Assimilation of satellite based rainfall estimates with the Canadian Precipitation Analysis

by

Bruce Friesen

A thesis submitted to the Faculty of Graduate Studies of

the University of Manitoba

in partial fulfilment of the requirements for the degree of

Master of Science

Department of Civil Engineering

University of Manitoba

Winnipeg, Manitoba

Abstract

The Canadian Precipitation Analysis (CaPA) produces a gridded product by assimilating data from stations and the Global Environmental Multiscale (GEM) model. This project assesses the performance of the satellite based rainfall estimates for Canada, and the results of their assimilation with CaPA. The satellite based estimates considered are those from the Climate Prediction Center Morphing method (CMORPH) and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN).

Relative to the Second Generation of Daily Adjusted Precipitation for Canada (APC2), all satellite products are shown to generally underestimate rainfall, however convective events result in an overestimation. Skill scores show that the satellite products possess the most skill for eastern Canada and decreasingly so westward. When assimilated with CaPA, the satellite products express decreased skill for light rainfall and potential improvements for larger events. While central Canada experiences the greatest improvements, all regions benefit the most from June through August.

Acknowledgements

I would like to thank my advisor, Dr. Peter Rasmussen, for the opportunity to be involved with this project, as well as an environment conducive to exploring and learning about new ideas. His guidance was essential for the completion of this project.

I would also like to thank Dr. Tricia Stadnyk from the Department of Civil Engineering, Dr. Ronald Stewart from the Department of Environment and Geography, and Dr. Vincent Fortin from Environment Canada for agreeing to be apart of my review committee, and taking the time to read my thesis and provide feedback.

For their quick response to requests, patience to field questions, and general guidance in operating CaPA, a major component of this project, I would like to express my gratitude to Mr. Bruce Davison, and again, Dr. Vincent Fortin.

Finally, I would like to acknowledged the Natural Sciences and Engineering Research Council and Manitoba Hydro for their support of this work and making this project possible.

Table of Contents

List of Figures

List of Tables

Part 1 | Introduction

Knowledge of precipitation intensity and distribution is vital to numerous hydrological applications, including but not limited to the forecasting of floods and droughts, agriculture, climatic studies, and use as a resource. Measurements of precipitation by the use of weather stations provide great temporal records, however the costs associated with their installation is prohibitive to large dense networks. Density is particularly important as the weather stations are representative of a point and simple interpolation cannot fully delineate an event, such as with convective events which are spatially small but contribute large volumes of rainfall.

Remote sensing technologies, such as ground radar or satellite based retrievals, can provide frequent and spatially complete estimates within the domain of the instrument. Not measuring rainfall itself, these technologies require some interpretation estimate precipitation, which introduces inaccuracies. Canada has a radar network mostly present in the southern region (Fortin et al., 2014), however it shares one obstacle with weather stations — it would be costly to implement over a country as large as Canada.

An advantage of satellite technology is its spatial extent. A combination of satellites can give a near global estimate while maintaining comparable temporal resolutions (Janowiak et al., 2001). Relevant to Canada, rainfall estimates from satellite retrievals can provide information in regions beyond that of the radar network.

The Canadian Precipitation Analysis (CaPA) creates a gridded real-time analysis of precipitation across Canada (Mahfouf et al., 2007). In its current form it assimilates gridded modelled estimates, referred to as the background field, with station data, however CaPA's methodology also allows for the assimilation with additional observation data. This project explores the use of satellite data with CaPA.

1.1 Objectives

Since evaluations of satellite rainfall products are uncommon over Canada, the first objective of this project compares select satellite products with station data. Following this, the satellite data are assimilated with CaPA in two different ways. First, it is used in place of the climate model, meaning that it is assimilated with station data using CaPA. Second, it is used as an additional data source, meaning that the modelled estimates are adjusted based on both station and satellite data. In either scenarios, the resulting performance is compared to the current operation implementation. More succinctly, the objectives of this project are as follows:

1. Quantify the error associated with the selected satellite products.
2. Evaluate the use of satellite data as a background field.
3. Evaluate the use of satellite data as an additional observation source.

1.2 Scope

Given Canada's latitude, the applicable satellite products are limited. Those used here are the Climate Prediction Center morphing method (CMORPH; Joyce et al., 2004) and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Hsu et al., 1997). They both use infrared and passive microwave data to provide estimates with relatively high spatial and temporal resolutions and completeness. Of rainfall products based on infrared and passive microwave data, CMORPH and PERSIANN are the most commonly discussed, are shown to perform well, and most importantly provide estimates over some portion of Canada (Ebert et al., 2007; Kidd et al., 2011). For CMORPH, high and low resolution datasets of the same methodology are used in this project, while for PERSIANN the high and low resolutions are from different methodologies even though they share a name. One satellite dataset that is commonly evaluated is the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007).

It cannot be used for this project as the latitudinal extent in 50°N to 50°S.

The temporal and spatial domains of this project are defined by the datasets being used. The different objectives are explored over ten years to allow for more confidence when specific regions or times are evaluated. The error quantification uses data from 2001 to 2010, while the analyses with CaPA are from 2002 to 2011—with the discrepancy due to the readily available input data for use with CaPA. For PERSIANN-CCS, data are only available from 2003 onwards. The spatial domain of this project is for Canada below 60°N, which is the upper limit of the satellite products.

1.3 Document structure

Figure 1.1 shows the work flow of the project, where grey represents a dataset, red a process, and yellow is the associated part in this thesis. This figure shows how the satellite data undergoes quality checks and has estimates over snow covered ground removed before they are used to meet the previously defined objectives. It also shows that the error quantification encourages an adjustment before the satellite data are used with CaPA.

This thesis is organized as a series of parts which are for the most part chronological after Parts 2 and 3, which merely provide contextual information. Specifically, Part 2 details the methodology of CaPA, the satellites and instrumentation relevant to this project, and how the retrievals are related to rainfall rates, with an emphasis on the information pertaining to satellite data as it is the new contribution in this project. Following that, Part 3 discusses the various datasets used in this project—both satellite and station data.

Part 4 handles the first objective where the errors in the satellite data are quantified and the performance relative to observations is evaluated, which is mostly completed through decomposing the error into its different components and various categorical skill scores. This part also provides findings regarding the quality control procedures that were implemented.

Part 4 also shows that some bias removal or adjustment should be performed to the satellite data before their implementation into CaPA. Part 5 details the implementation of the adjustment and the resulting changes in skill scores. It also discusses the removals of rainfall estimates over snow covered ground, which were in fact performed before the error quantification of Part 4.

The adjusted rainfall estimates are then assimilated with CaPA as a background field and an additional data source in Parts 6 and 7, respectively, addressing the second and third objectives. Both of these parts evaluate the results of the analyses in a similar manner. As a quantification of performance, categorical scores consistent with those used by Environment Canada are employed (Fortin et al., 2014; Lespinas et al., 2014; Mahfouf et al., 2007). The resulting scores are compared to the operational form to assess improvements. Within these two parts, the results become increasingly specific in that they start as scores across Canada, but are eventually broken down by month, region, and 6-hourly time step. The final part summarizes the results and makes recommendations for the use of satellite data across Canada.

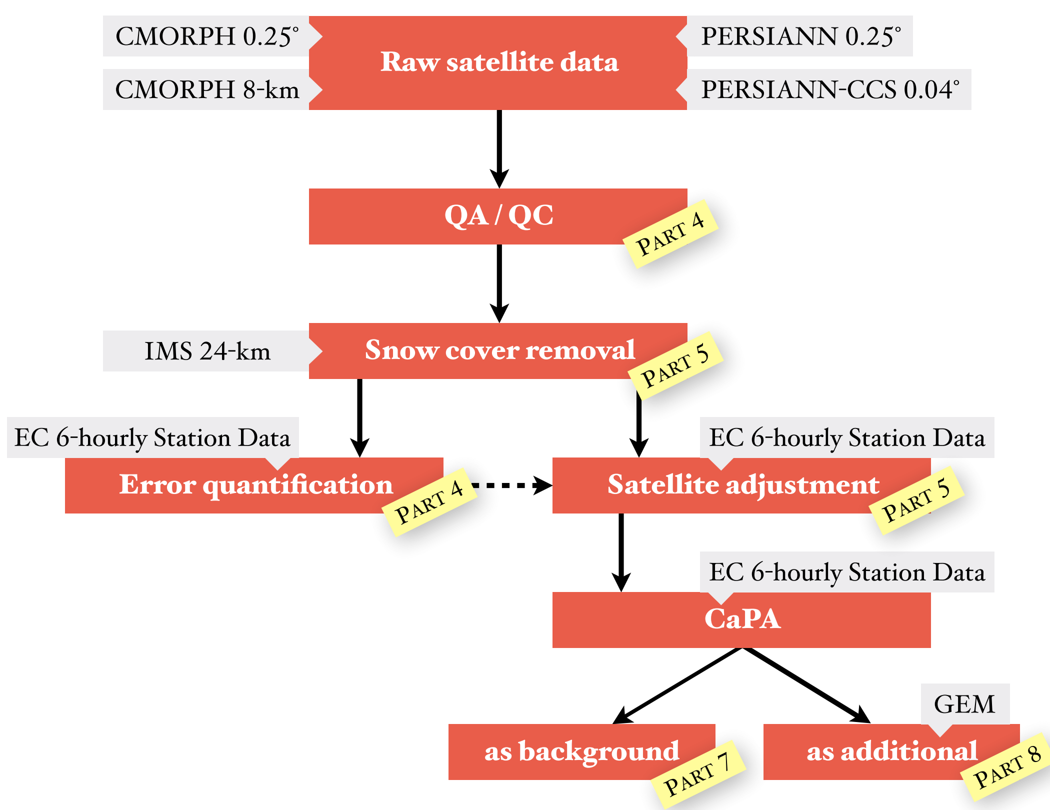


Figure 1.1: Project flow chart.

Part 2 | Background

Two major components of this project are CaPA and satellite data, both of which are detailed in the following sections. The first section is for CaPA. Since the objectives of this project do not involve changing CaPA's methodology and the current version serves the needs of the objectives, it is not explained in the fullest of detail. The last two sections are concerned with the satellite retrievals and how they relate to rainfall rates. Between these two sections, the current satellites, their instruments, and the methods to estimate rainfall are detailed. Only the instruments and methods relevant to the datasets mentioned in this project are included.

2.1 Canadian Precipitation Analysis (CaPA)

The creation of CaPA began in 2003 by the Meteorological Research Branch and the Canadian Meteorological Centre of Canada with the objective of creating a real-time gridded precipitation product for Canada. This is achieved through the statistical interpolation of station data with forecasts from the Global Environmental Multiscale (GEM; Côté et al., 1998) model which serves as a background field. The analysis is created by using simple kriging to interpolate the differences between GEM and the station data, which is then applied back to GEM. (Mahfouf et al., 2007). The main advantage of using the GEM model as a background field is that it is spatially and temporally complete. Analyses are performed in 6-hour increments (00:00, 06:00, 12:00, and 18:00 UTC) with the latest GEM grid at 10 km resolution, where the final analysis is the same resolution as the input. This project uses version 2.4b8 of CaPA which has undergone updates to the statistical interpolation method as well as kept up with the spatial resolution of GEM, which was originally 25 km. Changes in the methodology include the transformation applied to the input data from a log-normal to cubic-root, and also the formation of the variograms (Lespinas et al., 2014).

Prior to an assimilation, station data are collected from a variety of networks (see Section 3.5) and undergo numerous quality control checks. The first removes some measurements associated with snow. Automated stations are removed if the temperature is near or below 0°C, and manned stations are removed if the temperature is near or below 0°C with a wind speed is greater than roughly 2 m/s (Environment Canada, 2014b). The other quality control procedures are concerned with relative changes in space and time. If the frequency of recordings change for a given station, it will be removed as there may be an issue with the recordings. Where stations are close together a 'super station' is created, and the individual stations are compared to each other for inconsistencies. Finally, each station is compared with the leave-one-out cross-validation at its location to find large differences, potentially a result of measurement error. The correlation length is considered during this phase to avoid rejecting stations during small but intense events (Lespinas et al., 2014).

As mention previously, CaPA uses statistical interpolation to interpolate the differences between the observations and the background field in transformed space, which can be represented by the following equation from Daley (1991) and Mahfouf et al. (2007):

$$A_g = B_g + \sum_{s=1}^{n} W_{sg} \left(O_s - B_s\right) \tag{2.1}$$

where the analysis (A) of grid cell g equals the background value at that location (B_g) plus the weighted (W_{sg}) combination of differences between the observation (O) of station s and interpolated background values at that location (B_s).

The weights in Equation 2.1 minimize the analysis error variance by solving:

$$\left(C_B + C_O\right) W = C_b \tag{2.2}$$

as per Daley (1991), where C_B and C_O represent the background error covariance matrix at the station locations, and the observation error covariance matrix, respectively. Meanwhile, C_b is the error covariance between an observation O_s and grid point B_g. To account for the horizontal correlation present in the background field, a correlation function is used in combination with the background standard deviation as presented by Mahfouf et al. (2007) and shown in Equation 2.3, where r_{sg} is the distance between s and g, σ_B is the standard deviation of background error, and L is a characteristic length.

$$C_B = \sigma_B \left(1 + \frac{r_{sg}}{L}\right) \exp^{-\frac{r_{sg}}{L}} \tag{2.3}$$

Meanwhile, observations from stations are assumed to be uncorrelated in space:

$$C_O = \sigma_O \times I \tag{2.4}$$

as defined by Daley (1991), where σ_O is the standard deviation of observation errors. The solution of Equation 2.2 then requires the determination of three parameters: σ_B, σ_O, and L. While Mahfouf et al. (2007) fixed these parameters, the current methodology of CaPA recalculates them for every time step via a variogram analysis. The variogram is assumed to be isotropic, homogeneous, and an exponential model. The model fit is based on data from the current time step as well as those from previous time steps, with an exponential filter to give recent data more influence (Environment Canada, 2014a; Lespinas et al., 2014).

The inclusion of radar data as an additional data source has been explored and proposed to be used in the operational form. This project's assimilation of satellite data as an additional source is done in the same manner as that of radar data (Fortin et al., 2014). It should be noted that CaPA operates in such a way that when an observational dataset is gridded, horizontal correlation is accounted for in a similar manner as explained above for the background field.

2.2 Current satellites and instrumentation

The satellite rainfall estimates considered in this project are based on retrievals from satellites in two different orbits—low earth orbit (LEO) and geostationary earth orbit (GEO). Depending on the orbit, the temporal and spatial characteristics of the retrievals differ.

LEO satellites are defined by having an altitude less than 1,500 km (Capderou, 2005). For each orbit, the satellites cross the equator at the same local time, observing a given location roughly twice per day. Therefore, these satellites cannot provide spatially or temporally complete retrievals (Joyce et al., 2010; Kidd and Levizzani, 2011). For example, TMPA, which uses both passive microwave and infrared data from multiple satellites, covers roughly 80% of the Earth between 50°N and 50°S for a given 3-hour period (Huffman et al., 2010). In this context of this project, the LEO satellites house imagers and sounders to measure passive microwave emissions. For rainfall measurements in applications beyond the datasets used in this project, they can also retrieve infrared data. The TRMM satellite was the first to house satellite based radar (Kidd and Levizzani, 2011), which is not directly used in this project's satellite datasets.

GEO satellites are further from the Earth with an altitude of 35,786 km — a distance chosen such that an orbit takes the same amount of time as the Earth does to rotate (Capderou, 2005), and as such it is constantly observing the same location. Relevant to this project, these satellites house infrared sensors that can provide spatially (within their domain) complete measurements every 15 to 60 minutes depending on the instrumentation (Joyce et al., 2010; Kidd and Levizzani, 2011).

Table 2.1 lists the current satellites relevant to the data products used in this project, where the list is compiled of information from Huffman et al. (2010), Joyce et al. (2010), and World Meteorological Organization (2014). See Appendix A for full names of the agencies, satellites, and sensors.

2.3 Relating temperature brightness to rainfall

The retrievals from passive microwaves and infrared are of temperature brightness, so some algorithm needs to be employed to relate them to rainfall rates. The methodologies to do so can be categorized of emission, scattering or multichannel inversion types (Kidd and Levizzani, 2011; Kummerow et al., 1996). The ocean's emissivity is relatively lower and more homogeneous than that of land. The ocean then appears cool in passive microwave retrievals while the rainfall above it appears warmer, to the degree that lower frequencies (less than 37 GHz) can be used to separate rainfall occurrence and define a rainfall rate. Methods that look for an increase in temperature brightness to find rainfall are of the emission type. The greater heterogeneity and emissivity of land makes it an unsuitable region for emission methods, so instead an ice-scattering method is used. As the frequencies of the passive microwave instruments increase, they are increasingly scattered by ice particles, usually present in precipitating clouds. These methods then look for depressions in temperature brightness for rainfall as the

Table 2.1: Current satellite information relevant to this project.

Agency	Satellites	Sensor	Altitude (km)
NASA / JAXA	TRMM	TMI	402
EUMETSAT	Metop-A	AMSU / MHS	827
EUMETSAT	Metop-B	AMSU / MHS	827
NASA	Aqua	AMSR-E	705
NOAA	NOAA-15	AMSU	807
NOAA	NOAA-18	AMSU / MHS	854
NOAA	NOAA-19	AMSU / MHS	870
DMSP	F-13	SSM/I	850
DMSP	F-15	SSM/I	850
DMSP	F-16	SSMIS	850
DMSP	F-17	SSMIS	850
DMSP	F-18	SSMIS	850
EUMETSAT	Meteosat-7	MVIRI	35,786
EUMETSAT	Meteosat-8	SEVIRI	35,786
EUMETSAT	Meteosat-9	SEVIRI	35,786
EUMETSAT	Meteosat-10	SEVIRI	35,786
JMA	MTSAT-2	MTSAT-2 Imager	35,786
NASA / NOAA	GOES-13	Imager	35,786
NASA / NOAA	GOES-14	Imager	35,786
NASA / NOAA	GOES-15	Imager	35,786

emissions from the Earth are scattered before reaching the satellites. Multichannel inversion methods profile the vertical cloud structure by using multiple frequencies for a given location. The measured profile is compared to previously observed events to infer a rainfall rate.

In a more recent description of CMORPH, Joyce et al. (2010) outlines the methodologies to relate satellite retrievals of sensors used in CMORPH to rainfall rates. For the TMI instrument aboard TRMM, the methodology of Kummerow et al. (1996) is used which is of the multichannel inversion type. For a given location, multiple channels define the hydrometer profile, which is related to rainfall rates using the Goddard Profiling Algorithm (GPROF).

For the SSM/I instruments aboard the DMSP satellites, CMORPH uses estimates from the methodology of Ferraro et al. (2000). This method is of the scattering type, which looks for depressions in temperature brightness to identify rainfall. More specifically, it defines a scattering index which is the difference of the temperature brightnesses at 23 and 89 GHz, as the greater frequencies are scattered by the presence of ice particles. The scattering index is used to define the rainfall as convective or stratiform, as convective events will have more scattering. Then, the appropriate empirical relationship (convective to stratiform) is used to relate the scattering index to rainfall rate. Based on the temperature brightness over snow covered ground or arid regions, the estimates may be defined as indeterminate. The author notes that lower frequencies have difficulty detecting stratiform rainfall due to the smaller ice particles, and found that this is particularly true for western United States and western Europe because of the different rainfall processes in these regions relative to elsewhere on the continents. Although Ferraro et al. (2000) does not mention it, the older paper, Ferraro and Marks (1995), states that this method uses an emission type over the ocean.

Finally, for the AMSU instruments the methodology of Zhao and Weng (2002) is used which relates the ice water path (or vertically integrated water content) and ice particle diameters to rainfall rates based on cloud simulations. This method also removes estimates over snow covered ground based on a combination of temperature brightnesses at multiple frequencies and surface temperatures.

While Joyce et al. (2010) outlines each algorithm used for the input data in CMORPH, Hsu and Sorooshian (2008) only refers to the methodology of Ferraro and Marks (1995) for PERSIANN. Both CMORPH and PERSIANN use the same global infrared dataset from Janowiak et al. (2001).

As suggested earlier, snow covered ground is problematic for rainfall estimates from passive microwave retrievals because it scatters emissions in a similar manner as the particles at the tops of precipitating clouds. Very dry sand can cause similar issues (Ferraro et al., 2000; Wilheit et al., 1994; Zhao and Weng, 2002). The ice screening methodology of Section 5.2 is employed before the analyses and evaluations of this project.

Part 3 | Datasets

Numerous datasets are used throughout the duration of the project and serve as either sources of rainfall estimates or for intermediate processing. This section discusses these datasets, where following a brief overview each dataset is detailed in its own section.

In terms of satellite based rainfall, the objective of this project is to consider rainfall estimates from PERSIANN and CMORPH. These datasets are made available in a variety of temporal and spatial resolutions, and even different methodologies, as is the case for PERSIANN. Both CMORPH and PERSIANN produce a common 3-hour 0.25° estimate in near real-time, an estimate adjusted with observations (not available in real-time), and a higher resolution product. For PERSIANN, the higher resolution is unique in that it is the output of a different algorithm (cloud classification system), compared to CMORPH where the high resolution product follows the same methodology as the lower resolution product.

As a means to evaluate the satellite products and provide input into CaPA, two observational datasets are used. The first is the Second Generation of Daily Adjusted Precipitation for Canada (APC2) which is not available in real-time, but is the best estimate of precipitation for Canada, and as such is most applicable for evaluating the error and performance of the satellite products (Mekis and Vincent, 2011). The second dataset is a collection of stations from Environment Canada and co-operative networks at 6-hourly increments. These stations are used to rescale the satellite data, as input into CaPA, and a reference field for the final evaluations.

GEM produces gridded estimates of precipitation across Canada and is the background field

Table 3.1: Summary of dataset properties.

Dataset	Domain: Temporal	Domain: Spatial	Resolution: Temporal	Resolution: Spatial
CMORPH	1998 − present	60°N − 60°S	3 h, Daily	$0.25° \times 0.25°$
CMORPH	1998 − present	60°N − 60°S	30 min	8 km
PERSIANN	2001 − present	60°N − 60°S	3 h, 6 h	$0.25° \times 0.25°$
PERSIANN-CCS	2003 − present	60°N − 60°S	1 h	$0.04° \times 0.04°$
APC2	∼1900 − 2012	Canada	Daily	Point observations
EC 6-hourly data	2002 − 2011	North America	6 h	Point observations
GEM	2002 − 2011	North America	6 h	10 km
IMS	1997 − present	90°N − 0°	Daily	4 km, 24 km

in the operational form of CaPA. This project explores the use of satellite data in place of, and in addition to GEM.

Since the performance of satellite rainfall estimates is greatly affected by snow covered ground (Munchak and Skofronick-Jackson, 2013), the Interactive Multisensor Snow and Ice Mapping System (IMS) is used to remove these estimates. Although CMORPH does document a similar snow-screening algorithm, it is important that the same method is applied to both PERSIANN and CMORPH for fair comparisons (Joyce et al., 2004).

The properties of these datasets are summarized in Table 3.1, as they are available at this time of this project. For the temporal domain, only years where all months are available are included.

3.1 The Climate Prediction Center Morphing method

While the infrared retrievals from geostationary satellites are spatially complete and available every 15 to 60 minutes, the relationship to rainfall rates is not as strong as that for passive microwave data (Joyce et al., 2010). The passive microwave data are from orbiting satellites, and therefore only provide sparse coverage. CMORPH, which is run by NOAA, operates in such a way to take advantage of the strengths of these two sources — the spatial completeness of infrared data and accuracy of passive microwave data (Joyce et al., 2004). The more recent publication of Joyce et al. (2010) states that CMORPH uses infrared data from Meteosat, GOES, and MTSAT (formerly the Geostationary Meteorological Satellite) geostationary satellites and passive microwave data from NOAA, DMSP, TRMM and NASA's Aqua polar orbiting satellites.

Part of CMORPH methodology is handling the separate sources of passive microwave data. The infrared data are received as a global image and does not require such processing. Between the previously mentioned orbiting satellites, there are four sensors used for passive microwave measurements: TMI, AMSR-E, SSM/I, and AMSU. The sensors produce slightly different estimates of rainfall because they operate on different channels (frequencies) and therefore employ different methodologies to relative the retrievals to rainfall. To create one cohesive estimate, all other sensors are rescaled to the AMSR-E and TMI sensors via distribution matching, as these two sensors are of greater resolution and possess improved detection thresholds over the oceans (Joyce et al., 2010). Also, an advantage of using TMI is that the TRMM satellite orbits below the other satellites because it is at a lower altitude, providing many overlapping grid cells. For grid cells where two satellites provide an estimate, a priority list is used as as follows:

1. TMI
2. AMSR-E
3. SSM/I
4. AMSU

The general idea of CMORPH is that the more accurate passive microwave estimates are moved through time and space based on the movements observed in the infrared data, because multiple global infrared retrievals are made available between those of passive microwave data. The following procedure is a summary of that detailed by Joyce et al. (2004). First, the motion vectors are created between successive infrared images via spatial lag correlation. Then, a given passive microwave estimate is morphed for each infrared time step based on the motion vectors until the next passive microwave estimate is available. The same is then done in reverse, where the latest passive microwave estimate is morphed backwards in time using the opposite of the motion vectors to the original estimate. So, for each infrared time step there exists two rainfall estimates — one based on the first estimate morphed forwards in time, and the second based on the latest estimate morphed backwards in time. These two are combined via inverse distance weighting with respect to time. An important attribute of this methodology is that the rainfall estimates are only based on those from microwave data. This project uses CMORPH version 1.0.

3.2 Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks

Currently operated by the University of California Irvine, PERSIANN provides estimates from similar data as CMORPH. Specifically, infrared data from the GOES, Meteosat, and MTSAT geostationary satellites, and passive microwave data from TRMM, NOAA, and DMSP orbiting satellites (Hsu et al., 1997; Hsu and Sorooshian, 2008).

PERSIANN combines the strengths of the infrared and passive microwave data through the use of a neural network, which can take information from multiple sources to reproduce complex non-linear processes (Hsu et al., 1997) and ultimately produce a rainfall estimate. The global infrared images provide more coverage for a given 30 minutes than passive microwave, and as such, are used an input data into the neural network. The greater accuracy of the passive microwave data make them ideal for training the neural network's parameters as they becomes available (Hsu et al., 1997). The development and validations were performed over the Japanese Islands and Florida. Hsu and Sorooshian (2008) details the input data into the neural network for a given pixel as the:

- Mean and standard deviation of temperature brightness of surrounding pixels (3×3 and

5×5 windows as per Hsu et al. (1997))

- Classification of the extracted feature
- Mapping of the texture to rainfall rate

3.3 The PERSIANN Cloud-patch Classification System

A second PERSIANN dataset is used in this project which has a different methodology and resolution than that previously discussed. The concept of the PERSIANN Cloud-patch Classification System (PERSIANN-CCS) is to extract the different properties of the precipitation events before relating them to rainfall rates (Hong et al., 2004). Infrared data are used to define the properties of the events, while ground radar and passive microwave observations aid in relating the temperature brightnesses to precipitation rates (Hsu et al., 2007).

The following four steps are those taken to create a precipitation analysis (Hong et al., 2004; Hsu et al., 2007). The first step spatially separates the precipitation events for a given time step with a watershed-based method, which in this application was found to better delineate events than a constant brightness temperature, for example 253K. Once the events are spatially identified, features are extracted at each threshold temperature of 220K, 235K, and 253K. The consideration of three threshold temperatures provides additional information at varying altitudes. The recorded features of a given event are of three categories: coldness, geometric, and texture. The coldness of an event is defined by its minimum and mean temperature, and the geometric properties are its area and shape index. Finally, the texture features are measures of the events' temperature brightness standard deviation, local means and standard deviations, the gradient of brightness temperature, and the maximum second moment. All of this information is used as input into a self-organizing feature map, from which groups of cloud types are defined. The final step relates the temperature brightnesses to rainfall rates through an empirical equation for the specific group.

3.4 The Second Generation of Daily Adjusted Precipitation for Canada

APC2 is used as the reference field to quantify the error of the satellite products. The validity of the error quantification is dependent on the confidence of the reference field, and APC2 provides the best fit for this need as it not only consists of the most reliable stations, but also adjusts each station type for known deficiencies. The dataset is comprised of 464 stations across Canada, shown in Figure 3.1, with roughly 400 in the domain of this study. Some stations have been combined with those nearby to create complete time series. All the measurements are done by observers, not by automatic systems (Mekis and Vincent, 2011).

The adjustment method is the same for the various gauges, however its magnitude changes with gauge type. Type A are amongst the oldest used and transitioned from copper to plastic construction around 1965. Both materials are accounted for in the adjustment. Following Type A, Type B gauges were introduced in the 1970s and reduced errors associated with evaporation and retention losses (Devine and Mekis, 2008). Stations with these gauges are adjusted

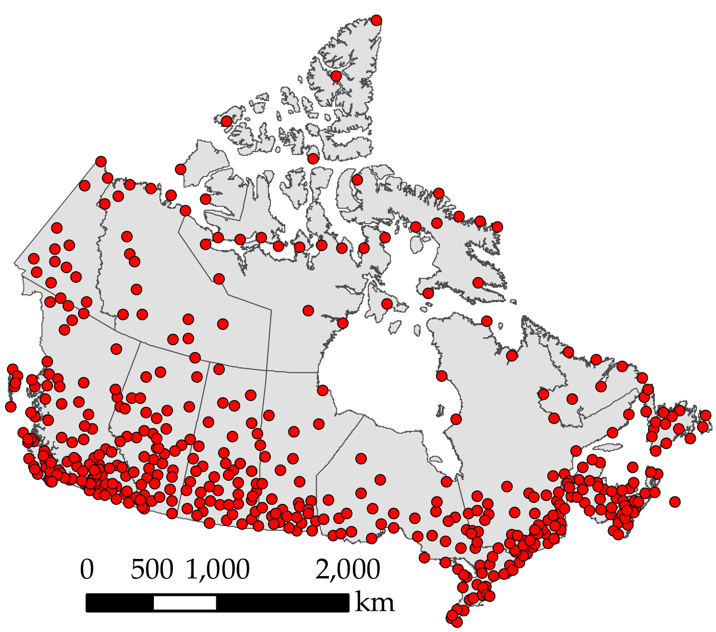


Figure 3.1: Stations present in the APC2 dataset.

in accordance with Equation 3.1:

$$R_a = (R_m + F_c + E_c + C_c) \times (1 + W_c) \tag{3.1}$$

where R_a and R_m are the adjusted and measured daily rainfall values, and F_c, E_c, C_c, and W_c are corrections for losses associated with wetting at the funnel, evaporation, and wetting in the receiver, respectively (Mekis and Vincent, 2011). The values of these corrections are from observations of a side-by-side experiment. Overall, the adjustment increases the magnitude of precipitation. The largest relative adjustment for rainfall is in northern Canada, because the adjustment increases trace events' magnitude which contribute more to total precipitation in this region compared to the rest of Canada. Below 60°N, the ratios of adjusted to measured rainfall are for the most part between 1.05 and 1.10 (Mekis and Vincent, 2011).

3.5 Environment Canada's compiled observational data

The station data used in CaPA are from multiple networks across North America that have been, and continue to be, compiled by Environment Canada. These networks include Canada's own synoptic stations, the Meteorological Aerodrome stations (METAR), the Météorologique Coopératif du Québec (RMCQ), and an American co-operative network, SHEF. From here on, the compiled dataset is referred to as EC 6-hourly data.

Canada's synoptic stations are shown in Figure 3.2 and have been classified as manual and automated stations using Environment Canada's filtering tools. The stations shown here are those used at some point for the analyses of Parts 6 and 7, totalling roughly 1150 stations. These stations report every six hours with precipitation measurements, and also temperature which is used in the quality control procedures (Lespinas et al., 2014). The measures at manual stations are usually from Type B gauges, while those from automated stations are from tipping bucket or Fischer-Porter gauges (Mahfouf et al., 2007).

RMCQ and SHEF are both co-operative networks. The stations in the SHEF network are located in the United States and report daily accumulations. As such, they are not used in this project because 6-hourly analyses are performed. Outside of Canada then, measurements

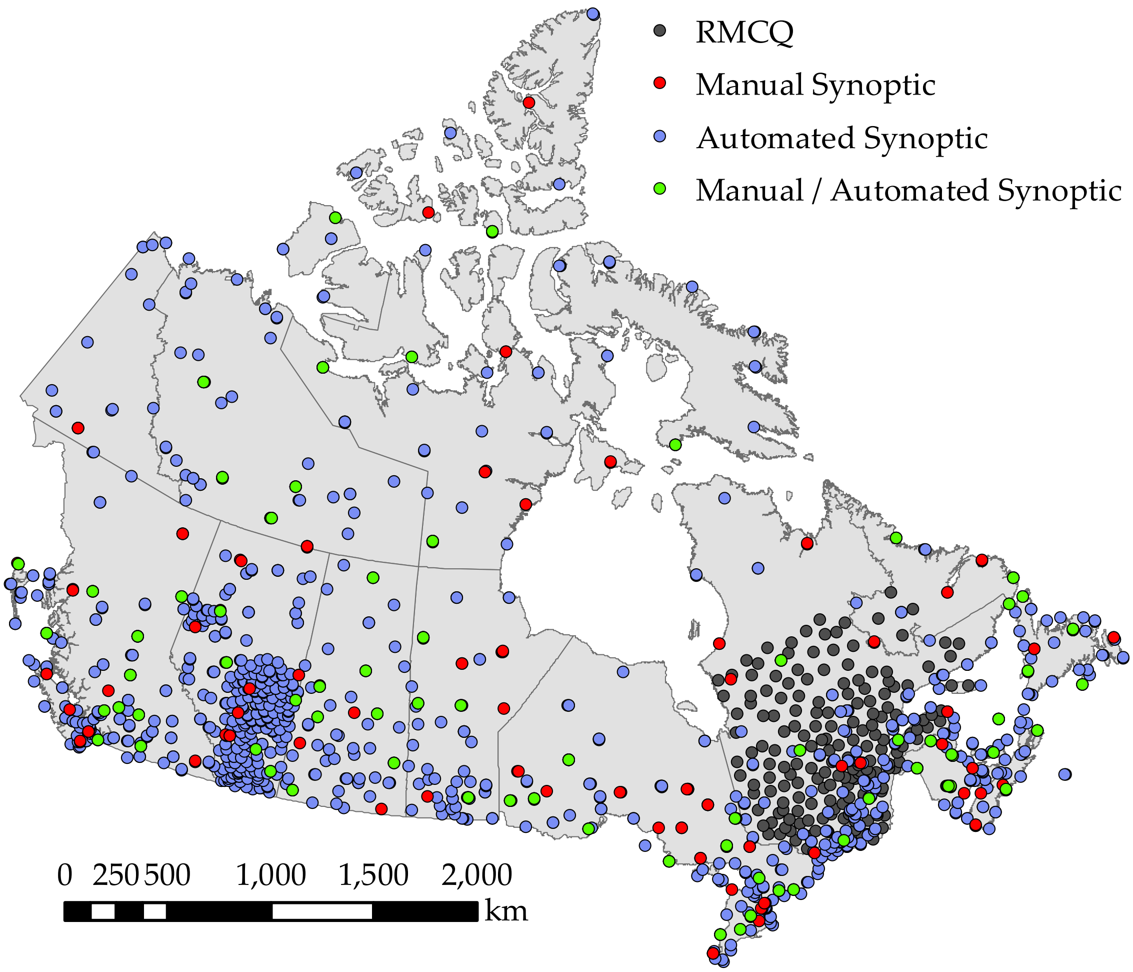


Figure 3.2: Stations used in CaPA from 2002 to 2011.

from the METAR are increasingly important. METAR stations are usually located at airports and only report precipitation if it has occurred (Lespinas et al., 2014).

Given the quality control procedures and the nature of the METAR network, the number of stations used for a given time step will vary throughout the analysis. Therefore, the evaluations of the analyses performed in Parts 6 and 7 are done with a bootstrapping methodology which accounts for spatial and temporal changes in station density, to give even weight to all time steps and locations. This methodology is described at the beginning of Part 6.

3.6 Global Environmental Multiscale model

GEM was developed by the Canadian Meteorological Centre and made operational in 1997 (Côté et al., 1998). In 2004, this atmospheric model was updated with improvements to various elements of the physics package regarding the cloud boundary layer, shallow convection, deep convection and orographic drag. Also, the resolution was increased from 24 km to 15 km, and more recently to 10 km (Mailhot et al., 2006). The latitude-longitude grid of GEM is such that the resolution increases over an area of interest, and the pole is rotated to limit distortion.

Since the estimates from GEM used in CaPA are from a short-term forecast, they are a balance of errors between those present during the spin-up phase and those from forecasting further into the future. Mahfouf et al. (2007) found that a suitable solution for a given 6-hour estimate is to take the difference between 12 and 6-hour accumulated forecasts, thus allowing for six hours of spin-up.

3.7 Interactive Multisensor Snow and Ice Mapping System

The IMS dataset (National Ice Center, 2008) classifies the Northern Hemisphere as sea, land, ice, or snow on 4 and 24 km grids daily. Data is available from 1997 to present and is created and published by the National Oceanic and Atmospheric Administration's National Environmental Satellite Data and Information Service (NOAA NESDIS).

While the 24-km product is available from 1997 to present, the higher resolution product is only available from 2004 onward. Since the satellite data in this project precedes 2004, the lower resolution product is used. This choice does come at a cost. The lower resolution product does not include smaller water bodies, and more importantly, cannot define the coast in the same amount of detail. As such, extra steps are taken in this project regarding the classifications along the coast, which is especially important where the land is covered with snow and no sea ice is present in the adjacent cell. An example of the 4-km product for February 1, 2012 is provided in Figure 3.3, where one can see the large number of water

Figure 3.3: IMS based snow and ice coverage for February 1, 2012.

bodies that are present because of the greater resolution.

A given time step is created by analysts who take the previous day's coverage and redefine it with the latest information. The changes are based on a variety of sources, including geostationary satellites, polar orbiting satellites, station data, and knowledge of the region's climatology. The preferable and most common source is visible imagery from polar orbiting and geostationary satellites operated by Europe, Japan, and various departments of the United States (Helfrich et al., 2007). Under cloudy conditions a combination of modelled results and observations are used, such as those from the Snow Data Assimilation System (SNODAS) and METAR (Helfrich et al., 2007).

Part 4 | Error quantification

Before assimilating satellite rainfall products with observations using CaPA, it is worthwhile assessing their performance over Canada. For both CMORPH and PERSIANN a high and low resolution product is tested. In the case of PERSIANN, the high resolution product (PERSIANN-CCS) has an altered methodology, so the differing results are dependent on more than just the resolution. The CMORPH resolutions are 0.25° and 8 km, while those for PERSIANN and PERSIANN-CCS are 0.25° and 0.04°, respectively. The results from this aid in both understanding the product's limitations in time and space, and in explaining the relative changes in skill as a result of using them with CaPA. To quantify the error and performance, a variety of techniques are detailed in the following sections alongside the results. Following information regarding the quality check performed on the satellite data, biases and categorical skill scores are explored through space and time. Then, the total error is decomposed, detailing origins of the error and their relative contribution.

It should be noted that all evaluations in this section are using the raw satellite rainfall data as they are made available. No adjustments have been made to the magnitude of estimates, however those over snow covered ground have been removed using the process described in Section 5.2. The spatial domain of the satellite products remains unchanged throughout the project, extending from 60°N to 30°N and 180°W to 30°W as shown in Figure 4.1. Unfortunately, the upper limit is defined by the data availability, which is the same amongst all of the products. The other extents were selected conservatively to reduce potential edge effects in future analyses and/or interpolation methods (see Section 5.3.1).

Figure 4.1 also shows two red lines placed at 115.5°W and 95°W which are used to define

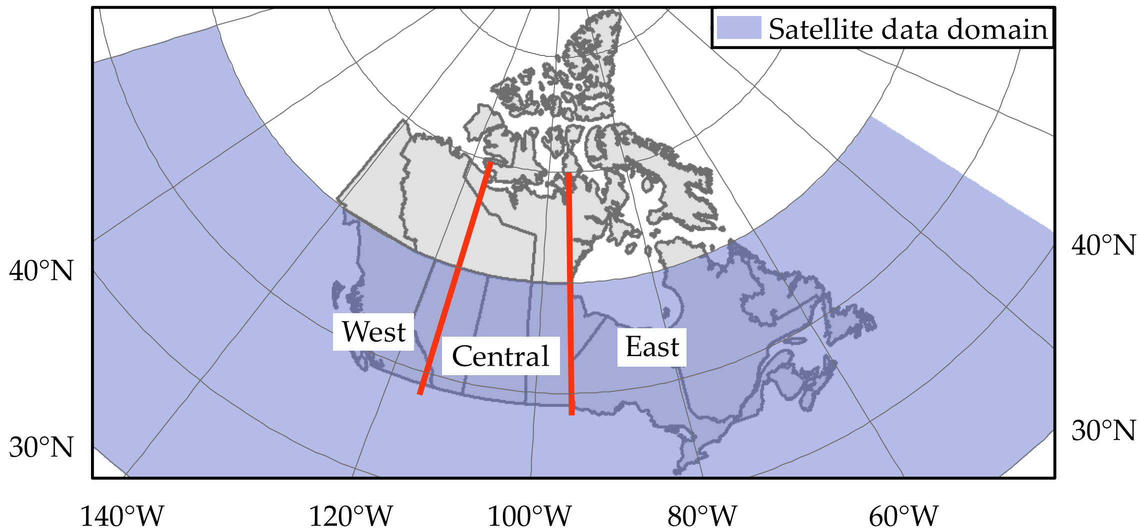


Figure 4.1: Spatial domain of the satellite rainfall products and definitions of western, central, and eastern Canada.

western, central, and eastern Canada throughout this project. These bounds are selected to align with the prairies in terms of longitude, as defined by the climate regions from Environment Canada (2014) and Mekis and Hogg (1999), which represent areas with similar characteristics. This definition also limits the Rocky Mountains to western Canada.

The temporal domain used here is unique to this part, as this is the only one which uses APC2. Specifically, it ranges from 2001 to 2010 and only considers June to October, where the monthly selection is due to limitations of the satellite products with snow cover (see Section 2.3 or Figure 5.4). These years are selected based on the availability of data between the satellite products and the observed data. The lower bound is due to the PERSIANN 0.25° product, while the upper bound is chosen to allow an analysis for a decade. PERSIANN-CCS data, however, are only available from 2003 onward. For the bias and categorical scores in Section 4.2, all available data for each dataset is used as it is an evaluation of the datasets themselves. The error decomposition in Section 4.3 is more of a relative comparison between products so it only considers data available amongst all of the datasets, 2003 to 2010. This decade is used in particular because is it the oldest the satellite products will allow and the number of stations in APC2 decreases with time (Mekis and Vincent, 2011).

Why would one use an observational dataset with decreasing amounts of data? The fundamental principal of quantifying error is that the reference dataset is taken as truth. For the reasons discussed in Section 3.4 APC2 is the best estimate of precipitation available for Canada, so it is used instead of the raw station data. It is, however, important to acknowledge the difficulty of measuring precipitation and the following analysis is entirely dependent on the accuracy of APC2.

Since APC2 is a daily precipitation product, the satellite based rainfall products are accumulated to daily measurements. To accomplish this, an accumulation method similar to that used for PERSIANN internally before public release is applied. The only difference is in the thresholds for missing data. In the data used here, if at least three-quarters of the pixels being accumulated are present, the accumulated precipitation, P, can be found by Equation 4.1, where the known values are scaled to compensate for the missing data. Otherwise the pixel is considered to be missing. Although this may be introducing some inaccuracies, it was implemented primarily due to the high resolution products. Requiring a pixel to have estimates for all accumulated time steps is quite strict since these daily accumulations are from 30 minute and 1 hour estimates. For the internal PERSIANN accumulation, the rule of thumb is to only accumulate if two-thirds of the pixels have data (Braithwaite, 2013). In Equation 4.1, i is the number of time steps to be accumulated, n is the number of time steps with data, and p_k is the precipitation of the pixel at time step k.

$$P = \frac{i}{n} \sum_{k=1}^{i} (p_k) \tag{4.1}$$

4.1 Quality control findings

While Equation 4.1 works fine when missing data in the satellite products is, for the most part, random in space and time, a problem arose with a reoccurring artefact in the PERSIANN data. Specifically, the same area was found to routinely be missing data. A conservative example of this is shown in Figure 4.2 where a group of grid cells are missing on May 22, 2001. Through visual inspection of raw infrared data, this region is a gap in the mosaic of infrared data if some

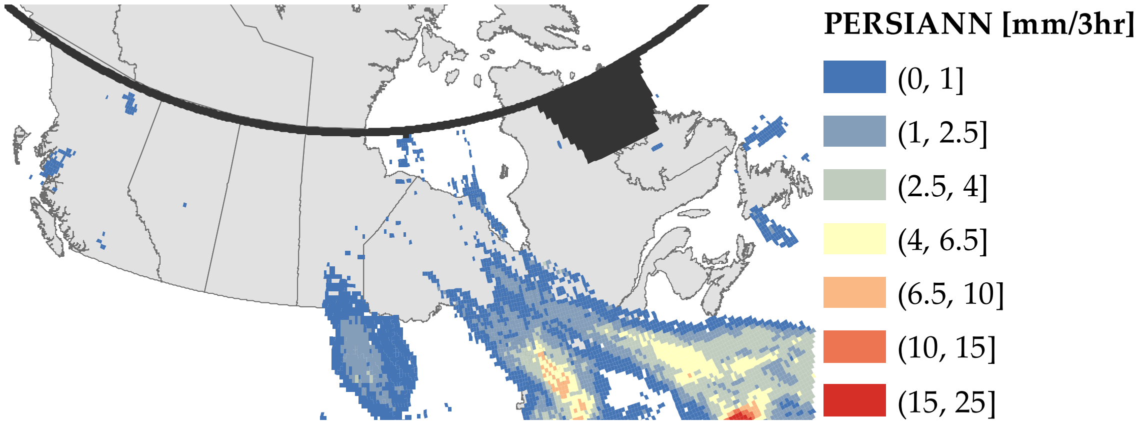


Figure 4.2: PERSIANN 0.25° precipitation estimate for May 22, 2011, where black indicates missing data.

satellites do not report any data, which would directly effect the results from the PERSIANN algorithm, as infrared data alone are used as input. When an infrared image is missing for CMORPH, a spatial interpolation of vectors is performed (Joyce et al., 2004). While this will not produce missing data, it may decrease performance in this region.

To determine the frequency of these events, the number of grid cells with missing data within 80°W, 55°W, 30°N, and 60°N is found for every 3-hourly time step and plotted in Figure 4.3. It is clear that CMORPH is unaffected by this artefact, only having a few time steps for which noticeable data were missing.

PERSIANN however, contains more missing data. The quantity of time steps missing all data in Figure 4.3a may be artificially greater than from the raw data because it was found that some PERSIANN files contain all zeros, presumably either misclassified missing data or an issue with the PERSIANN algorithm itself. Either way, these time steps were detected and considered missing after verifying them with the accompanying image files.

Regarding the artefact over eastern Canada, there is a clear signal in Figure 4.3a where numerous time steps experience the same portion of missing data (roughly 10%). Those time steps with larger portions of missing data are due to the missing data shown in Figure 4.2 continuing southward to form a point. To handle this, missing data in each time step is quantified with

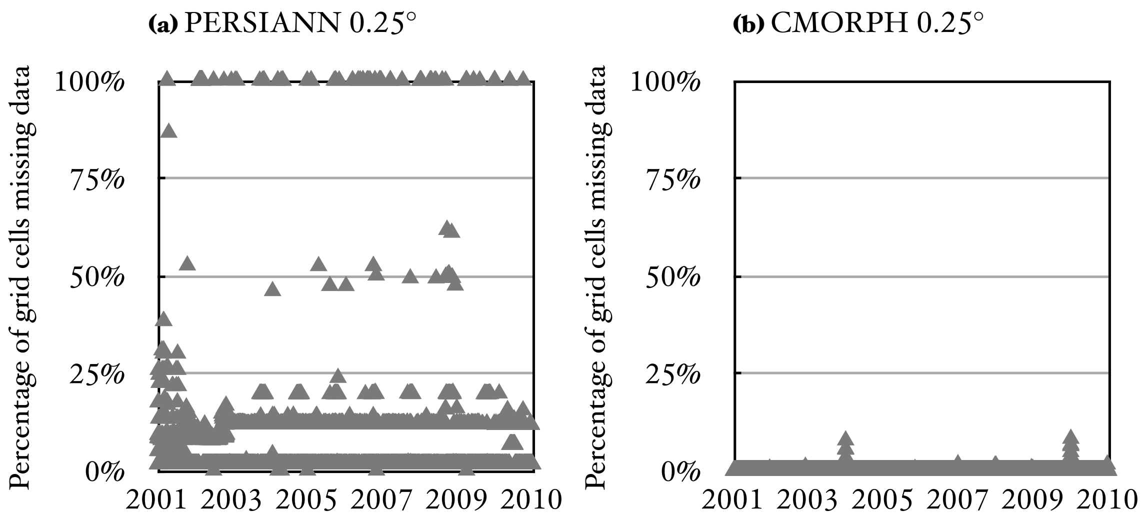


Figure 4.3: Percentage of grid cells missing data between 80°W, 55°W for each 3-hour time step within June through October, 2001 to 2010.

between 80°W, 55°W, and if the portion of missing data matches that signal, Equation 4.1 is not applied within the larger of the problematic regions.

A final note specific to the PERSIANN 0.25° product is that on occasion, in error, it will estimate large amounts of precipitation around the globe. Specific to the domain of this project, this occurred in November of 2009. Kidd et al. (2011) also acknowledged this occurring over Europe. To find these errors, one can simply accumulate precipitation over the domain and notice a large anomaly.

4.2 Bias and categorical scores

Few evaluations of CMORPH and PERSIANN over Canada exist, and most information is in the form of generalizations for high latitudes or cold climates. The results from this section will then need to be compared to the conclusions from other works that can be transferred to Canada with some degree of certainty. What type of information can be transferred? The most applicable is the performance of the satellite precipitation products for specific types of precipitation. For example, performance can change depending if the event is convective, stratiform, light, heavy, or warm. Also, sharing a border with the United States, which is well studied in this regard, allows for one to check for continuity of results along the border.

In evaluating the satellite products over the United States, there are trends in the spatial and temporal performance, consistent between multiple authors. In terms of bias, CMORPH and PERSIANN have been shown to underestimate precipitation in the winter, with mixed results in the summer (Sapiano and Arkin, 2009; Tian et al., 2007). The underestimation in the winter is attributed to the issues with the detection of precipitation over snow or ice covered ground (see Section 2.3) or even snowfall over bare ground (Munchak and Skofronick-Jackson, 2013).

When looking more closely at the spatial patterns of performance, it is found that central United States overestimates precipitation, while the northwestern and northeastern regions underestimate (Tian et al., 2007; Ebert et al., 2007). It is worth noting that Sapiano and Arkin (2009) has shown in the Great Plains, that while CMORPH and PERSIANN have a positive bias in the summer, they also possess greater skill with convective storms. This increase in skill is also observed by Tian et al. (2007) on a larger spatial domain. The increase in skill is important because a given satellite product can overestimate precipitation or possess less bias, but if it is not also increasing the skill, it is not as meaningful. Ebert et al. (2007) suggests that the overestimation is due in part to gauge analysis missing short and intense convective cells, since the satellite products agreed well with radar data, which are spatially continuous.

Beyond central North America, Tian et al. (2007) observed that in general the correlation of CMORPH is greatest in eastern United States. Areas of decreased correlation are attributed to missing stratiform, warm, and/or shallow rainfall, due to the algorithms relating satellite retrievals to rainfall rates having detection difficulties, also found by Ferraro et al. (2000).

While evaluations over the United States have given insight into the performance of specific precipitation regimes and patterns over North America, considering the performance over Europe may be just as important. Even though it does not have the proximity of the United States, it does share the issues Canada has with satellite products at mid to high latitudes. Kidd et al. (2011) detail many of the problems with satellite products at these high latitudes, where issues involve the low intensity events, mixed precipitation, small amounts of stations, and strong annual fluctuations in climate between seasons. Light rainfall is particularly troublesome in that the satellite products are currently tuned for precipitation near the equator with warm climates and moderate to high rainfall. The majority of the precipitation falling at mid to high latitudes is relatively light, so where this type of precipitation would contribute small errors near the equator, the errors are much greater at higher latitudes.

In terms of information that may be transferred to Canada other than the general underestimation and missing light precipitation, Kidd et al. (2011) and Ebert et al. (2007) too found that CMORPH was the best amongst satellite products, coming from an increased probability of detection and similar false alarm ratios relative to the other products. Similar to that over the United States, Kidd et al. (2011) also found an overestimation due to convective events and issues with an underestimation of precipitation on the windward coast of Europe.

Over Australia, Ebert et al. (2007) found that the satellite products underestimate the moist onshore flow, attributing it to this type of rainfall having warmer cloud tops, and therefore less ice particles to cause ice-scattering (see Section 2.3). The authors found similar performance of the satellite products over Australia as are present elsewhere in the world, specifically issues detecting non-convective precipitation and CMORPH having the highest correlation. Uniquely, Ebert et al. (2007) suggest that some of the issues regarding convective storms may be due to the sparse temporal resolution of the passive microwave scans, in that they are not able to properly capture the development of some fast growing storms.

4.2.1 Methodology

The evaluations performed in this section are of two forms: bias and categorical scores. The biases calculated here are for June through October from 2001 to 2010, and are either in the forms of differences or ratios. Equations 4.2 and 4.3 define the two types as they are used throughout the rest of this project. Given these definitions, a positive difference and a ratio less than one represent an overestimation. The ratios are defined in this way to be consistent with the adjustment methodology presented in Sections 5.3.2 and 5.3.1.

$$\text{Bias} = \text{Satellite} - \text{Observed} \tag{4.2}$$

$$\text{Bias Ratio} = \frac{\text{Observed}}{\text{Satellite}} \tag{4.3}$$

While it may seem unnecessary to use both differences and ratios, they show different aspects of bias. In taking the difference, the resulting bias is relative to the normal precipitation of that point or region. For example, a station which receives more precipitation is capable of possessing a greater bias compared to a stations which generally receives less precipitation. When using ratios, the normal precipitation for a given location is less relevant because the ratio is describing the overestimation or underestimation relative to that area.

Considering the datasets used in this project and discussed in Part 3, they are of different formats. The satellite products are in the form of gridded data, continuous in space and of varying resolutions, while the observational datasets are point observations, representing the stations. To evaluate these two different formats against each other, each station is paired with its co-located grid cell. Then each pair is classified as a hit, miss, false alarm, or non-event based on Table 4.1. For this section, each station is paired with its co-located grid cell and then classified, regardless of whether multiple stations share a grid cell. Given the general low station density of APC2, no stations shared grid cells for the 8-km or 0.04° products. For the 0.25° products, five stations (roughly 1.3%) share grid cells with another.

It is important to note that this method cannot fully capture the performance of the gridded product. With the gridded product, the estimate is an average of the precipitation over the grid cell, while the station is a point observation. One can then image a scenario where the

Table 4.1: Classification of categorical scores for a given precipitation threshold T or bin B.

		Observation	
		$\geq T \quad (\in B)$	$< T \quad (\not\in B)$
Forecast	$\geq T \quad (\in B)$	Hit (H)	Miss (M)
	$< T \quad (\not\in B)$	False Alarm (F)	Non-event (N)

maximum of an event is not fully realized in the gridded product, meanwhile a station could (assuming it is in the appropriate location and the measurement is of minimal/no instrumental error). Based on this, one may expect some degree of underestimation in comparing a gridded product to a point observation. This also occurs in comparing the different resolutions, since the lower resolution is not be able to produce the high magnitude events, especially since they usually occur over small areas.

There is also the issue of spatial representativeness. Gridded products allow one to see precipitation features which cannot be seen by stations, unless the station network is sufficiently dense. So in comparing a station to its co-located grid cell, it is possible that the gridded product is matching the spatial features of the event very well, but is offset slightly in space, and therefore receives a low score. This is especially true for datasets with high temporal or spatial resolutions (Rossa et al., 2008).

Two different definitions for this classification are presented in Table 4.1: those based on a threshold T, and those based on a bin B. For thresholds, the classifications represent increasingly extreme events. Depending on the spatial and temporal domain being considered, the maximum threshold has generally been set as 25 mm/day or 50mm/day for the comparisons to APC2, and 25 mm/6-hour or 50 mm/6-hour for the CaPA assimilations. These upper thresholds are based on the associated confidence intervals, where the decrease in frequency of extreme events results in large amounts of uncertainty and therefore less meaningful results. When one is measuring skill scores by rainfall thresholds, it is possible that a sudden change in classifications for higher thresholds influences all of the scores below. This is not the case for binned data; however, the classifications are dependent on the size of the bins. In this project, the bins have been chosen to be consistent with those currently used by Environment Canada, so comparisons with existing works can be made.

The classifications resulting from Table 4.1 for a given bin or threshold can be used to calculate numerous scores. In this section, four categorical scores are used to describe the performance of the satellite products: equitable threat score (ETS), probability of detection (POD), false alarm rate (FAR), and frequency bias index (FBI). All of the categorical scores calculated here are done by considering all stations for all days in one calculation resulting in no averaging, in comparison to calculating a score for each day and averaging through time.

ETS is calculated using Equation 4.4 and is a measure of correctly forecasted events (hits) to all others, less those from random chance (Equation 4.5). Given that it is an overall score, it is used from here on to quantify the skill of the satellite products or the analyses from CaPA. So, when skill is stated, it is referring to the relative changes in the equitable threat scores. Possible values range from negative one-third to positive one, with one being a perfect score, zero implying no skill, and less than zero implying the analysis is worse than random.

FBI is a ratio of the forecasted event frequency to those observed, and is defined by Equation 4.6. With this definition, above zero is an overestimation of events and below zero is an underestimation. Equation 4.7 is used to calculate POD, which is a ratio describing the amount of observed precipitating events detected by the forecast, ranging from zero to one, where one is a perfect score. Lastly, FAR is a ratio of false alarms to non-precipitating days as defined by the observations, and is calculated used Equation 4.8. While an ideal FAR is zero, caution should be taken at the larger thresholds or bins because the majority of classification will tend to be non-events, resulting in a low FAR, but not necessarily great overall performance by the forecast.

$$ETS = \frac{H - H_r}{H + F + M - H_r} \tag{4.4}$$

$$\text{where,} \quad H_r = \frac{(H + F)(H + M)}{H + M + F + N} \tag{4.5}$$

$$FBI - 1 = \frac{H + F}{H + M} - 1 \tag{4.6}$$

$$POD = \frac{H}{H + M} \tag{4.7}$$

$$FAR = \frac{F}{F + N} \tag{4.8}$$

4.2.2 Results

To give an overall impression of the bias present in the satellite products, Figure 4.5 shows the mean monthly bias from June to October, calculated from 2001 to 2010 (2003 to 2010 for PERSIANN-CCS). As mentioned in Section 4.2.1, the magnitude of the bias is related to the normal precipitation for the station, meaning that this is influenced greatest by regions of large normal precipitation, namely western Canada, and eastern Canada to a lesser extent as shown in the normal precipitation values in Figure 4.4. Also, station density is assumed to be constant, which mostly is the case for the southern portion of the projects domain (except Ontario) as shown previously in Figure 3.1.

Caveats aside, Figure 4.5 shows the trends in bias for Canada. In being such a generalized plot, it is difficult to confidently infer sources of error, so that will be held until Section 4.3. Regardless, the overall underestimation of rainfall from the satellite products is clear and consistent with the previously discussed works. Sohn et al. (2009) studied these products over the Korean Peninsula and also found them to underestimate. The author specifically noted that the GPROF database, which relates passive microwave readings to rainfall, does not represent mid-to-high latitude precipitation, and therefore produces biased results.

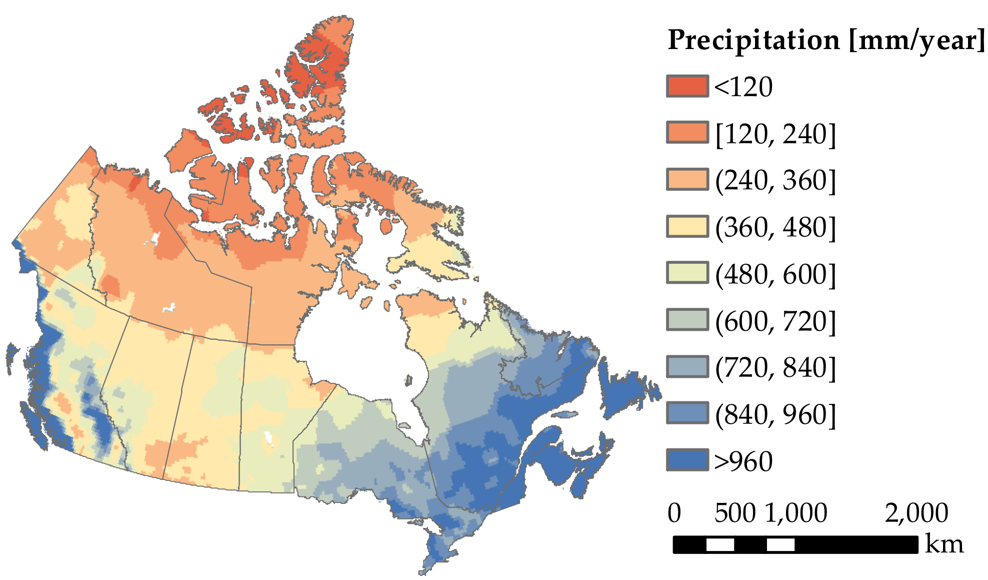


Figure 4.4: Normal annual precipitation considering 1951 to 2000 based on GPCC data (Meyer-Christoffer et al., 2011).

At the monthly scale, Figure 4.5 shows that the satellite products generally behave similarly, especially for the two CMORPH resolutions, which produce nearly the same trend in bias. For PERSIANN, however, the difference in the resolution and algorithm has a clear impact on the bias. Among all products, there is a reduction in negative bias in the beginning of the summer. This reduction in July and August is most likely due to the strong change in central Canada, shown in Figure 4.6 and discussed in greater detail in Section 4.3.2.

In comparing the bias of CMORPH and PERSIANN, CMORPH usually possesses less bias. The exception is during October where PERSIANN is better than CMORPH in this regard. This is not the first time that this has been observed around the world, as Kidd et al. (2011) has shown that PERSIANN overestimates precipitation over Europe in the September - October - November and December - January - February seasons compared to CMORPH.

PERSIANN's reduction in negative bias is most likely due to the performance of infrared data with mixed precipitation and snow cover. But why would PERSIANN's performance be more closely related to infrared than CMORPH? Even though they both use infrared and passive microwave data, CMORPH's rainfall magnitudes are solely based on passive microwave data, while PERSIANN uses infrared as input in its neural network. This is why some authors refer to CMORPH as a passive microwave product and PERSIANN as an infrared product.

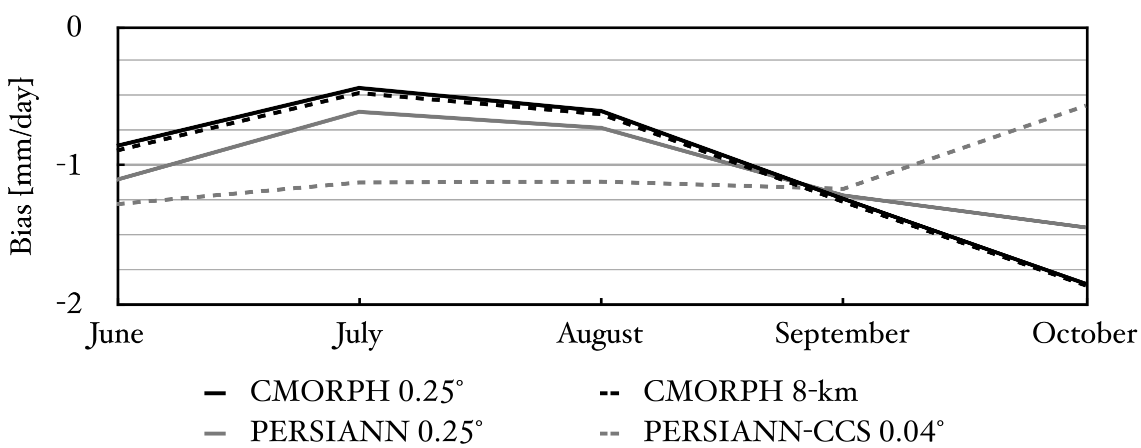


Figure 4.5: Monthly basis from 2001 to 2010 for all stations in APC2, where bias is defined as satellite less observed.

Regarding the issue of estimates over snow covered ground, Ebert et al. (2007) has shown how infrared data products can appear to be beneficial over mountainous regions in winter as the frequency of events increase, and thus the bias too. Although estimates over snow cover have been removed for this project, mixed precipitation in the fall remains, which is most likely the cause. The increase in bias in October for PERSIANN may then be a result of mixed precipitation. The increase in frequency that Ebert et al. (2007) discuss with infrared products can also be seen in October in central Canada in Table 4.3 where the PERSIANN products shift from error in hits and misses to false alarms. The formation and purpose of this table is discussed in Section 4.3.

As previously discussed, convective events are a common cause for overestimation in CMORPH and PERSIANN. Figure 4.5 shows the potential for this to be the case for Canada as well, with a noticeable reduction in negative bias for July and August. To further test this, bias has been explored in both space and time. Mean monthly ratios have been calculated for June, July and August from 2001 to 2010 and are presented in Figure 4.6. For each station in APC2, observed precipitation is accumulated to monthly values along with its co-located grid cell. A ratio is taken between the two (Equation 4.3) for each year and then averaged over the ten years. The result is a mean ratio for each station between APC2 and a satellite product. For better presentation and handling of outliers the ratios are smoothed using a thin plate spline across the project's domain, resulting in Figure 4.6. The stations included on the maps are those used in defining the thin plate spline (ie. had enough data), and can be used to see areas of extrapolation. The 0.25° products are shown here, but the same methodology was applied to the high resolution products, and as Figure 4.5 implies, PERSIANN-CCS does not show any spatial trends. For reference, plots of the high resolution product are shown in Figure B.1.

Considering Figures 4.5 and 4.6 together, it is quite clear that the reduction in negative bias in June, July, and August is a result of an overestimation in the Canadian Prairies during the same months which are prone to convective conditions (Brooks et al., 2003). In terms of the frequency of observed convective events in the Canadian Prairies, Raddatz and Hanesiak (2008) identified convective events with significant rainfall accumulation and lightning data.

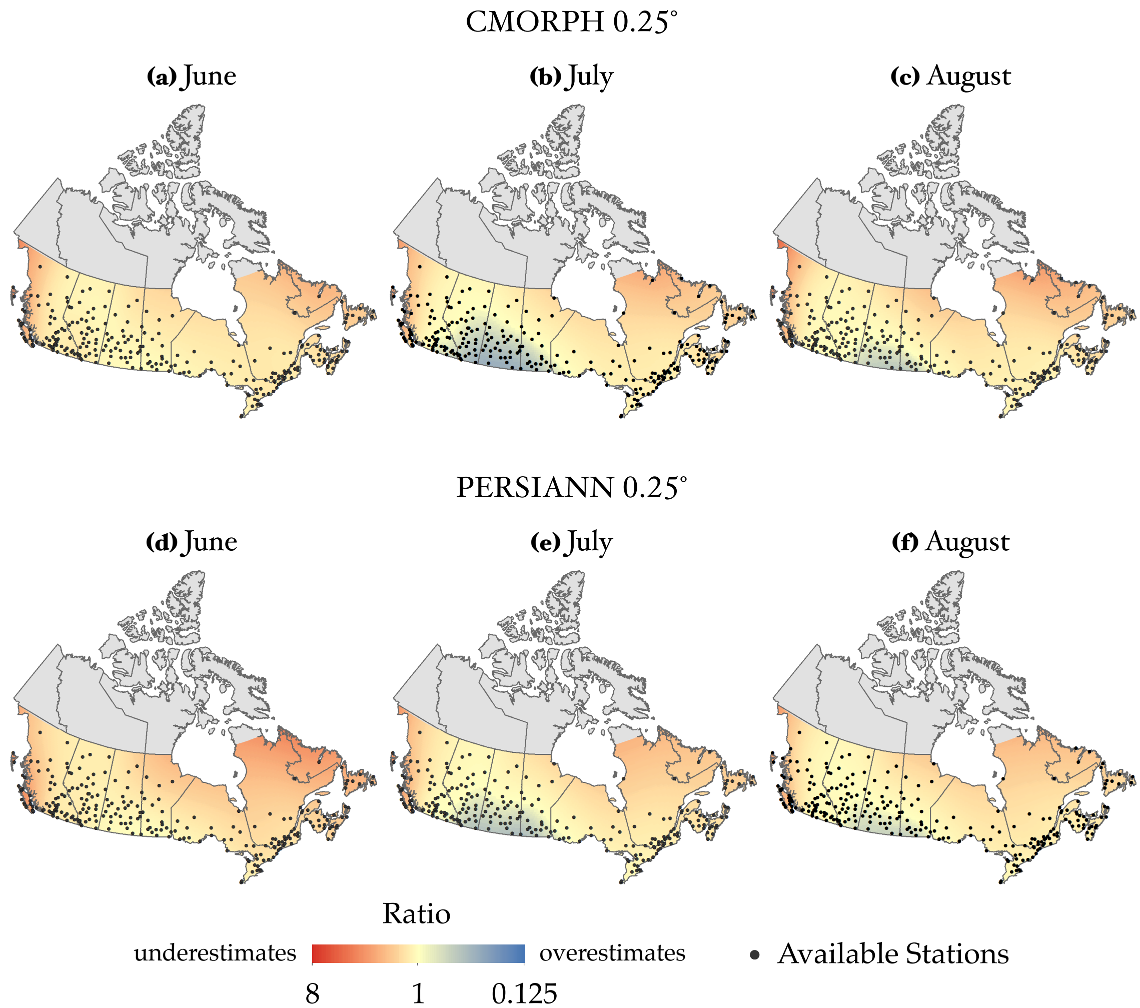


Figure 4.6: Mean monthly ratios of observed/satellite between the 0.25° CMORPH and PERSIANN products and APC2 from 2001 to 2010. Applicable stations are shown in black.

The authors found that from 2000 to 2004, "in June, 74% of the events were solely or partially convective, in July 85%, and in August 79%" (p. 1610). These monthly frequencies match very well with the relative degrees of overestimation in Figure 4.6, in that they are strongest in July followed by August. It then seems that these overestimations are due to the presence of convective storms and increased ice-scattering, as noted elsewhere in the world.

While bias provides some information regarding the satellites products' magnitudes of estimation, equitable threat score measures their predictive capability. Similar to before, this is explored in space and time based on both threshold and binned values from 2001 to 2010 (2003 to 2010 for PERSIANN-CCS) using Equation 4.4 and the previously discussed methodology of Section 4.2.1. For precipitation above 1 mm/day across Canada, Figure 4.7 shows that equitable threat score is for the most part in the range of 0.2–0.3. Over Australia, Ebert et al. (2007) evaluated equitable threat scores for multiple satellite products and models during 2004 to 2005. The authors found the median equitable threat scores for CMORPH and PERSIANN to be roughly 0.25 and 0.35, respectively, during the December-January-February summer season. For precipitation exceeding 20 mm/day, the scores presented here are between the 50^{th} and 75^{th} percentiles defined by the authors. Hence, the equitable threat scores presented here are not unique to this project.

In Section 5.1, the same data are also used to produce probabilities of detection and false alarm rates. From the plots discussed in that section, one can conclude that the equitable threat scores are more closely related to the probabilities of detection as opposed to false alarm rates, in that the false alarm rates are constant while probability of detection changes with equitable threat score. While Kidd et al. (2011) did not show this exactly, one can infer from authors' graphs that the skills of PERSIANN and CMORPH are also more closely tied to probability of detection, since, as was the case here, the skill scores and probabilities of detection change while false alarm ratios are mostly constant.

Spatially, in Figure 4.7, the ranking of satellite performance remains unchanged, with the 0.25° CMORPH product at the top. Similar to the correlations calculated by Tian et al. (2007), there is better agreement with the satellite products on the east coast and decreasing westward. As mentioned previously, problems in western Canada may be related to those seen in Australia and Europe, where windward coasts tend to underestimate precipitation due to missing stratiform events (Ebert et al., 2007; Ferraro et al., 2000; Kidd et al., 2011). Orographic effects may be at play as well. The binned equitable threat scores in Figure B.3 provide some additional insight into the high and low skill scores in eastern and western Canada, respectively. For all of the products, the equitable threat scores below 10 mm/day are essentially the same for all regions. It is only above this magnitude that they start to vary.

On a monthly basis, the difference in binned scores in Figure B.2 are difficult to see, so these plots are delegated to the appendix and plots based on threshold values are presented here. Figure 4.8 shows that a common trend between the two CMORPH resolutions and the 0.25° PERSIANN product is that skill increases to peak in July and August, and decreases afterwards. Amongst all the satellite products, the 0.25° CMORPH product continues to have the best results, still consistent with the works discussed at the beginning of this chapter.

While the 0.25° CMORPH product has the greatest skill, it also has the greatest variability month-to-month, caused by the relatively low skill in October. Interestingly, PERSIANN

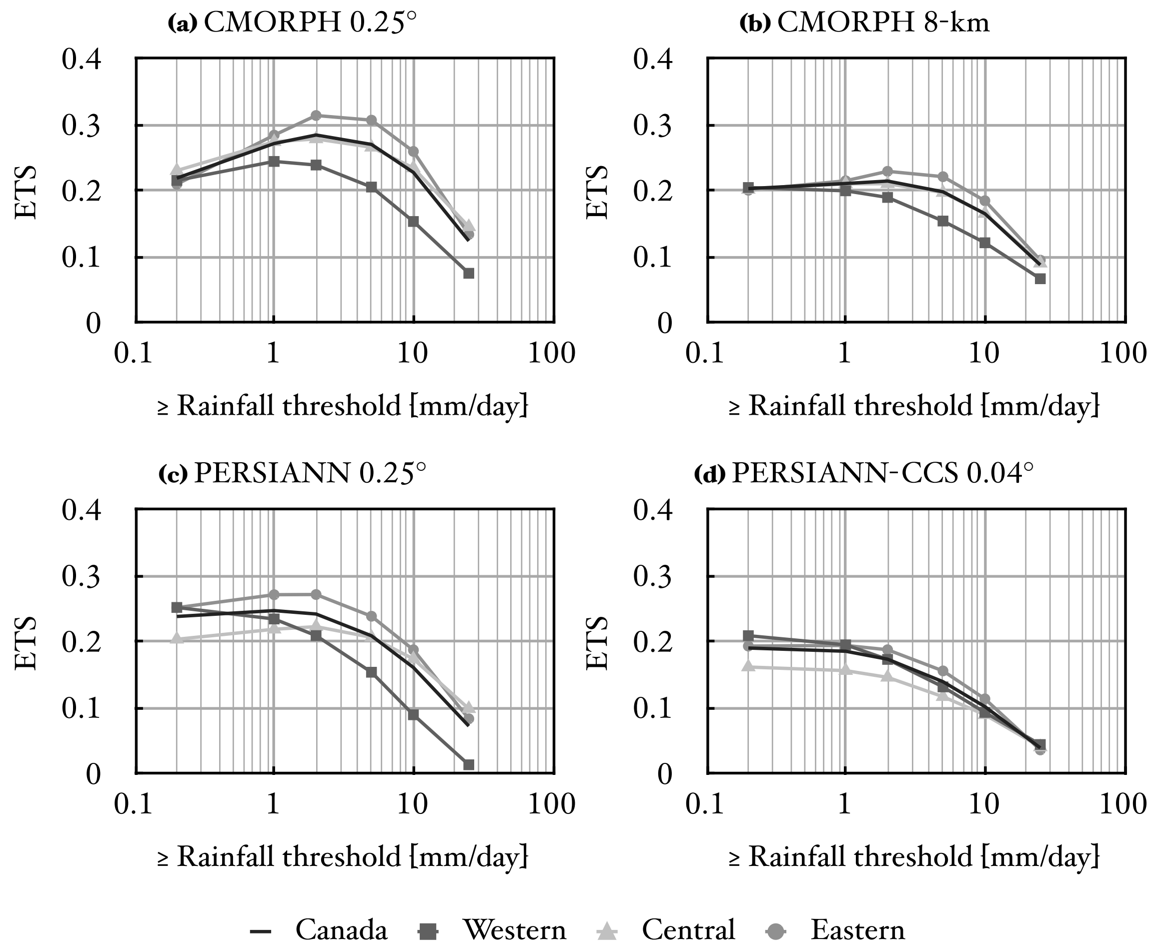


Figure 4.7: Regional equitable threat scores of the satellite rainfall products versus APC2 observations for June through October, 2001 to 2010 (2003 to 2010 for PERSIANN-CCS).

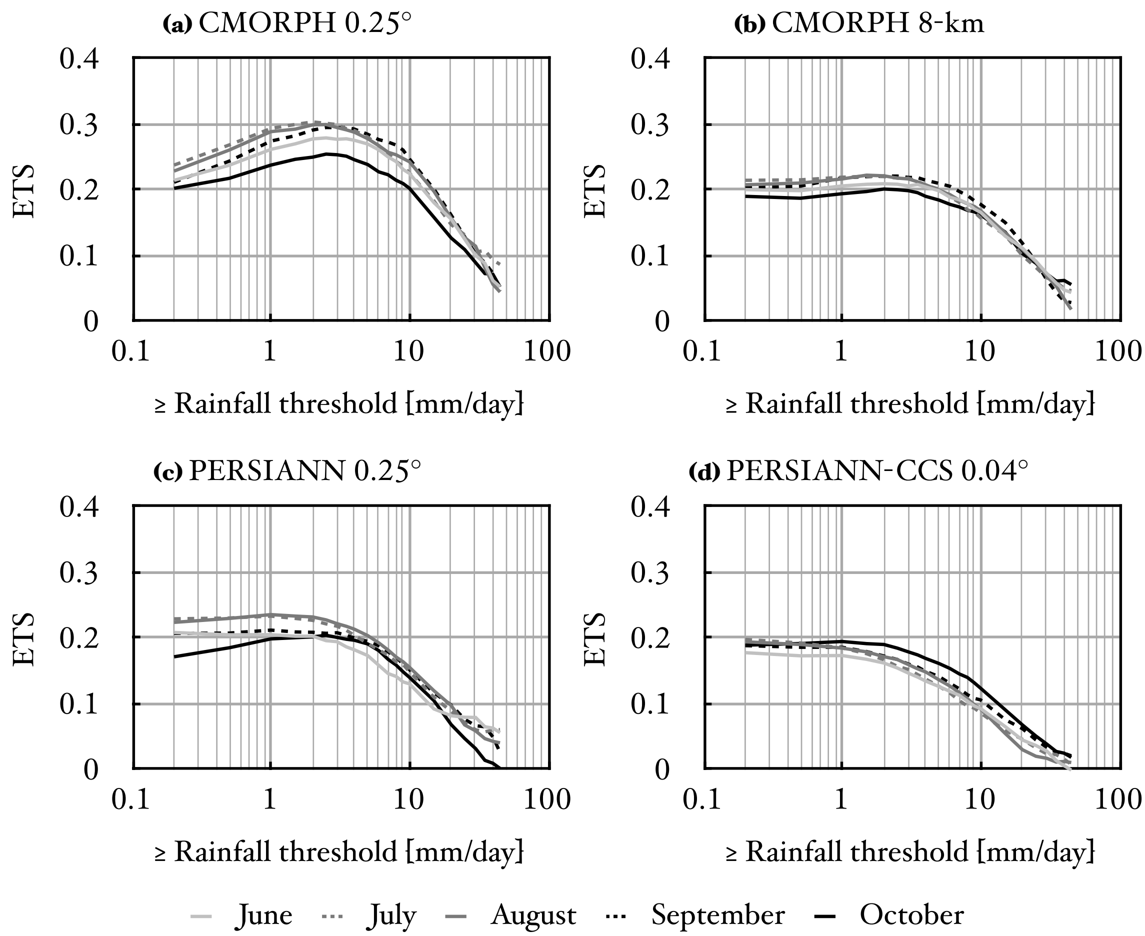


Figure 4.8: Monthly equitable threat scores of the satellite rainfall products versus APC2 observations across Canada from 2001 to 2010 (2003 to 2010 for PERSIANN-CCS).

does not have a reduction of skill in October, consistent with the discussion earlier in this section regarding the capabilities of infrared products with mixed precipitation and some degree of snow cover. Meanwhile, PERSIANN-CCS product does not perform well at all.

An interesting attribute of the CMORPH 0.25° product is that equitable threat scores as function of binned rainfall or rainfall thresholds have a peak, as opposed to the 0.25° PERSIANN product which never experiences this. Figure 4.9 shows frequency biases for the satellite products relative to APC2 across Canada from 2001 to 2010. The frequency bias indices are calculated using Equation 4.6, where zero is a match between observed and satellite event

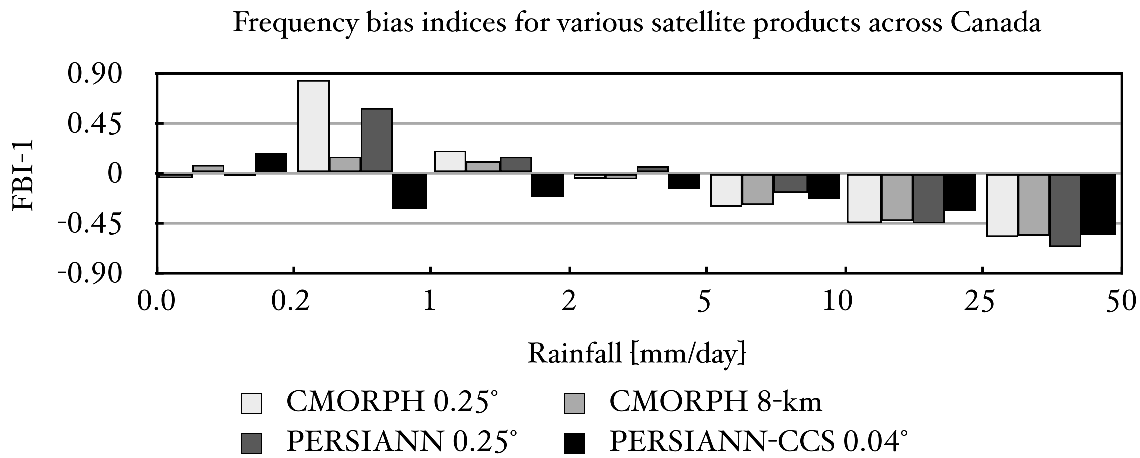


Figure 4.9: Frequency bias indices less one for the satellite products relative to APC2 for stations across Canada from 2001 to 2010 (2003 onward for PERSIANN-CCS).

occurrence. Figure 4.7 shows that for rainfall above 2 mm/day with the 0.25° CMORPH product, it is at peak performance. The frequency bias of CMORPH above 2 mm/day in Figure 4.9 begins to become negative, meaning it is underestimating occurrence. All satellite products are similar for rainfall between 2 and 50 mm/day, but it is between 0.2 and 2 mm/day that they behave differently. Comparing the 0.25° CMORPH and PERSIANN products, CMORPH overestimates the frequency of these small events to a greater degree than PERSIANN, which is coupled with a negative effect on CMORPH's ability to predict light rainfall. For rainfall between 2 and 50 mm/day, the frequency biases of these two products are similar in that neither is consistently worse, however CMORPH always has an improved equitable threat score (Figure 4.7 or Figure 4.9), speaking to its superiority over PERSIANN.

Figure 4.9 also allows for one to compare the effects of resolution, particularly the CMORPH products as the only difference is the resolution. The largest difference between the two is for light rainfall (0 to 2 mm/day) where the lower resolution overestimates the frequency of events much more than the high resolution product. This is somewhat intuitive as the occurrence of light rainfall, either at the edge of larger systems or small and light events, can be more accurately delineated with the high resolution product, since rainfall in a portion of a larger grid cell inaccurately implies rainfall over the larger area.

4.2.3 Key points

- All satellite products underestimate precipitation across Canada.
- Bias improves from June to August in the Canadian Prairies, in accordance with convective storms to the degree that it affects the bias found across Canada for all months considered.
- Equitable threat scores across Canada peak in July and August.
- Equitable threat scores are the best in eastern Canada and decrease westward.
- Varying skill scores in eastern and western Canada are dependent on rainfall greater than 10 mm/day.
- The 0.25° CMORPH product performs the best across Canada.
- The decrease in skill of CMORPH for light rainfall is accompanied with a large overestimation of rainfall occurrence.
- The 8-km CMORPH product does not overestimate light rainfall occurrence as much as its lower resolution counterpart.

4.3 Error decomposition

Considering Figure 4.6b for the ratios of CMORPH during July, the overestimation is clear, but does not fully describe its nature. Using the classifications defined in Table 4.1, this overestimation can be from either hits or false alarms. Depending on the origin, hits or false alarms, the performance of the satellite is very different. In the case of false alarms, the satellite product would not be producing anything meaningful, but for hits, the satellite product could be performing very well and simply have a bias, which is much easier to deal with. Therefore, the first portion of this section quantifies the errors associated with the different error components, hits, misses, and false alarms.

Tian et al. (2009) performed such an analysis over the United States and found that hit bias and missed precipitation are the largest contributors to the overall bias. As the author states, "in summer, positive hit bias, up to 50%, dominates the total errors for most products" (p. 1), two of which are CMORPH and PERSIANN. Considering the authors' plots, the positive hit bias is in central United States, and given the cited literature, presumably the result of convective events. Spatially, the area of satellite-based precipitation affected by convective events in Canada (Figure 4.6) is much less than that of the United States shown in the previously discussed literature, so the dominance of positive hit bias presented in this project is expected to be less. In eastern United States, there is a large negative hit bias coupled with missed precipitation. False alarms are generally present across the continent. While the trends for CMORPH and PERSIANN are similar in this analysis, CMORPH has less error from false alarms. From the analysis presented here, one can expect Canada to have a positive hit bias in Central Canada, missed precipitation in western and eastern Canada, and a fairly constant distribution of false alarms.

To take this concept one step further, the second portion in this section uses the methodologies of Tian et al. (2013) and AghaKouchak et al. (2012) to separate the hits component into random and systematic error. The results from this analysis quantify the uncertainties of the satellite products, and the potential effectiveness of a bias correction or magnitude adjustment.

AghaKouchak et al. (2012) found that in the summers of 2005 to 2007, CMORPH has more random error than PERSIANN in central United States. In western United States, data are sparse because of the Rocky Mountains, but the error is mostly systematic in nature. Finally, in northeastern United States, CMORPH is fairly well balanced between systematic and random error, while PERSIANN is dominated by systematic error. While AghaKouchak et al. (2012) derived systematic and random error from a linear relationship, rainfall errors are proportional to the magnitude of precipitation, so a linear relationship may not be the best model to use.

Prior to separating hit errors into systematic and random components, the methodology defined by Tian et al. (2013) is used to test a linear and non-linear model, where the non-linear model's error component scales with the observed precipitation. The author tested a linear and non-linear model in Oklahoma from 2005 to 2007 with data from TMPA, whose raw data are the passive microwave component of CMORPH and PERSIANN. The author found that these data are, not surprisingly, better represented by the non-linear model, in that the residuals possessed much less systematic change.

4.3.1 Methodology

Quantifying the error from hits, misses, and false alarms is done in three ways: percent of component error, bias, and frequency. Note that component error is purposefully different than total error. At least in this project, total error refers to differences between the satellite products and the observations. In this definition false alarms and misses cancel out with each other, and with portions of the hit bias, meaning total error is not the sum of the components. Thus the term component error is used, defined as the accumulation of absolute differences between the satellite products and observations, as shown in Equation 4.12.

To calculate the percent contribution to component error, each observational and satellite pair in space and time is classified using Table 4.1 with a threshold just above zero. Specifically, events are recognized if any magnitude of precipitation greater than zero is present. The hit, miss, and false alarm components are then accumulated in space s and time t by Equations 4.9 to 4.11 respectively, and divided by the component error, E_{comp}, to find their contribution.

$$E_H = \frac{\sum_{s=1}^{n} \sum_{t=1}^{m} (|H_{s,t}|)}{E_{comp}} \cdot 100 \tag{4.9}$$

$$E_M = \frac{\sum_{s=1}^{n} \sum_{t=1}^{m} (M_{s,t})}{E_{comp}} \cdot 100 \tag{4.10}$$

$$E_F = \frac{\sum_{s=1}^{n} \sum_{t=1}^{m} (F_{s,t})}{E_{comp}} \cdot 100 \tag{4.11}$$

$$E_{comp} = \sum_{s=1}^{n} \sum_{t=1}^{m} (|S_{s,t} - O_{s,t}|) \tag{4.12}$$

In Equations 4.9 to 4.12, S and O represent satellite and observed precipitation values, and n and m represent the number of stations and time steps, respectively. Some of the plots in Section 4.3.2 presenting this idea limit s to stations in specific regions of Canada, or t to specific months. As mentioned at the beginning of this part, observations are from the APC2 dataset, and the regions of Canada are defined by those in Figure 4.1.

The bias for a given classification is found by averaging the events associated with it, for a given location and/or time. The frequency is in the form of a percentage, representing the portion of Table 4.1 for the given classification in the spatial/temporal domain.

To test the application of a linear or non-linear model and more importantly the presence of systematic and random error, the methodology of Tian et al. (2013) is used so that results can also be checked for continuity at the Canada and United states border. The two models are defined by Equations 4.13 and 4.14, respectively, for an observation i, since all stations and time steps are treated equally in this test. Taking the natural logarithm of Equation 4.14 results in Equation 4.15, which is a linear model with transformed observations and estimates. The fitting parameters, a and b, in Equations 4.13 and 4.15 can be found through least squared error optimization.

$$S_i = aO_i + b + \varepsilon_i \tag{4.13}$$

$$S_i = bO_i^a e^{\varepsilon_i} \tag{4.14}$$

$$\ln(S_i) = \ln(b) + a\ln(O_i) + \varepsilon_i \tag{4.15}$$

After fitting the models, the residuals can be plotted as functions of observed data to look for any trends. If all of the systematic error is removed via the model, the residuals should not express any trends. On the plots, standard deviations of binned residuals are used to visually check if they are constant, as they should be.

Once a model has been selected, the actual random and systematic error can be expressed spatially. This is done through calculations of mean squared error on a station-by-station basis. That is, a model is fit for each station and root mean squared errors for systematic and random residuals are calculated by Equation 4.17 through time, t. Their final presentation is as a percent of total mean squared error.

$$MSE_{total} = MSE_{systematic} + MSE_{random} \tag{4.16}$$

$$\frac{\sum_{t=1}^{m}(S_t - O_t)^2}{m} = \frac{\sum_{t=1}^{m}(\hat{S}_t - O_t)^2}{m} + \frac{\sum_{t=1}^{m}(S_t - \hat{S}_t)^2}{m} \tag{4.17}$$

4.3.2 Results

The contribution of errors from hits, misses, and false alarms are shown in Figure 4.10. Classifications are based on daily accumulations of the satellite products with the APC2 dataset for June through October, 2003 to 2010. Figure 4.10a shows that hit bias is the largest contributor to the satellite products' error. While this may be promising because there is apparently some detection skill, the ability to correct these errors is dependent on the presence of random error, which is explored near the end of this section.

While false alarm contributions are for the most part constant amongst the satellite products, misses vary. Both PERSIANN products are more affected by misses than their CMORPH counterparts. The presence of error from misses is consistent with the general underestimation discussed in Section 4.2.2, as well observation that CMORPH overestimates the frequency of light rainfall more so than PERSIANN. While the hits portion of Figure 4.10 can contribute positive or negative errors, it is most likely negative as the false alarms and misses are of similar portions (cancelling each other out) and the satellite products underestimate rainfall overall.

Figure 4.10 also shows trends for the different resolutions. One would expect a higher resolution product to experience a decrease in the frequency of false alarms and increase in misses, as there is less spatial averaging. This is not explicitly apparent for any of the regions shown in Figure 4.10, but the values shown are dependent on both the magnitude and frequency. This is why frequency and magnitude are separated later in this section. Regardless, the higher resolution products are much more effected by misses than their low resolution counterparts across all regions of Canada. This increase comes at the cost of hits, whose portion reduces while false alarm's portion remains constant. As mentioned previously, false alarms were ex-

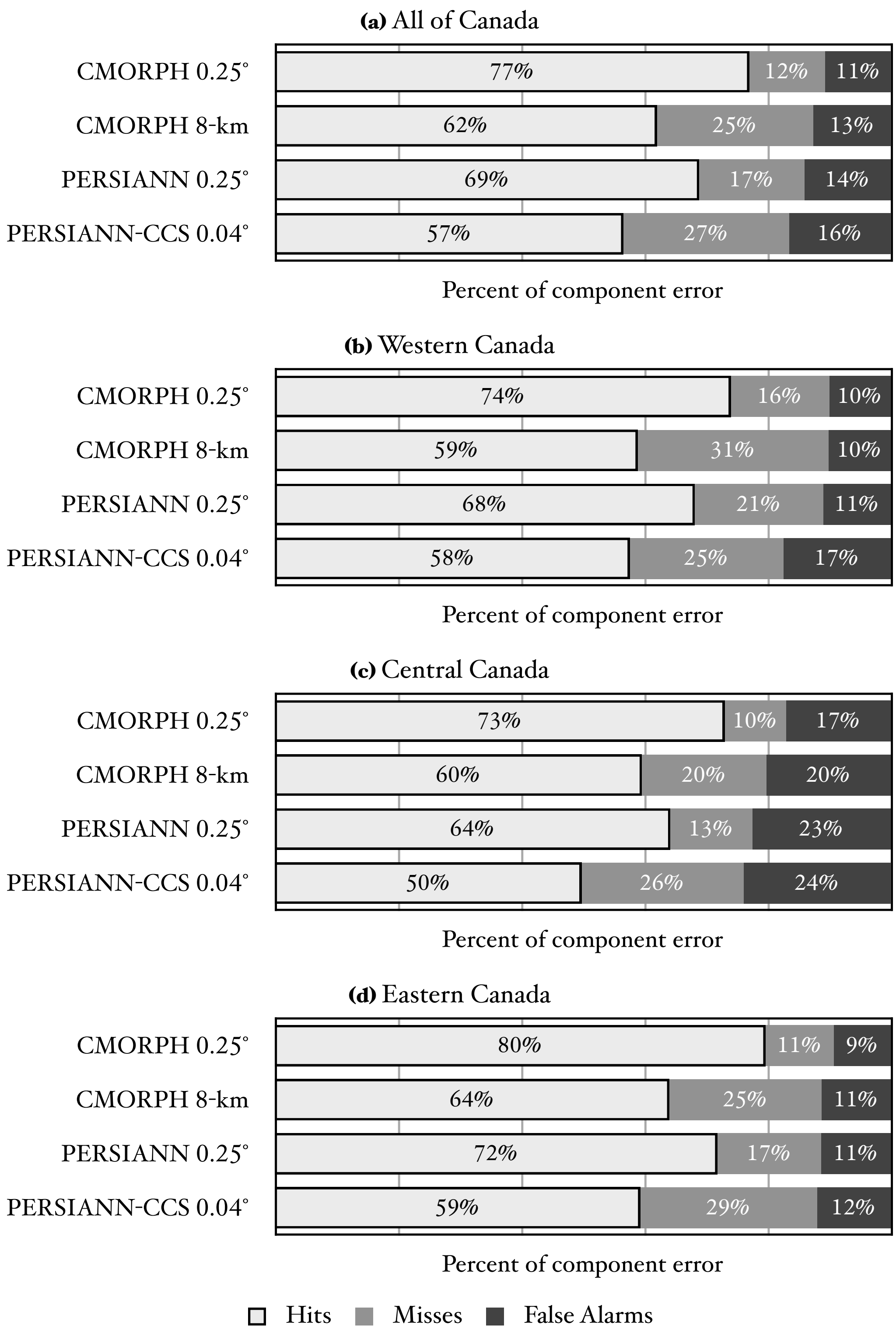


Figure 4.10: Percent contribution of categorized error to total error, when comparing satellite based precipitation products to APC2 for June through October, 2003 to 2010.

pected to decrease, which is not the case in Figure 4.10, as the higher resolution products are greater if there is a difference.

Between regions the contributions may change, but they are relatively the same amongst satellite products. For example, the 0.25° CMORPH product is always the most effected by hits compared to all other products for a given region. The largest regional change is that of central Canada where there is a large increase in the contribution from false alarm error, consistent with the previously findings of convective storms causing an overestimation. While a positive hit bias may also being causing this overestimation, false alarms can now be said to be playing a role.

To further investigate this noticeable change in Central Canada, Figure B.4 applies the same methodology to just central Canada through time. Considering the CMORPH products first, there is a change in the contributions to component error in July and August (the months known to have the greatest convective activity in central Canada). During these two months exclusively, false alarm contribution increases from roughly 15% to 21%. Hits and misses only appear to change as a result of false alarms. The PERSIANN products are not as clearly different in July and August, in that July, August and September are all similar, at least in this representation of error. Not seeing such a strong change relative to CMORPH is consistent with the mean ratios of Figures 4.6 and B.1 where the PERSIANN products were less affected by convective storms. It seems then that in central Canada, convective storms cause an overestimation, at least in part, due to false alarms. As discussed previously, Ebert et al. (2007) suggest that the overestimation is in part due to gauges missing the small and and intense convective storms, as the satellite products agreed between with the spatially complete radar data.

For all of the satellite products, the contribution of error from hits, misses, and false alarms is quite difference in October. The overall bias in Figure 4.5 shows that both PERSIANN products have reduced negative bias in this month. Figure B.4e shows that for October in central Canada, the CMORPH products have a large increase in error from missed precipitation, at the cost of hit bias. Meanwhile, the PERSIANN products have a large increase in false alarms, mostly at the cost of hit bias. This fits well with the previous discussion regarding the performance of passive microwave and infrared products over snow and possibly mixed precipitation. Again, Ebert et al. (2007) discussed that the infrared products overestimate the frequency of precipitation in mountainous regions and snow cover, while passive microwave have issues with detection. This appears to be the case in Figure B.4e, but, like previously mentioned, these are combined measures of frequency and magnitude.

To further detail the origin of the error, the frequency and mean of the errors are calculated for each dataset and classification. Table 4.2 presents the mean value for each classification for a given dataset and region, as well as that classification's portion within the classification table, shown in parentheses. Table 4.3 presents the same information for central Canada from June through October.

Regarding the previous curiosity of both false alarm and miss errors increasing with resolution, Table 4.2 shows why that is the case, and that the change in resolution actually acts as expected. For both CMORPH and PERSIANN, albeit to differing degrees, the high resolution product has a greater frequency of misses and a lower frequency of false alarms. It also shows that the reason for the previous counter-intuitive results is because in the case of false alarms, the frequency drops while the magnitude increases.

Table 4.2 also shows that the reason for the increase in false alarm error in Central Canada (Figure 4.10c) is different for the CMORPH and PERSIANN products. For the PERSIANN 0.25° product, the increase is from both an increase in frequency and an increase in mean error, which is not necessarily preferable. Meanwhile, both CMORPH products' false alarm frequency remains constant across Canada, and the magnitude increases.

Considering hits, the frequency decreases slightly, but the mean error decreases dramatically, from roughly -3.5 mm/day Canada wide, to -1.0 mm/day. So it appears at this temporal scale that the overestimation of convective events by CMORPH is not an overestimation of frequency, but an overestimation of the magnitude itself, as shown by both false alarm and hit

Table 4.2: Mean error and frequency (shown in parenthesis) for the decompositions presented in Figure 4.10 of the satellite rainfall products against APC2 for June through October, 2003 to 2010. Mean error is in units of mm/day.

		Hits	Misses	False Alarms
Canada	CMORPH 0.25°	−3.42 (31%)	−2.82 (10%)	1.08 (24%)
	CMORPH 8-km	−3.36 (24%)	−3.79 (17%)	2.39 (14%)
	PERSIANN 0.25°	−3.51 (27%)	−3.17 (14%)	2.17 (16%)
	PERSIANN 0.04°	−3.28 (22%)	−4.13 (19%)	3.90 (12%)
Western	CMORPH 0.25°	−2.95 (30%)	−2.91 (11%)	1.00 (24%)
	CMORPH 8-km	−2.83 (23%)	−3.51 (18%)	1.94 (12%)
	PERSIANN 0.25°	−3.13 (27%)	−3.12 (14%)	1.65 (15%)
	PERSIANN 0.04°	−1.24 (22%)	−3.60 (19%)	4.26 (11%)
Central	CMORPH 0.25°	−1.06 (27%)	−2.49 (9%)	1.37 (24%)
	CMORPH 8-km	−0.94 (21%)	−3.35 (14%)	2.72 (15%)
	PERSIANN 0.25°	−1.17 (25%)	−2.72 (11%)	2.23 (21%)
	PERSIANN 0.04°	−1.50 (18%)	−3.67 (18%)	3.63 (14%)
Eastern	CMORPH 0.25°	−3.94 (34%)	−2.73 (10%)	1.04 (24%)
	CMORPH 8-km	−3.96 (26%)	−3.88 (18%)	2.40 (13%)
	PERSIANN 0.25°	−3.91 (29%)	−3.23 (15%)	2.18 (14%)
	PERSIANN 0.04°	−4.25 (23%)	−4.42 (21%)	3.46 (11%)

magnitudes. This is the better of the two situations because a satellite product that reduces bias with increased false alarms frequency or magnitude is not producing a meaningful forecast. Also, bias in a particular region or to a type of precipitation is easier to correct, than having to spatially add or remove estimates.

In terms of trends across Canada, the frequencies of misses and false alarms are fairly constant with fluctuations in the magnitudes. Of missed precipitation, the mean magnitude is the lowest in central Canada, which can be partially attributed to the decreased normal precipitation in this region. Hit bias and frequency show the greatest spatial variability in that there is a relatively large reduction in negative bias in central Canada compared to either coast. Along with the decrease in hit bias, the frequency of hits is also reduced.

The general trends discussed here are for the most part in good alignment with the work of Tian et al. (2009) who decomposed errors over the United States. Near the border (closest to what is presented here), the authors observed large fluctuations in hit bias, from an underestimation on the west coast, to an overestimation in central United States, and back to a much larger underestimation on the east coast from both 0.25° PERSIANN and CMORPH products. Table 4.3 does not show an overestimation in central Canada because it ranges from June through October, while Tian et al. (2009) limited the analysis to the June-July-August season, which was shown to overestimate in Canada in Figure 4.6. The underestimation on the coasts is present, as well as the relatively greater negative hit bias on the east coast. For false alarms, the authors had the greatest amount in central United States with PERSIANN being more affected, as was the case in Table 4.2 where magnitude is greatest in central Canada and that of PERSIANN is greater than CMORPH, 2.23 mm/day compared to 1.37 mm/day respectively. One difference with the authors is the amount of missed precipitation on the east coast. Where the authors show CMORPH having more missed precipitation, the results here show that PERSIANN has more.

As done before, the error in central Canada is explored in greater detail from month to month and is included in Table 4.3. The main purpose of this table is to explore the error components in June, July, and August, when the satellite products are overestimating convective events, which are not significant enough to be fully captured in Table 4.2. Describing the components in this much detail allows for one to see the positive hit bias for all the satellite products, except PERSIANN-CCS, which does not show the same effects from convective storms. Not surprisingly, the hit bias in this table follows that of the mean ratio plots, in that the greatest hit bias is during July, followed by August and June, which follows the general amount of convective events in the Canadian Prairies (Raddatz and Hanesiak, 2008). During July when the positive hit bias is found, the frequency of hits and misses decrease while the frequency

Table 4.3: Same as Table 4.2 but for monthly values in central Canada.

		Hits	Misses	False Alarms
June	CMORPH 0.25°	−1.28 (35%)	−2.83 (9%)	1.55 (23%)
	CMORPH 8-km	−1.49 (28%)	−3.61 (16%)	3.20 (15%)
	PERSIANN 0.25°	−1.95 (30%)	−3.04 (14%)	2.29 (17%)
	PERSIANN 0.04°	−2.65 (22%)	−3.96 (22%)	3.43 (11%)
July	CMORPH 0.25°	0.66 (31%)	−2.76 (6%)	1.88 (28%)
	CMORPH 8-km	0.99 (24%)	−4.29 (12%)	3.62 (19%)
	PERSIANN 0.25°	0.19 (26%)	−3.07 (11%)	2.92 (19%)
	PERSIANN 0.04°	−1.35 (18%)	−4.50 (18%)	3.56 (13%)
August	CMORPH 0.25°	−0.45 (29%)	−2.68 (7%)	1.58 (28%)
	CMORPH 8-km	−0.29 (23%)	−3.81 (13%)	3.20 (18%)
	PERSIANN 0.25°	−0.69 (25%)	−3.00 (11%)	2.56 (19%)
	PERSIANN 0.04°	−2.32 (18%)	−4.19 (18%)	3.39 (12%)
September	CMORPH 0.25°	−1.83 (25%)	−2.10 (11%)	0.97 (22%)
	CMORPH 8-km	−1.72 (20%)	−2.61 (16%)	1.99 (13%)
	PERSIANN 0.25°	−1.59 (23%)	−2.48 (12%)	1.87 (20%)
	PERSIANN 0.04°	−0.70 (18%)	−3.15 (17%)	3.37 (14%)
October	CMORPH 0.25°	−2.71 (14%)	−2.07 (11%)	0.85 (21%)
	CMORPH 8-km	−2.59 (10%)	−2.33 (15%)	1.43 (12%)
	PERSIANN 0.25°	−1.97 (17%)	−1.91 (8%)	1.50 (31%)
	PERSIANN 0.04°	−0.32 (14%)	−2.40 (12%)	4.43 (23%)

and magnitudes of false alarms increase. While, as mentioned before, the increase in false alarm frequency and magnitude is not ideal, the increase in hit bias magnitude is so much greater, and its frequency is still the greatest amongst all the classification.

An aspect of the high resolution products is that there is less spatial averaging, so they can estimate greater magnitudes of precipitation. This is apparent across Canada in Table 4.2

where hits and false alarms are of greater magnitudes than the 0.25° product. While this may seem opposite of Figure 4.5, where the higher resolution products have greater negative bias, Table 4.2 shows that this negative bias is the result of the combination of magnitude and frequency. For example, the 8-km CMORPH product has greater magnitudes of hit and false alarm bias, but its frequency is less than that of the 0.25° product.

Before separating hit errors into systematic and random components, the applicability of a linear or non-linear model, given by Equations 4.13 and 4.14 respectively, is tested. The benefits of the non-linear model being that the errors scale with observed precipitation, since the errors associated with larger magnitudes of precipitation will mostly be larger as well.

Using daily data from APC2 as observations and satellite estimates from the 0.25° CMORPH product from June through October, 2001 to 2005, the two models are fit exclusively to hits and the residual are plotted in Figure 4.11. The full temporal range is not used because 2001

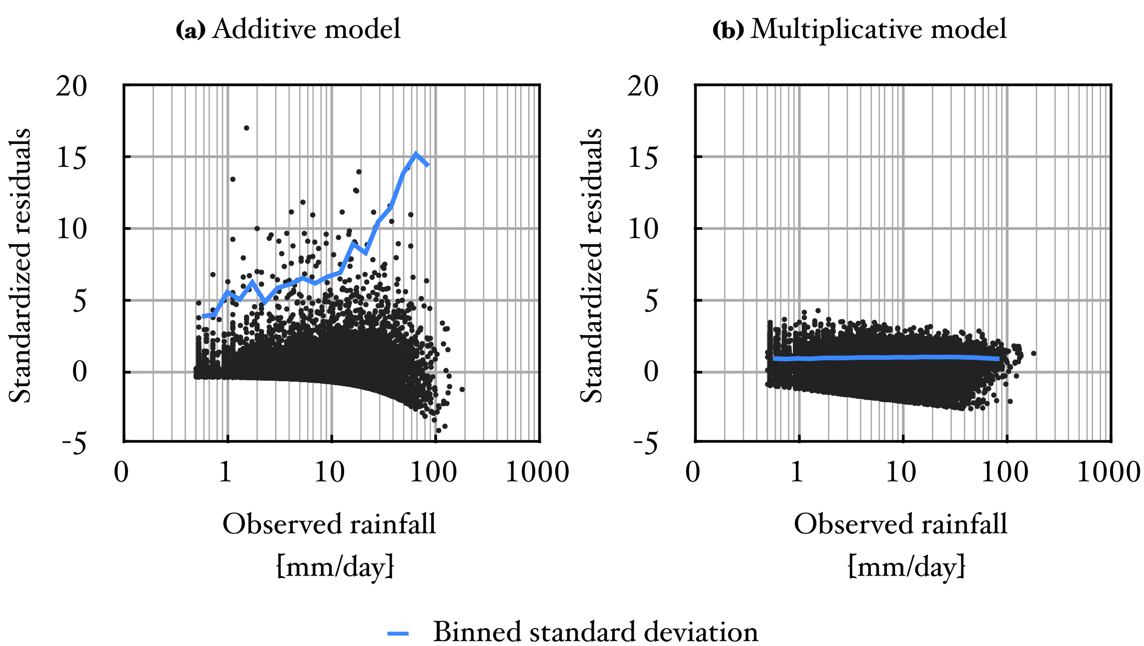


Figure 4.11: Standardized residuals resulting from applying additive and multiplicative models to daily 0.25° CMORPH rainfall estimates and APC2 measurements. Applied June through October, 2001 to 2005 with a 0.5 mm/day threshold to define hits.

through 2005 contains enough data for this test. The plots of standardized residuals as a function of observed precipitation show there is much more spread in the additive model compared to the multiplicative model at greater observed magnitudes. On the same axis, standard deviations for binned residuals are shown in blue and increase, which suggests that some systematic error is present in the residuals which should be random. As such, the non-linear model, or effectively transforming the data into log-space first, is the preferred option for the datasets considered here. It is assumed that this is applicable to the other satellite datasets tested in this project. The curve at the bottom of Figure 4.11a is because the residuals have a lower limit based on the fitted model. It is curved because it has been plotted in semilog space, for no other reason than to make the plots on the same scales.

For each station, a non-linear model is fit to all the data in the applicable temporal domain for the dataset, and the resulting mean squared random error is shown in Figure 4.12. The common thread that PERSIANN-CCS is fairly unresponsive in these analysis remains true here. The random and systematic contribution to hit error for PERSIANN-CCS is roughly equal. The other three datasets, however, show common trends. Western Canada uniquely consists mostly of systematic error, which is promising for future bias corrections. Eastern Canada is fairly balanced between systematic and random error, but in some parts of the Maritimes there is variability in the presence of random error, especially in the CMORPH products. The most distinctive region again is central Canada, where lots of random error is present.

The three satellite products (excluding PERSIANN-CCS) have been shown to have a large change in hit bias in July and August in central Canada. It is possible that this large amount of random error in central Canada is a result of poor fitting, since the hit bias changes so dramatically with time. To check this, July and September months are tested individually with the 0.25° products in Figure 4.13. Some stations have been removed depending on the presence of missing data. The patterns for random error in the individual months are similar to those calculated over the entire period, in that western Canada is mostly systematic error, central Canada is mostly random, and eastern Canada is a mix. The influence of convective storms in July is, however, much clearer. Both satellite products have more random error in

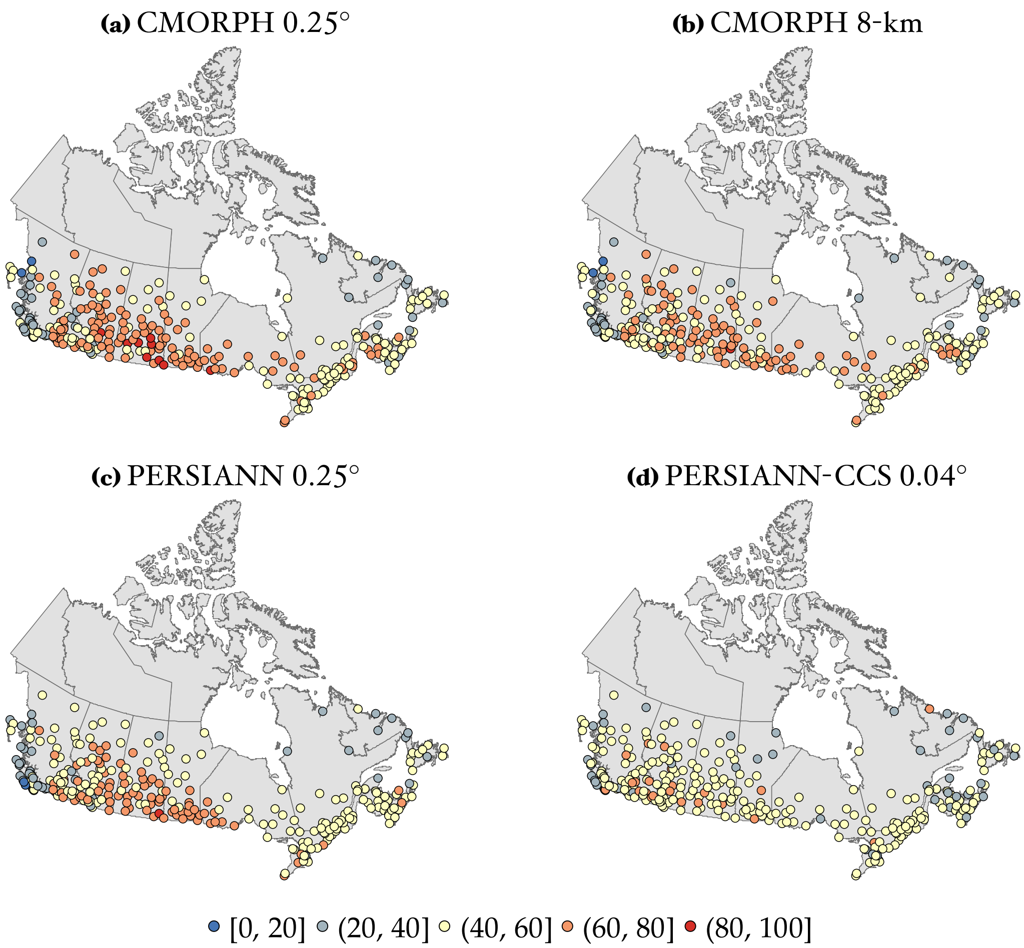


Figure 4.12: Percentage of random error resulting from the multiplicative model for June through October, 2001 to 2010 at each station.

July in central Canada, where convective events have been shown to have an effect, compared to September or the overall temporal domain. Therefore, while the overall quantification of systematic and random error is not negatively affected by the temporal change in hit bias, there is some change from month to month depending on the portion of events being convective. As such, future bias corrections or adjustments should consider each month separately to more effectively remove the bias.

4.3.3 Key points

- Hit bias is the largest contributor to error, followed by missed precipitation.
- The higher resolution products have less error from hits, but more from missed rainfall.

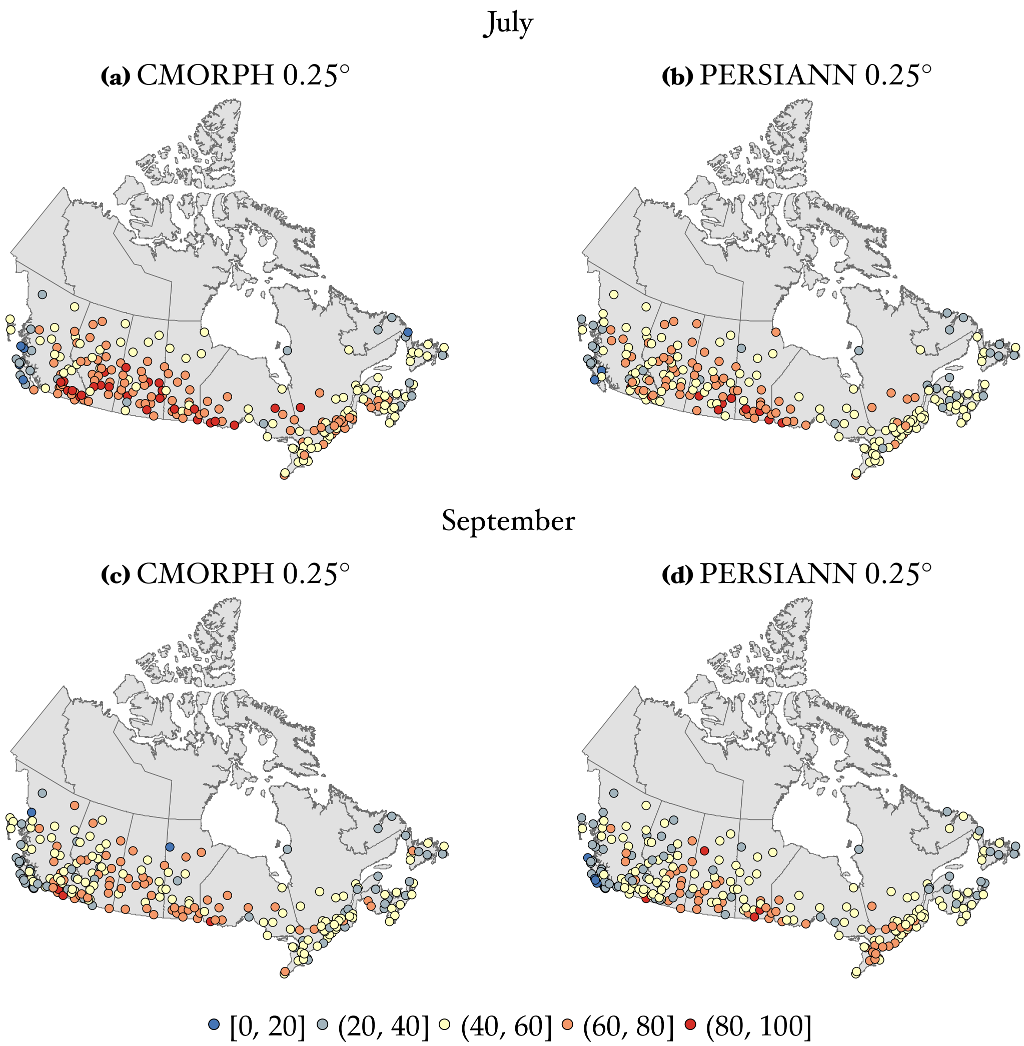


Figure 4.13: Same as Figure 4.12, but for July and September only.

- Hit bias is most negative in eastern Canada.
- Hit bias is most positive in central Canada for all products but PERSIANN-CCS.
- All products' hit bias is temporally correlated with convective storms in central Canada — frequency is not.
- Errors from false alarms are greater in central Canada mainly due to magnitude.
- CMORPH's false alarms in terms of both magnitude and frequency increase temporally in central Canada with convective storm occurrence.
- PERSIANN's false alarms are temporally correlated with convective storm occurrence in central Canada with regards to magnitude only, not frequency.

- The high resolution products have fewer false alarms and more misses than their low resolution counterparts, but can be masked by the mean magnitudes of the errors.
- A model whose error term scales with observed precipitation is better at separating systematic and random error, further demonstrating the importance of transforming precipitation data.
- The satellite products produce more systematic error for western Canada eastern Canada than central Canada.
- Although the trends in systematic and random error are similar through time, the magnitudes change on a monthly basis (mostly in central Canada), suggesting that future bias correction or adjustments consider the temporal changes in the data.

Part 5 | Satellite precipitation data preparation

In terms of workflow, the error quantification of Part 4 is independent of the work done with CaPA. It does, however, serve as a means to explain CaPA's results in Parts 6 and 7. It also provides some guidance in the pre-processing step, especially in detailing the presence of bias across Canada which should be dealt with prior to the assimilation. This section details the processes applied to the satellite data before they are assimilated with CaPA.

Since the error quantification uses the APC2 dataset as a reference field, while CaPA uses EC 6-hourly data, the equitable threat scores from the two with satellite data are first compared. This sets expectations of the satellite data's performance relative to the EC 6-hourly observations, and provides some insight into the effects of the adjustments made in the APC2 dataset.

The algorithm for removing estimates over snow covered ground is then detailed. This algorithm is applied to the satellite data before the error quantification of Part 4 and before the assimilation with CaPA. The effectiveness of the algorithm is compared with that of CMORPH, which zeros out such estimates. This removal also aids in defining the temporal domain of the project, June to October, based on remaining data availability.

Finally, two methodologies for adjusting the satellite data to be in better accordance with the reference data fields are tested. One uses the APC2 dataset to define a bias that is removed from the satellite products, while the other rescales the satellite data to be consistent with accumulated observed rainfall over the previous thirty days. The better performing of the two is applied to the satellite products before they are assimilated with CaPA.

The methods are evaluated based on their ability to reproduce the reference field. The equitable threat scores for these evaluations in the following sections, as well as for the CaPA analyses, are based on leave-one-out cross-validations, in which each station is sequentially removed, the analysis is performed, and the value at the location of the removed station is recorded for evaluation. Also, in cases where temporal resolutions differ, the lesser is accumulated so they are equivalent. This is required as the performance is dependent on the temporal resolution, even more than spatial resolution.

5.1 Differences between using APC2 and 6-hourly data as a reference field

From here on in the project, the primary reference dataset is the EC 6-hourly data, which are currently being used as input data in CaPA (see Section 3.5). The role of APC2 from here on is only as a means to potentially adjust the satellite data, should the methodology prove beneficial. The change in observed data is based on the remaining objectives. Where Part 4 required the reference dataset to be closest to the truth to best quantify error, the remaining work is concerned with changes in skill relative to the currently operational form. As such, the same observed data are used.

It is not expected that changing the reference dataset will greatly impact the ranking of the satellite precipitation products. The magnitudes of the skill scores will change, however, as the EC 6-hourly data has not undergone any adjustments as the APC2 data has. This section exists for two purposes. First, to show that the skill scores relative to the APC2 data (calculated previously) should not be compared with those from CaPA analyses in the following parts. Secondly, to propose new expectations for the performance of the satellite products when the EC 6-hourly data are used instead.

For each comparison, the EC 6-hourly data from manual stations are accumulated to daily values, where the exclusive use of manual stations is chosen for their reliability as well as their fairly even distribution across Canada, as shown in Figure 3.2. The satellite precipitation products at their two resolutions (0.25°/8 km for CMORPH, and 0.25°/0.04° for PERSIANN and PERSIANN-CCS) are compared via equitable threat scores to both APC2 and the accumulated EC 6-hour data for the months of June through October, from 2002 to 2010. Once again, PERSIANN-CCS is from 2003 to 2010.

Since it better shows the differences between the resolutions and datasets, Figure 5.1 presents the equitable threat scores as a function of threshold precipitations, following the methodology of Section 4.2.1. Binned values of equitable threat score, probability of detection, and false alarm rate are also presented for CMORPH and PERSIANN in Figures C.1 and C.2, respectively.

Two distinct trends are apparent in Figure 5.1 and constant across the magnitude spectrum. First, the satellite precipitation products are in better accordance with the accumulated EC 6-hourly data as opposed to APC2. Second, the 0.25° products have more skill by this measure than their high resolution counterparts.

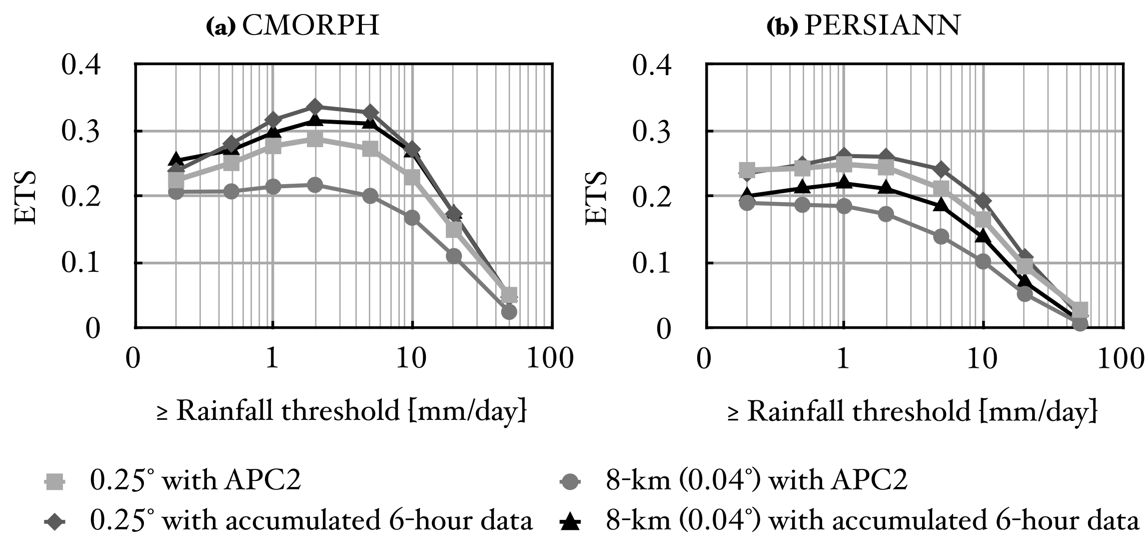


Figure 5.1: Equitable threat scores across Canada for differing spatial resolutions of satellite data (PERSIANN-CCS shown in parentheses) and different observational data for June through October from 2002 to 2010. EC 6-hourly data from manual stations has been accumulated to daily values to be fairly compared with the daily APC2.

While the performance of the high and low resolution products has been discussed in Part 4, it is yet to be determined why the EC 6-hourly accumulated data results in greater skill for the satellite products. If one compares the same satellite product and resolution to the two different observational sources, the difference in skill must be from the adjustment, because almost everything else is being held constant, given that the observed data are for the most part from the same origin. Combining the facts that both PERSIANN and CMORPH tend to underestimate precipitation across Canada, as discussed in Section 4.2, and that the adjustments in APC2 have the net effect of increasing precipitation magnitudes in the domain of this project (Mekis and Vincent, 2011), one can conclude that the relatively greater skill is from evaluating a satellite product that underestimates rainfall with underestimated observational data. By increasing the magnitude of the observed data to better represent reality, the gap between the satellite products and the observations also increases, reducing the skill.

The binned equitable threat scores for CMORPH and PERSIANN in Figures C.1a and C.2a, respectively, generally show the same trends as Figure 5.1. The only discrepancy present in the binned data different than the above conclusions is that for rainfall greater than 25 mm/day, where the 8-km CMORPH product performs slightly better than the 0.25° CMORPH product when using EC 6-hourly data as a reference. Figure C.1 suggests that this increase in skill is related to the probability of detection. For both CMORPH and PERSIANN, probability of detection is more closely related to equitable threat score than false alarm rate, based on their individual changes relative to equitable threat score.

To further investigate this increase in skill, the differences in probability of detection between the two resolutions are plotted in Figure 5.2 for the different regions of Canada. Firstly, it shows that for probability of detection, the higher resolution product is preferable at greater magnitudes. This is likely a result of larger magnitude events covering smaller spatial areas, as the 8-km product can better delineate them. Also, it shows that for rainfall greater than 30 mm/day, central Canada has the greatest detection of all the regions. It is worth noting that there is a lot of variability in probability of detection in this region.

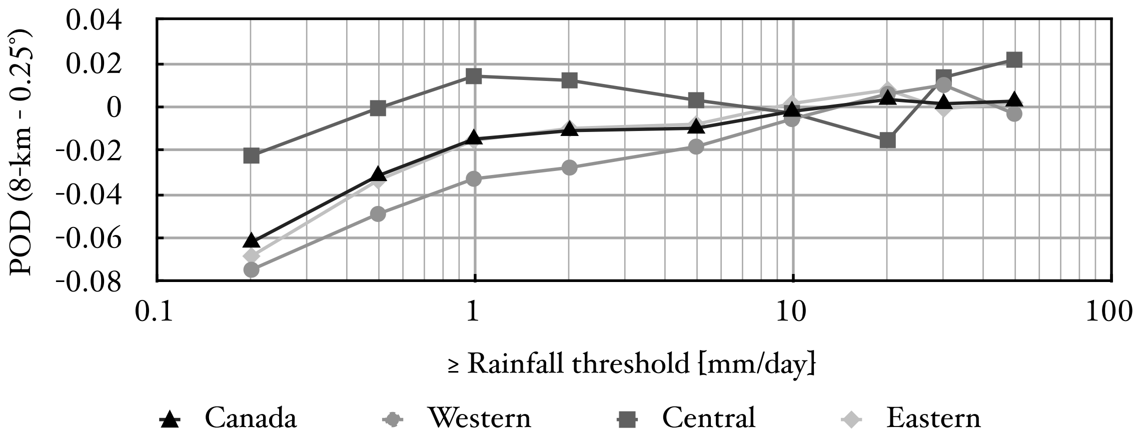


Figure 5.2: Differences in probability of detection between the 8-km and 0.25° CMORPH products for June to October, 2002 to 2010 using daily accumulations of the 6-hourly data as a reference.

The caution remains that the equitable threat scores between varying observational datasets presented in this project should not be compared directly. Across the magnitude spectrum, the satellite products have greater skill scores with the EC 6-hourly data, compared to APC2. In terms of resolution, the 0.25° products are more skilful than their higher resolution counterparts. This is not to say that the high resolution product should be discarded. Those used here are not only of greater spatial resolution, but also of greater temporal resolution, and can provide useful information that the lower resolution products cannot convey, such as event growth, event decay, or more detailed spatial features. While the skill may be lower, more is being asked of the higher resolution products, so it may not be fair to compare them in this manner. Ultimately, the resolution of the satellite product should be chosen to best match the needs of the project it is being used for.

5.2 Snow estimate removal algorithm

To remove precipitation estimates over snow covered land the IMS dataset is used, which is detailed in Section 3.7. It classifies the Northern Hemisphere daily as land, snow, water or ice on either a 4 km or 24 km grid, and is publicly available in a real-time — fitting the real-time nature of CaPA. This methodology is applied with the 24 km product primarily because it is the only resolution which covers the entire temporal domain of this project.

This section details how the estimates over snow covered ground have been removed. As a means of evaluation, the results from the methodology proposed here are tested against the 0.25° CMORPH product, which has estimates over snow covered land removed before it is made available (Joyce et al., 2004). In CMORPH, the removal process has not replaced the estimates with missing data, but with zeros and as such they still need to be removed. Additionally, there is some benefit in being able to apply the same methodology across all of the products.

5.2.1 Methodology

Although the overall algorithm for removing precipitation estimates where a second dataset specifies ice or snow is quite simple, special care needs to be be taken to ensure its effectiveness. The first issue is that while the satellite products are on a standard WGS84 latitude/longitude grid, the IMS data are on a polar stereographic projection centred at 90°N. This means that the grid cells do not perfectly align, so some spatial interpretation is required.

Given that the centre points for all grid cells are known, one can find the closest cell in the IMS dataset for every cell in the satellite precipitation product, as it will have the greatest spatial overlap. While this works well for most of Canada, it produces problems in areas near open water, where the closest grid cell in the IMS dataset is a mix of land and water.

The method implemented here ultimately treats the co-located grid cells near water different than those inland. For every satellite grid cell the closest grid cell in the IMS data is found, and if it is classified as snow or ice the precipitation estimate is removed. At the same time, a 3×3 window is defined about the IMS grid cell and checked using the logical statement of Equation 5.1 where & and | are the logical 'and' and 'or' operators, respectively. While the concern is only regarding snow and sea ice, the 'sea' term in Equation 5.1 is added to ensure that the expanded window is only applied on the problematic coastal regions. The main application of this windowing technique is in areas where the land is covered in snow, the nearby water is open, and the closest IMS grid cell is off-land.

$$\text{Sea Ice} \mid (\text{Sea} \;\&\; \text{Snow}) \tag{5.1}$$

5.2.2 Results

Since the CMORPH products replace estimates over snow cover with zeros, this information can be used to evaluate the relative effectiveness of the methodology of Section 5.2.1. Figure 5.3a shows the percentage of grid cells within Canada below 60°N equalling zero from 2001 through 2005 before and after the algorithm is applied. The replacement of estimates with zeros in the winter months is clear. After applying the above methodology, the sharp increase in the quantity of zeros is replaced by a sharp decrease, as they are now being considered missing data. From this figure, it seems as though the methodology is able to replicate that used by Joyce et al. (2004) to some degree, but with removing estimates instead of replacing them with zeros. Figure 5.3b also demonstrates this effectiveness by showing what percentage of grid cells in Canada were removed accurately (zero) or falsely (non-zero) over the same period. Few estimates were falsely removed, on average 1.4%, in comparison to those accurately removed.

To show the overall effect of removing estimates over snow covered land, Figure 5.4 shows daily percentages of stations with co-located satellite data from 2001 through 2005. The raw

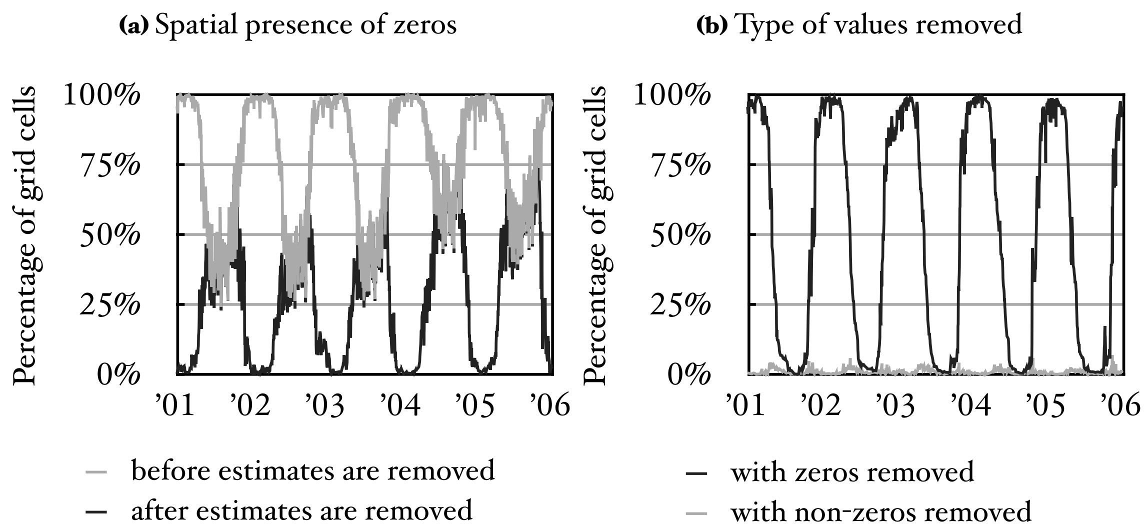


Figure 5.3: Effects of the snow cover removal methodology applied to daily estimates from the 0.25° CMORPH product from 2001 through 2005 on the presence of zeros.

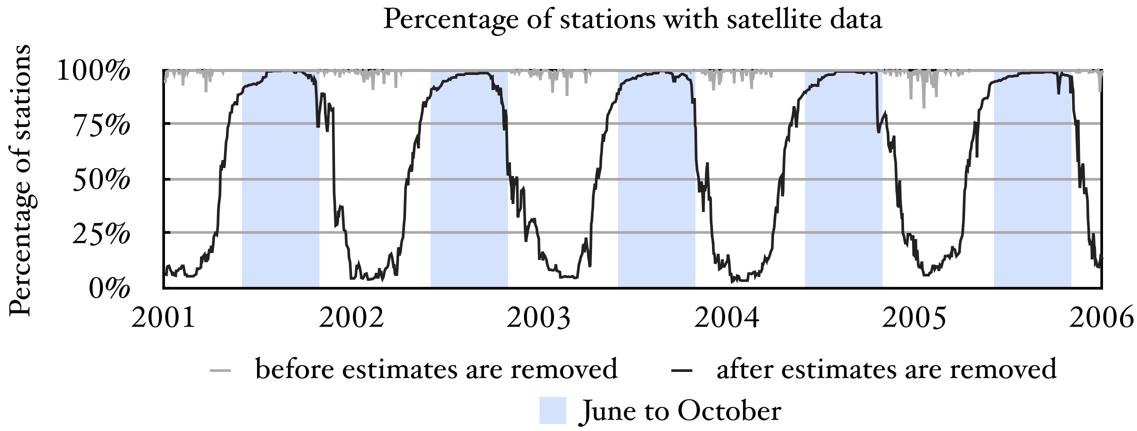


Figure 5.4: Daily percentages of stations with co-located data in the 0.25° CMORPH product before and after the removal methodology is applied from 2001 through 2005.

CMORPH data are quite complete, with most missing data occurring in the summer months, likely because the replacement of data in the winter with zeros is also replacing missing data since these estimates, if they exist, would be removed regardless. By replacing the zeroed data with missing data, the seasonal trend is apparent in black.

The monthly domain for this project (June to October) is shown in Figure 5.4 in blue, which has been chosen such that the majority of stations have satellite data to compare with. This figure also helps in explaining the sudden change in error contribution in Figure B.4e (discussed in Section 4.3.2), in that the remaining data decreases rapidly at the end of October. So while missing data may have an effect in some years, it shows that there must be mixed precipitation over bare land in October, which was a suggested cause for the sudden change in error contribution, as the ground is covered with snow shortly after.

5.3 Dataset adjustment methodology

Before the satellite products are assimilated with observations using CaPA, an attempt is made to remove the bias shown in Part 4. While in a reanalysis scenario one would be able to rescale the satellite data to match the calender month's accumulations, the focus here is on a real-time application. As such, the following methodologies use either historic information to remove the bias, or that from the recent past to make an adjustment.

5.3.1 Past thirty days

While using historical data provides benefits in terms of the quantity of information, it cannot handle recent changes in the datasets, should they occur. As shown in Part 3, the quantity of stations changes through time, either through decommission or the inclusion of additional networks, which would change the pre-calculated bias ratios. Perhaps more significantly, the satellite data are always changing as satellites are replaced or methodologies are updated. By only considering the previous thirty days, any changes in the datasets will be considered. It also means that the methodology is sensitive to quick changes in climate or data quality errors. Since only the recent past is considered, one cannot say that the bias is being removed, so from here on this methodology will be referred to as adjusting or rescaling the satellite data.

This adjustment is similar to the final step used by Huffman et al. (2007) in creating the TRMM 3B42RT product, which itself is a combination of satellite estimates (infrared and passive microwave) and stations, released in real-time. A key difference is the station density. Where Huffman et al. (2007) had a gridded observational dataset, this project only makes use of station data. To account for this difference in datasets, this project incorporates interpolation.

At every station being considered and its co-located satellite grid cell, estimates are accumulated over the previous 30 days and ratios of observed to satellite are calculated. These ratios are then interpolated in log-space to the satellite dataset's resolution, which should produce better results than interpolating the observed data first, as there is expected to be less spatial variability and no issues with zero rainfall. The interpolated ratios are then applied to the satellite data at its original temporal resolution. The result is that the accumulations of the satellite and observed data are the same over the previous thirty days.

This does, however, assume that the bias in the satellite data is constant for a given 30 days. Unfortunately, mean monthly ratio maps in Figures 4.6 and B.1 show that bias in central Canada changes on a monthly basis. This, combined with the relatively large portion of random errors, suggests that the adjustment will have the smallest improvements in this region. Such evaluations are made in Section 5.5.

In terms of the specifics of the methodology, data are removed from both the adjusted and satellite data if either is missing it for a given station and time step prior to accumulating. After this, if a station is missing more than one-third of its data, it is not used in the interpolation. While this limit is selected somewhat subjectively, Figure 5.5 aided in its selection. For every time step from June to October from 2002 through 2011, the portion of available data (non-missing) is found for each stations over the previous thirty days. Percentages for each station, for each time step, are then binned resulting in Figure 5.5 for the 0.25° CMORPH product and 6-hourly observations. Requiring two-thirds of data is based on a balance of having enough data to provide a reliable ratio, but also not to reject too many stations. At the high end of Figure 5.5 there is a non-linear decrease in frequency, where 60% is at the tail of this decline so the majority of worthwhile stations will be included.

Before the ratios are interpolated, an upper bound is enforced. The limit chosen here is four, which was also used by Lin and Wang (2011), who assimilated satellite data and observations over Canada through a combined multiplicative and additive methodology.

Figure 5.6 shows the results from CaPA when using CMORPH as an additional observational

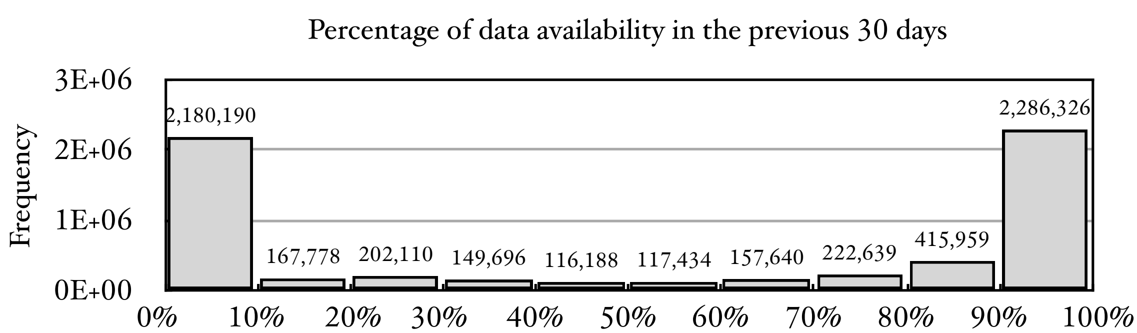


Figure 5.5: Binned station counts based on their percent of available data over the previous thirty days, from June to October, 2002 through 2011. The EC 6-hourly data are used as observations and the 0.25° CMORPH product is used as satellite data.

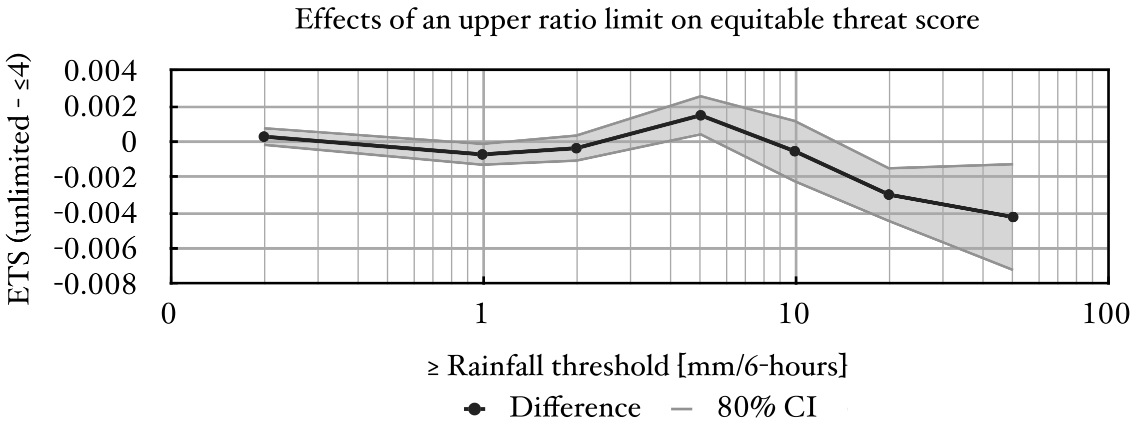


Figure 5.6: Changes in equitable threat score as a result of imposing an upper limit of four to the interpolated ratio field on CaPA analyses. The analysis is run with GEM as the background field, and the CMORPH 0.25° product as an additional data source for June to October, from 2002 through 2006.

source from June to October, 2002 through 2006 (trial run) with and without an upper ratio limit imposed. The scores are the result of a leave-one-out cross-validation of all available stations Canada wide, and plotted with 80% confidence intervals. To interpolate the ratios in this particular analysis, a cubic polynomial is used. The effectiveness of this interpolation method is further tested in Section 5.4. The values in the plot are the differences between an adjustment with no limit less one with an upper bound of four. With this definition, a negative score indicates the application of an upper bound to be beneficial. While the scores in Figure 5.6 are minor, they follow an intuitive trend. For precipitation events greater than roughly 12 mm/6-hour, it is beneficial to apply an upper limit, as it is apparently reducing an over adjustment of the satellite data. While this difference is statistically significant, those at lower magnitudes are not.

5.3.2 Mean monthly ratios

To remove bias, the satellite data are scaled such that the long-term monthly accumulations match those of the reference dataset. For this method APC2 is used because it is available over most of the temporal domain of the project, and it provides the best estimate of bias. While a more accurate solution would be to scale the satellite data for each calender month retroactively, this project is concerned with a real-time application of satellite data. As such, monthly accumulations at each station are projected based on those from the same month in previous years.

To start, missing data in either the observed or co-located satellite data is removed from both datasets. This demonstrated another advantage of using APC2 instead of the 6-hourly, in that the APC2 data have fewer missing days, excluding those as a result of station decommission. Even still, the satellite data will introduce missing data, so the same criterion as in Section 5.3.1 is used, where two-thirds of monthly data is required to accumulate. Although the justification for this limit was formed while using the EC 6-hourly data, its application here is more conservative since fewer observations are missing. If anything, it may be rejecting stations in exchange for consistent and stable ratios through time. At each station, data for every month from 2001 through 2010 are accumulated and a ratio of observed to satellite data is calculated. For a given month, the ratios for each year outside of three standard deviations are removed, and if at least five ratios (representing five years) remain, they are averaged and interpolated spatially in log-space to the resolution of the satellite product.

These requirements for mean monthly ratios prior to interpolation are found to work well. For example, Figure C.3 shows the mean ratios of APC2 to 0.25° CMORPH data for each station as a function of longitude for July, from 2001 to 2010. The figure shows how the ratios vary across Canada with longitude with a smooth change in variability, while remaining free of outliers.

The final result is a ratio field for each month, which can be multiplied to a satellite product time step of the same month. As with the methodology in Section 5.3.1, two interpolation/smoothing techniques are considered and tested in Section 5.4.1.

5.4 Dataset adjustment selection

From the methods described in Section 5.3, two selections are made here. The first is regarding the transfer of ratios at each station to that of the satellite data's gridded resolution, in which the use of a cubic polynomial and thin plate spline are considered. Amongst simple interpolation techniques, there was not a significant difference in the resulting equitable threat scores in comparing the adjusted satellite field to the reference data. From this group, cubic polynomial was selected. Kriging was also considered, but the resulting equitable threat score from a trial evaluation was comparable to that from the cubic polynomial. With minimal differences in skill, the additional computation time is a deterrent. The inclusion of a thin plate spline is based around the idea that in its smoothing, it will filter out anomalies and produce a better result in areas of high variability.

The selection is based on the equitable threat scores of the satellite data, after they has been adjusted with a given alternative, with the reference field. Two difference reference fields are used here based on the methodologies of the adjustments, APC2 and the EC 6-hourly data. When methodologies using APC2 and EC 6-hourly data are compared against each other, the temporal range is made consistent, ranging from 2002 to 2010, and the EC 6-hourly data is accumulated to daily.

5.4.1 Interpolation/smoothing techniques

To compare the techniques, Figure 5.7 shows the difference in equitable threat scores for given rainfall thresholds. The scores are again based on a leave-one-out analysis, measuring the ability of the adjustment methodology to reproduce the reference field, after it has been applied to the satellite products. For this section, the CMORPH 0.25° product is used exclusively for its few missing days and because it has been shown to be the preferred dataset and resolution so far. Subfigure 5.7b shows the results considering all stations and just manual stations, as either can be applied with the method, and it may be beneficial to use fewer stations in the adjustment if they are of superior quality, even if the lower quantity impacts the interpolation.

For the mean monthly ratio technique shown in Subfigure 5.7a, the change in equitable threat score for the two interpolation techniques is minimal relative to that of Subfigure 5.7b for the other method. A noticeable trend in Subfigure 5.7a is that the cubic polynomial is more skilful

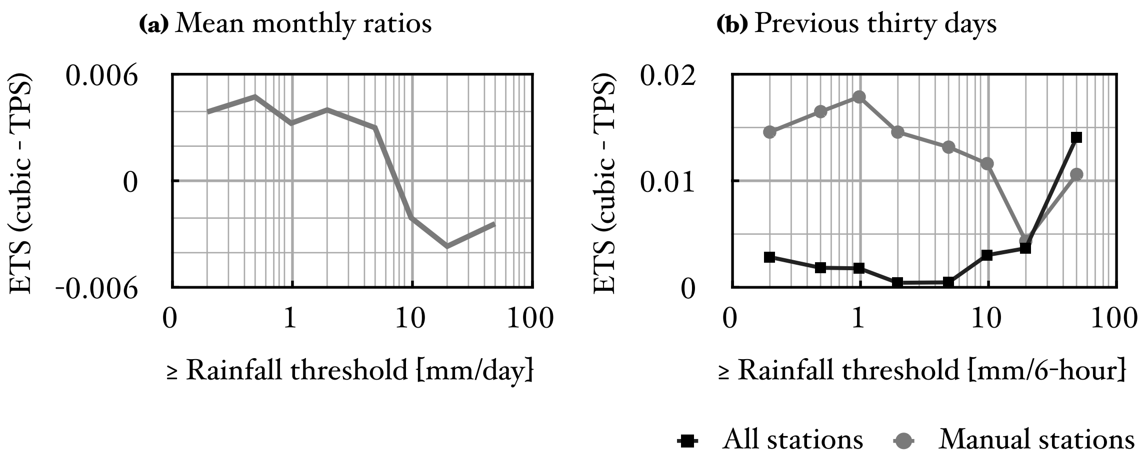


Figure 5.7: Differences in equitable threat scores by applying different interpolation/smoothing techniques to adjust the 0.25° CMORPH product with APC2 and 6-hourly data from June to October for 2001 through 2011.

at lower magnitudes than the thin plate spline, and the opposite is true for greater magnitudes. Since the ratios are from accumulated rainfall, a single large event can greatly influence the results. However, the methodology of mean monthly ratios, as described above, should be quite smooth through space as the monthly accumulations are averaged over ten years. This type of spatial description is well suited with that of the thin plate spline, which will produce a smoother surface. That said, the cubic polynomial is expected to produce more skilful results as it is preserving the nodal values in the interpolation. So if a sharp change in ratios is present, the cubic polynomial will be able to reproduce it while the thin plate spline will be minimally effected. So while the cubic polynomial is producing results closer to the reference field, the thin plate spline is producing what are potentially more realistic and conservative results. This may be the case in Subfigure 5.7a at greater magnitudes, because this is one of the extremes and is uniquely comprised of few events, so the effects will be more pronounced. So while slightly larger ratios are interpolated through space and causing nearby overestimations, the thin plate spline is not doing it to the same degree.

For the second adjustment methodology shown in Figure 5.7b, the cubic polynomial interpolation is better than the thin plate spline at all rainfall thresholds. This may seem contradictory to the previous methodology, but the manual stations (which are more closely related to the APC2 dataset) in the EC 6-hourly data show a similar trend, in that the difference generally decreases at larger magnitudes. The interesting aspect of this subfigure is the cubic polynomial interpolation is strongly preferred by the manual stations, unlike that for all stations. This may speak to the quality of the manual stations relative all stations. If the manual stations possess less variability in the ratios in space, the cubic interpolation would perform better as it is preserving nodal values and a given station is not greatly negatively effecting the area around it.

Regarding the intended purpose of this section, an interpolation/smoothing technique needs to be selected for each adjustment. Considering Subfigure 5.7a, the differences in interpolation/smoothing methodologies are ultimately not large enough to influence a selection of adjustment method. As such, the thin plate spline smoothing is selected for its effectiveness with extreme events. For the previous thirty days adjustment in subfigure 5.7b, regardless of which stations are used, cubic polynomial interpolation will be used from here on.

5.4.2 Adjustment method

The adjustments from the mean monthly ratio methodology of Section 5.3.2 using thin plate spline smoothing, and the previous thirty days methodology of Section 5.3.1 using cubic polynomial interpolation are applied to the CMORPH 0.25° product to test their effectiveness. As before, the evaluation is based on the ability of the adjusted satellite product to reproduce the reference field with a leave-one-out cross-validation. The results are shown in Figure 5.8 for June through October from 2002 to 2010. The year 2001 has been removed as it is not available in the EC 6-hourly data. Also, the EC 6-hourly data have been accumulated after the adjustment so the two methods are tested at daily resolutions.

There are two apparent ways that the mean monthly ratio methodology could be evaluated, using either APC2 or EC 6-hourly data as the reference field. While the latter is closer to what would be the eventual implementation, the former is used here because this section intends to test the methods themselves. The latter would be artificially low because, as demonstrated in Section 5.1, the 6-hourly and APC2 data produce difference scores for a given satellite

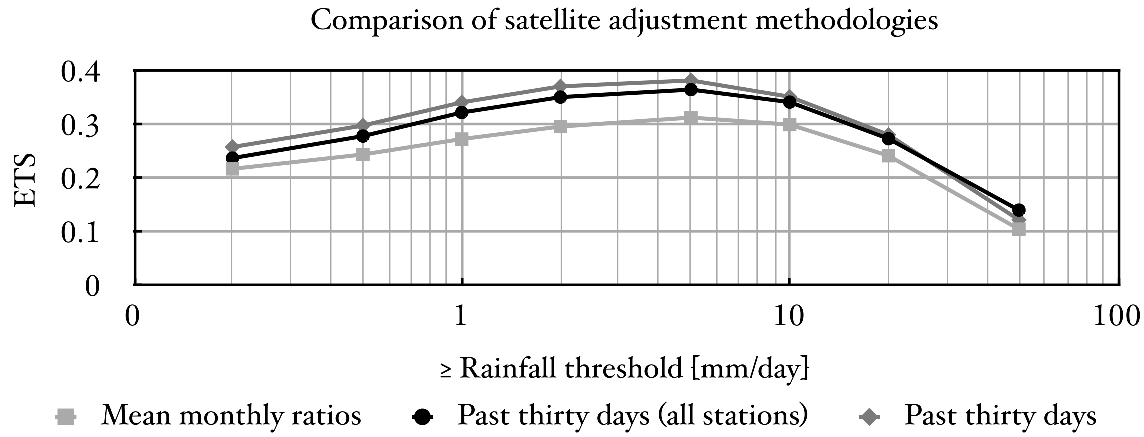


Figure 5.8: Equitable threat scores of the 0.25° CMORPH product adjusted by different methodologies from June to October for 2002 through 2010.

product, and as such the ratios would be correcting the satellite data to a different dataset. However, if the method itself can be shown to be effective, then its implementation can be rationalized even if it produces a lower score against the 6-hourly data, because the satellite data would be closer to the 'truth' provided by the APC2 data.

As Figure 5.8 shows, whether the mean monthly ratio methodology uses a thin plate spline or cubic polynomial interpolation, the differences in equitable threat scores are not enough to influence the conclusions of this section. Across the magnitude spectrum, performing an adjustment to the previous thirty days is preferable to mean monthly ratios, based on this evaluation. This is because it operates over a short temporal range, and as such can adjust for smaller differences which are lost in the bias correction.

There was some concern that performing the adjustment to the previous thirty days with just manual stations would negatively affect the skill as the number of stations is greatly reduced. This is fortunately not the case, as shown in Figure 5.8.

From Figure 5.8, the adjustment to the satellite data from here on will consider the previous thirty days and use cubic polynomial interpolation. Whether to use all stations or just manual stations in the adjustment is explored in Section 6.1 with the satellite data's assimilation with CaPA.

5.5 Dataset adjustment performance

Section 5.4 detailed the selection of an adjustment methodology, where the optimal adjustment was found to be based on ratios at each station based on accumulations from the previous thirty days, which are interpolated through space and applied back to the satellite data. Now that this has been established, the effects of this adjustment can be explored in more detail, specifically its effectiveness relative to the findings in the error quantification of Part 4.

First, the equitable threat scores are separated into the regions of Canada as defined previously in Figure 4.1 and can be seen in Figure 5.9. Subfigure 5.9a shows the scores for the CMORPH 0.25° product relative to the EC 6-hourly data for June through October from 2002 to 2011, while Subfigure 5.9b shows the differences between the scores of the raw and adjusted satellite data. As has been done previously, the data plotted is from a leave-one-out cross-validation, and the scores are a measure of the ability of the satellite data to replicate the reference field.

In terms of the spatial trends, the adjustment increases the scores across the magnitude spec-

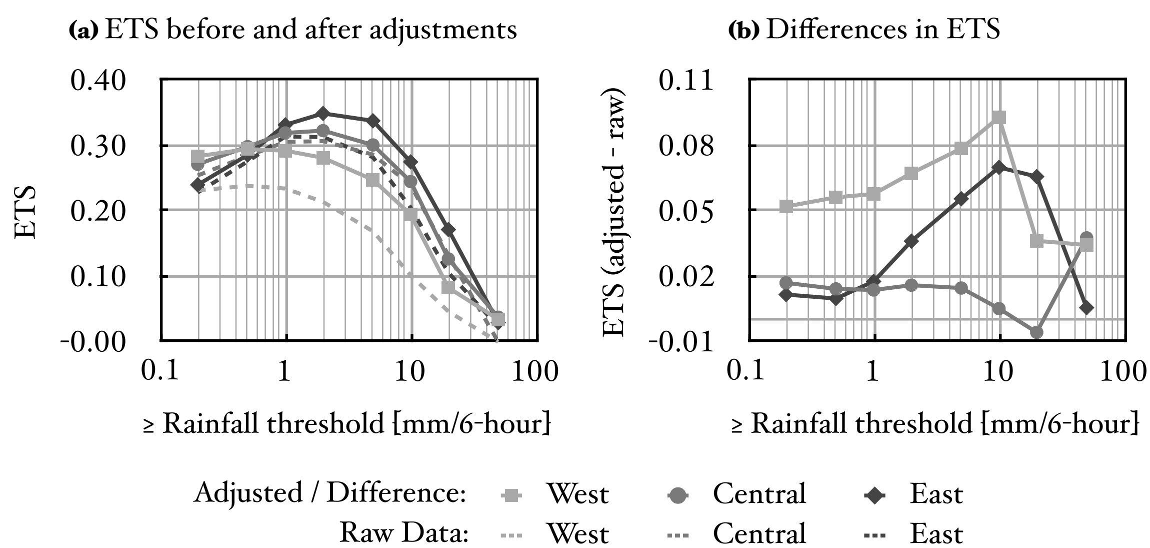


Figure 5.9: Equitable threat scores before and after applying an adjustment to the 0.25° CMORPH product based on the past 30 days of data for June through October from 2002 to 2011 relative to the manual stations from the EC 6-hourly dataset.

trum to varying degrees. The scores for the regions are similar, relative to each other, before and after the adjustment. That is, western Canada has the worst performance and eastern Canada is slightly better than central Canada. After the adjustment, the difference between central and eastern Canada is greater. This is related to the nature of the error in the satellite data. As Subfigure 5.9b shows, the largest gains from the adjustment are for western Canada followed by eastern Canada. Compared to these, central Canada did not benefit a lot from the adjustment, however it was performing well before the adjustment, so it still manages to perform better than western Canada.

The different amounts of change in equitable threat scores for the regions of Canada are best explained by the proportions of systematic and random error in the raw satellite data. In Section 4.3 the random and systematic error were quantified for all of the satellite products relative to the APC2 dataset. The general trend for the 0.25° CMORPH product was to have mostly systematic error on the west coast, mostly random error in central Canada, and an even amount in eastern Canada. Figure 5.10 shows the same type of analysis for the CMORPH 0.25° product, but for systematic error and with manual stations from the 6-hourly data for

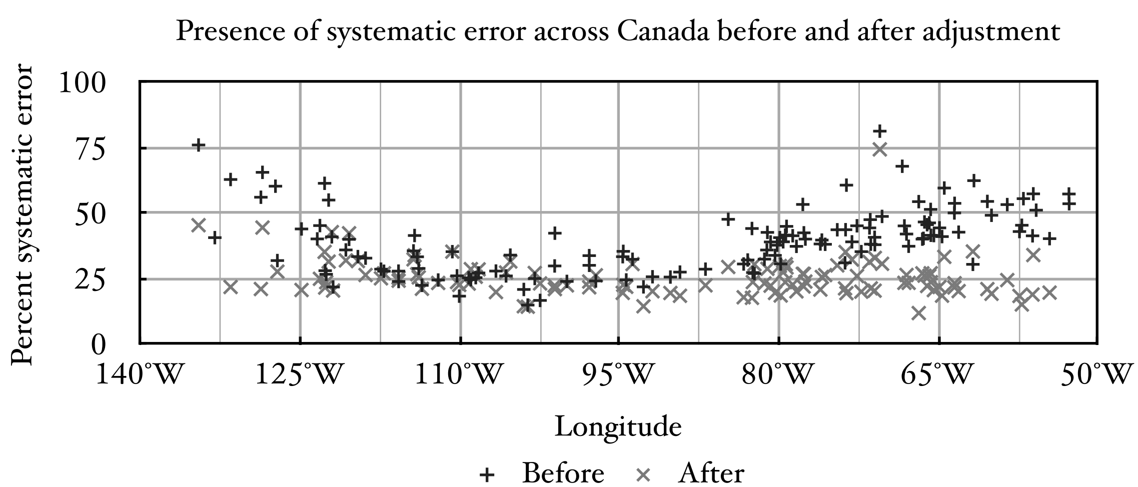


Figure 5.10: Percentage of systematic error before and after the adjustment between the 0.25° CMORPH product and the EC 6-hourly observational data at each manual station across Canada for June through October from 2002 to 2011.

June through October from 2002 to 2011 as a function of longitude before and after the adjustment. The trends for EC 6-hourly data are similar to that of the APC2 dataset, as mentioned previously. While it appears that western and eastern Canada have similar percentages of systematic error, eastern Canada has concentrations of stations at roughly 80°W and 65°W with 42% systematic error, while those in western Canada are spread out in terms of longitude.

Relating this figure to the changes in equitable threat scores, one can see that regions with large amounts of systematic error benefited greatest from the adjustment, namely eastern and western Canada. Central Canada which is comprised mostly of random error does not benefit as much from the adjustment. That said, the adjustment can only be expected to reduce systematic error and not random error. While this observation is based off of the systematic error before the adjustment, Figure 5.10 also shows the systematic error after the adjustment. Not surprisingly, the portion of systematic error is reduced with the adjustment methodology. It is, however, only effective up to a certain degree, in that it cannot remove all of the systematic error because the definition of systematic error differs from the adjustment methodology.

5.6 Key points

- The equitable threat scores of the satellite products are different when compared with the APC2 data or the EC 6-hourly data.
- The satellite products are in better agreement with the 6-hourly datasets, most likely because they both underestimate rainfall.
- Estimates of rainfall over snow covered ground are removed with minimal false alarms and good agreement with the methodology applied internally with CMORPH.
- Based on the availability of data after estimates over snow covered ground are removed, the applicable range for the satellite products in Canada (below 60°N) is June though October.
- Two methods for adjusting the satellite data are detailed and evaluated, based either on

historic monthly mean values or the previous thirty days.

- Cubic interpolation was found to be the preferable for the adjustment considering the previous thirty days.
- The adjustment considering the previous thirty days produced greater skill in the satellite product relative to the reference field, and will be implemented before the satellite data is assimilated using CaPA.
- The performance of the adjustment is closely related to the presence of systematic and random error in the satellite products across Canada.
- Central Canada, which consists mostly of random error, benefited least from the adjustment while western Canada benefited the most.

Part 6 | Satellite data as a background field

In its current form, CaPA assimilates the output from GEM with observations every six hours. One application of the satellite data could be in place of GEM, meaning the satellite data would be assimilated with observational data. In this scenario, two evaluations can be made. The first is the change in skill of the satellite products as a result of the assimilation. This evaluation should be straightforward, with an increase in skill overall. The second is the applicability of using the satellite data instead of GEM. Given that the satellite products have shown trends in performance with space and time, the skills relative to GEM are also explored in space and time. The following sections perform such comparisons to test the validity of replacing GEM with one of the satellite products.

The satellite data used here are adjusted to be consistent with the previous thirty days, as detailed in Subsection 5.3.1. The stations considered in the adjustment are not constrained to those in Canada. Any station within Canada and the United States which passes the quality controls are used in the adjustment. In other words, the satellite data are adjusted over their domain (30°N to 60°N) and not just Canada. Doing so better represents an operational form, as the adjustment in places like southern Ontario are influenced by the United States stations on either side. Also, strong features such as the overestimation in central Canada, shown in Figure 4.6 extend into the United States, so including United States stations ensure continuity close to the border. In the evaluation between the satellite products and GEM, data from a leave-one-out cross-validation is used for manual Canadian stations below 60°N (limited by the satellite data's domain).

With data from the leave-one-out cross-validation, the bootstrapping method applied in previous CaPA studies is used here to find various skill scores and their associated uncertainties. For each of the 100 bootstrap samples created, a location and time are randomly selected, from which a search radius is created with an exponential distribution in space and geometric distribution in time. All analysis values within the radius are collected, and if the density is greater than 0.001 stations/km^2 only a subset is retained. This is repeated until the number of analysis values in the sample equals those in the original domain. From the samples, expected skill scores and 80% confidence intervals are calculated (Lespinas et al., 2014).

As skill scores in this part, frequency bias indices, equitable threat scores, and departures from partial means are used as functions of binned and threshold rainfall magnitudes. The frequency bias indices and equitable threat scores presented here are calculated the same way as before, with Equations 4.6 and 4.4, respectively. Departure from partial mean (DPM) is the only score used here for the first time. It is simply a measure of the mean rainfall below a given threshold and is plotted as the difference between that of the analysis and the reference field.

The comparisons in this part between GEM and the satellite products are made through the use of two types of plots. Within the main text, equitable threat scores are calculated as functions of rainfall thresholds, and in the appendix, differences in equitable threat scores between the satellite products and GEM are presented with 80% confidence intervals. The threshold based plots in the main text are for more of a generalized evaluation, providing some insight into the magnitudes of skill present in the analyses, while the differences presented in the appendix are intended for comparisons. For both plots, the scores are the expected values from the bootstrapping method, and therefore account for temporal and spatial densities.

In the following two parts, satellite data as a background field and as an additional source, the evaluations are of similar regions and times. They both look at the skill scores overall, regionally, monthly, and 6-hourly. While each part contains some additional evaluations, these core evaluations are all performed in the same manner. They are all derived from analyses performed from 2002 to 2011, for the months of June through October, and of manual stations within Canada. For regional, monthly or 6-hourly evaluations, subsets of the entire temporal and spatial domain are used. The scores are based on the previously detailed bootstrapping method, and depending on the plot, will also show the 80% confidence intervals to identify statistically significant changes.

6.1 Adjusting with manual stations versus all stations

The adjustment of satellite data with the methodology of Section 5.3.1 can be completed with all of the stations or a subset there of. While using a subset of the available stations more emphasis is placed on the interpolation, but the sole use of the best stations may produce better results — a balance of quality and quantity. In this section, two groups of stations within Canada are defined: all available stations and manual stations. The effects of using them for interpolation and as a reference field is explored.

To be consistent with the current works on CaPA, the final evaluations of CaPA assimilation are based solely on manual stations (Lespinas et al., 2014). The following considers the differences in equitable threat scores in which both subsets (manual or all stations) are used in the adjustment. The analyses are for all of the satellite products from June through October, for 2002 to 2011 (2003 and onward for PERSIANN-CCS). It should be stated that while the equitable threat scores are based on a leave-one-out cross-validation, no stations were removed in the adjustment processes. The results from which are apparent in the following discussions.

Figure 6.1 shows the differences in equitable threat scores where all the satellite products are adjusted with all or just manual stations and used as background fields in an assimilation using CaPA over the previously stated temporal domain. From this figure, one can see that using the entire observational network has minimal effect when compared with just manual stations. The only significant change is for rainfall greater than 20 mm/6-hour in the PERSIANN 0.25° analysis.

The sudden change in equitable threat score for rainfall greater than 50 mm/6-hour for the

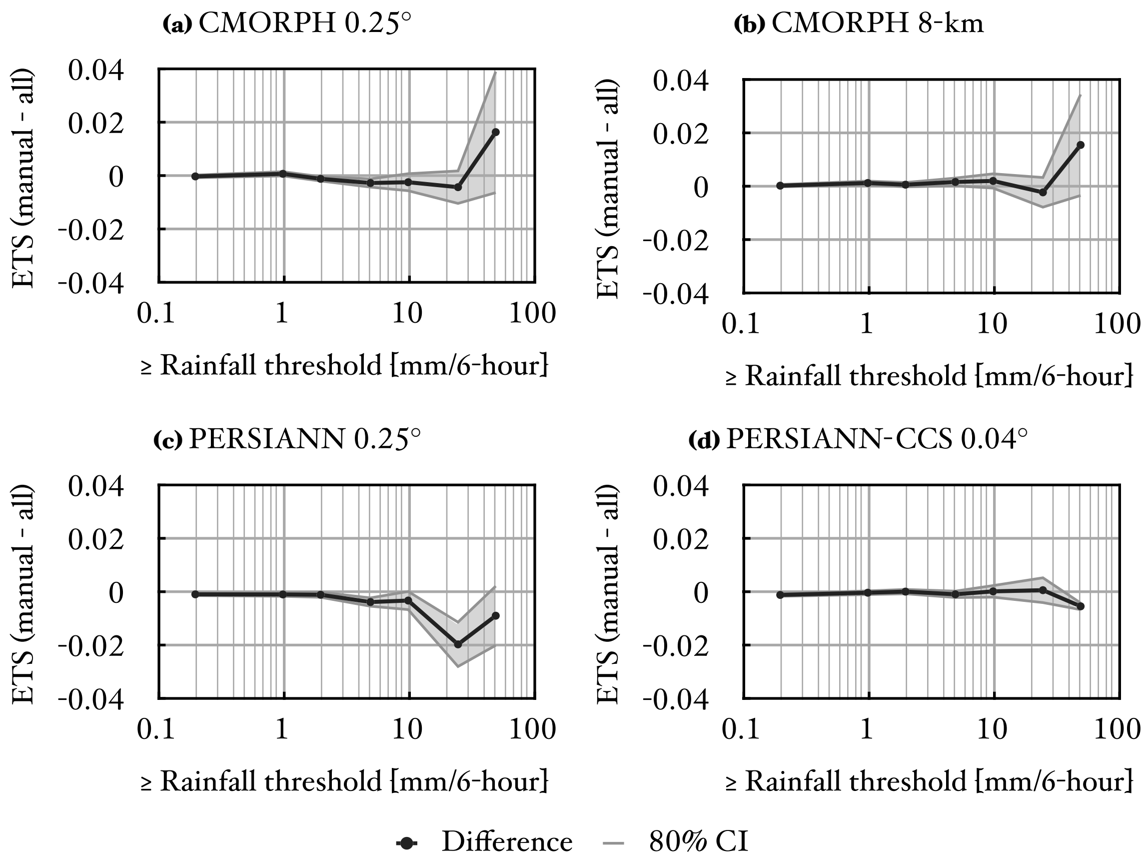


Figure 6.1: Differences in equitable threat scores and 80% confidence intervals from adjusting the satellite products with either manual or all stations. The scores are from a leave-one-out cross-validation of CaPA, where each satellite product is used as a background field for June through October, 2002 to 2011 (2003-2011 for PERSIANN-CCS). As a reference field, only manual stations are used.

CMORPH products is related to the nature of the stations used. For both resolutions, adjusting with all stations produces a greater frequency bias above 10 mm/6-hour than the adjustment with just manual stations (Figure D.2). Considering the station data alone, the empirical distribution functions over the same time period show that the manual stations do not report as many extreme events (Figure D.1), and as such are not introduced into the adjustment. The increased equitable threat score for the adjustment using manual stations relative to that from all stations in Figure 6.1 then suggests that these extreme events reported by all stations are negatively affecting the adjustment for CMORPH. The increase in skill for great magnitudes of precipitation is not statistically significant, but is observed across multiple datasets (all but PERSIANN-CCS).

This increase in skill for an adjustment with manual stations is also present in Figure 6.2, which is the same as Figure 6.1, but the scores are calculated against all stations. In doing so, an adjustment considering all stations is preferable across all datasets and rainfall thresholds because, unlike the previous, the adjustment here with manual stations is being evaluated at points which did not have observational data in the adjustment. That is, the adjustment with manual stations is being evaluated at points based purely on interpolation. This then is not a fair test, but that is not the intention of this section, which is to explore the effects of using all stations or just manual stations in the adjustment and the final evaluation of CaPA. Based on the previous two figures, performing an adjustment based on all stations will not negatively

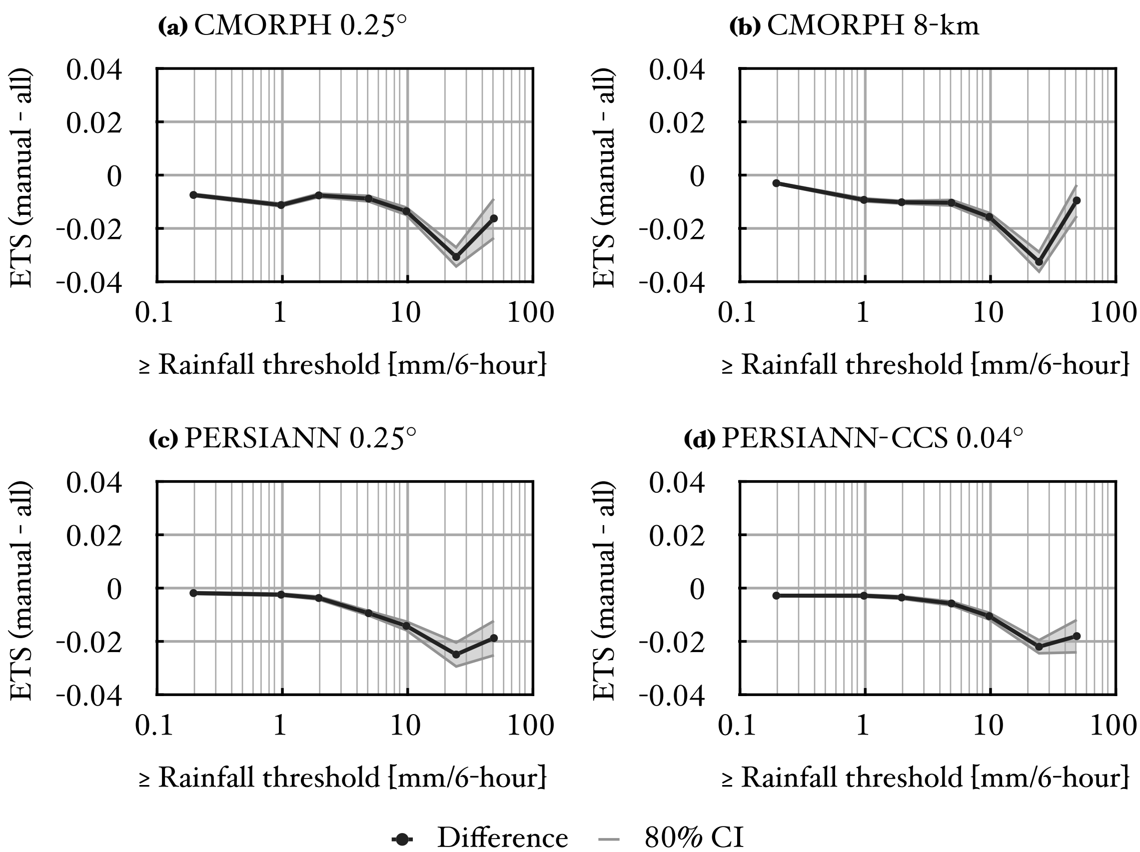


Figure 6.2: Same as Figure 6.1, but using all stations as the reference field.

affect the evaluation of CaPA results in a statistically significant way. For the remainder of this project, evaluations are of satellite data adjusted relative to all stations, and compared to manual stations to be consistent with previous CaPA literature.

6.2 Visual comparison of the raw and assimilated satellite products

Samples of raw data for three of the satellite products are shown in Figure 6.3, specifically, July 13, 2006 from 06:00 to 12:00 UTC. Comparing the 0.25° CMORPH and PERSIANN products in Figures 6.3a and 6.3c, respectively, the first thing one notices is the spatial extent of rainfall. This CMORPH time step consists of many smaller light rainfall events, where PERSIANN's estimates are more spatially focussed. This seconds Figure 4.9 which shows that the 0.25° CMORPH product uniquely overestimates light rainfall frequency amongst the considered satellite products. The spatial presence of events are generally consistent between CMORPH and PERSIANN, which is not entirely surprising since the estimates are derived from the same satellite retrievals. It is interesting, however, that the gradient of rainfall rates outward from the centre of a given event is steeper in CMORPH's estimates. Which is to say that CMORPH transitions from heavy rainfall to light over a shorter distance.

The differences between the low and high resolution CMORPH products are shown in Figures 6.3a and 6.3b, respectively. Due to the decrease in resolution, the lower resolution product is smoother and spatially contains more rainfall. Also, the maximum rainfall estimates are reduced due to the averaging over a larger spatial area. A benefit of the high resolution product is that the small rainfall events across the country are less obvious since they are only from a couple of grid cells in the high resolution product.

Figure 6.4 shows the same date as Figure 6.3, but for eastern North America with the raw data on the left and the CaPA assimilations with the satellite data as a background field on the right. Also, the station data used in the assimilation are shown as point observations on the same scale. Comparing the raw satellite data with the stations shows that the most significant event is spatially common between the two datasets. Subjectively, the raw CMORPH data appears to better match the stations for the largest rainfall event in its most intense region. However, the spatially smaller, but still heavy rainfall to the lower left is better represented by PERSIANN. In fact, for light rainfall in the lower left region, PERSIANN matches quite well, while CMORPH misses some light rainfall.

As a result of these products' assimilation with station data using CaPA, they both experience a change in rainfall rate and frequency. For both CMORPH and PERSIANN, the spatially

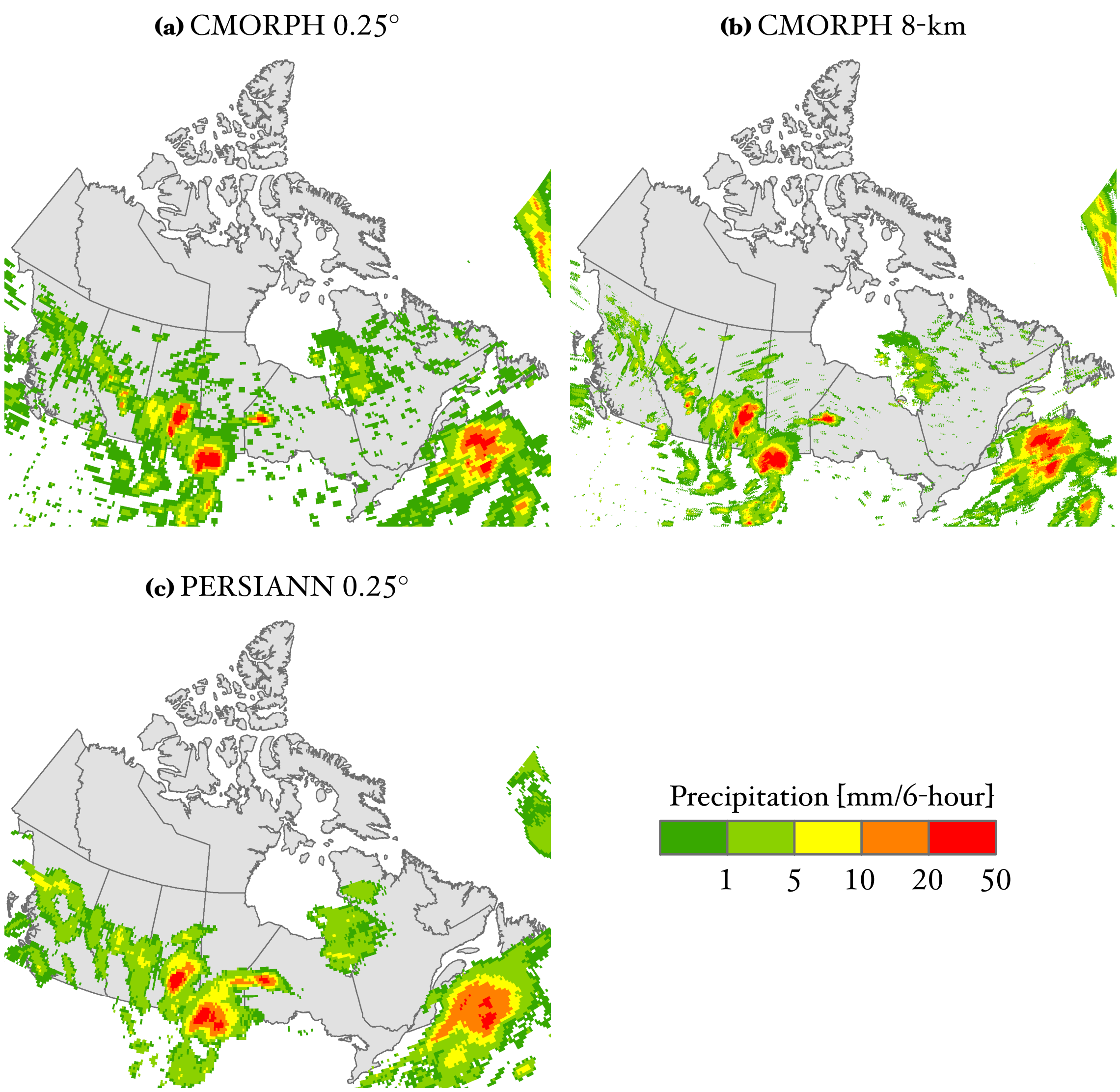


Figure 6.3: Raw satellite data products for July 13, 2006 from 06:00 to 12:00 UTC.

extent of the event increases along with adjustments for the missed rainfall. Changes of this form are not surprising as the satellite products were shown to generally possess a negative bias (underestimate) in Section 4.2.2. The second largest event to the lower left, which was well

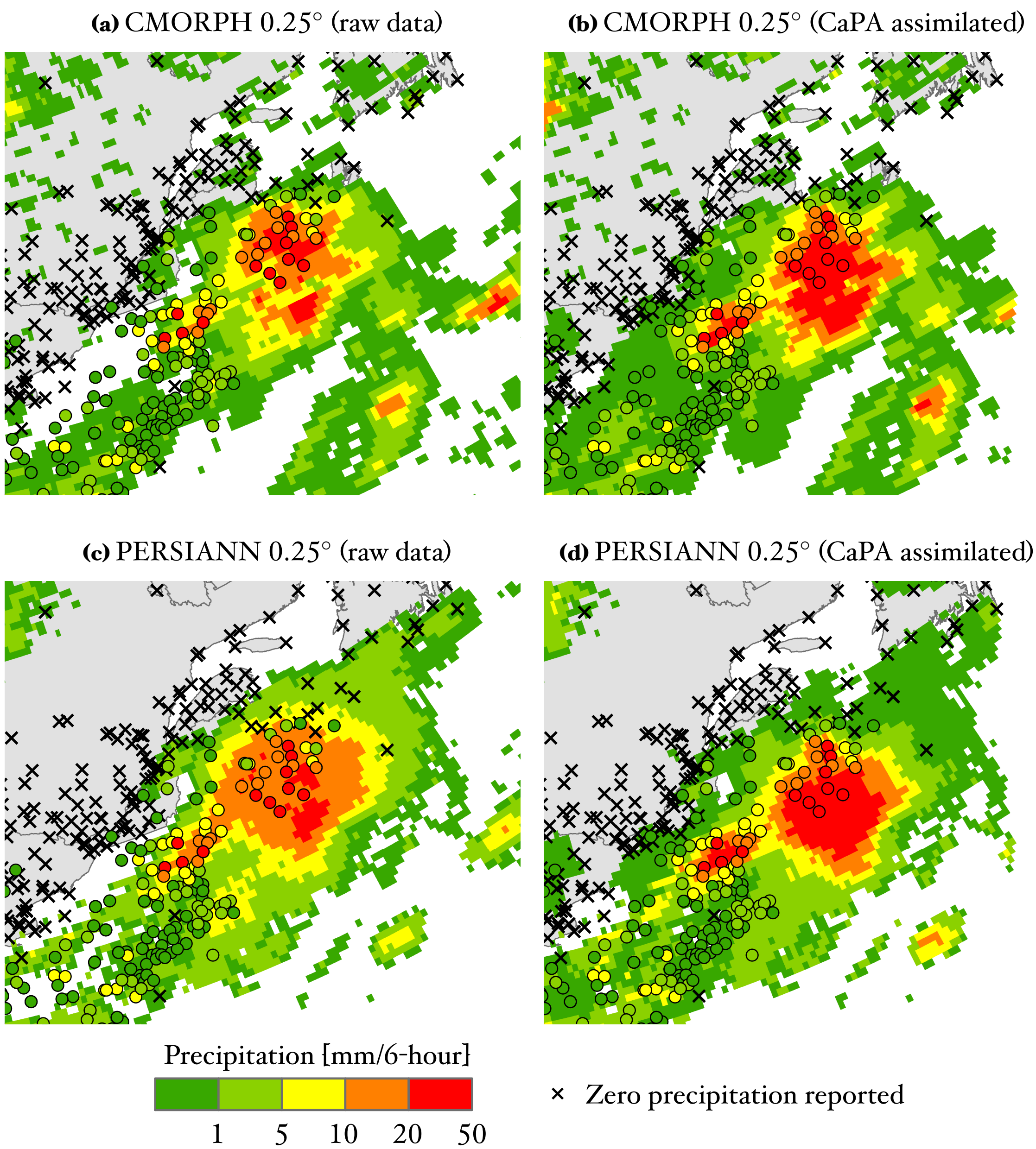


Figure 6.4: Satellite rainfall estimates for July 13, 2006 from 06:00 to 12:00 UTC over eastern Canada. The left figures show raw estimates, and the right showing the assimilations with CaPA using the satellite data as a background field. Stations used in the assimilation are also shown.

detected by PERSIANN and much less so by CMORPH, is present in both final assimilations.

6.3 Overall

Since CaPA assimilates a gridded dataset with observations, its effectiveness can be shown by the skill scores of the background field before and after. Plots in this section show just that, but this relative change is not the main objective of this part, which is to evaluate the use of a satellite product instead of the current background field, GEM.

As an initial evaluation of replacing GEM with the satellite products, the results are compared via equitable threat scores over the entire temporal and spatial domain of the project. This evaluation can be seen in Figure 6.5, where leave-one-out cross-validations of the assimilations with CaPA are performed from June through October, for 2002 to 2011. As implied above, the equitable threat scores for each satellite product before and after the assimilation are plotted alongside that of assimilated GEM. While Figure 6.5 is a function of rainfall thresholds, Figure D.3 in the appendix compares the differences between the equitable threat scores of the differences between the satellite products and GEM versus binned rainfall with 80% confidence intervals

From Figure 6.5, one can see that CaPA is quite successful in increasing the skill of all the satellite products across all rainfall thresholds. In these plots, 'before' is the skill of the satellite products that have been rescaled to the previous thirty days, as detailed previously. The assimilation was particularly helpful with the 0.25° PERSIANN product, which was clearly inferior to its CMORPH counterpart before the analysis, but much closer afterwards. However, CMORPH still produces greater skill scores.

In terms of performance relative to GEM, Figure D.3 shows that potential improvement is mainly for greater magnitudes, although uncertainty also increases. Previously, binned equitable threat scores for the satellite products across Canada in Figure B.2, albeit daily accumulations, show that the satellite products possess greater skill with increasing rainfall magnitudes. This was discussed in Part 4, where satellite products generally have difficulties with light rain-

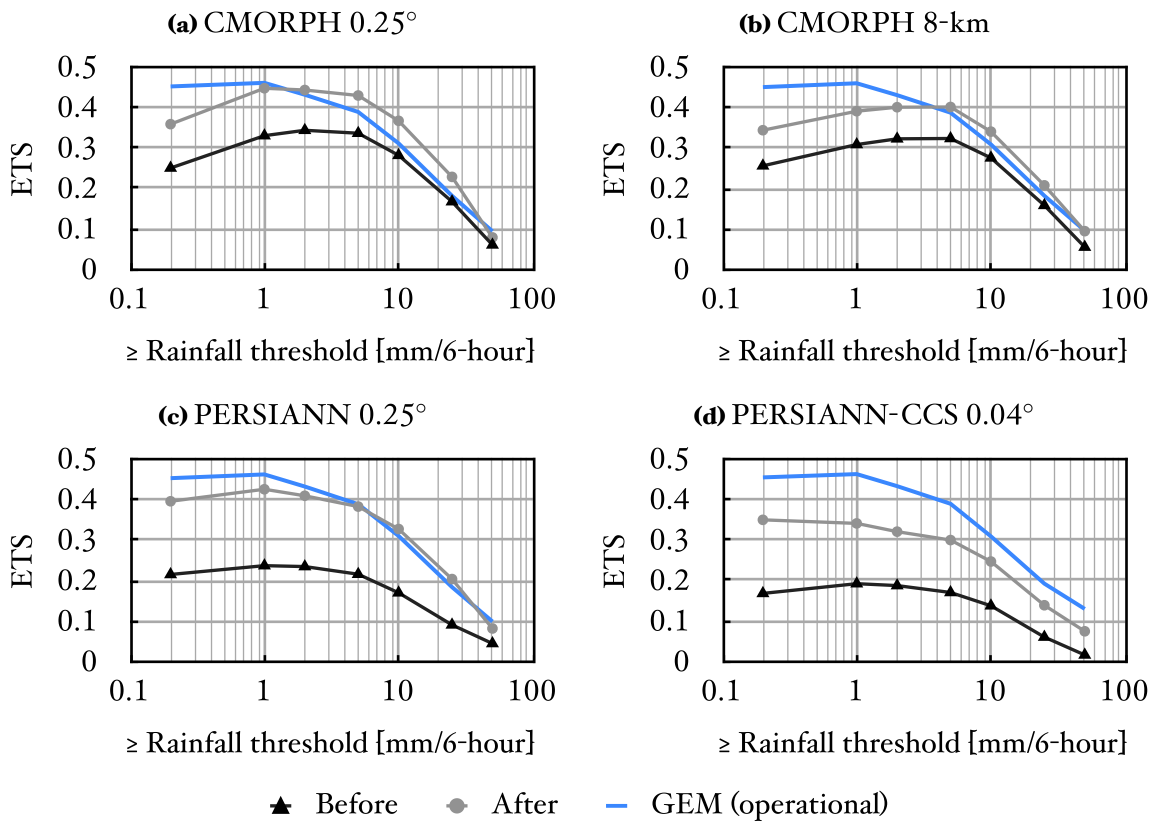


Figure 6.5: Equitable threat scores of various satellite products before and after assimilation using CaPA, with the operational form of CaPA with GEM plotted in blue. Scores represent June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

fall. Hence, it should come as no surprise that if the satellite were to be preferable to GEM, it would be for greater magnitudes of rainfall.

Of the satellite products, the 0.25° CMORPH product posses the greatest skill, which was shown to the be case previously, prior to any assimilation. For all months and across Canada it is significantly better than GEM for rainfall between 10 mm/6-hour and 50 mm/6-hour, above which it possess too much uncertainty. The 0.25° PERSIANN product is also statistically significant for this range, but to a lesser degree. As has always been the case thus far, the PERSIANN-CCS product produces the lowest equitable threat scores.

Between Figures 6.5 and D.3, an interesting comparison is that of the 0.25° PERSIANN and the 8-km CMORPH products. The threshold based evaluation of Figure 6.5 suggests that the 8-km CMORPH product produces more skill than the 0.25° PERSIANN product, but as previously stated, only PERSIANN shows significant improvement over GEM. This is because of the relative performances in the final bin of Figure D.3 alone, where the 8-km CMORPH product outperforms the 0.25° PERSIANN product, although there is a great amount of uncertainty associated with this magnitude of rainfall.

While assimilating the satellite products with observations using CaPA has greatly increased their skill for all magnitudes of rainfall, it is not unanimously greater than that of GEM. There are, however, magnitudes of rainfall where the satellite products perform better than GEM, hinting that there may be specific regions or times in which the satellite products are preferable. Ideally, the poor performance as a background field is limited to a specific region or time, so the applicability of satellite data can be beneficial for particular studies. The remaining sections in this part explore the relative increases in skill between the satellite products and GEM for spatial trends across Canada, temporal trends (monthly and 6-hourly), and during convective events, which were shown in Part 4 to be the most promising.

6.4 Regionally

Figure 6.6 shows the same data used in Section 6.3, but separated for regions of Canada as defined previously in Figure 4.1. That is, equitable threat scores between the satellite products and GEM are found for June through October from 2002 to 2011 against manual stations in Canada. This figure shows how improvements relative to GEM are not only for greater magnitudes, but possess spatial dependence.

In eastern Canada, the satellite products generally showed the most skill when compared with station data, but apparently GEM also has this property in that the satellite products are minimally better. The binned data in Figure D.4c of the appendix show that any significant increase is for the 0.25° CMORPH product for rainfall between 10 mm/6-hour and 50 mm/6-hour. Above this the changes are not significant, and below it GEM is strongly preferable.

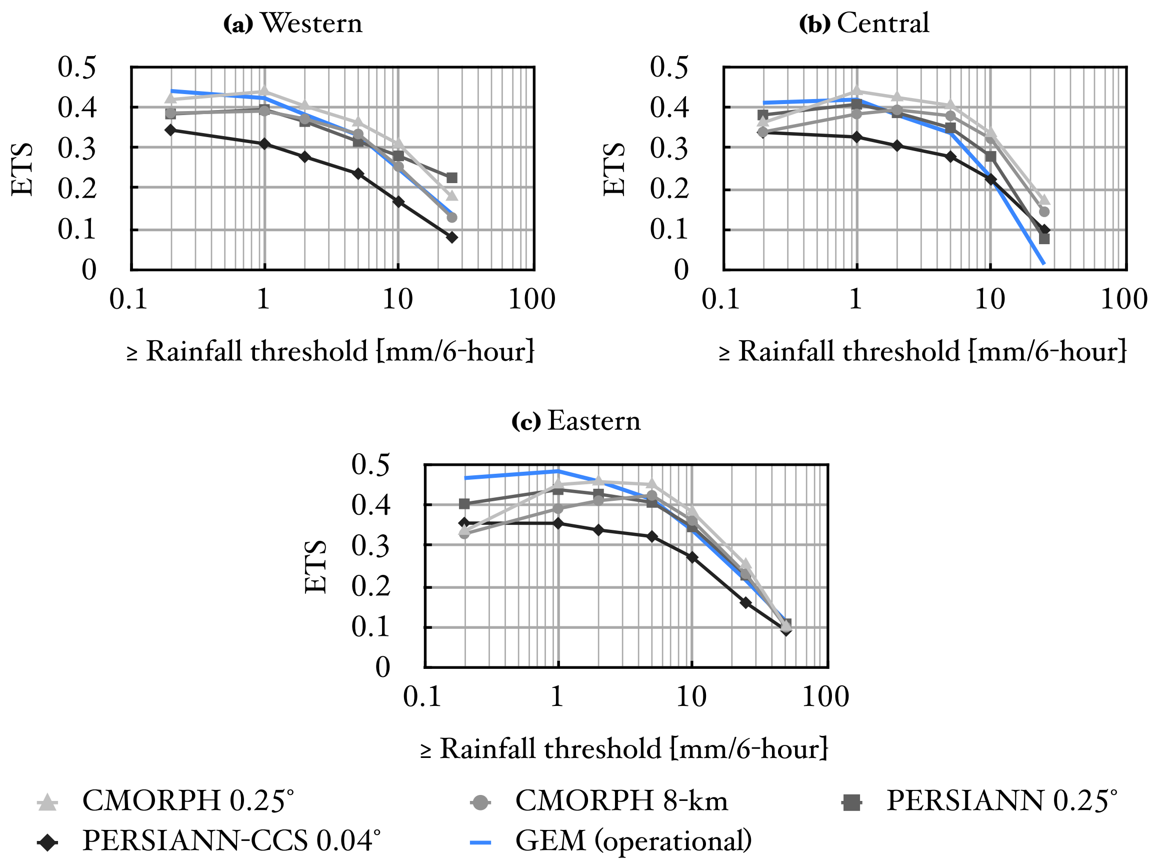


Figure 6.6: Regional equitable threat scores of the assimilated satellite products with the operational form of CaPA with GEM plotted in blue. Scores represent June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

The second best performing region of Canada for the satellite products, as proposed in Part 4, is in central Canada. Figure 6.6b shows that for rainfall greater than 2mm/6-hour, all of the satellite products produce greater equitable threat scores. The binned data in Figure D.4b show that this increase in skill is from greater magnitudes of rainfall. Unlike that in eastern Canada, the skill is so much greater for the rainfall between 10 mm/6-hour and 50 mm/6-hour, that it is clearly statistically significant based on the 80% confidence intervals. Even for the PERSIANN-CCS product, whose performance has not been well, it shows statistically significant improvements compared to GEM for rainfall between 25 mm/6-hour and 50 mm/6-hour. The 0.25° CMORPH product, which is generally the best performing, alone shows statistically significant improvement for rainfall between 5 mm/6-hour and 50 mm/6-hour. Given the potential of the satellite products in this region, the scores related to convective events is explored in greater detail in Section 6.6.3.

Finally for western Canada, where the satellite products had the most difficulty estimating precipitation, Figure 6.6a suggests that the 0.25° CMORPH product produces slightly greater skill for rainfall above 2 mm/6-hour, while the 0.25° PERSIANN product is limited to the extreme precipitation. This is seconded in the binned data of Figure D.4b where the 0.25° CMORPH product is only significantly better than GEM for rainfall between 10 mm/6-hour and 25 mm/6-hour, above which it is beneficial but has too much uncertainty to conclude significant change. The 0.25° PERSIANN product is interesting in that the performance of the other satellite products relative to stations increase with magnitude, but often peaks and decreases slightly for the truly extreme rainfall. In Figure D.4b, the skill of this product relative to GEM is only increases, to the point where for rainfall between 25 mm/6-hour and 50 mm/6-hour it is the only product showing significant improvement. Based on this previously found performance of the PERSIANN products, this score should only be considered with caution.

6.5 Monthly

In Subsection 4.2.2 the monthly performance of the unadjusted satellite products was explored and presented in Figure 4.8. The general trend was for the satellite products to perform the best in July and August, with comparable performance for greater magnitudes in September. The variability of performance was greatest for the 0.25° CMORPH product, mostly due to poor performance in October. Figure 6.7 shows the monthly performance of the satellite products after they have been assimilated with stations via CaPA from 2002 to 2011 (2003 onward for PERSIANN-CCS). For reference, the assimilation of GEM and stations is plotted in blue for the same time period. Month to month, the skill scores of the satellite products after assimilation are similar for a given product, presumably because those months which had lower scores benefited most, similar to before where the performance gap between the 0.25° CMORPH and PERSIANN products decreased with the assimilation. The trends of each satellite product for varying magnitudes remains unchanged with the assimilation. For example, the 0.25° CMORPH product continues to have a prominent peak in performance, whereas the other products are flat at lower magnitudes before decreasing for greater magnitudes.

Based on the raw data, the satellite products may be the most applicable in July, August, and September. Figure 6.7 shows that relative to GEM the satellite products perform well during July and August for rainfall greater than 2 mm/6-hour. The above average performance of the 0.25° CMORPH product expresses itself here by having greater equitable threat scores, and therefore beneficial over GEM at even lower magnitudes of rainfall. The satellite products also perform well in June for greater rainfall rates, but not to the same degree as July or August. By September, the performance of GEM is great enough that it outperforms the satellite products. Finally, in October none of the satellite products produce enough skill to be better than GEM.

What could be causing this increased skill in June, July, and August for the satellite products at greater magnitudes? The best explanation is based on the presence of convective events. Subsection 4.2.2 not only showed an overestimation by the satellite products in regions where convective storms are known to occur, but overestimations in proportion to the frequency of observed convective events. Furthermore, these overestimations, which were mostly limited to July and August, were great enough to have an influence in the overall monthly bias of Figure 4.5, making it possible here as well. Considering Figures 6.6 and 6.7 together, the mostly applicable time and space for use of satellite products instead of GEM, if at all, would be in central Canada during July, August, and possibly June — locations and times consistent with intense short-lived convective events. This further warrants the exploration of the effects of convective events on the performance of the satellite products relative to GEM, which is presented in Subsection 6.6.3

While Figure 6.7 provides a smooth trend in performance through time, the significance of

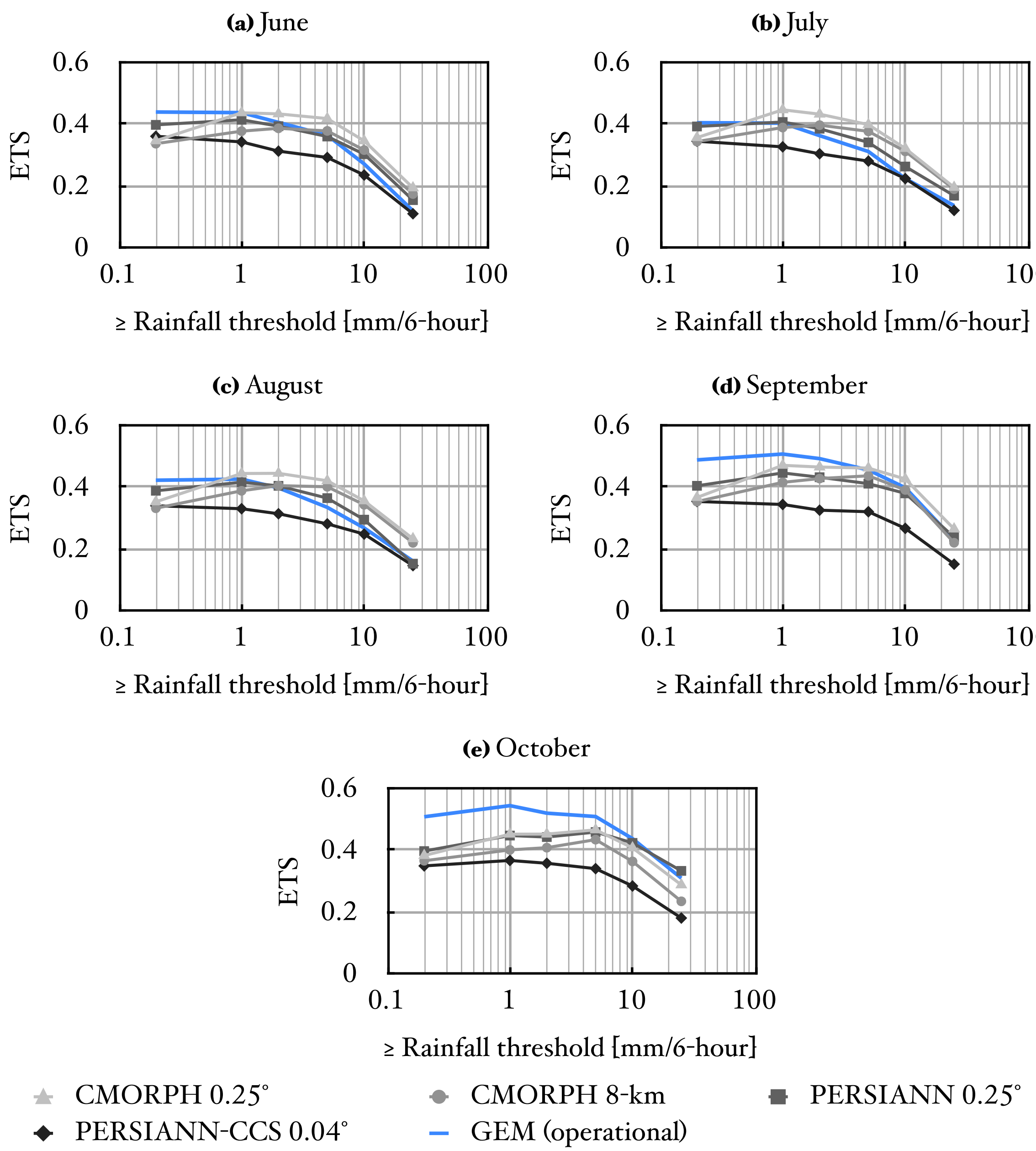


Figure 6.7: Monthly equitable threat scores of the assimilated satellite products with the operational form of CaPA with GEM plotted in blue across Canada from 2002 to 2011 (2003 onward for PERSIANN-CCS).

binned differences relative to GEM in Figure D.5 explores them in greater detail. The results in this figure are similar in that the lower resolution CMORPH product produces the greatest skill scores even when considering statistical significance, but also in that the skill of the satellite products is derived from the greater rainfall rates.

What Figure D.5 shows that Figure 6.7 does not is the nature of performance at the high end. During June, the increased performance relative to GEM is mostly from rainfall rates of 25 to 50 mm/6-hour, where it is statistically significant only for CMORPH (both resolutions). The lower 10 to 25 mm/6-hour bin shows smaller increases in skill, but are only significant for the 0.25° CMORPH product. During July, the skill for estimating rainfall between 10 to 25 mm/6-hour are not only greater for the satellite products, but possess relatively less uncertainty. As a result, all satellite products but PERSIANN-CCS show significant gains compared to GEM. For rainfall in the bin above this, the same satellite products show less improvements relative to GEM and are of greater uncertainty. So for rainfall between 25 to 50 mm/6-hour, only the CMORPH 0.25° product shows little statistically significant improvement, if at all. Considering statistical significance, August is the best performing month, as scores are greater for the satellite products relative to GEM for rainfall between 10 and 50 mm/6-hour than for any other month. This greater skill is accompanied by similar or less uncertainty. For rainfall between 10 and 50 mm/6-hour, both CMORPH resolutions show exceptional skill, and even PERSIANN-CCS, which has been often disregarded, is similar in performance to GEM. So between July and August, which had similar improvements over GEM when evaluated against rainfall thresholds, August is the best performing.

During September and October, when the satellite products do not perform nearly as well as GEM overall, they are the most comparable for the greater magnitudes. Both 0.25° products are not significantly different than GEM for rainfall between 10 and 50 mm/6-hour. The same is only true for rainfall between 25 and 50 mm/6-hour in October.

6.6 6-hourly

The assimilation of the satellite products with observations using CaPA are performed on 6-hourly time steps, and until thus far, all time steps have been treated equal. The following subsections explore the performance of the satellite products relative to GEM for the individual increments.

6.6.1 Across Canada

To start, equitable threat scores for the satellite products have been calculated for each 6-hour time step and plotted in Figure 6.8. As before, the scores are based on data for June through October, from 2002 to 2011. Consistent with the previously found skill scores, the

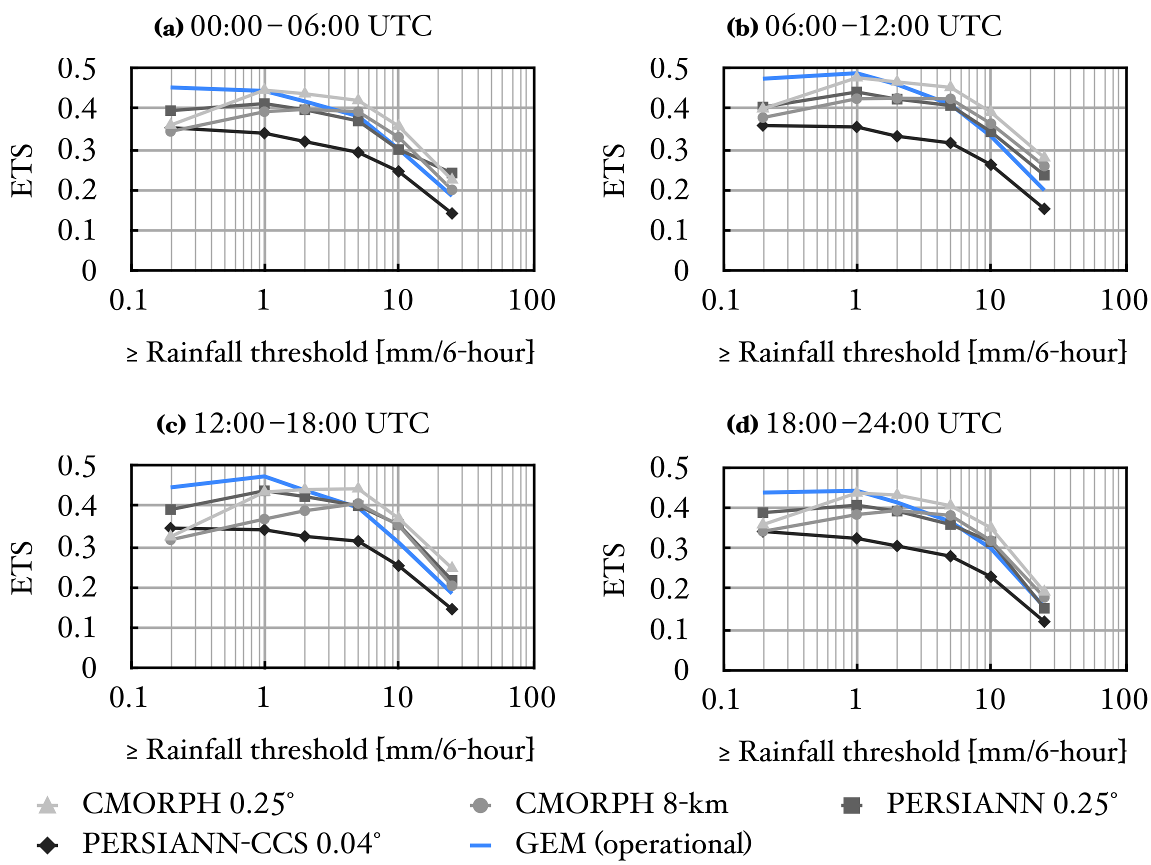


Figure 6.8: Equitable threat scores across Canada for assimilated satellite products and GEM for each 6-hour time step in the temporal domain of June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

satellite products outperform GEM, if at all, for greater rainfall magnitudes. What this figure shows that is new is that there is some trend in the equitable threat scores through time for a given day. For example, GEM has the most skill between 06:00 and 12:00 UTC. The satellite products too, generally produce their greatest skill at this time, as well as between 12:00 and 18:00 UTC.

For a given satellite product, the performance of the different time steps is small but consistent. Figure 6.9 shows the same data but plotted for the low resolution products, with that of GEM for all time steps as a reference. For CMORPH in Figure 6.9a, the performance for rainfall above 2 mm/6-hour is always greatest between 06:00 and 12:00 UTC, followed by 12:00 to 18:00 UTC. This increased performance is clearer for PERSIANN in Figure 6.9b. While the performance of the satellite products is similar between time steps, there is something to be said for the consistency between 06:00 and 18:00 UTC relative to the rest, also the fact that it is observed for two different products.

Returning back to the more important comparison, which is that of the satellite products relative to GEM. As in the previous sections, the binned differences in equitable threat scores have been calculated for the same subsets as in Figure 6.8, and are presented in Figure D.6. Relative to GEM, only the lower resolution (0.25°) products show significant improvements

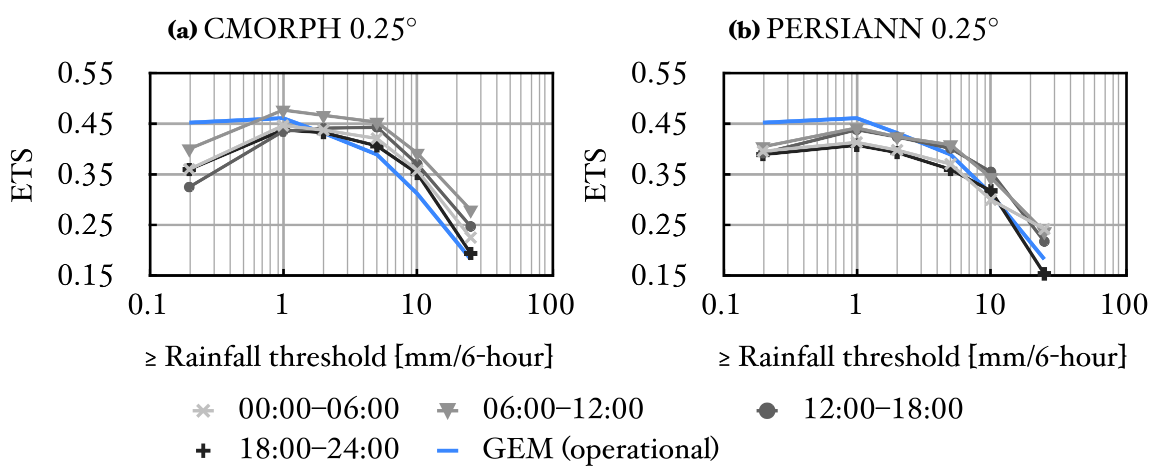


Figure 6.9: The same data as Figure 6.8, but plotted for a given 0.25° satellite product.

in equitable threat score. Uncertainty aside, nearly all products but PERSIANN-CCS show more skill than GEM for rainfall between 10 and 50 mm/6-hour. The only exception being 0.25° PERSIANN between 18:00 and 24:00 UTC, which is slightly worse.

The previously discussed trends in the satellite products remain. The 0.25° CMORPH product continues to be the best of the satellite products as it is significantly better than GEM between 10 and 25 mm/6-hour in all time steps, while the low resolution PERSIANN is only significantly better between 12:00 and 18:00 UTC in this bin. As Figure 6.9 implies, the best time steps for the satellite products is between 06:00 and 18:00 UTC, during which time the low resolution CMORPH product produces significantly better skill scores for rainfall between 25 and 50 mm/6-hour. The low resolution PERSIANN product is only significantly better than GEM in this bin and between 00:00 and 12:00 UTC.

6.6.2 Resolution dependence

While the plots thus far considering 6-hourly time steps have shown that the high resolution satellite products do not perform as well as their low resolution counterparts, other measures of performance highlight the advantages of the high resolution products.

Figure 6.10 shows frequency bias indices for the satellite products post assimilation, as functions of binned precipitation. They are calculated via the previously described bootstrapping methodology, so changes in spatial and temporal density are accounted for. Below 2 mm/6-hour, all the satellite products show the same trend in that they all overestimate the frequency of events. The smallest bin is very close to zero because of the inclusion of 0 mm/6-hour. Above 2 mm/6-hour, the performance of the satellite products begin to vary relative to each other as a function of rainfall magnitude. The PERSIANN 0.25° product is somewhat unique for rainfall between 10 and 50 mm/6-hour, in that it underestimates rainfall occurrence to greater degrees with increasing magnitude, relative to observations. This is in contrast to the other three products, all of which increase the number of rainfall occurrences relative to observations with increasing rainfall magnitude. An interesting property of the high resolution products is that they both ultimately overestimate rainfall frequency for the largest observed rainfall. Given that the two CMORPH products underwent the same assimilation, the differences presented in Figure 6.10 should be due to the grid size alone. Since the rainfall in a high resolution product is an average of a smaller area, it should be able to reproduce the magnitudes of the extreme events, which appears to be the case here.

Previously, rainfall between 06:00 and 12:00 UTC was a common time for the 0.25° CMORPH and PERSIANN products to perform well. This period also has unique properties in frequency bias indices for the satellite products. For this time step, the overestimation of events less than 2 mm/6-hour is less relative to other time steps, and the overestimation is greater between 10 and 50 mm/6-hour.

While the purpose of Figure 6.10 was to compare the high and low satellite resolution, GEM has been added as a reference. It is similar to the satellite products below 2 mm/6-hour as it underestimates the frequency of events. Where it deviates is above this magnitude, in which it continues to increasingly underestimate with increasing magnitude.

The effects of increased rainfall frequency can also be seen in measures of the departures from partial means, shown in Figure 6.11, following the methodology put forth at the beginning of this part. The values plotted here are the differences between a given analysis and that of the observed data (manual stations only). With this definition, positive values result from the analysis having a greater mean rainfall, negative values from the analysis having a lesser mean rainfall, and zero being ideal. The trend closely follows that of Figure 6.10, as expected, since an increased frequency results in a greater mean for a given threshold. In being plotted alongside GEM, the low resolution products appear to be preferable to GEM when considering precipitation under roughly 10 mm/day, above which they underestimate more so than GEM. The high resolution satellite products, however, produce means closer to the station data for the majority of thresholds due to the increased frequency of greater magnitudes. Consequently, the trends in the low resolution satellite products and GEM are similar in that they overestimate small rainfall and underestimate greater rainfall, however, the satellite products underestimate to a greater degree for all rainfall magnitudes. This means that

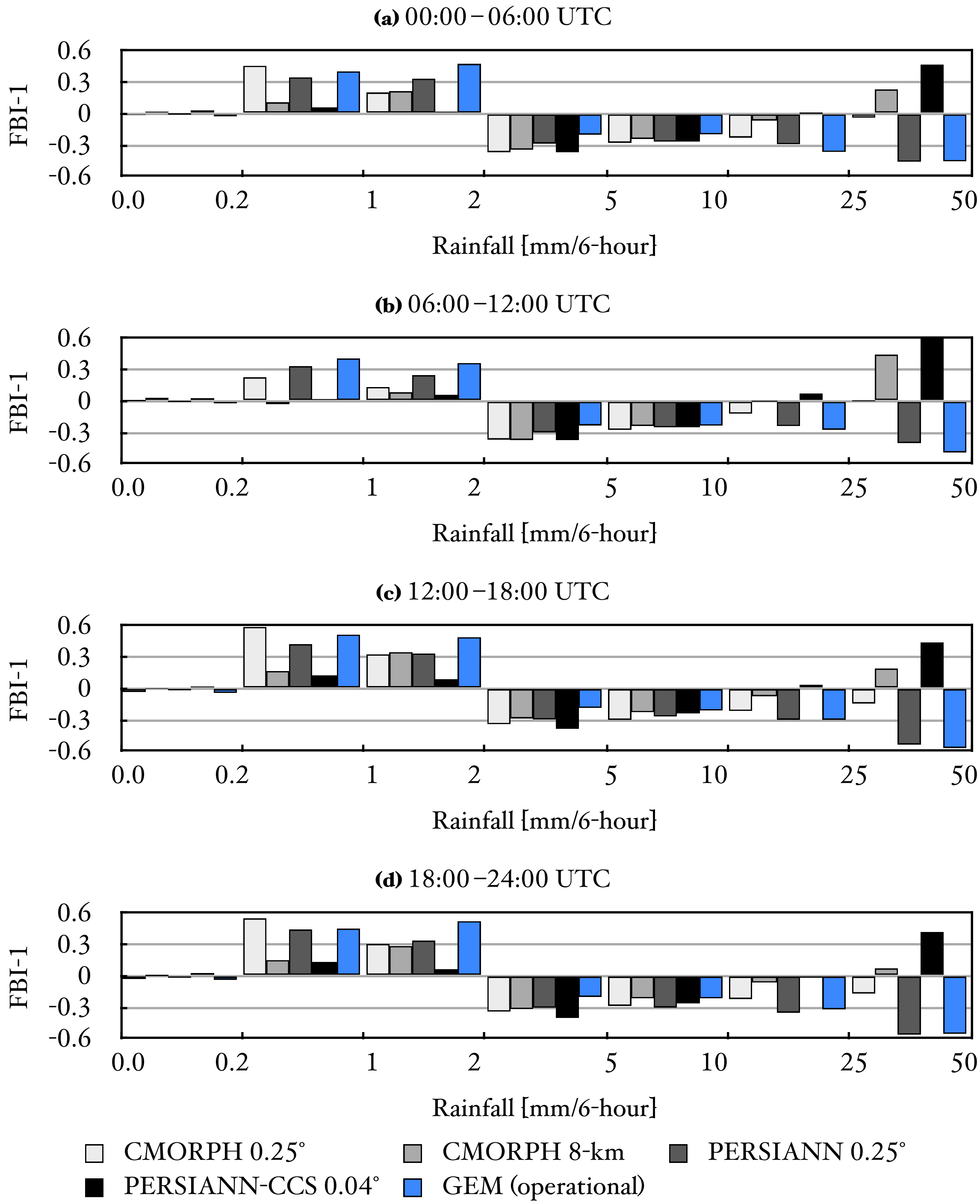


Figure 6.10: Frequency bias indices for various satellite products after being assimilated using CaPA, separated by time steps. Each plot uses data from all manual stations across Canada for June through October, 2002 to 2011 for the given time period. (2003 to 2011 for PERSIANN-CCS).

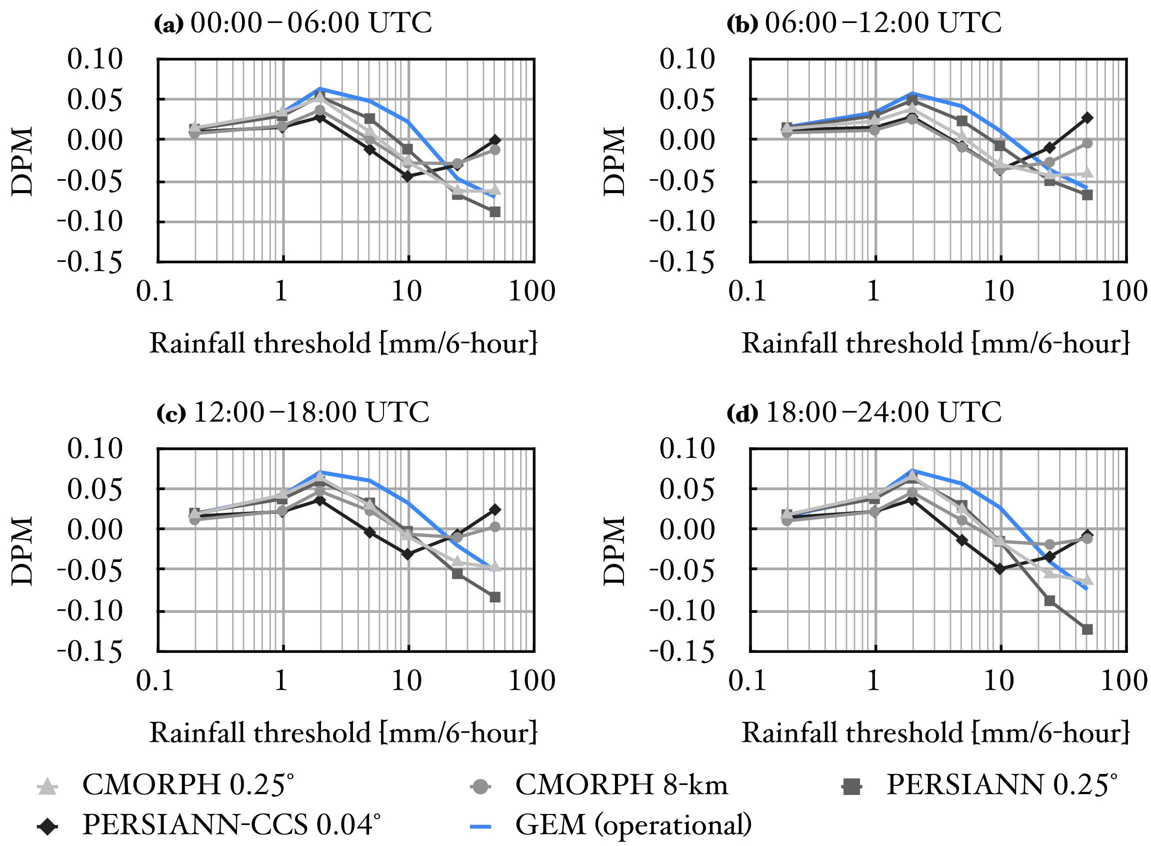


Figure 6.11: Binned departures from the partial mean relative to manual stations in Canada for June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS). The scores are based on means less than the rainfall threshold.

the low resolution satellite products are preferable only for smaller rainfall magnitudes. The high resolution products also underestimate small events, but overestimate larger events, and as such, are preferable across the entire rainfall magnitude spectrum by this measure.

6.6.3 Regionally

Finally, the results from the satellite datasets are explored as both functions of space and time, whereas before it was one or the other. Initially separating them independently aided in creating generalized summaries of the performance, where as here it is quite specific. Although similar conclusions could be drawn from this section, the scores presented here are based on fewer stations and therefore contain much more uncertainty.

To explore the satellite data in space and time, the same definitions are used as before. Assimilated satellite data and data from manual stations are separated by region and time step before calculating the scores. The overall temporal domain remains the same, being June through October, 2002 to 2011. In the interest of reducing clutter, this section only evaluates the 0.25° products, as evaluating the high resolution products in the way presented here will provide minimal information. This is because only equitable threat scores are used in this subsection, for which the high resolution products have been shown to be inferior to their lower resolutions counterparts. The equitable threat scores as function of rainfall thresholds are presented in Figures 6.12, 6.13, and 6.14 for western, central, and eastern Canada, respectively.

There are two worthwhile notes regarding these figures. First, as before, the scores are influenced by changes for greater magnitudes of rainfall, and as such the satellite data may appear preferable for a range of rainfall, but in fact it may only be for the greatest magnitude. This is why binned data are eventually presented and discussed. Second, the timezones indicated for the subfigures are in UTC for all regions. So for example, while the time period of 18:00 to 24:00 UTC is the afternoon in central Canada, it contains some morning rainfall in western Canada. All of the subfigures in this subsection are oriented in a similar matter such that the top row is during the night, and the bottom row is during the daytime. The terminology and orientation do not change for these types of plots from here on, and are shown in Table 6.1.

In using these figures to compare equitable threat scores, one can determine if a given product is favourable for a specific time or place. Specifically, this is intended to check if the satellite products after assimilation possess additional skill in the afternoon in central Canada, relative

Table 6.1: Sub-daily time periods and definitions.

00:00 – 06:00 UTC Evening	06:00 – 12:00 UTC Night time
12:00 – 18:00 UTC Morning	18:00 – 24:00 UTC Afternoon

to other regions and times. For both CMORPH and PERSIANN, they do not appear to express additional skill in estimating rainfall in central Canada, as the performance between regions for each time step follow the overall trends discussed previously, specifically eastern Canada having the most skill followed by central Canada. Similarly, the time steps generally follow the trend of improved performance between 06:00 and 18:00 UTC. If, for example, convective storms resulted in increased skill for this type of evaluation, one may see that while overall eastern Canada performs better than central Canada, a specific time step (afternoon/evening) would alone produce greater skill. This is, however, not the case. As to why there is no distinct change in skill for the satellite, a proper test would limit the temporal domain to July and August, as these are the months when convective storms are most prevalent in central Canada. It is entirely possible that any advantages the satellite products had for convective events are now hidden by the improved skill elsewhere as a result of assimilating with observations.

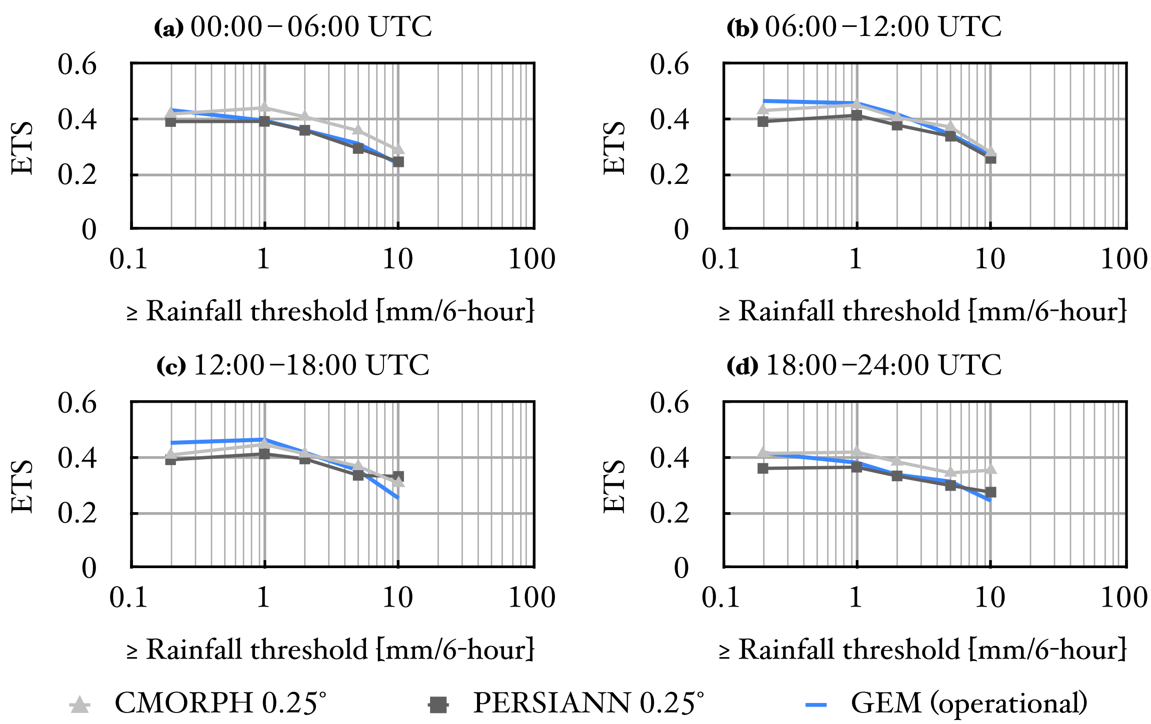


Figure 6.12: Equitable threat scores between assimilated satellite and GEM data for each 6-hour time step from June through October, 2002 to 2011 for manual stations in western Canada

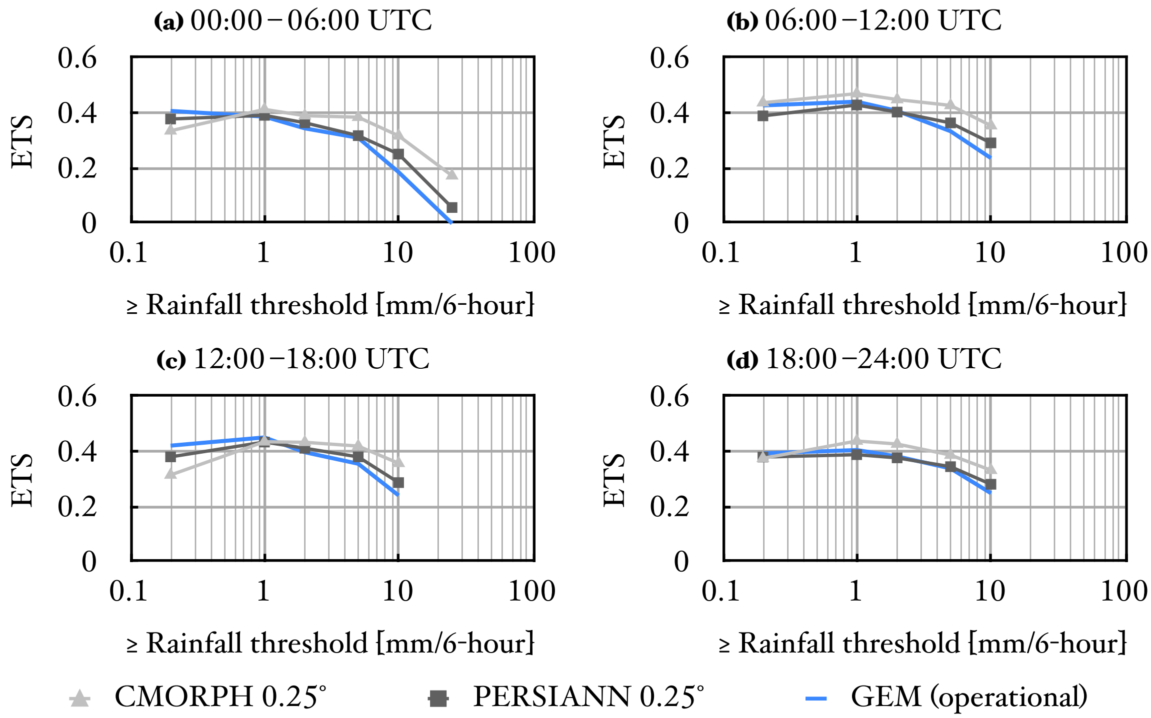


Figure 6.13: Same as Figure 6.12 but for central Canada

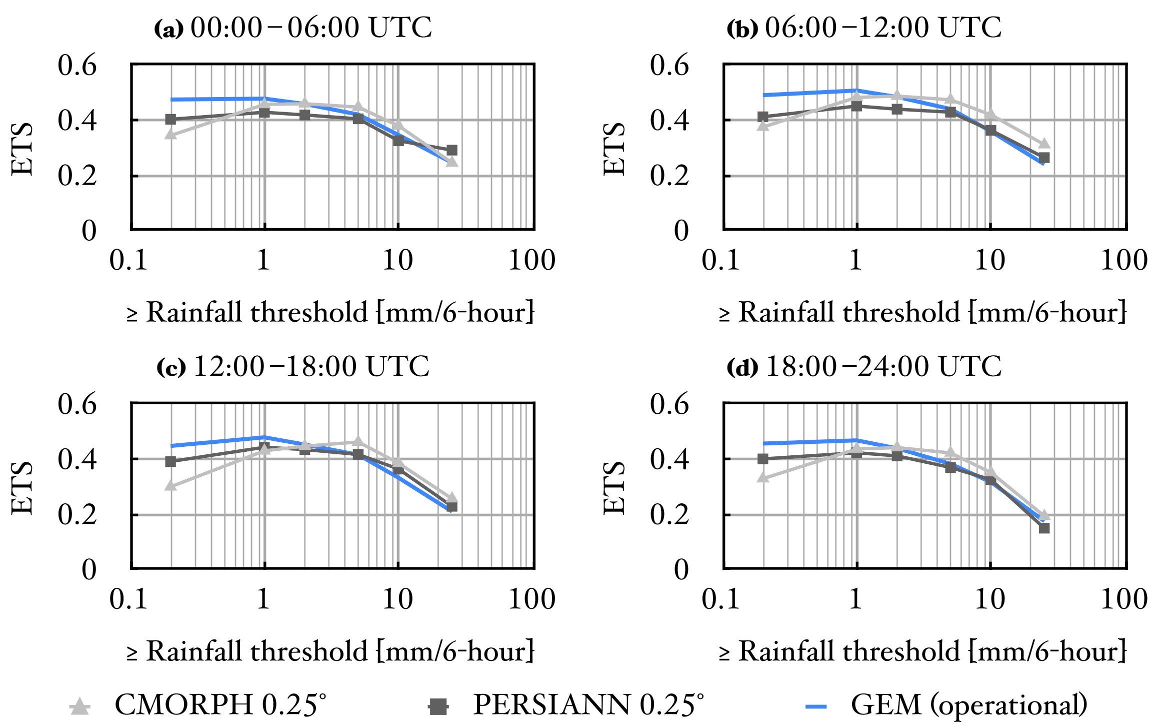


Figure 6.14: Same as Figure 6.12 but for eastern Canada

Relative to GEM, both of the satellite products do show benefits for specific time steps in central Canada. Figures D.7, D.8, and D.9 show the binned differences between GEM and the satellite products for western, central, and eastern Canada, respectively. All of the comparisons between GEM and the satellite products have shown the satellite products to be worse for light rainfall and possibly better for greater magnitudes. When considering the analyses in this great of detail, this property is not always present.

Similar to before, the satellite products possess the greatest skill in eastern Canada, however GEM does as well, so there is minimal advantage to using the satellite products in this region. The same cannot be said for all time steps in central and western Canada. During the night time and morning (06:00 to 18:00 UTC) in western Canada, the satellite products behave similar to eastern Canada and do not perform well, but not to the same degree. Across all rainfall magnitudes during the afternoon and evening, the satellite products are for the most part comparable to GEM, meanwhile CMORPH shows additional skill for greater magnitudes between 18:00 to 24:00 UTC.

The differences in central Canada are the most interesting in that CMORPH performs well during the night time, and both satellite products perform well in the afternoon. During the night time (06:00 to 12:00), for no rainfall less than 25 mm/6-hour is CMORPH significantly worse than GEM, and it is significantly better between 10 and 25 mm/6-hour. In the afternoon (18:00 to 24:00), both satellite products are barely worse below 1 mm/6-hour, above which they are equivalent to or better than GEM. The most appropriate explanation for this unique performance in central Canada is the combined effects of GEM having difficulties reproducing convective events (Lespinas et al., 2014) and the overestimation of the satellite products in terms of both frequency and magnitude (see Subsection 6.6.2 or 4.3.2).

6.7 Key points

- Adjusting the satellite products in accordance with all or manual stations shows no statistically significant differences in the results when compared against manual stations.
- The CMORPH products show an increase in skill for rainfall above 50 mm/6-hour

when adjusting with manual stations.

- Rainfall measurements from manual stations do not possess the extreme events found when considering all stations.
- The large events when considering all stations negatively affect the CMORPH adjustment, however it is not significant (based on 80% confidence intervals).
- Results from assimilation with CaPA are tested against manual stations for consistency with previous literature.
- Since there is no significant change in skill between adjusting with all or just manual stations when compared to manual stations, an adjustment including all available stations will be performed for future flexibility should an evaluation against all stations be warranted.
- Assimilation of the satellite products with CaPA improves their skill.
- Overall, the satellite products only have potential to be preferable to GEM at greater rainfall magnitudes, consistent with the previous findings of low performance for smaller events.
- CMORPH continues to be the best performing of the satellite products and can be significantly better than GEM for the larger rainfall magnitudes.
- While the satellite products had the greatest skill in eastern Canada, relative to GEM the region of greatest improvement is in central Canada.
- The satellite products provide the greatest skill relative to GEM in August across Canada.
- The higher resolution satellite products produce greater frequencies of large events, relative to both of the low resolution equivalents.
- Equitable threat scores for the high resolution satellite products do not benefit from such overestimations.
- For all times steps in eastern Canada, the satellite products have the most difficultly with light rainfall and do not show more skill for greater magnitudes, relative to GEM.
- Western Canada shows no significant difference in the afternoon or evening relative to GEM, except for greater rainfall in the afternoon, where using a CMORPH is beneficial. Satellite products in other time steps decrease the skill.

- Central Canada is the most applicable region for satellite data.
- During the afternoon in central Canada, both CMORPH and PERSIANN are similar to GEM, with CMORPH expressing significant skill for greater rainfall.
- During the night in central Canada, CMORPH is either not significant different than GEM, and significantly better for greater rainfall.

Part 7 | Satellite data as an additional data source

Where Part 6 explored the use of satellite data instead of GEM, this section considers using the satellite data as an additional observational source. That is, the GEM product is assimilated with observations and satellite data based on the idea that the satellite products can provide additional information in the areas between stations. The effectiveness of including satellite data is dependent on its performance relative to that of GEM.

The following sections compare the assimilations including satellite data to the current operational form where GEM is assimilated with station data alone. In contrast to Part 6, the plots in the following sections present the differences in equitable threat scores for readability, as the changes in skill are not as large. The following evaluations focus on the same regions and times as Part 6. The first is an overall evaluation summarizing the performance of all results within the temporal and spatial domain. Following this, the results evaluated for specific locations and times which have been shown to possess unique performance. These include regions of Canada as defined by Figure 4.1, months within the temporal domain, and each 6-hourly time step.

The results from the assimilation are based on the same principles as Part 6 in that they are the results of a leave-one-out cross-validation and have the same bootstrapping methodology applied to account for changes in spatial and temporal densities, as well as 80% confidence intervals. All of the results are subsets of an analysis against manual stations within Canada below 60°N from June through October for 2002 to 2011 (2003 onward for PERSIANN-CCS). The equitable threat scores are plotted as functions of rainfall threshold in the main text, with binned scores in the appendix.

7.1 Overall

To serve as a general evaluation of the analyses assimilating GEM with both station and satellite data, all data within the temporal and spatial domain are used. Subfigure 7.1a shows the skill of the GEM data assimilated with stations using CaPA, and Subfigure 7.1b shows the departures from Subfigure 7.1a as a result of including various satellite products as additional data in the assimilation. In Figure 7.1, the skill is plotted as a function of rainfall threshold.

For rainfall below 0.2 mm/6-hour, the inclusion of satellite products worsen the analysis to a small degree. CMORPH, which has shown to have issues with light rainfall (Figure 4.7; Kidd et al., 2011) causes the greatest decrease in skill. However, for rainfall above 1 mm/6-hour, the performance of the satellite products relative to each other start to differ. All products but PERSIANN-CCS benefit the analysis to some degree. The low resolution CMORPH in particular performs the best, followed by its high resolution counterpart. The inclusion of

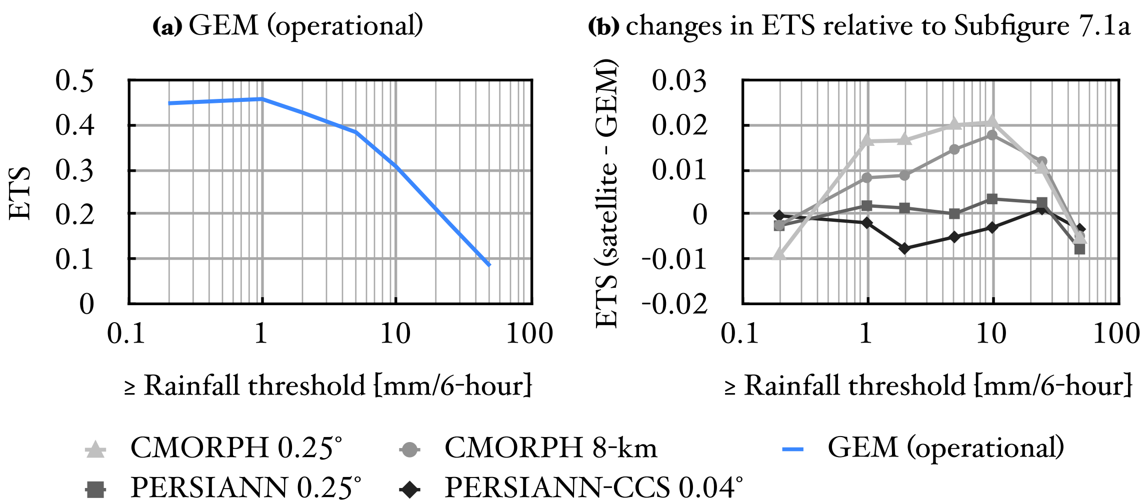


Figure 7.1: Subfigure 7.1a shows the skill of the analysis when no satellite data are used, from which the differences in Subigure 7.1b are calculated. Subfigure 7.1b shows the differences in skill when various satellite products are included in the analysis as additional data sources. Both are based on data from June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

satellite data do not provide benefits for rainfall above 50 mm/6-hour. The following sections consider skill through various subsets of time and space to determine when and where these advantages and disadvantages occur.

Similar to the previous part, the differences are also plotted as functions of binned rainfall in Figure E.1 in Appendix E and include 80% confidence intervals. When the 0.25° CMORPH product is included, the skill of rainfall less than 1 mm/6-hour deteriorates significantly, but with increasing rainfall rates the differences improve to the point where for rainfall between 5 and 50 mm/6-hour the skill improves significantly. The 8-km CMORPH product is similar, but the positive influence is skewed to greater rainfall rates. Meanwhile, the inclusion of the 0.25° PERSIANN product produces slight increases in skill, but they are never significant. For the greater magnitude bins, PERSIANN-CCS is never able to improve the skill of the analysis above its operational form.

The changes in skill from including satellite data in the assimilation are consistent with the general performance of the raw data, in that the low resolution CMORPH product is most beneficial and PERSIANN-CCS provides no improvement. Also, the greatest amount of skill is added to the analysis for greater magnitudes of rainfall, as that is when they perform the best when compared with observations. For very large rainfall (greater than 50 mm/6-hour), the satellite products worsen the analysis, but such conclusions contain a great amount of uncertainty. That said, the improvements of including any of these satellite products are small, roughly 4% at their peak. This is directly related to the skill of the satellite products relative to that of GEM assimilated with stations.

7.2 Regionally

In taking the same data from Figure 7.1a and separating them into the previously defined regions of Figure 4.1, spatial differences in skill scores for analyses including satellite data are plotted and presented in Figure 7.2.

In Section 7.1, the differences in skill scores from including the various satellite products produced some trends, such as reducing the skill for lower and very large rainfall. The regional scores of Figure 7.2 show that these overall trends are mostly dictated by central and eastern Canada. Inclusion of the 0.25° PERSIANN product is different in western than other regions of Canada as it increases skill for rainfall above 10 mm/6-hour, otherwise there is no change or a decrease in skill. In the same region, the inclusion of either CMORPH product increases the skill relatively constantly for increasing rainfall thresholds.

In central and eastern Canada, the differences in equitable threat scores from CMORPH are similar to the overall evaluation and each other in that there is a peak in performance for rainfall greater than 10 mm/6-hour. In the overall evaluation, the inclusion of low resolution CMORPH data decreases the skill for rainfall greater than 0.2 mm/6-hour. In Figure 7.2, it can be seen that this decreased skill is present only in eastern Canada.

The binned data in Figure E.2 show similar trends. In western Canada, the cause of the constant improvement in skill from CMORPH is apparent, as skill is added for rainfall between 0 and 1 mm/6-hour. For these rates in other regions, the analysis is either neutral or worse than the operational form. When including the low resolution CMORPH product, the analysis improves between 0 and 50 mm/6-hour, and significantly so for multiple bins. With the exception of rainfall between 1 to 2 mm/6-hour, inclusion of the high resolution CMORPH product also improves the analysis, although only significantly so for smaller events (less than 1 mm/6-hour). For rainfall magnitudes reported in Figure E.2, using PERSIANN has neutral effects on equitable threat scores. Using PERSIANN-CCS does not change light rainfall, but significant decreases for rainfall between 5 and 25 mm/6-hour (that between 25 and 50 mm/6-hour is not plotted).

For rainfall less than 5 mm/6-hour, the performance in central Canada is a transition between eastern and western Canada. Where in western Canada the inclusion of satellite data are beneficial, in eastern Canada they are not. As a result, in central Canada it is relatively neutral, especially when considering 80% confidence intervals. For the analysis with the low resolution PERSIANN product, no meaningful improvements are present. Both CMORPH resolutions

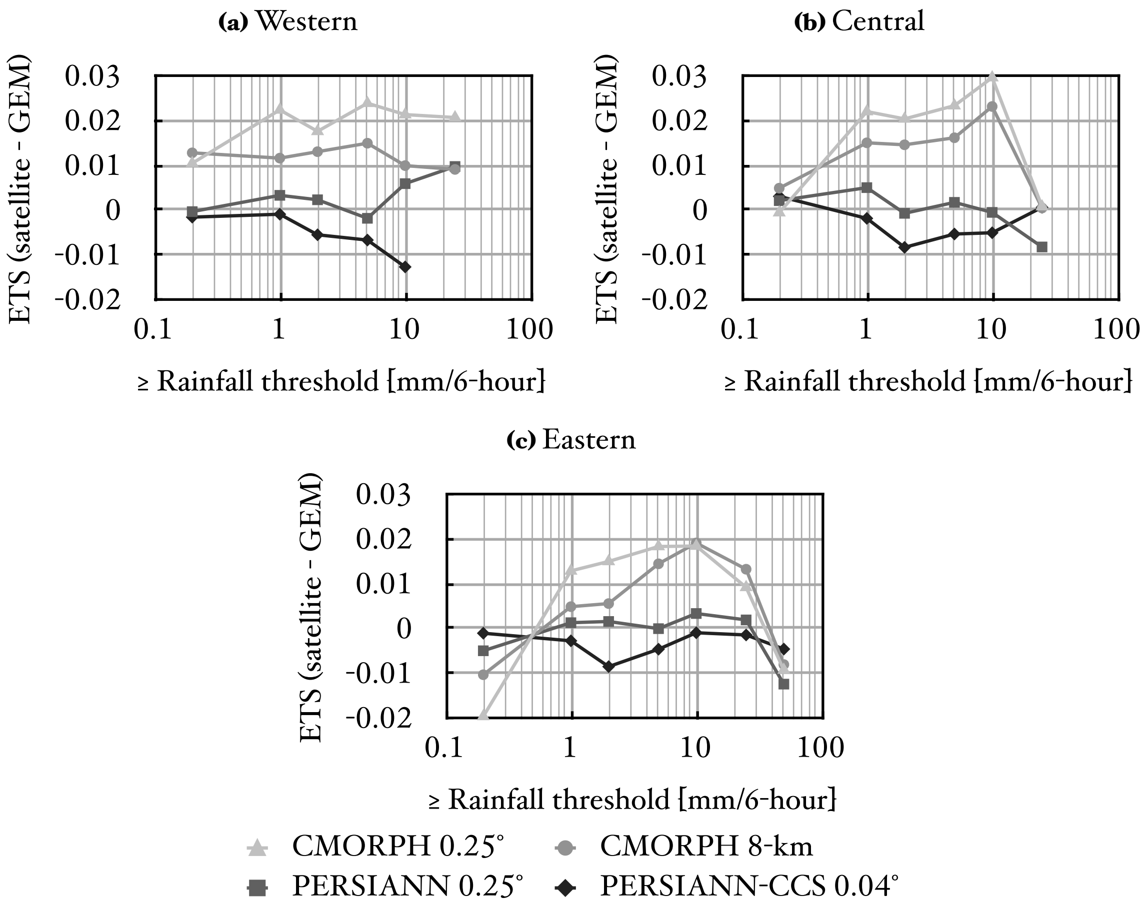


Figure 7.2: Regional differences in equitable threat scores when various satellite products are included in the assimilation using CaPA. Scores are from data of June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

add skill to their analyses to similar degrees for rainfall below 25 mm/6-hour, in that they are fairly neutral in their impact, but significantly positive between 10 and 25 mm/6-hour. The differing improvements previously in Figure 7.2 for these two is based on rainfall above this, where the 0.25° is neutral in its effect, and the 8-km product reduces skill.

The magnitudes of change in equitable threat scores from including satellite data in eastern Canada are uniquely statistically unfavourable for rainfall less than 2 mm/6-hour. So much in fact, that it influences the overall evaluation of Figure E.1. Interestingly, for rainfall between 5 and 50 mm/6-hour the analysis is improved, even more than in central Canada, however a reduction in skill for rainfall above 50 mm/6-hour masks this in Figure 7.2. While general trends in performance are similar between CMORPH and PERSIANN, they are exaggerated (much worse or much better) when using either resolution of CMORPH.

7.3 Monthly

As the monthly skill scores of the raw satellite data in Figure 4.8 have shown, the performance of the satellite data have some temporal dependence. As such, changes in equitable threat scores from including satellite data in the assimilation of GEM and observations are evaluated for every month within the temporal domain. Figure 7.3 shows this for manual stations across Canada for a given month. The scores plotted here, again, are the result of a leave-one-out cross-validation and the previously detailed bootstrapping methodology. Not surprisingly, during months where the raw satellite data perform the best, July and August, the assimilation benefits the most. Also consistent with the raw data, the low resolution CMORPH product changes the most between months.

Comparing the monthly in Figure 7.3 with the overall scores in Figure 7.1, one can see that the overall scores are mostly representative of those in July and August. While the increase in equitable threat score during July is relatively high (greater than 0.25), the overall score is less because of the poor performance during September and October. June has a unique trend in that for rainfall greater than 25 mm/6-hour, the satellite data better the analysis, however for July and August, there is a decreasing trend for these events.

Figure 7.4 shows the frequency bias indices of the assimilation of GEM and station data, as well as those when the 0.25° CMORPH product is included. Regarding June, the increase in equitable threat scores for rainfall greater than 25 mm/6-hour is accompanied by an increase in event frequency relative to the assimilation without satellite data. During July, the assimilation with satellite data results in a greater underestimation of event frequency, however larger events in the following months produce increasingly similar frequencies. It is interesting that the inclusion of the 0.25° CMORPH product in July and August further underestimates the frequency of events while increasing the equitable threat scores.

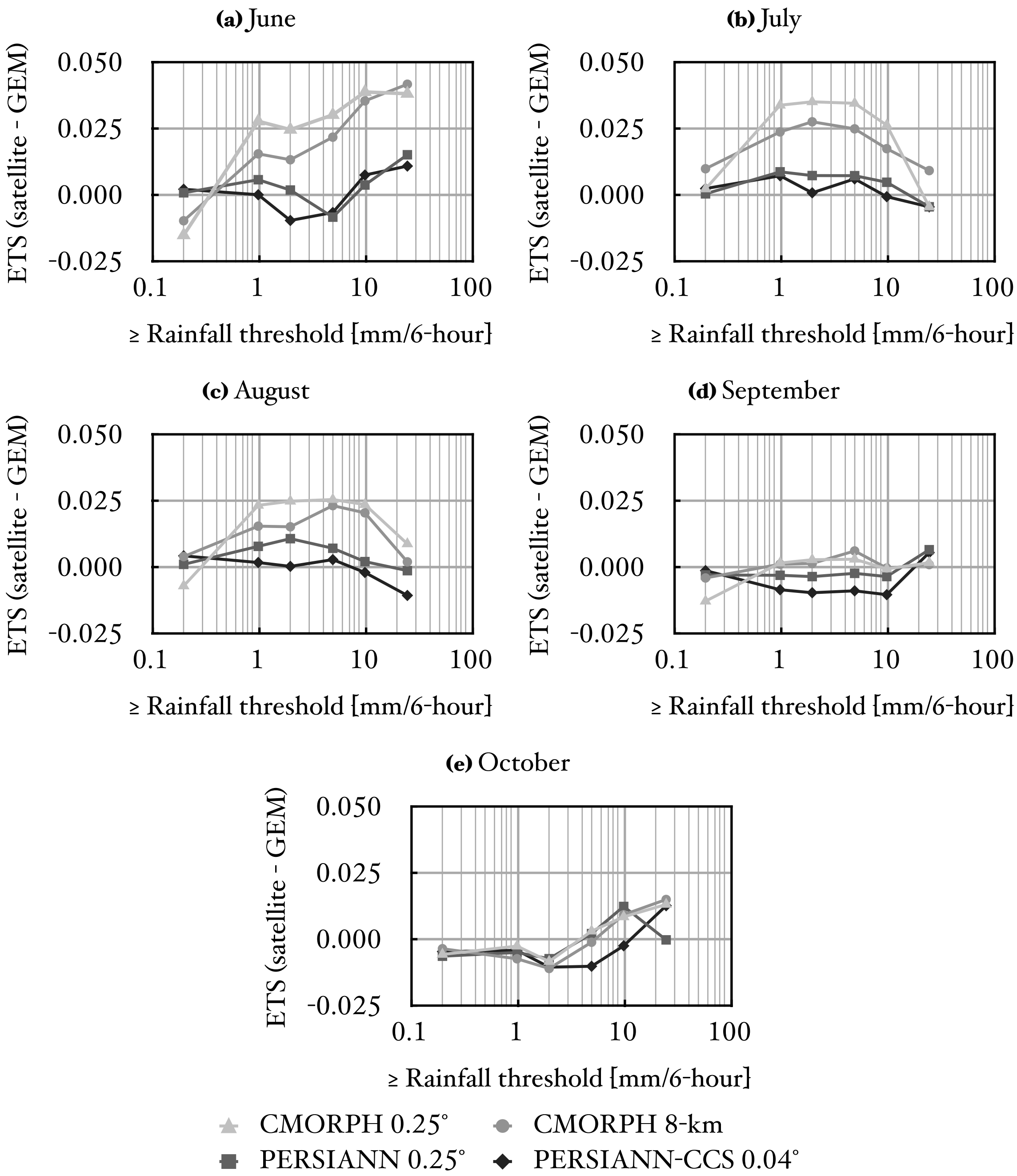


Figure 7.3: Monthly differences in equitable threat scores between analyses with and without the inclusion of satellite data in the assimilation of GEM data for June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

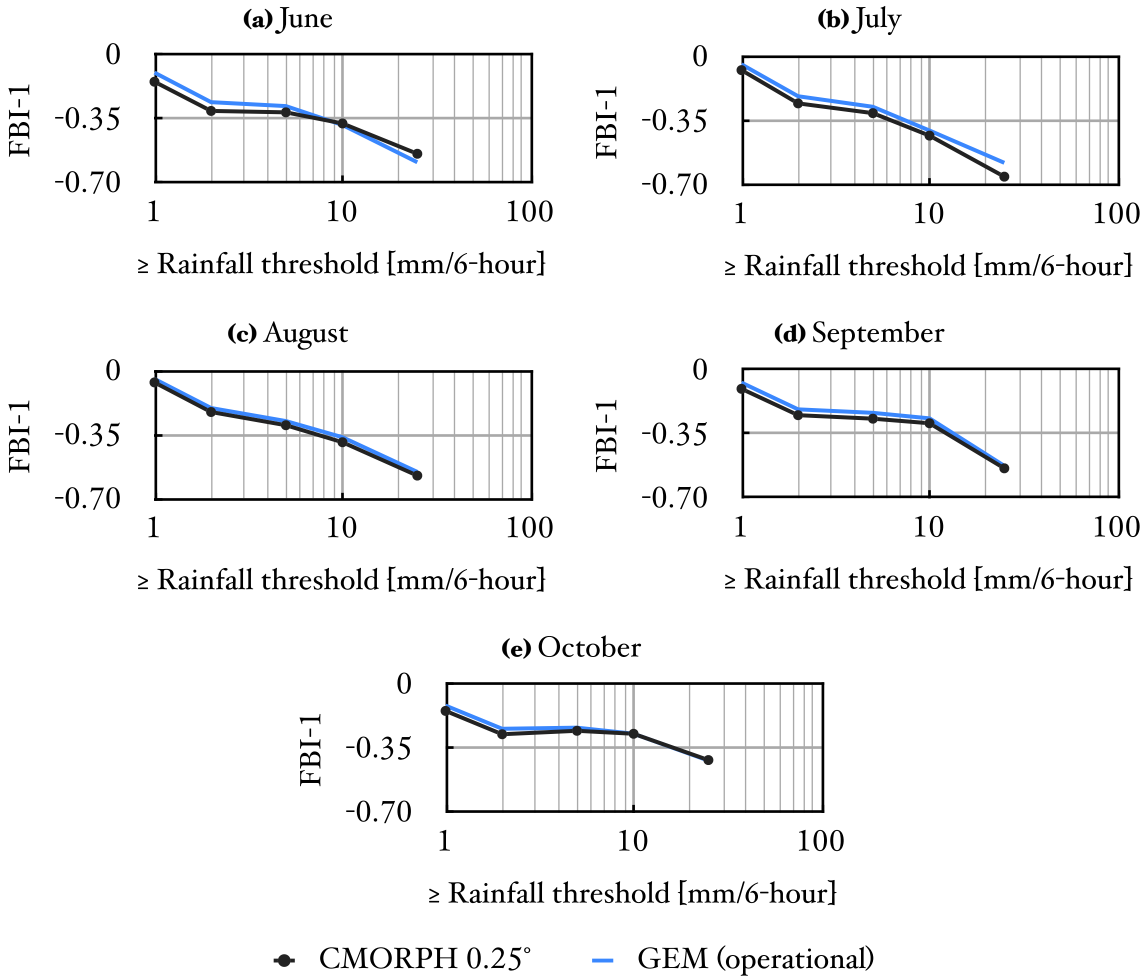


Figure 7.4: Monthly frequency bias indices for GEM assimilated with observations and GEM assimilated with observations and 0.25° CMORPH data from 2002 to 2011.

Regarding the increased equitable threat scores in June, the large increase is accompanied with some of the largest uncertainty in any of the analyses. Also, such improvements are not seen when the same satellite data are used as a background field. It is most likely a limited number of stations which benefited particularly well in this analysis.

In the overall evaluation of Section 7.1 a concern was that for rainfall greater than 0.2 mm/6-hour, the inclusion of satellite data decreased the skill of the analysis. Then, considering the data spatially in Section 7.2 revealed that this is only the case for eastern Canada (excluding PERSIANN-CCS). In Figure 7.3 the PERSIANN products have a neutral effect for rainfall greater than 0.2 mm/6-hour, except for October where they are worse. For the same magnitudes, the CMORPH products are worse for all months except July, and for the 8-km product during August. Then so far, one can say that the products appear to be providing the most benefit during June through August in central and western Canada. July performs very well in both the magnitude of improvement, but also the range of rainfall being improved.

The binned equitable threat scores in Figure E.3 show that while inclusion of satellite data during June adds large amounts of skill for rainfall above 10 mm/6-hour, they also cause large amounts of uncertainty. The influence of adding CMORPH results in significant decreases in skill for rainfall between 0 and 1 mm/6-hour and significant increases above 10 mm/6-hour. The PERSIANN products never significantly increase in skill during June.

As previously discussed, the satellite products perform well in July, and in this application never significantly worsen the analysis. While both PERSIANN products do not produce significant increases in skill, both CMORPH resolutions do for most bins. Notable differences during July relative to other months include the positive performance for rainfall less than 1 mm/6-hour, which has generally been difficult for the satellite products. Also, inclusion of the 8-km CMORPH product performs well compared to that from the 0.25° product, especially for the lowest and greatest rainfalls in the plot. Previously, the performance of the high resolution product has not been as great as that from the lower resolution.

August performs similar to July, but the improvements are not as great. That said, those from the CMORPH products are still statistically significant. The remaining months rarely show any significant improvement. The only increases in equitable threat scores during October are for rainfall greater than 10 mm/6-hour, but even then it is only barely significant for the lower resolution between 10 and 25 mm/6-hour.

While the improvements in this part are small, by considering more specific regions in the last two sections, the improvements have increased for some areas and times. If the scores are consistent in providing benefits across the rainfall magnitude spectrum, an applicable location and time to use satellite data can be proposed, even if it is a subset of the project's domain.

7.4 6-hourly

As with the previous part, the analyses are also evaluated 6-hour time steps in an attempt to identify any diurnal trends. This section contains two evaluations, one for stations across Canada, and the other for specific regions of Canada. The results from this section serve to further define an applicable location and time for the assimilation of GEM and stations to also include satellite data.

7.4.1 Across Canada

The evaluation in this section calculates scores from manual stations across Canada from June through October for 2002 to 2011 (2003 onward for PERSIANN-CCS), separated only by time step. The equitable threat scores for these time steps are shown in Figure 7.5, which are organised in a similar manner to the previous part. The first row is, for the most part, during the night, and the second row is during the daytime.

For the different 6-hour time steps, the PERSIANN data are not very interesting — the 0.25° product adds minimal skill, if any at all. PERSIANN-CCS only adds skill for rainfall greater than 25 mm/6-hour during the daytime. Both CMORPH products worsen the analysis for rainfall less than 0.2 mm/6-hour, but improve it for rainfall greater than 1 mm/6-hour, reaffirming the previously shown difficulties with light rainfall. As observed previously, the 8-km CMORPH product is producing comparable improvements to the 0.25° product. In Figure7.5, for three of the four time steps it is adding more skill to the analysis for rainfall greater than 25 mm/6-hour. During the afternoon and evening (18:00 to 06:00 UTC), both resolutions also reach their peak contribution to skill at lower thresholds (1 mm/6-hour), as opposed to 5 mm/6-hour for the night and early morning. From these plots, the magnitude from which the increase in skill is derived cannot be determined, and will need to be considered in evaluations versus binned rainfall. It is worth noting that between 18:00 and 06:00 UTC the satellite products add the most skill across the greatest spectrum of magnitudes (greater scores at lower thresholds), and in the evening (00:00 to 06:00 UTC) produce the greatest increase overall.

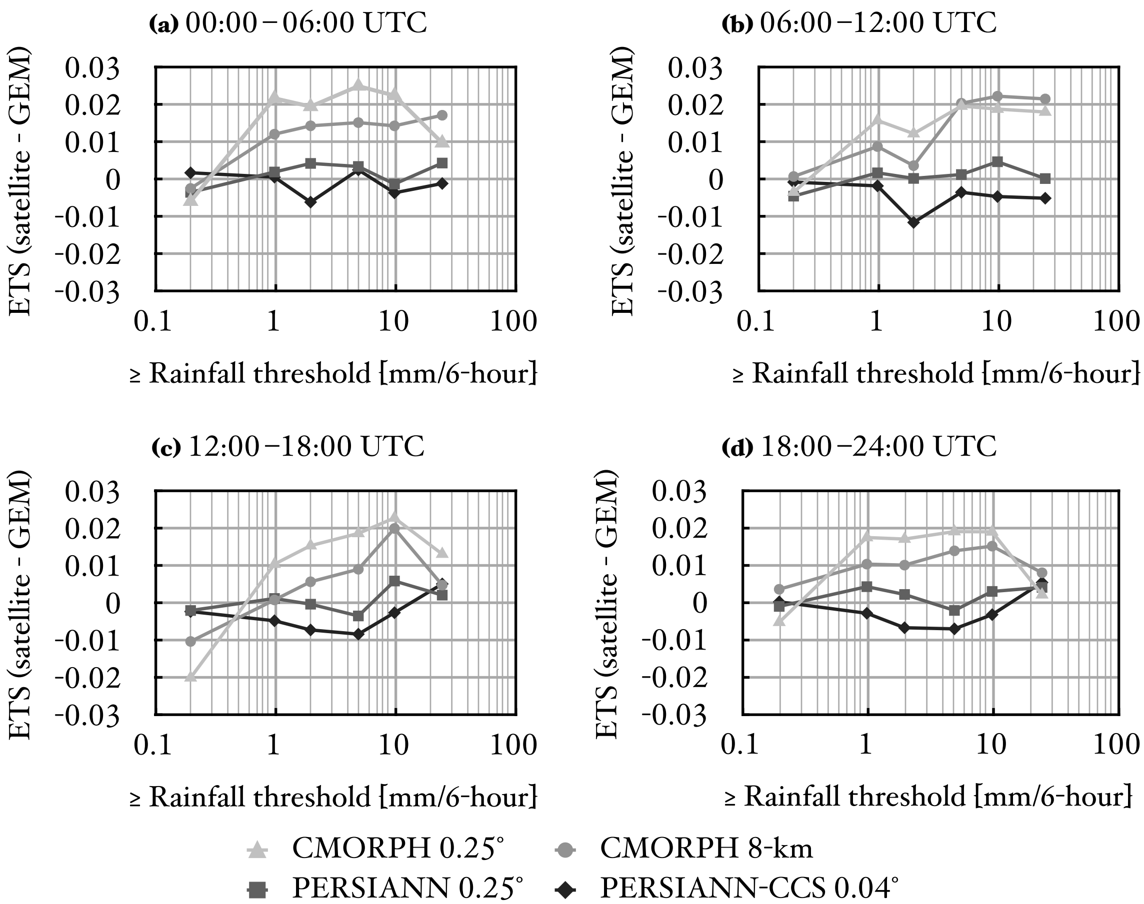


Figure 7.5: Equitable threat scores for assimilated satellite products and GEM relative to Canadian manual stations for each 6-hour time step over the temporal domain of June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

The same data are plotted in Figure E.4 of the appendix versus binned rainfall values. The changes in skill from the assimilation of PERSIANN data for different time steps across Canada as a function of binned data are the same as those as a function of threshold rainfall, in that they are minimal. In this context, the CMORPH data provide greater changes in skill. Interestingly, Figure E.4 shows that the skill added in the afternoon and evening relative to the other times is not from the large events, but rather from smaller events. During these times, rainfall between 0 and 5 mm/6-hour provides no change in skill, which then amplifies the significant increase in skill for rainfall between 10 and 50 mm/6-hour (between 5 and 50 mm/6-hour for 18:00 to 24:00 UTC). This is in contrast to 06:00 to 12:00 UTC, and to a lesser degree 12:00 to 18:00 UTC, where rainfall less than 5 mm/6-hour worsens the analysis, even though skill between 25 and 50 mm/6-hour demonstrates the greatest amount of improvement.

7.4.2 Regionally

In the greatest detail for this part, results from the analyses of GEM assimilated with station and satellite data have been separated by 6-hourly time steps and regions of Canada. The changes in equitable threat scores relative to the analysis without satellite data are shown in Figures 7.6, 7.7, and 7.8 for western, central, and eastern Canada respectively, as a function of rainfall threshold. These plots contain data from June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

In General, the PERSIANN products do not add as much skill to the analyses as either CMORPH resolution, but in eastern Canada or during the morning in central Canada they do not worsen the analysis for light precipitation to the same degree as either CMORPH. The changes in skill for the low resolution PERSIANN product usually remains close to zero, with a few notable changes consistent across multiple rainfall thresholds. During the evening in central Canada, PERSIANN adds skill for rainfall above 0.2 mm/6-hour, and more so with increasing thresholds. The same is true during the afternoon in eastern Canada.

By including CMORPH in western Canada, rainfall above 0.2mm /6-hour benefits during all time steps, with the greatest improvements in the evening, and to a lesser degree the afternoon. The largest improvements to the skill of the analyses are for rainfall greater than 10 mm/6-hour during the morning and afternoon. Larger rainfall during the evening and night do not benefit from including either CMORPH product. In central Canada during the morning, the analyses is made worse for rainfall greater than 0.2 mm/6-hour, but benefits for any threshold greater than that. The evening is the only other time step which does not benefit, albeit to a small degree. Meanwhile, this time step also sees the largest increase in skill for greater thresholds of any location and time. The trends in eastern Canada do not change dramatically with time. Equitable threat scores for rainfall less than 0.2 mm/6-hour are lower as a result of including CMORPH, however above 1 mm/6-hour the analysis benefits. The only notable change amongst the time steps is the additional skill added for rainfall during the night above 25 mm/6-hour, relative to other times.

Differences in binned equitable threat scores for the same data are plotted in Figures E.8, E.9, and, E.10 in the appendix for western, central, and eastern Canada respectively. Including the 0.25° PERSIANN product for western Canada shows a slight increase for lower magnitudes (less than 5 mm/6-hour) in the evening and afternoon. All other time steps are either not significant or worse as a result. Central Canada shows no notable improvements. In eastern Canada, including PERSIANN results in a similar effect as CMORPH, where for rainfall less than 2 mm/6-hour, the analysis worsens. Although it is quite small, it may be worth noting a significant improvement in the afternoon between 5 and 50 mm/6-hour, only because it exists over multiple rainfall bins.

The binned differences for both CMORPH resolutions are much greater than those of PERSIANN. Usually, the 0.25° product is capable of adding more skill than the 8-km product. In western Canada, the 0.25° CMORPH product, and sometimes the 8-km product, show significant improvements for lighter rainfall (less than 5 mm/6-hour) in the evening. Similarly, there are significant improvements above 5 mm/6-hour in the morning and afternoon. Additions to skill are found in central Canada during the evening between 10 and 25 mm/6-hour, and to a lesser degree between 2 and 5 mm/6-hour. Some of the lightest rainfall (less than 0.2 mm/6-hour) benefits in the afternoon and late at night. Eastern Canada sees improvements during the greatest events (25 to 50 mm/6-hour) during the night and in the morning. For rainfall between 10 to 25 mm/6-hour, significant improvements are during the night, morning, and afternoon. As summarized in Figure 7.2c for eastern Canada and shown in more detail here, all time steps experience a decrease in skill for rainfall less than 1 mm/6-hour from assimilating GEM with stations and satellite data using CaPA.

Figures E.8, E.9, and E.10 in the appendix show 6-hourly frequency bias indices for western, eastern, and central Canada respectively. While some locations, times, and bins experience changes in frequency bias from including satellite data, the majority have minor changes. There does not appear to be a direct correlation between the changes in equitable threat scores and frequency of events, even though this was found previously to be the case in June. For example, the inclusion of CMORPH data in eastern Canada results in a decrease of skill for rainfall less than 1 mm/6-hour, which coincides with an increase in frequency bias between 0.2 and 1 mm/6-hour. However this is not always the case as the frequency bias in central Canada also sees an increase in frequency between 0.2 and 1 mm/6-hour, but only experiences a decrease in skill for half of the time steps. Also, western Canada benefits most for rainfall between 10 and 25 mm/6-hour in the morning and afternoon, but the frequency of events between all time steps show no change. The effects of frequency bias on equitable threat scores may just be a monthly occurrence, which cannot be seen when all months are considered to-

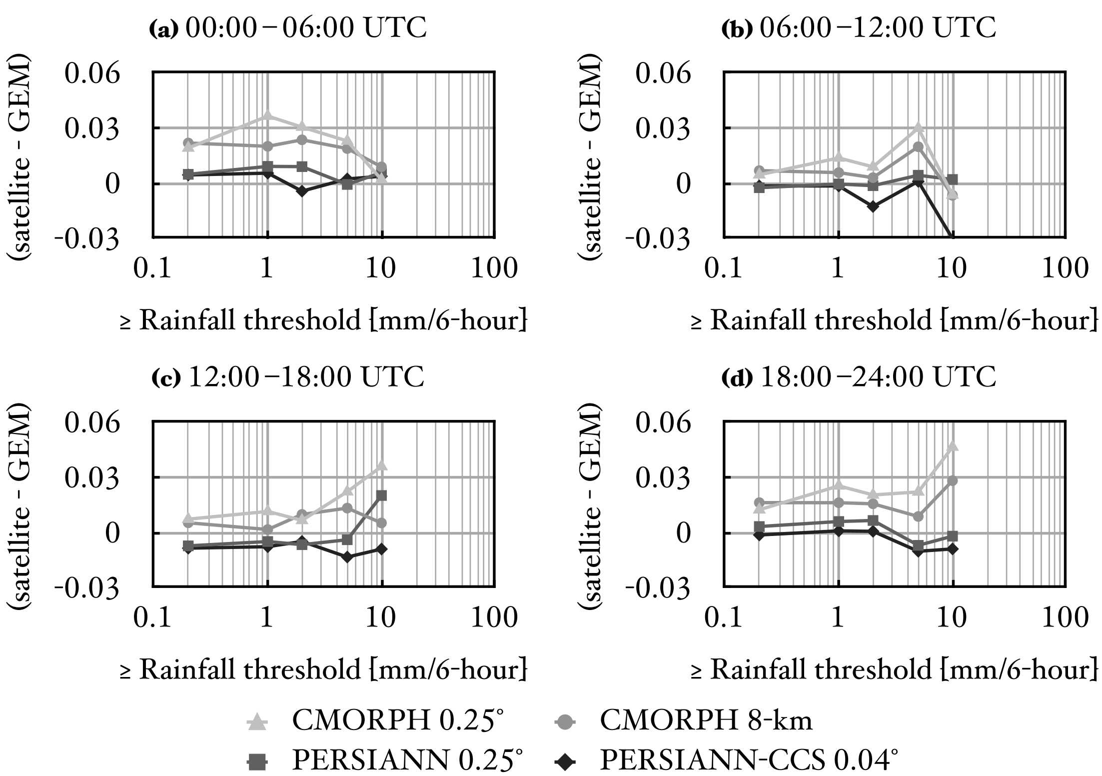


Figure 7.6: Differences in equitable threat scores for western Canada between satellite products and GEM after assimilation via CaPA. Scores are separated by time step and include data for June to October, 2002 to 2011 in western Canada.

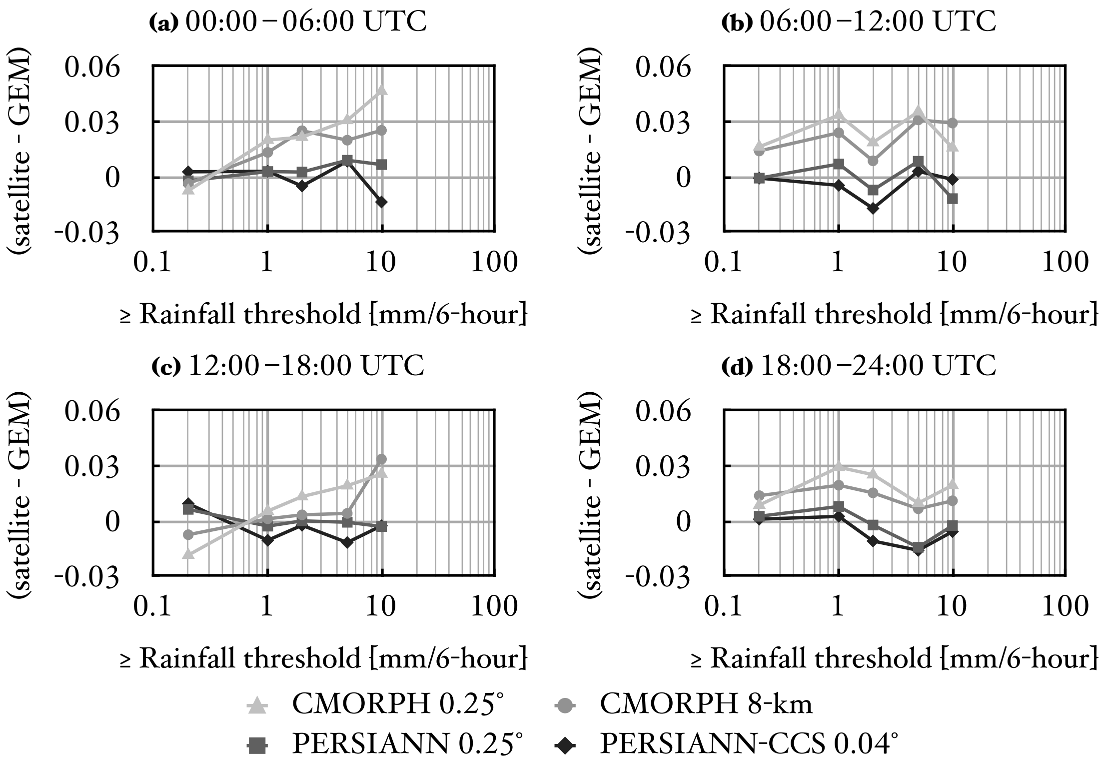


Figure 7.7: Same as Figure 7.6, but for central Canada.

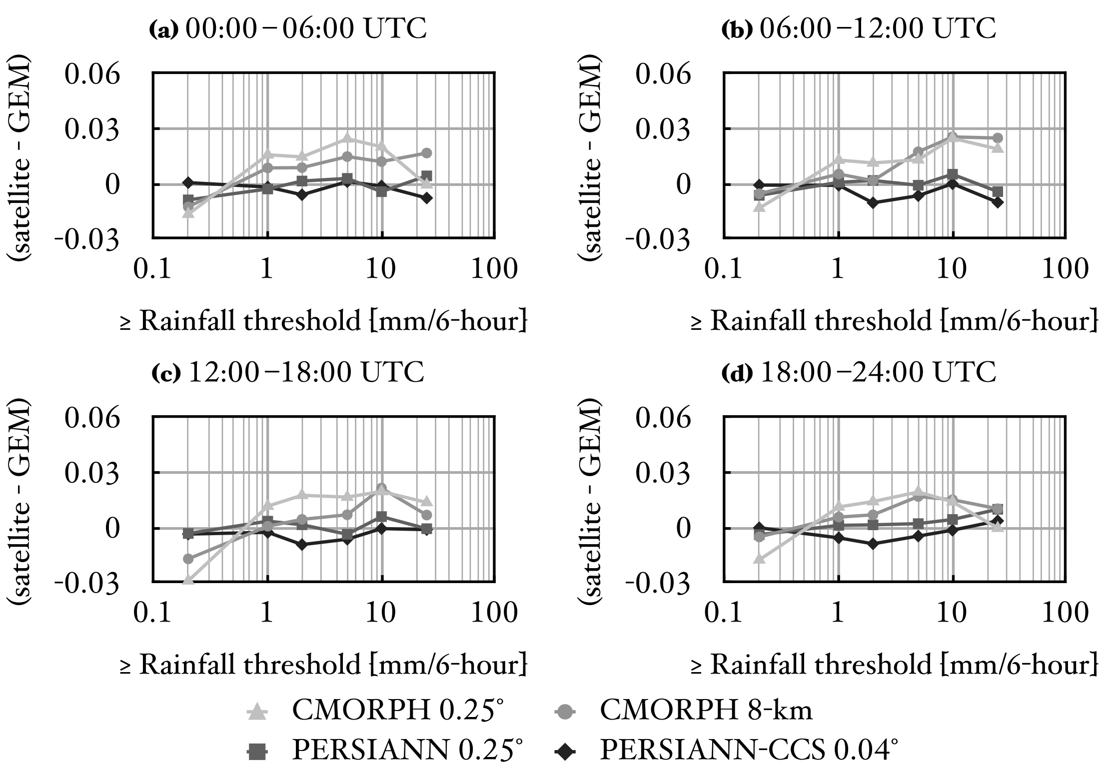


Figure 7.8: Same as Figure 7.6, but for eastern Canada.

gether. The error decomposition of Section 4.3 showed the relations between occurrences of errors between months. That said, the assimilations with satellite data are changing the nature of the analyses. Frequency bias in western Canada remains mostly unchanged, however the frequency of light events (0.2 to 1 mm/6-hour) in central and eastern Canada increase. This increase is only for CMORPH products in central Canada and all PERSIANN and CMORPH products in eastern Canada, with CMORPH having a greater effect. With regards to these changes, the smallest changes in frequency are during the night in central and eastern Canada.

7.5 Satellite data as a background field versus as an additional data source

In Parts 6 and 7 satellite data are used with the CaPA assimilation methodology and compared to the currently operational form which uses GEM. In Part 6 the satellite data took place of GEM, and in Part 7 satellite data were used to adjust GEM along with station data. The two different approaches provided locations and times where the use of satellite data are both advantageous and disadvantageous. While some trends in the changes of skill from using satellite data are similar between the two approaches, such as that at lower or greater magnitudes of rainfall, the magnitudes of the changes are quite different.

When satellite data are used as an additional data source, the magnitudes of changes are not as great as when they were used as a background field, simply because replacing the background field is a more drastic change regarding data and methodology. While this means that the improvements are not as prevalent, the decreases in skill are not as great either. So in eastern Canada where replacing the background field greatly reduced the skill for light rainfall, it did not reduce to the same degree when the satellite data were used as an additional data source. For rainfall above 5 or 10 mm/6-hour, using CMORPH as an additional data source produces more significant increases in skill than using them as a background field.

Central Canada, shown in Figures 6.13 and 7.7, experiences similar effects for time steps where the skill decreased at lower magnitudes (less than 5 mm/6-hour). Using satellite data in this region as another source instead of a background reduces the significant increases for large events (between 10 and 25 mm/6-hour) to where they only exist in the evening.

Between the two approaches, western Canada appears to benefit the most from using either satellite product as an additional data source instead of as a background field, based on Figures 6.13 and 7.7. The latter produced significantly worse scores during the night and morning relative to GEM, but with the former, skills are generally either the same or better relative to GEM.

7.6 Key points

- Inclusion of the 0.25° CMORPH product shows the greatest potential for improving the analysis.
- Improvements from using the satellite products are small.
- CMORPH worsens the skill for light rainfall, but significantly increases it for greater rainfall rates across Canada.
- PERSIANN does not significantly increase the skill across Canada.
- Inclusion of CMORPH increases skill for all magnitudes in western Canada, and PERSIANN only does for greater rainfall rates.
- The reduction in skill for light rainfall is mostly because of eastern Canada, where skill is significantly reduced for light rainfall and increased for large rainfall.
- CMORPH products cause a significant increase in skill in central Canada between 10 and 25 mm/6-hour.
- Western Canada benefits the most from including CMORPH for the greatest range of rainfall rates.
- Inclusion of satellite data add the most skill across a range of rainfall rates in July and August, when the satellite products perform the best relative to observed data.
- September and October provide minimal significant improvements.
- Inclusion of PERSIANN consistently increases the skill of analyses for thresholds greater than 1 mm/6-hour during the evening of central Canada and afternoon of eastern Canada.

- Inclusion of CMORPH in western Canada improves rainfall above 0.2 mm/6-hour for all time steps, especially the evening.
- Large events in western Canada benefit during the morning and afternoon from including CMORPH
- The most skill is added in central Canada during the evening for rainfall greater than 10 mm/6-hour.
- CMORPH adds skill in eastern Canada for large rainfall during the night an morning, although all time steps benefit for slightly lower rainfall rates.
- Although using satellite data as an additional data source provide only small changes in analysis skill, they provides more reliable improvements, in that the decreases in skill are not as prevalent.

Part 8 | Concluding remarks

With the methodology originally presented by Mahfouf et al. (2007), and recently updated by Lespinas et al. (2014), CaPA produces a gridded precipitation estimate for Canada in real-time by assimilating GEM with observation stations via statistical interpolation. While the details of the statistical interpolation have changed over time, so have the considered datasets, as the experimental version (soon to be operational) assimilates the previously mentioned datasets with radar (Fortin et al., 2014).

This project focussed around the inclusion of satellite based rainfall estimates within the pre-existing CaPA methodology to determine when and where changes in the analysis skill occur. The motivation for this idea is based off of works such as those by Ebert et al. (2007), who found that under certain conditions, satellite based rainfall estimates can be preferable to numerical weather products. This inclusion of satellite based estimates is achieved through the application of the works of Joyce et al. (2004), Hsu et al. (1997), and Hong et al. (2004) who make available CMORPH, PERSIANN, and PERSIANN-CCS, respectively. These three satellite products make estimates available as far as 60°N. With CaPA's methodology, the satellite based estimates can be assimilated in two different ways, which defined the second and third objectives for this project as presented in Section 1.1. The first objective evaluates the performance of these products relative to the APC2 observations which have been quality controlled and adjusted by Mekis and Vincent (2011).

Prior to any evaluations, estimates over snow covered ground were removed based on the statements made by Ebert et al. (2007), Ferraro et al. (2000), Joyce et al. (2004), and Munchak and Skofronick-Jackson (2013), noting these conditions as a limitation of satellite based rainfall estimates. The removal procedure implemented here cross-referenced the satellite products with the IMS dataset made available by the National Ice Center (2008). This removal also aided in defining the applicable range for satellite based rainfall estimates over Canada to be June through October.

8.1 Objectives

8.1.1 Quantify the error associated with the selected satellite products

The evaluations of the satellite products relative to APC2 gave results similar to those observed in the United States, Europe, and Australia by other authors. For example, these products exhibited a general underestimation of rainfall magnitude which was also observed in the United States by Sapiano and Arkin (2009) and Tian et al. (2007). Of the satellite products, CMORPH performed the best in Canada, which was also shown to be the case elsewhere in the world by Ebert et al. (2007) and Kidd et al. (2011). Nonetheless, CMORPH possessed difficulties estimating light rainfall—a characteristic Kidd et al. (2011) observed in Europe.

As Tian et al. (2007) observed spatial trends in performance of these products over the United States, a similar evaluation was warranted for Canada. It was found that the satellite products produce some of the greatest skill scores in eastern Canada, followed by central Canada, and finally western Canada. This trend is consistent with the finding of Tian et al. (2007), who along with Ferraro et al. (2000) attribute the decrease in performance to the presence of stratiform events, whose rainfall, unlike that of convective events, cannot be as easily detected by the algorithms relating the retrievals to rainfall rates. While the frequencies of missed events in western Canada are not less than other regions, the errors resulting from such misses are greater.

Central Canada was shown to possess unique characteristics. During July and August, the satellite products overestimated rainfall in a manner that is spatially continuous with the overestimations found in the United States by Ebert et al. (2007) and Tian et al. (2007). Compared to other regions during June through October, the frequency of hits and false alarms are similar but the magnitudes are differ — suggesting the bias increase is more closely related to a change in estimated magnitude than frequency. The bias present in the satellite data warranted some adjustment prior to its assimilation. Overall skill scores increased as a result of the adjustment, with central Canada benefiting the least as it was found to be mostly comprised of random error.

8.1.2 Evaluate use of satellite data as a background field

In comparing the assimilation of satellite and station data using CaPA to the operational form with the same skill scores used by Lespinas et al. (2014), it is found that the satellite products cannot reproduce light rainfall with the same amount of skill as GEM. For greater magnitudes, however, the satellite products can provide benefits.

Relative to the assimilation with GEM, eastern Canada is the most affected by poor estimates of light rainfall. Meanwhile, western Canada benefits overall in the afternoon. Central Canada is the most applicable region for satellite data, as using CMORPH provides significant improvements for greater rainfall and no decreases for light rainfall in the early morning and afternoon. That said, other time steps are plagued with poor performance of light rainfall.

8.1.3 Evaluate the use of satellite data as an additional observation source

Relative to GEM, the satellite products conditionally estimate rainfall with greater skill, but even then the differences are not large. As such, the inclusion of satellite data as another data source does not improve the analysis' skill greatly. Even though it is of small magnitudes, some statistically significant increases in skill are present for western and central Canada during July and August. Specifically, western Canada benefits in all time steps but early morning, and central Canada benefits only in the afternoon. The skill in eastern Canada improves for greater rainfall rates, but it is greatly decreased for light rainfall.

8.2 Recommendations

Of the satellite products tested here, CMORPH is unanimously better than PERSIANN. Of the CMORPH products, the lower 0.25° resolution produces greater equitable threat scores, although other categorical scores differ.

The application of satellite data must undergo some quality control measures to ensure realistic measurements. Also, its application should be restricted to times where the ground is not covered by snow. Finally, some bias removal or adjustment should be performed, keeping in mind the strong changes in bias in central Canada during the summer.

The most applicable regions and times are central and western Canada during July and August, of which certain time steps benefit more than others (for example, the afternoon in central Canada). Of the assimilation methods, that where satellite data are used as an additional source is preferable as the satellite data are not required to be spatially complete, assuming the goal is a single Canada-wide analysis. The availability of satellite data only below 60°N cannot be used to create an estimate over all of Canada alone. If one is concerned about a specific region, then using satellite data as a background field can also be considered.

8.3 Future developments of satellite products

Two concerns with the application of satellite data are the spatial domain and the skill of the estimates, both of which are continually being addressed. While the evaluations of the satellite products presented here show how the satellite products are on the verge of being able to improve CaPA analyses, that may change in the near future. In early 2014 the Global Precipitation Measurement (GPM) mission launched, putting a new satellite in low earth orbit to replace TRMM, which houses the GPM Microwave Imager (GMI) capable of covering more frequencies than its predecessor (Munchak and Skofronick-Jackson, 2013). This satellite orbits 60°N to 60°S, compared to TRMM's 30°N to 30°S, from which it is reasonable to expect that the higher latitudes will benefit especially from coverage of the core observatory.

CMORPH will be undergoing changes as it enters its second generation. Using a Kalman Filter based CMORPH algorithm in combination with additional information from the Climate Forecast System Reanalysis (CFSR) to create the motion vectors, a global product (pole to pole) is produced. Preliminary tests have shown improved estimates with heavy and orographic precipitation (Xie and Joyce, 2014). The authors specify that this product is a precipitation estimate, as opposed to the current version (used in this project) which is a rainfall estimate.

Bibliography

AghaKouchak, A., Mehran, A., Norouzi, H., and Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, 39(9):1–4.

Braithwaite, D. (2013). Personal communication. Center for Hydrometeorology and Remote Sensing (CHRS), University of California, Irvine.

Brooks, H. E., Lee, J. W., and Craven, J. P. (2003). The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmospheric Research*, 67–68(0):73–94.

Capderou, M. (2005). *Satellites: Orbits and missions*. Springer, Paris.

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

Daley, R. (1991). *Atmospheric data analysis*. Cambridge University Press, Cambridge, UK.

Devine, K. A. and Mekis, É. (2008). Field accuracy of Canadian rain measurements. *Atmosphere-Ocean*, 46(2):213–227.

Ebert, E. E., Janowiak, J. E., and Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, 88(1):47–64.

Environment Canada (2014). Climate trends and variations bulletin - Annual 2013. Ottawa.

Environment Canada (2014a). mist.cfg. Source code provided by Environment Canada.

Environment Canada (2014b). obscapa.cfg. Source code provided by Environment Canada.

Ferraro, R. R. and Marks, G. F. (1995). The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements. *Journal of Atmospheric and Oceanic Technology*, 12(4):755–770.

Ferraro, R. R., Weng, F., Grody, N. C., and Zhao, L. (2000). Precipitation characteristics over land from the NOAA-15 AMSU sensor. *Geophysical Research Letters*, 27(17):2669–2672.

Fortin, V., Roy, G., and Donaldson, N. (2014). Assimilation of radar QPE in the Canadian Precipitation Analysis (CaPA). In *Proceedings of the International Symposium Weather Radar and Hydrology*, Washington, DC. American Society of Civil Engineers Environmental & Water Resources Institute.

Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., and Kasheta, T. (2007). Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrological Processes*, 21(12):1576–1586.

Hong, Y., Hsu, K.-L., Sorooshian, S., and Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43(12):1834–1853.

Hsu, K.-L., Gao, X., Sorooshian, S., and Gupta, H. V. (1997). Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, 36(9):1176–1190.

Hsu, K.-L., Hong, Y., and Sorooshian, S. (2007). Rainfall estimation using a cloud patch classification map. In Levizzani, V., Bauer, P., and Turk, F., editors, *Measuring Precipitation From Space*, volume 28 of *Advances In Global Change Research*, pages 329–342. Springer Netherlands.

Hsu, K.-L. and Sorooshian, S. (2008). Satellite-based precipitation measurement using PERSIANN system. In Sorooshian, S., Hsu, K.-L., Coppola, E., Tomassetti, B., Verdecchia, M., and Visconti, G., editors, *Hydrological Modelling and the Water Cycle*, volume 63 of *Water Science and Technology Library*, pages 27–48. Springer Berlin Heidelberg.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J. (2010). The TRMM Multi-satellite Precipitation Analysis (tmpa). In Gebremichael, M. and Hossain, F., editors, *Satellite Rainfall Applications for Surface Hydrology*, pages 3–22. Springer Netherlands.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55.

Janowiak, J. E., Joyce, R. J., and Yarosh, Y. (2001). A realtime global half-hourly pixel-resolution infrared dataset and its applications. *Bulletin of the American Meteorological Society*, 82(2):205–217.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, 5(3):487–503.

Joyce, R. J., Xie, P., Yarosh, Y., Janowiak, J. E., and Arkin, P. A. (2010). CMORPH: A "morphing" approach for high resolution precipitation product generation. In Gebremichael, M. and Hossain, F., editors, *Satellite Rainfall Applications for Surface Hydrology*, pages 23–37. Springer Netherlands.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R. J., Hsu, K.-L., and Braithwaite, D. (2011). Intercomparison of high-resolution precipitation products over northwest Europe. *Journal of Hydrometeorology*, 13(1):67–83.

Kidd, C. and Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, 15(4):1109–1116.

Kummerow, C., Olson, W. S., and Giglio, L. (1996). A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *Geoscience and Remote Sensing, IEEE Transactions on*, 34(5):1213–1232.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., and Stadnyk, T. (2014). Performance evaluation of the Canadian Precipitation Analysis (CaPA). Manuscript submitted for publication.

Lin, A. and Wang, X. L. (2011). An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada. *Journal of Geophysical Research: Atmospheres*, 116(D21111):1–19.

Mahfouf, J., 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., Méthot, A., Patoine, A., Plante, A., Rahill, A., Robinson, T., Talbot, D., Tremblay, A., Vaillancourt, P., Zadra, A., and Qaddouri, A. (2006). The 15 km version of the Canadian regional forecast system. *Atmosphere-Ocean*, 44(2):133–149.

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

Mekis, É. and Vincent, L. A. (2011). An overview of the Second Generation Adjusted Daily Precipitation Dataset for trend analysis in Canada. *Atmosphere-Ocean*, 49(2):163–177.

Meyer-Christoffer, A., Becker, A., Finger, P., Rudolf, B., Schneider, U., and Ziese, M. (2011). GPCC climatology version 2011 at 0.25°: Monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data.

Munchak, S. J. and Skofronick-Jackson, G. (2013). Evaluation of precipitation detection over various surfaces from passive microwave imagers and sounders. *Atmospheric Research*, 131(0):81–94.

National Ice Center (2008). IMS daily northern hemisphere snow and ice analysis at 4 km and 24 km resolution. Boulder, CO: National Snow and Ice Data Center.

Raddatz, R. L. and Hanesiak, J. M. (2008). Significant summer rainfall in the Canadian prairie provinces: modes and mechanisms 2000–2004. *International Journal of Climatology*, 28(12):1607–1613.

Rossa, A., Nurmi, P., and Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In Michaelides, S., editor, *Precipitation: Advances in Measurement, Estimation and Prediction*, pages 419–452. Springer Berlin Heidelberg.

Sapiano, M. R. P. and Arkin, P. A. (2009). An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology*, 10(1):149–166.

Sohn, B. J., Han, H.-J. . J., and Seo, E.-K. . K. (2009). Validation of satellite-based high-resolution rainfall products over the Korean Peninsula using data from a dense rain gauge network. *Journal of Applied Meteorology and Climatology*, 49(4):701–714.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H. (2013). Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophysical Research Letters*, 40(10):2060–2065.

Tian, Y., Peters-Lidard, C. D., Choudhury, B. J., and Garcia, M. (2007). Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *Journal of Hydrometeorology*, 8(6):1165–1183.

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., Hsu, K.-L., Turk, F. J., Garcia, M., and Zeng, J. (2009). Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research: Atmospheres*, 114(D24):1–19.

Wilheit, T., Adler, R., Avery, S., Barrett, E., Bauer, P., Berg, W., Chang, A., Ferriday, J., Grody, N., Goodman, S., Kidd, C., Kniveton, D., Kummerow, C., Mugnai, A., Olson, W., Petty, G., Shibata, A., and Smith, E. (1994). Algorithms for the retrieval of rainfall from passive microwave measurements. *Remote Sensing Reviews*, 11(1-4):163–194.

World Meteorological Organization (2014). Observing systems capability analysis and review tools. Accessed July 2014. Database.

Xie, P. and Joyce, R. J. (2014). Developing the Second Generation CMORPH: A prototype. Presented at the European Geosciences Union General Assembley, Austria, Vienna.

Zhao, L. and Weng, F. (2002). Retrieval of ice cloud parameters using the Advanced Microwave Sounding Unit. *Journal of Applied Meteorology*, 41(4):384–395.

Appendix A | Current satellites and instrumentation acronyms

The following defines the acronyms in Table 2.1:

Table A.1: Current satellite information.

Acronym	Definition
NASA	National Aeronautics and Space Administration
JAXA	Japan Aerospace Exploration Agency
NOAA	National Oceanic and Atmospheric Administration
DMSP	Defense Meteorological Satellite Program
JMA	Japan Meteorological Agency
Metop	Meteorological operational satellite
MTSAT	Multifunction Transport Satellite
GOES	Geostationary Operational Environmental Satellite
TMI	TRMM Microwave Imager
AMSU	Advanced Microwave Sounding Unit
MHS	Microwave Humidity Sounders
AMSR-E	Advanced Microwave Scanning Radiometer − Earth Observing System
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager/Sounder
MVIRI	Meteosat Visible Infra-Red Imager
SEVIRI	Spinning Enhanced Visible Infra-Red Imager

Appendix B | Error quantification

B.1 Bias and categorical scores

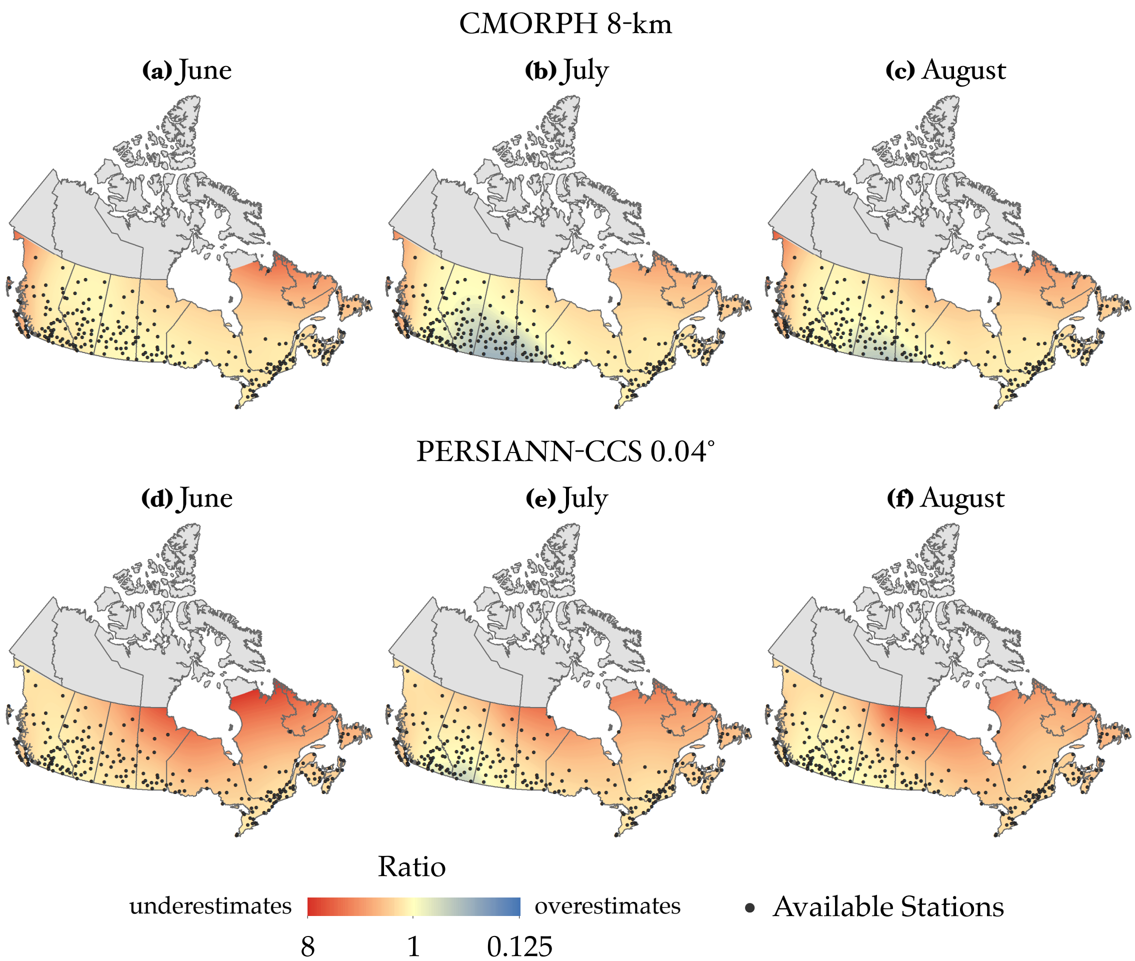


Figure B.1: Mean monthly ratios of observed/satellite for the 8-km CMORPH and PERSIANN-CCS products and APC2 from 2001 to 2010. Applicable stations are shown in black.

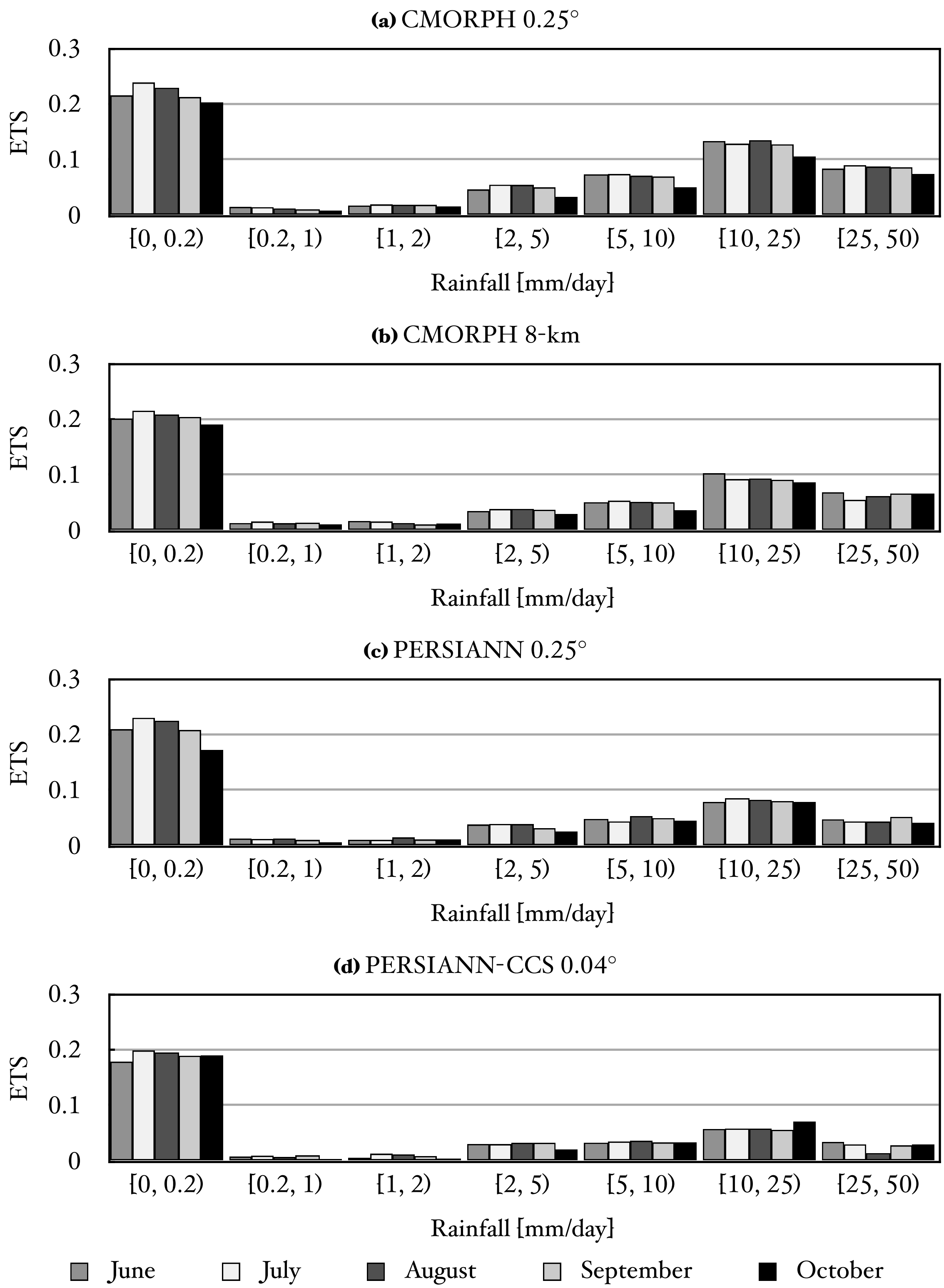


Figure B.2: Monthly equitable threat scores for the satellite rainfall products across Canada versus APC2 observations from 2001 to 2010.

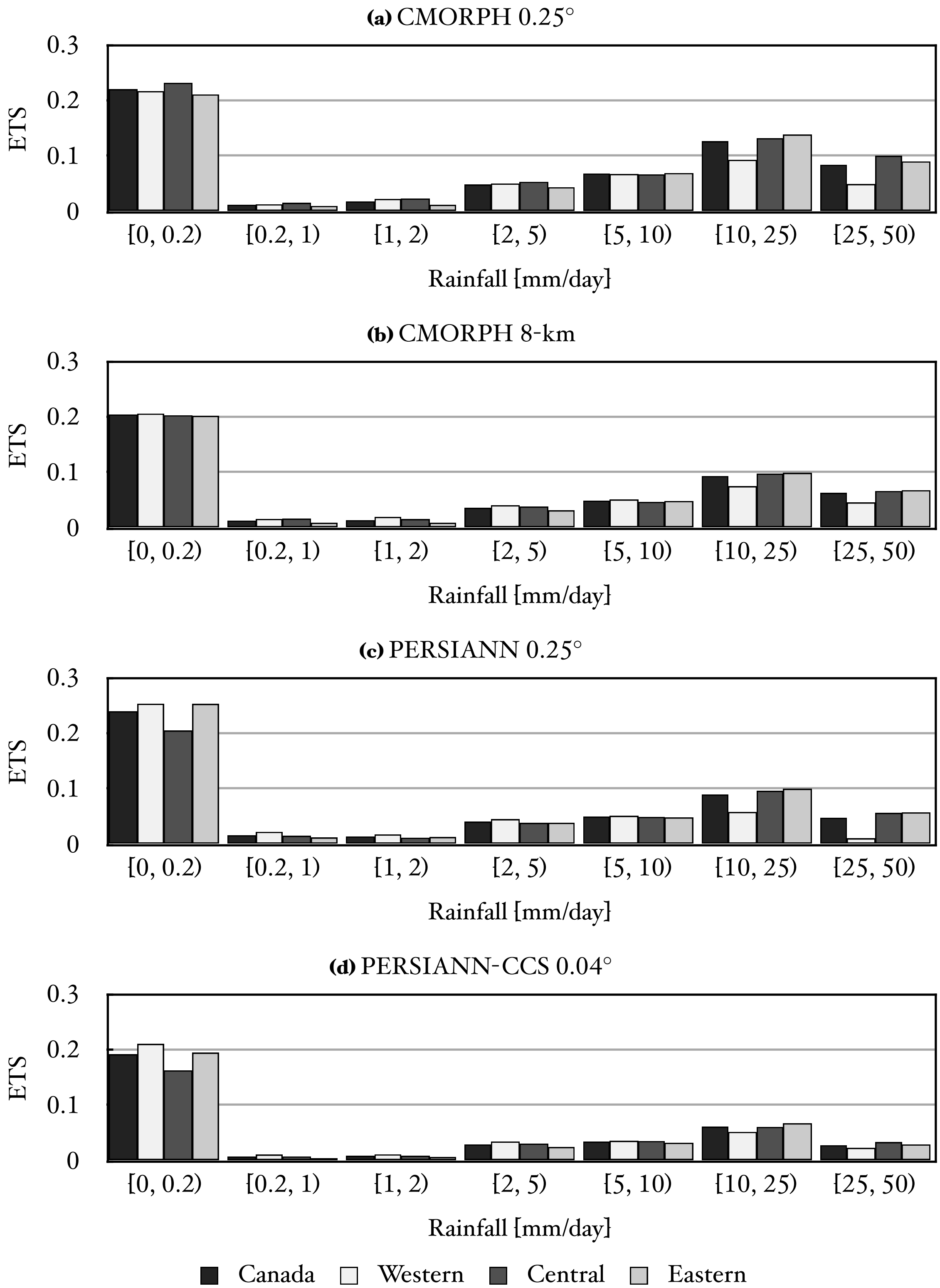
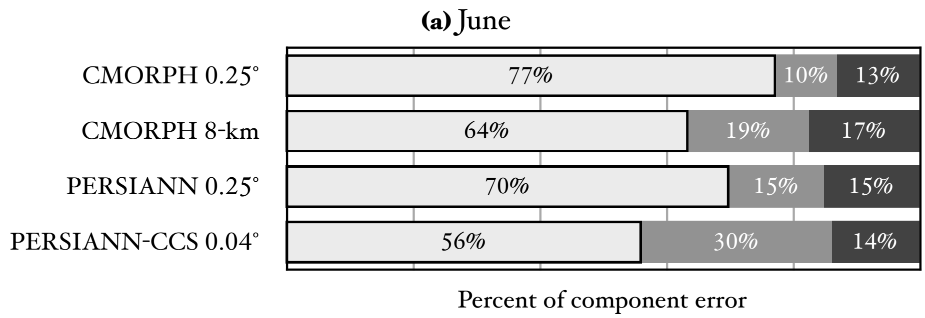
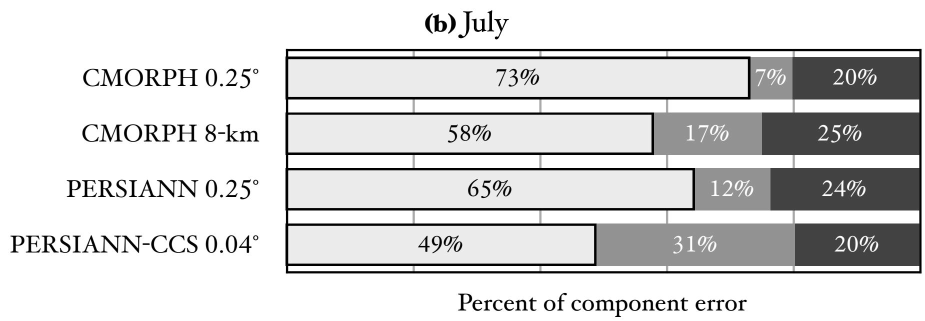
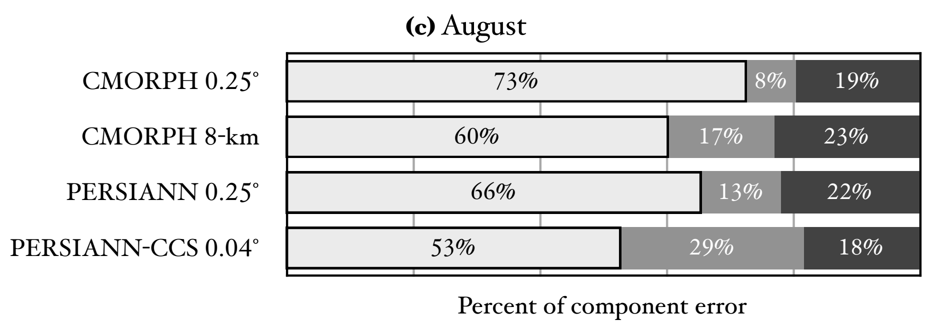


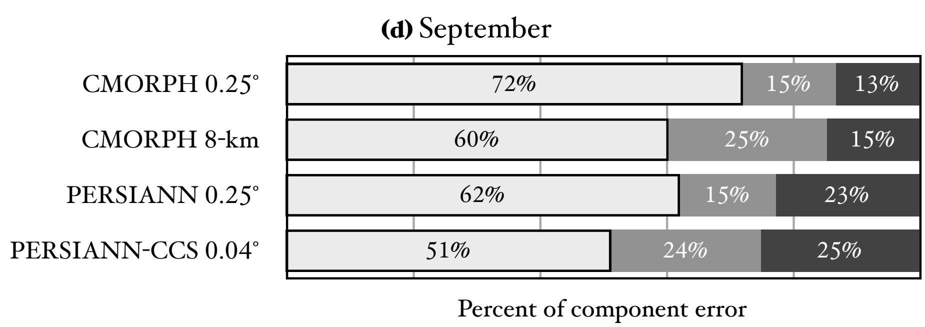
Figure B.3: Regional equitable threat scores for the satellite rainfall products versus APC2 observations for June through October, 2001 to 2010.

B.2 Monthly hit, miss, and false alarm contributions to total error









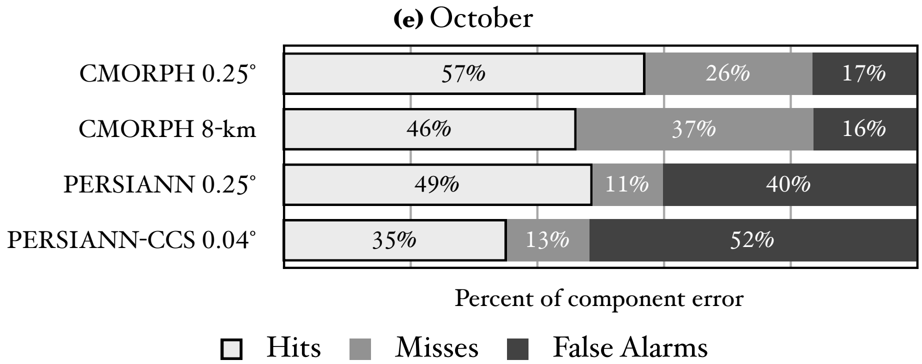


Figure B.4: Percent contribution of categorized error to total error in central Canada when comparing satellite based precipitation products to APC2 for June through October, 2001 to 2010.

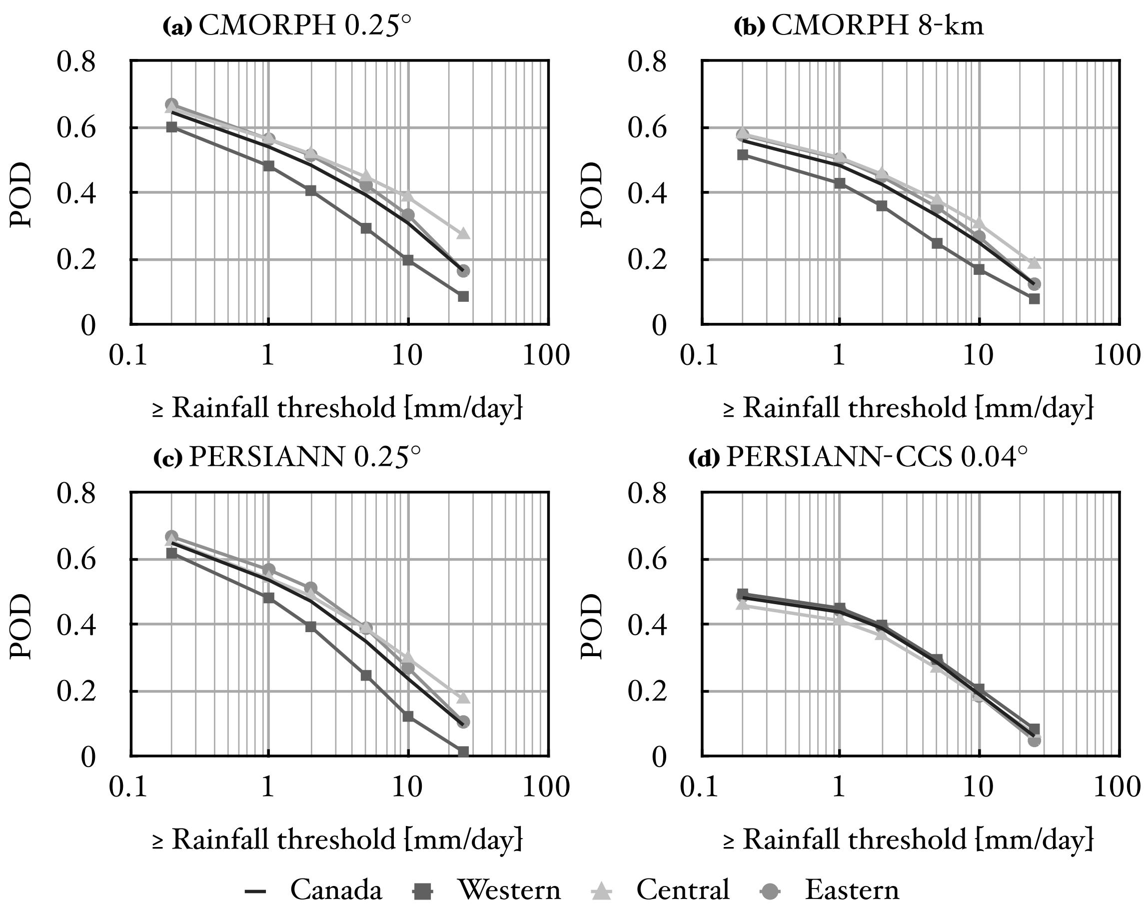


Figure B.5: Probabilities of detection for various satellite products with APC2 for June to October from 2001 to 2010 (2003 to 2010 for PERSIANN-CCS).

Appendix C | Satellite precipitation data preparation

C.1 Discrepancies between APC2 and 6-hourly data

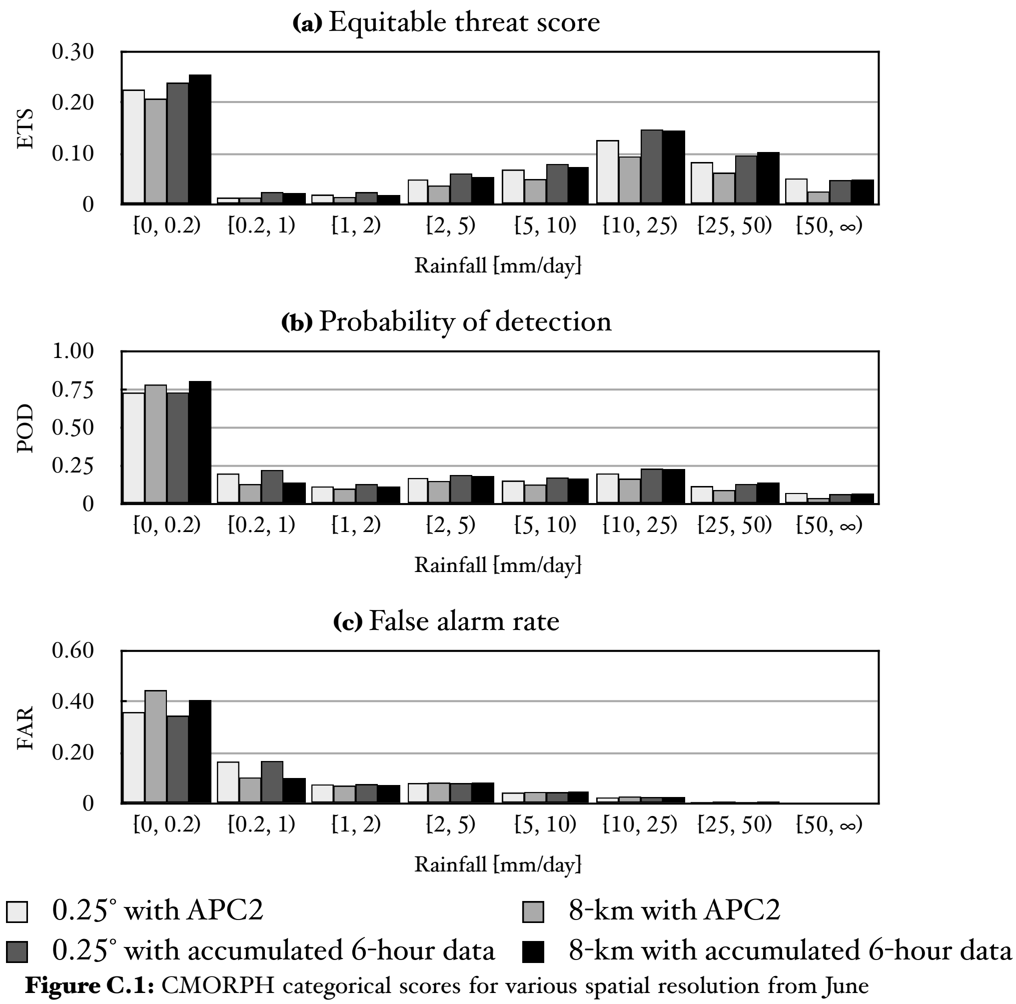


Figure C.1: CMORPH categorical scores for various spatial resolution from June through October, 2002 to 2010. As observations, 6-hourly data from manual stations has been accumulated to daily values to be fairly compared with the daily APC2.

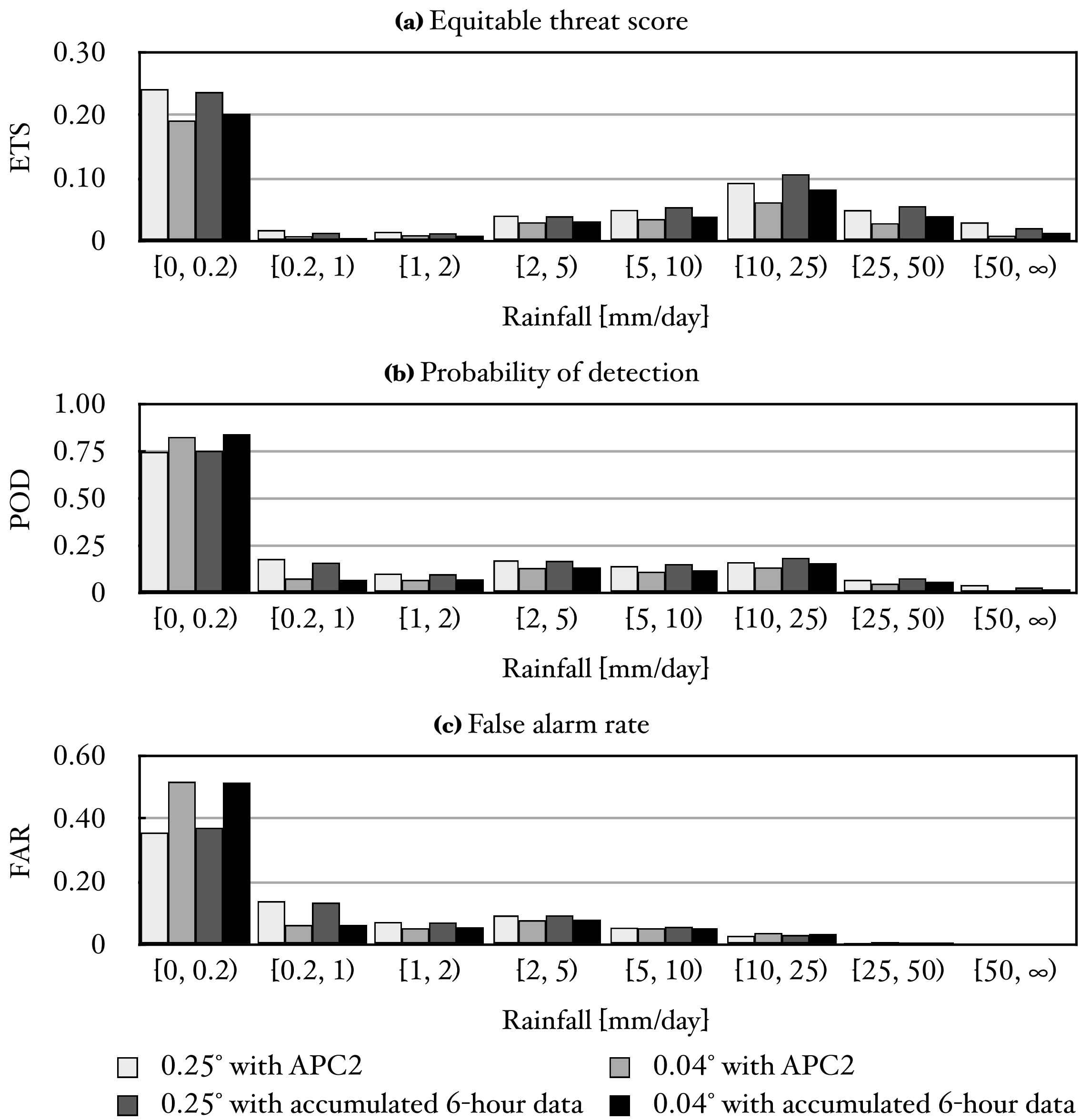


Figure C.2: Same as Figure C.1, but for PERSIANN.

C.2 Dataset adjustment selection

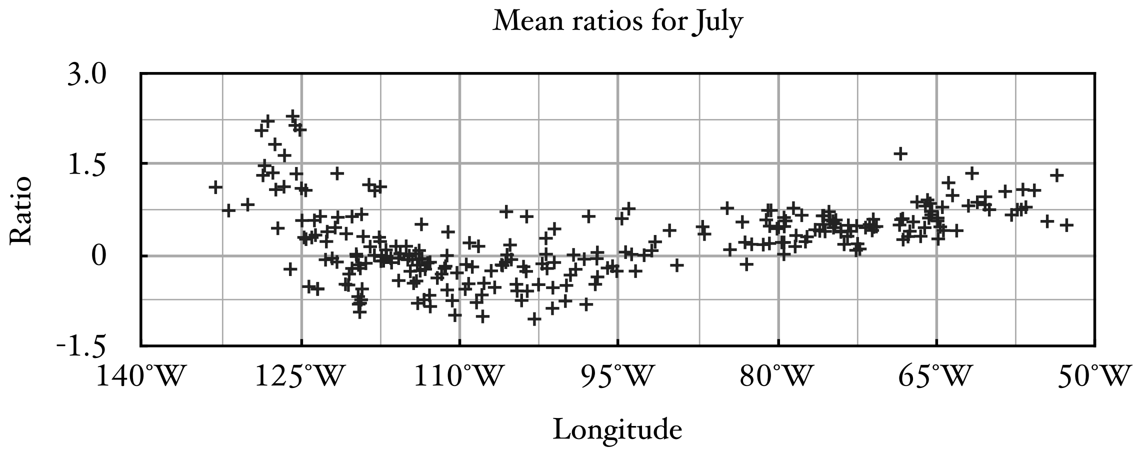


Figure C.3: Ratios of observed to satellite data between APC2 and the CMORPH 0.25° product for July from 2001 to 2010 for each station as a function of longitude.

Appendix D | Satellite data as a background field

D.1 Adjusting with manual stations versus all stations

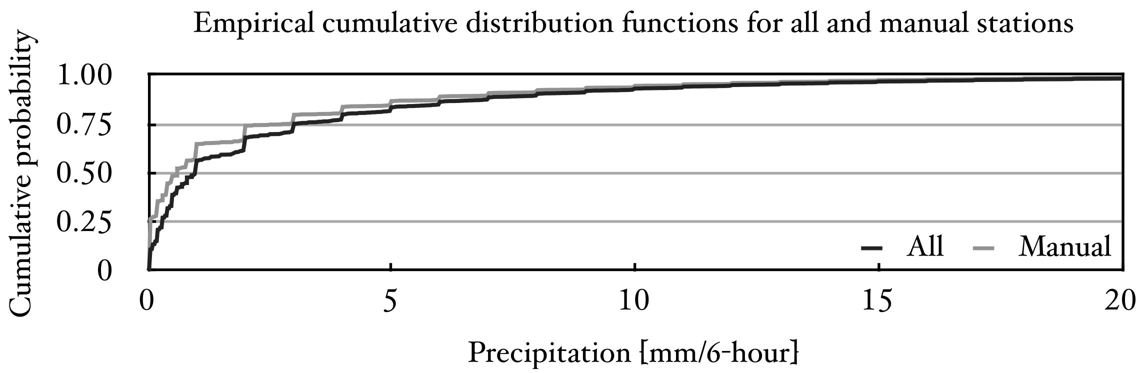


Figure D.1: Empirical distribution functions for 6-hourly data at manual and all stations within Canada for June through October from 2002 to 2011. The plot has been truncated at 20 mm/6-hour.

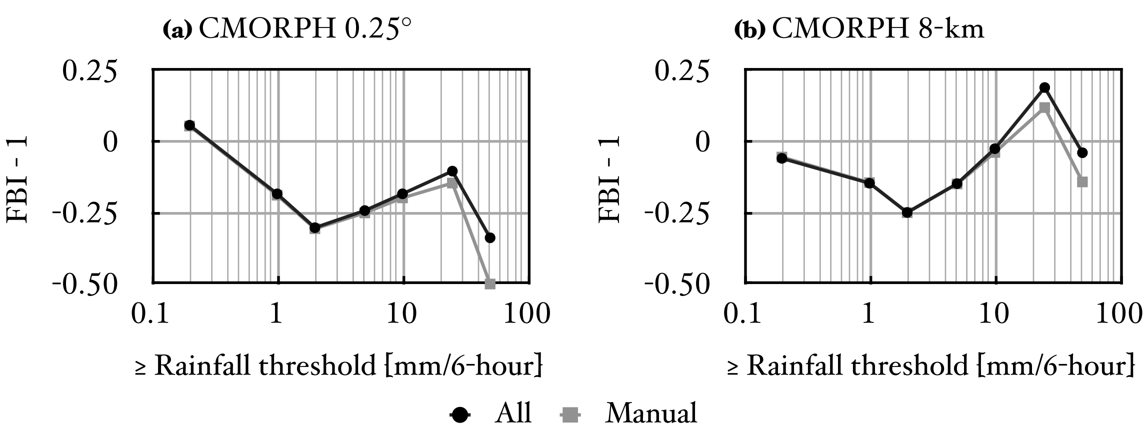


Figure D.2: Frequency bias indices for 0.25° and 8-km CMORPH products for June through October, 2002 to 2011. For each resolution, adjustments with all and manual stations are compared against manual stations.

D.2 Overall

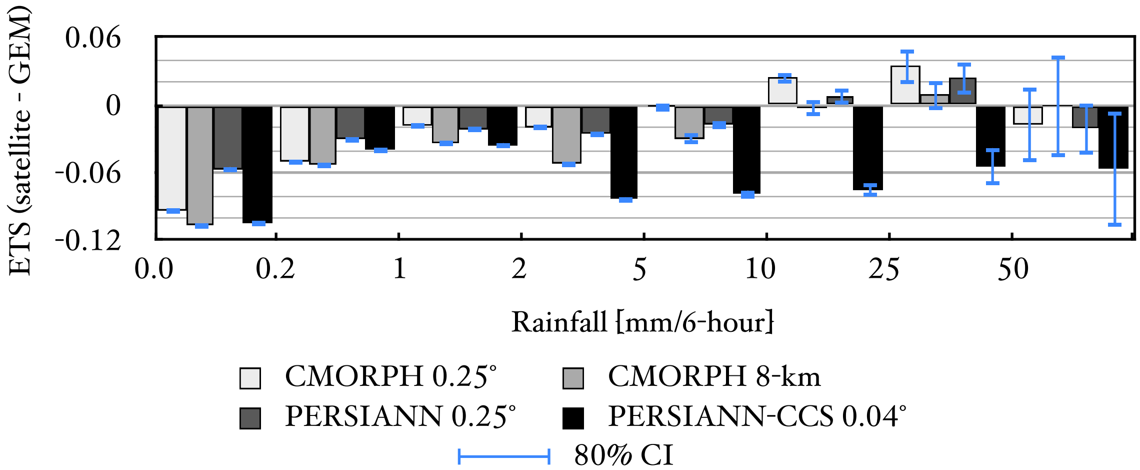


Figure D.3: Differences in equitable threat scores of CaPA assimilation results for using the satellite products as background fields with 80% confidence intervals. The analysis represents June though October, 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

D.3 Regionally

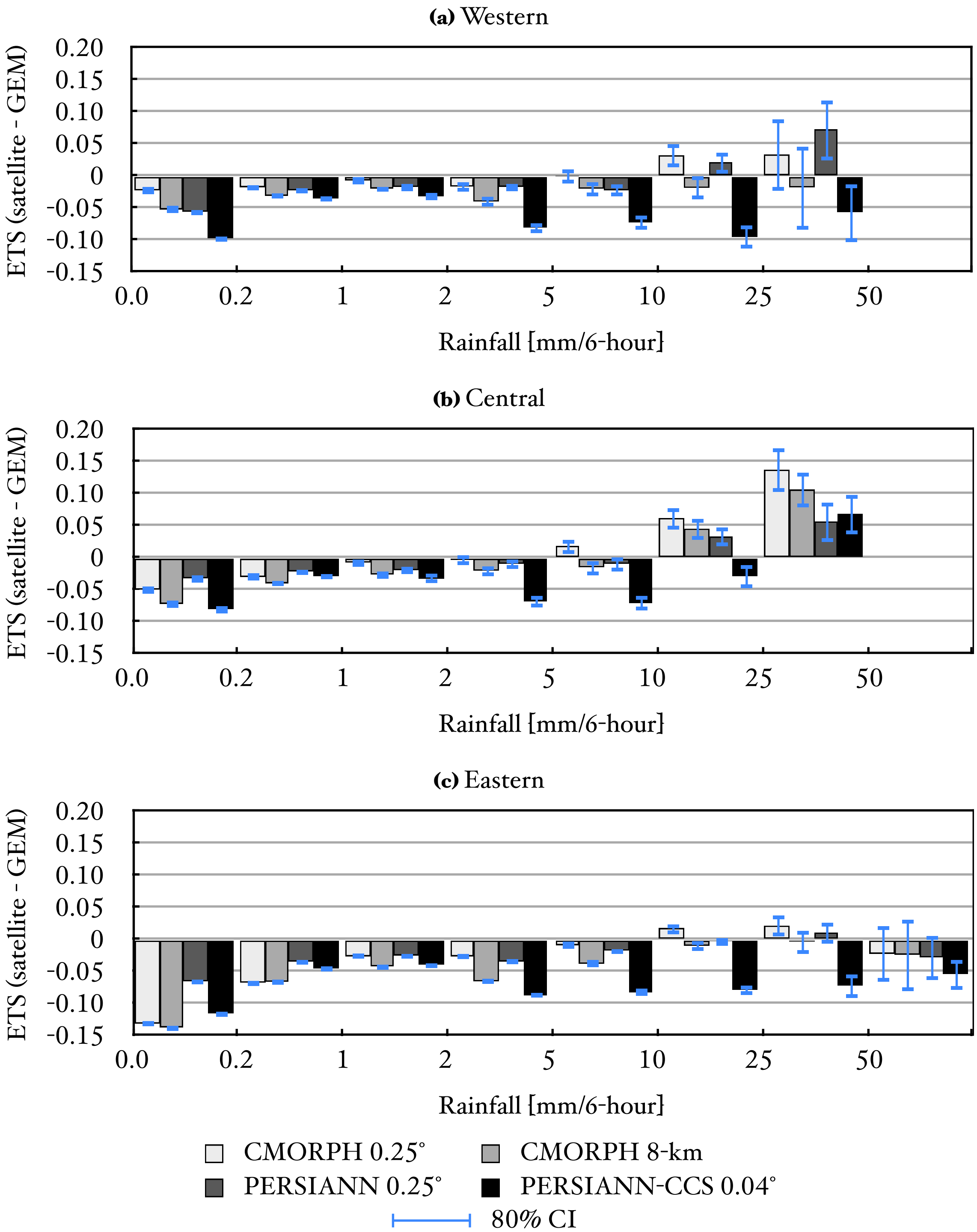
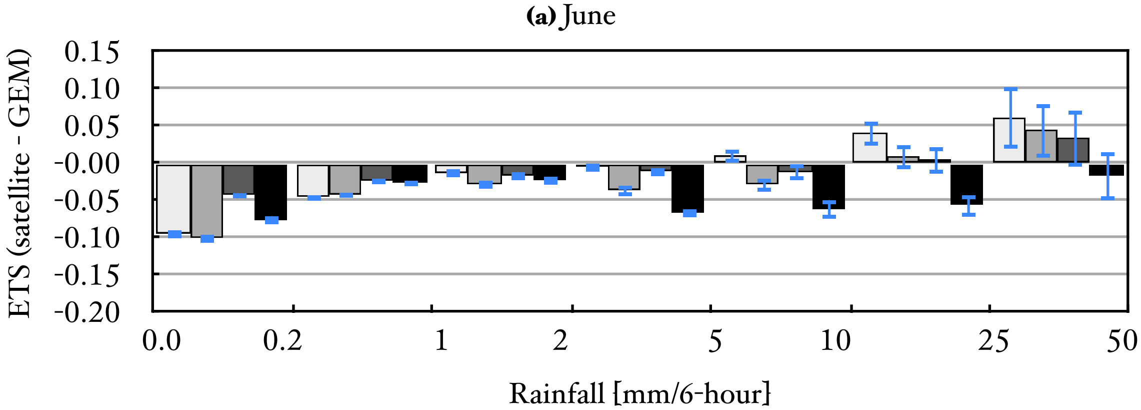
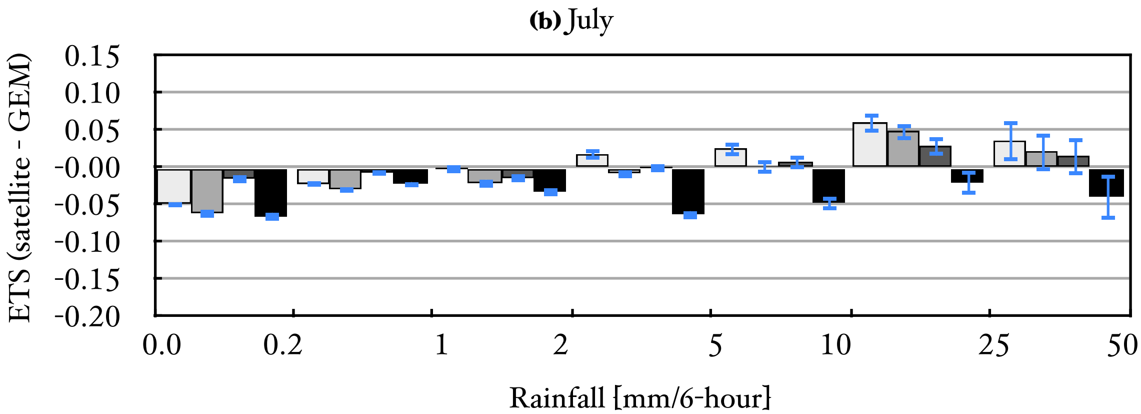
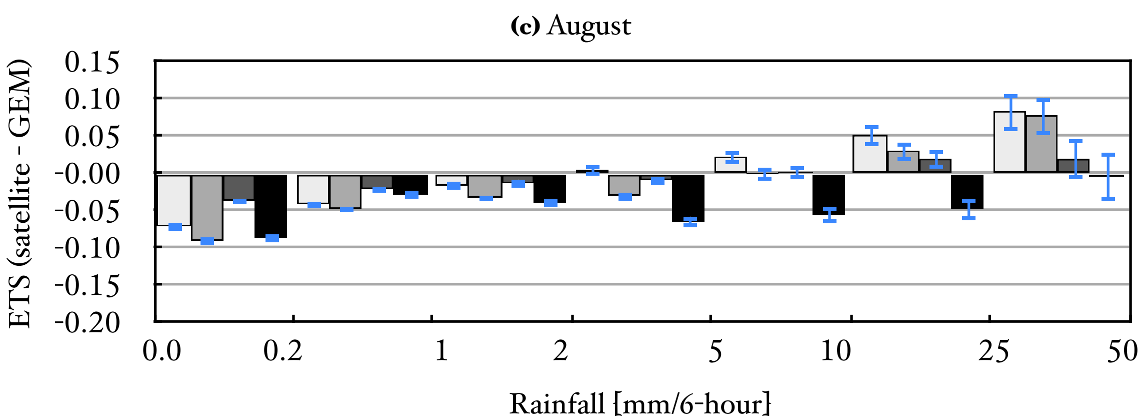


Figure D.4: Regionally differences in equitable threat scores from assimilating the satellite products with stations using CaPA relative to the operational form, with 80% confidence intervals. The analysis represents June though October, 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

D.4 Monthly


(a) June
ETS (satellite - GEM)
0.15
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.20
0.0
0.2
1
2
5
10
25
50
Rainfall [mm/6-hour]


(b) July
ETS (satellite - GEM)
0.15
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.20
0.0
0.2
1
2
5
10
25
50
Rainfall [mm/6-hour]


(c) August
ETS (satellite - GEM)
0.15
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.20
0.0
0.2
1
2
5
10
25
50
Rainfall [mm/6-hour]

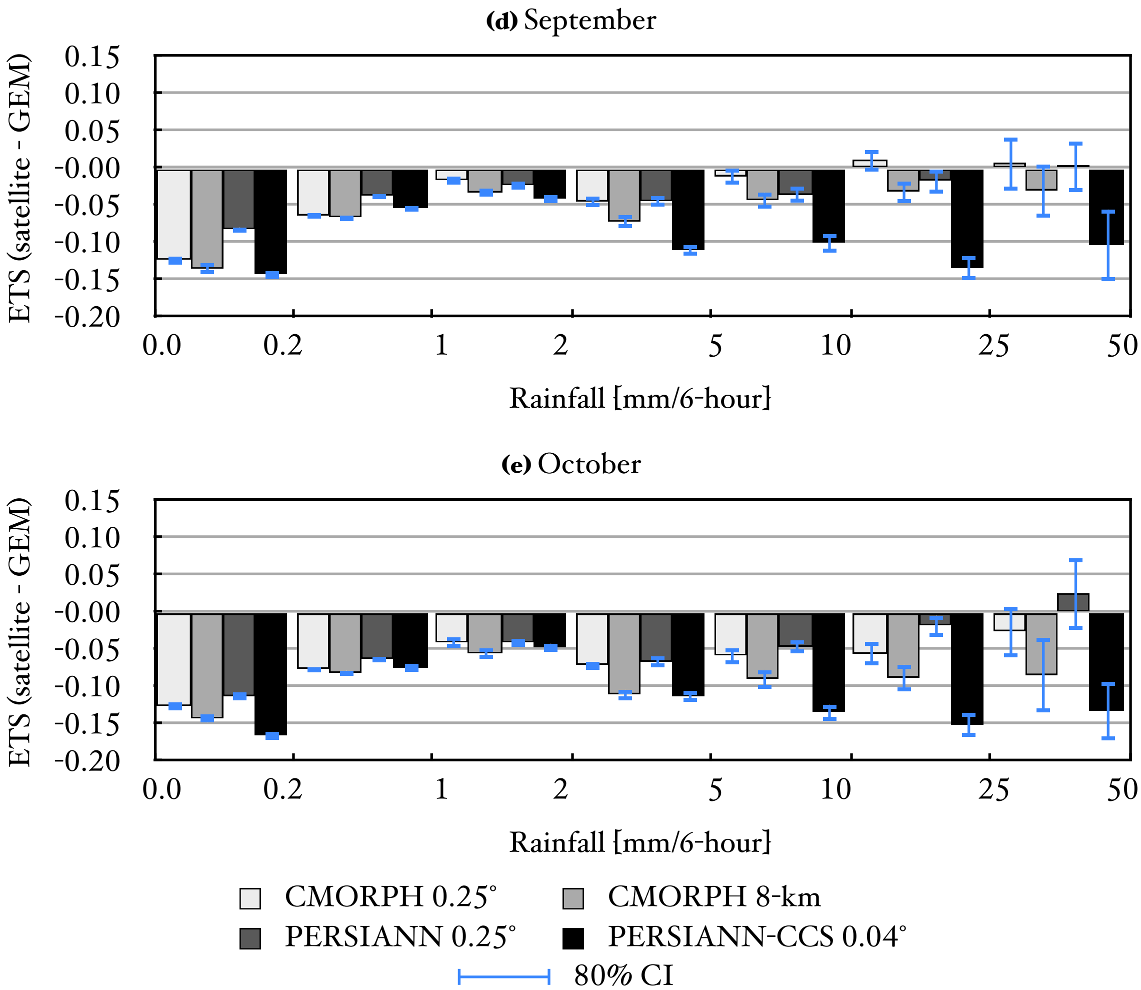


Figure D.5: Monthly differences in equitable threat scores of CaPA assimilation results for using the satellite products as background fields. The differences and 80% confidence intervals are relative to the operational form. The analysis includes months from 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

D.5 6-hourly

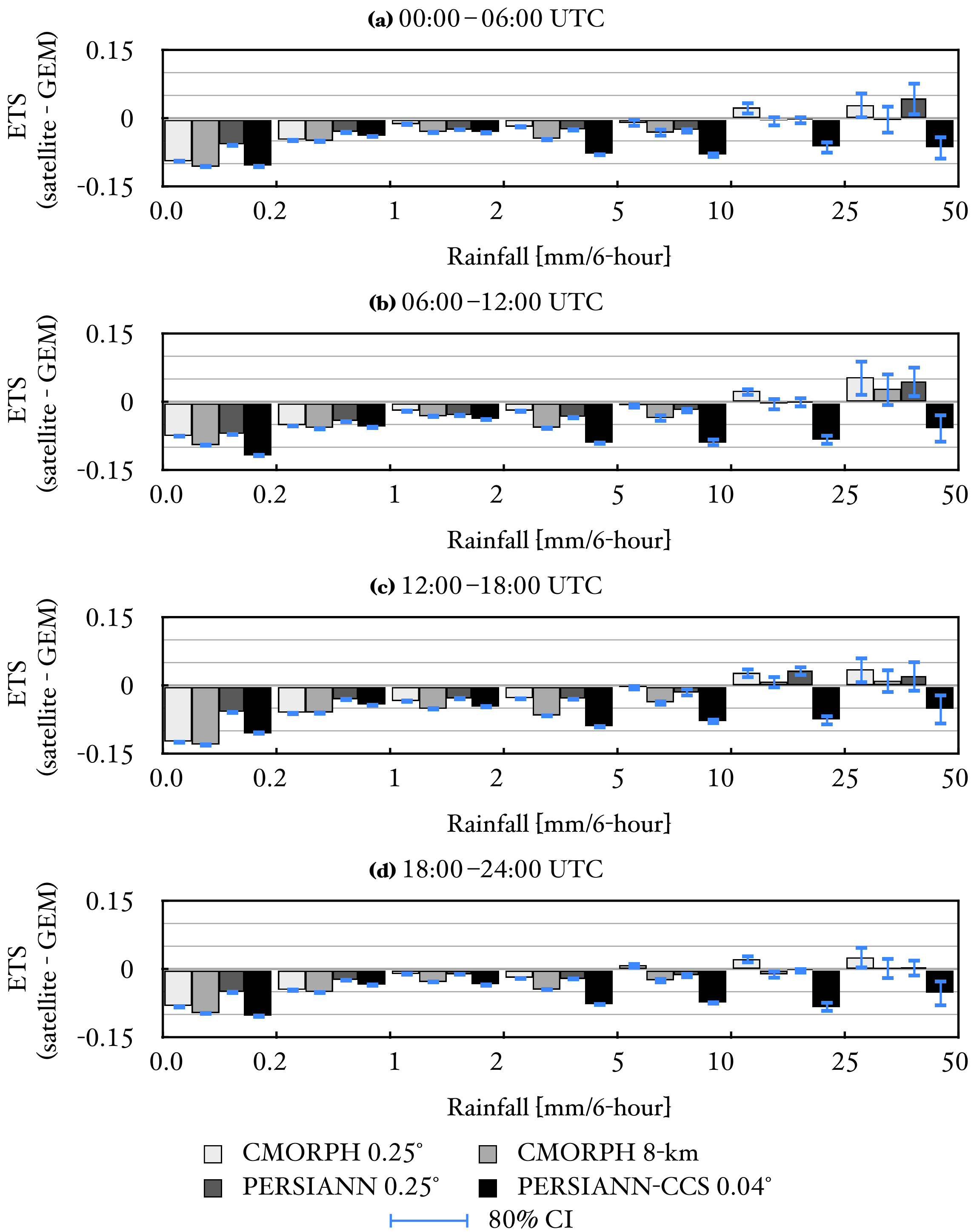


Figure D.6: Binned differences in equitable threat scores for various satellite products assimilated with stations less that from GEM for manual stations in Canada for June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

D.5.1 Regionally

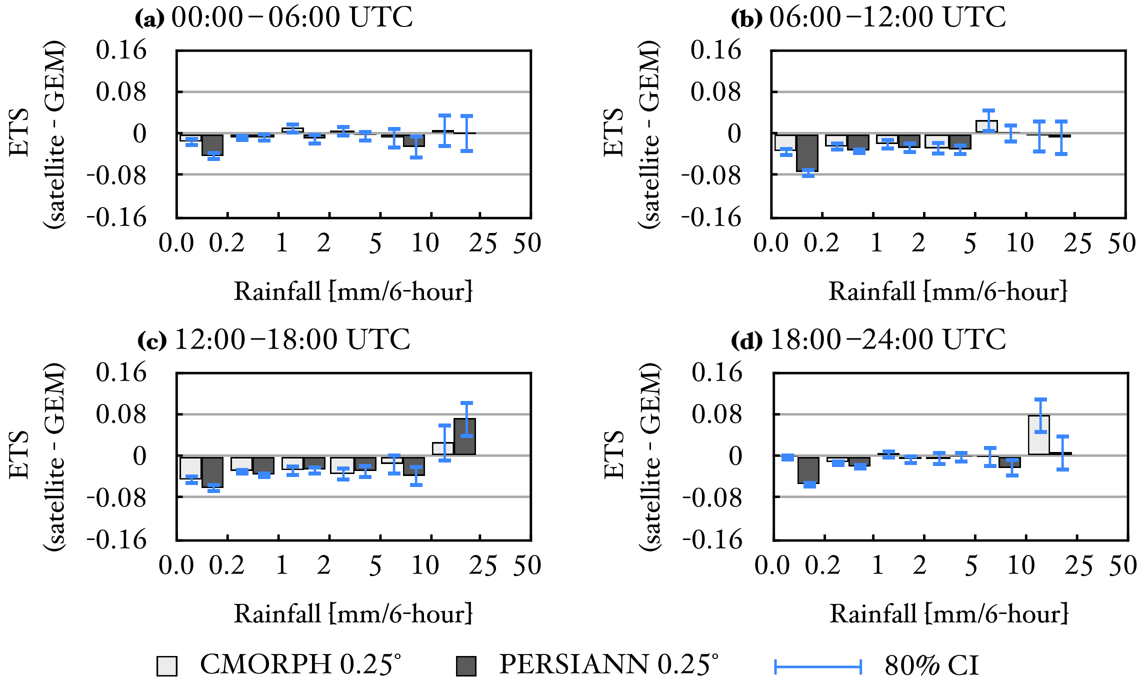


Figure D.7: Binned differences in equitable threat scores between satellite products and GEM after assimilation via CaPA. Scores are separated by time step and include data for June to October, 2002 to 2011 in western Canada.

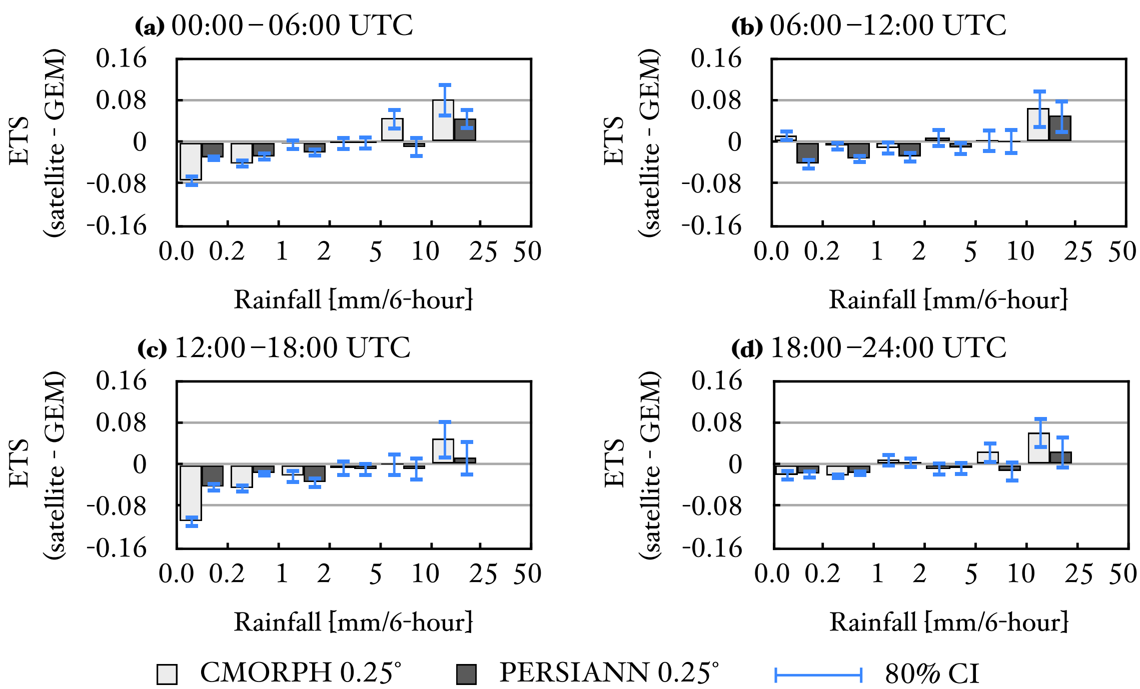


Figure D.8: Same as Figure D.7 but for central Canada.

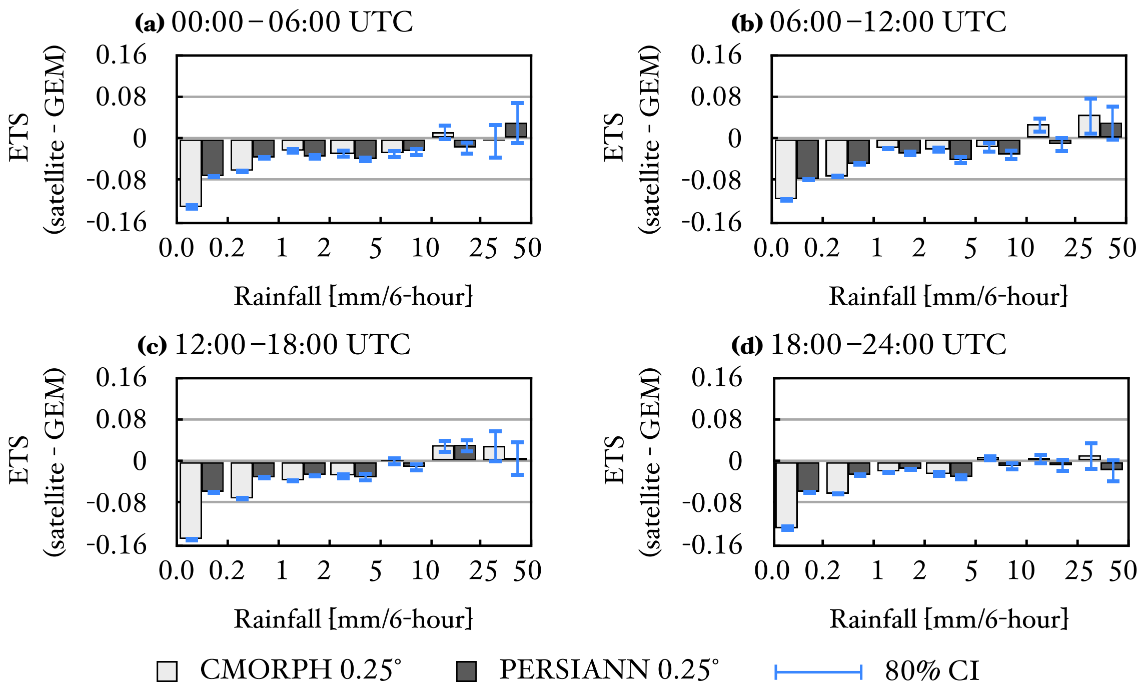


Figure D.9: Same as Figure D.7 but for eastern Canada.

Appendix E | Satellite data as an additional data source

E.1 Overall

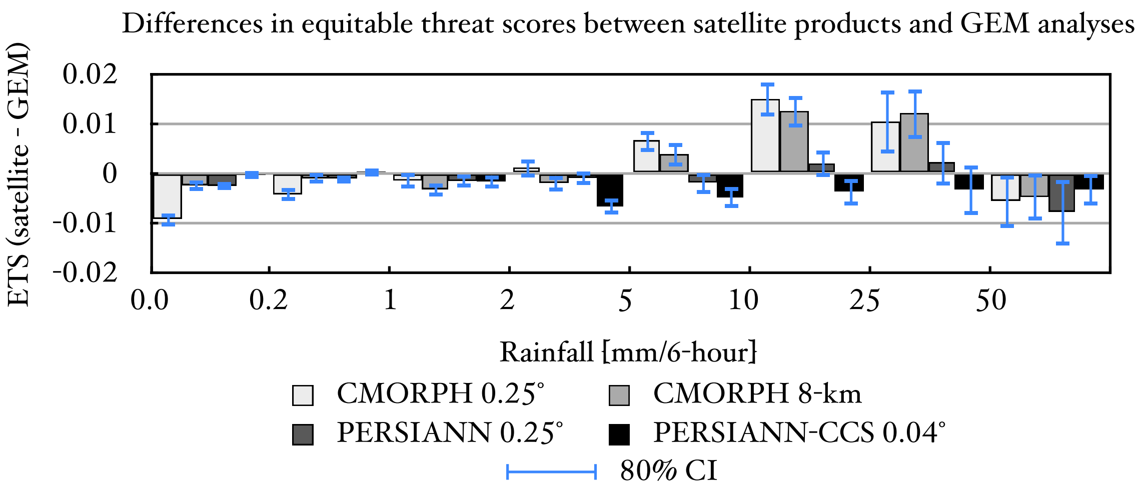


Figure E.1: Binned differences in equitable threat scores across Canada from using satellite data as an additional data source in the CaPA assimilation. The analysis represents June though October, 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

E.2 Regionally

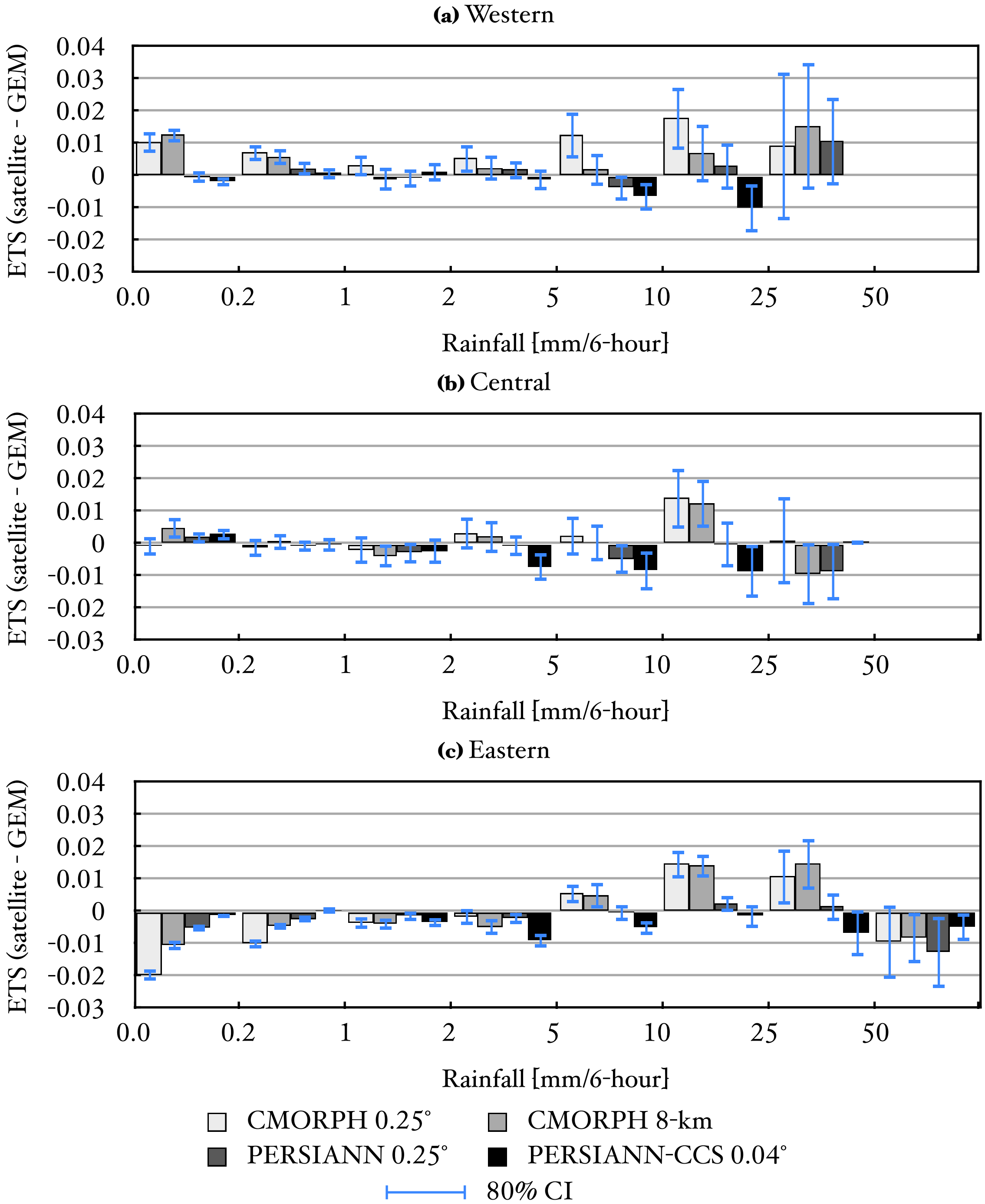
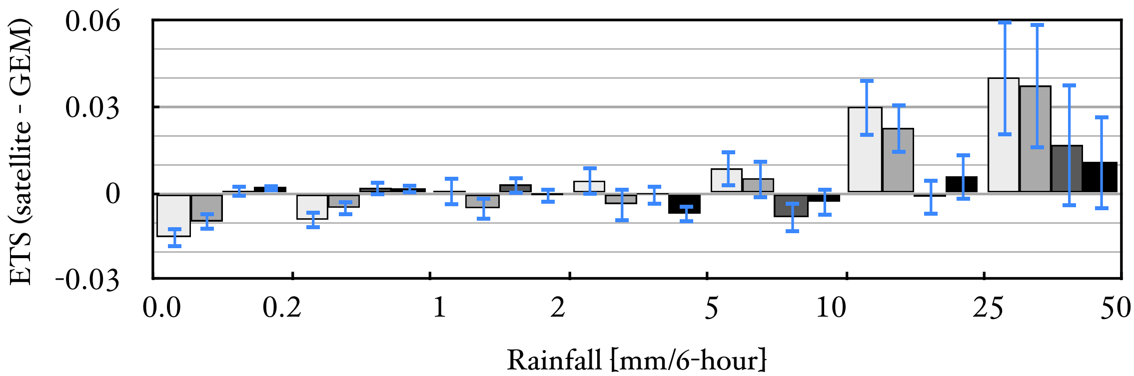


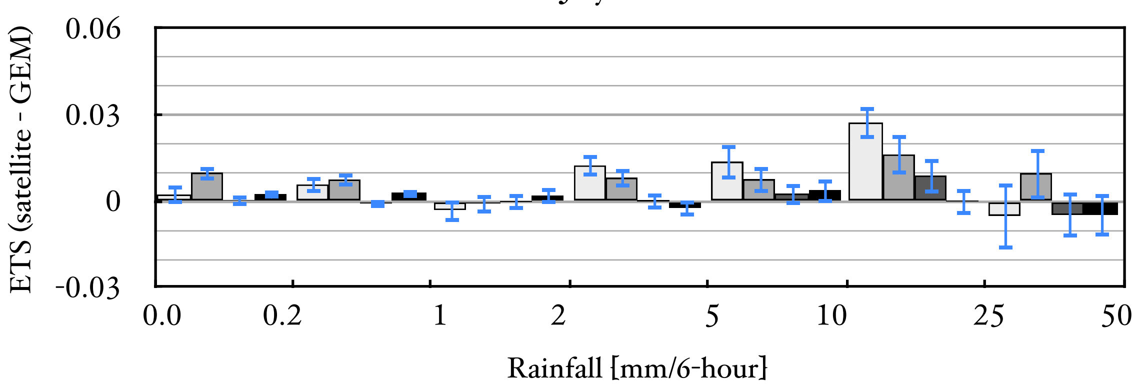
Figure E.2: Binned regional differences in equitable threat scores from using satellite data as an additional data source in the CaPA assimilation. The analysis represents June though October, 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

E.3 Monthly

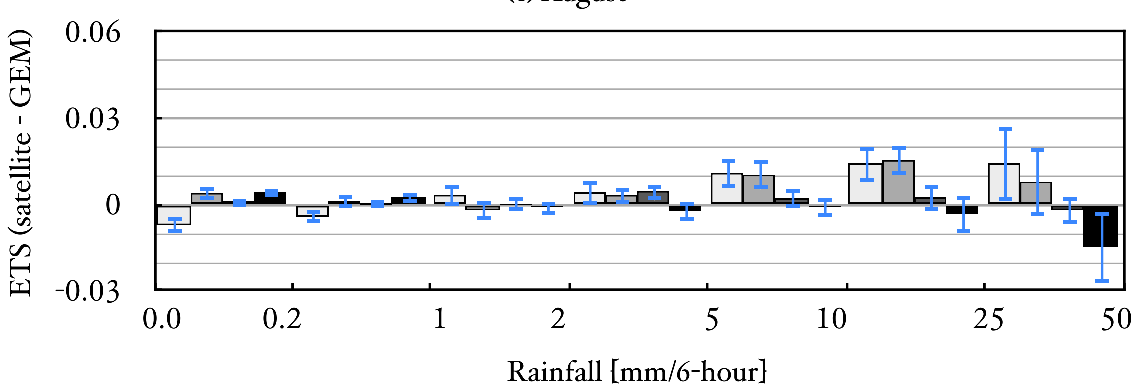
(a) June



(b) July



(c) August



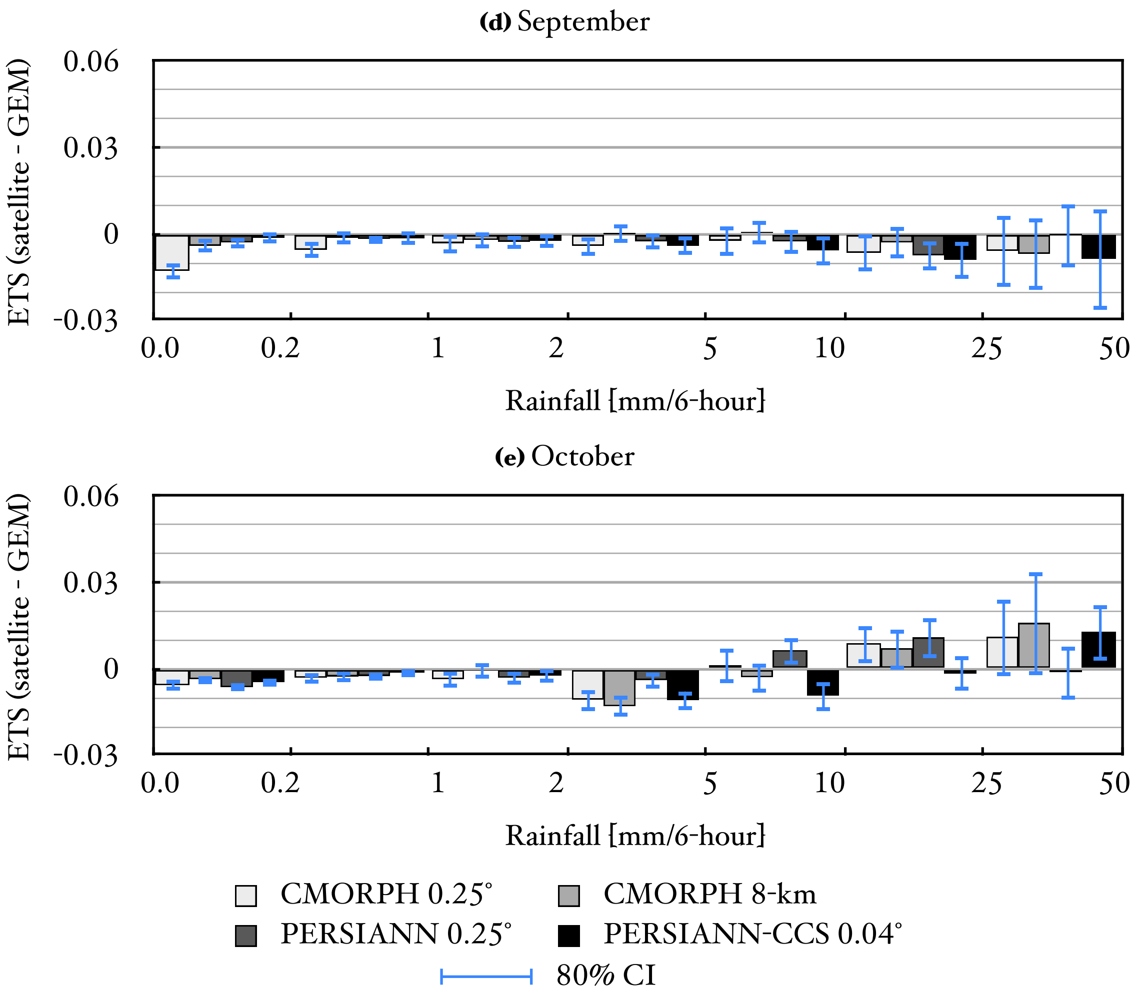


Figure E.3: Monthly binned differences in equitable threat scores from using satellite data as an additional data source in the CaPA assimilation from 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

E.4 6-hourly

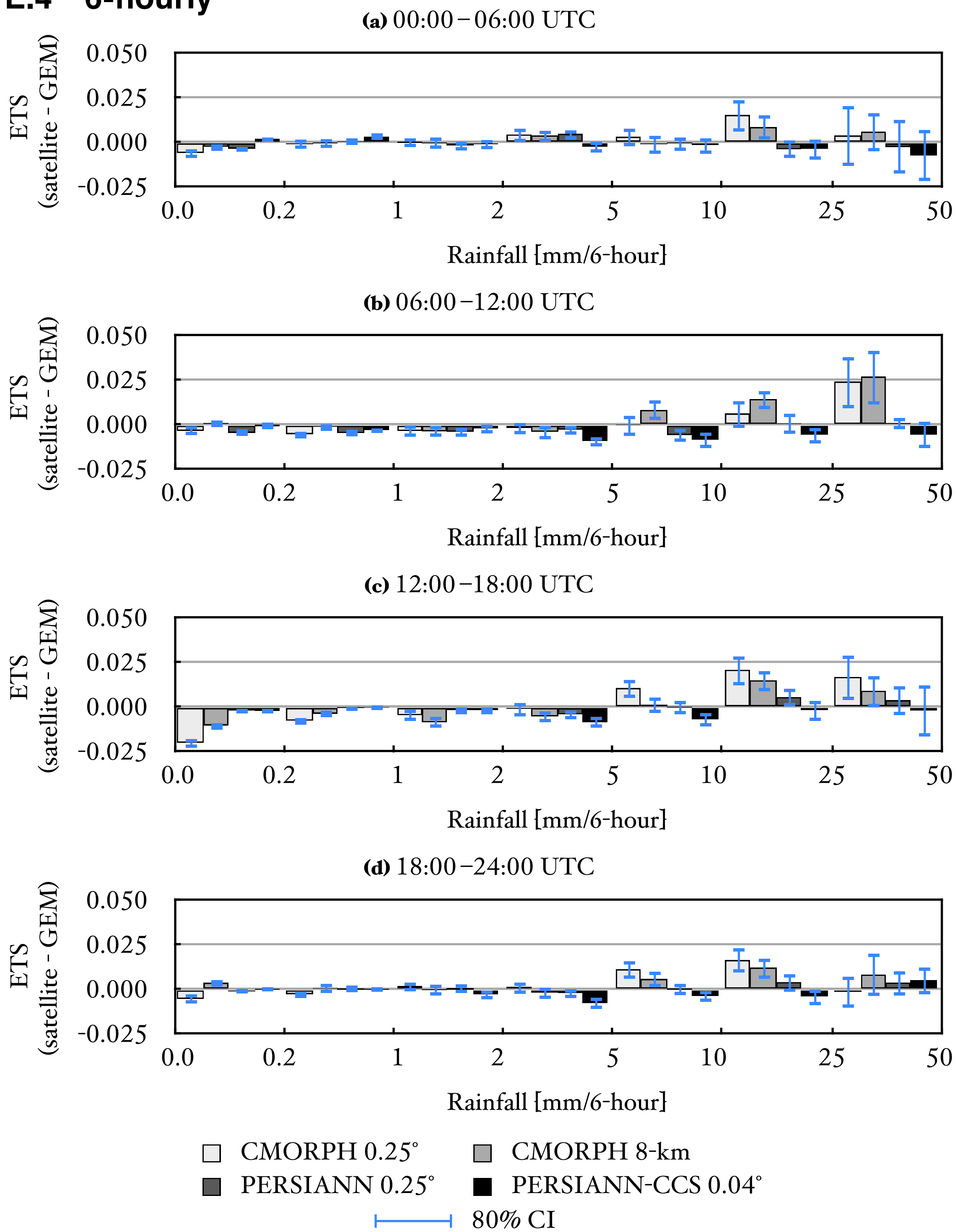


Figure E.4: Binned differences in equitable threat scores for each analysis time step across Canada from using satellite data as an additional data source in the CaPA assimilation. The analysis represents June though October, 2002 to 2011 (2003 to 2011 for PERSIANN-CCS).

E.4.1 Regionally

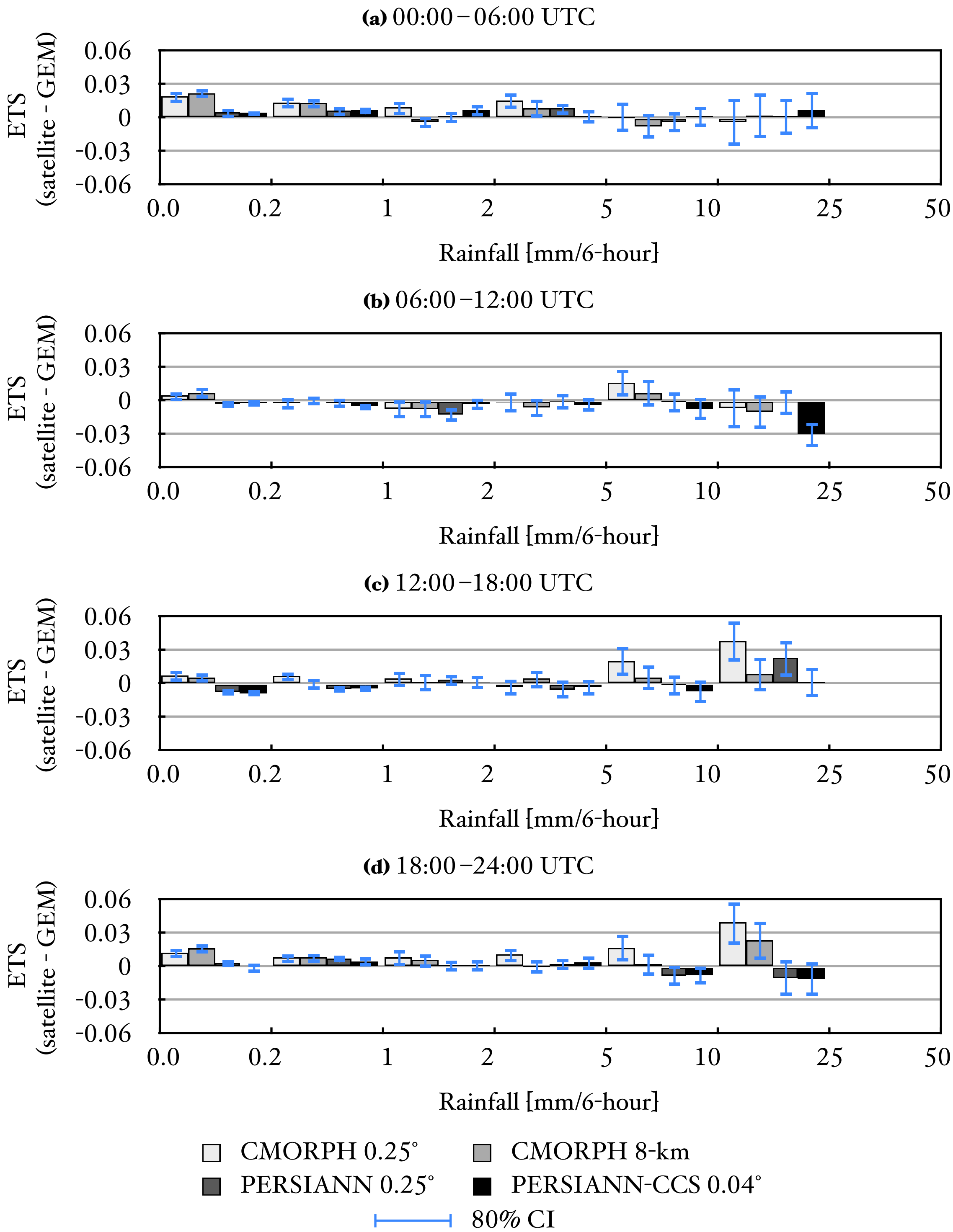


Figure E.5: Binned differences in equitable threat scores between satellite products and GEM after assimilation via CaPA. Scores are separated by time step and include data for June to October, 2002 to 2011 in western Canada.

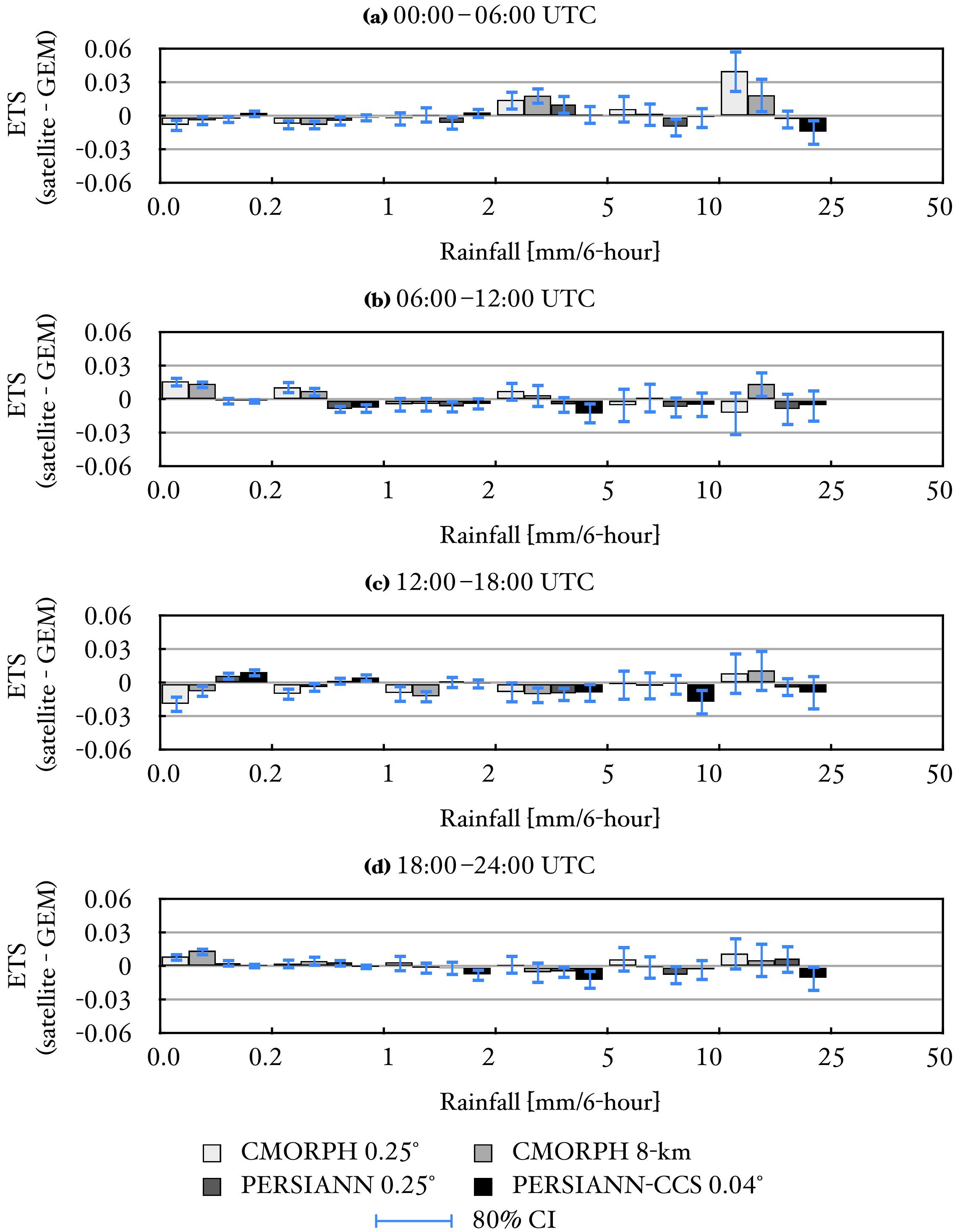


Figure E.6: Same as Figure E.5 but for central Canada.

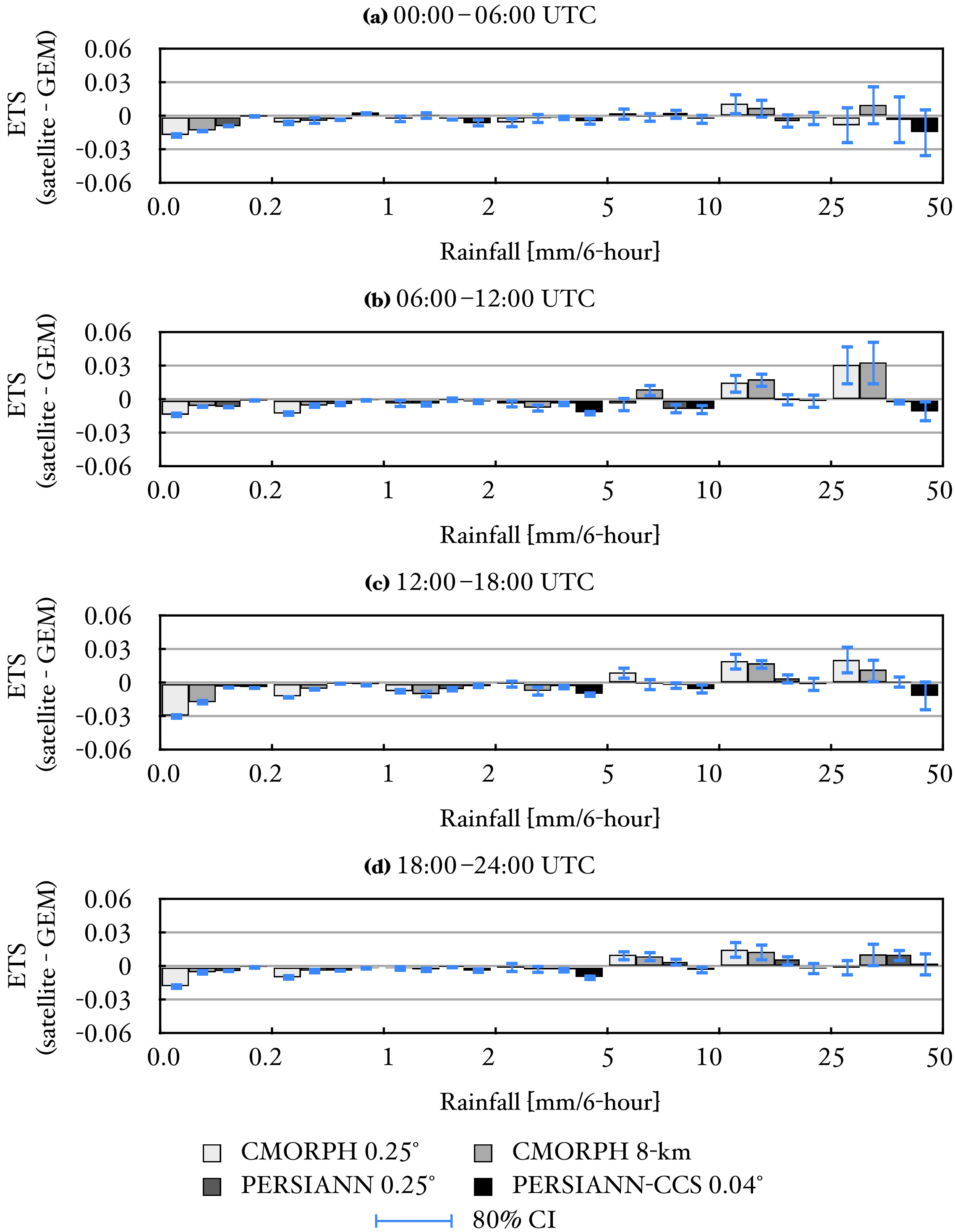


Figure E.7: Same as Figure E.5 but for eastern Canada.

E.4.2 Regional FBI

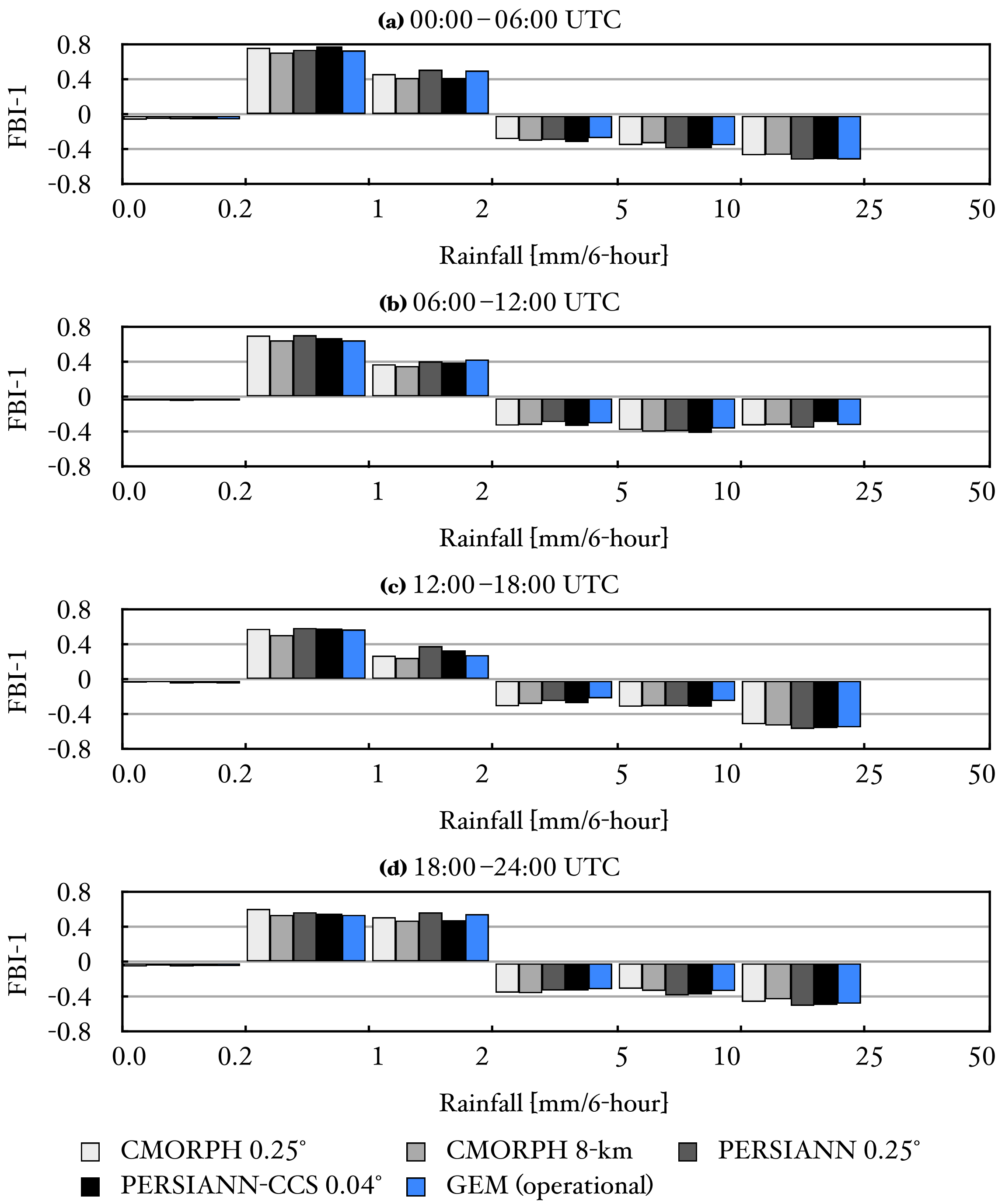


Figure E.8: Binned frequency bias indices of each time step for GEM assimilated with stations and various satellite products for manual stations across Canada from June through October, 2002 to 2011 (2003 onward for PERSIANN-CCS).

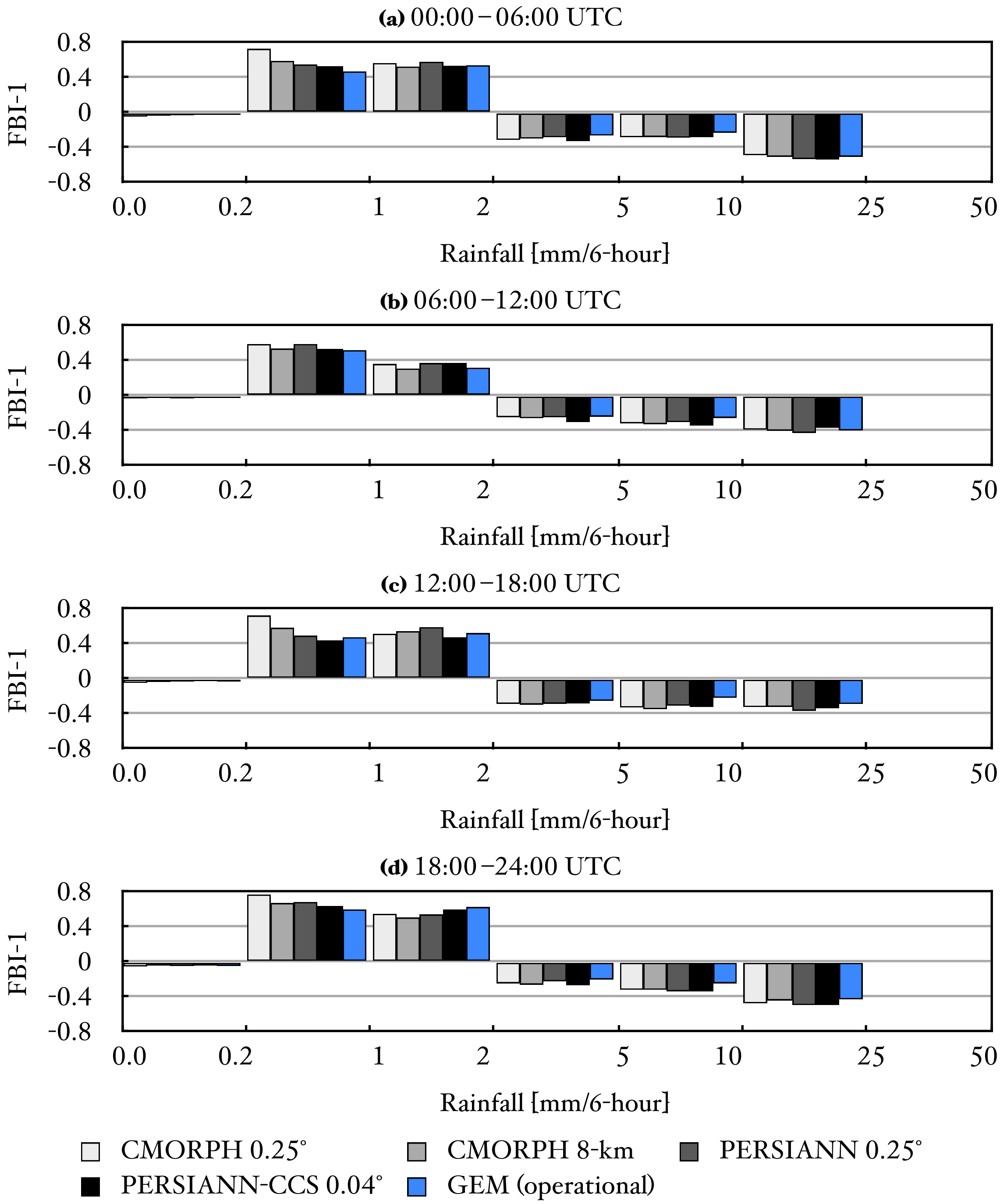


Figure E.9: Same as Figure E.8, but for central Canada.

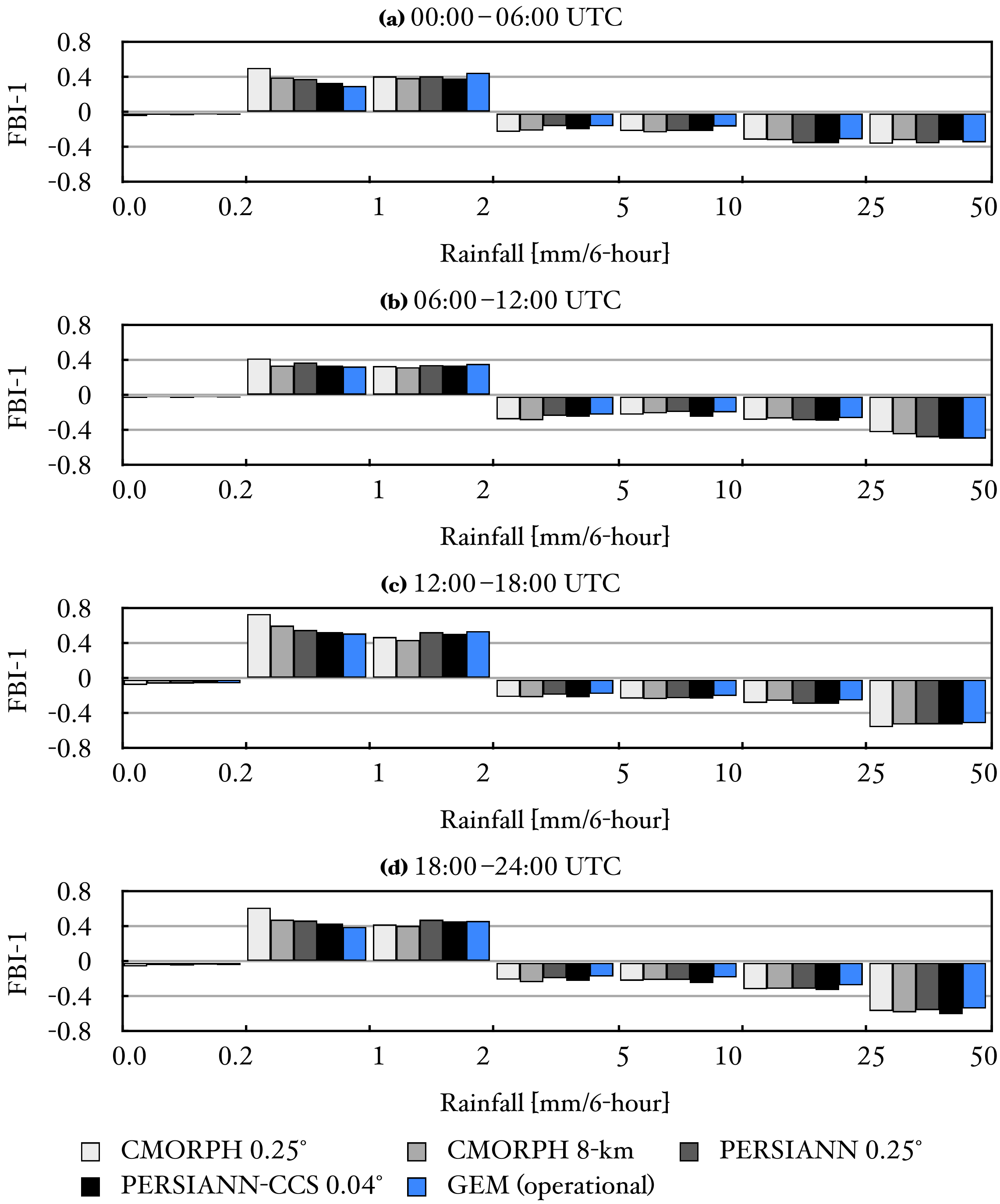


Figure E.10: Same as Figure E.8, but for eastern Canada.