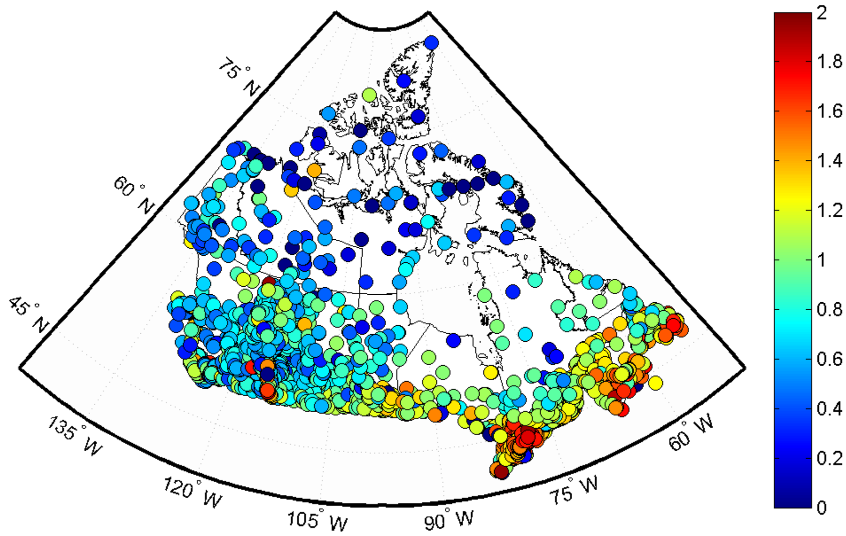


Figure D.5: Bias correction ratio of NARR in May

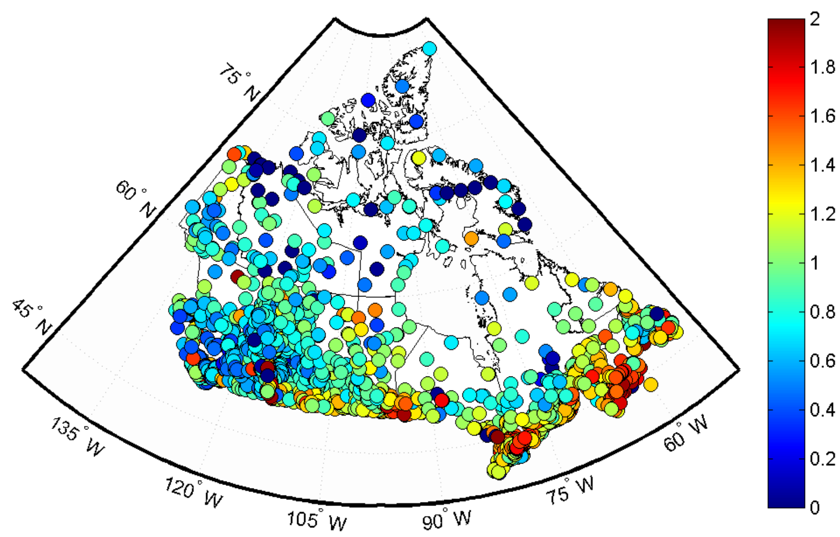


Figure D.6: Bias correction ratio of NARR in June

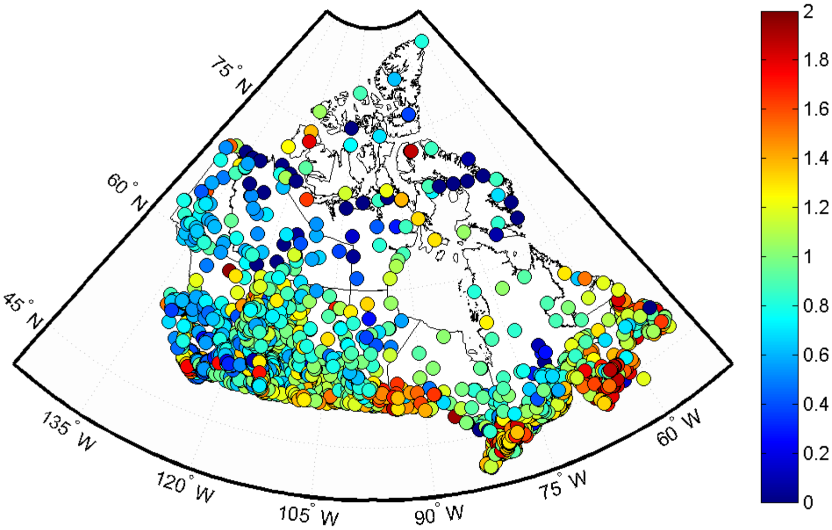


Figure D.7: Bias correction ratio of NARR in July

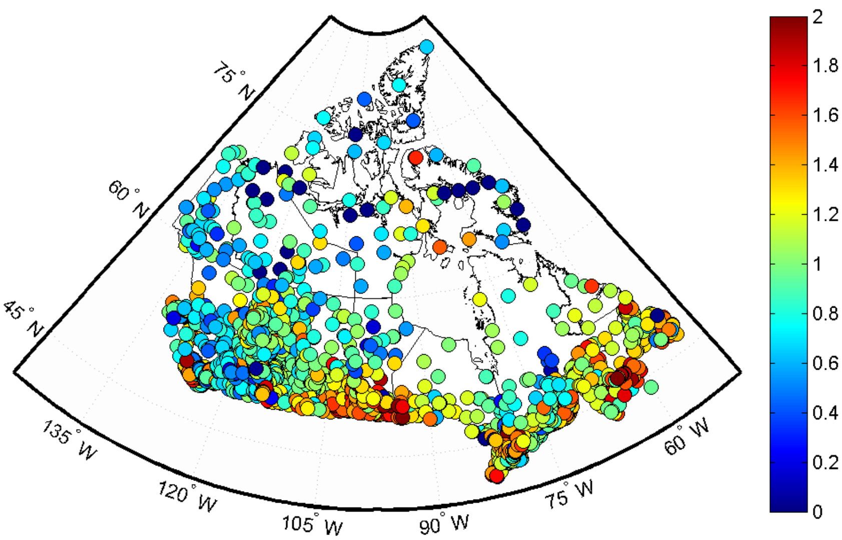


Figure D.8: Bias correction ratio of NARR in August

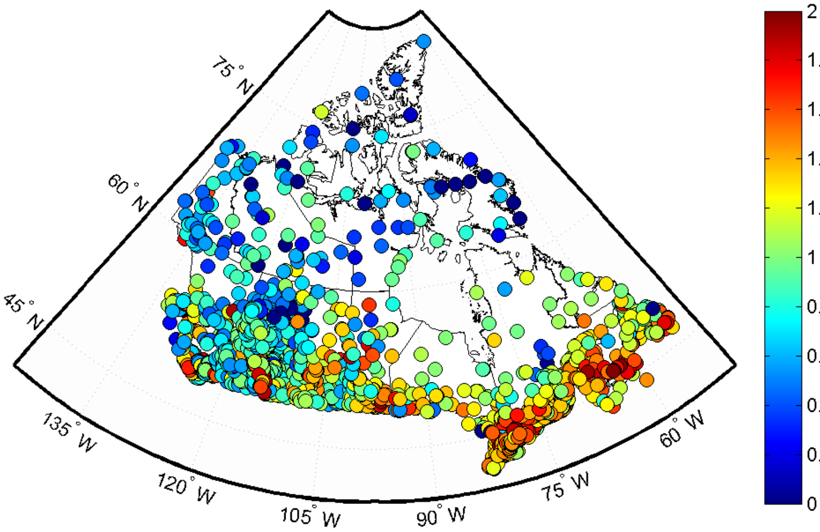


Figure D.9: Bias correction ratio of NARR in September

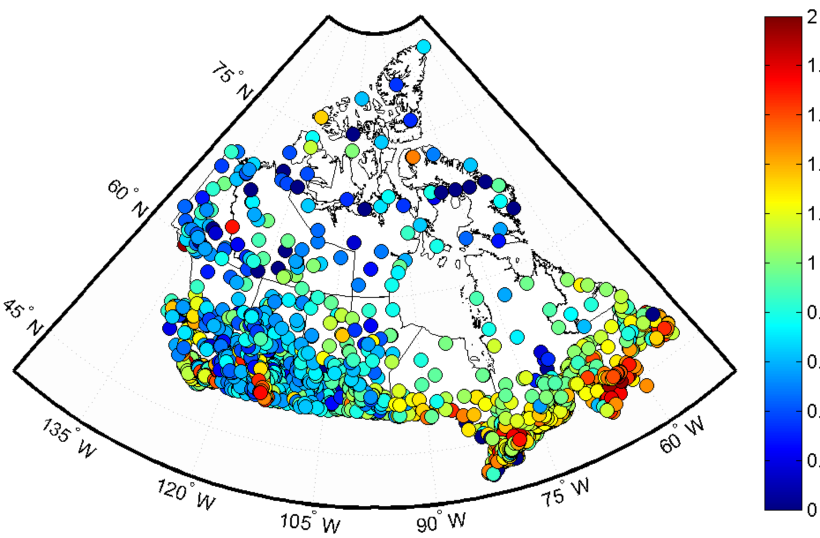


Figure D.10: Bias correction ratio of NARR in October

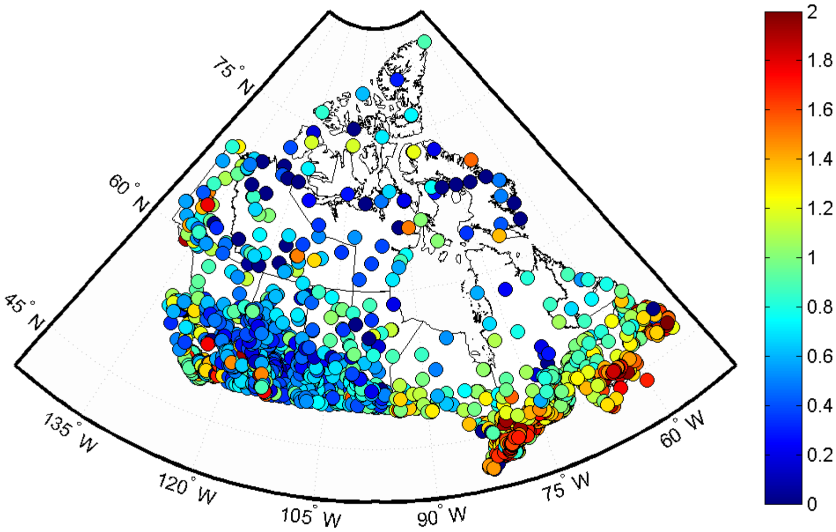


Figure D.11: Bias correction ratio of NARR in November

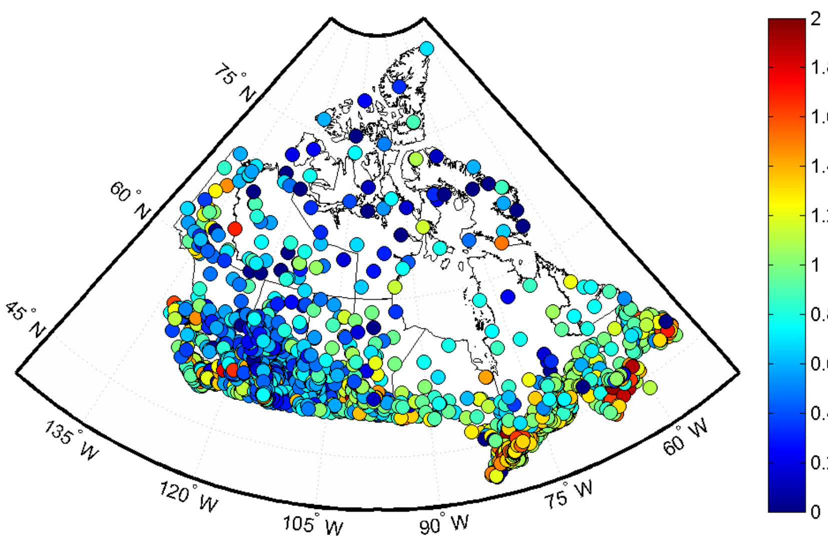


Figure D.12: Bias correction ratio of NARR in December

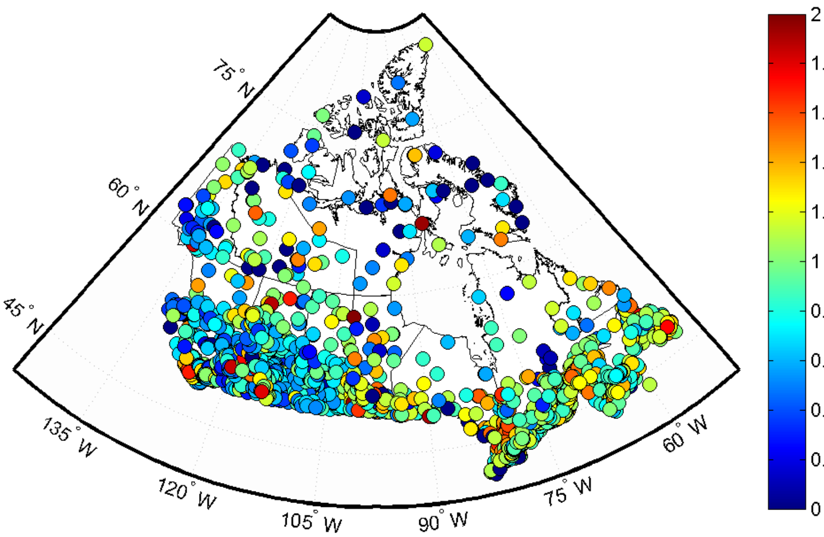


Figure D.13: Bias correction ratio of ERA-Interim in January

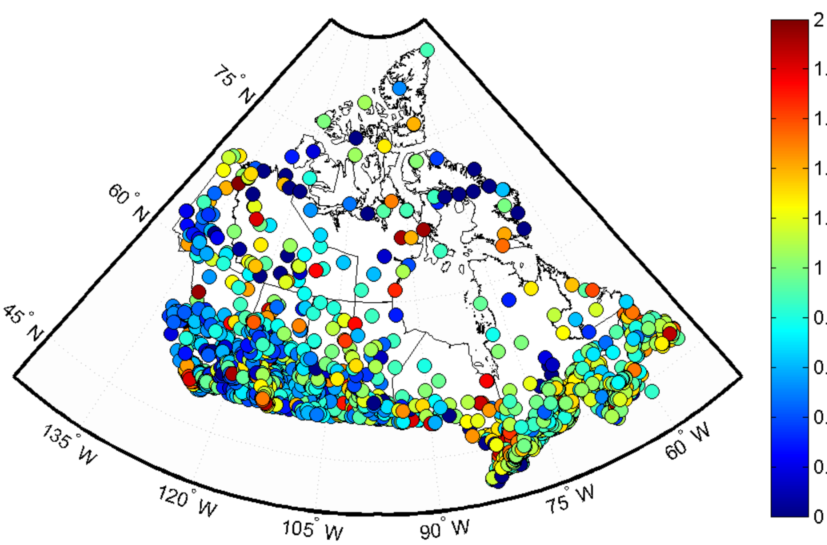


Figure D.14: Bias correction ratio of ERA-Interim in February

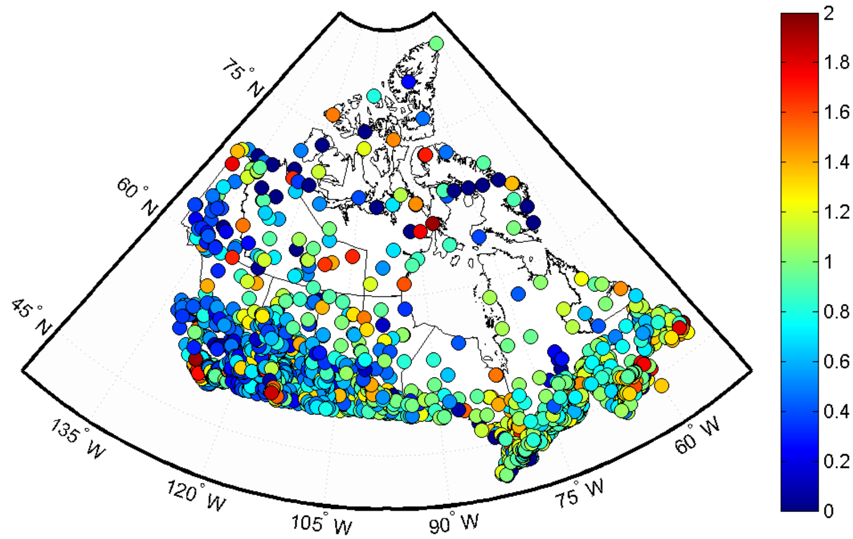


Figure D.15: Bias correction ratio of ERA-Interim in March

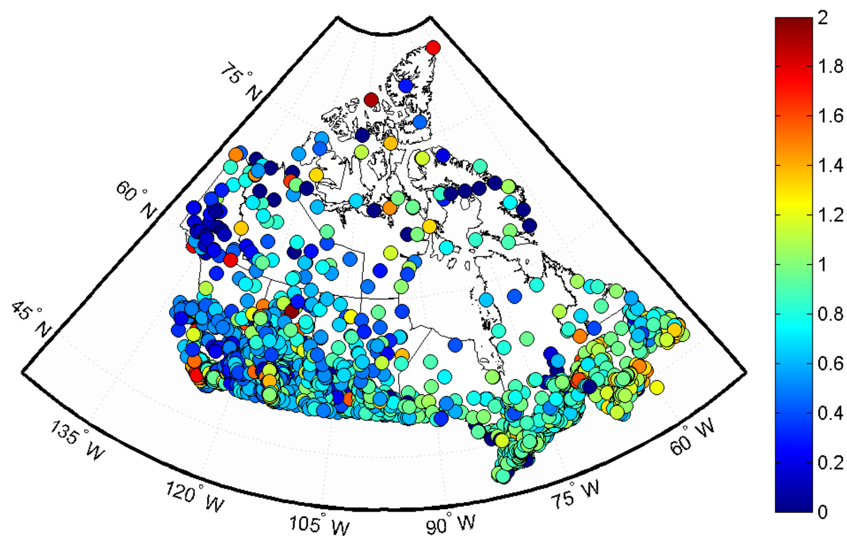


Figure D.16: Bias correction ratio of ERA-Interim in April

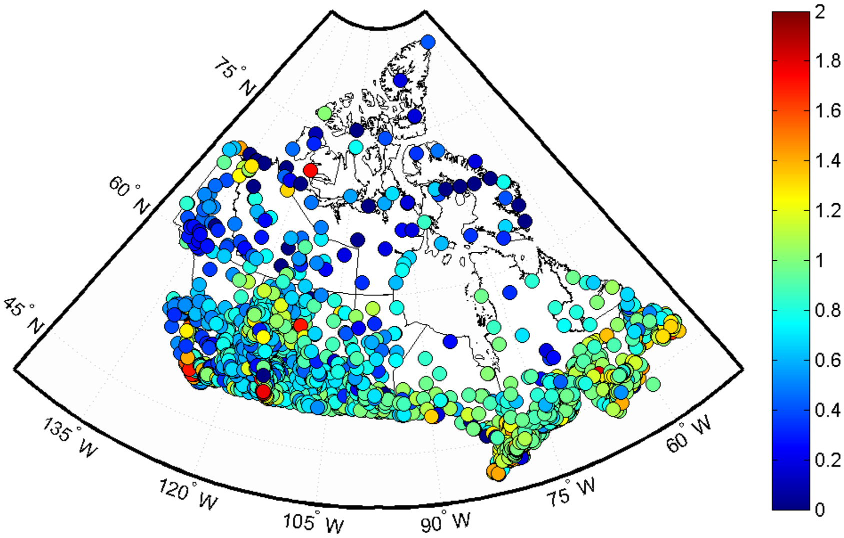


Figure D.17: Bias correction ratio of ERA-Interim in May

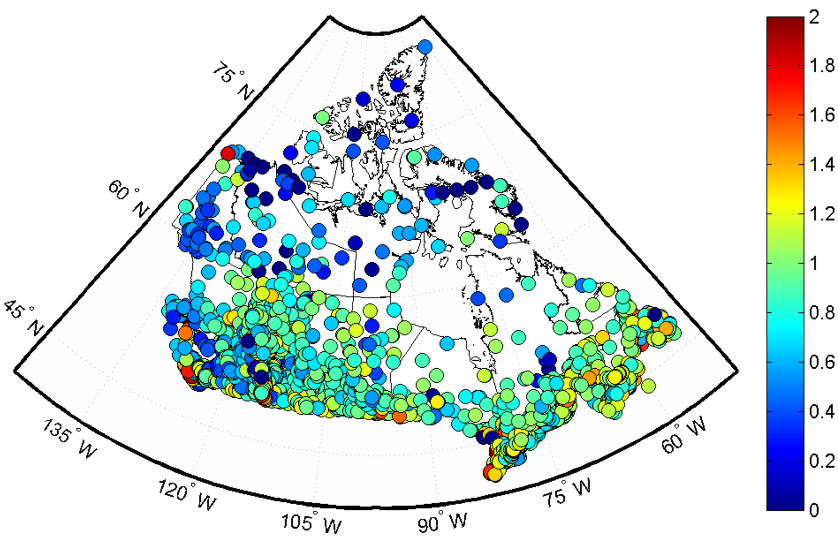


Figure D.18: Bias correction ratio of ERA-Interim in June

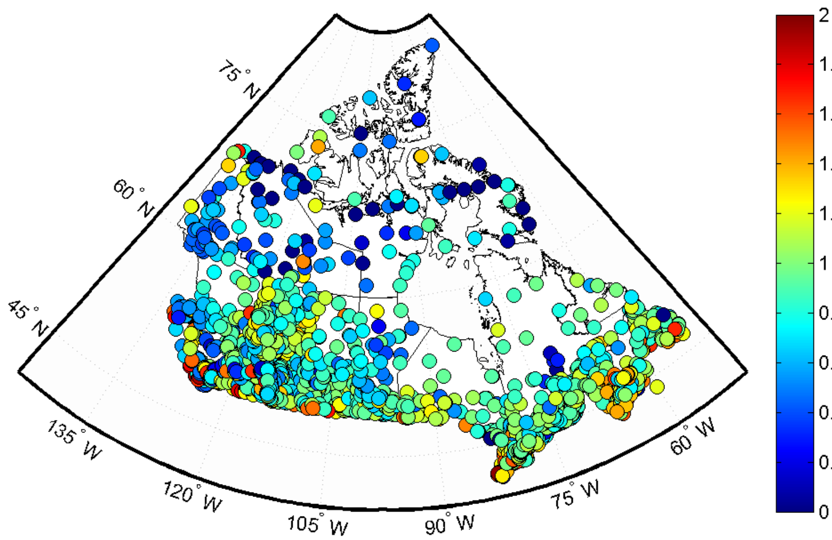


Figure D.19: Bias correction ratio of ERA-Interim in July

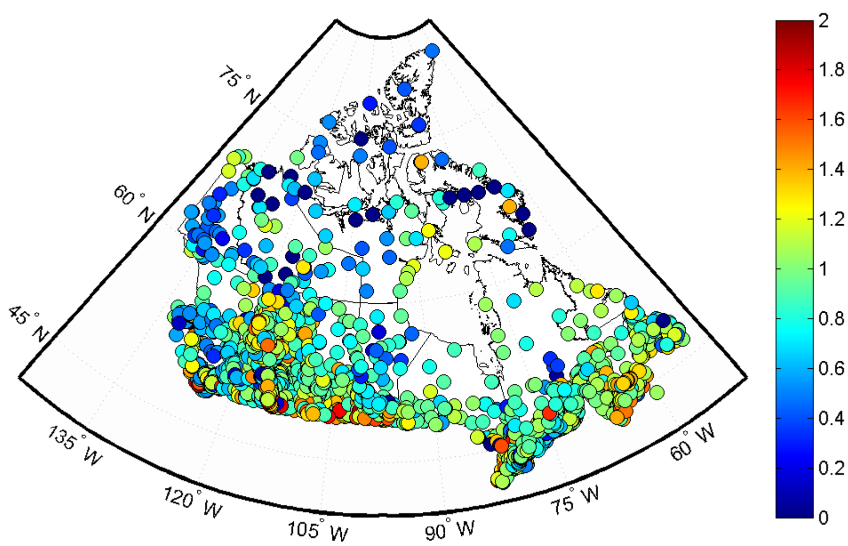


Figure D.20: Bias correction ratio of ERA-Interim in August

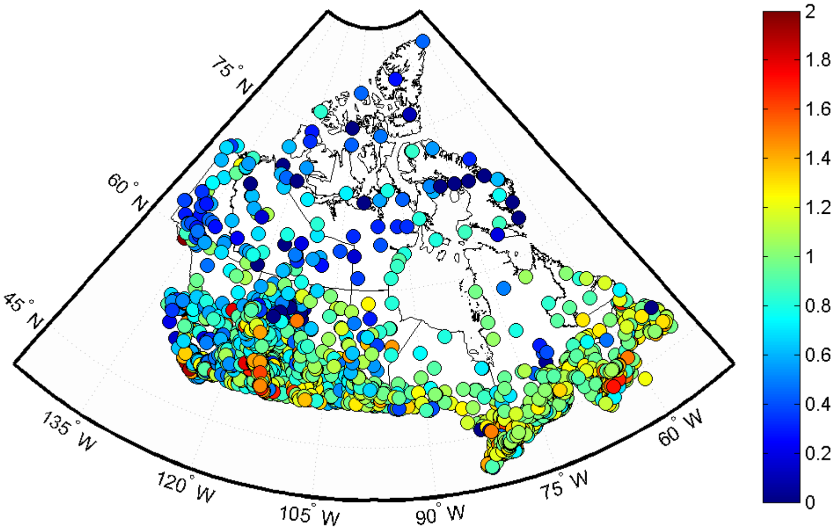


Figure D.21: Bias correction ratio of ERA-Interim in September

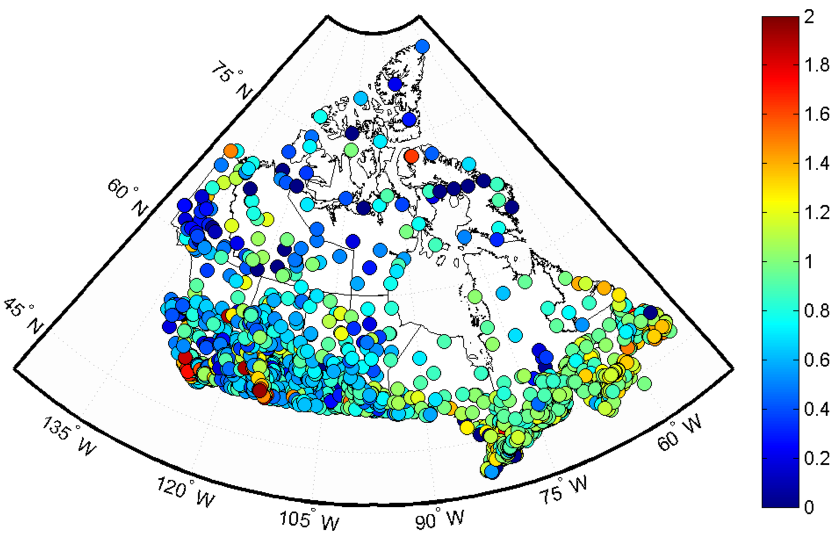


Figure D.22: Bias correction ratio of ERA-Interim in October

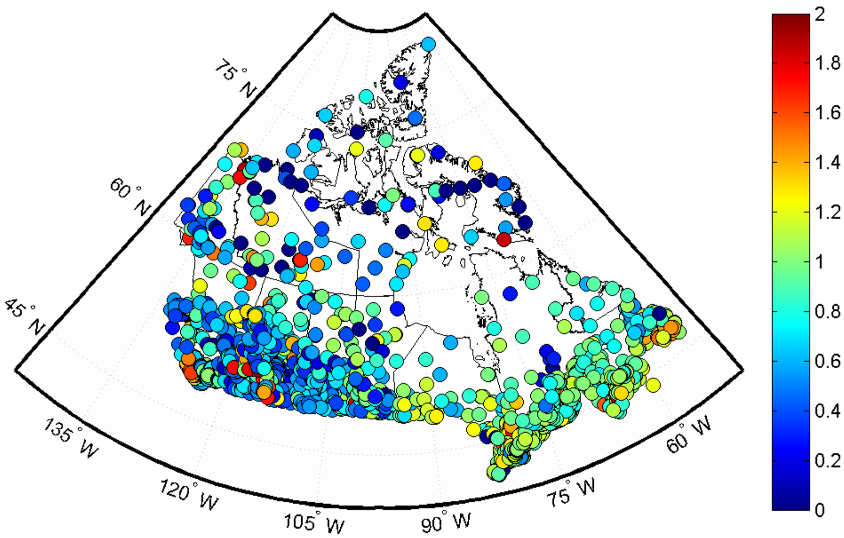


Figure D.23: Bias correction ratio of ERA-Interim in November

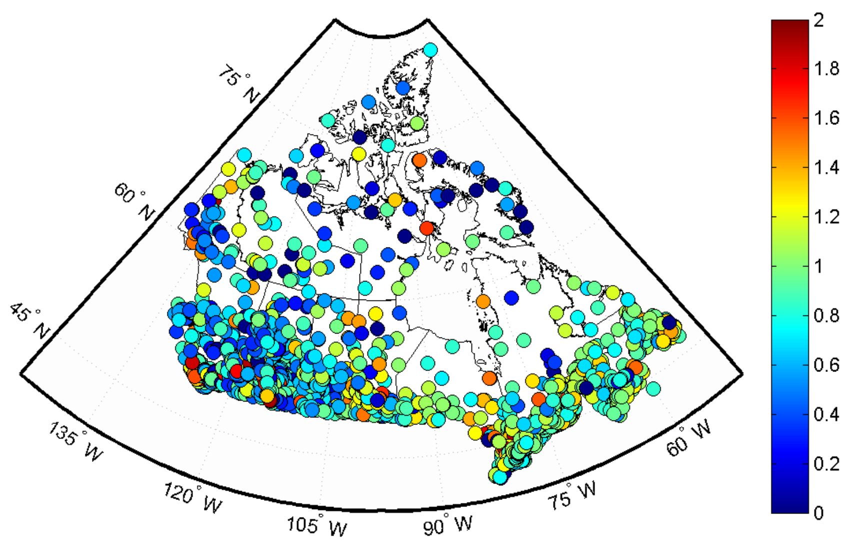


Figure D.24: Bias correction ratio of ERA-Interim in December