Summer 2016

Prediction of North Atlantic tropical cyclone activity and rainfall

Beda Nidhi Luitel
University of Iowa

PREDICTION OF NORTH ATLANTIC TROPICAL CYCLONE ACTIVITY AND RAINFALL

By

Beda Nidhi Luitel

A thesis submitted in partial fulfillment
of the requirements for the Master of Science
degree in Civil and Environmental Engineering in the
Graduate College of
The University of Iowa

August 2016

Thesis Supervisor: Assistant Professor Gabriele Villarini

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Beda Nidhi Luitel

has been approved by the Examining Committee for
the thesis requirement for the Master of Science degree
in Civil and Environmental Engineering at the August 2016 graduation.

Thesis Committee: _____
Gabriele Villarini, Thesis Supervisor

_____
Allen Bradley

_____
Witold F Krajewski

ACKNOWLEDGEMENTS

I want to thank my advisor, Dr. Gabriele Villarini for his guidance, continuous encouragements and constructive suggestions throughout the course of this project as well as for providing me the opportunity to participate in this research at IIHR-Hydroscience & Engineering.

My sincere thanks to National Science Foundation (NSF) for providing for providing financial support under grant No. AGS-1262099, and Award NA14OAR4830101 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

I would like to thank my thesis committee members: Dr. Gabriele Villarini, Dr. Witold F Krajewski, Dr. Allen Bradley.

I want to thank my family for continuous support and love in all difficult situations. I want to thank Munir Nayak for his help and all my friends.

ABSTRACT

Among natural disasters affecting the United States, North Atlantic tropical cyclones (TCs) and hurricanes are responsible for the highest economic losses and are one of the main causes of fatalities. Although we cannot prevent these storms from occurring, skillful seasonal predictions of the North Atlantic TC activity and associated impacts can provide basic information critical to our improved preparedness. Unfortunately, it is not yet possible to predict heavy rainfall and flooding associated with these storms several months in advance, and the lead time is limited to few days at the most. On the other hand, overall North Atlantic TC activity can be potentially predicted with a six- to nine-month lead time.

This thesis focuses on the evaluation of the skill in predicting basin-wide North Atlantic TC activity with a long lead time and rainfall with a short lead time. For the seasonal forecast of TC activity, we develop statistical-dynamical forecasting systems for different quantities related to the frequency and intensity of North Atlantic TCs using only tropical Atlantic and tropical mean sea surface temperatures (SSTs) as covariates. Our results show that skillful predictions of North Atlantic TC activity are possible starting from November for a TC season that peaks in the August-October months.

The short term forecasting of rainfall associated with TC activity is based on five numerical weather prediction (NWP) models. Our analyses focused on 15 North Atlantic TCs that made landfall along the U.S. coast over the period of 2007-2012. The skill of the NWP models is quantified by visual examination of the distribution of the errors for the different lead-times, and numerical examination of the first three moments of the error

distribution. Based on our results, we conclude that the NWP models can provide skillful forecasts of TC rainfall with lead times up to 48 hours, without a consistently best or worst NWP model.

PUBLIC ABSTRACT

North Atlantic tropical cyclones (TCs) and hurricanes are responsible for a large number of fatalities and economic damage. Skilful seasonal predictions of the North Atlantic TC activity can provide basic information critical to our improved preparedness. First half of this study focuses on the development of statistical-dynamical seasonal forecasting systems for different quantities related to the frequency and intensity of North Atlantic TCs. TCs are associated with multiple hazards, including strong winds, storm surges, heavy rainfall and flooding. While the effects of winds and surge are mostly felt along the coastal areas near the landfall location, heavy rainfall and flooding are responsible for significant damage over much larger areas, even hundreds of kilometers from the coast. So a skillful forecast of rainfall associated with North Atlantic TCs has higher significant for emergency management services as well as in water resource management. In the second half of this study, we have performed the skill evaluation of five state-of-the-art numerical weather prediction (NWP) systems in forecasting rainfall from North Atlantic TCs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Motivation

North Atlantic tropical cyclones (TCs) represent a major catastrophic natural phenomenon affecting large areas of North and Central America. These storms generally occur during late spring into late fall every year, with peak activity in August-October. According to the National Climatic Data Center (NCDC), over the period 1980-2014 North Atlantic TCs were responsible for 31 events that caused losses in excess of $1 billion (http://www.ncdc.noaa.gov/billions/), corresponding to 23.3% of all the weather and climate disasters affecting the United States. An even more staggering number is associated with the economic losses: TCs were responsible for $417.9 billion (Consumer Price Index adjusted to 2011 dollars value), which correspond to 47.4% of all the losses caused by all the natural disasters responsible for $1+ billion combined (Smith and Katz 2013). Strong winds, storm surges, heavy rainfall and flooding associated with TCs are the major threats to life and infrastructure not only in coastal areas but miles inland as well (e.g., Elsner and Kara 1999; Blake et al. 2007; Changnon 2008; Lin et al. 2010; Villarini et al. 2011; Rappaport 2000, 2014; Lin et al. 2014; Villarini et al. 2014). According to the Spatial Hazard Events and Losses Database for the United States (SHELDUS), hurricanes and tropical storms were responsible for $267 billion losses for the period of 1960-2014 (see Figure1-1), which correspond to about 1/3 of all the losses over that period. More than 50% of the fatalities associated with TCs between 1970 and 2004 were caused by fresh water flooding (http://www.nws.noaa.gov/os/water/ahps/pdfs/InlandFloodBrochure7F.pdf\). Hurricane Ivan (2004) alone accounted for two-thirds of the total flood insurance payments

made by the federal government in that year, impacting 23 different states (Czajkowski et al. 2013). Over the period 1963-2012, Rappaport (2014) showed that almost 50% of U.S landfalling TCs have at least one fatality related to rain, while the storm surge was responsible for 49% of all the fatalities directly attributable to these storms. Blake et al. (2011) provide a summary of the deadliest and costliest TCs from 1851 to 2010. While many of the storms that caused the largest number of fatalities generally occurred in the earlier part of the twentieth century, the majority of the TCs responsible for the highest losses are concentrated in the last decade of the twentieth century and in the first decade of the twenty-first century. The staggering numbers represent a combination of different factors, related to both the hazard (i.e., TCs) and the increasing exposure (i.e., highly populated areas along the coast, with a large concentration of properties and wealth).

However, TCs are not just a threat to coastal areas, but affect much larger areas of the United States. In particular, while strong winds and storm surge mostly represent a hazard for the coastal areas (e.g., Czajkowski and Done 2014), heavy rainfall and flooding impact inland areas, hundreds of kilometers from the storm track (e.g., Villarini et al. 2011). Villarini et al. (2014) showed that U.S. landfalling TCs are responsible for major flooding over Florida, the U.S. East Coast from South Carolina to New Jersey and the northeastern United States; they also found that TCs can cause major flooding over the central United States, affecting in particular major urban areas including Chicago and Detroit. Although precipitation directly associated with TCs is less than 25% of the annual precipitation even in the most affected regions, the impacts in terms of extremes can be significant (e.g., Kunkel et al. 2010 ; Jiang and Zipser 2010; Barlow 2011).

Despite these negative socio-economic impacts, landfalling TCs have also been found to play a significant role as "drought busters" (Elsberry 2002; Kam et al. 2013; Maxwell et al. 2012; Maxwell et al. 2013). Torrential rainfall associated with TCs occasionally can have the effect of breaking a prolonged drought by recharging reservoirs and elevating soil moisture. In these situations TC rain events mitigate one environmental stressor, even as the potential for damage associated with the extreme rainfall, high-speed wind and ocean surge remain (Kam et al. 2013; Maxwell et al. 2013).

A better understanding and a skilful forecast of these storms is therefore necessary to help reduce the socio-economic consequences of these events and to provide basic information critical to improved preparedness.

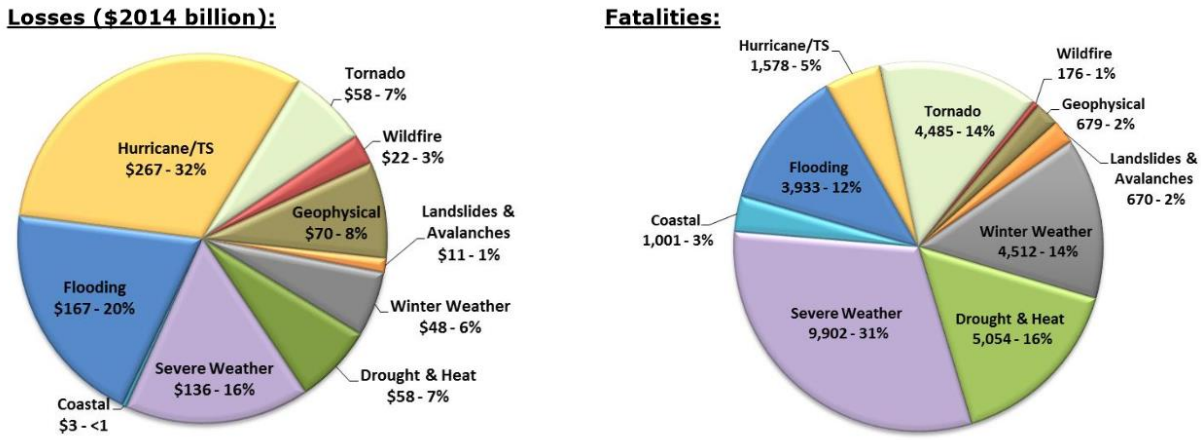

Figure 1-1: Monetary losses by hazard type (1960-2014) [Source: http://hvri.geog.sc.edu/SHELDUS/index.cfm?page=reports].

1.2 Objectives

Ideally, we would like to be able to provide skillful regional predictions of TC activity and associated impacts with a long lead time (Vecchi and Villarini 2014). Figure 1.2 highlights what the most hydrologically impactful forecasts would be and different lead times. We would want to be able to predict flooding (or at least rainfall) associated with these storms at a regional level with a long lead time, say months and years. However, this is still not possible if not at a much shorter lead time (on the order of hours and days). What we can potentially predict with a much longer lead time (i.e., from months to years; Gray 1984a, b; Gray et al. 1992; Lehmiller et al. 1997; Vitart and Stockdale 2001; Camargo et al. 2007; Camargo and Barnston 2009; Smith et al. 2010; Villarini et al. 2010; Villarini and Vecchi 2012, 2013; Klotzbach 2014; Murakami et al. 2016) is the basinwide activity. Therefore, while we would want to be in the top-right corner of Figure 1.2 (i.e., long-lead predictions of hydrologic impacts of TCs), the state of the science mostly allows us to move along the x- and y-axes: we can obtain long-lead predictions of a quantity that may not be very useful from an impact perspective (e.g., number of hurricanes in the North Atlantic during a given season) or we can have forecasts of rainfall associated with TCs but with a much shorter lead time.

Figure 1.2 summarizes the goals of this study (see black crosses).

1. To develop seasonal forecasting systems for the basinwide North Atlantic TC activity.

Basinwide TC activity can be summarized by four metrics related to the frequency, duration and intensity of these storms: number of tropical storms and hurricanes, power dissipation index (PDI) and accumulated cyclone energy (ACE) (details about each of these

metrics will be provided in Chapter 2.3). Here we will examine the possibility of predicting these four metrics with a long lead time (8-10 months before the beginning of the TC season). Our hypothesis is that skillful predictions with such a long lead time are possible by developing statistical-dynamical models in which the TC activity is tied to the state of the oceans and their intrinsically longer predictability. Recently, there have been some encouraging improvements towards accomplishing this task (e.g., Smith et al. 2010; Vecchi et al. 2013; Villarini and Vecchi 2013).

We will obtain predictors related to the North Atlantic TC activity from six GCMs of the North American Multi-Model Ensemble (NMME) project (Kirtman et al. 2014) with a lead time from one month to one year. Because of the availability of multiple GCMs, our second hypothesis is that ensemble averaging and merging will lead to an improved skill compared to predictions from individual GCMs, especially at longer lead times. For instance, Vecchi et al. (2011) found that the skill of the hurricane forecasting slightly increased when they merged covariates related to the temperature of the oceans forecasted from the Geophysical Fluid Dynamics Laboratory (GFDL–CM2.1) and the National Centers for Environmental Prediction's Climate Forecast System (NCEP–CFS). Threfore one of the important objectives of the first part of the thesis is to examine whether and for what lead times we are able to increase the skill of individual models by developing a multi-model ensemble (MME) forecasting system.

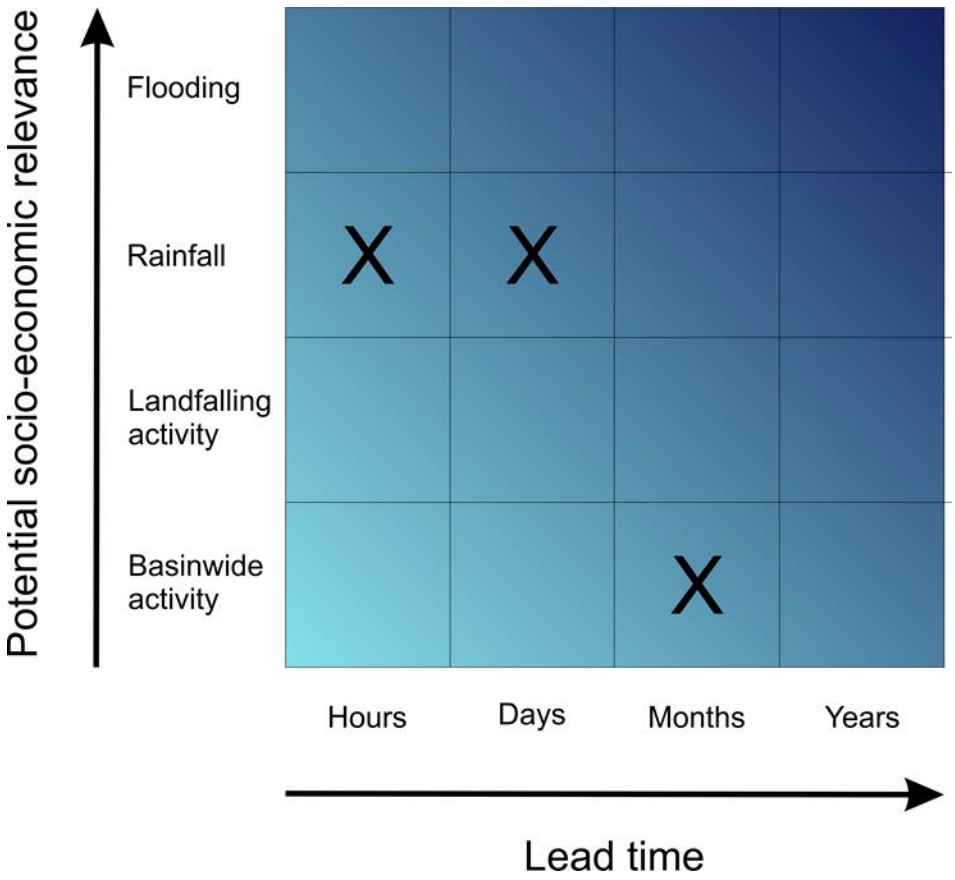

Figure 1-2: Research objective of this study vs. desired predictability. The black crosses indicate the quantities and lead times considered in this thesis.

2. To evaluate the skill of NWP models in forecasting TC rainfall.

NWP models provide forecasts of different quantities, including precipitation, with a lead time from few hours to up to two weeks. While the outputs of these models are routinely used in forecasting the track of TCs, much less is known about their capability in forecasting the rainfall associated with these storms. Moreover, how rapidly the performance of these models in forecasting TC rainfall deteriorates with increasing lead time is not known. Our goal is to evaluate the skill of NWP models in forecasting TC rainfall by quantifying their errors with respect to different "observational data."

In this thesis, Chapter 2 focuses on the development of forecasting systems for long-lead predictions of basinwide TC activity, while Chapter 3 deals with the evaluation of the skill of NWP models in short-term forecasting of TC rainfall. Chapter 4 summarizes the main conclusions of this work and provides the seasonal forecast of North Atlantic TC activity for the upcoming 2016 season.

CHAPTER 2: MULTI-MODEL ENSEMBLE FORECASTING OF NORTH ATLANTIC TROPICAL CYCLONE ACTIVITY

2.1. Introduction

The history of seasonal forecasting of the North Atlantic TC activity using statistical or dynamical models can be traced back to William Gray's work in the first half of the 1980s (Gray 1984a, b), while (Camargo et al. 2007) describe some of the developments in the seasonal forecasting methodology ever since. Currently, there are a number of institutions and research groups around the globe that routinely issue seasonal forecasts of North Atlantic TC activity, generally from April to August (e.g., Camargo and Barnston 2009; Klotzbach 2014; Vecchi et al. 2014; Camp et al. 2015). Issuing a seasonal forecast of TC activity 9-10 months before the start of a particular season is still challenging (Vecchi and Villarini 2014), but there has been recent progress towards accomplishing this task (e.g., Smith et al. 2010; Vecchi et al. 2013; Villarini and Vecchi 2013). Therefore, one of the research questions we address in this chapter is related to the evaluation of the forecast skill for different quantities related to TC activity with lead times starting from November of the year prior to the season to forecast; our focus will be on basin-wide number of tropical storms and hurricanes, power dissipation index (PDI) and accumulated cyclone energy (ACE) (see next section for a description of these quantities). We will accomplish this task by building hybrid statistical-dynamical models that use only tropical Atlantic and tropical mean SSTs to describe the variability exhibited by the observational record. As discussed later on, these predictors are selected because they reflect the importance of both

[1] Adapted from: Villarini, G., B. Luitel, G.A. Vecchi, and J. Ghosh, Multi-model ensemble forecasting of North Atlantic tropical cyclone activity, submitted to Climate Dynamics, 2016.

local and non-local effects on the genesis and development of TCs in the North Atlantic basin (e.g., Latif et al. 2007; Swanson 2007; Vecchi et al. 2008, 2011, 2013; Villarini et al. 2010; Zhao et al. 2009, 2010; Ramsay and Sobel 2011; Villarini and Vecchi 2012, 2013).

Another important research question that we are addressing in this chapter is related to the possibility of increasing the skill of individual models by merging them. As mentioned in Chapter 1, Vecchi et al. (2011) found that the skill in forecasting hurricane activity slightly increased when they merged covariates (SSTs) forecasted from the GFDL–CM2.1 and NCEP–CFS. Here we examine whether and for what lead times we are able to increase the skill of the individual models by developing a MME forecasting system. MME forecasting is a relatively new line of research and believed to provide potentially large improvements in weather and climate forecasting in particular for longer lead times (e.g., Palmer et al. 2005; Yun et al. 2005). Studies showed that multi-model techniques can help in reducing model errors, while the ensemble technique can help in normalizing uncertainties associated with initial conditions (e.g., Richardson 2000; Hagedorn et al. 2005; Yun et al. 2005; Zhu 2005; Tebaldi and Knutti 2007). Equal-weight average, weighted average and Bayesian methods are some of the more widely used techniques to combine individual models in the process of building MME models (Robertson et al. 2004). Here we consider different MME techniques as a function of lead times and TC quantity, highlighting the improvements we can achieve in forecasting North Atlantic TC activity.

This chapter is organized as follows. Section 2.2 will describe the data we use, followed by the methodology in Section 2.3; the results are in Section 2.4 while Section 2.5 will summarize this chapter.

2.2 Data

We consider four quantities to summarize the North Atlantic TC activity: hurricane frequency, frequency of tropical storms lasting more than 2 days, PDI and ACE. Hurricane and tropical storm frequencies tell us how active a season is, whereas PDI and ACE incorporate storm duration, frequency and intensity over an entire TC season (Bister and Emanuel 1998; Bell and Chelliah 2006; Emanuel 2007; Villarini and Vecchi 2012, 2013). We use the National Oceanic and Atmospheric Administration's (NOAA) National Hurricane Center's best-track database (HURDAT2) to obtain hurricane and tropical storm counts, and use HURDAT-2's maximum wind speed to compute ACE and PDI (Landsea and Franklin 2013). ACE and PDI are calculated similar to Villarini and Vecchi (2012), where ACE (PDI) is computed by taking the sum of the square (third power) of the wind speed exceeding 17 m/s during the lifetime of each storm during each TC season and the value is adjusted by a factor of 10^9 (10^{11}). We use observed data from 1958 onward to limit the effect of inhomogeneity in the data in the earlier part of the TC record.

As in previous dynamical and statistical modeling studies (e.g., Camargo and Barnston 2009; Zhao et al. 2009, 2010; Villarini et al. 2010; Vecchi et al. 2011, 2013; Villarini and Vecchi 2012), we use tropical Atlantic and tropical mean SST as our predictors. As reference SST data, we use monthly SST, observations from Met Office Hadley Centre (Rayner et al. 2003). From the Hadley Center's gridded datasets, we compute time-series of SST, averaged over the North Atlantic TC main development region (10°N - 25°N and 80°W- 20°W; SST_{Atl} hereafter) and over the tropical belt (30°S -30°N around the globe; SST_{Trop} hereafter). Similar to Villarini and Vecchi (2013), seasonal SST anomalies are

computed with respect to the climatology for 1982-2005, averaged over the peak hurricane period of August to October.

Table 2-1: Summary of GCMs from the NMME project used in this study.

Model	Version	Acronyms used in this study	Lead times (months)	Number of ensemble available
Geophysical Fluid Dynamics Laboratory	GFDL-CM2.1	GFDL	12	10
	GFDL-CM2.5-A06	GFDL_A06	12	12
	GFDL-CM2.5-B01	GFDL_B01	12	12
NASA Global Modeling and Assimilation Office	NASA-GMAO-062012	NASA	9	11
The Center for Ocean-Land-Atmosphere Studies–Rosenstiel School of Marine and Atmospheric Science	COLA-RSMAS-CCSM3	COLA	12	6
Canadian Centre for Climate Modeling and Analysis	CCCma – CanCM4	CMC2	12	10

In developing our forecasting system, we use SST forecasts from a set of retrospective predictions by six experimental seasonal-to-interannual prediction systems from the NMME (Kirtman et al. 2014). Table 2-1 summarizes the NMME models we considered, together with the acronyms we use to refer to them in the text. Rather than using each member from each model, we only focus on the ensemble mean of the different members. The skill of the GCMs in forecasting SSTs with different lead times is evaluated using different statistical metrics quantifying the degree of similarity [Pearson correlation coefficient (Corr. Coeff.) and Spearman correlation coefficient] or dissimilarity [the root

mean square errors (RMSE) and the mean absolute errors (MAE)] between observations and forecasts.

2.3 Methodology

We use Poisson regression to forecast the seasonal frequency of North Atlantic hurricane and tropical storms (e.g., Elsner and Schmertmann 1993; Elsner and Jagger 2006; Villarini et al. 2010; Vecchi et al. 2011). The expected value (λ) of the Poisson regression model is a function of two explanatory variables, SST_{Atl} and SST_{Trop}; increasing values of SST_{Atl} would be associated with an increase in storm activity, while increasing values of SST_{Trop} would lead to an overall decrease in TC activity. The model selection and the choice of these two predictors for tropical storms are based on Villarini et al. (2010). The λ parameter of the Poisson regression model is linearly related to the predictors via a logarithmic link function as follows:

$$\lambda_q = \exp\left(\beta_0 + \beta_{Atl} \cdot SST_{Atl} + \beta_{Trop} \cdot SST_{Trop}\right) \tag{2.1}$$

where the subscript q refers to the different TC quantity (subscript "H" for hurricanes and "TS" for tropical storms).

For hurricane counts (λ_H), we use the same β coefficients as in Vecchi et al. (2011) (β_0 = 1.707, β_{Atl} = 1.388 and β_{Trop} = -1.521). Vecchi et al. (2011) trained the statistical model not only on observations but also on dynamical model outputs across a broad range of climate states.

In forecasting the number of tropical storms, we follow the retroactive validation method discussed in Mason and Baddour (2008). More specifically, we estimate the

parameters for λ_{TS} in Equation 2.1 based on the observations available over the 1958-1981 period. We then use the Atlantic and tropical SST forecasts to obtain the forecast distribution for 1982. We then re-estimate the parameters including the 1982 data, and issue the forecast for 1983. We continue in this way across all the available years. A time series of the model coefficients for the Poisson parameter λ_{TS} [Eq. (2.1)] for the tropical storms is provided in Figure 2-1 (panels a and b). The time series reveals that the value of model's intercept (β_0) increased around 1995 and remained almost constant at the higher value after that, consistently with Elsner et al. (2004).

For the seasonally integrated ACE and PDI indices, we follow the approach described in Villarini and Vecchi (2012, 2013). They used a gamma distribution which has two parameters, μ and σ to describe the year-to-year variability in normalized PDI and ACE. Based on the parameterization of the gamma distribution in the Generalized Additive Model in Location, Scale, and Shape (Rigby and Stasinopoulos 2005, 2007), the expected value is equal to the location parameter μ, and the variance is equal to $\mu^2\sigma^2$. Similar to Villarini and Vecchi (2012, 2013), we model the μ parameter as a linear function of SST_{Atl} and SST_{Trop} via a logarithmic link function:

$$\mu = \exp(\gamma_0 + \gamma_{Atl}.SST_{Atl} + \gamma_{Trop}.SST_{Trop}) \quad (2.2)$$

while the σ parameter is kept constant. A time series of the coefficients for the gamma model is presented in Figure 2-1 (panels c-f). Similar to Elsner et al. (2004) and Villarini and Vecchi (2013), there appears to be a step change around 1995. It is also interesting to notice that there appears to be another potential shift starting from around 2010, according to which we may be headed towards a quieter period in terms of TC activity (this is present in the time series of the coefficients for both the Poisson and gamma models). While it is

too early to make conclusive statements because of the limited number of years, this is something that it is worth taking into account as we move forward with the examination of the North Atlantic TC record.

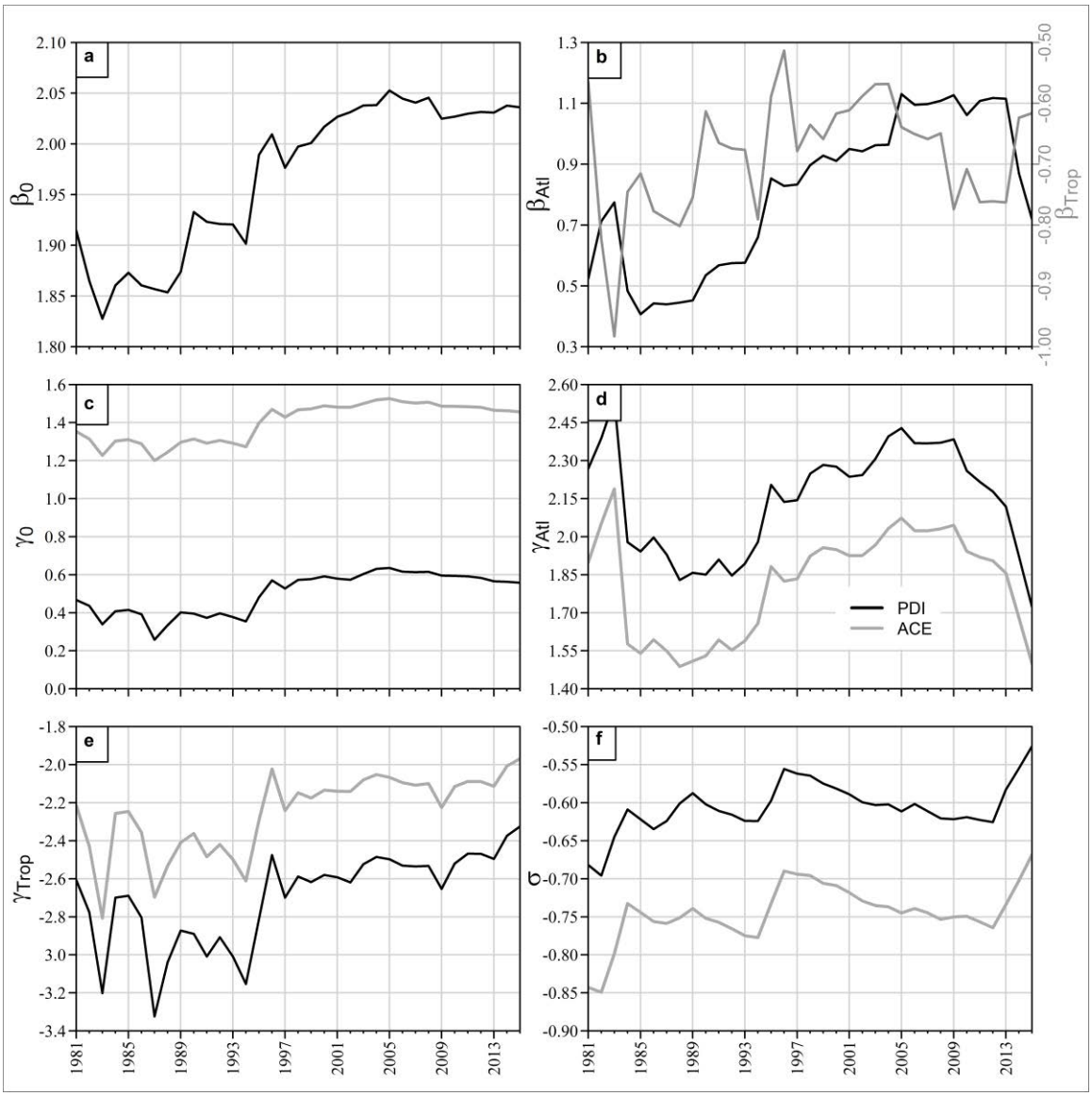

Figure 2-1: Time series of the model coefficients for the model parameter shown in Eq. (2.1) and Eq. (2.2). Top panels (a and b) refer to the λ parameter [Eq. (2.1)] for tropical storm forecast. Middle and bottom panels (c-f) describe the location parameter μ [Eq. (2.2)] and scale parameter σ over the period 1981–2015 for ACE (gray line) and PDI (black line).

As discussed in Section 2.2, we have SST forecasts from six GCMs. Figure 2-2 shows the skill of each individual prediction system and their average in forecasting seasonal SST_{Atl} (left panels) and SST_{Trop} (right panels). We evaluate the skill using different metrics and for different lead times, from August-initialized (i.e., the shortest lead time) to November of previous year-initialized forecasts (i.e., the longest lead time). Overall, these models are very skillful in forecasting the tropical mean and tropical Atlantic SSTs; the correlation coefficient between forecasts and observation is very high (larger than 0.8), with small values of RMSE and MAE, in particular at the shorter lead times. The forecast skill for SST_{Trop} tends to decrease for increasing lead times faster than the forecast of SST_{Atl}, which tends to preserve the skill observed at the shortest lead time longer. The skill for the majority of the models is comparable, with the exception of COLA (red line), which shows the worst performance among the six GCMs for all lead times. Because not a single model exhibits consistently higher skill for all lead times and SST quantities, a weighted model, where weights for each prediction system are computed as a function RMSE for each lead times, might be helpful in optimizing the overall skill.

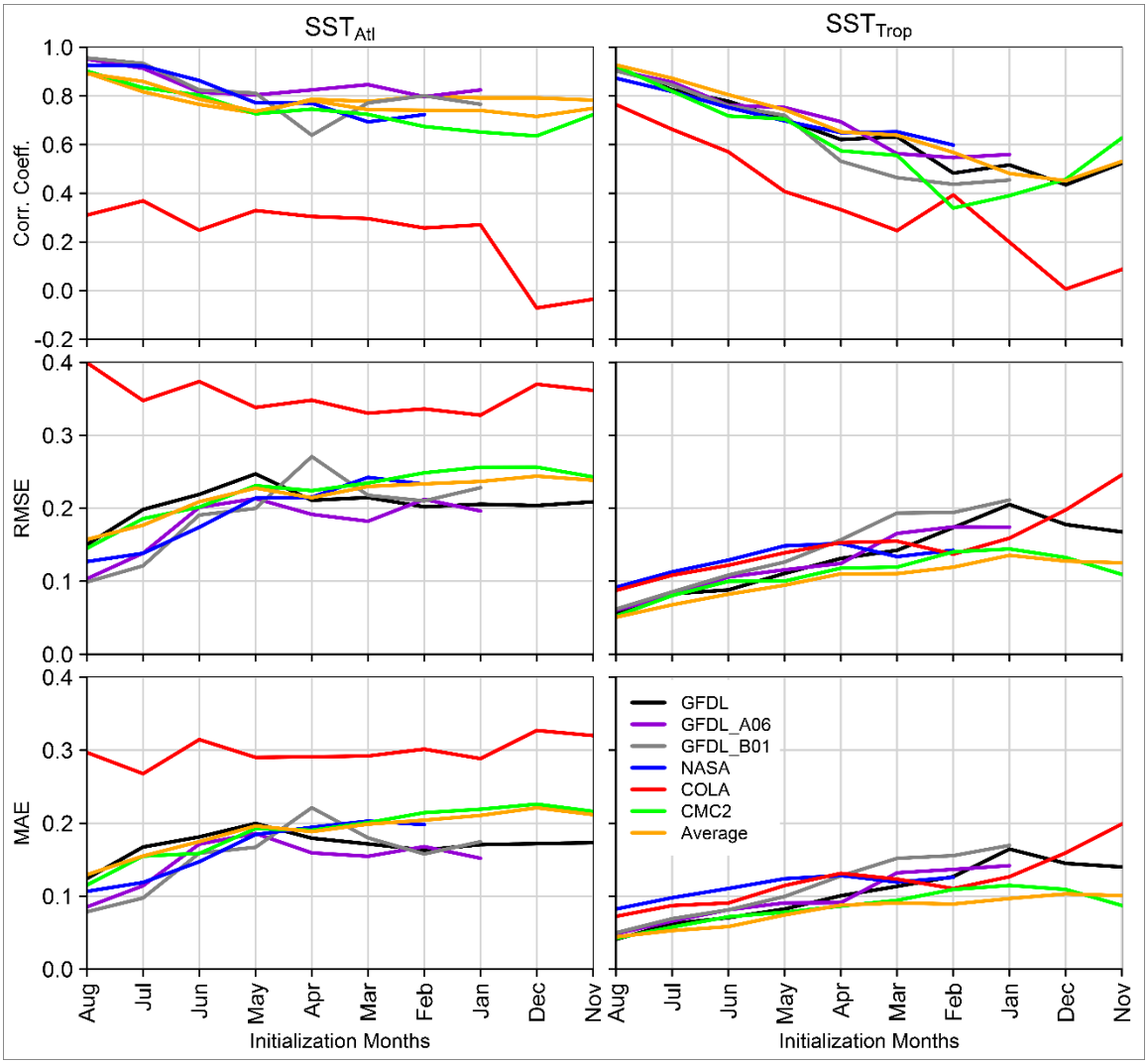

Figure 2-2: Skill evaluation of each GCM in forecasting SST_{Atl} (right panels) and SST_{Trop} (left panels) compared to SSTs from Met Office Hadley Centre. Pearson correlation coefficient (top row), RMSE (middle row) and MAE (bottom row) are the three metrics used in the evaluation of seasonal SST anomalies for different lead times (i.e., August being the shortest lead time). The black lines represent the average of the different GCMs' forecasts.

There are different ways to combine individual models to build MME [refer to Robertson et al. (2004) for details]. The three different approaches we use to merge the forecasts from the individual GCMs are as follows:

1. Equally weighted (arithmetic) average method, where predictors from each GCM are given equal weights.
2. Weighted average method, where predictors from each prediction system are weighted based on their RMSE with respect to the reference SST anomalies. For example let us consider SST_{Atl} and one of the initialization months; the RMSE for the i^{th} GCM is computed as:

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{n}(F_{i,j} - O_j)^2}{n}} \tag{2.3}$$

where n is the number of years used for the forecast; $F_{i,j}$ and O_j are the forecast and observed SST_{Atl} for the j^{th} year and the i^{th} GCM.

Using RMSE, the weight for each model is computes as:

$$w_i = \frac{\frac{1}{RMSE_i}}{\sum_{i=1}^{k}\frac{1}{RMSE_i}} \tag{2.4}$$

where i ranges from 1 to the number of available GCM forecasts for that year.

Then the weighted SST_{Atl} value for each year is computed as:

$$(SST_{Atl})_j = \sum_{i=1}^{k} w_i F_{i,j} \tag{2.5}$$

where j = 1 to n years and i = 1 to k GCMs.

3. Weighted average to build the empirical distribution of the quantity of interest, where the weights for each GCM are given based on the RMSE between the observed values of the TC metric of interest and the λ parameter of the Poisson distribution (for hurricanes and tropical storms) or the μ parameter of the gamma distribution (for PDI

and ACE). For instance, for PDI and one of the initialization months, the RMSE for the i^{th} GCM is computed as:

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{n}(\mu_{i,j} - O_j)^2}{n}} \quad (2.6)$$

where $\mu_{i,j}$ and O_j are the expected value from the gamma regression and observed PDI for the j^{th} year and the i^{th} prediction system, respectively; the weights are computed similar to the weights for SSTs explained above (see equation 2.4).

Once the weights are computed, we build the forecast distributions for each of the four metrics (hurricane and tropical storm counts, ACE and PDI) for every year based on Monte Carlo simulations, giving weights to each GCM depending on how well it was able to reproduce the observational record. We termed this weighted MME model as "Hierarchical model."

Figure 2-3 summarizes all the weights for each lead time using the different weighting schemes described above. Weights in Figure 2-3 reflect the skills of each GCM based on the RMSE (Figure 2-2, middle row). To take into account the relatively poor performance of the COLA model with respect to the other GCMs, we also consider weighting schemes excluding COLA.

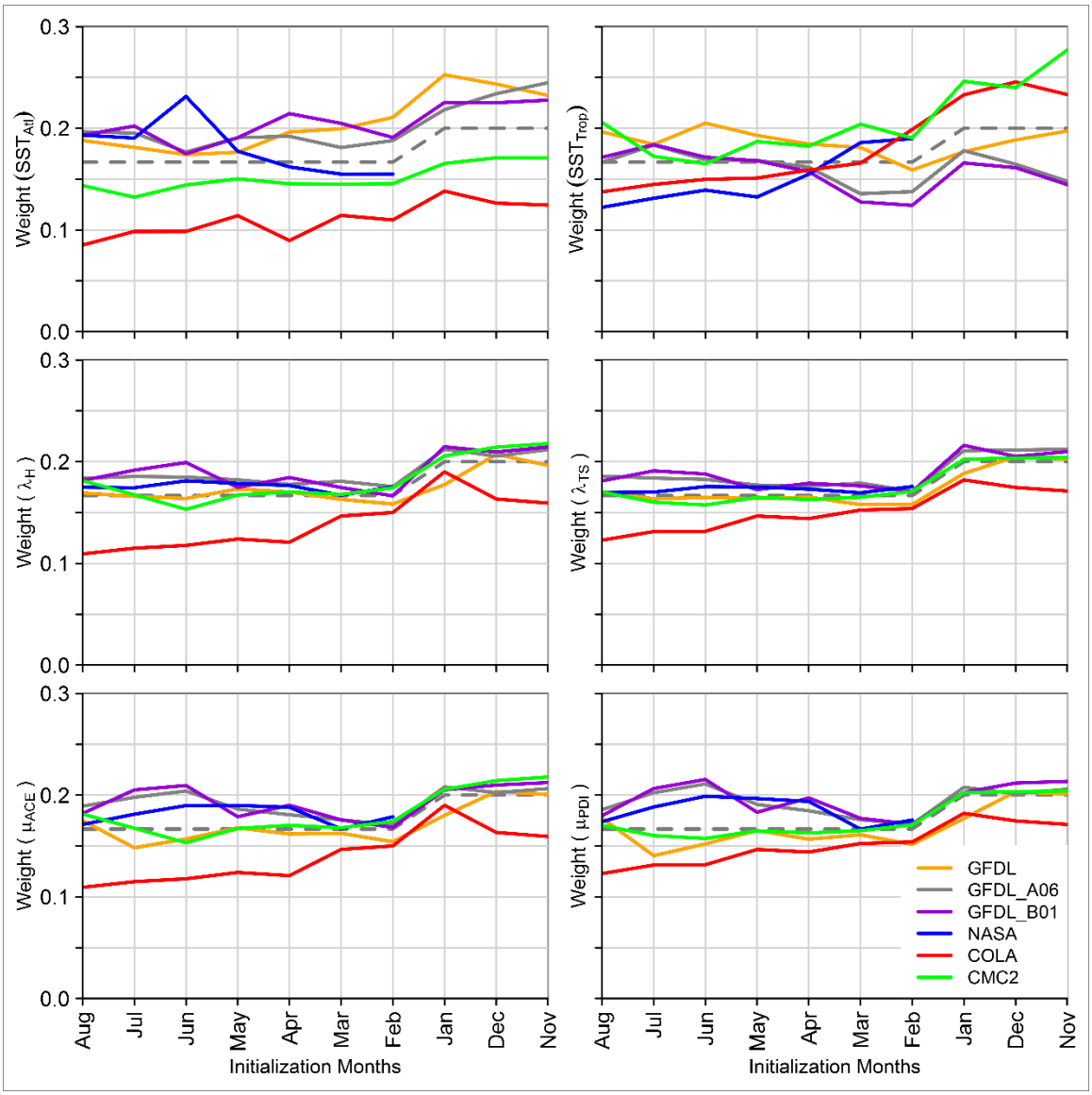

Figure 2-3: Top panel: Weights for each predictor with respect to RMSE between the predicted SST anomalies from each GCM and the reference SST anomalies. Middle and bottom panels: weights with respect to the λ parameter (for hurricanes and tropical storms) and the μ parameter (for ACE and PDI) for the four different TC metrics. The dotted lines represent equal weights (1/6 for August to February initialization and 1/5 after that). The sudden jump in weights for the initializations months of January, December and November is because NASA has only nine-month lead times (see Table 2-1), so the proportion of the weights for the remaining models increases.

We quantify the skill of these forecasts using both deterministic and probabilistic verification metrics. We use the median of the Poisson or gamma distributions as our

reference forecast value. To quantify the accuracy of the forecasts, we use the Spearman correlation coefficient between observed TC metric and the corresponding median forecast. Because of the probabilistic nature of our forecasting system, we use the probabilistic verification approach described in Laio and Tamea (2007) for PDI and ACE (continuous variables). We compare the distribution of $z_i = P_i(x_i)$ [x_i is the observed value for the i^{th} year, and $P(\cdot)$ is the forecast distribution] to a standard uniform distribution using the Kolmogorov-Smirnov test. A good forecast system should not exhibit too much or too little dispersion, with the uniform distribution being the target distribution (see also Wilks (2011)).

2.4 Results

We use the Poisson regression model discussed in the previous section to perform the retrospective seasonal forecasting of hurricane and tropical storm frequencies for the period of 1982-2015. Figures 2-4 and 2-5 show the retrospective forecasts for different initialization months (the shortest lead time is initialized in the month of August and the longest lead time is initialized in the month of November of the previous year) using the weighted model in which the weights are computed based on how well the GCMs can reproduce tropical Atlantic and tropical mean SSTs. These simple and parsimonious Poisson regression models are able to capture the interannual variability in seasonal hurricane and tropical storm counts very well. Even as early as November of the previous year, most of the observations are within the 5^{th} and 95^{th} percentiles (90% confidence interval) and as the lead time decreases, the median forecast tends to follow the observed values more closely.

For ACE and PDI, we use the gamma regression model to perform retrospective seasonal forecasting over the period of 1982-2015. Similar to what observed for the number of tropical storms and hurricanes, Figures 2-6 and 2-7 highlight the ability of these models in capturing year-to-year variability in these two TC quantities. The overall visual agreement increases as we reduce the lead time. Moreover, there is much more inter-annual variability at the shorter lead times. The time series in Figures 2-4-2-7 exhibit a larger activity post-1995, and it is unclear whether the current relatively quiet few last years are a possible indication that we are currently in another regime shift towards reduced TC activity, or whether it is just part of the year-to-year variability (see also results in Figure 2-1 and relative discussion). Complete results for all the individual GCMs and MME schemes are provided in the Appendix A, (Figures A-1 - A-10 for the number of hurricanes, Figures A-11 - A-20 for the number of tropical storms, Figures A-21 - A-30 for ACE and Figures A-31 - A-40 for PDI). Overall, these results indicate that it is possible to obtain skillful forecasts of North Atlantic TC activity several months prior to the TC season.

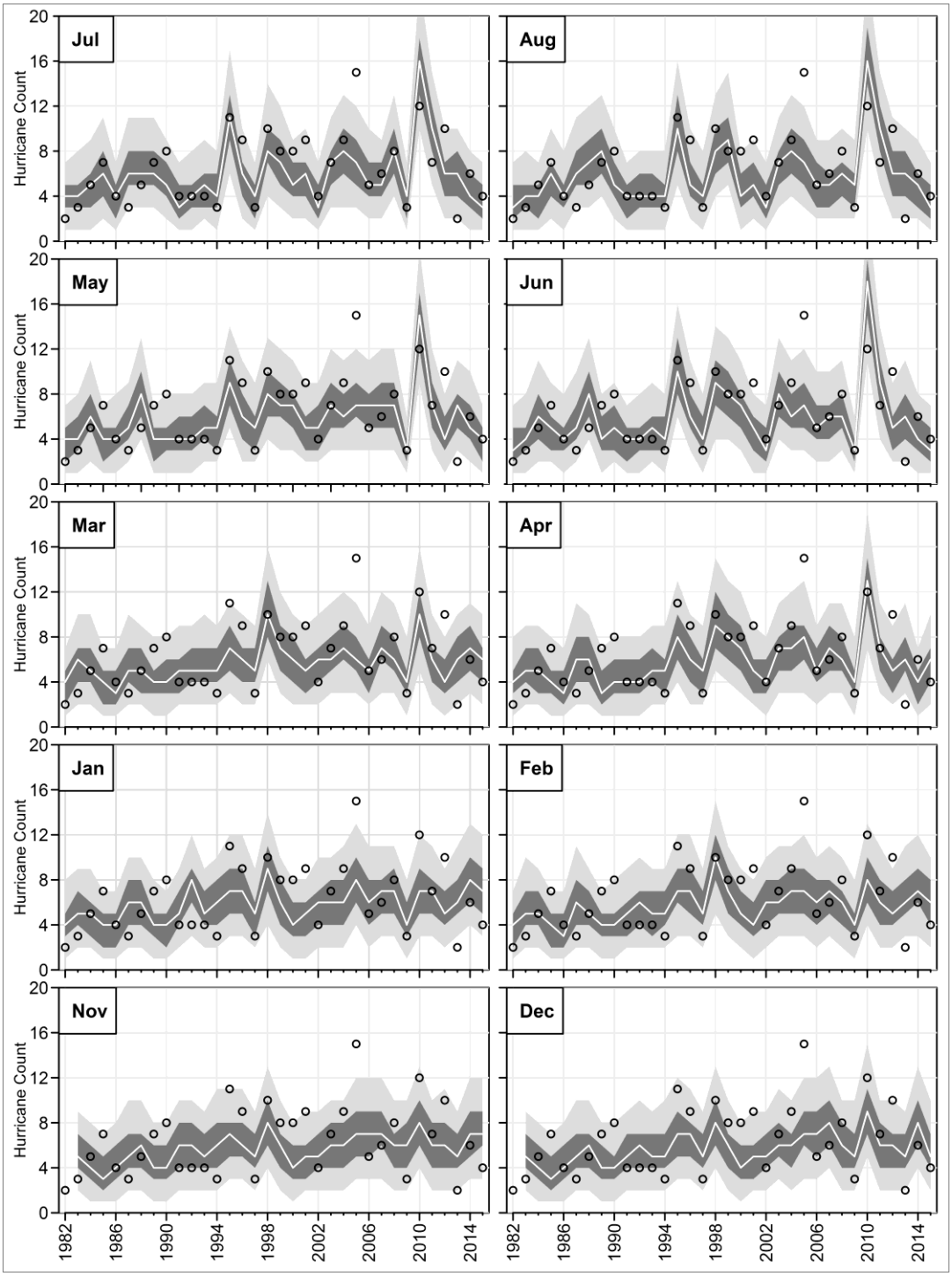

Figure 2-4: Retrospective seasonal forecast of hurricane using the model in which the two SST predictors are weighted, initialized from November of the previous year to August of the same year. In each panel, the circles are the observed values, the white line is the median (50th percentile). The outer light gray area represents the region between the 5th and 95th percentiles, while the middle dark gray area represents the region between the 25th and 75th percentiles.

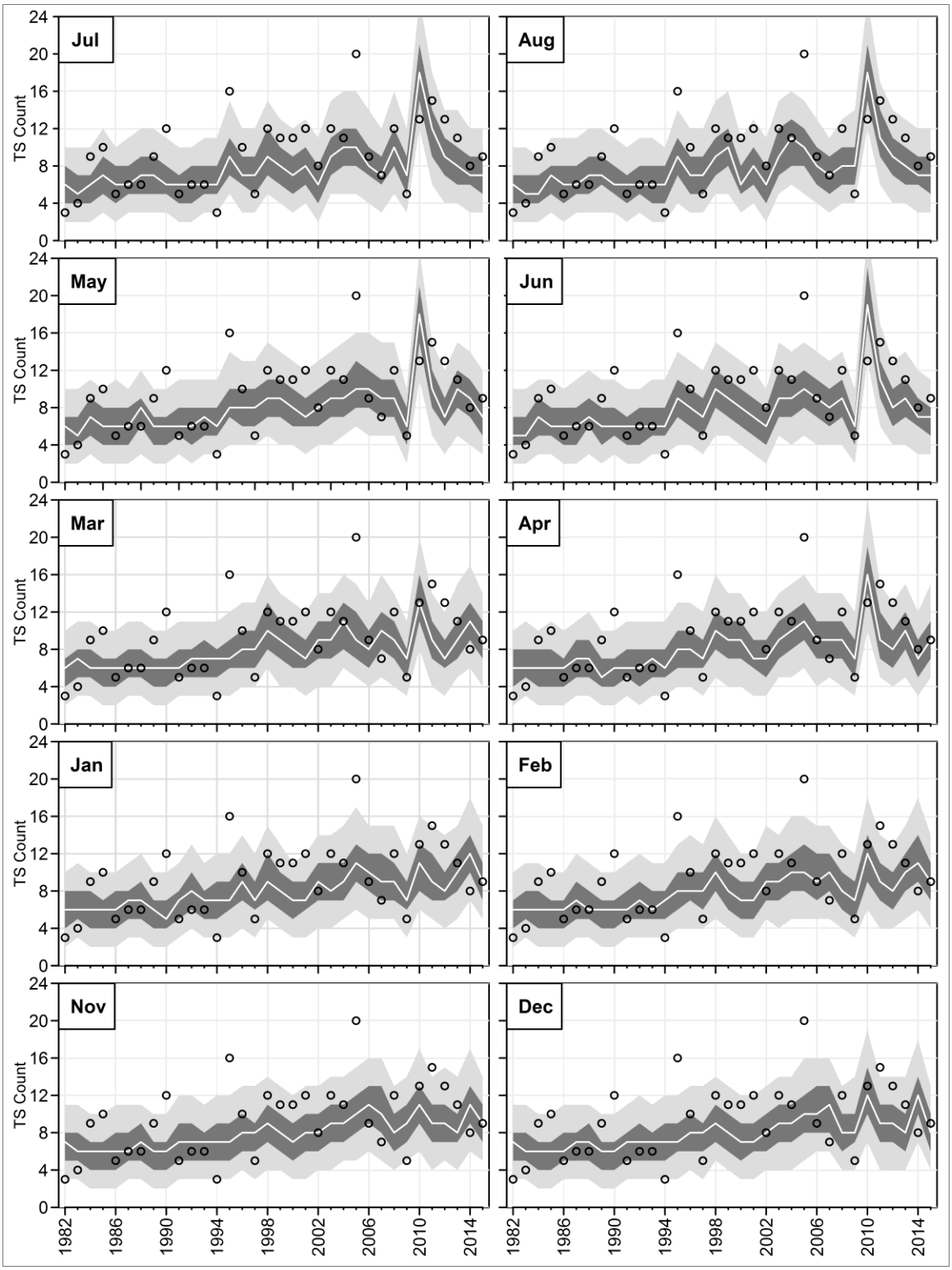

Figure 2-5: Same as Figure 2-4, but for tropical storms.

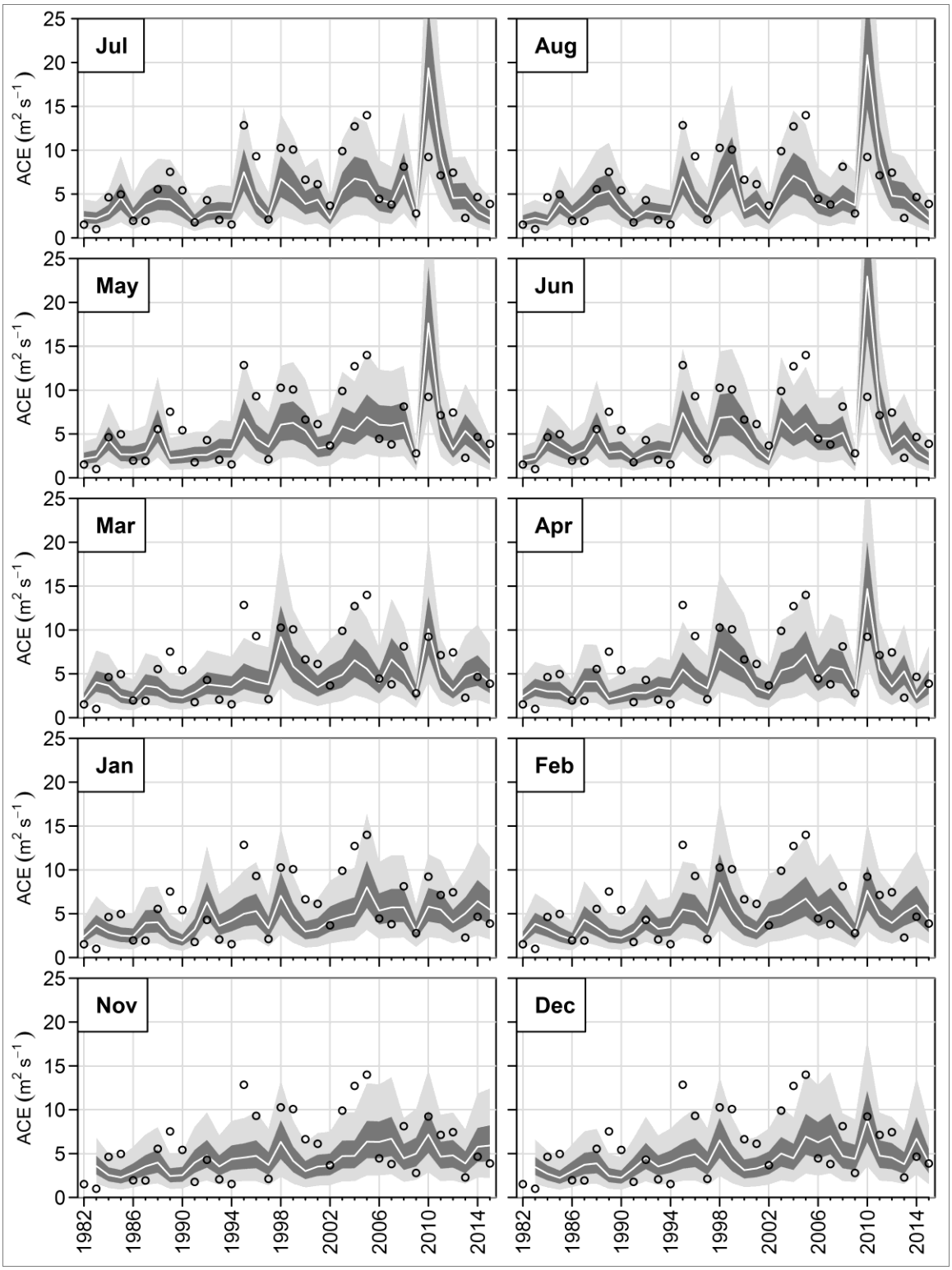

Figure 2-6: Same as Figure 2-4, but for ACE.

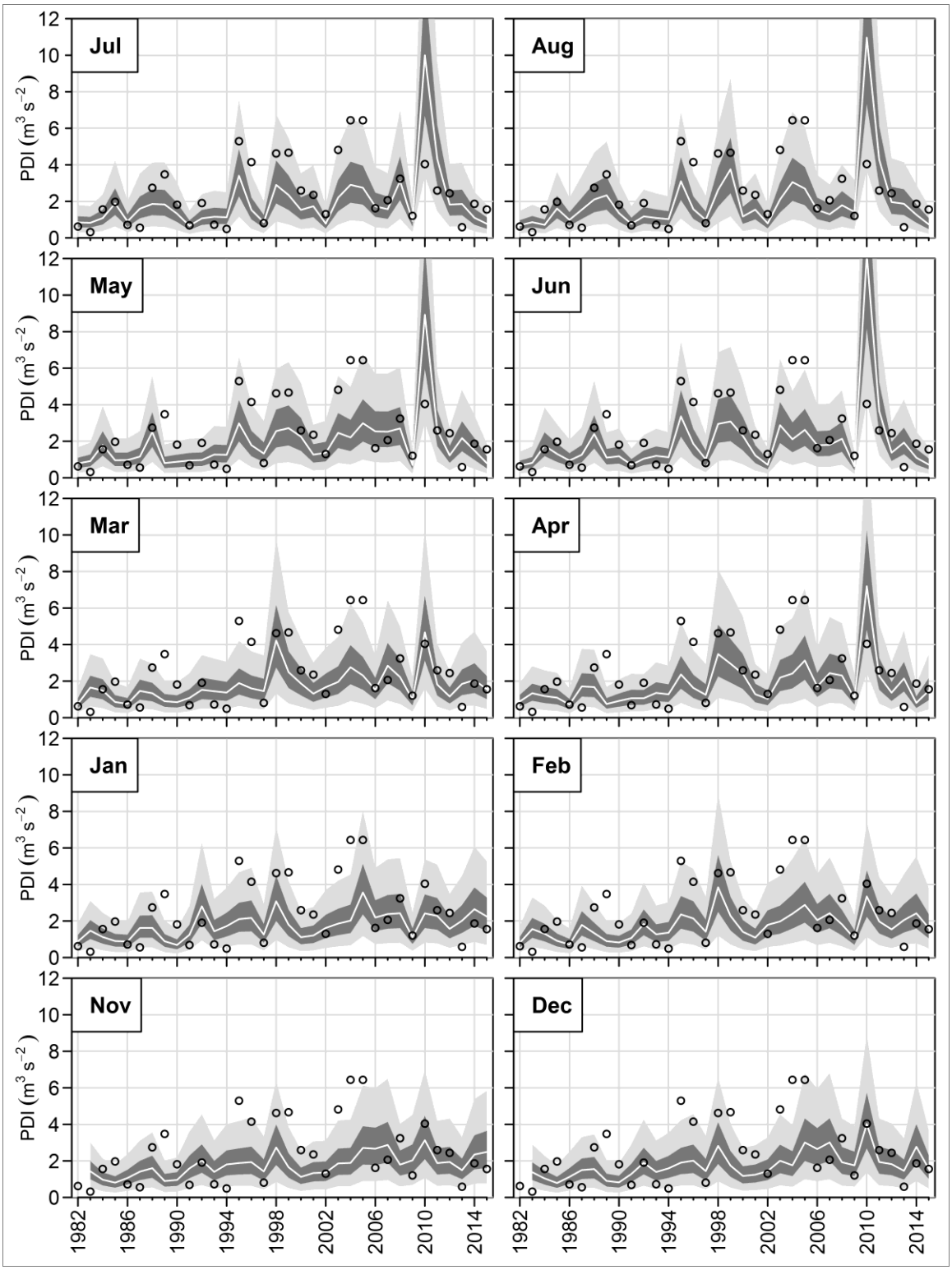

Figure 2-7: Same as Figure 2-4, but for PDI.

Because one of the main goals of this work is to compare the results obtained using individual models and the merged MME models, we used Spearman correlation coefficient to quantify the skill in retrospective forecast for all the individual and weighted models. Figure 2-8 shows the Spearman correlation coefficients between the 50th percentiles (median) of the forecasted values and the observed values for different initialization months for each of the individual prediction systems and MME models in forecasting all four metrics (hurricane and tropical storm counts, ACE and PDI). Here the models from M01 to M06 represent the forecasts based on the individual GCMs described in Table 2-1, models from M07 to M11 represent different MME schemes described in Section 2.3, while M12 in the top row of each panel represents five–year persistence as a measure of ‘‘null skill’’ as recommended in World Meteorological Organization (2008). When we compare the skills of individual models (M01–M06) with MME models (M07–M11), we found that MME models are less dependent on lead times. In other words, the skills are more consistent for all lead times in the case of MME models as compared to the individual models even if we consider equally weighted model. The skill values we obtain are large, with correlation coefficients larger than 0.7-0.8 for several lead times. We also have relatively large correlation values for long lead times, in particular for the weighted models. The hierarchical model is the one that shows the highest performance across a large number of lead times. For instance, the November- or December-initialized forecasts have correlation coefficients in excess of 0.5 (see Figures 2-8 and 2-9). Unlike Figure 3 in Vecchi et al. (2011) and Figure 5 in Villarini and Vecchi (2013) where there was a peak in skill in the March-April initialization, we find a gradual increment in forecasting skill as

the lead time decreases. This can be credited to the improved ability of the most recent versions of the dynamical prediction systems, as well as to the merging technique.

To further highlight the improvements associated with margining forecasts from different GCMs, in Figure 2-9 we show the differences in Spearman correlation coefficient between the equally weighted model (M07) and the other models for each of the TC quantities and for different lead times. Overall, it is clear that weighting the models leads to improved skill over the individual GCMs, in particular at longer lead times. Moreover, the hierarchical model formulation (M11) leads to improved results over the equal-weights model, in particular for the November- and December-initialized forecasts.

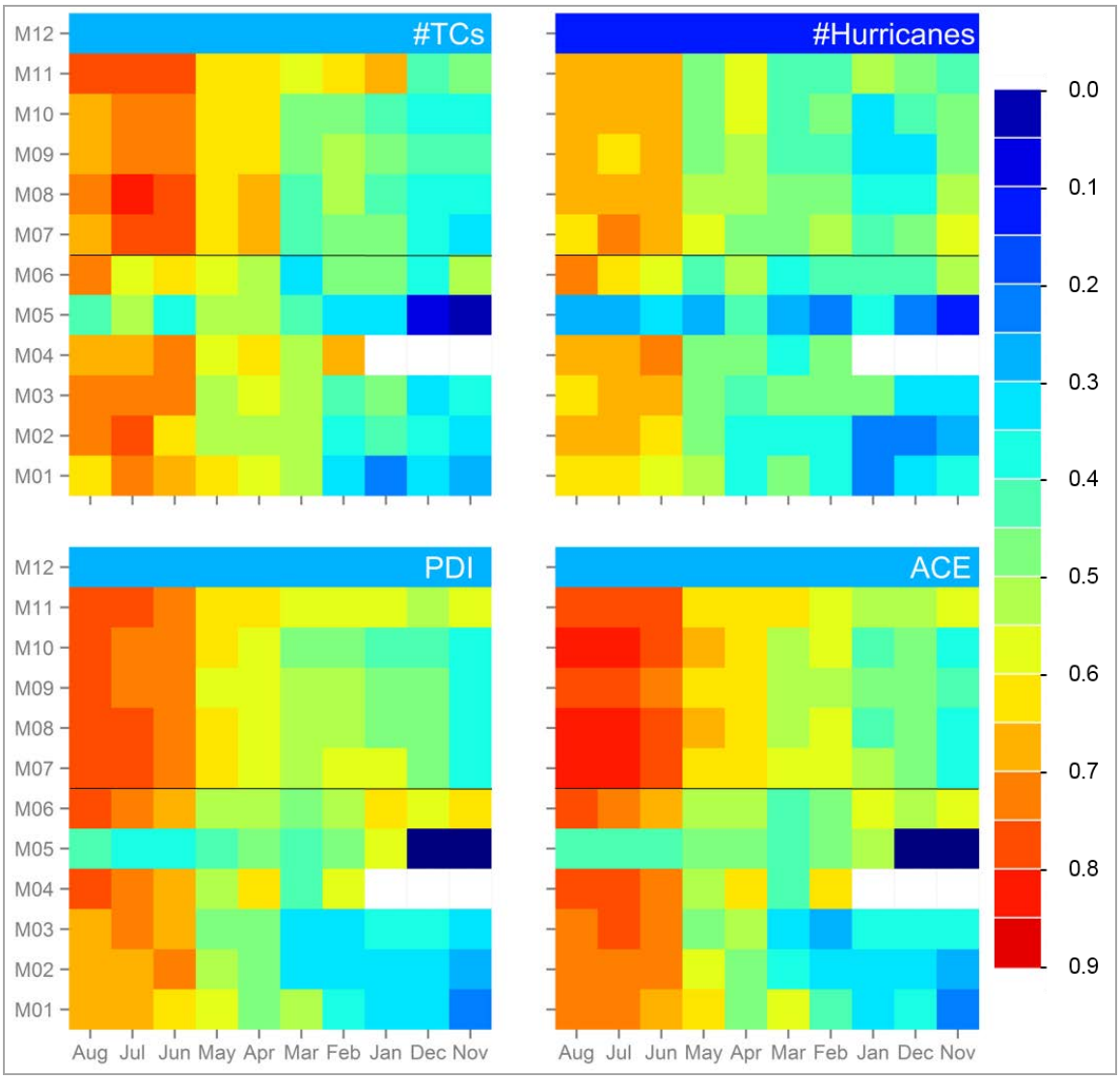

Figure 2-8: Spearman correlation coefficient between the 50th percentiles (median) of the forecasted and the observed values for different initialization months. Here M01= GFDL, M02 = GFDL-B01, M03 = GFDL-A06, M04 = NASA, M05 = COLA, M06 = CMC2, M07 = equally weighted model, M08 = weighted model, M09 = equally weighted model without COLA, M10 = weighted model without COLA, M11 = hierarchical model and M12 = Five-year persistence.

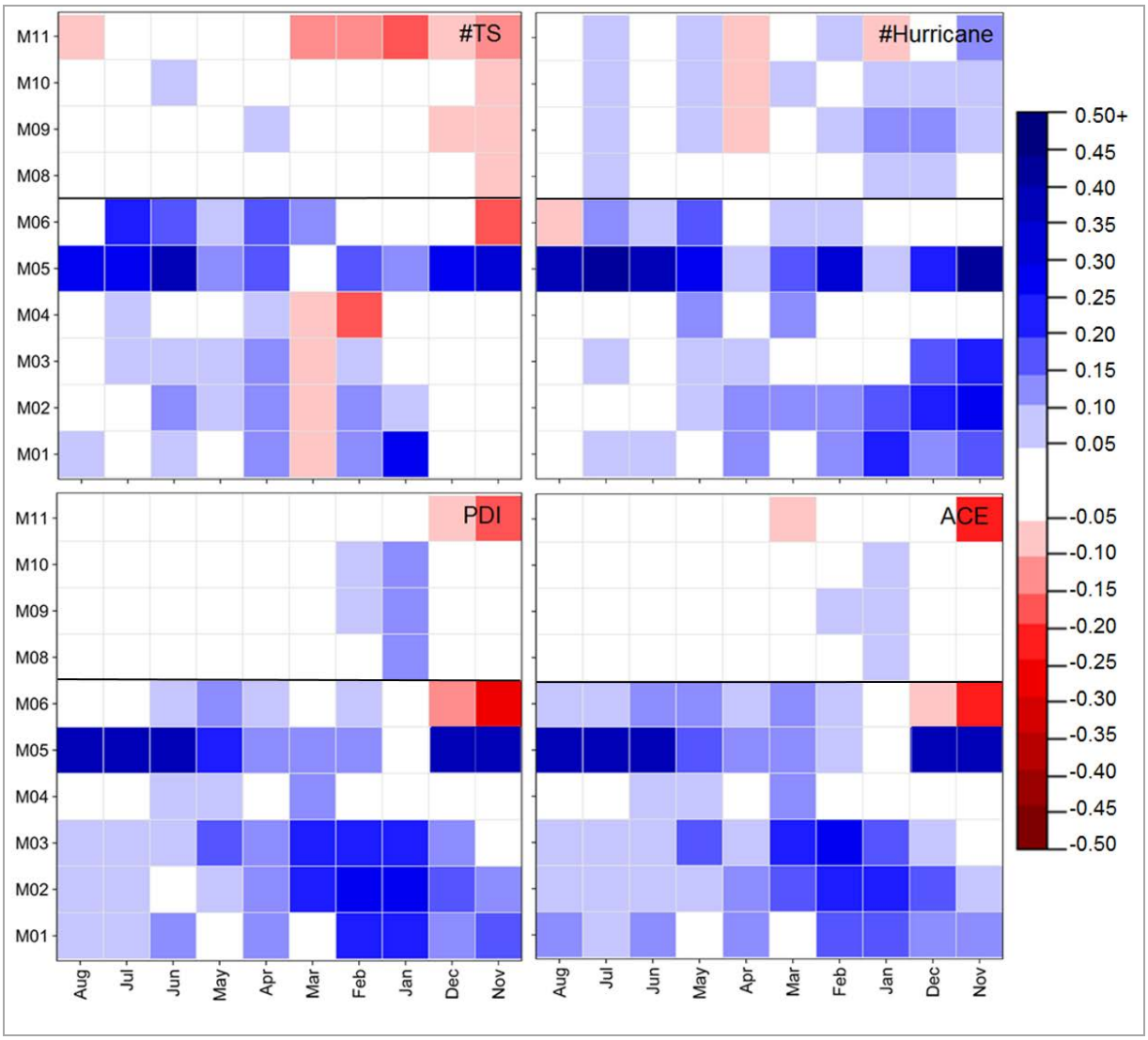

Figure 2-9: Differences in the rank correlation with respect to the equally weighted model (M07). M01-M06 (below the black line) represent individual GCMs, while M8-M11 represent MME models similar to Figure 2-8. Positive (negative) values refer to a better (worse) performance by the equally weighted model M07.

Finally we have performed probabilistic verification of the forecasts as described in Laio and Tamea (2007). The results in Figure 2-10 highlight that the best skill in terms of probabilistic verification is at the shorter lead times and for the hierarchical MME.

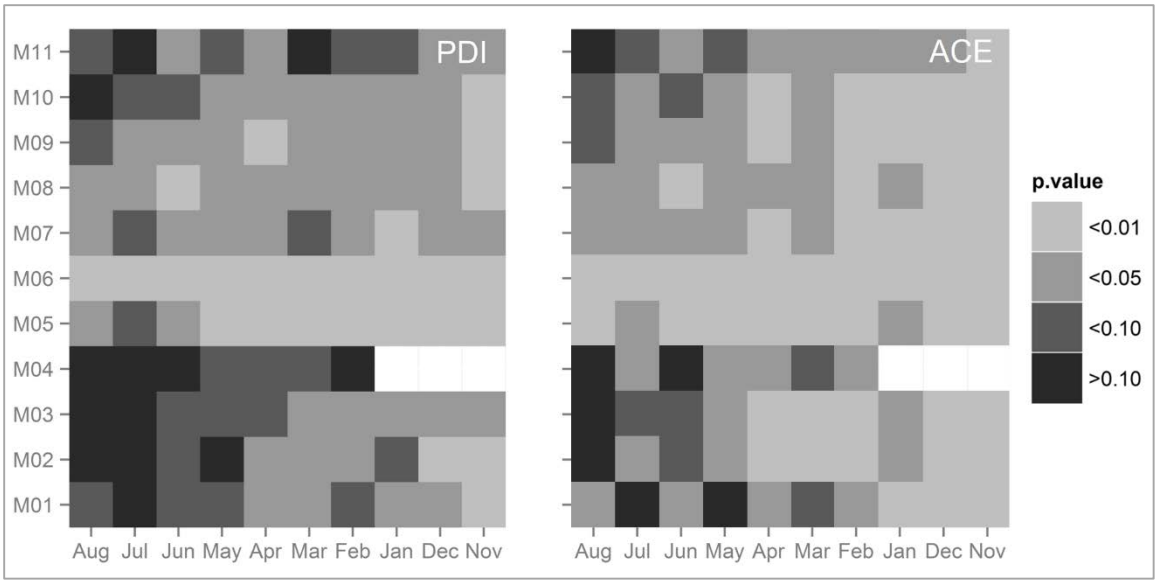

Figure 2-10: Probabilistic verification of the forecasts of PDI and ACE using the Kolmogorov-Smirnov test at different significance levels (90%, 95% ad 99%). Here M01= GFDL, M02 = GFDL-B01, M03 = GFDL-A06, M04 = NASA, M05 = COLA, M06 = CMC2, M07 = equally weighted model, M08 = weighted model, M09 = equally weighted model without COLA, M10 = weighted model without COLA, M11 = Hierarchical model. The larger the p-value, the lesser the evidence of departure from the uniform distribution, which is the target to avoid a too narrow or too wide forecast distribution.

2.5. Summary of Chapter 2

In this chapter, we have examined the retrospective skill in forecasting North Atlantic TC activity using four metrics: seasonal hurricane and tropical storm frequencies, ACE and PDI over the period of 1982-2015. Each of the hybrid statistical–dynamical models were built using SST_{ATL} and SST_{TROP} as the only two predictors. We used SSTs from the ensemble mean of six GCMs from the NMME project. Further, we built MME models using three different merging schemes, in which weights were the same for all the GCMs, or based on their capability in forecasting the two predictors, or based on how well they were able to reproduce the observed TC metric.

Overall, these simple hybrid-dynamical forecasting systems are able to reproduce well the year-to-year variations in the observations for all the four metrics. The skill overall decreases as we increase the forecast lead time. The long-lead skill (i.e., November and December-initialized forecasts) improves thank to the weighting schemes developed in this study. By using the weighting scheme procedures in building MME models, we can normalize the uncertainties associated with our covariates; this leads to improved predictions of SST even for the initialization months of November of the previous year, which in turn results in improvements in forecasting skills of TC activity. This is particularly true for the hierarchical model, in which the skill is retained for longer lead time.

CHAPTER 3: EVALUATION OF THE SKILL OF NUMERICAL WEATHER PREDICTION MODELS IN FORECASITNG RAINFALL ASSOCIATED WITH TROPICAL CYCLONES[2]

3.1 Introduction

NWP models provide forecasts of a number of weather-related variables (e.g., precipitation and temperature at different levels) for different lead-times (e.g., Lorenc 1986; Bougeault et al. 2010). However, quantitative information about the skill of NWP models in forecasting TC rainfall is still limited (e.g., Marchok et al. 2007; Mohanty et al. 2014). For a skillful prediction of TC rainfall, the models must predict the strength and distribution of rainfall rate and wind fields along with the track and intensity of the TC system (see Halperin et al. (2013) for a discussion on TC genesis forecasting of North Atlantic TCs). Therefore, precipitation forecasts from NWP models in general and for TCs in particular are inherently uncertain and subject to three types of error: localization, timing and intensity of precipitation events (e.g., Marchok et al. 2007). In this chapter, our goal is to evaluate the skill of NWP models in forecasting TC rainfall by quantifying their errors with respect to a reference (rain gauge-based) dataset. Moreover, five additional "observational" (remote sensing-based) datasets are also compared to the reference dataset: the skill of the NWP models in forecasting TC rainfall is quantified for different lead-times, and discussed in light of the performance of these "observational" products.

---

[2] Adapted from: Luitel, B., G. Villarini, and G.A. Vecchi, Can numerical weather prediction models skillfully forecast rainfall from U.S. landfalling tropical cyclones?, submitted to Journal of Hydrology, 2016.

In this chapter, the description of data and methodology is provided in Section 3.2, followed by results and discussion in Section 3.3. Section 3.4 summarizes the main points of the study and concludes the chapter.

3.2 Data and Methodology

We use gridded daily precipitation from National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) as the reference data. CPC precipitation data are obtained by interpolating rain gauge measurements to provide a gridded daily product for the continental United States with 0.25° × 0.25° spatial resolution (Higgins et al. 2000; Wehner et al. 2010). For the North Atlantic TC track information (date, time, latitude and longitude of all recorded storms with a 6-hour resolution) we use the HURDAT-2 (Landsea and Franklin 2013) introduced in Chapter 2.

We evaluate the forecast rainfall produced by five state-of-art NWP models: European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza et al. 2007) UK Met Office (UKMO; Bowler et al. 2008), National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1997), China Meteorological Administration (CMA), and Canadian Meteorological Center (CMC; Houtekamer et al. 2009). Data for NWP models have been archived from the THORPEX Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010). To benchmark the skill of these NWP models, we consider rainfall estimates from five remote sensing products (one ground based radar and four satellite-based rainfall products). Stage IV multi-sensor precipitation dataset is produced by NOAA-NCEP (Lin and Mitchell 2005). It has ~4-km and hourly resolution, and is obtained by merging ground-based radars across the United States, and rain gauge measurements are

used to perform bias correction. The four satellite-based rainfall products we use are: Tropical Rainfall Measuring Mission - Multi-satellite Precipitation Analysis [TMPA; both real-time (TMPA_RT) and research version (TMPA_RV); Huffman et al. 2010)]; Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2000); CPC MORPHing Technique (CMORPH; Joyce et al. 2004). These satellite-based products have a resolution coarser than Stage IV (3-hourly and 0.25-degree), and TMPA-research version is the only product for which a monthly bias correction with respect to rain gauges is applied.

The rainfall products have different spatial resolution. Stage IV has the highest spatial resolution (~4km), the four satellite-based remote sensing products and the CPC data have 0.25 × 0.25 degree lat/lon spatial resolution, whereas all the NWP model output we have explored are on a 0.5 degree lat/lon spatial resolution grid (the default resolution in the TIGGE archive). Because our focus is on the NWP models, all the products were regridded into 0.5-degree resolution to established uniformity in the analyses, with the expectation that both the agreement between observational estimates and the forecasting skill will increase as we coarsen the spatial resolution (e.g., Lavers and Villarini 2013). Results are based only on rainfall over land.

We focus on the evaluation of the precipitation associated with North Atlantic TCs that affected the continental United States over the period 2007–2012, and examine 15 storms that came within 500 km of the coast of the United States. We consider TC-rainfall the rainfall that occurred within a 500-km buffer around the centre of circulation of a given storm. We compare the storm total rainfall obtained from the CPC data (our reference) against the rainfall forecasts from the five NWP models. To quantify how close (or far)

these forecasts are from the reference data, we also use rainfall estimates from the five remote sensing products to give a range of potentially acceptable results.

Here we use Hurricane Irene (2011) as an example of our approach and methodology (Figure 3-1; the results for the other 14 storms are in Appendix B, Figures B-1 - B-14). Our approach is to examine the skill of the forecasts starting from the first time the center of circulation of the storm is within 500 km from the U.S. coastline. We will refer to this as the "0-hour lead-time." In the example in Figure 3-1, the "0-hour lead-time" represents the storm total rainfall from 27 August 2011 at 0 UTC to 31 August 2011 at 0 UTC. We will refer to the "12-hour lead-time" the forecast for the period from 27 August 2011 at 0 UTC to 31 August 2011 at 0 UTC initialized on 26 August 2011 at 12 UTC. Longer lead-times will follow the same rationale.

We compute rainfall errors with respect to the reference data (CPC) at each 0.5-degree pixel by subtracting CPC rainfall accumulations from the rainfall accumulations obtained from each of the remote sensing products and the NWP models at each 0.5-degree pixel. An examination of the skill of NWP in forecasting TC rainfall is based on both visual and quantitative evaluations. Visual inspection was performed by plotting storm total rainfall and rainfall errors for each of the storms for all the products. Statistical analysis of the rainfall errors was performed by computing the probability density function (pdf) and the first three moments of the error distributions. For each of the statistical measures we used in the process of skill evaluation, we develop envelopes from the five remote sensing products and then used them as the range to quantify and evaluate the skills of the NWP models.

3.3 Results

Skill evaluation of the NWP models is based on visual examination of the storm total rainfall fields and on the quantitative analysis of the error characteristics. As an example of the type of analyses we have performed, we focus on the results for Hurricane Irene (2011). According to the NHC report (http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf), Irene caused 41 direct deaths in the United States, among which six fatalities were due to storm surge, 15 were related to high wind and 21 were due to rainfall-induced floods. Based on the National Flood Insurance Program, the NHC reported that the total damage caused by the Irene was $15.8 billion out of which $7.2 billion of the damage was related to inland flooding and surge only (Avila and Cangialosi 2011; McCallum et al. 2012). Hurricane Irene affected large areas of the eastern United States from North Carolina to Maine (Figure 3-1).

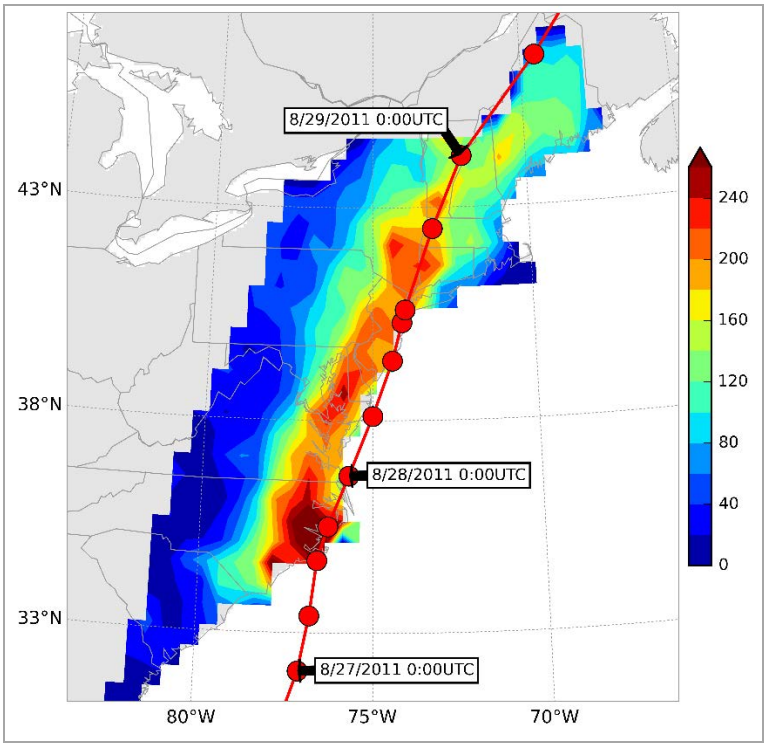

Figure 3-1: Rainfall accumulation (in mm) for Hurricane Irene (August 26-30 2011) based on the CPC dataset. The red curve with dots represents the storm track.

Figure 3-2 summarizes the total rainfall accumulation during Hurricane Irene according to the five remote sensing products and the five NWP models (the results for the other 14 storms are in Appendix B, Figures B-15-B-28). These results show that there is a relatively large range of variability in the "observational products" (top row in Figure 3-2). As a first step, each panel in Figure 3-2 is visually compared with CPC rainfall accumulation (Figure 3-1). This visual comparison provides qualitative information about the capability of each of these products in estimating rainfall associated with TCs. Overall, Stage IV is the product that more closely resembles the observational data both in terms of magnitudes and location of the areas with the larger rainfall accumulations (see also Villarini et al. (2011)). On the other hand, the satellite-based estimates tend to have smaller rainfall values generally spread over larger areas (this is particularly true for PERSIANN). We will use this variability in rainfall estimates from observational systems (ground- and space-based sensors) as a way of bounding what we can consider acceptable for the NWP models: we will deem as satisfactory NWP rainfall forecasts that are within the range of outcomes from the remote-sensing products.

The TC-rainfall forecasts appear to capture reasonably well the observed storm total rainfall both in location and magnitude. This is particularly true for the shortest lead-time, with the performance decreasing as we increase the lead-time to five days. As mentioned before, this has to do with the fact that we are expecting the models to not only correctly forecast the rainfall fields around the center of circulation of the storms, but also to correctly track these storms (e.g., Marchok et al. 2007). This issue is clear, for instance, for the 5-day lead-time, where the models were not able to correctly forecast the storm track.

Our expectations of the rainfall forecasts from the NWP models are admittedly very high, but necessary in order to improve our confidence in them.

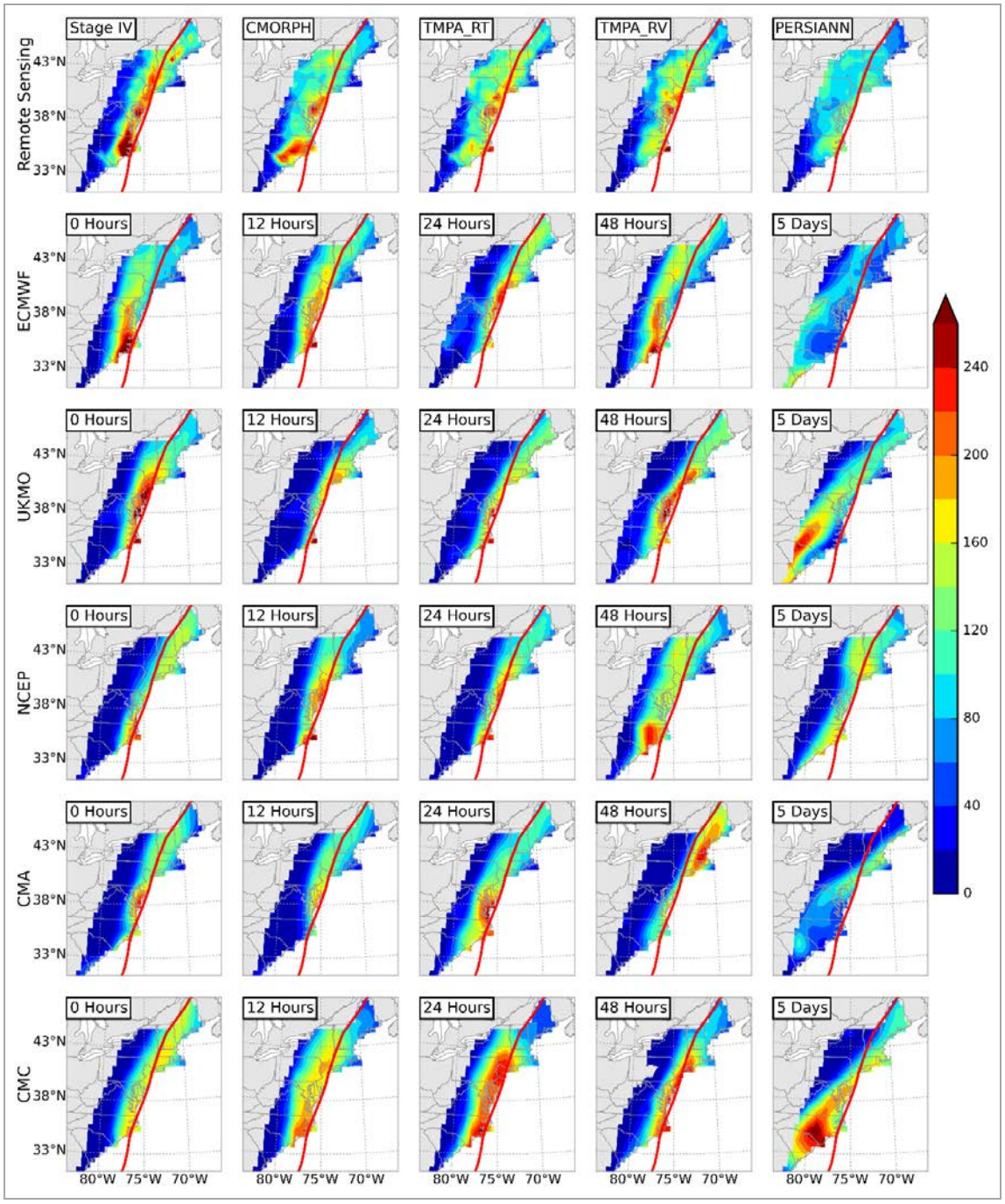

Figure 3-2: Rainfall accumulation (in mm) for Hurricane Irene (August 26-30 2011) based on remote sensing (top row) and NWP models (second to last rows). For the NWP models, lead-times increases left to right from 0 hours to 5 days. The red lines represent the storm track.

Furthermore, we compute rainfall errors using an additive formulation by subtracting the storm total rainfall from CPC from each of the remote sensing products and NWP models. The results in Figure 3-3 are for Hurricane Irene, while those for the other 14 TCs are presented in Appendix B (Figures B-29 – B-42). Stage IV shows the smallest discrepancies with respect to CPC, likely due to the bias correction performed using rain gauges. On the other hand, the results for the four satellite-based products show areas with consistent over- and under-estimation, similar to what observed for the NWP models. While for the forecasting models these patterns can be ascribed to errors in the forecasted tracks, it is harder to come up with an explanation for the satellite products. The magnitude of the differences between CPC and the NWP models increases as we increase the lead-time, with results up to 2-day lead-times that are comparable to the range of accuracy (both quantity and location wise) of the "observational" products. That is, the NWP models up to the 48-hour lead-time were at least as skillful at predicting rainfall as the least accurate remote sensing product we used. These findings are generally consistent across all the 15 TCs considered in this study. Among the NWP models, we did not see any one model performing consistently as the best or worst for all lead-times.

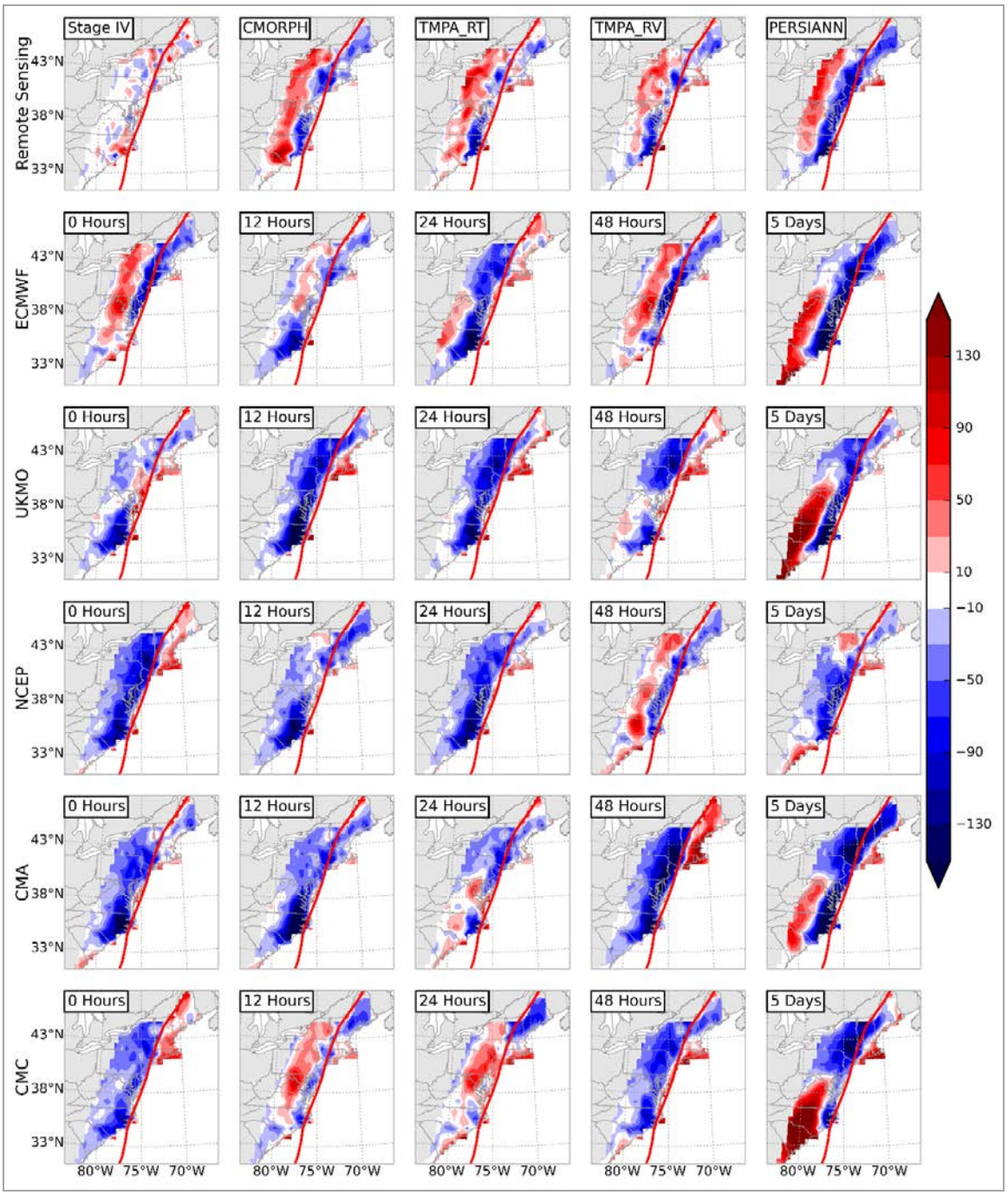

Figure 3-3: Rainfall errors (in mm) for Hurricane Irene (August 26-30 2011) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hours to 5 days. The red lines represent the storm track.

A more quantitative examination of the skill of NWP models is achieved by computing the pdf of the rainfall errors from each product. Figures 3-4 – 3-8 show the pdf of the rainfall errors for models ECMWF, UKMO, NCEP, CMA and CMC respectively for all 15 storms, together with the results obtained using the remote sensing products ("observational" products). For a given lead-time, if the pdf from any NWP model falls within the range of variability from the observational products (grey envelope in Figures 3-4 - 3-8), we can infer that the model at that lead-time was capable of predicting the TC rainfall with a skill comparable to the range of results from our five "observational" products. Taking into account all the challenges associated with the forecasting of TC rainfall (i.e., correctly forecast both the storm track and the rainfall around the center of circulation of these storms), the results in Figures 3-4 – 3-8 are rather promising. Overall, the error distributions for these 15 TCs are comparable to the error distributions from the remote sensing products. This is particularly true for the shortest lead-times, with the results for the 5-day lead that appear to be much smoother than the others. These results are consistent across all NWP models (Figures 3-4 – 3-8).

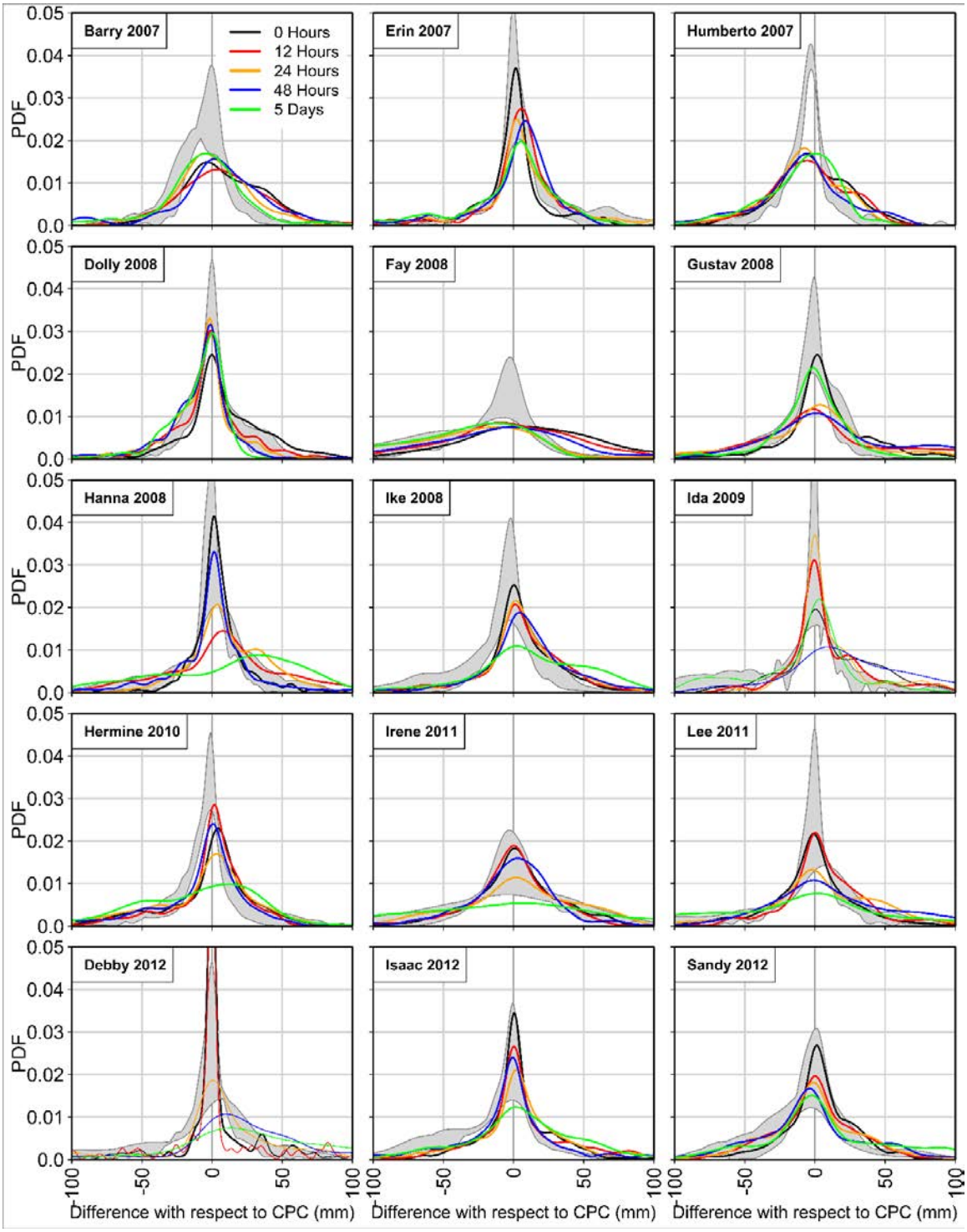

Figure 3-4: Probability density functions (pdfs) of rainfall errors exhibited by the ECMWF model for different forecast lead-times. The gray shaded envelopes represent the range of pdfs obtained from the rainfall errors of the five remote sensing products.

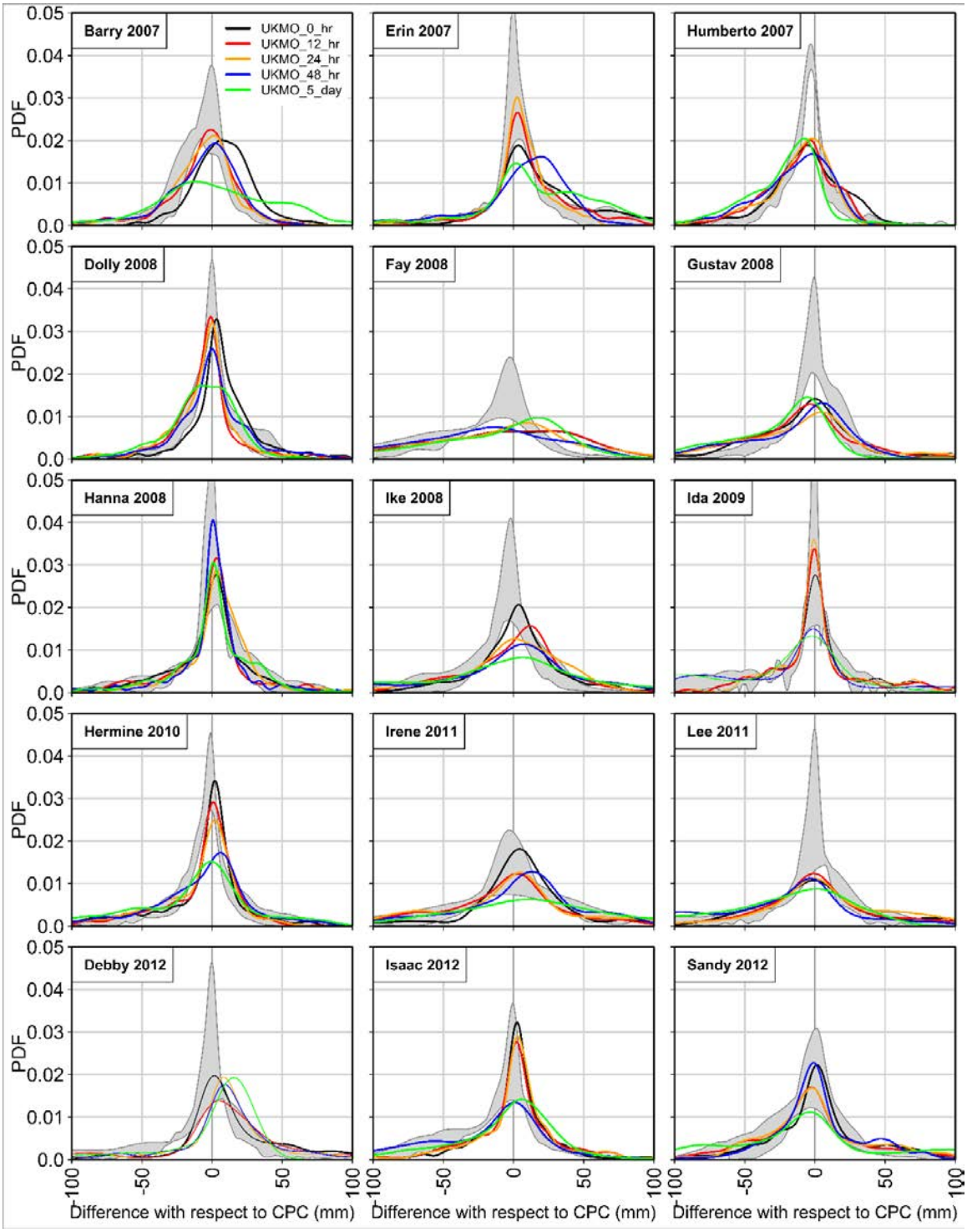

Figure 3-5: Same as Figure 3-4, but for UKMO model.

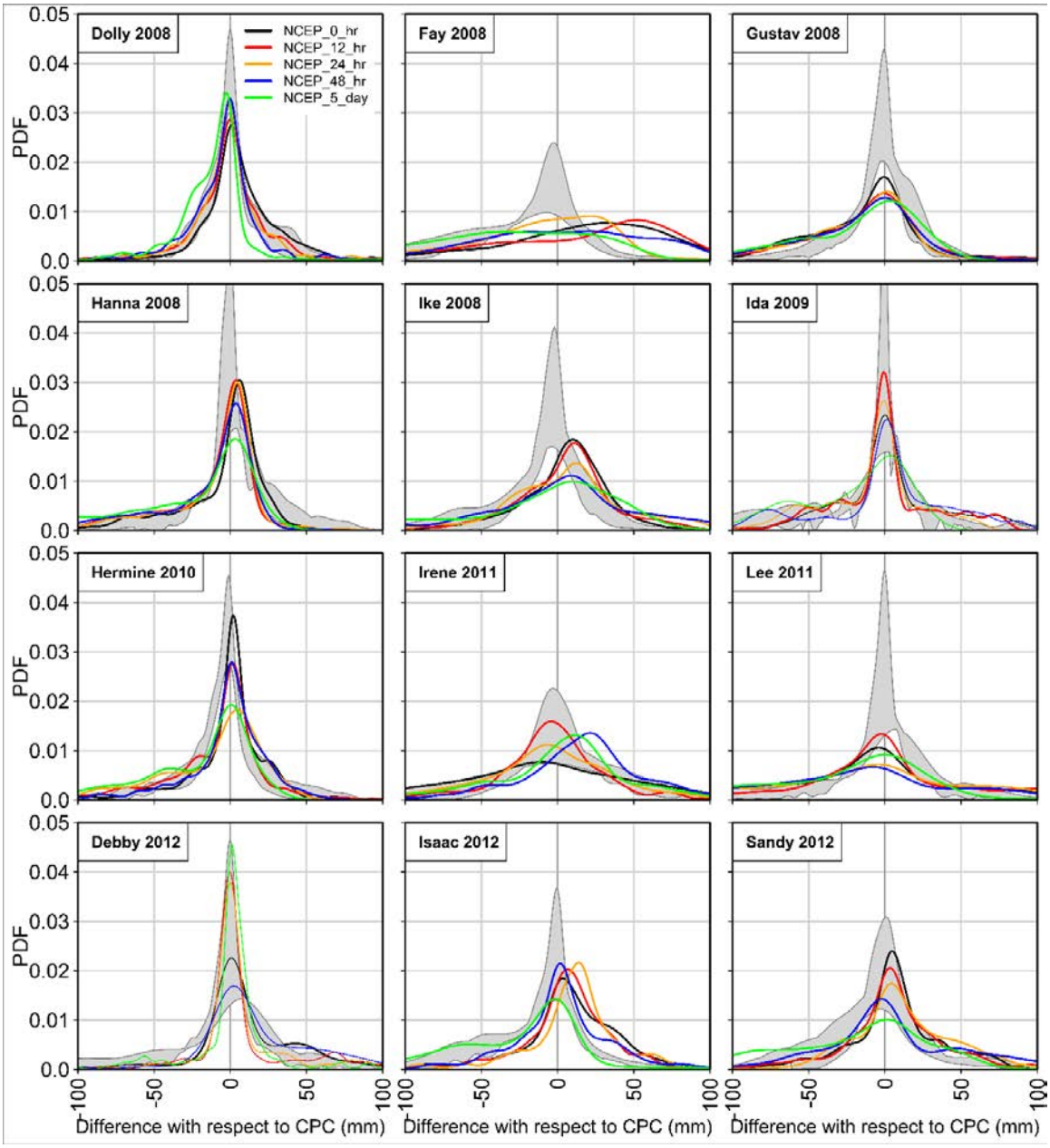

Figure 3-6: Same as Figure 3-4, but for NCEP model.

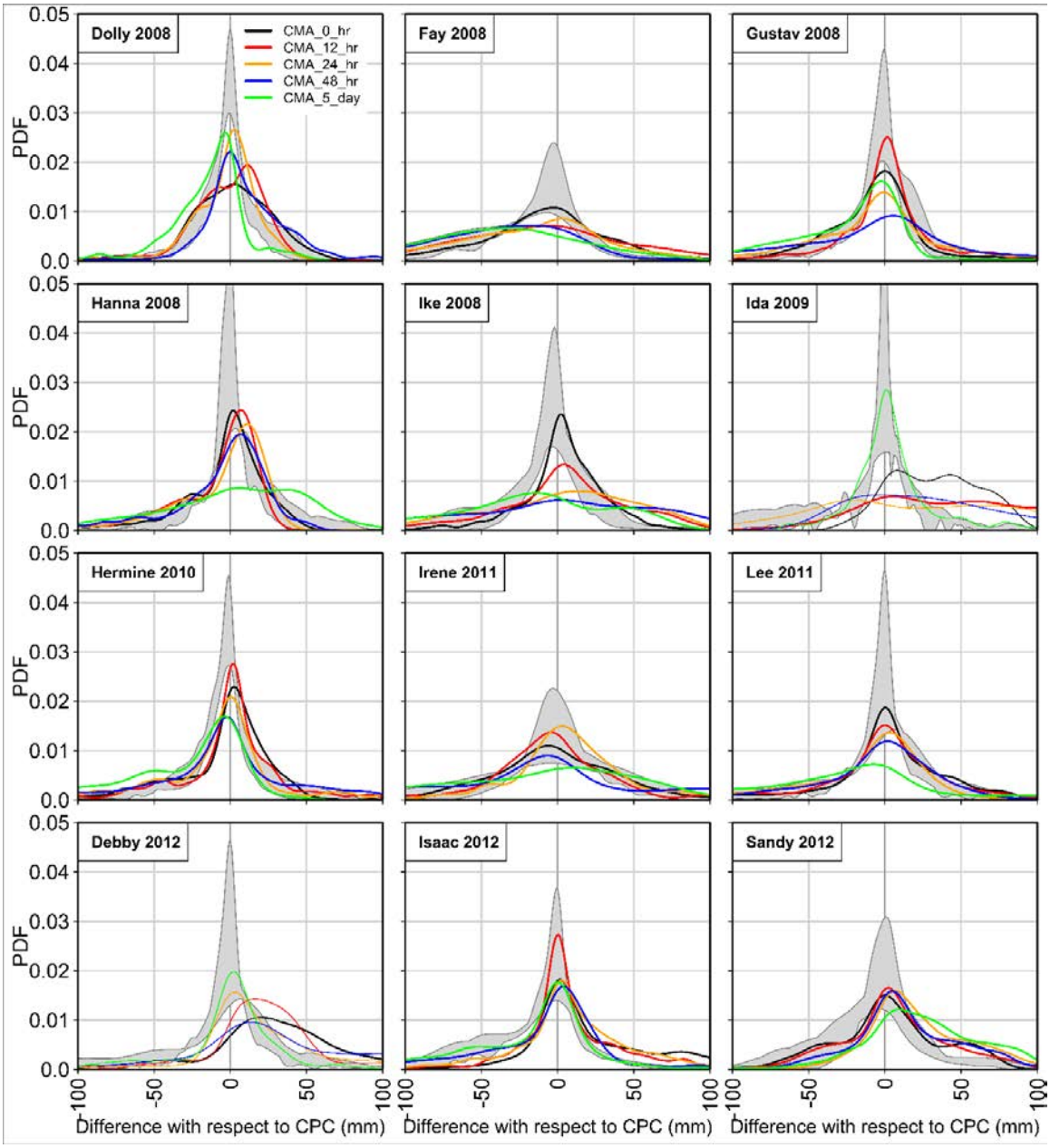

Figure 3-7: Same as Figure 3-4, but for CMA model.

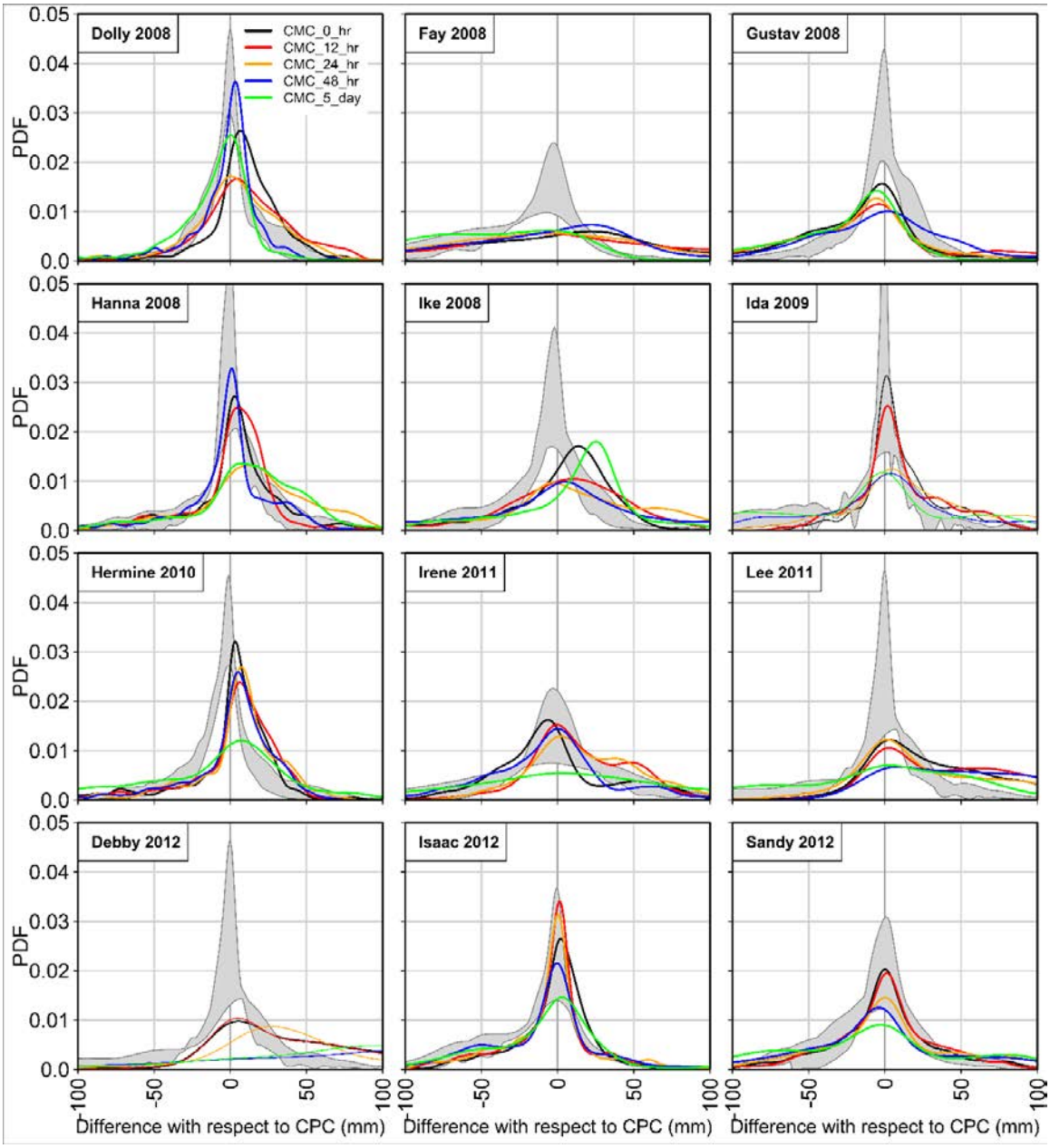

Figure 3-8: Same as Figure 3-4, but for CMC model.

In addition to the entire error distribution, we have also computed the first three moments (mean, standard deviation and skewness) of the errors, and compared them across NWP models and remote-sensing products (Figure 3-9). A model is considered to have better performance if the expected value of the errors is close to zero (unbiased). In the left

column of Figure 3-9, the expected values of the errors are closer to zero for the ECMWF, followed by UKMO and NCEP. In general, we would consider a "successful" forecast one for which the statistical properties of the error are within what we obtain from the remote-sensing products. For instance, in case of Superstorm Sandy (2012), UKMO, ECMWF and CMC have the average error close to zero across different lead times, with error characteristics that are almost better than what we obtain from the remote sensing products. More generally, with the exception of Tropical Storm Lee (2011) and Tropical Storm Debby (2012), the average errors from the NWP models are close to zero (or at least within the uncertainties from the remote sensing products) up to the 2-day lead time.

The standard deviations of the errors from the NWP models show, in general, more variability than what obtained from the remote-sensing products (Figure 3-9, middle column). The main exception is represented by the ECMWF, which shows variability that is comparable with that of the observational products. Therefore, these results indicate that the distribution of the errors, while generally unbiased for the NWP models, tends to be flatter than what we would expect from a range of remote sensing products.

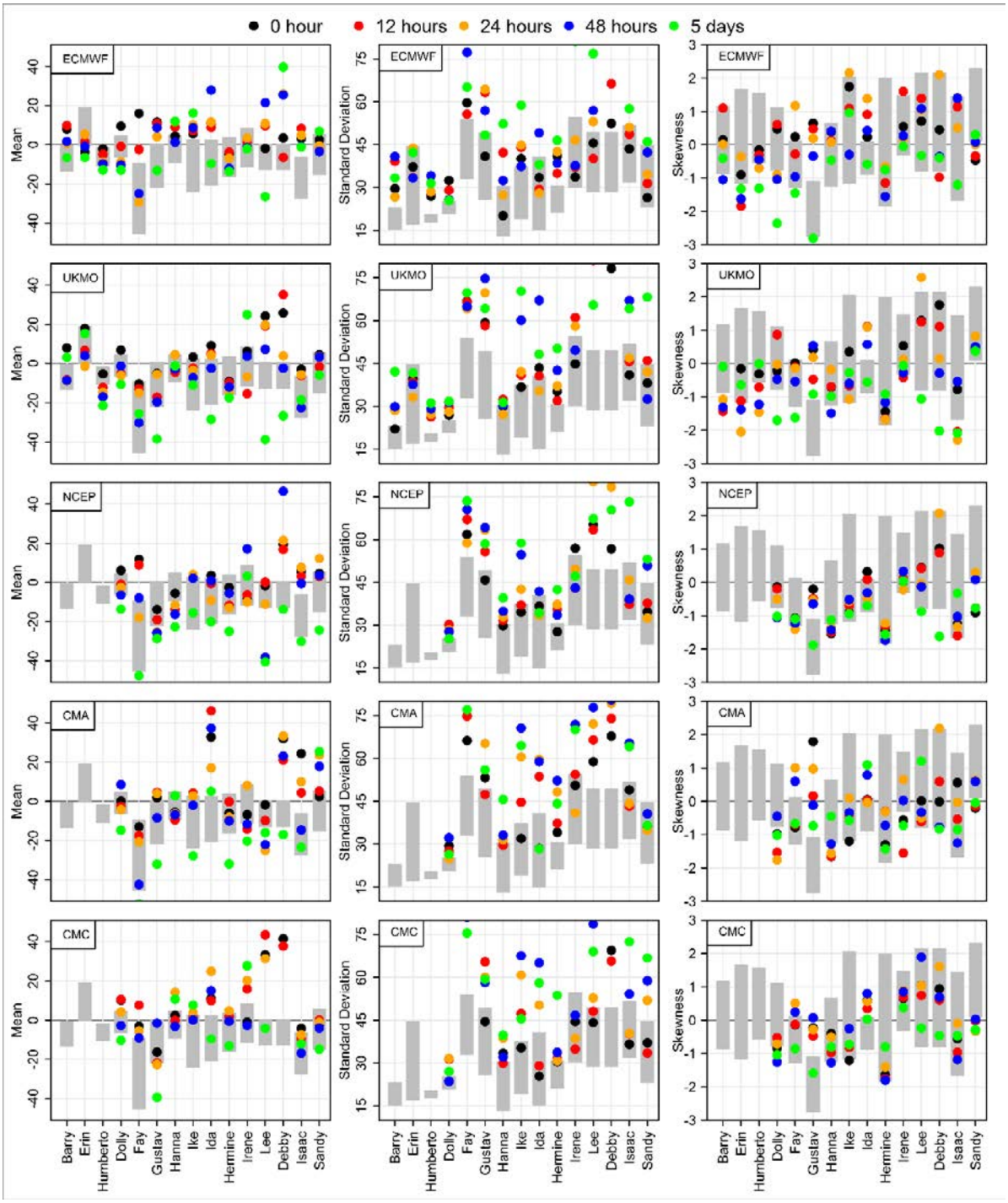

Figure 3-9: Mean (in mm; left column), standard deviation (in mm; middle column) and skewness (right column) of the rainfall errors for different lead-times and NWP models. The gray rectangular boxes in the background represent the range of values for each statistics based on the five remote sensing products. No forecast information for NCEP, CMA, and CMC was available for Tropical Storm Barry (2007), Tropical Storm Erin (2007) and Hurricane Umberto (2007).

Finally, the skewness of the errors from the NWP models is well within the results obtained from the observational data across the different models and lead times (Figure 3-9, right column). Depending on the TC, NWP model or lead time, the error distributions tend to be skewed, in particular negatively skewed. Therefore, our quantitative evaluation of the forecast errors suggests that the NWP models are capable of representing the error structure obtained from remote sensing products, in particular for lead times up to two days, even though the error distribution tends to be flatter than what observed from the observational datasets.

3.4 Summary of Chapter 3

In this chapter we have examined the skill of five state-of-the-art NWP models [European Centre for Medium-Range Weather Forecasts (ECMWF), UK Met Office (UKMO), National Centers for Environmental Prediction (NCEP), China Meteorological Administration (CMA), and Canadian Meteorological Center (CMC)] in forecasting rainfall associated with 15 U.S. landfalling TCs during the 2007-2012 period. These forecasts with a lead time up to five days were compared against gridded rain gauge based measurements. We used rainfall estimates from five remote-sensing products as a way of quantifying the fidelity we can expect from different observational datasets. The main findings of this study can be summarized as follows:

- Among the remote sensing products, Stage IV showed the closest resemblance with the CPC dataset, likely because of the use of rain gauge information for bias correction. Among other remote sensing products, PERSIANN consistently underestimated the TC rainfall.

- Overall, the performance of the NWP models was comparable to the difference between remote-sensing based products and the reference dataset, particularly at shorter lead-times. While quantitatively the forecasts were comparable to the reference data, a substantial error in the precipitation predictions appears to have been related to predictions of the storm track itself. Track-relative analyses as those performed in Marchok et al. (2007) would alleviate this issue, but were not performed here because of the lack of forecast track.

In interpreting these results, it is important to remember that we are evaluating the performance for a very challenging target: we are asking the NWP models not only to correctly forecast the storm track, but also to correctly characterize the rainfall distribution around this track. When we consider that the errors from the remote sensing products ("observations") are comparable to the forecast errors, the results from the NWP are rather encouraging. Overall, our findings indicate that skillful forecasting of storm total rainfall associated with North Atlantic TCs making landfall along the U.S. coast is possible up to two days prior to landfall.

CHAPTER 4: CONCLUSIONS

The overall objective of this thesis was to evaluate the predictability of North Atlantic TCs for long lead times, while we focused on short lead times when examining the skill of NWP models in forecasting rainfall associated with these storms.

We achieved our first objective by developing statistical-dynamical seasonal forecasting systems. Each of the hybrid statistical–dynamical models were built using tropical Atlantic and tropical mean SSTs as the only two predictors. The TC forecasts were based on SST forecasts from six GCMs from the NMME project. We also built multi-model ensemble averages using three different merging techniques as discussed in Section 2.3. For a TC season that peaks in August-October, we showed that it is possible to obtain skillful forecasts of basin-wide TC activity as early as November of the previous year; this means that we can skillfully forecast the upcoming season as the current one is still coming to an end. Our findings also clearly show the advantage of using a multi-model ensemble average to improve the results with respect to the predictions from individual models. Based on our examination of three different weighting schemes to compute the ensemble averages, we found that they all lead to improvements, in particular at the longer lead times. Overall, we contend that it is possible to skillfully forecast basin-wide activity, with the multi-model procedure that would be recommended whenever possible.

Based on these forecast models, we here provide the forecast of North Atlantic TC activity for the 2016 season initialized from November 2015 to March 2016 (Tables 4-1 - 4-4). With few exceptions, almost all of the models point to an above average 2016 season in terms of basin-wide hurricane or tropical storm counts (Tables 4-1 and 4-2), whereas ACE and PDI values (Tables 4-3 and 5-4) are around the 1982-2015 average. Considering

that the hierarchical models provided the overall best retrospective skill, we can forecast a median activity of 7 hurricanes and 11 tropical storms (using the median as reference). This forecast is likely highly influenced by the very strong El Niño we have just experienced. Following a strong El Niño, trade winds can become stronger than normal and an abnormal accumulation of cold water can occur in the central and eastern Pacific resulting in a strong La Niña ((Hoerling et al. 1997). During La Niña, the North Atlantic TC activity is generally above average with higher probability of more damages in the central and eastern United States (Pielke Jr; Landsea 1999; Villarini et al. 2014; Vitart and Anderson 2001). Our forecasts point to above average activity, even though the values of the TC metrics considered here have been generally reducing as we reduce the lead time.

Table 4-1: Summary of the 2016 season median hurricane forecast. The results in the first six rows are forecasts according to the six individual GCMs described in Table 1-1 and the remaining five models are the forecasts according to the different MME models described in Section 2.3. The March-initialization (2nd column) is the shortest lead time, and lead times increase as we move right, with the longest initialization in November 2015. Black filled cells indicate forecasts above the average value from the 1982-2015 value (6.32).

Hurricane	Initialization Month				
	Mar	Feb	Jan	Dec	Nov
GFDL	12	12	13	11	13
GFDL_B01	5	7	8	9	9
GFDL_A06	8	8	9	10	12
NASA	5	6	NA	NA	NA
COLA	7	NA	6	4	6
CMC2	6	6	6	6	5
Average	7	8	8	8	8
Weighted	7	8	9	9	9
Average w/o COLA	7	8	8	8	9
Weighted w/o COLA	7	8	8	8	9
Hierarchical	7	7	8	8	9

Table 4-2: Same as in Table 4-1 for tropical storms. The average number of tropical storms over the period of 1982-2015 is 9.24.

TS	Initialization Month				
	Mar	Feb	Jan	Dec	Nov
GFDL	14	14	15	14	15
GFDL_B01	9	11	11	12	12
GFDL_A06	11	12	12	13	14
NASA	10	11	NA	NA	NA
COLA	10	NA	9	7	9
CMC2	10	10	10	10	9
Average	11	11	11	11	12
Weighted	11	11	12	12	12
Average w/o COLA	11	11	12	12	12
Weighted w/o COLA	11	11	12	12	13
Hierarchical	11	11	11	11	12

Table 4-3: Same as in Table 4-1 for ACE. The average value of ACE over the period of 1982-2015 is 5.8.

ACE	Initialization Month				
	Mar	Feb	Jan	Dec	Nov
GFDL	8.9	8.2	9.5	7.7	9.4
GFDL_B01	3.4	4.3	5.1	6.2	8
GFDL_A06	5.6	5.9	6.7	7.7	9.3
NASA	3.5	3.9	NA	NA	NA
COLA	5.1	NA	4.4	3.1	4.4
CMC2	3.5	4.1	4.3	4.3	3.2
Average	4.7	5.12	5.7	5.5	5.9
Weighted	4.6	5.1	6.1	6.3	6.3
Average w/o COLA	4.7	5.0	5.9	5.9	6.3
Weighted w/o COLA	4.7	5.2	6.2	6.5	6.7
Hierarchical	4.8	4.9	5.5	5.6	5.9

Table 4-4: Same as in Table 4-1 for PDI. The average value of PDI over the period of 1982-2015 is 2.4.

PDI	Initialization Month				
	Mar	Feb	Jan	Dec	Nov
GFDL	4.1	3. 7	4.3	3.4	4.3
GFDL_B01	1.3	1.7	2.1	2.7	2.5
GFDL_A06	2.4	2.5	2.9	3.4	4.2
NASA	1.4	1.5	NA	NA	NA
COLA	2.2	NA	1.8	1.2	1.8
CMC2	1.4	1.6	1.7	1.7	1.2
Average	1.9	2.1	2.4	2.3	2.5
Weighted	1.9	2.1	2.6	2.7	2.7
Average w/o COLA	1.9	2.1	2.5	2.5	2.7
Weighted w/o COLA	1.9	2.2	2.7	2.8	2.9
Hierarchical	2.0	2.1	2.3	2.4	2.5

Our second objective of evaluating the skill in short-term forecasting of rainfall associated with TCs was achieved by considering five state-of-the-art NWP models. Our focus was on 15 U.S landfalling TCs over the period of 2007-2012. We quantified the skill of these models by visual examination of the distribution of the errors for the different lead-times, and numerical examination of the first three moments of the error distribution. Moreover, to assess whether a NWP forecast could be potentially useful, we used rainfall estimates from five remote-sensing rainfall products for the same 15 storms to aid with the decision. Our results were rather encouraging, considering the complexity of the problem: we were not only asking the NWP models to forecast the rainfall distribution around the TC track, but also to correctly track the future path of these storms. Despite the large complexity of the problem, the NWP models performed relatively well, with skillful rainfall forecasts with a lead time up to about two days.

This work represents a first step towards a more comprehensive forecasting system of TC activity. Ideally, we are interested in skillful predictions of TC impacts (e.g., heavy rainfall and flooding) at the local scale with a long lead time. While this is currently not possible, significant progress has been made in this direction in the recent years (e.g., Murakami et al. 2016), with more progress expected by improving our understanding of the physical processes controlling TC activity, tracking and impacts.

APPENDIX A: SUPPLEMENTAL MATERIALS FOR CHAPTER 2

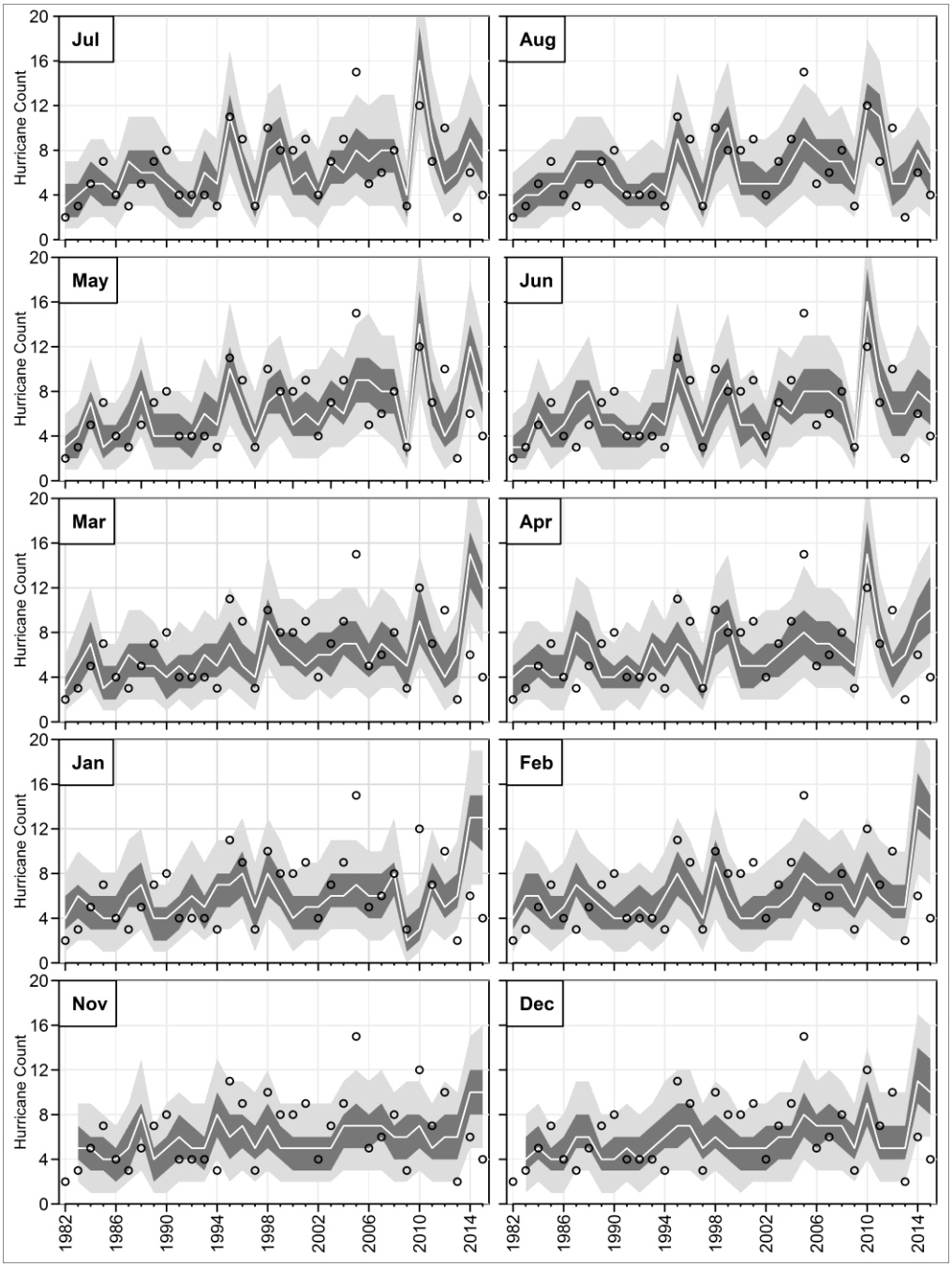

Figure A-1: Retrospective seasonal forecast of hurricane using GFDL model, initialized from November (of the previous year for each hurricane season) to August (of the same year). In each panel, the circles are observed values, the white line is the median (50th percentile) for each season. The outer light gray area represents the region between the 5th and 95th percentiles, while the middle dark gray area represents the region between the 25th and 75th percentiles.

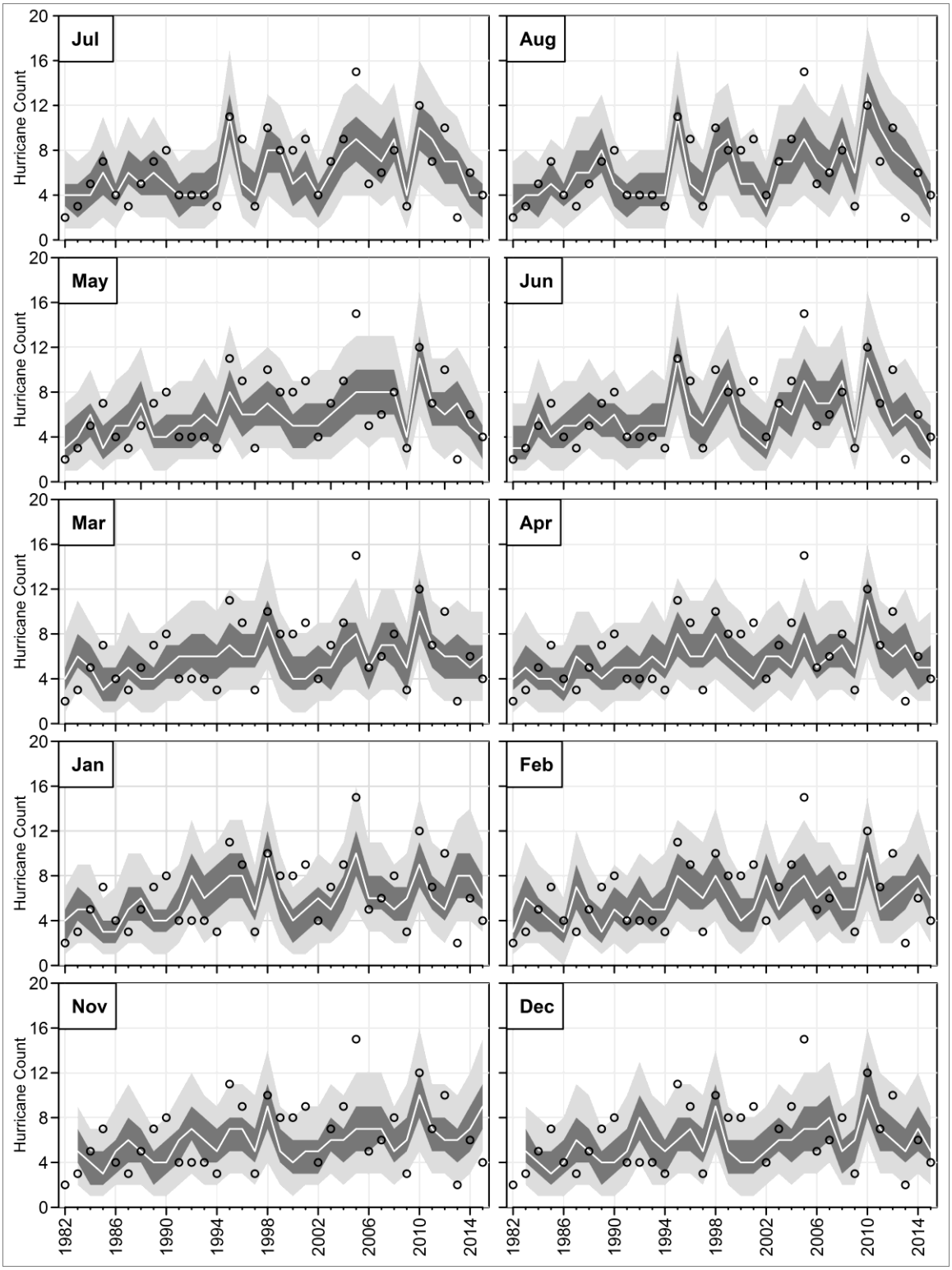

Figure A-2: Same as in Figure A-1, but using GFDL-B01.

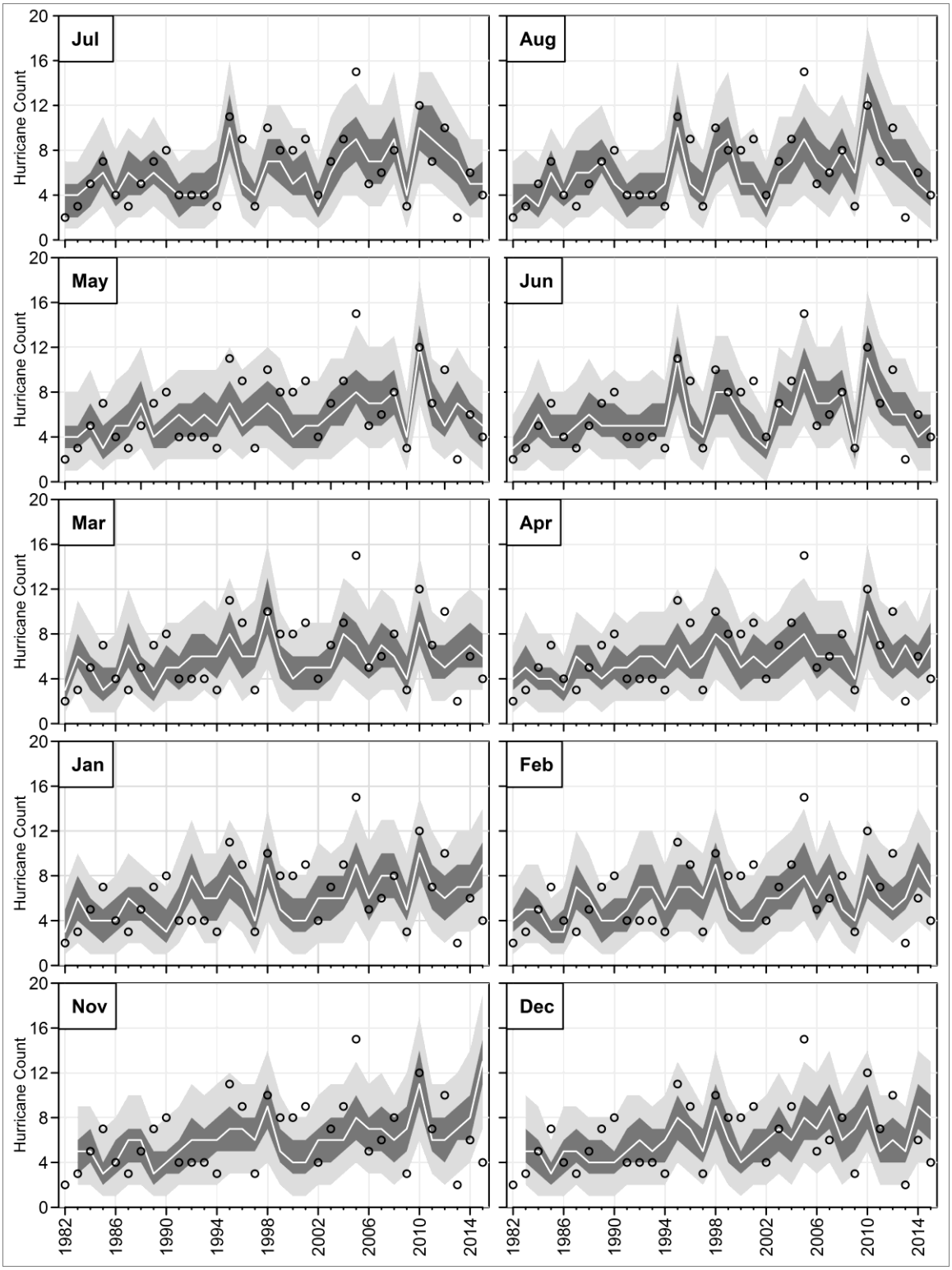

Figure A-3: Same as in Figure A-1, but using GFDL-A06.

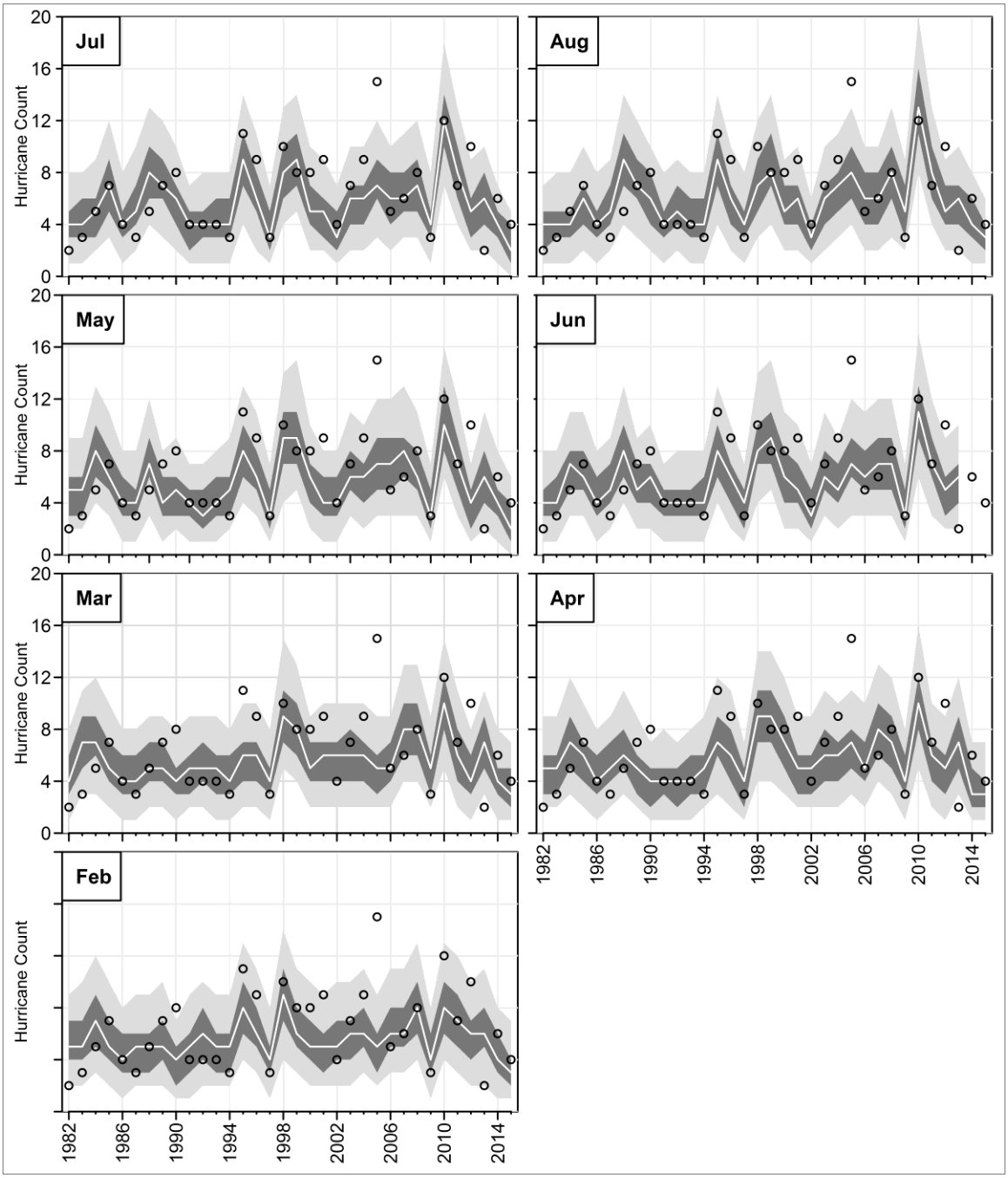

Figure A-4: Same as in Figure A-1, but using NASA.

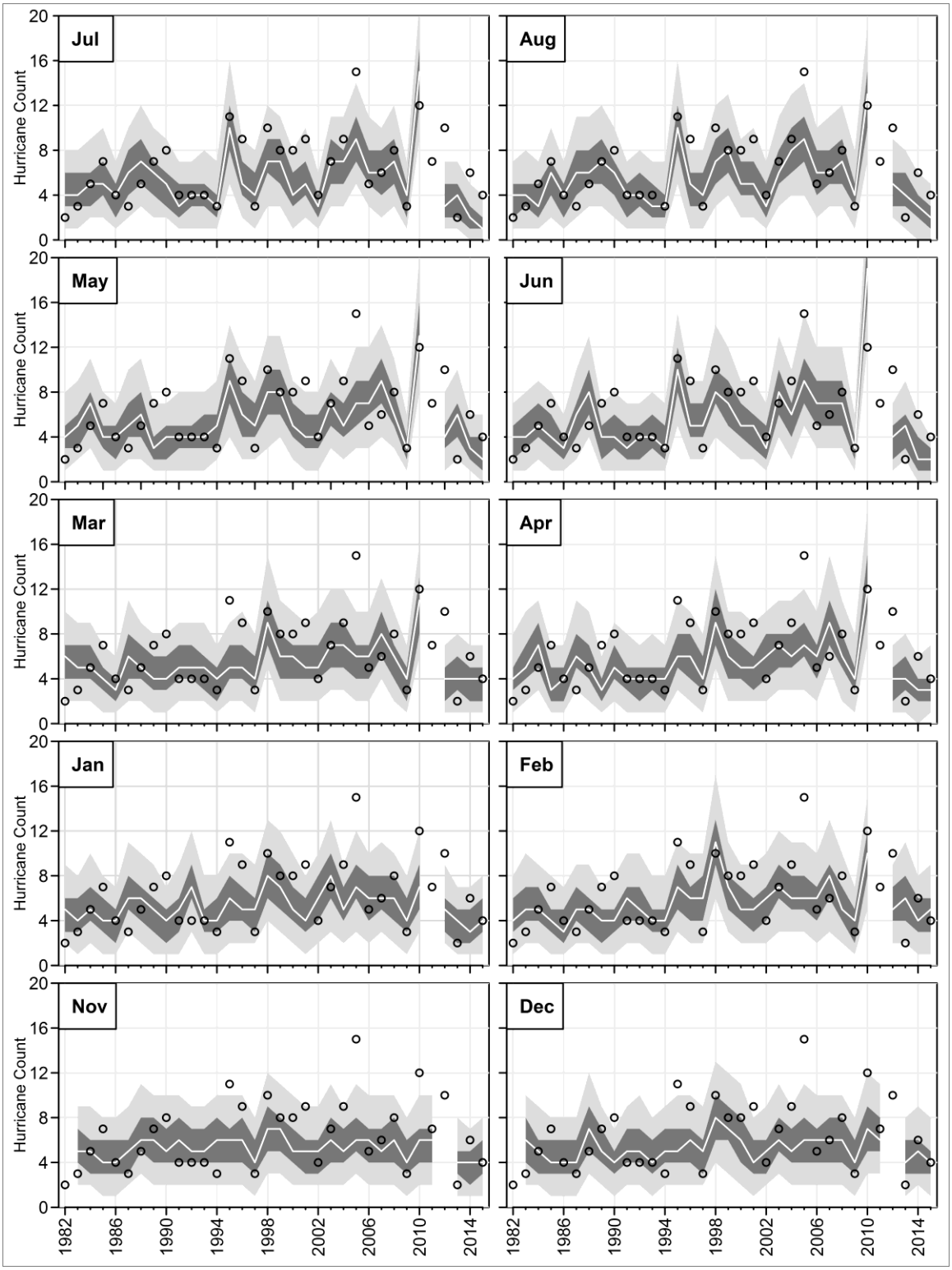

Figure A-5: Same as in Figure A-1, but using CMC2.

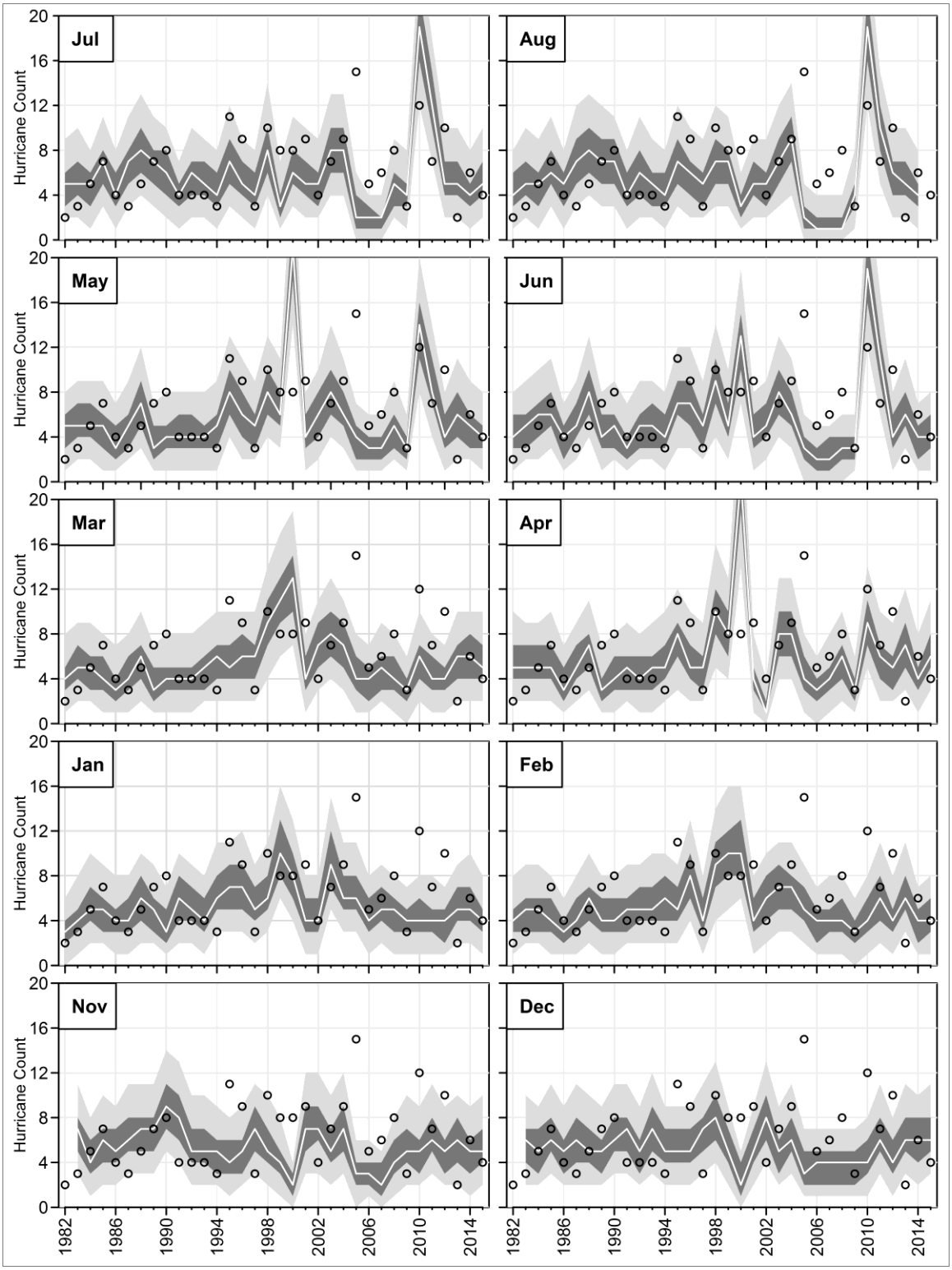

Figure A-6: Same as in Figure A-1, but using COLA.

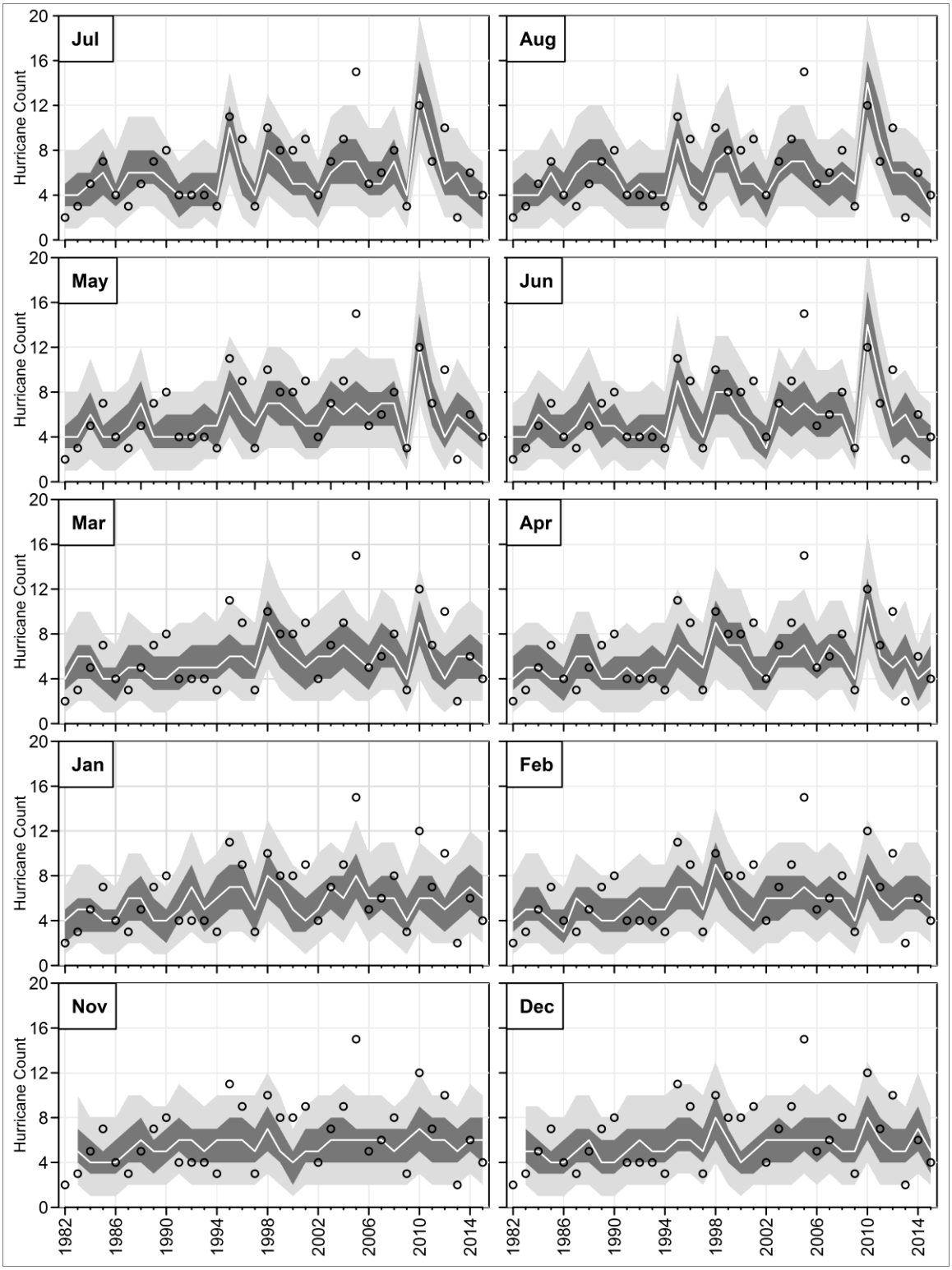

Figure A-7: Same as in Figure A-1, but using model in which the predictors are equally weighted.

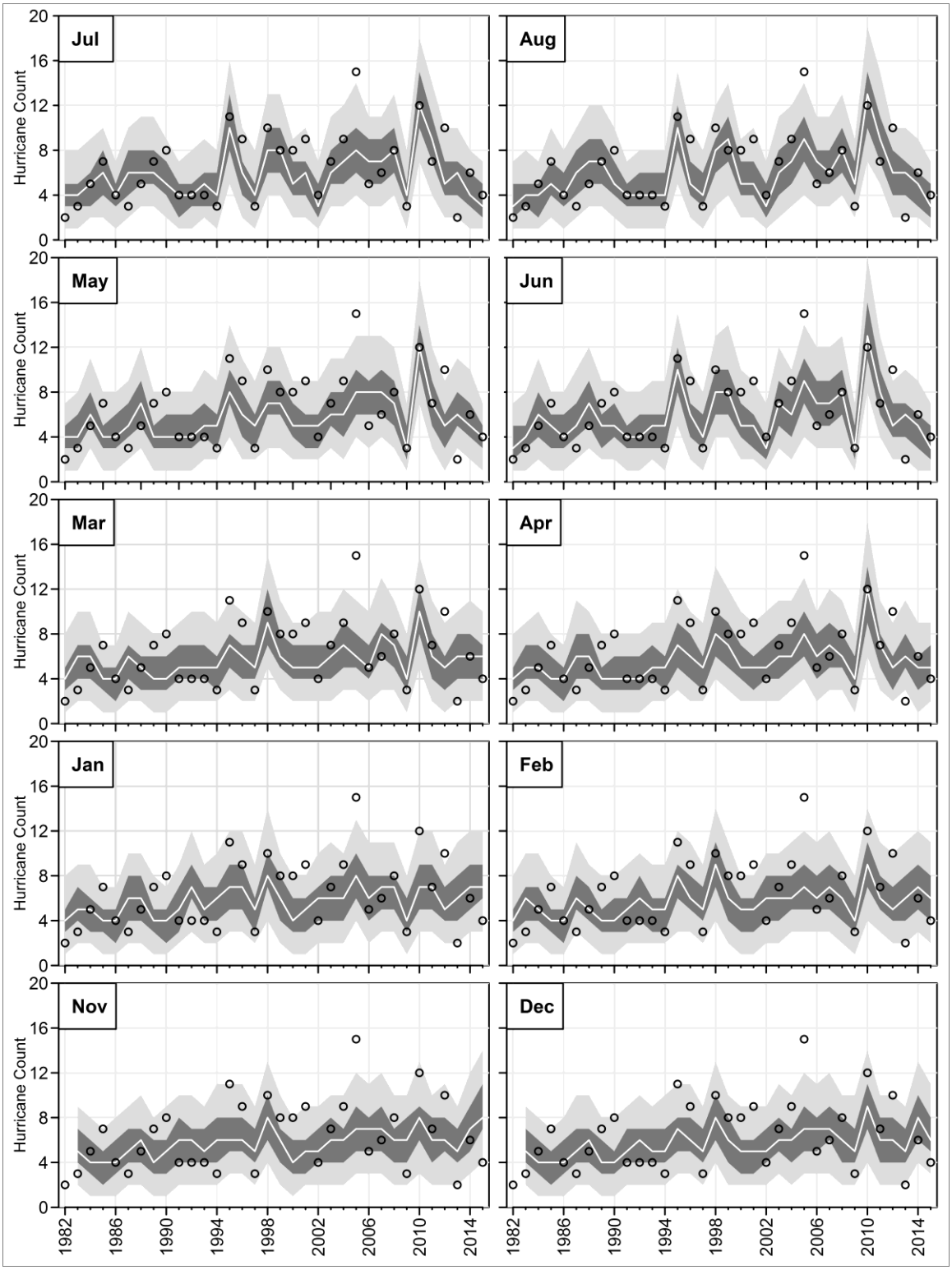

Figure A-8: Same as in Figure A-1, but using model in which the predictors are equally weighted without COLA.

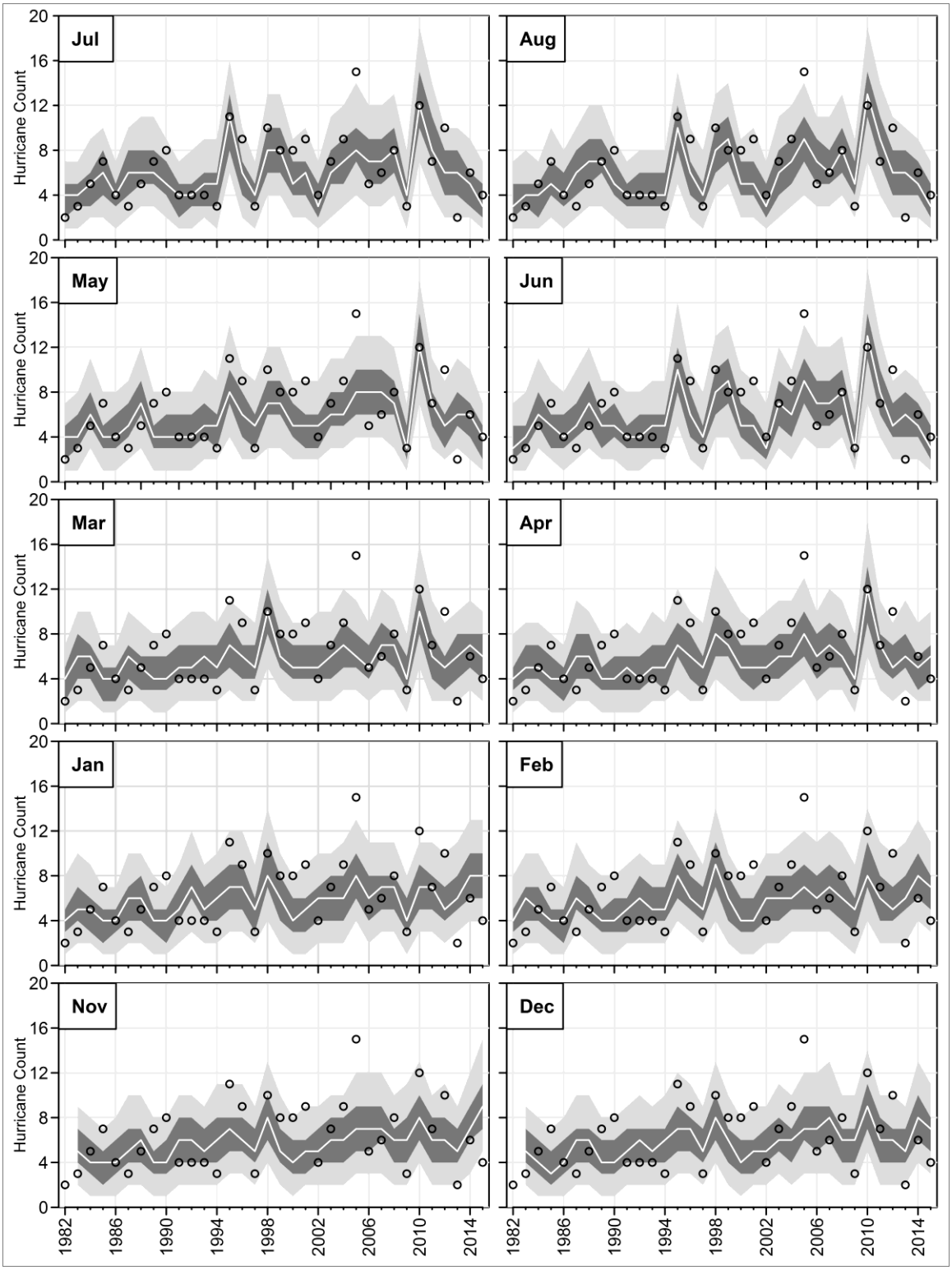

Figure A-9: Same as in Figure A-1, but using model in which the predictors are weighted without COLA.

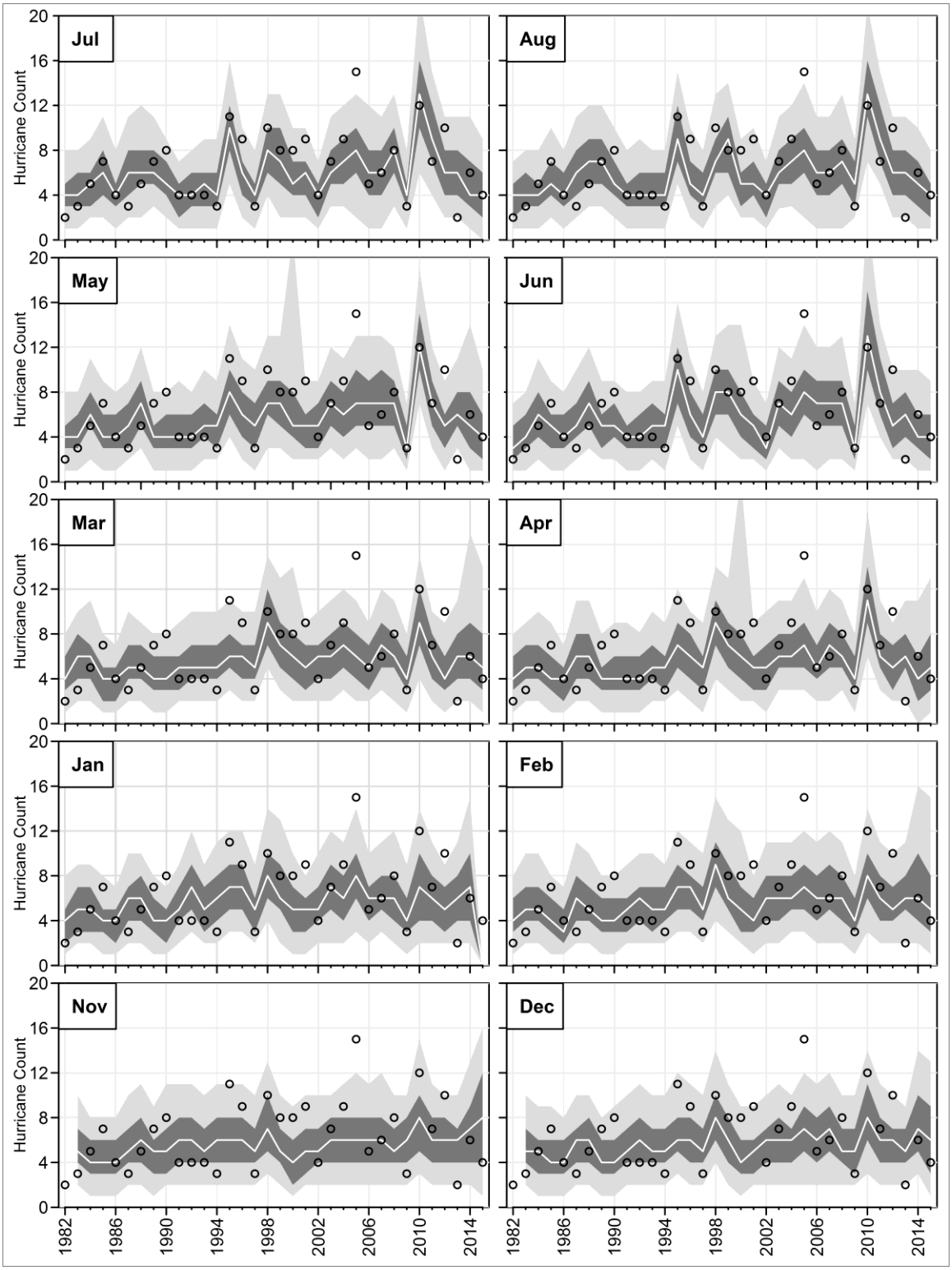

Figure A-10: Same as in Figure A-1, but using Hierarchical model.

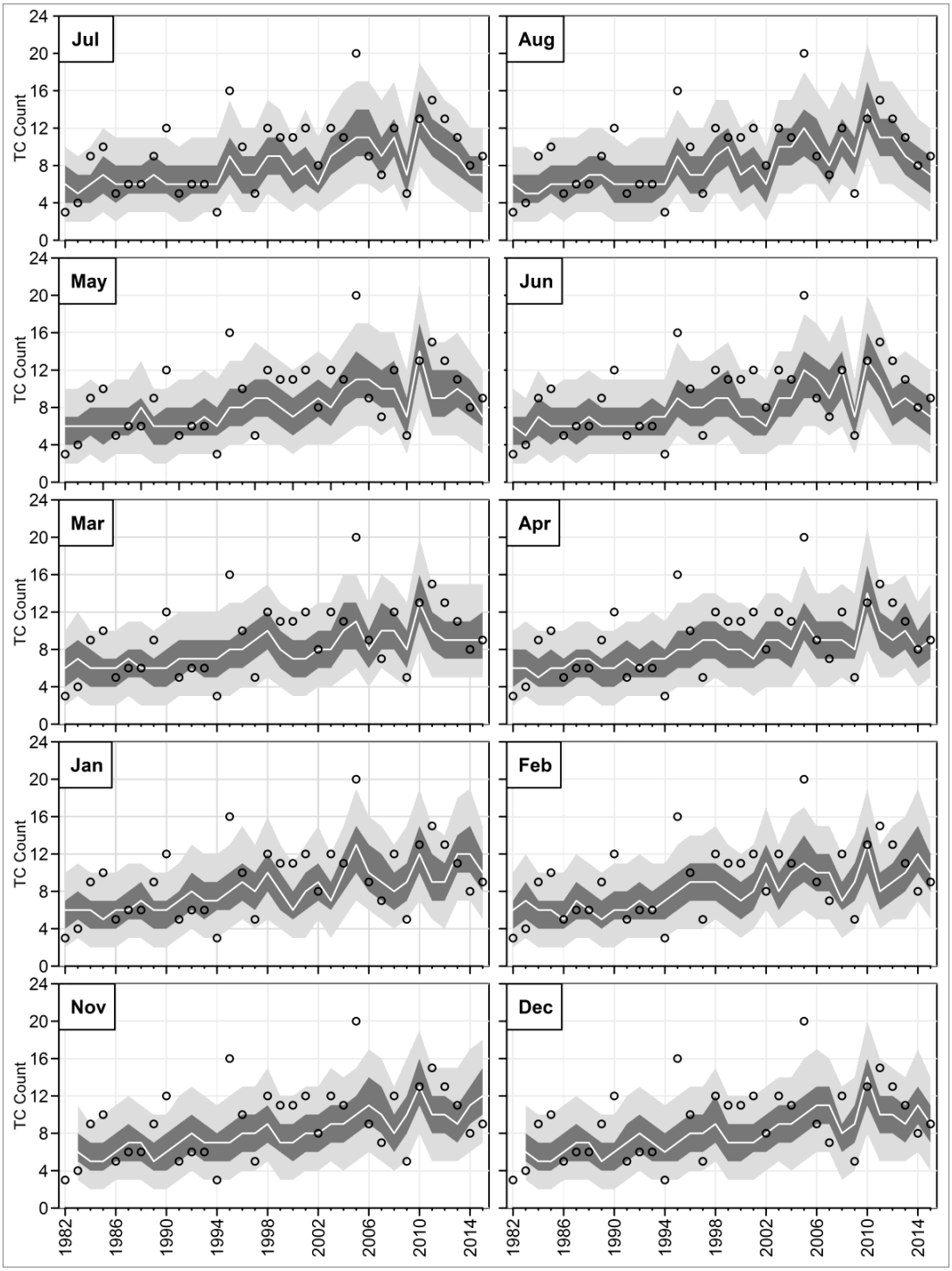

Figure A-11: Retrospective seasonal forecast of TC using GFDL model, initialized from November (of the previous year for each hurricane season) to August (of the same year). In each panel, the circles are observed values, the white line is the median (50th percentile) for each season. The outer light gray area represents the region between the 5th and 95th percentiles, while the middle dark gray area represents the region between the 25th and 75th percentiles.

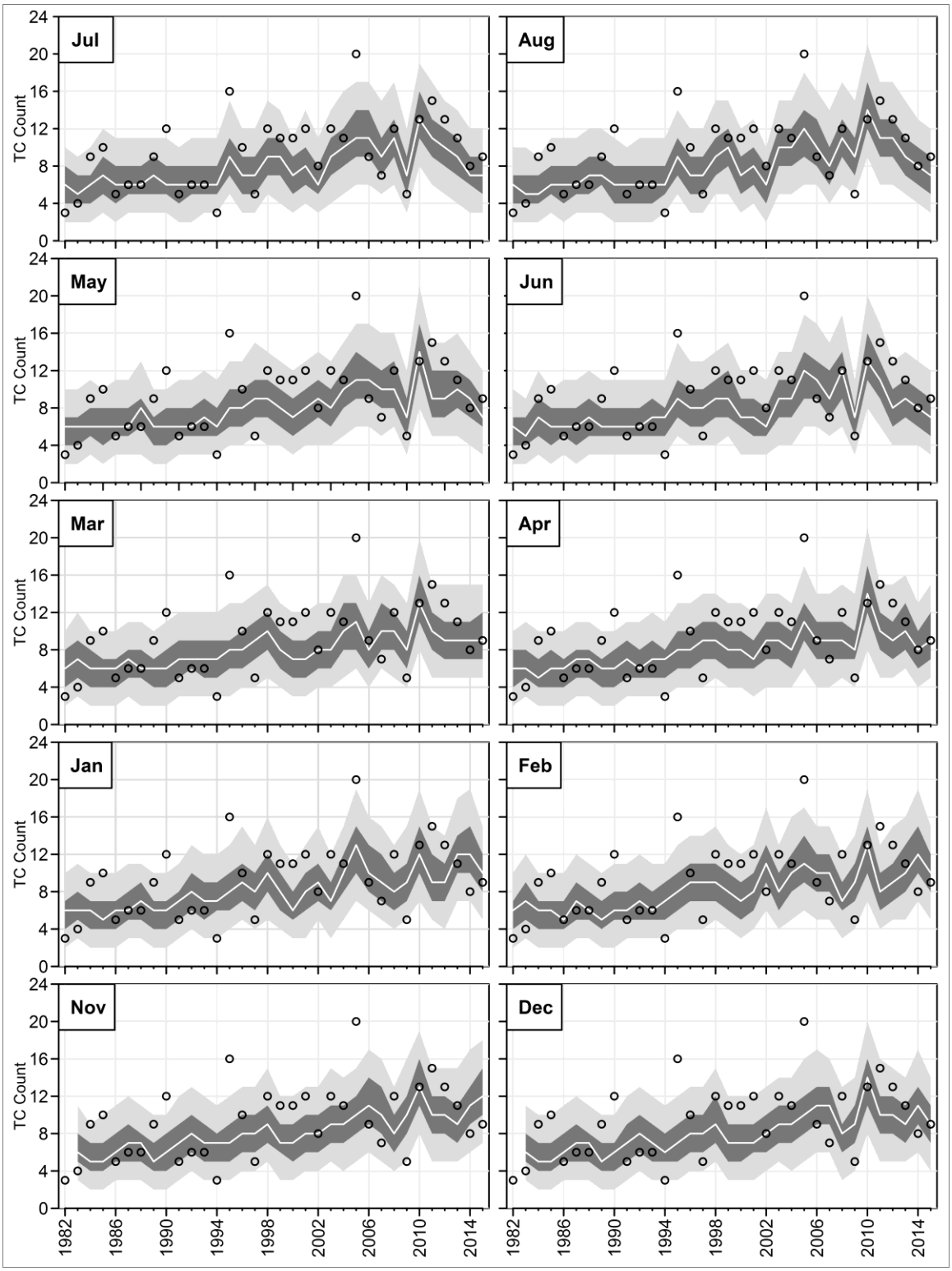

Figure A-12: Same as in Figure A-11, but using GFDL-B01.

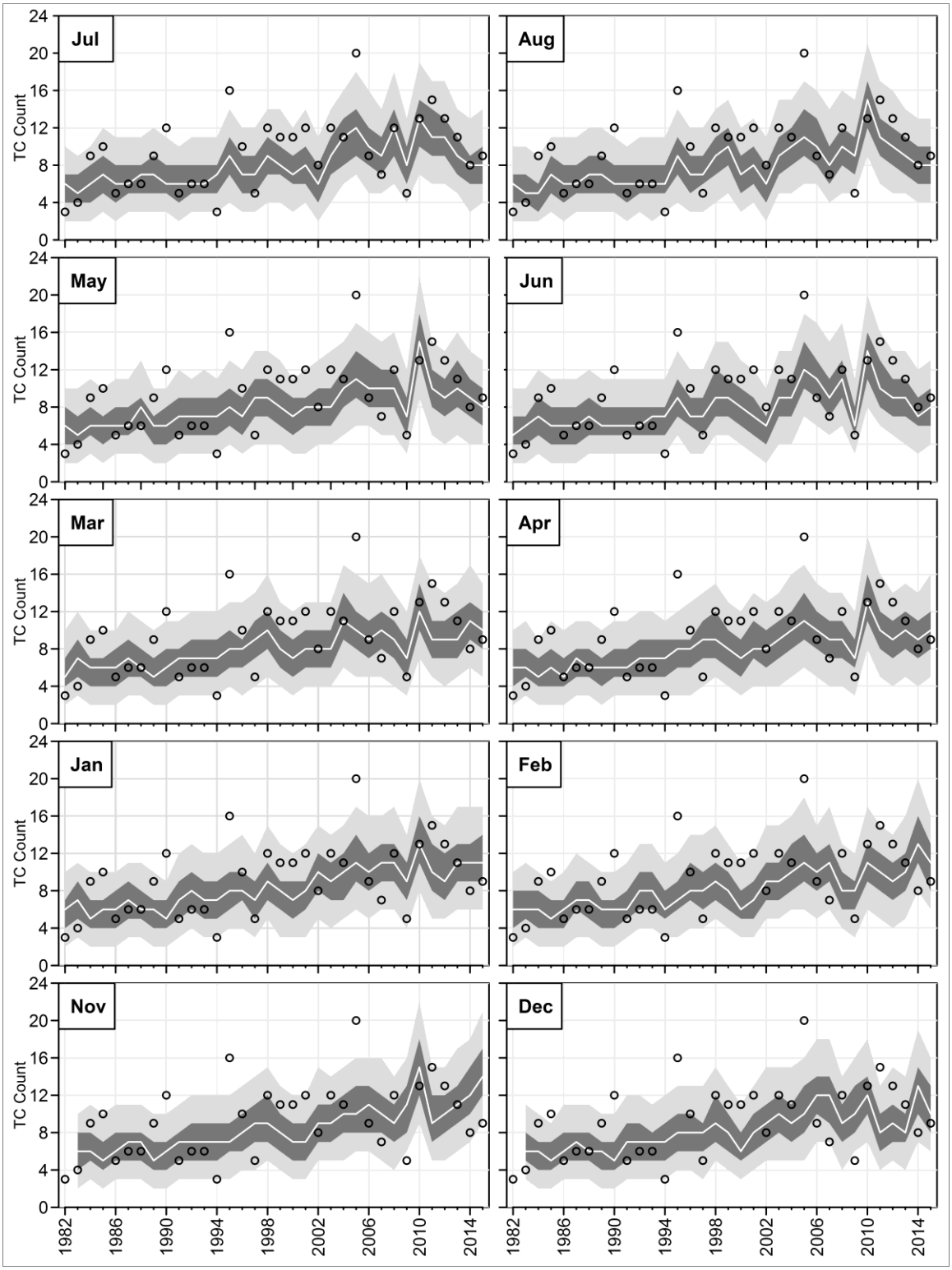

Figure A-13: Same as in Figure A-11, but using GFDL-A06.

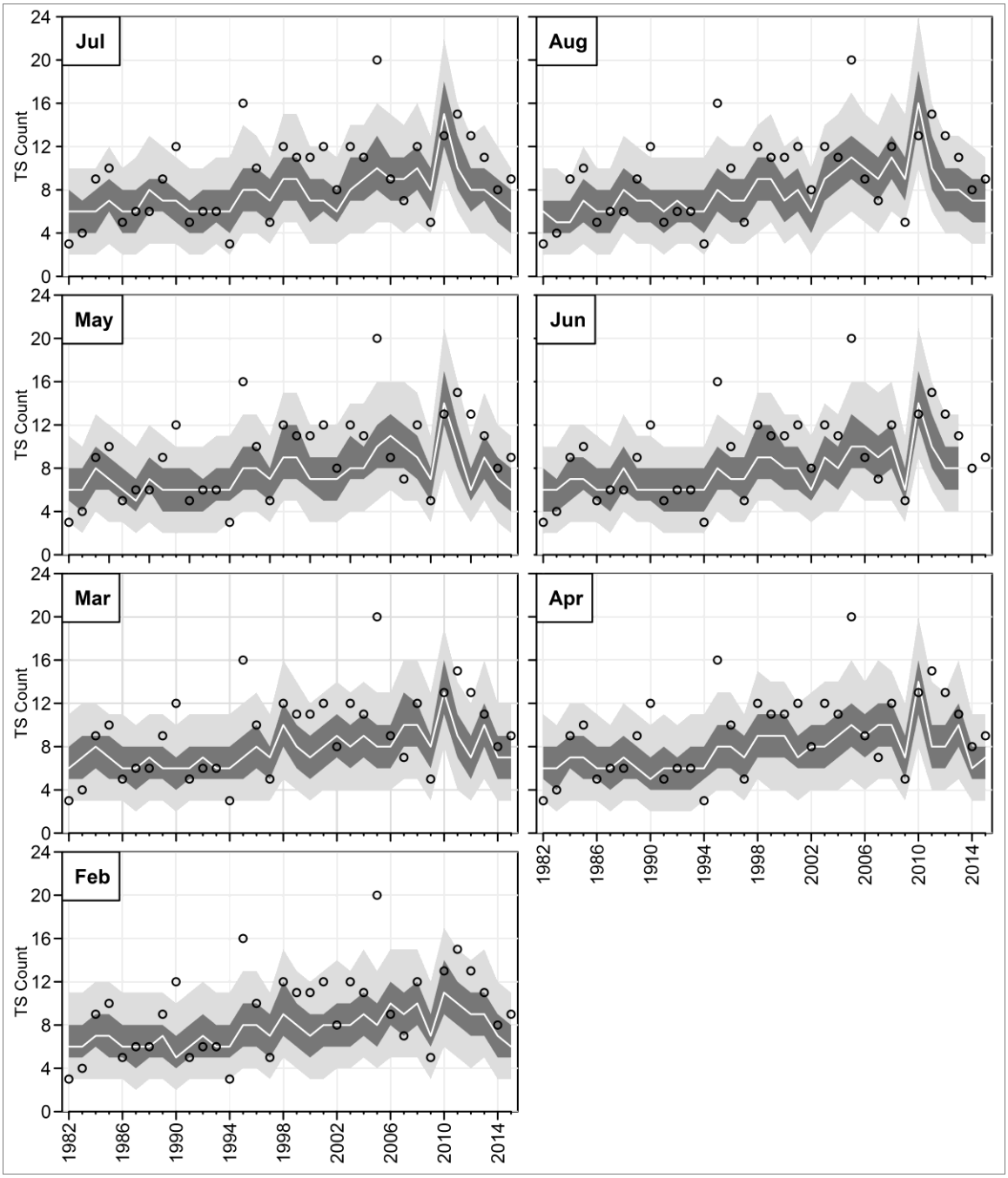

Figure A-14: Same as in Figure A-11, but using NASA.

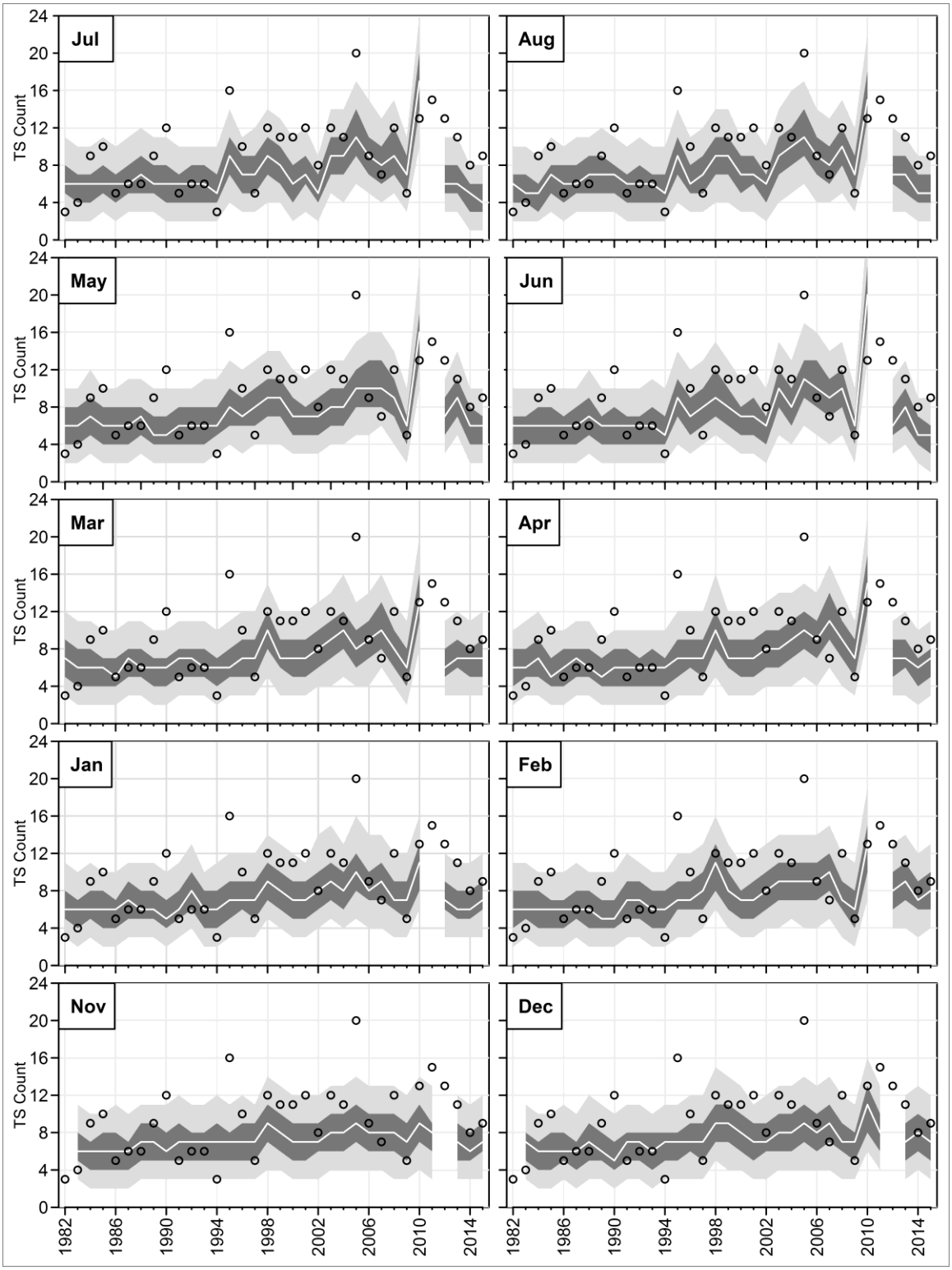

Figure A-15: Same as in Figure A-11, but using CMC2.

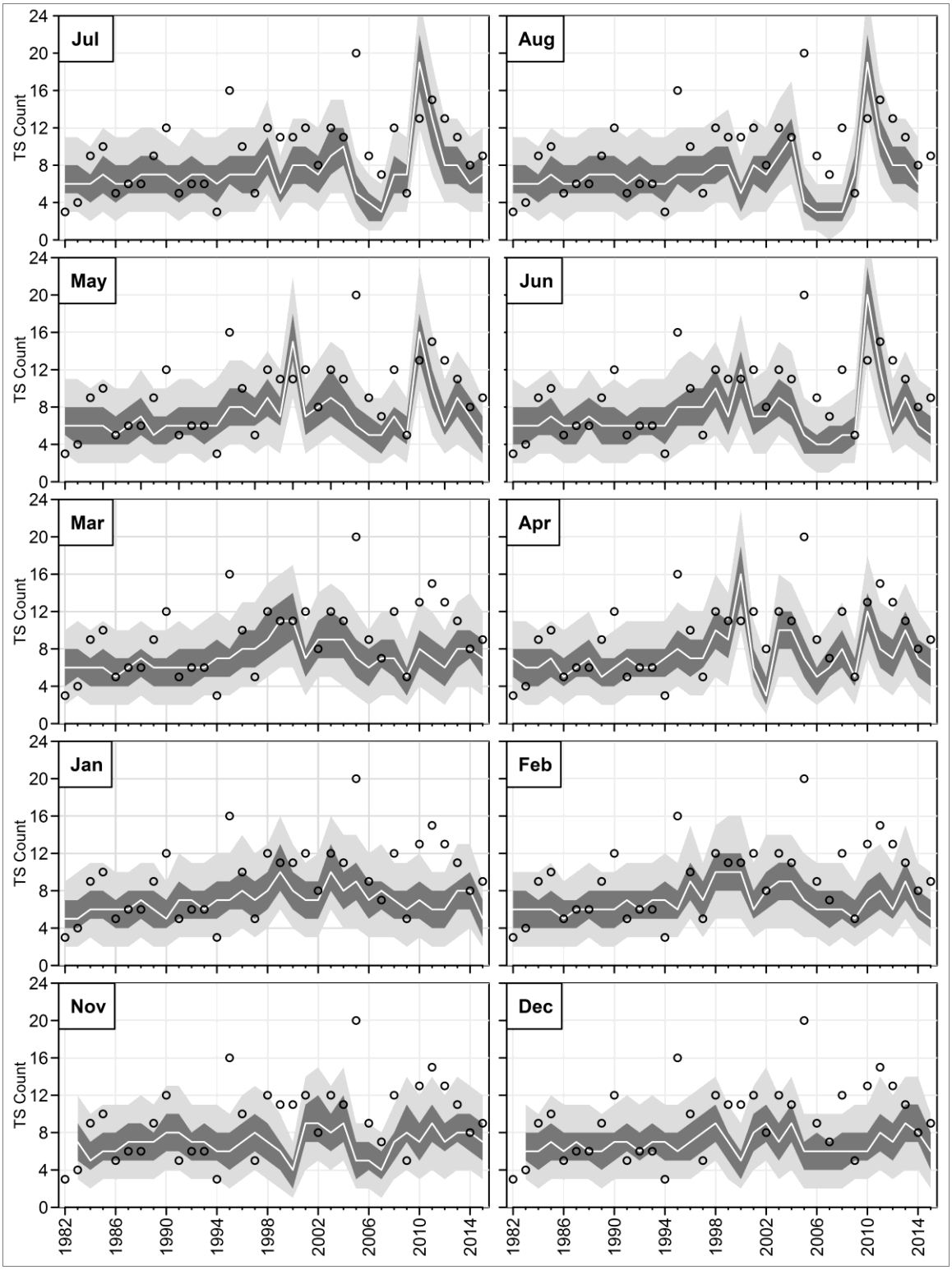

Figure A-16: Same as in Figure A-11, but using COLA.

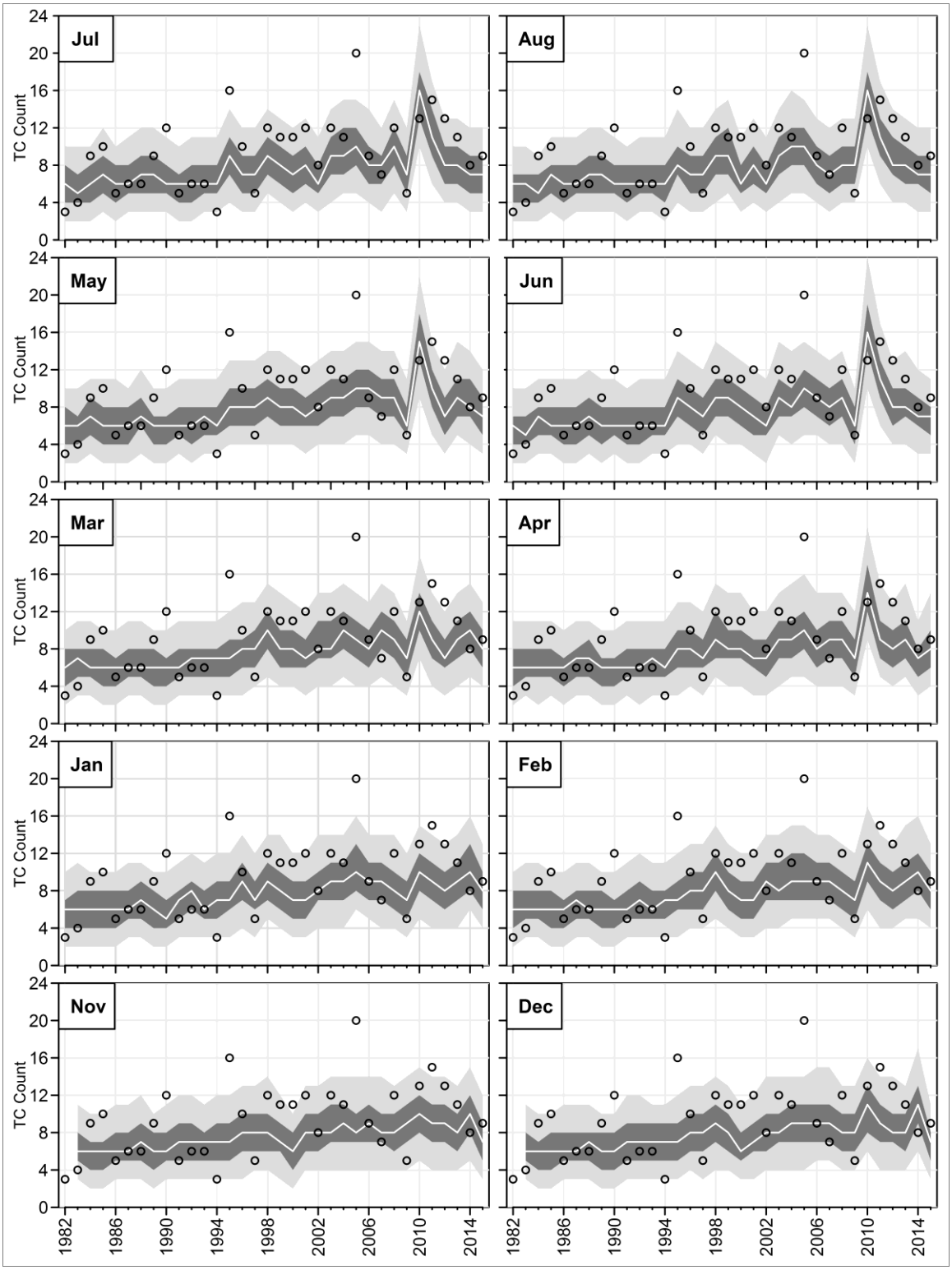

Figure A-17: Same as in Figure A-11, but using model in which the predictors are equally weighted.

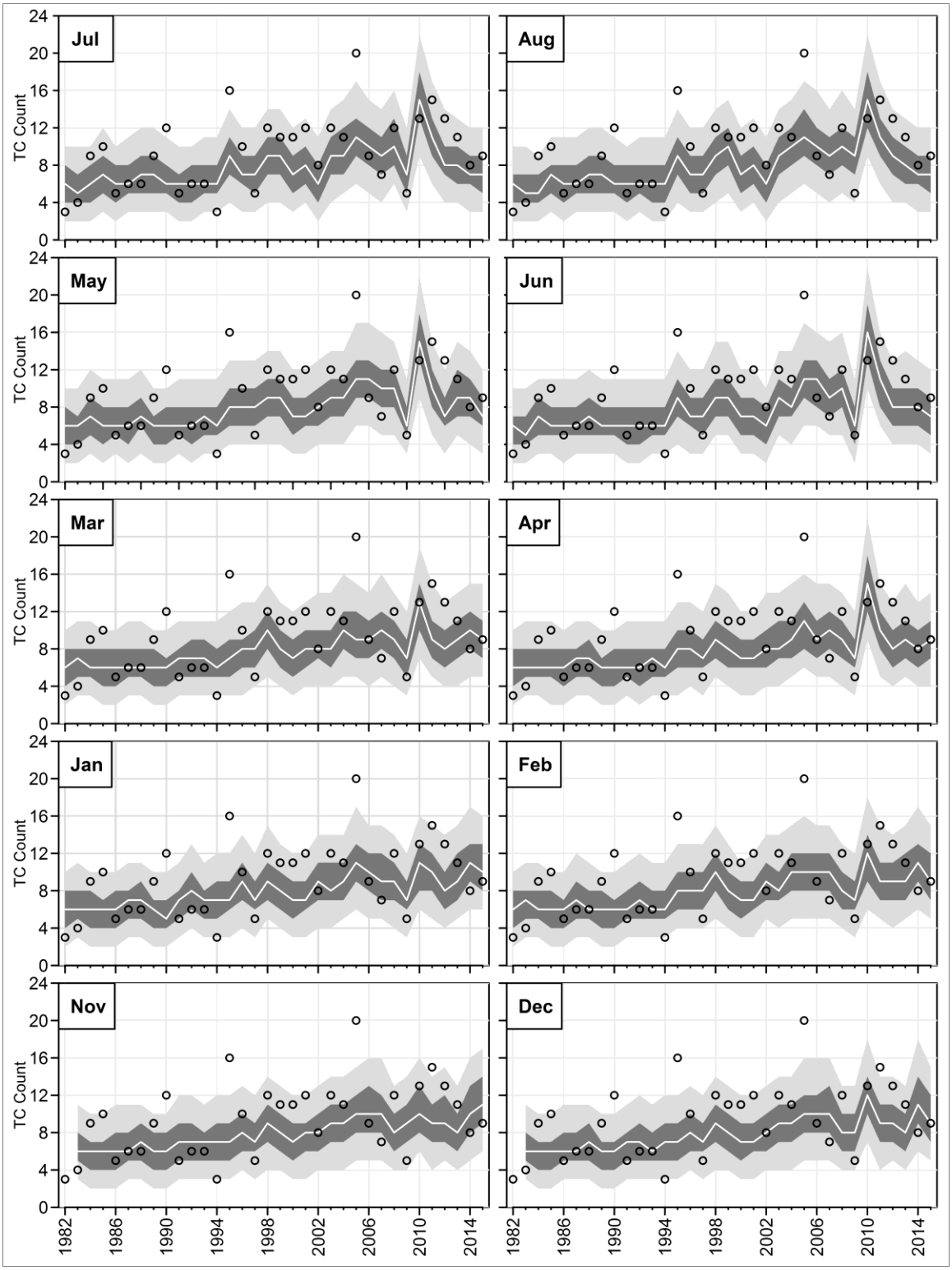

Figure A-18: Same as in Figure A-11, but using model in which the predictors are equally weighted without COLA.

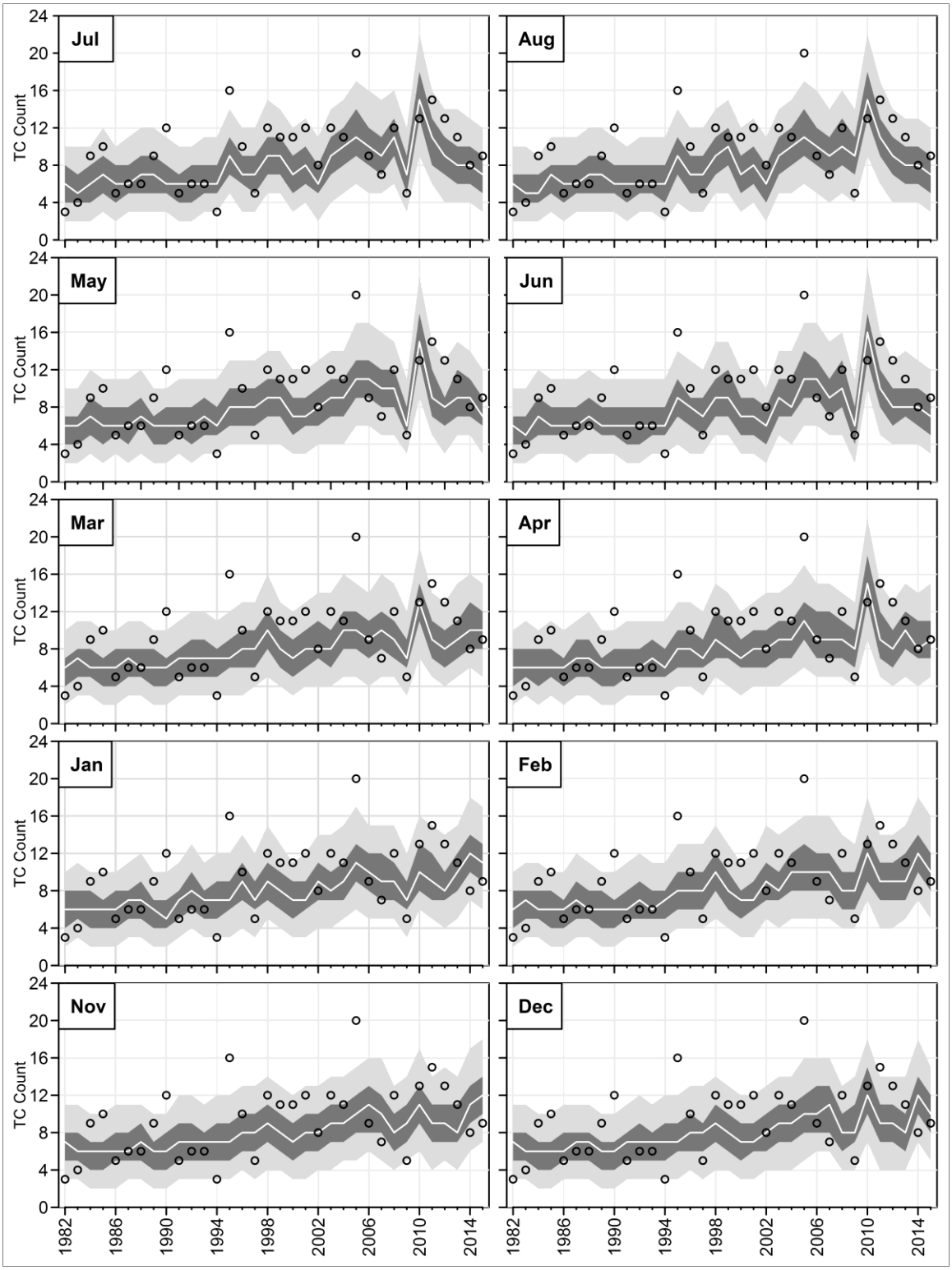

Figure A-19: Same as in Figure A-11, but using model in which the predictor are weighted without COLA.

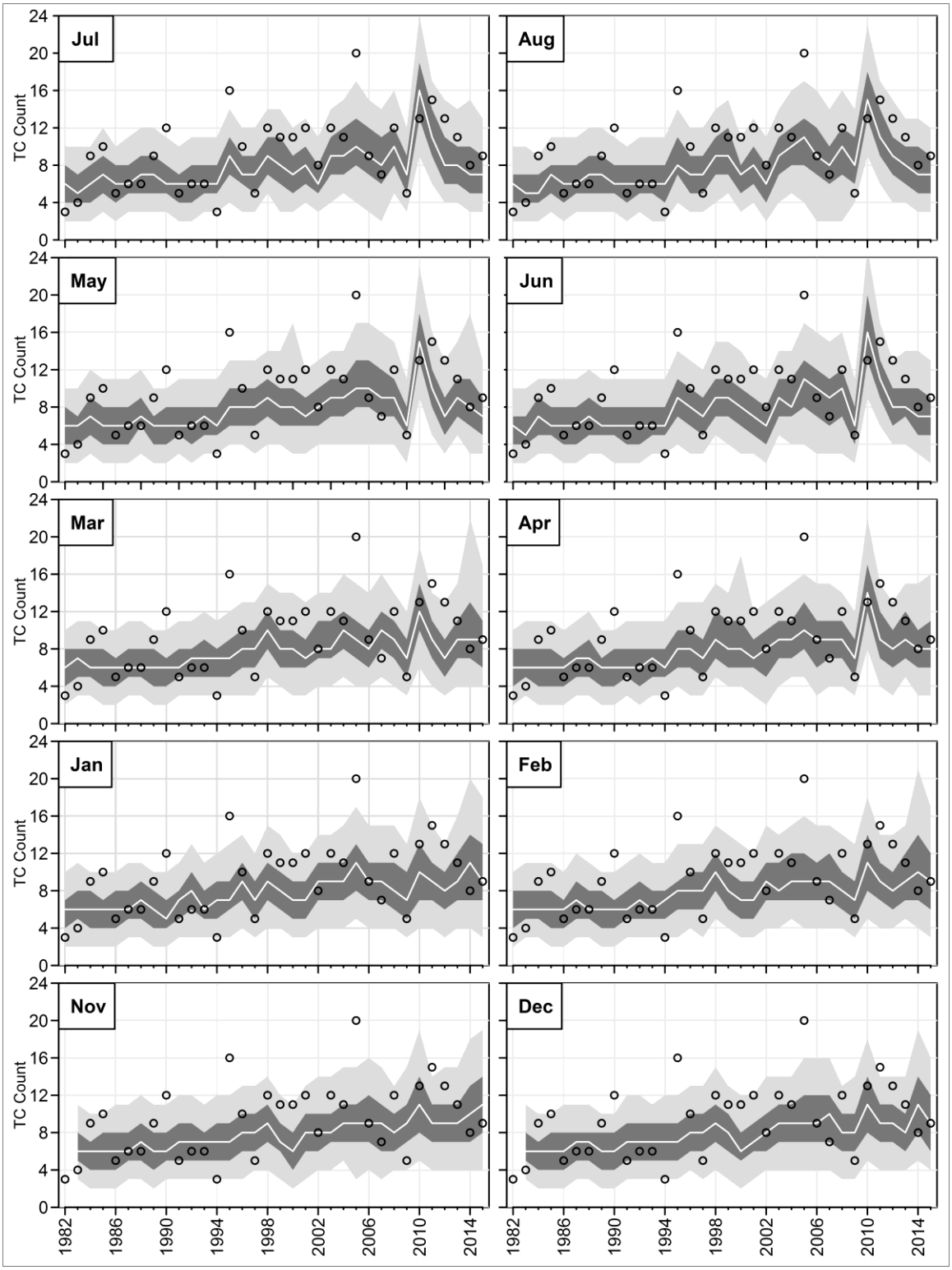

Figure A-20: Same as in Figure A-11, but using Hierarchical Model.

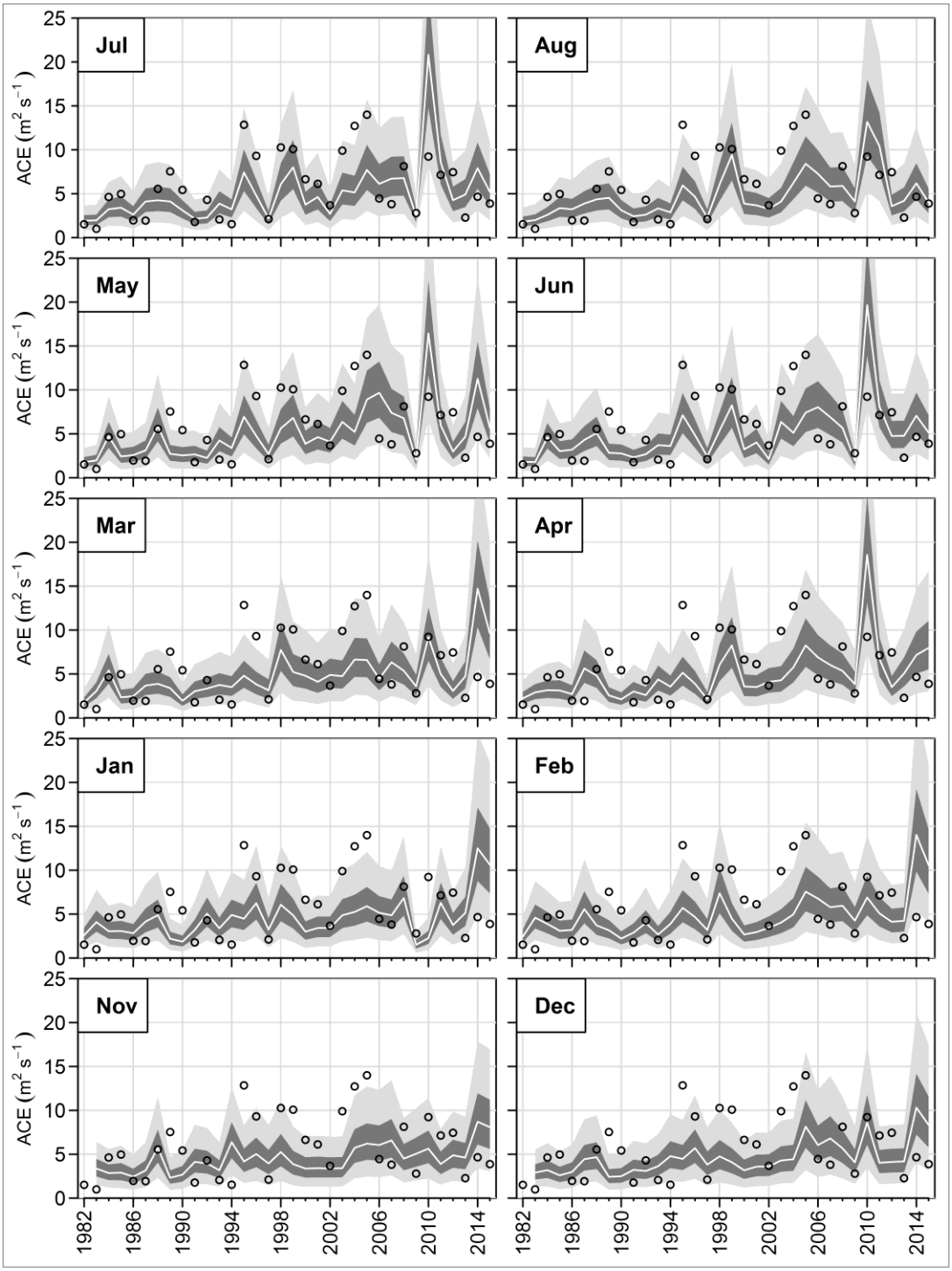

Figure A-21: Retrospective seasonal forecast of ACE using GFDL model, initialized from November (of the previous year for each hurricane season) to August (of the same year). In each panel, the circles are observed values, the white line is the median (50th percentile) for each season. The outer light gray area represents the region between the 5th and 95th percentiles, while the middle dark gray area represents the region between the 25th and 75th percentiles.

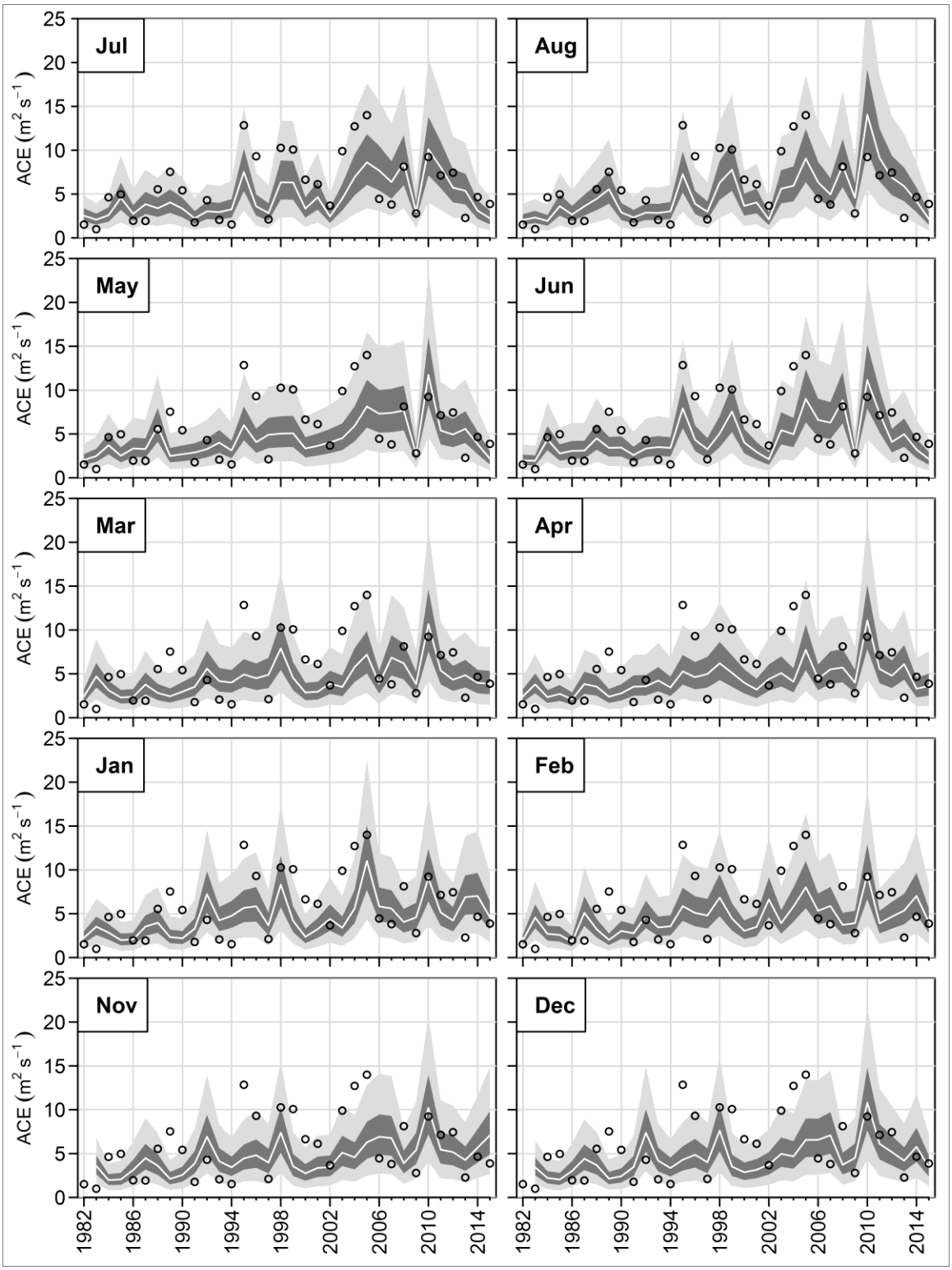

Figure A-22: Same as Figure A-21, but using GFDL-B01.

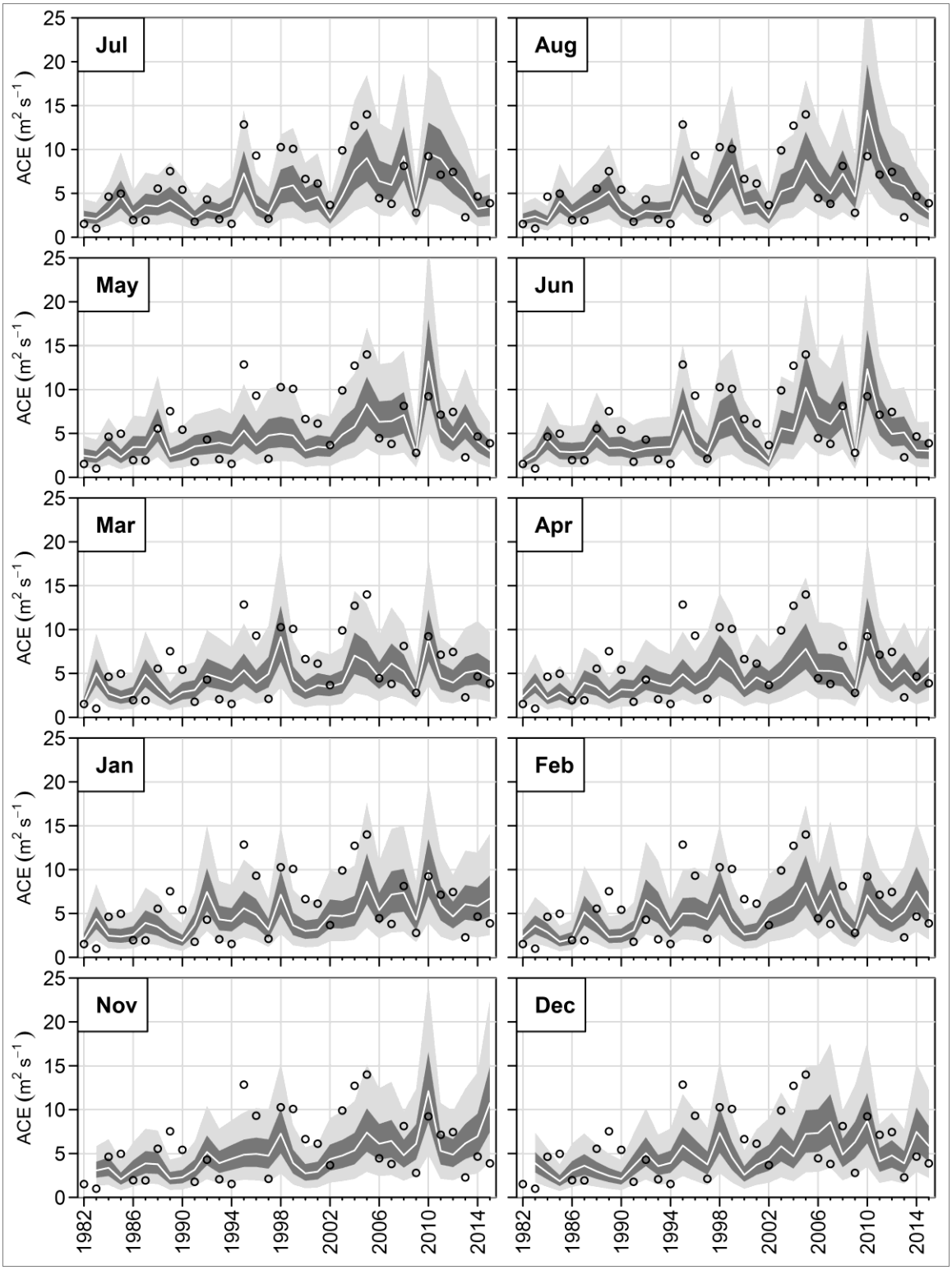

Figure A-23: Same as Figure A-21, but using GFDL-A06.

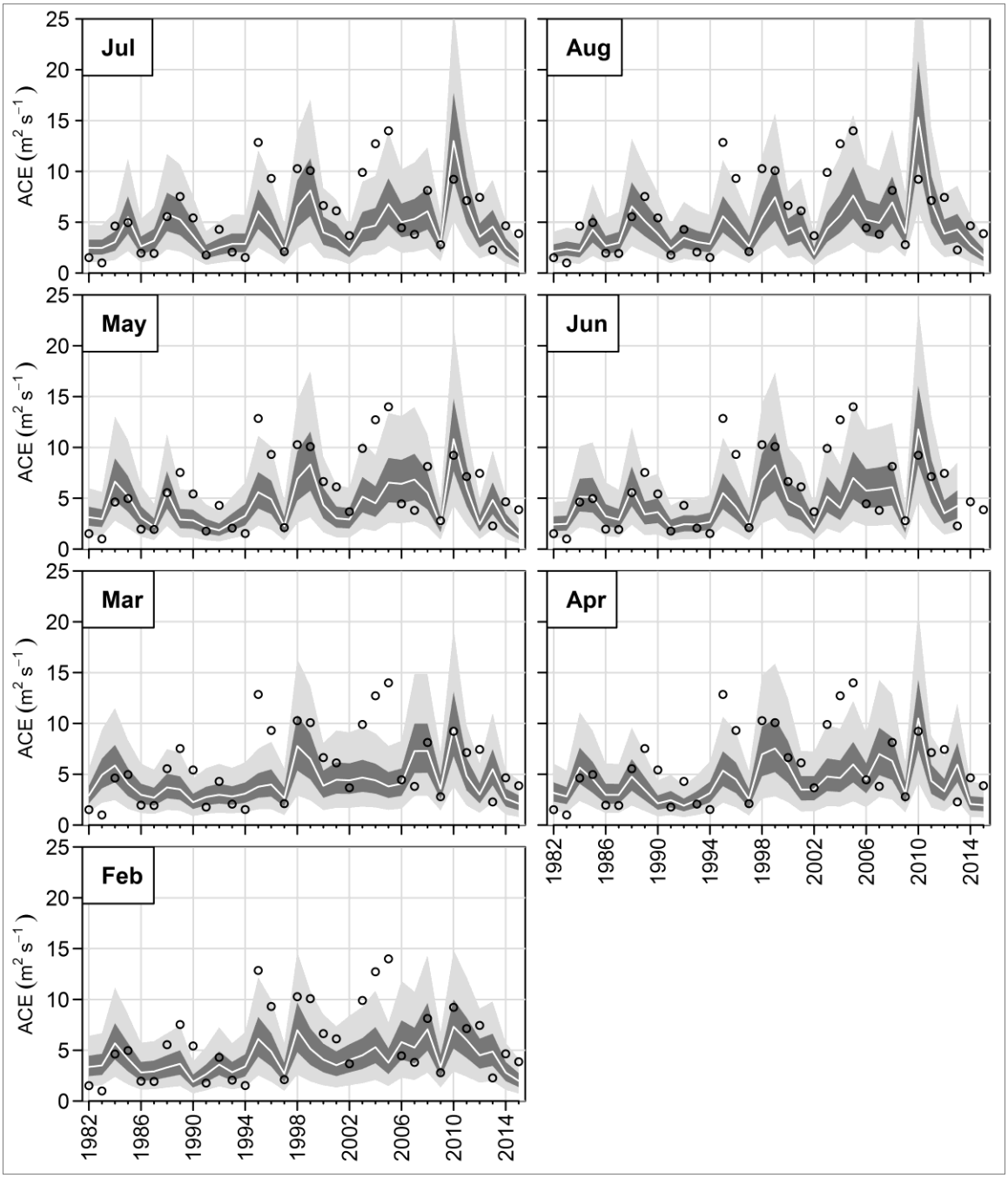

Figure A-24: Same as Figure A-21, but using NASA.

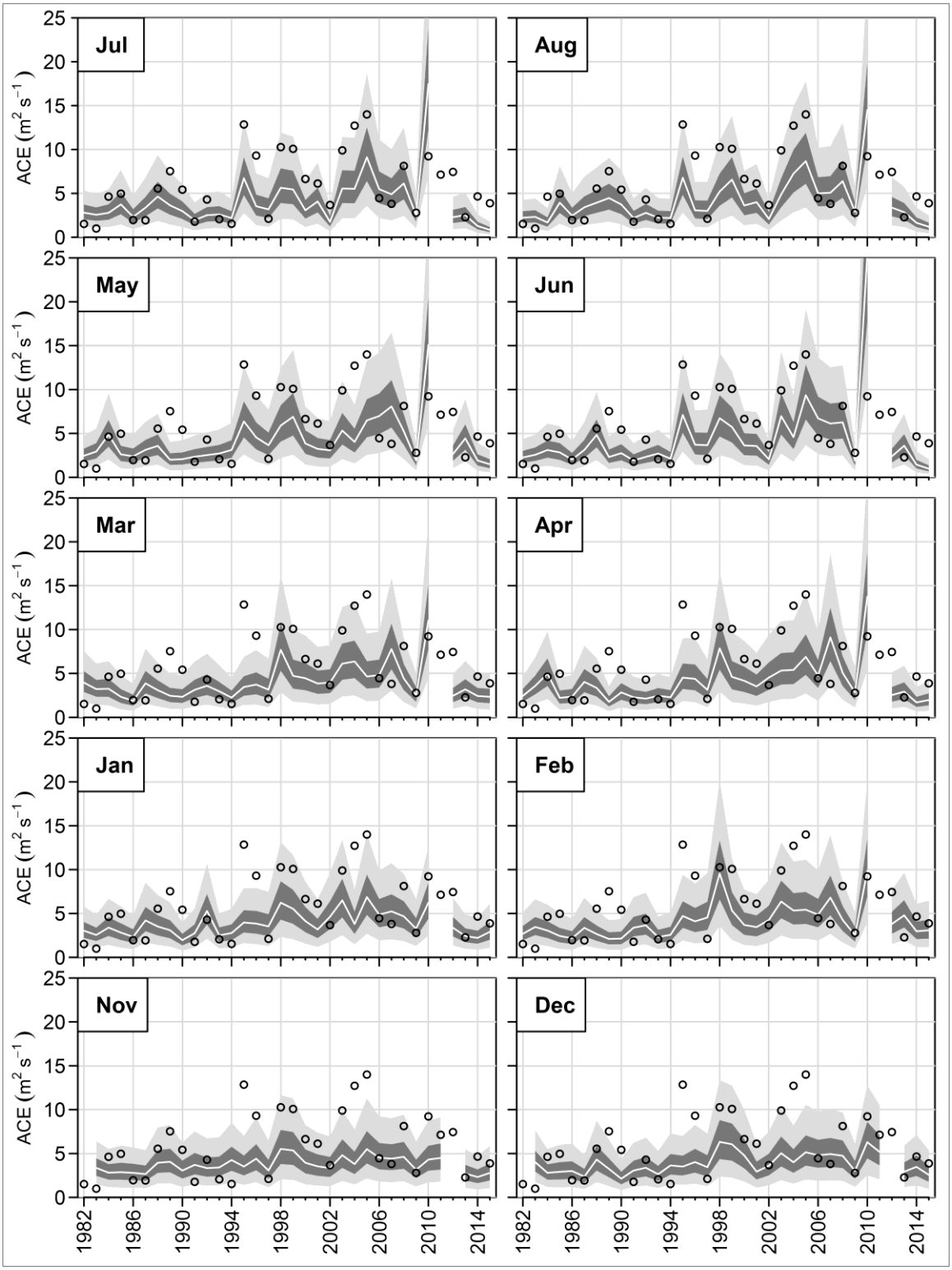

Figure A-25: Same as Figure A-21, but using CMC2.

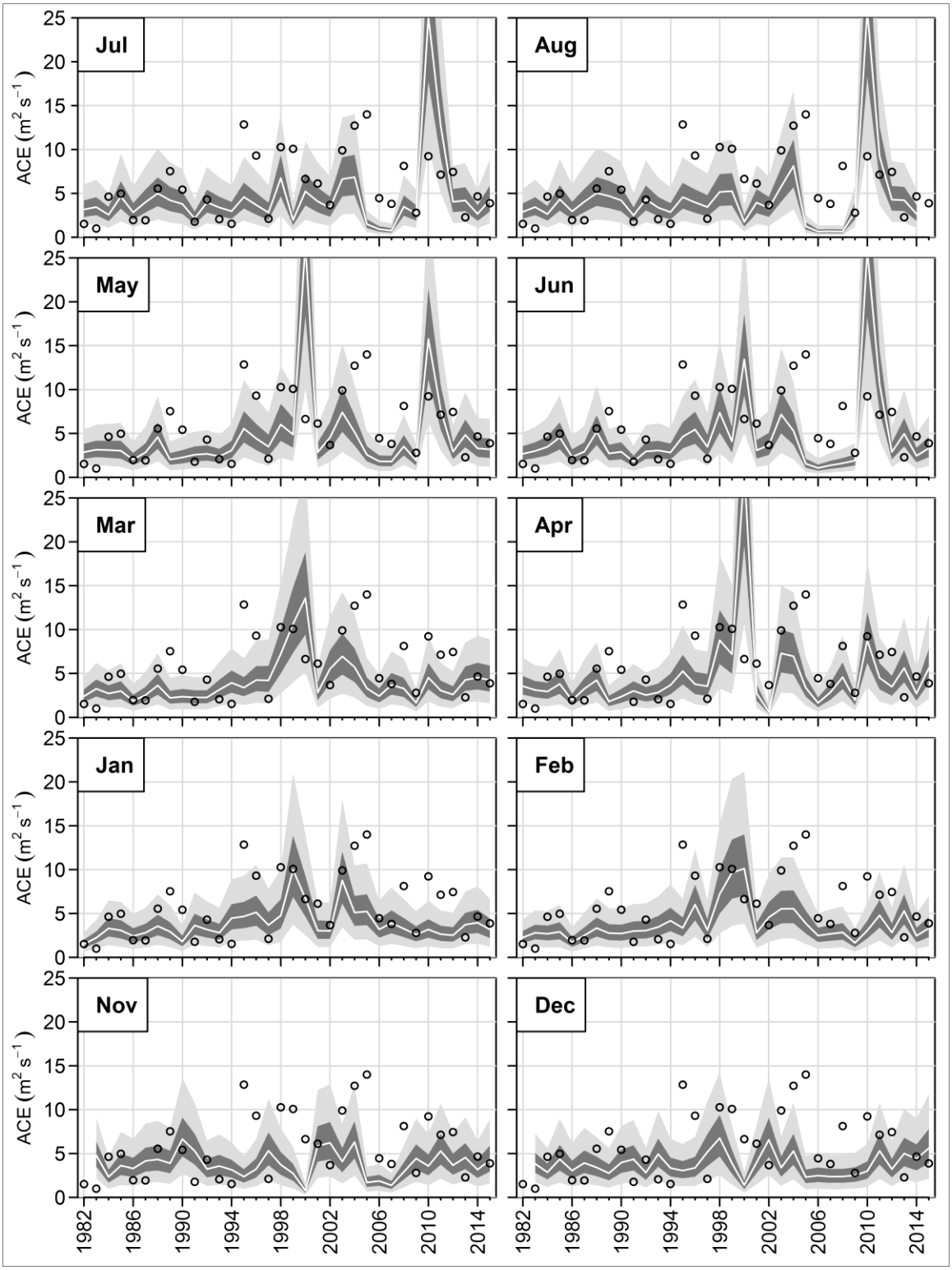

Figure A-26: Same as Figure A-21, but using COLA.

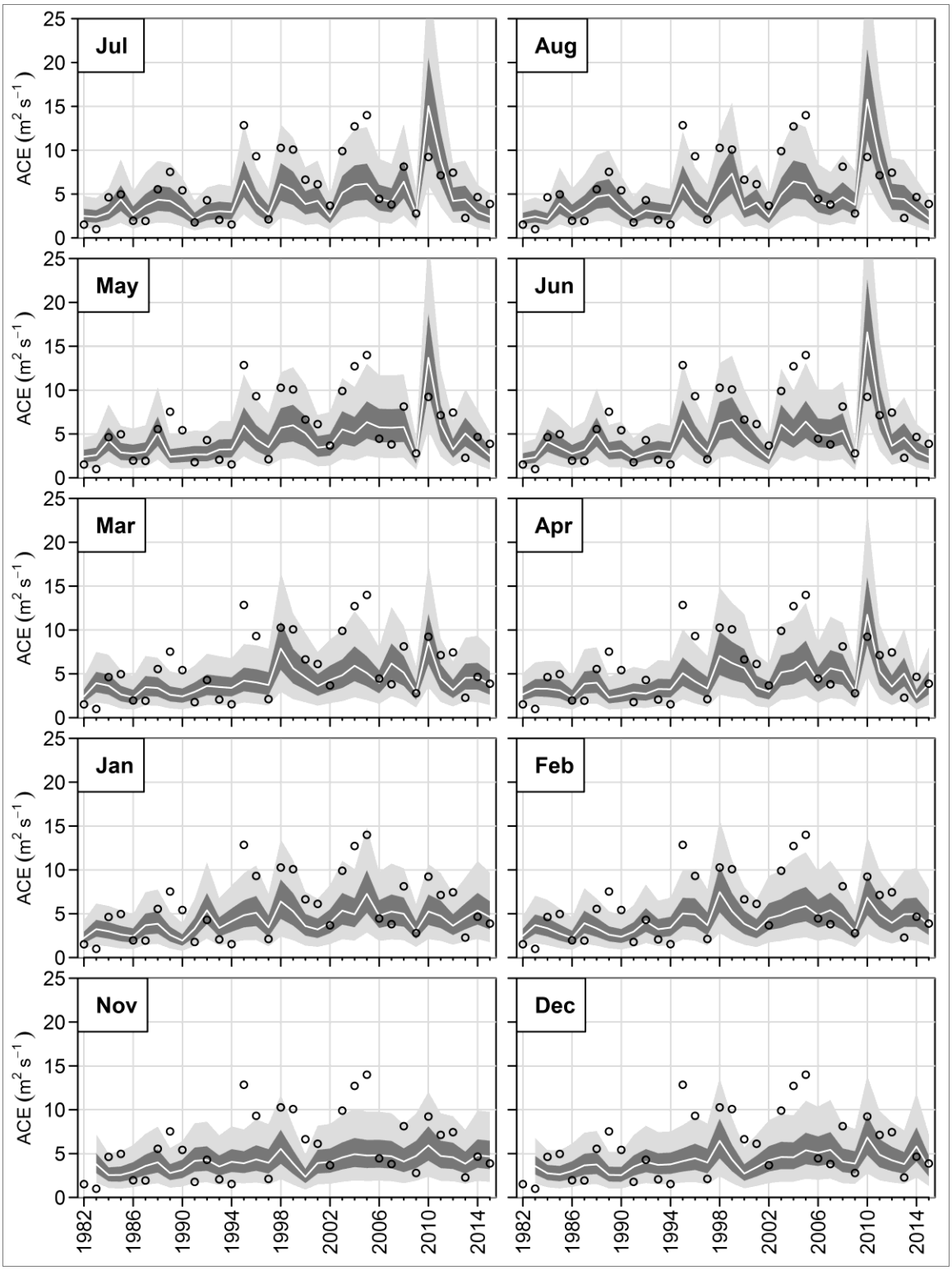

Figure A-27: Same as Figure A-21, but using model in which predictors are equally weighted.

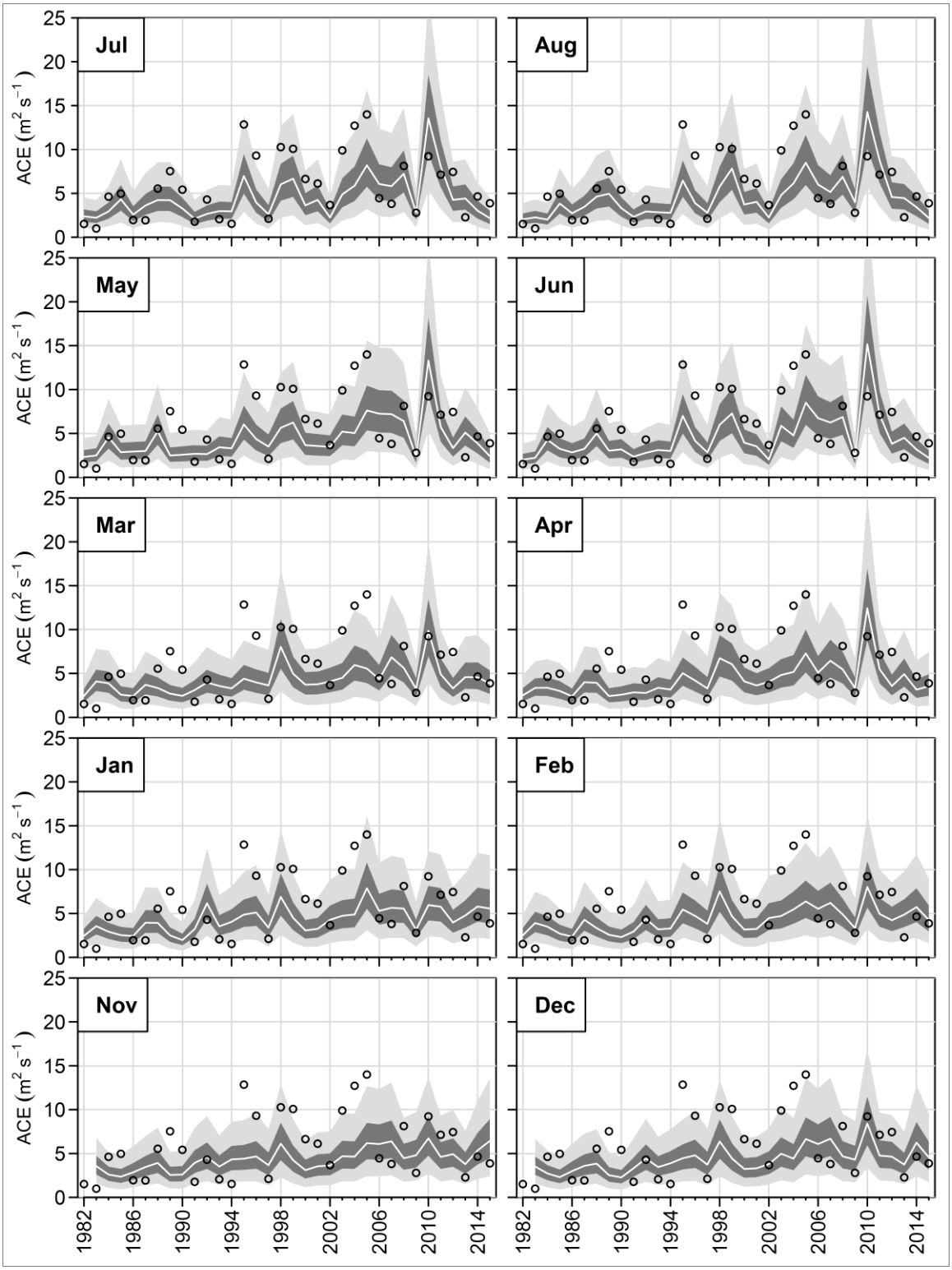

Figure A-28: Same as Figure A-21, but using model in which the predictors are equally weighted without COLA.

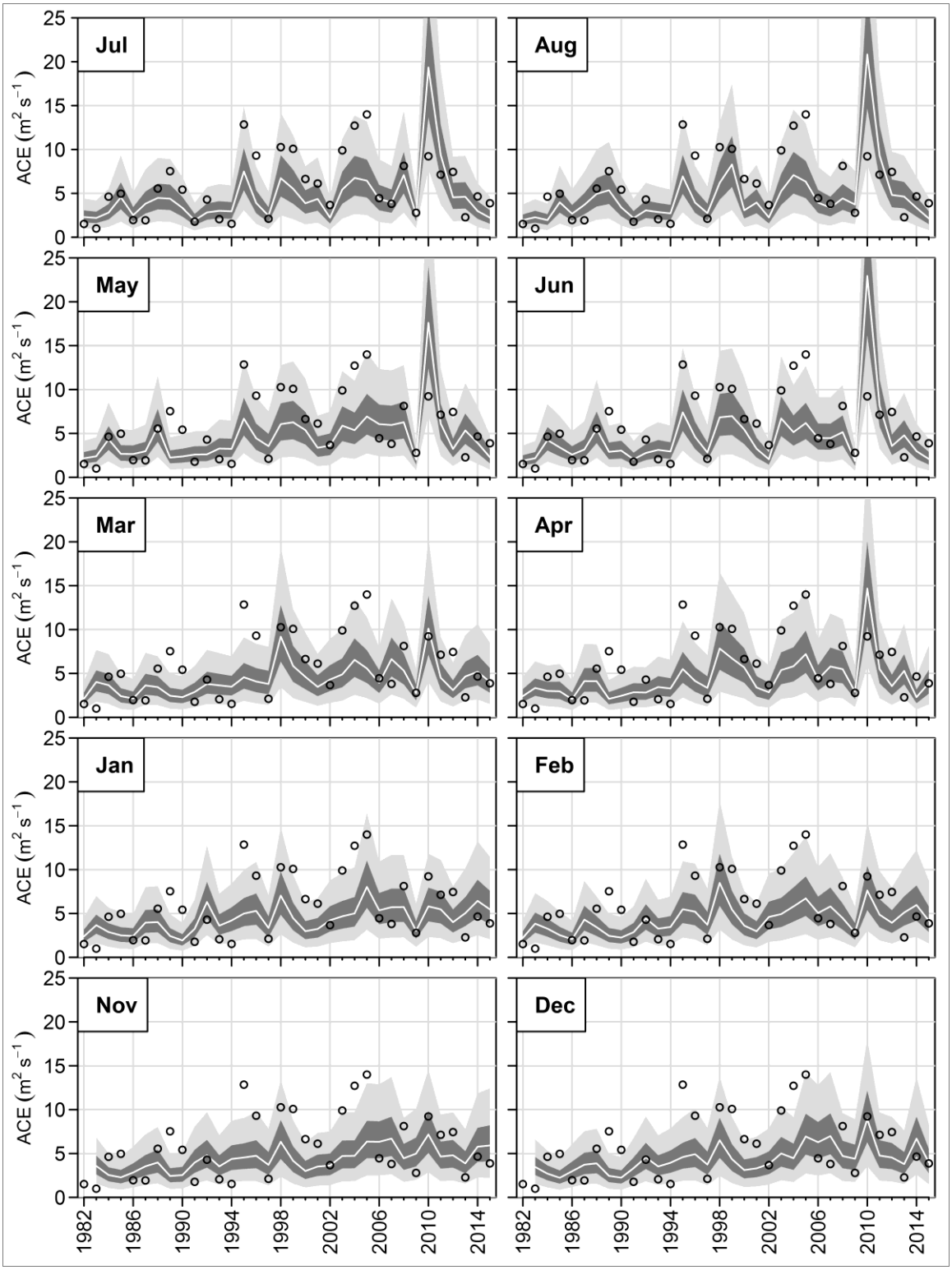

Figure A-29: Same as Figure A-21, but using model in which the predictors are weighted without COLA.

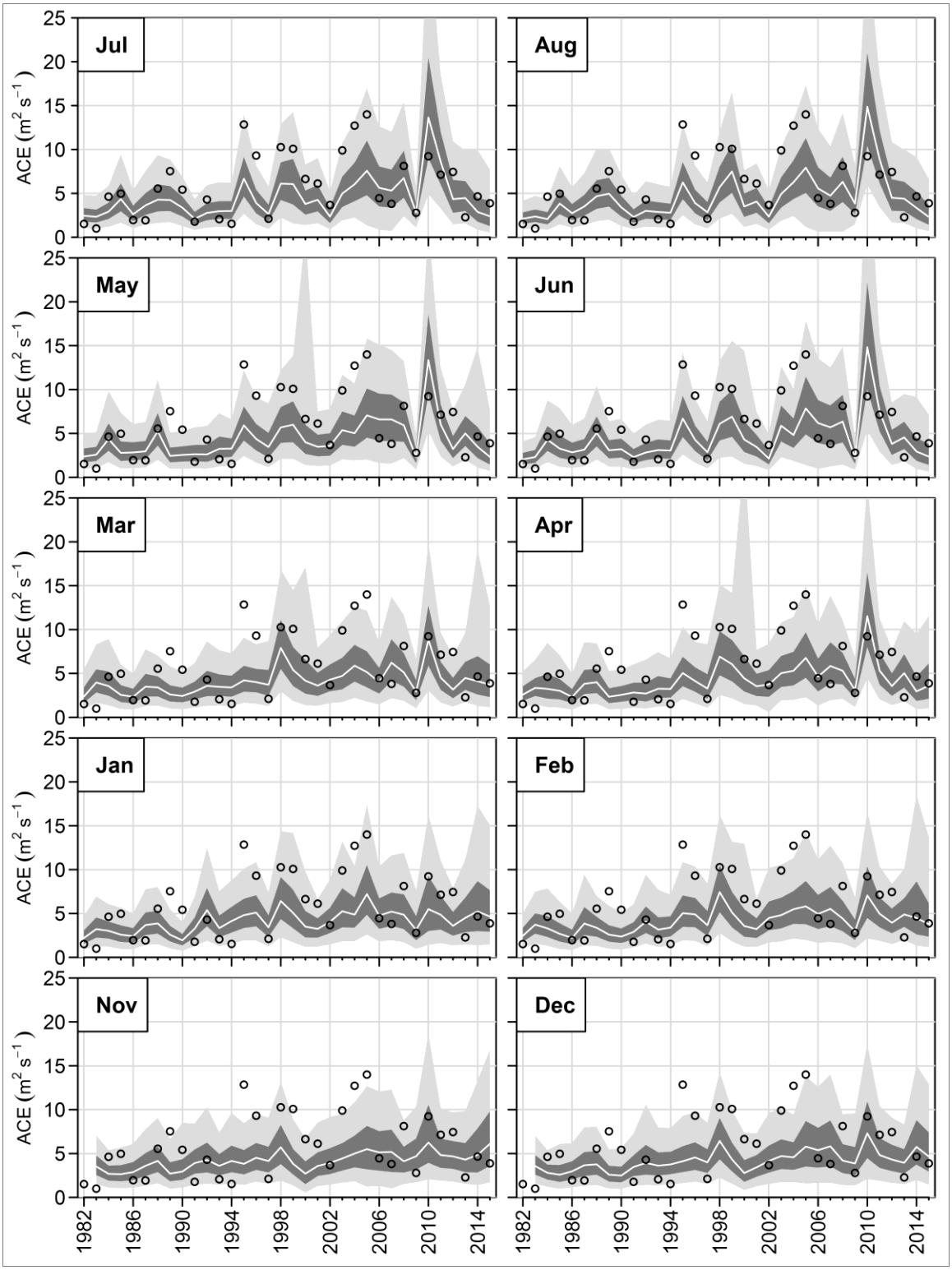

Figure A-30: Same as Figure A-21, but using Hierarchical model.

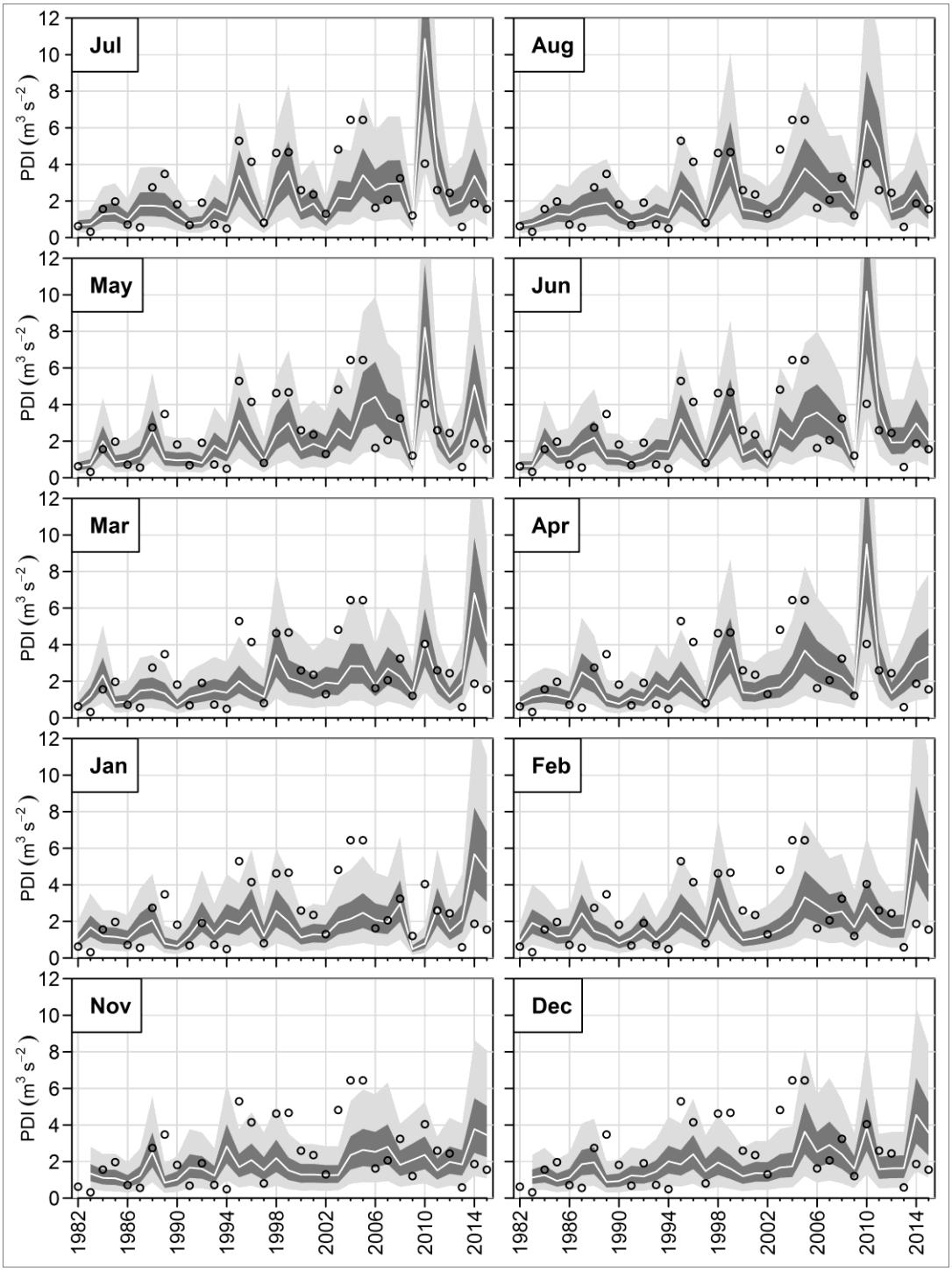

Figure A-31: Retrospective seasonal forecast of PDI using GFDL model, initialized from November (of the previous year for each hurricane season) to August (of the same year). In each panel, the circles are observed values, the white line is the median (50th percentile) for each season. The outer light gray area represents the region between the 5th and 95th percentiles, while the middle dark gray area represents the region between the 25th and 75th percentiles.

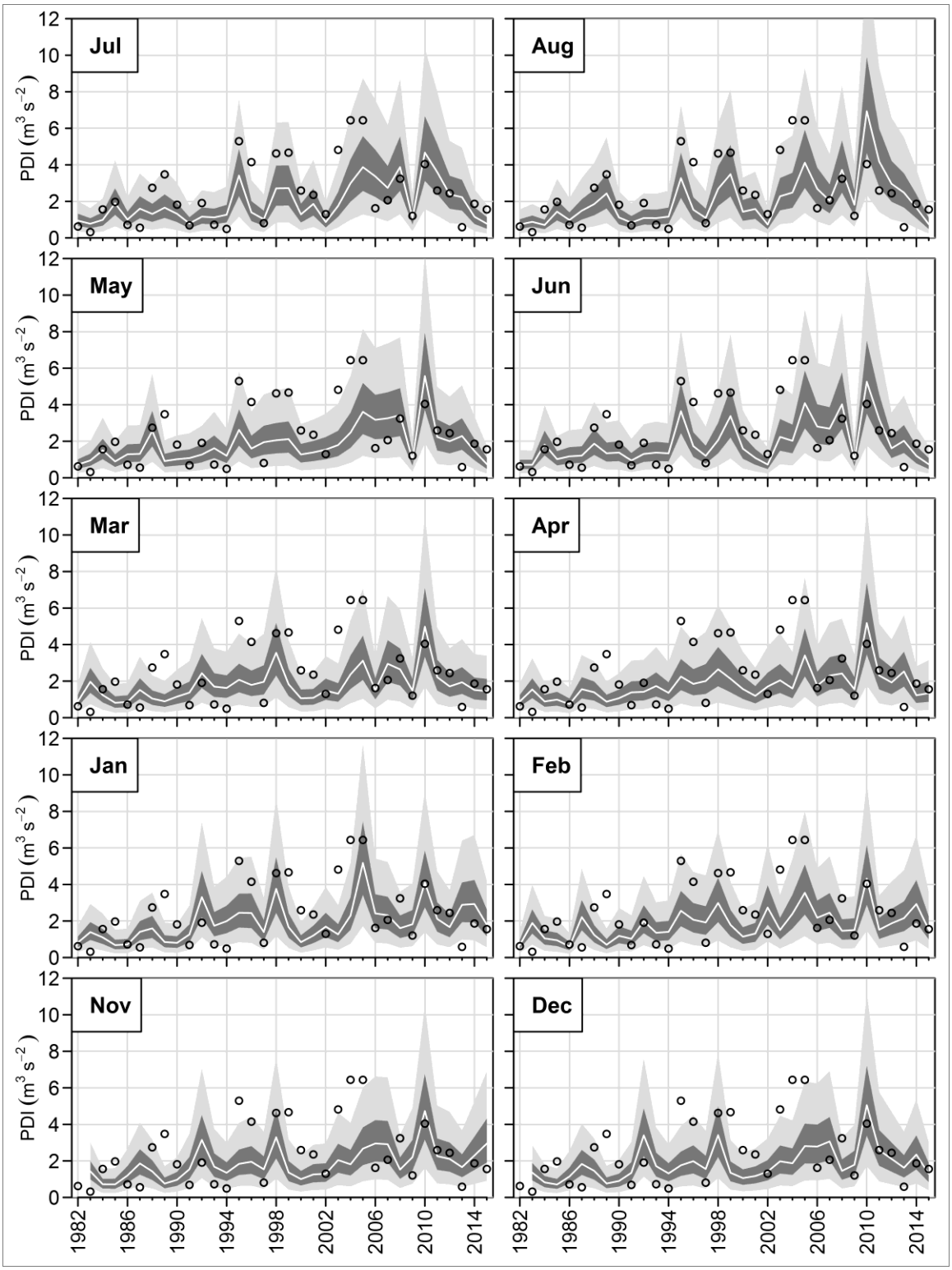

Figure A-32: Same as Figure A-31, but using GFDL-B01.

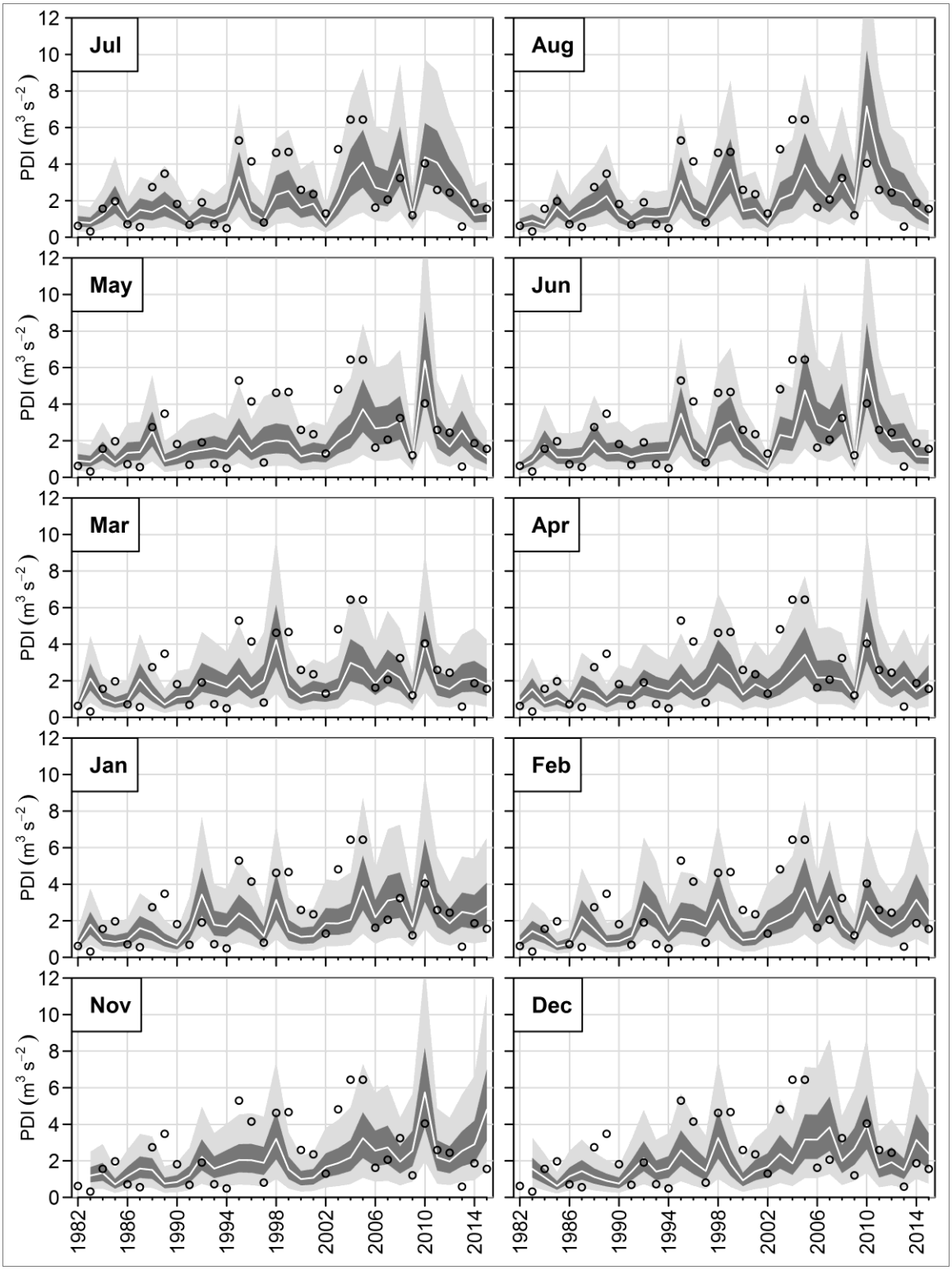

Figure A-33: Same as Figure A-31, but using GFDL-A06.

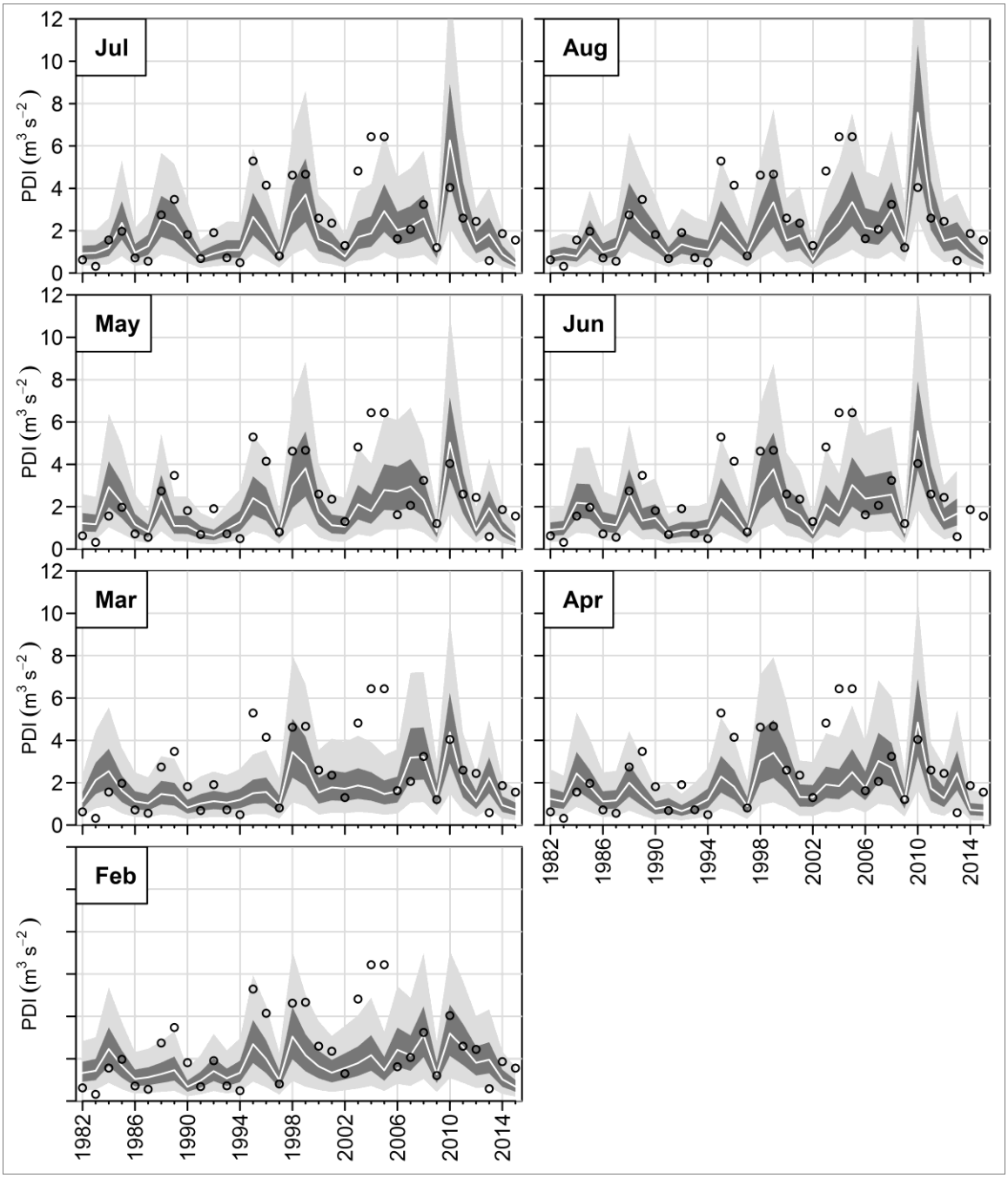

Figure A-34: Same as Figure A-31, but using NASA.

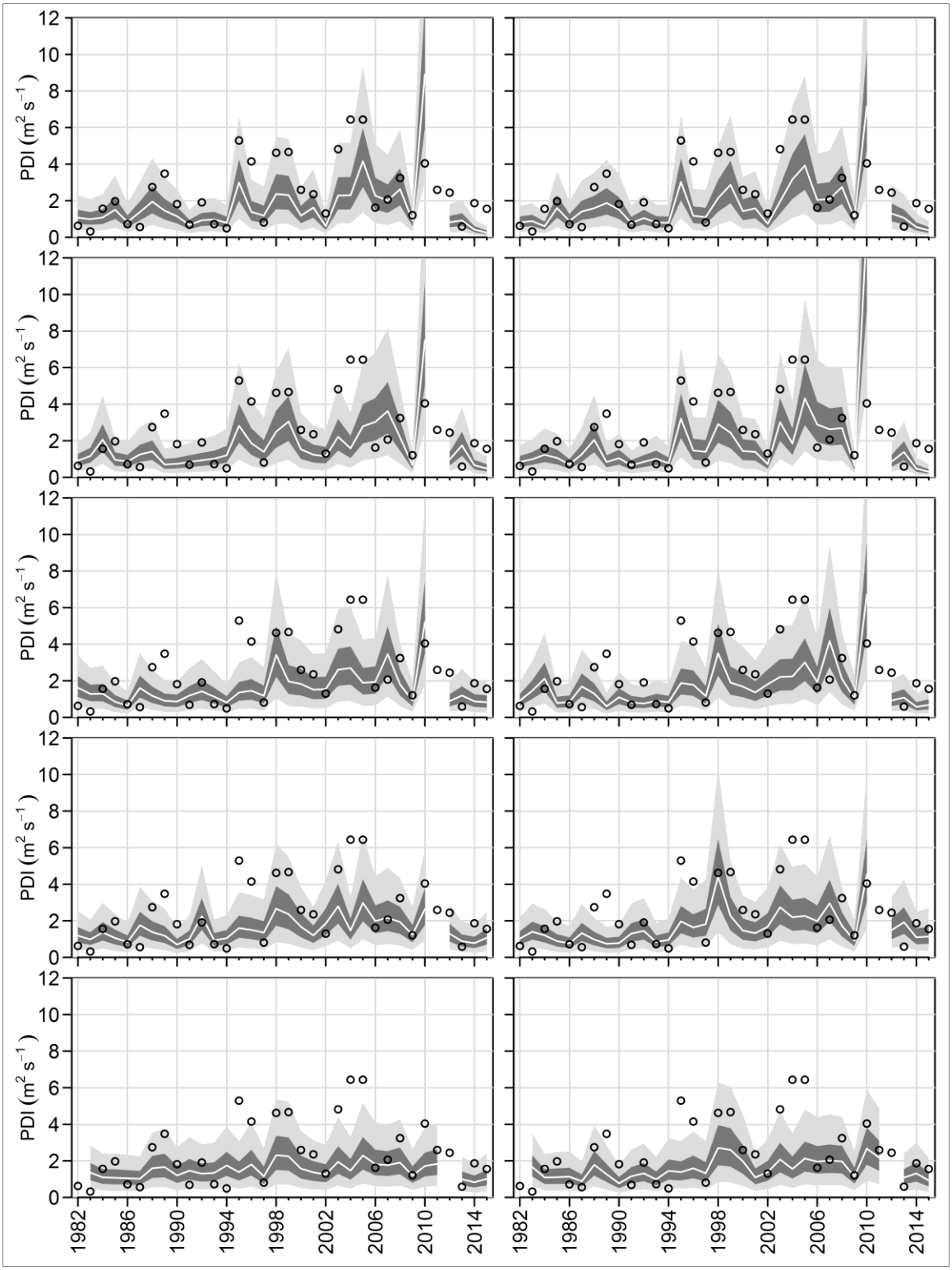

Figure A-35: Same as Figure A-31, but using CMC2.

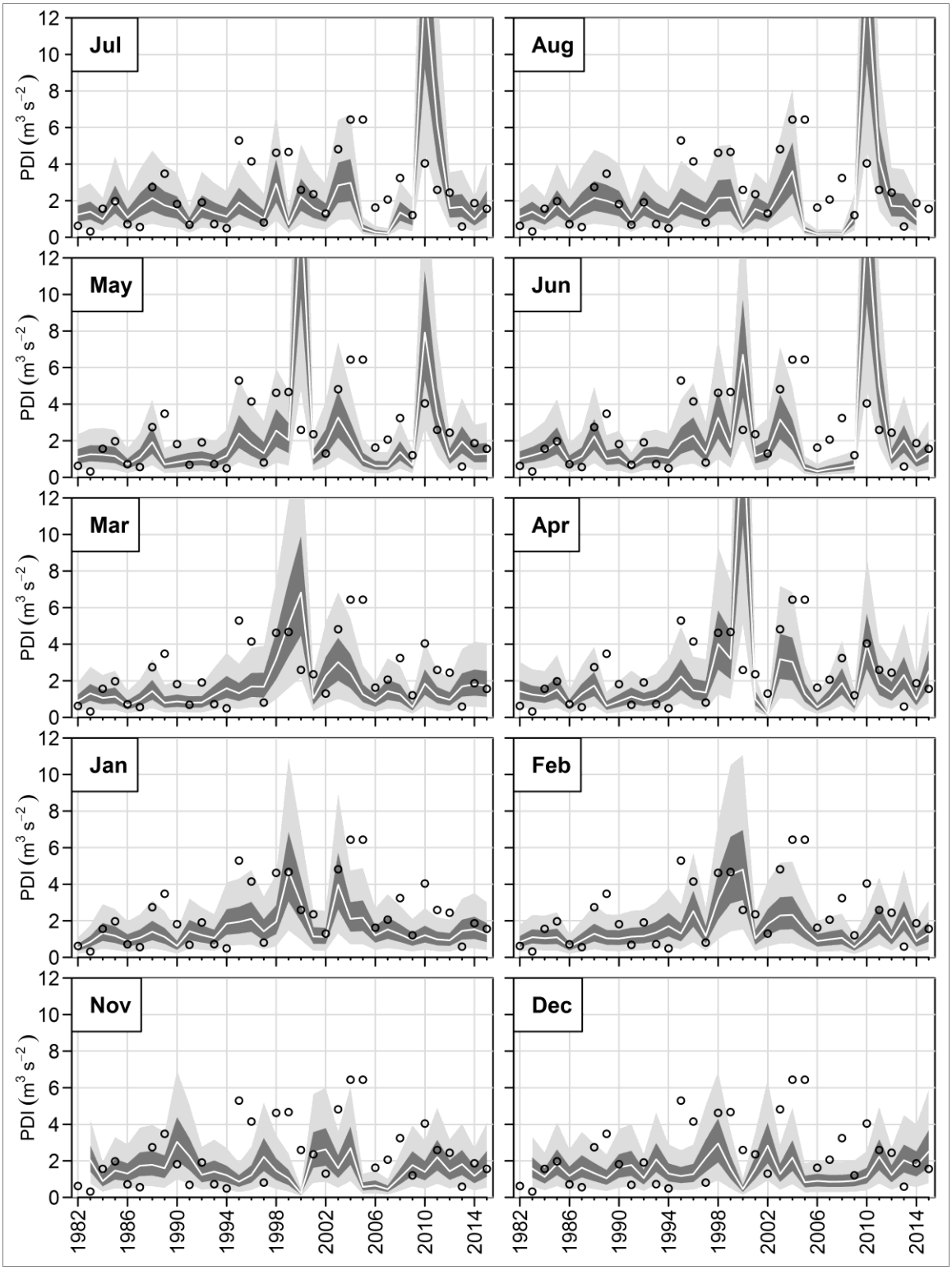

Figure A-36: Same as Figure A-31, but using COLA.

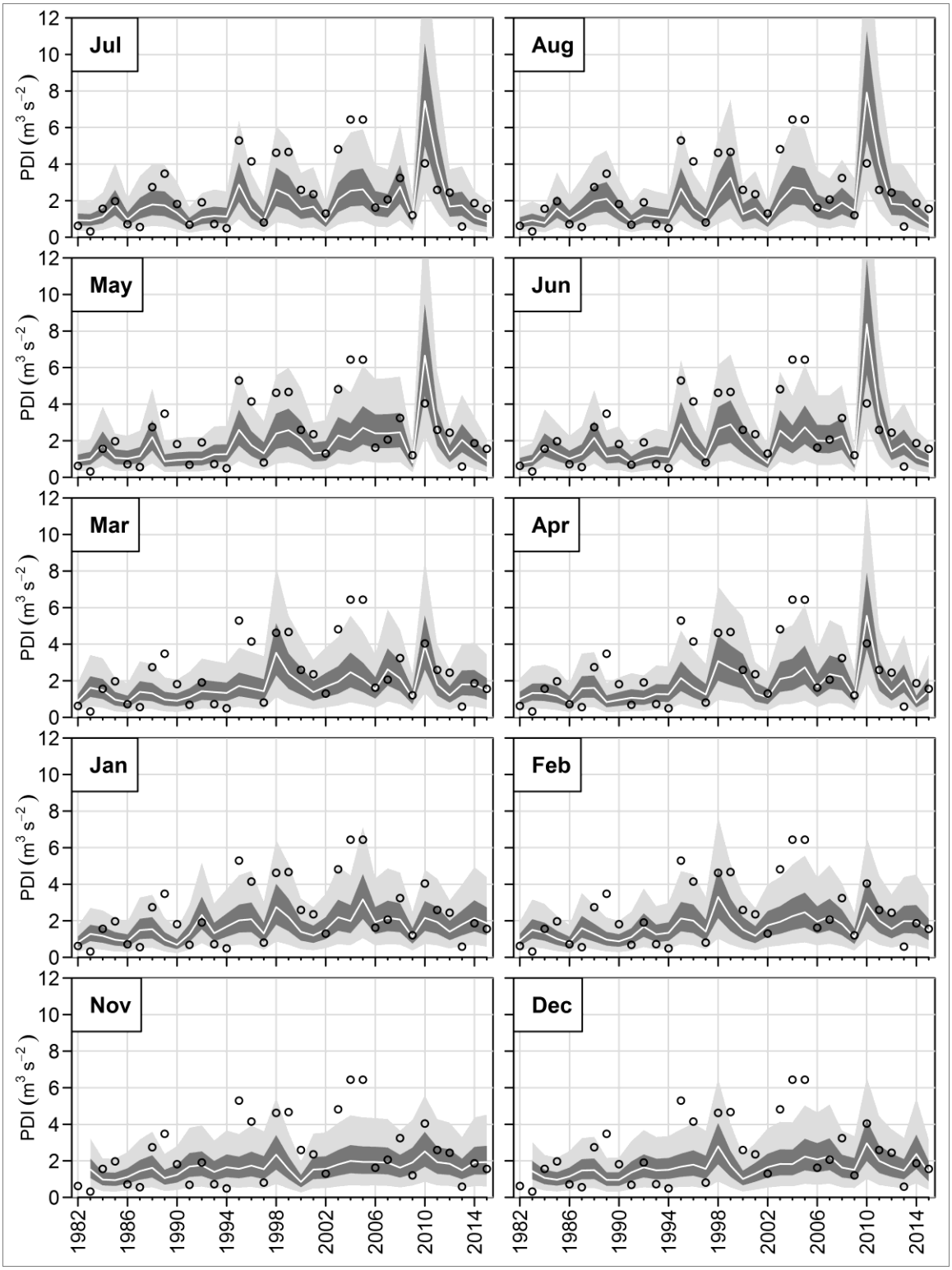

Figure A-37: Same as Figure A-31, but using model in which the predictors are equally weighted.

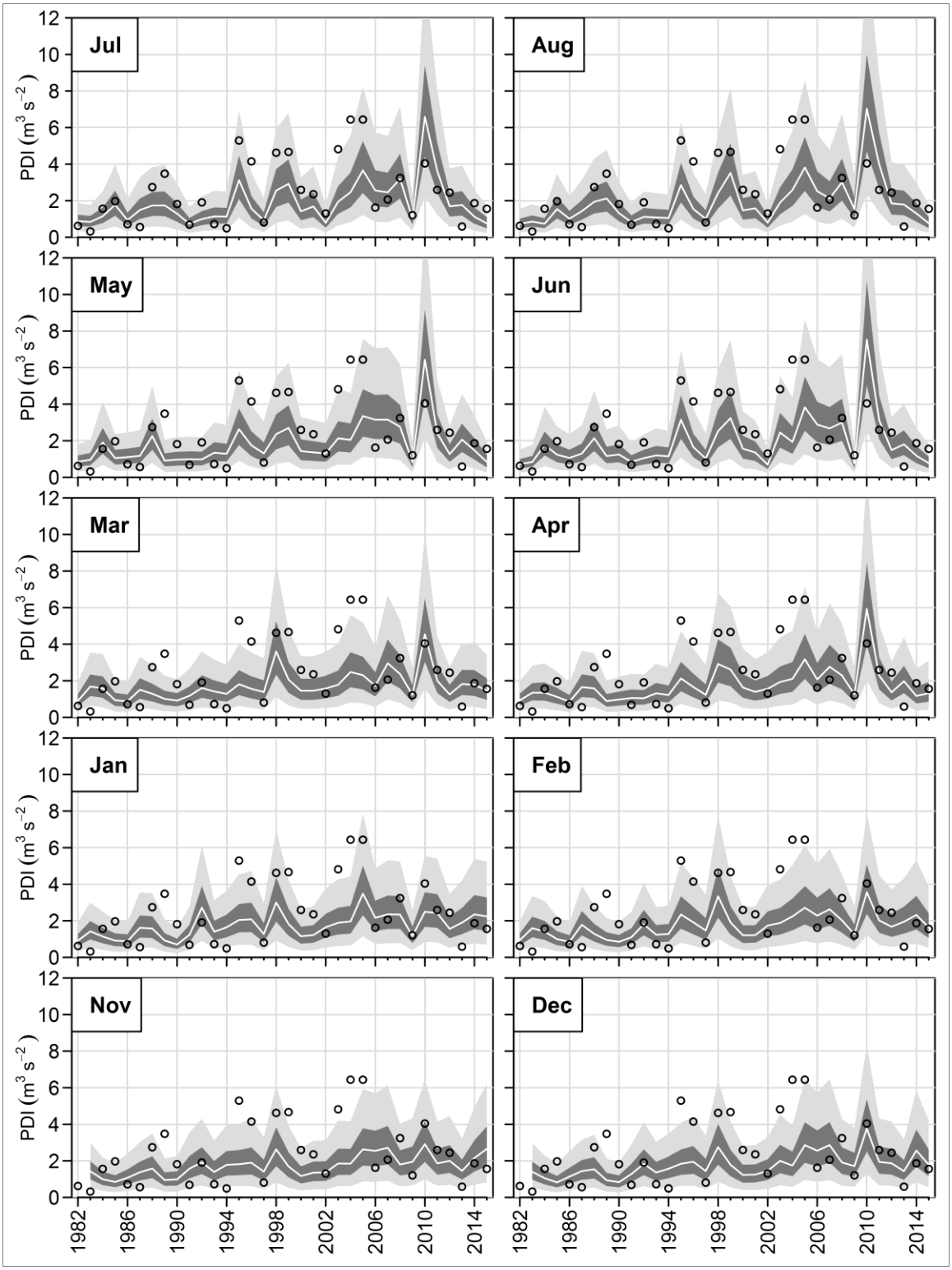

Figure A-38: Same as Figure A-31, but using model in which the predictors are equally weighted without COLA.

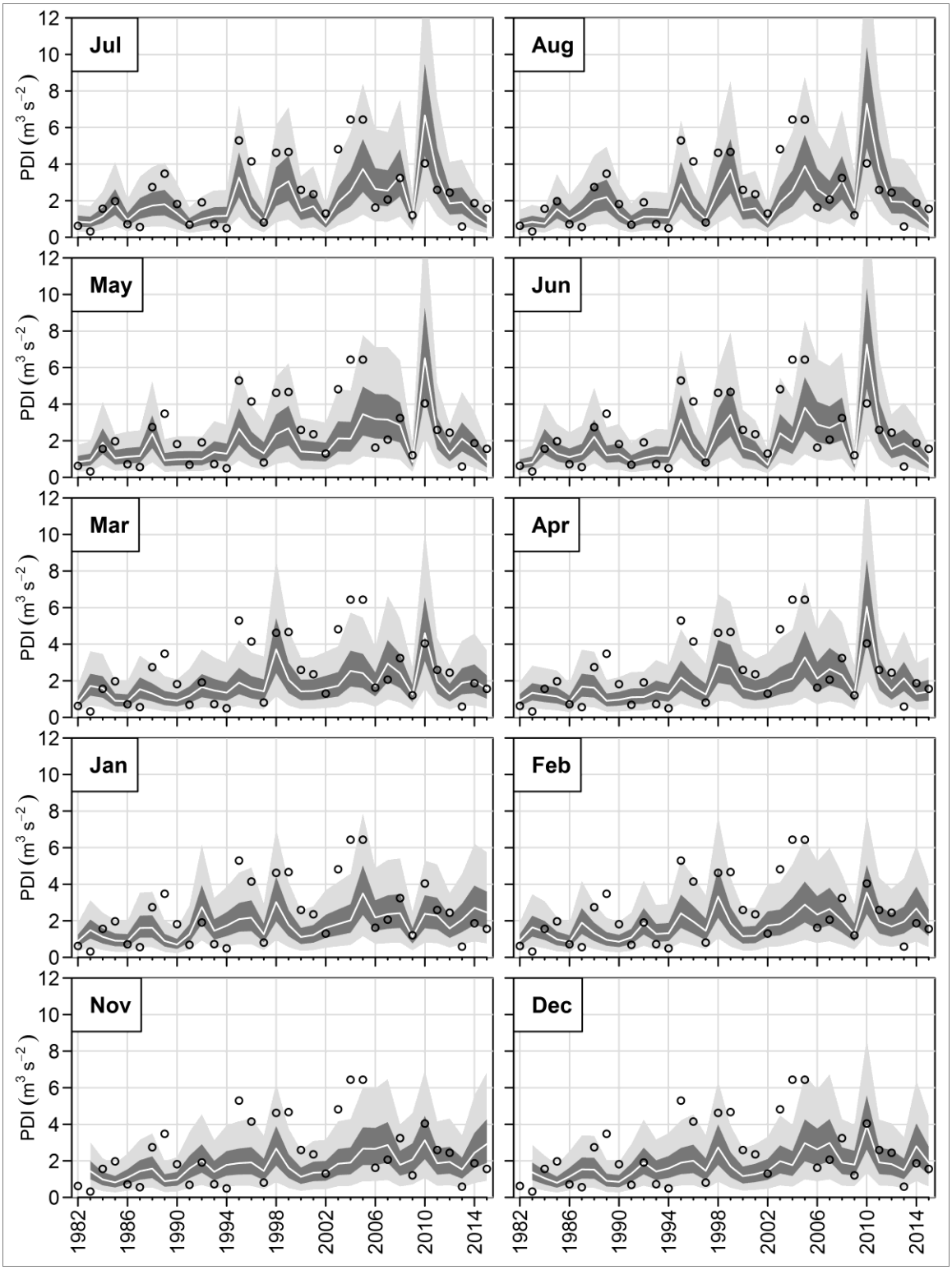

Figure A-39: Same as Figure A-31, but using model in which the predictors are weighted without COLA.

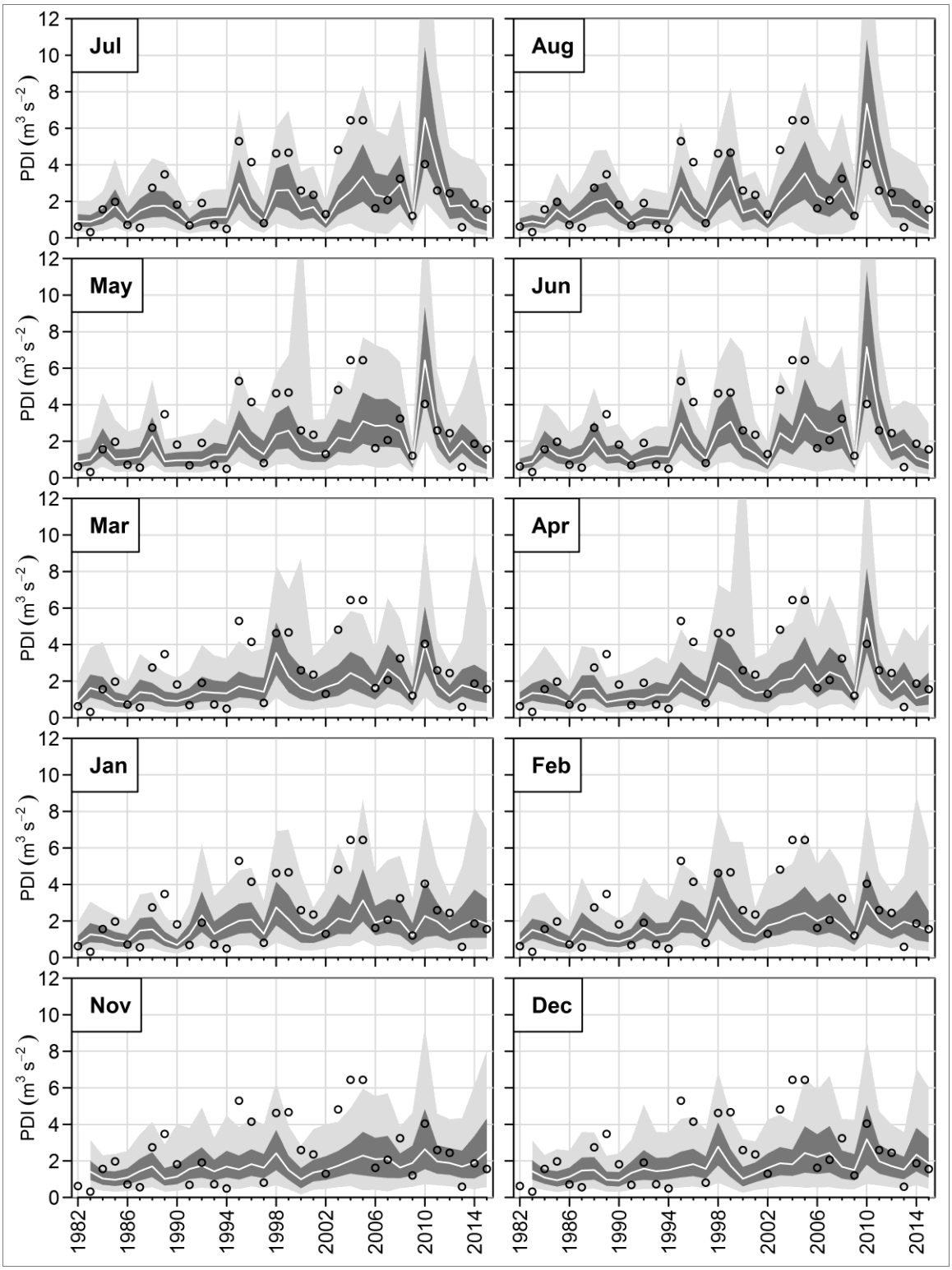

Figure A-40: Same as Figure A-31, but using Hierarchical model.

APPENDIX B: SUPPLEMENTAL MATERIALS FOR CHAPTER 3

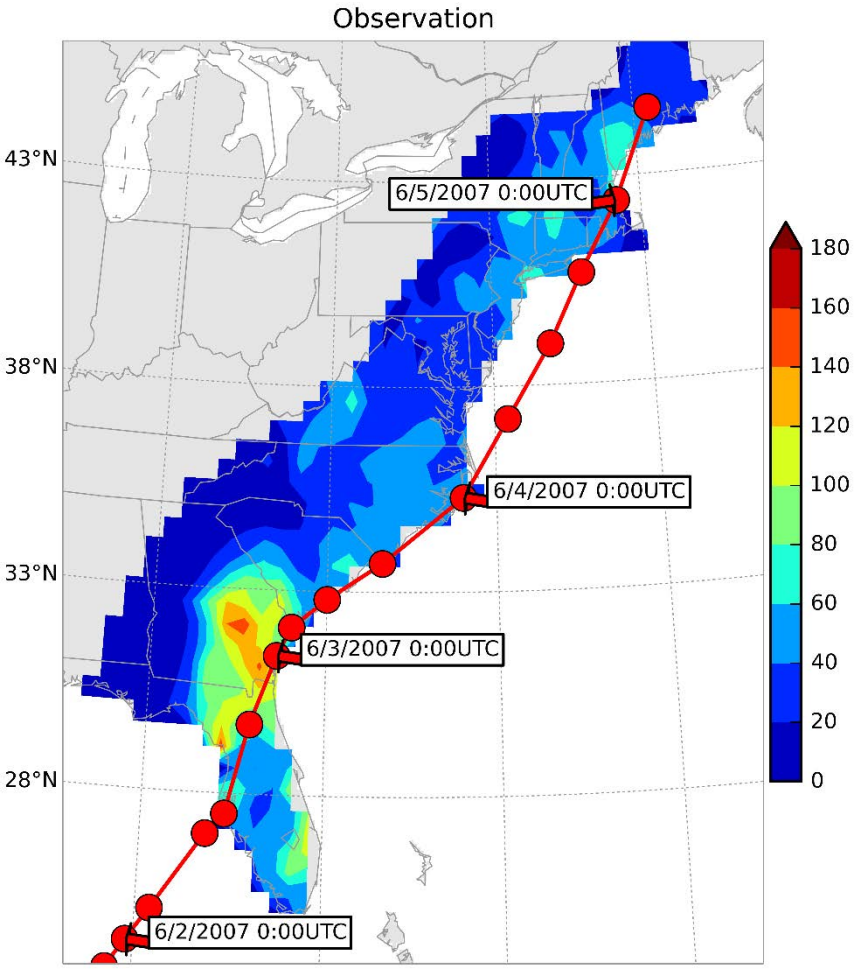

Figure B-1: Rainfall accumulation (in mm) for Tropical Storm Barry (June 2-6 2007) based on the CPC dataset. The red line with dots represents the storm track.

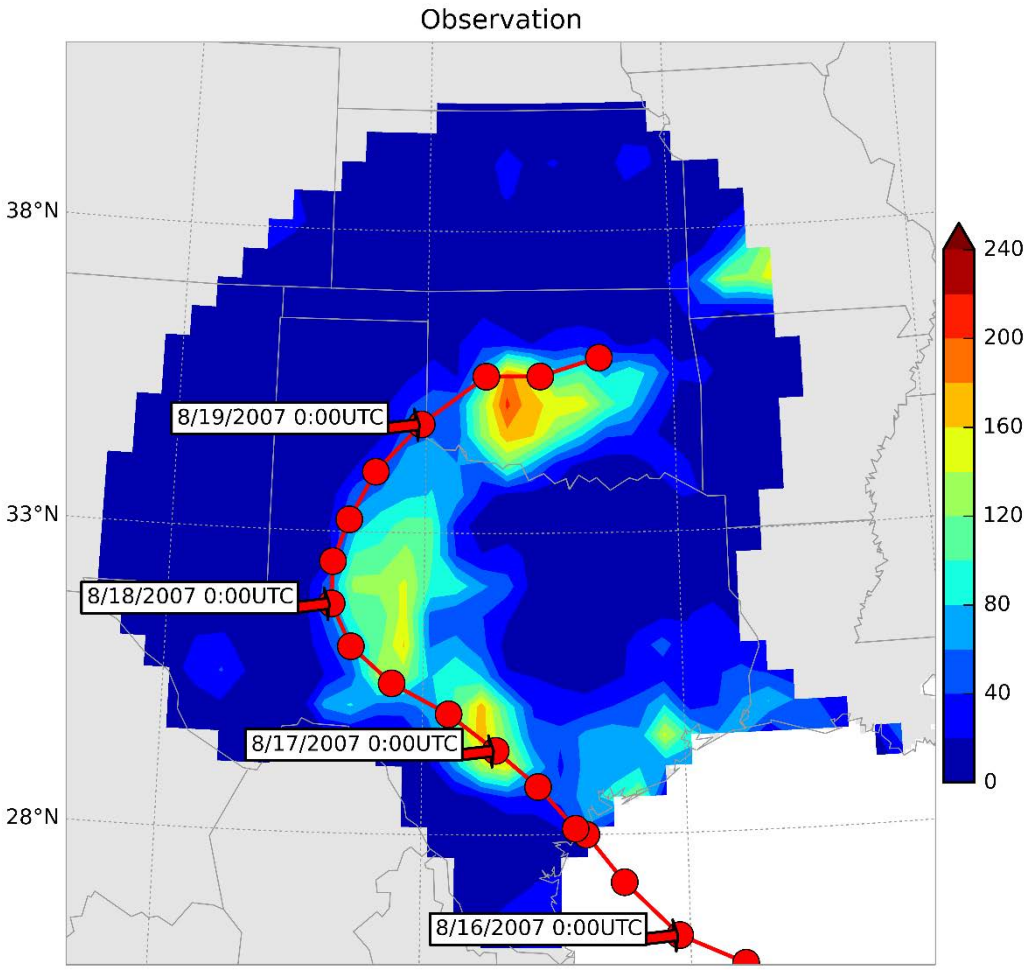

Figure B-2: Rainfall accumulation (in mm) for Tropical Storm Erin (August 16-21 2007) based on the CPC dataset. The red line with dots represents the storm track.

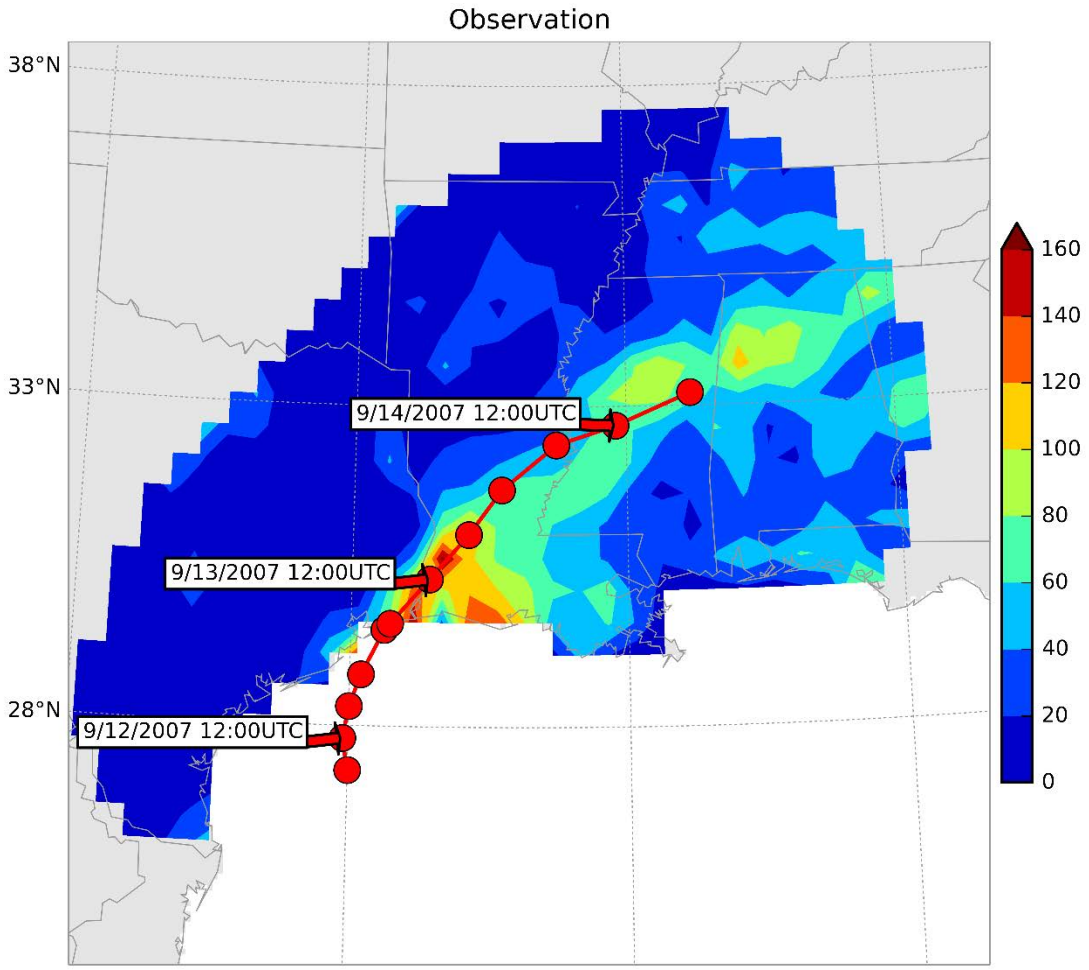

Figure B-3: Rainfall accumulation (in mm) for Hurricane Humberto (September 11-15 2007) based on the CPC dataset. The red line with dots represents the storm track.

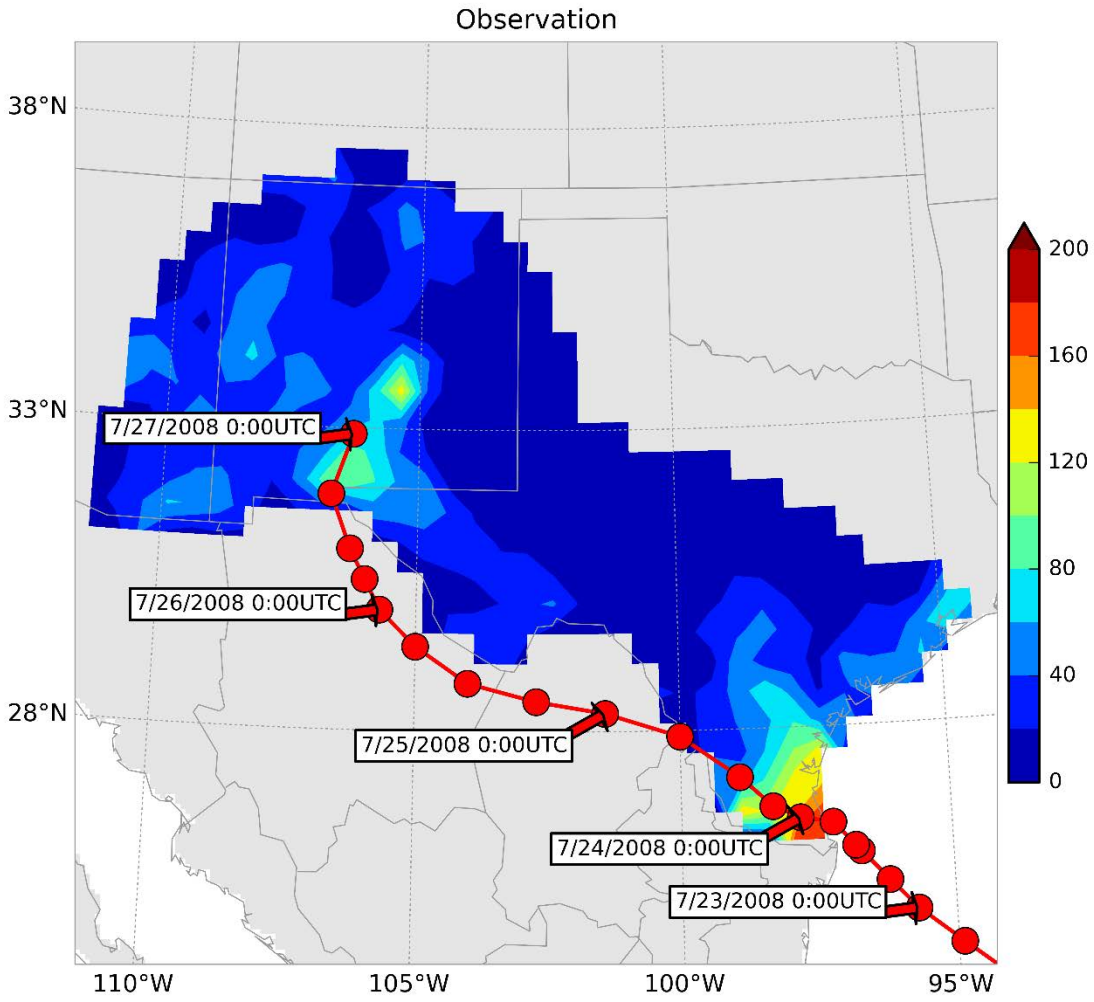

Figure B-4: Rainfall accumulation (in mm) for Hurricane Dolly (July 23-28 2008) based on the CPC dataset. The red line with dots represents the storm track.

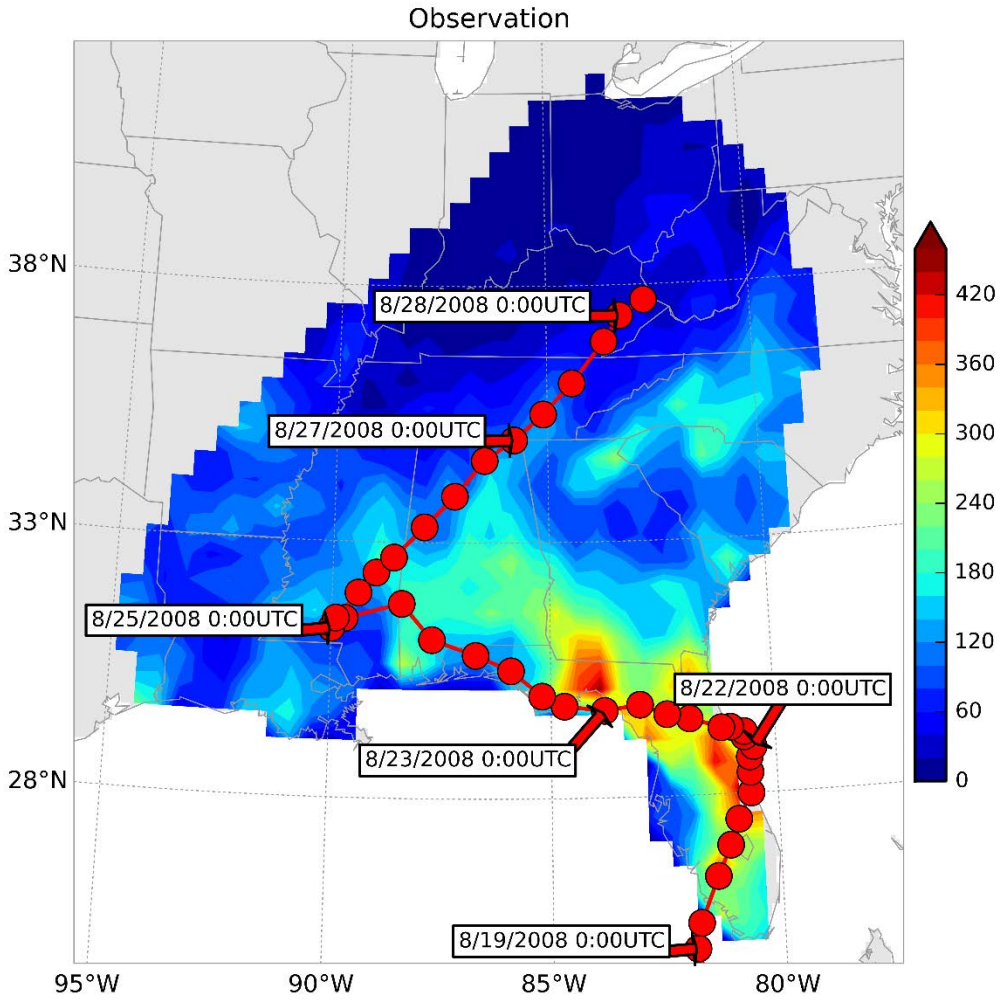

Figure B-5: Rainfall accumulation (in mm) for Tropical Storm Fay (August 19-28 2008) based on the CPC dataset. The red line with dots represents the storm track.

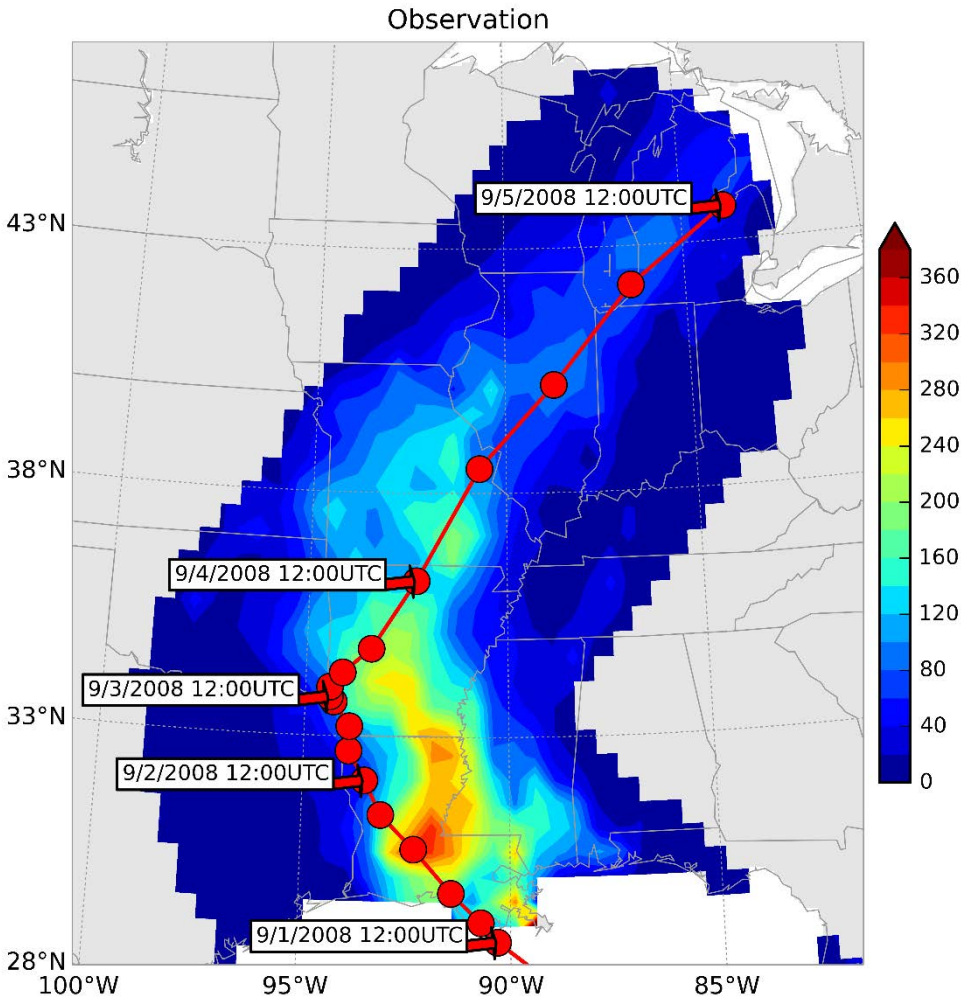

Figure B-6: Rainfall accumulation (in mm) for Hurricane Gustav (August 31- September 5 2008) based on the CPC dataset. The red line with dots represents the storm track.

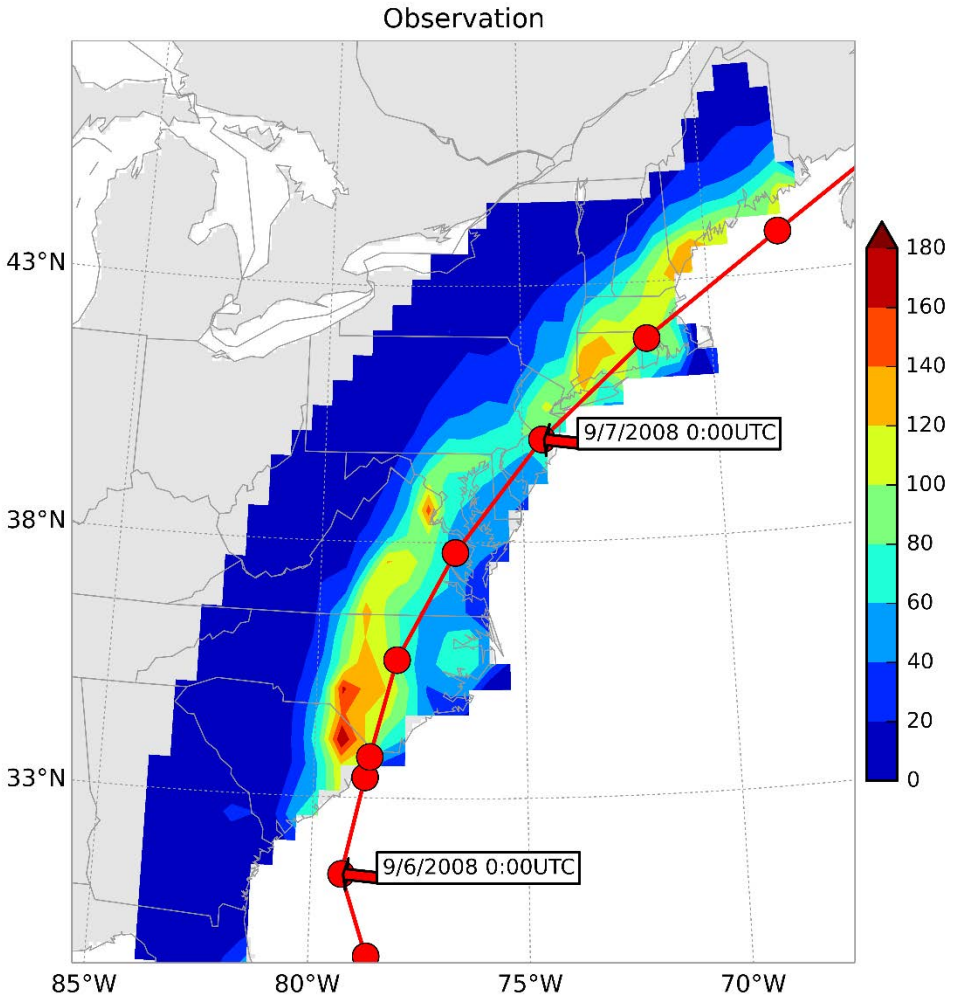

Figure B-7: Rainfall accumulation (in mm) for Tropical Storm Hanna (September 5-8 2008) based on the CPC dataset. The red line with dots represents the storm track.

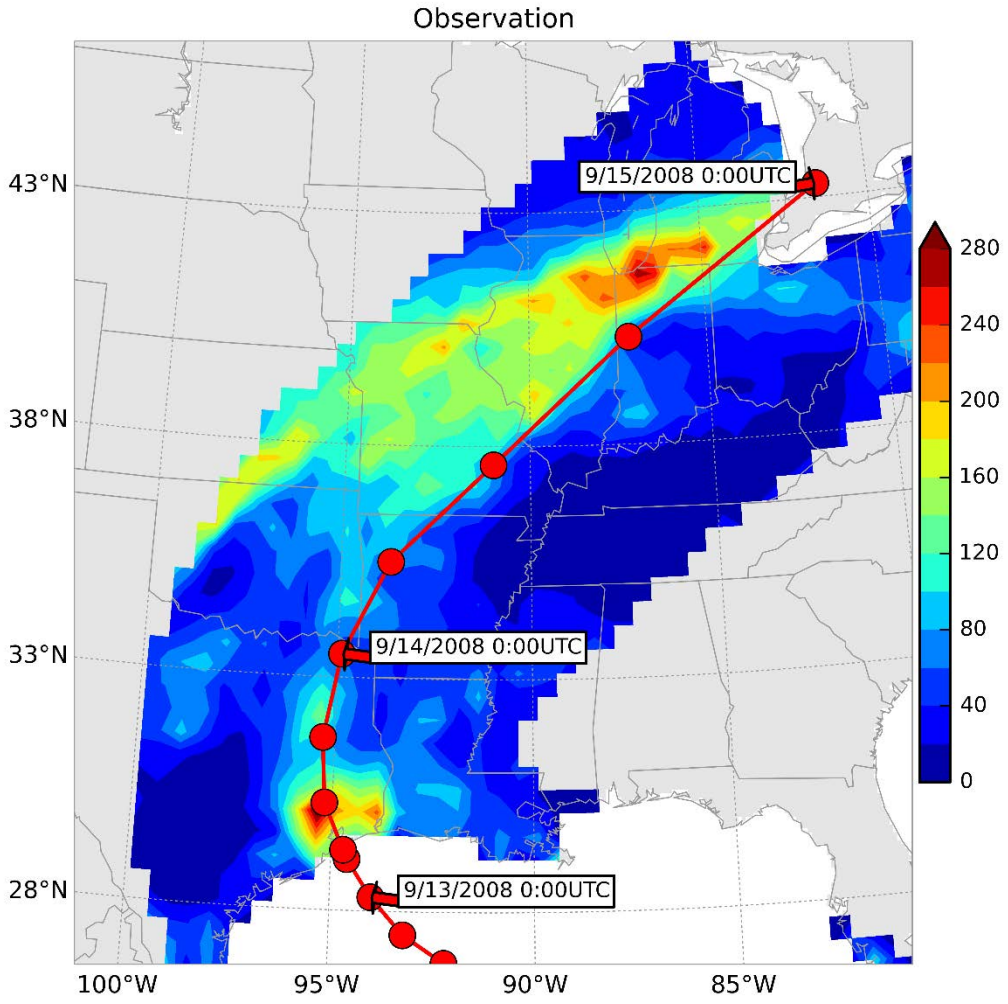

Figure B-8: Rainfall accumulation (in mm) for Hurricane Ike (September 12-15 2008) based on the CPC dataset. The red line with dots represents the storm track.

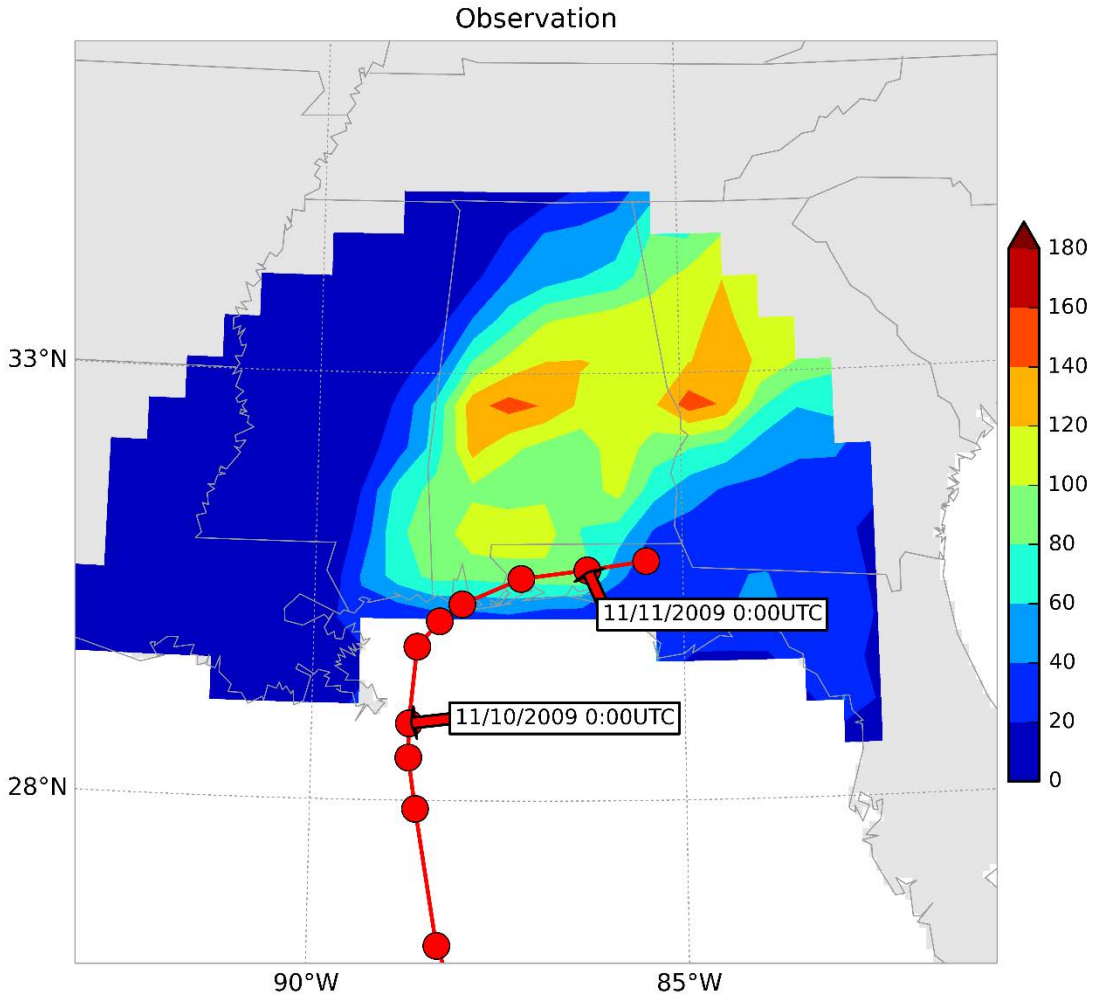

Figure B-9: Rainfall accumulation (in mm) for Hurricane Ida (November 10-12 2009) based on the CPC dataset. The red line with dots represents the storm track.

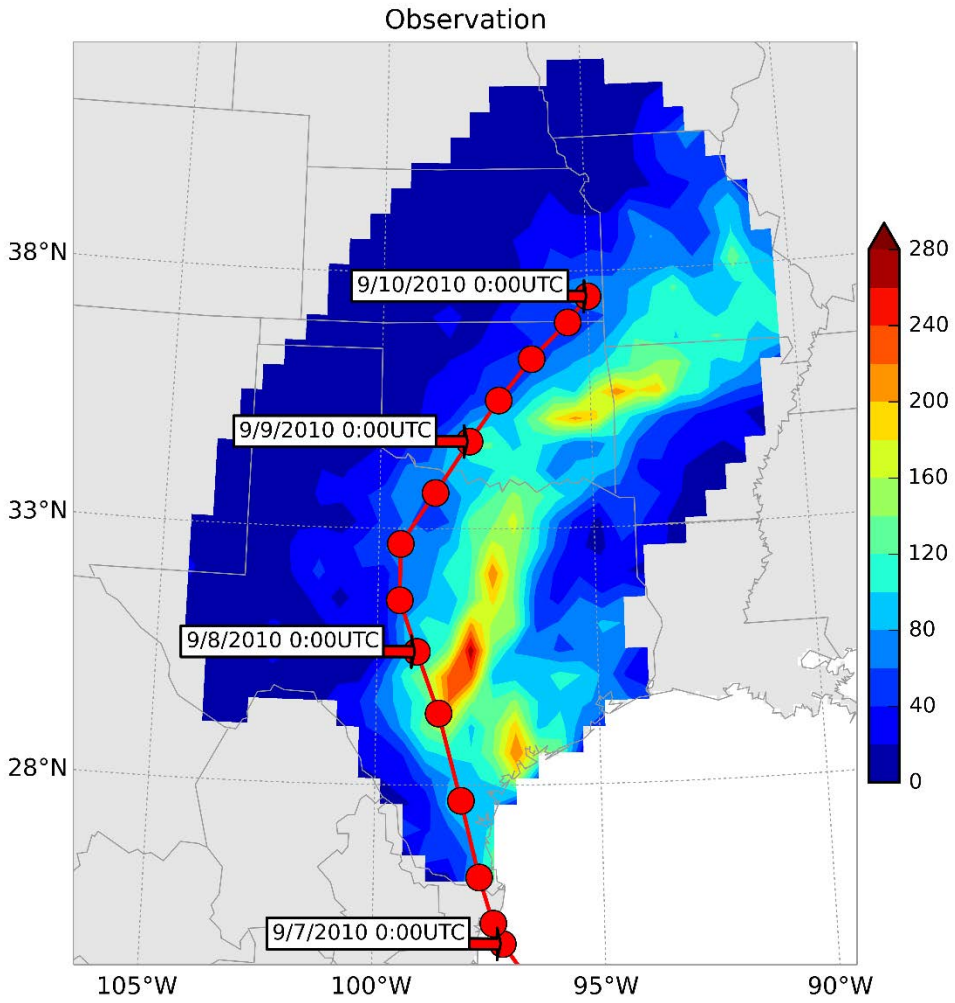

Figure B-10: Rainfall accumulation (in mm) for Tropical Storm Hermine (September 6-11 2010) based on the CPC dataset. The red line with dots represents the storm track.

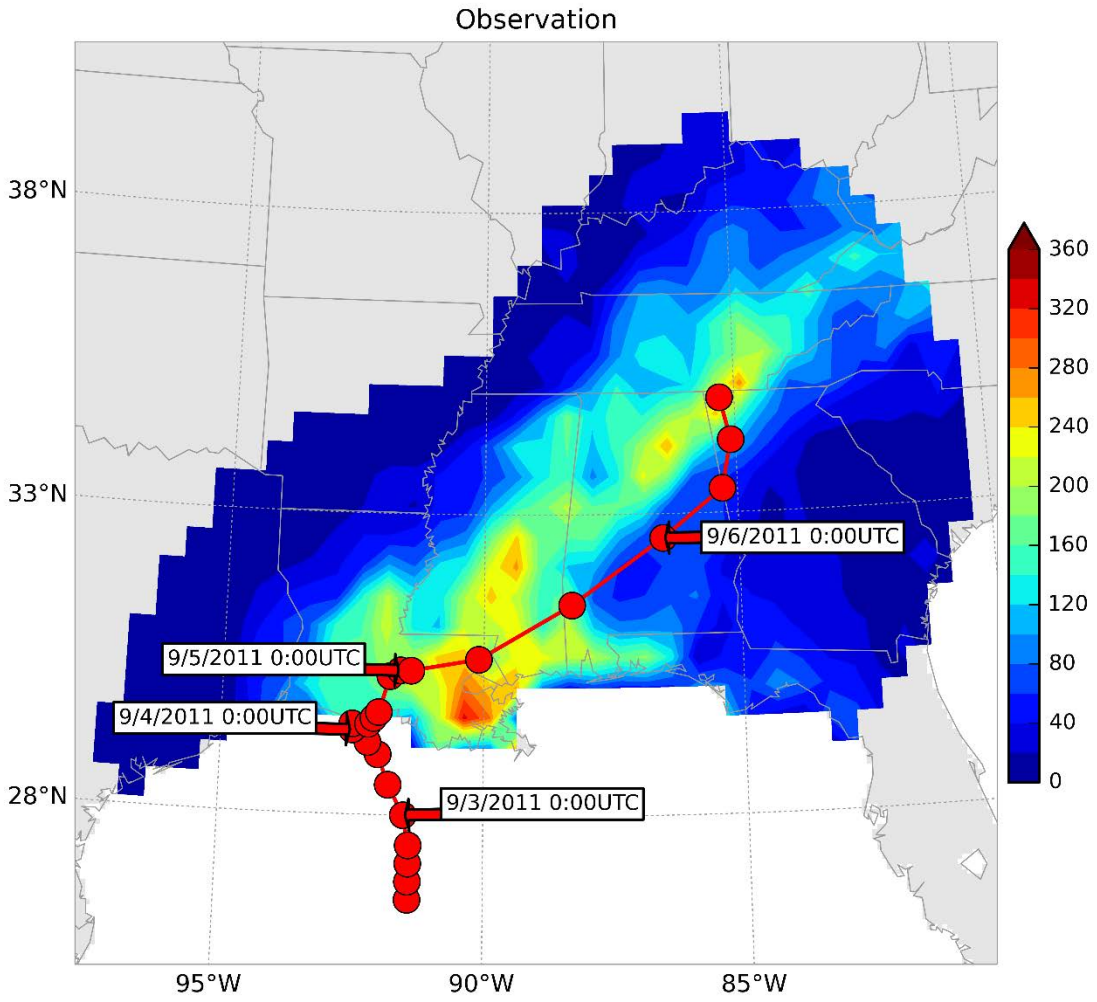

Figure B-11: Rainfall accumulation (in mm) for Tropical Storm Lee (September 3-8 2011) based on the CPC dataset. The red line with dots represents the storm track.

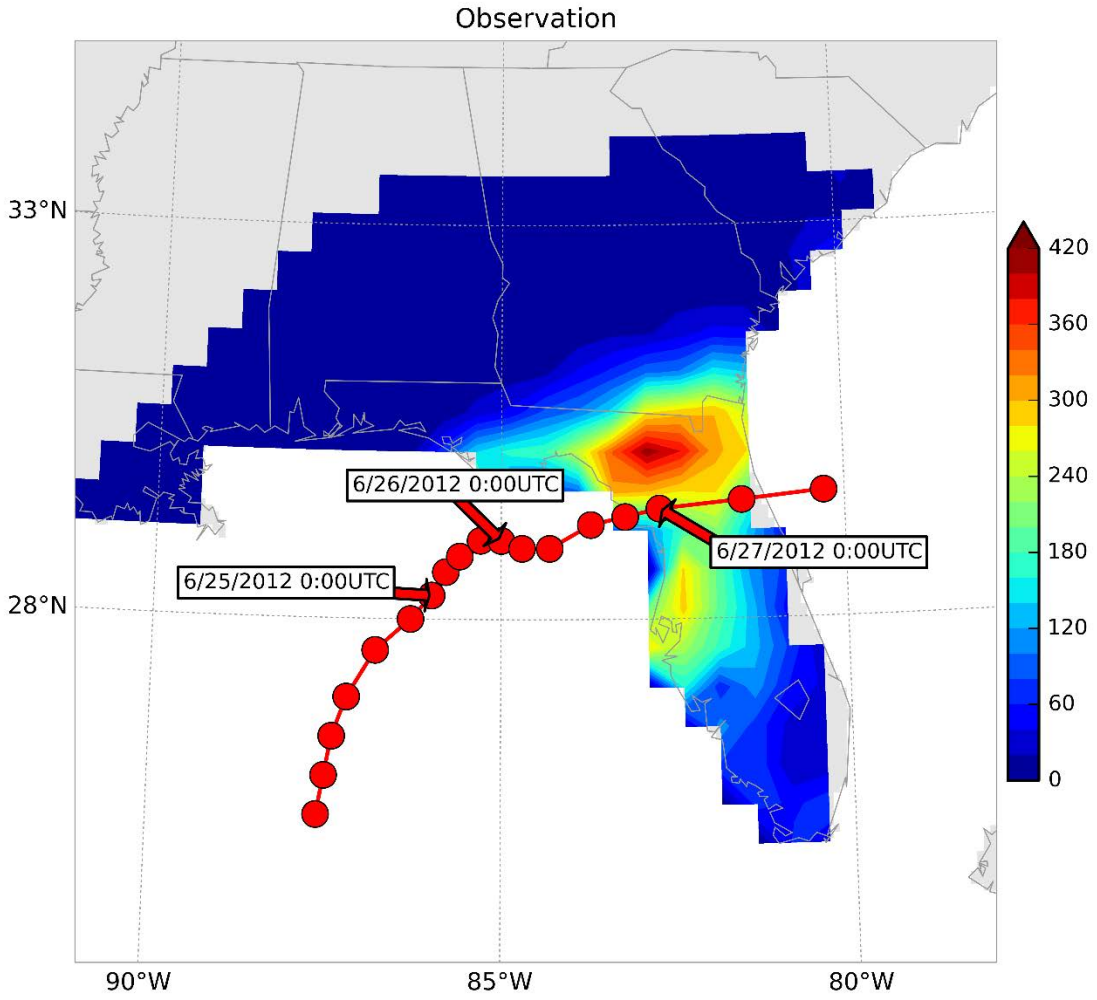

Figure B-12: Rainfall accumulation (in mm) for Tropical Storm Debby (June 25-28 2012) based on the CPC dataset. The red line with dots represents the storm track.

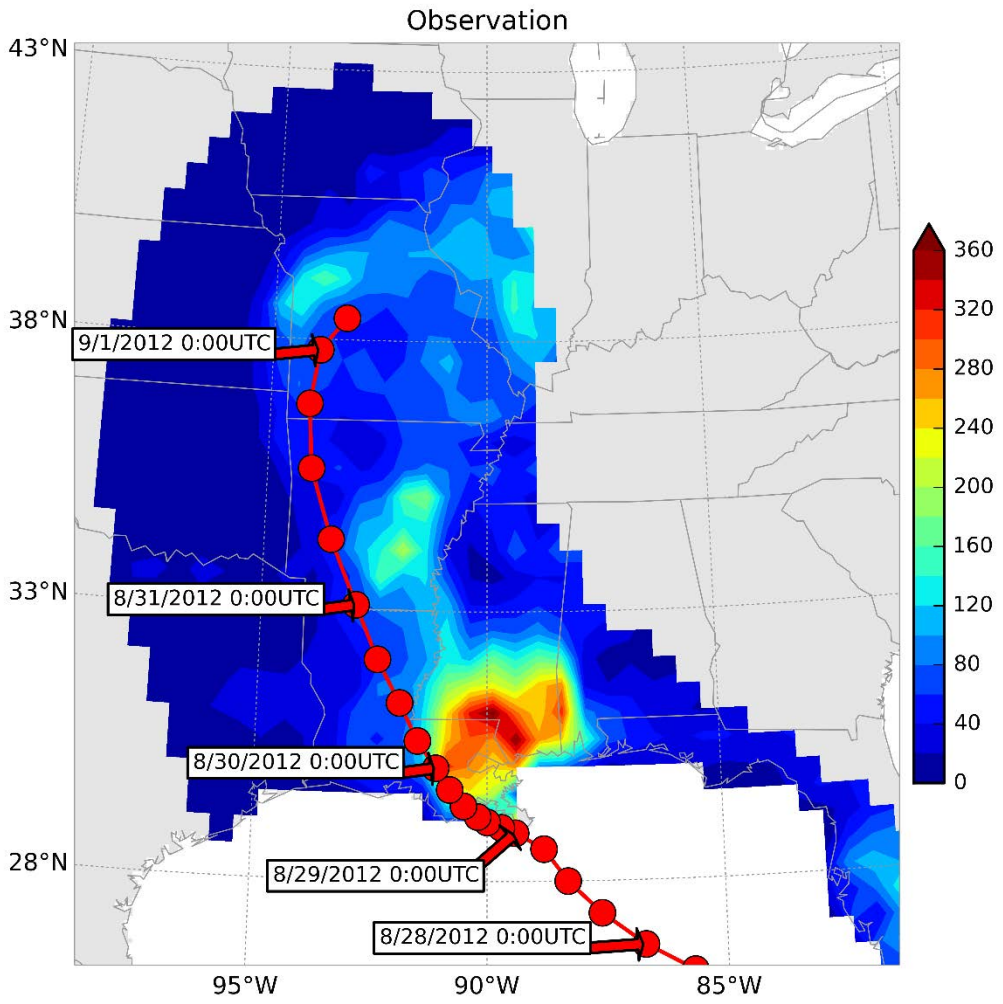

Figure B-13: Rainfall accumulation (in mm) for Hurricane Isaac (August 28- September 2 2012) based on the CPC dataset. The red line with dots represents the storm track.

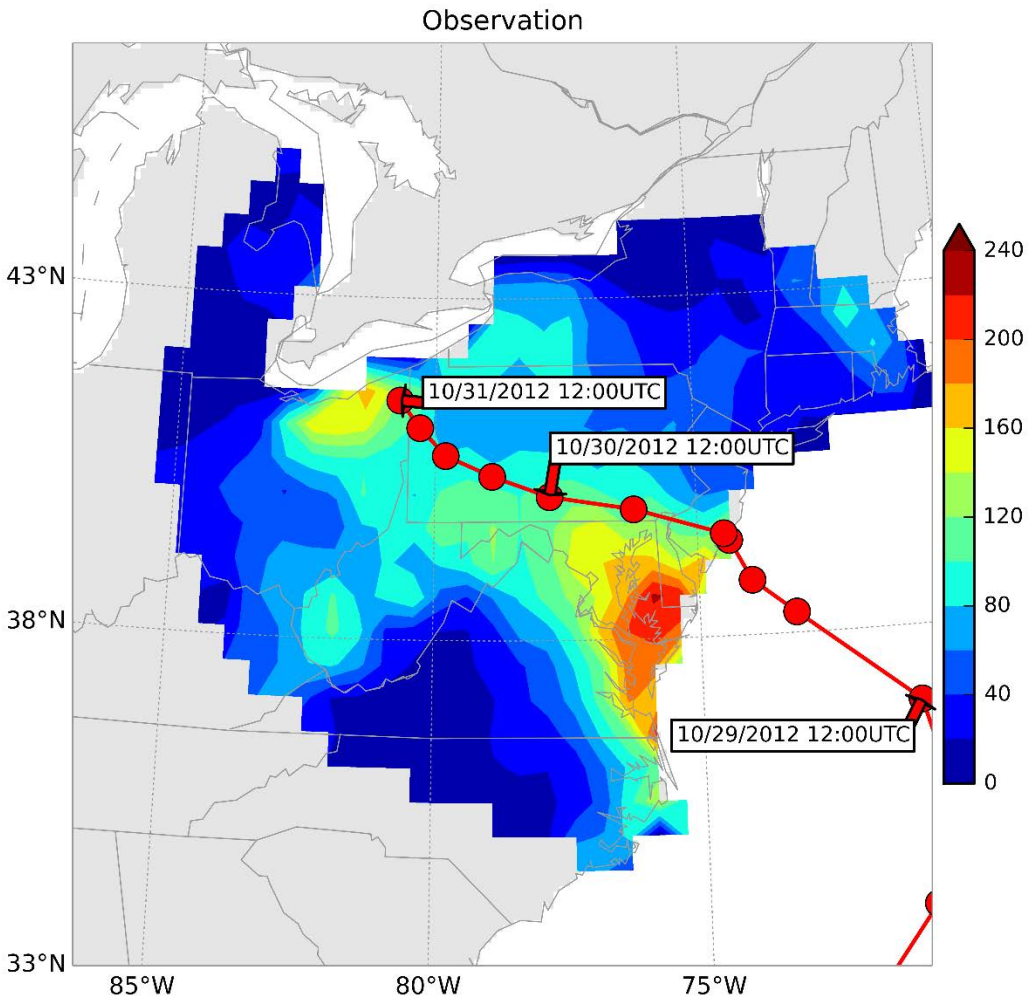

Figure B-14: Rainfall accumulation (in mm) for Hurricane Sandy (October 28-31 2012) based on the CPC dataset. The red line with dots represents the storm track.

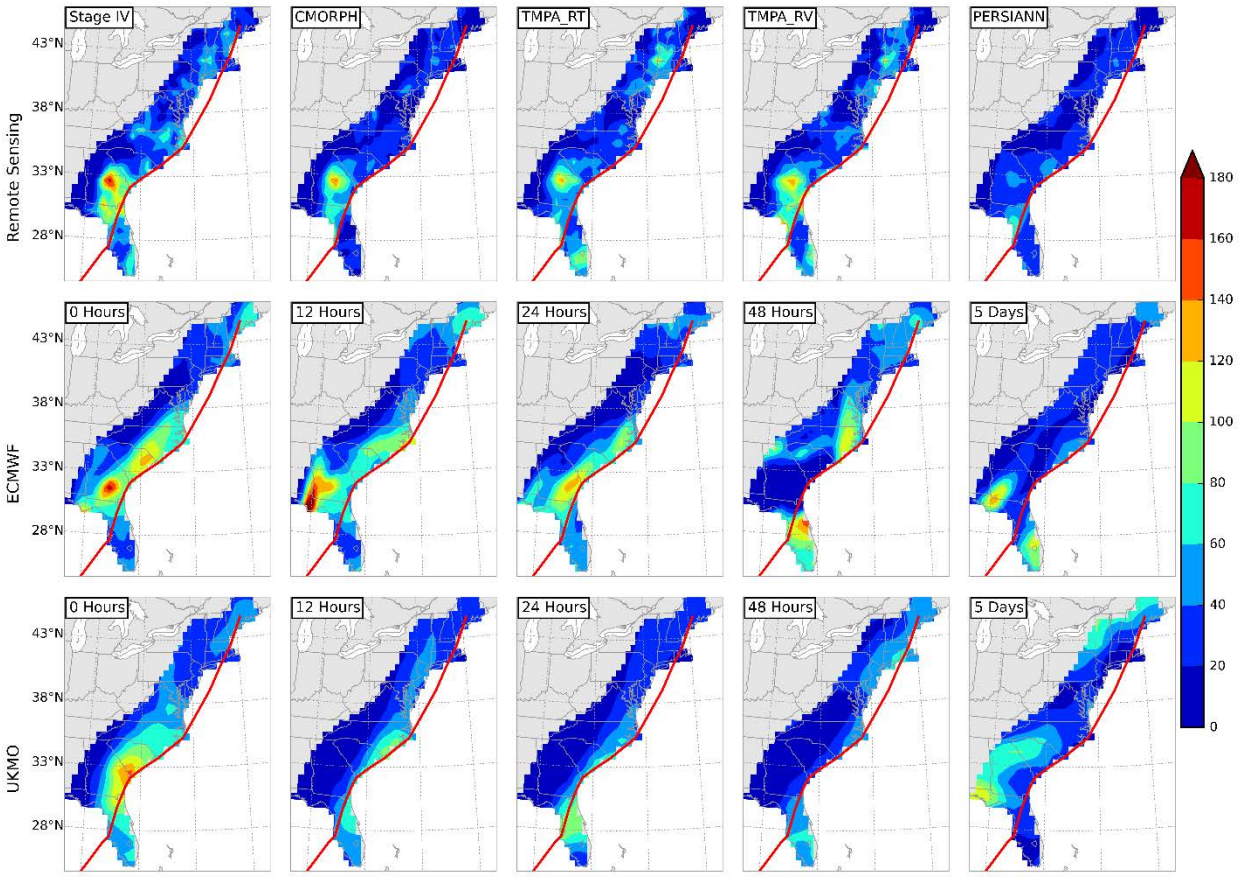

Figure B-15: Rainfall accumulation (in mm) (in mm) for Tropical Storm Barry (June 2-6 2007) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

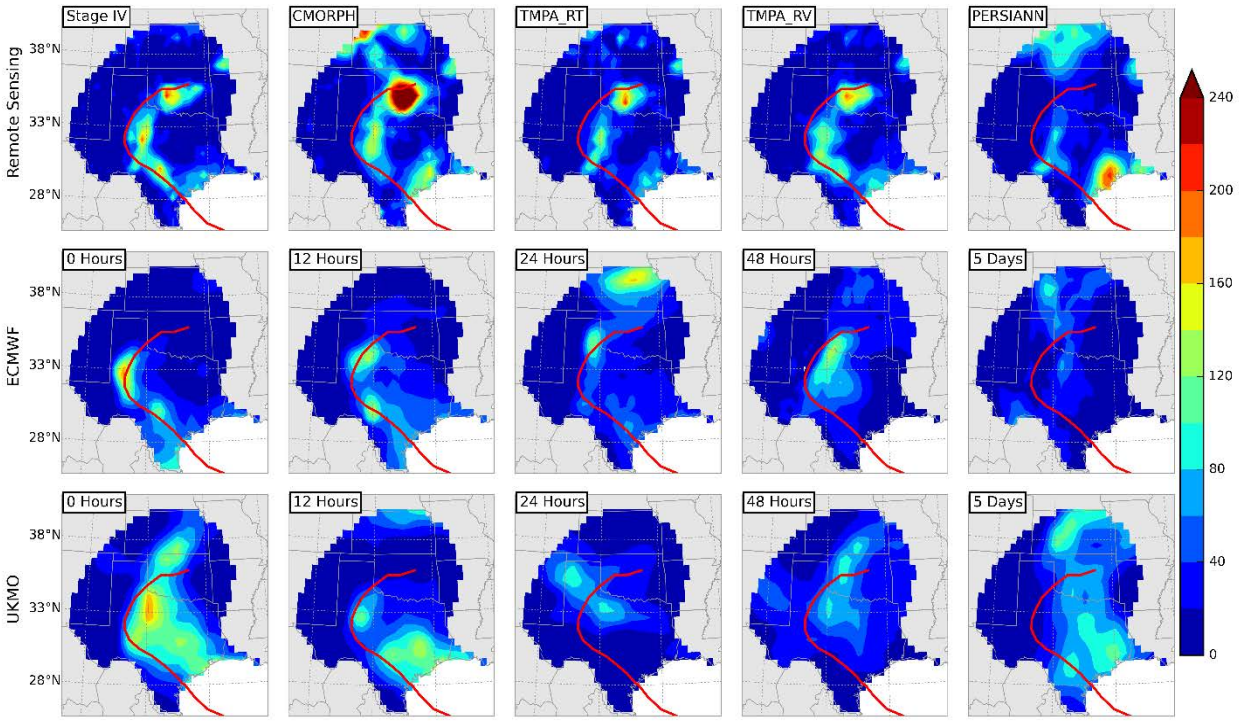

Figure B-16: Rainfall accumulation (in mm) (in mm) for Tropical Storm Erin (August 16-21 2007) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

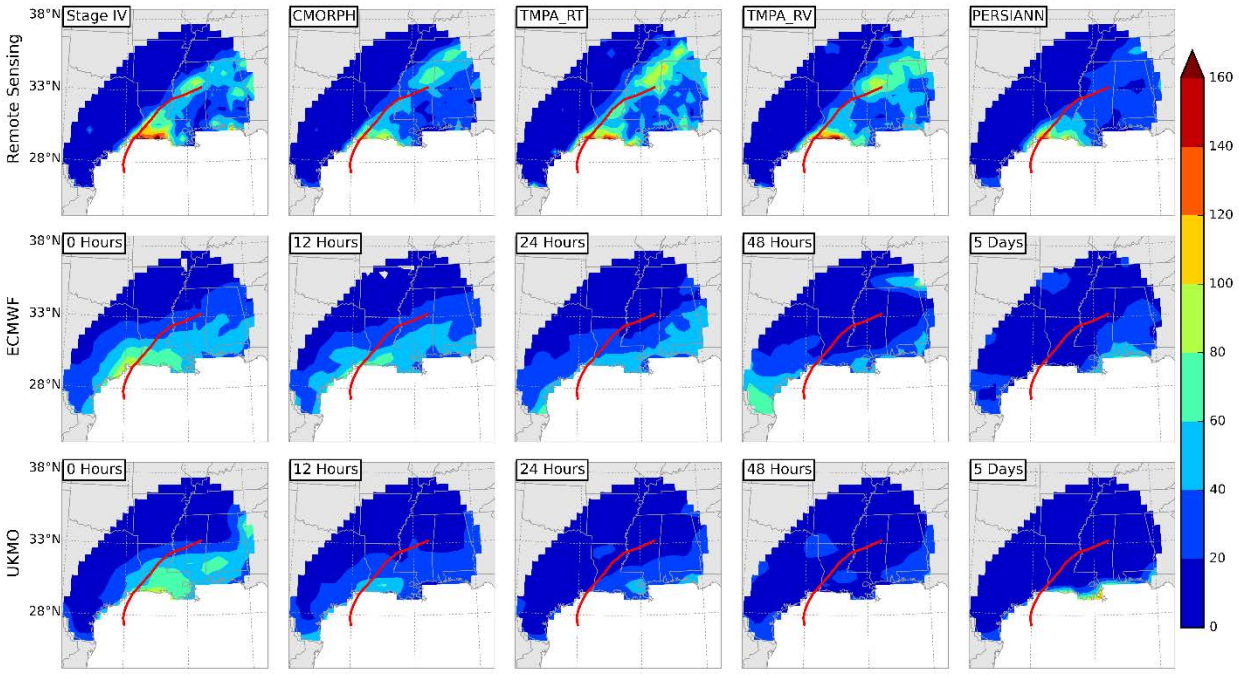

Figure B-17: Rainfall accumulation (in mm) (in mm) for Hurricane Humberto (September 11-15 2007) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

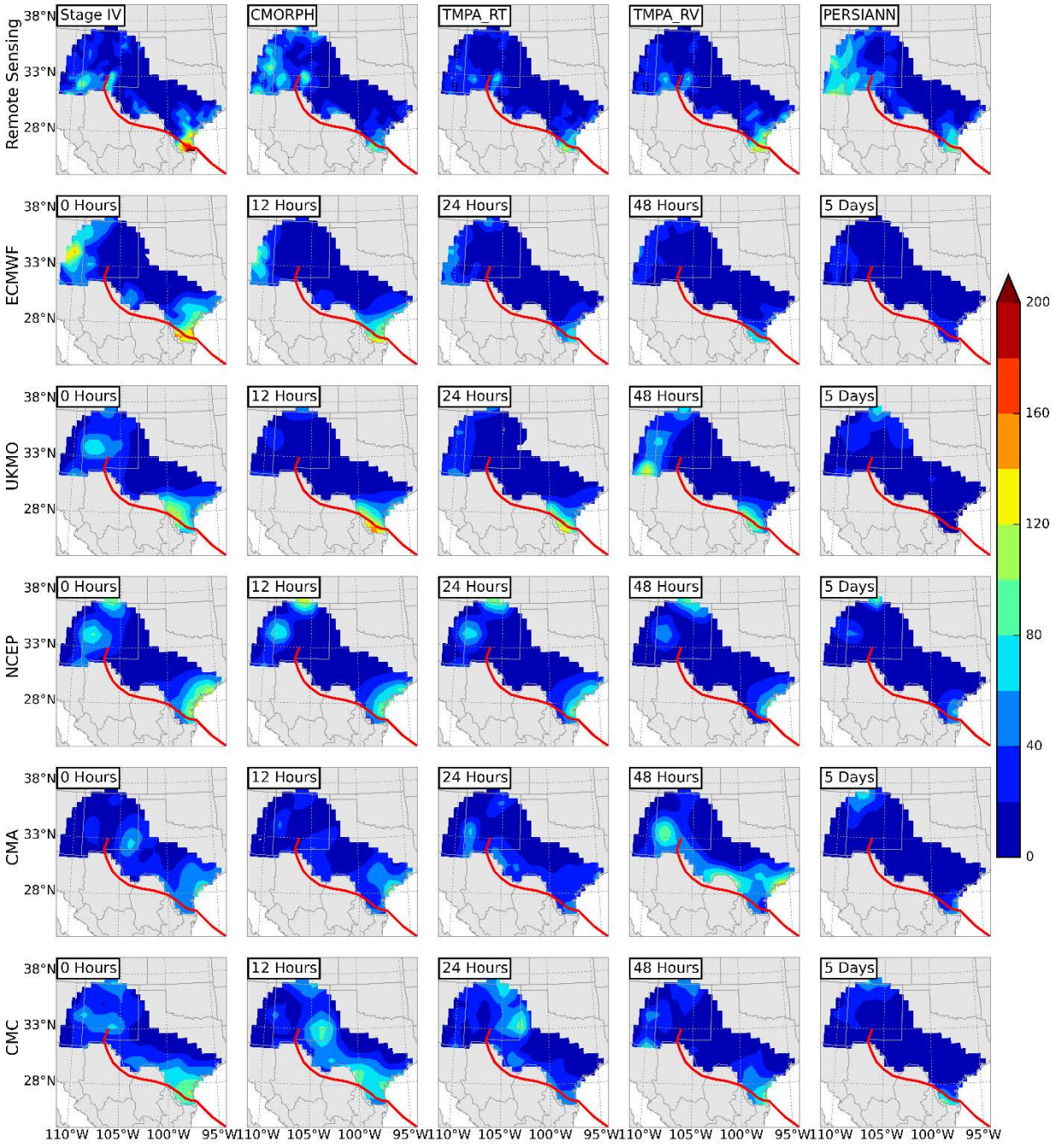

Figure B-18: Rainfall accumulation (in mm) (in mm) for Hurricane Dolly (July 23-28 2008) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

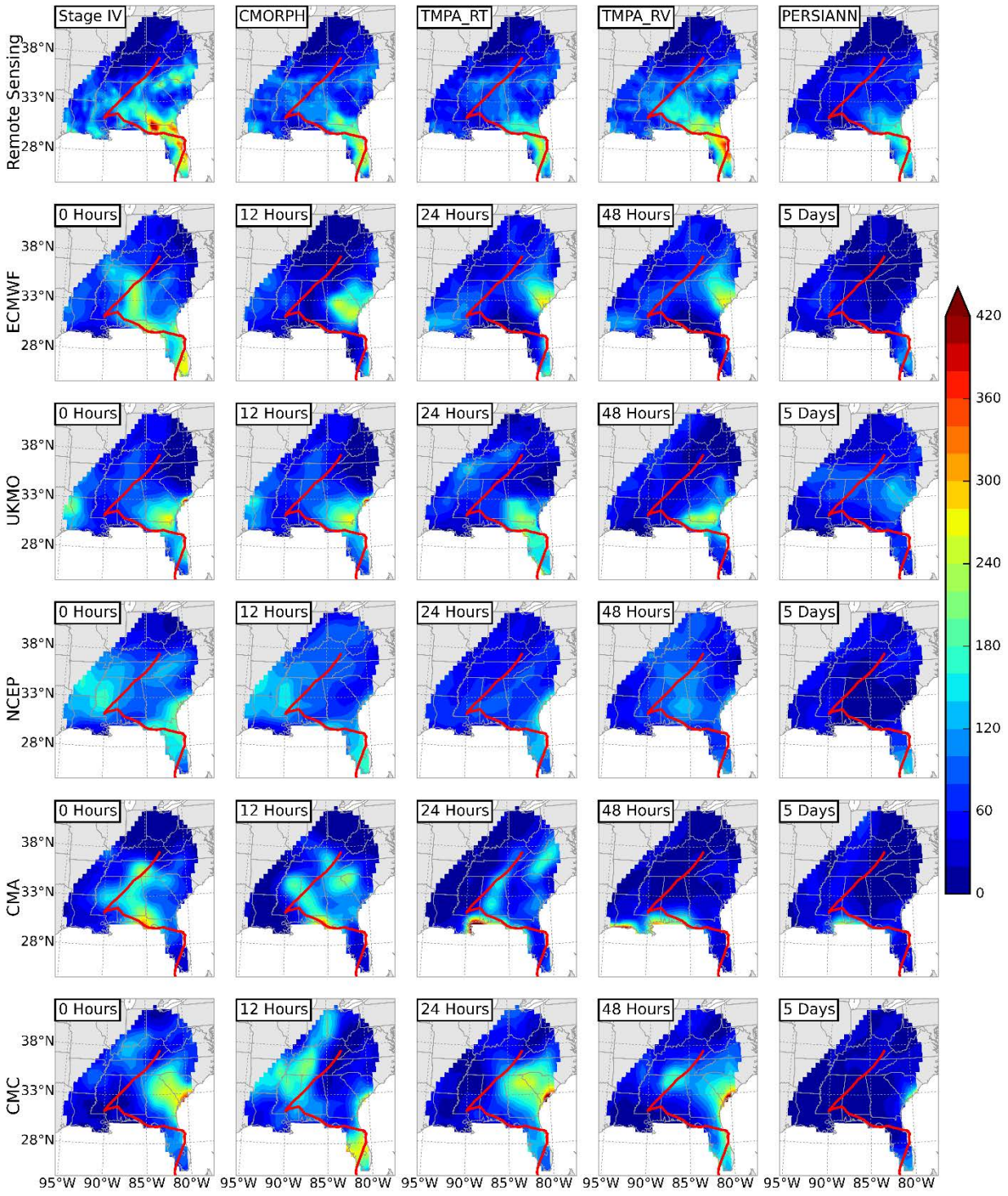

Figure B-19: Rainfall accumulation (in mm) (in mm) for Tropical Storm Fay (August 19-28 2008) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

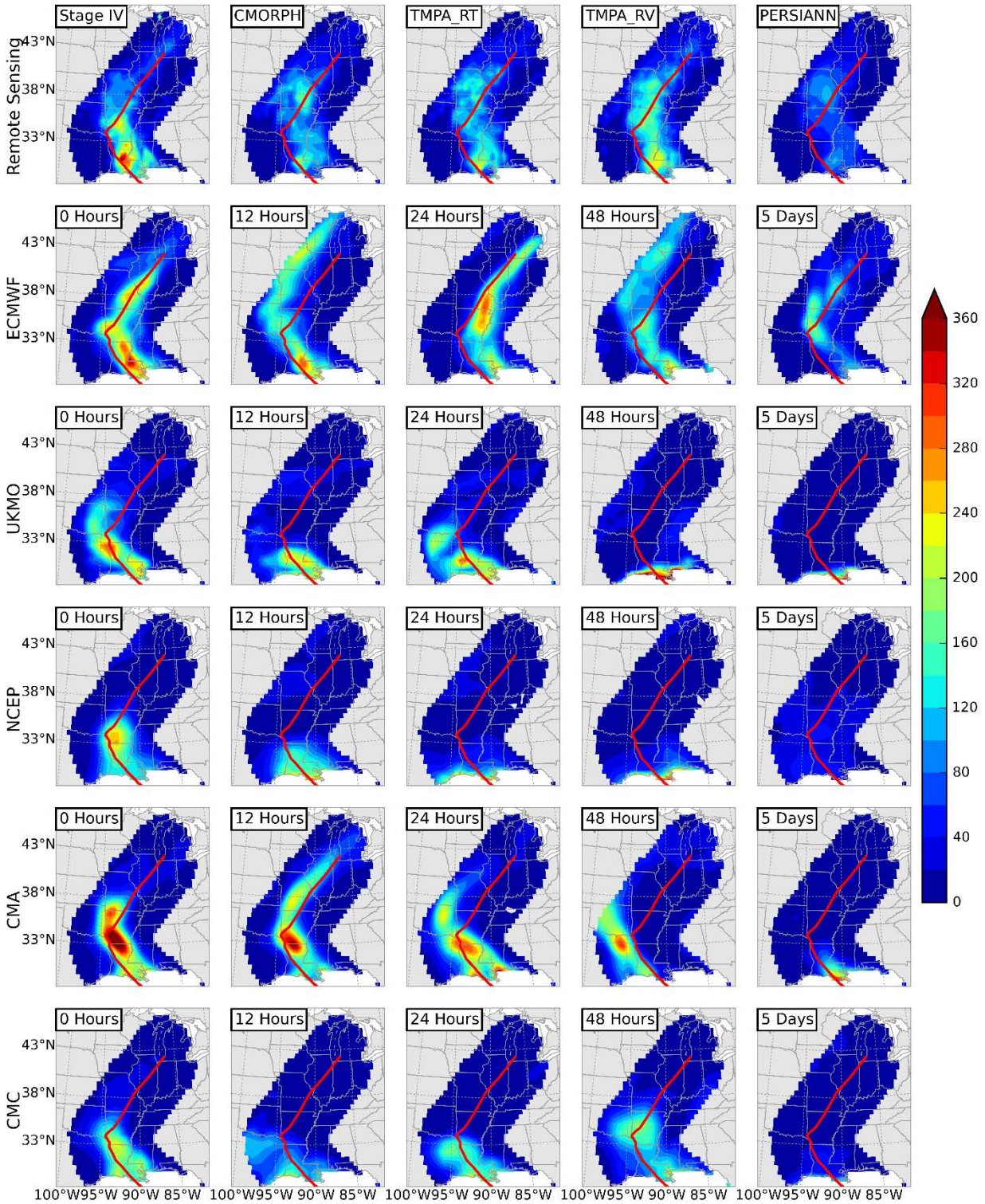

Figure B-20: Rainfall accumulation (in mm) (in mm) for Hurricane Gustav (August 31-September 5 2008) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

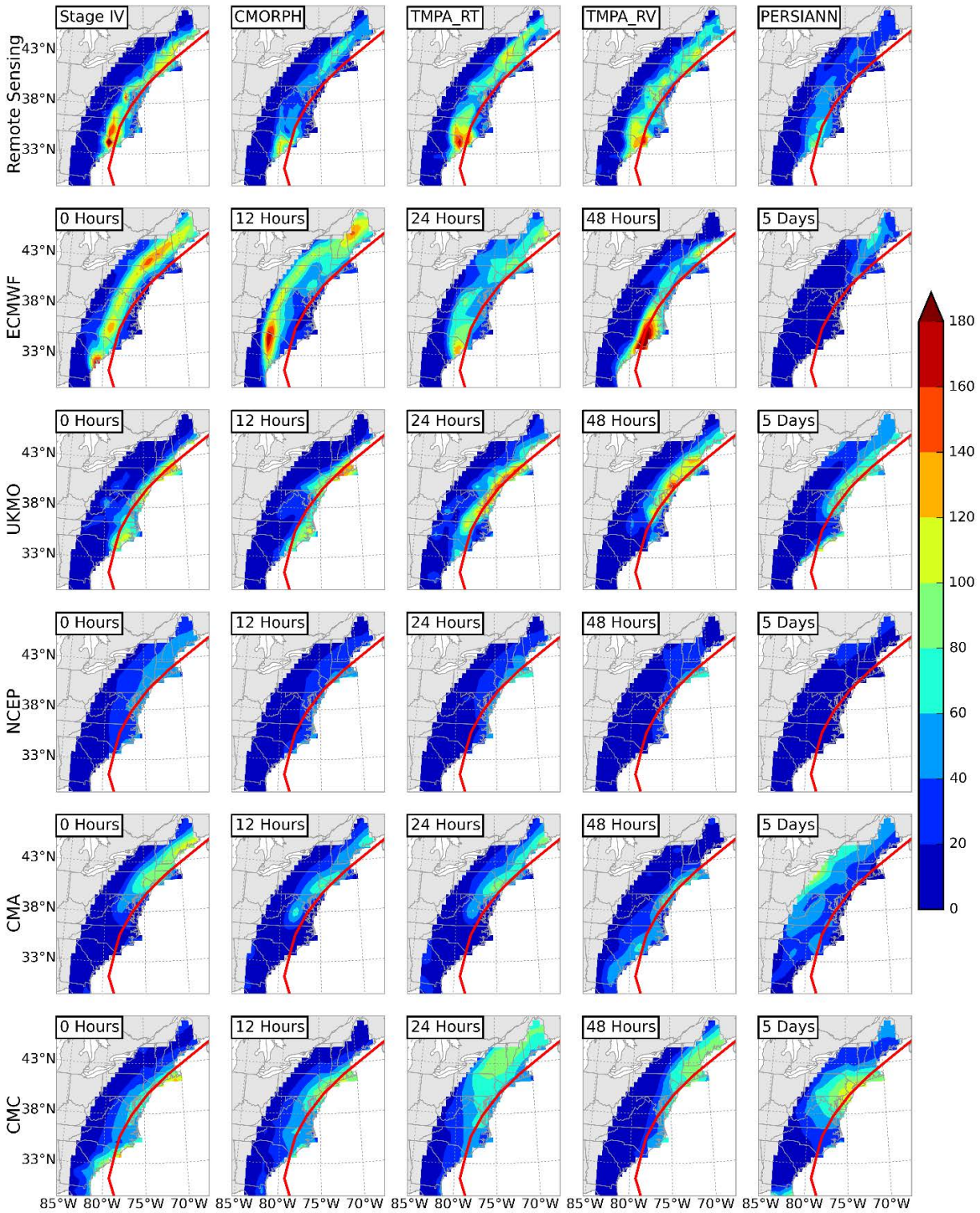

Figure B-21: Rainfall accumulation (in mm) (in mm) for Tropical Storm Hanna (September 5-8 2008) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

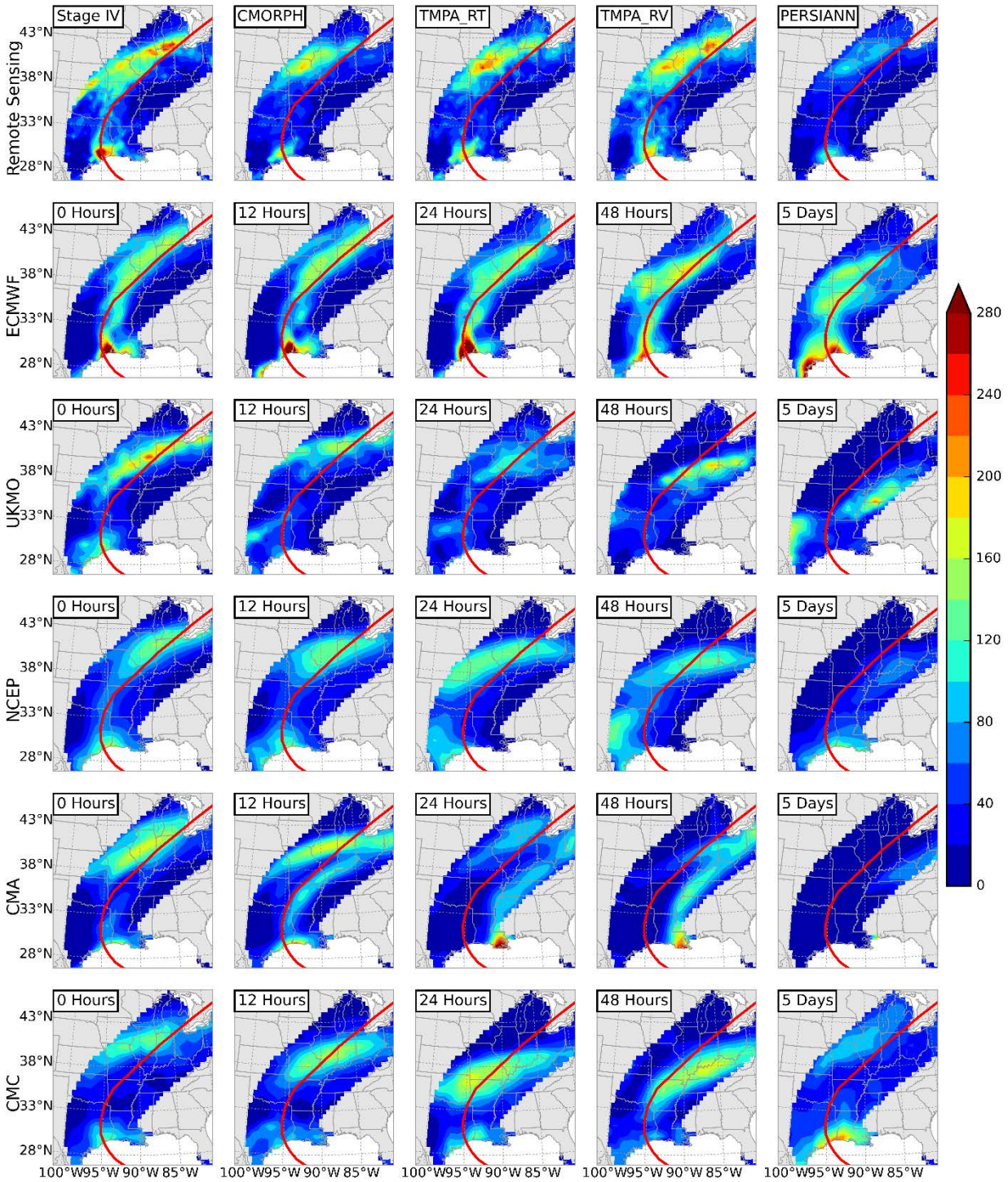

Figure B-22: Rainfall accumulation (in mm) (in mm) for Hurricane Ike (September 12-15 2008) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

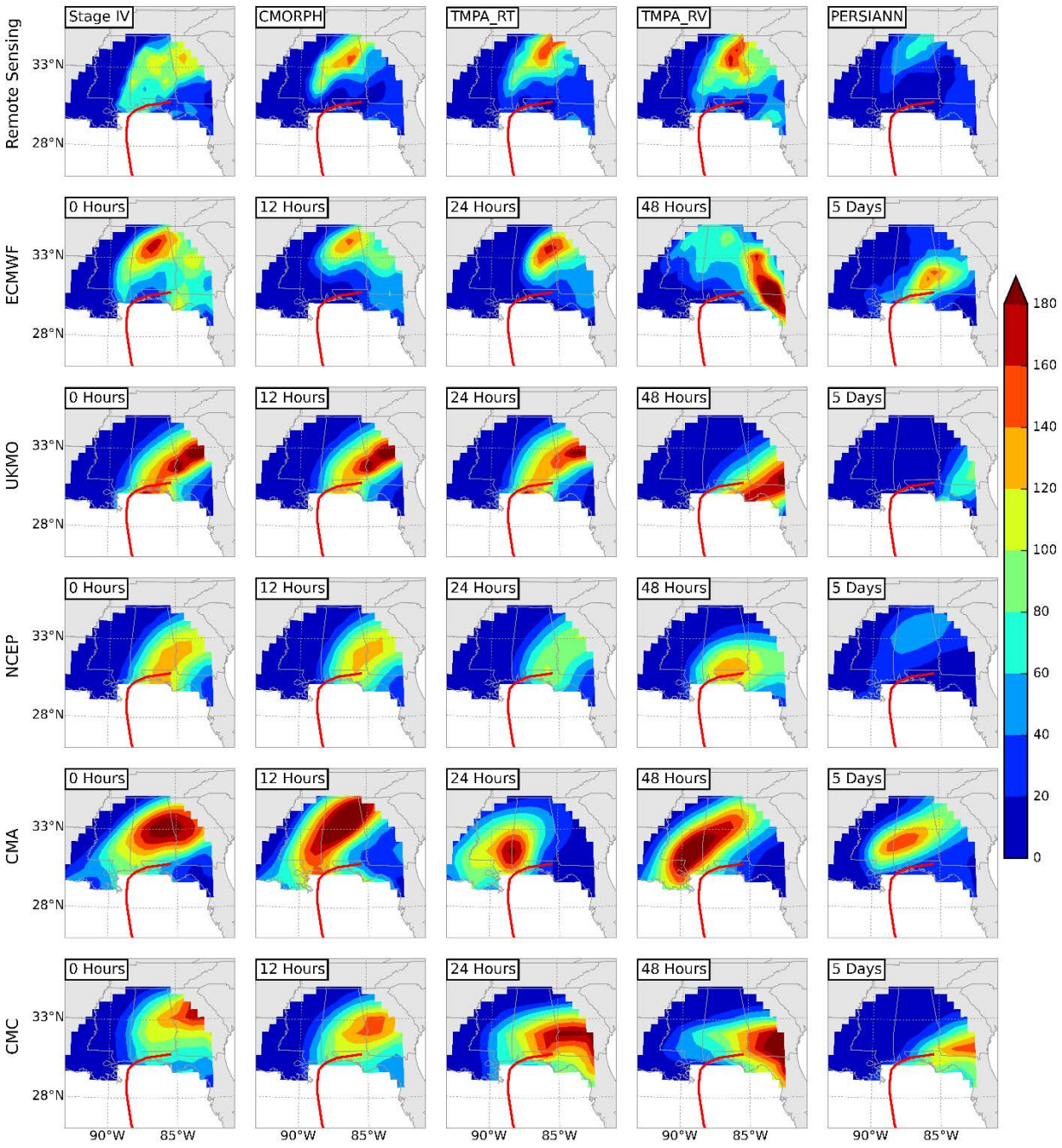

Figure B-23: Rainfall accumulation (in mm) (in mm) for Hurricane Ida (November 10-12 2009) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

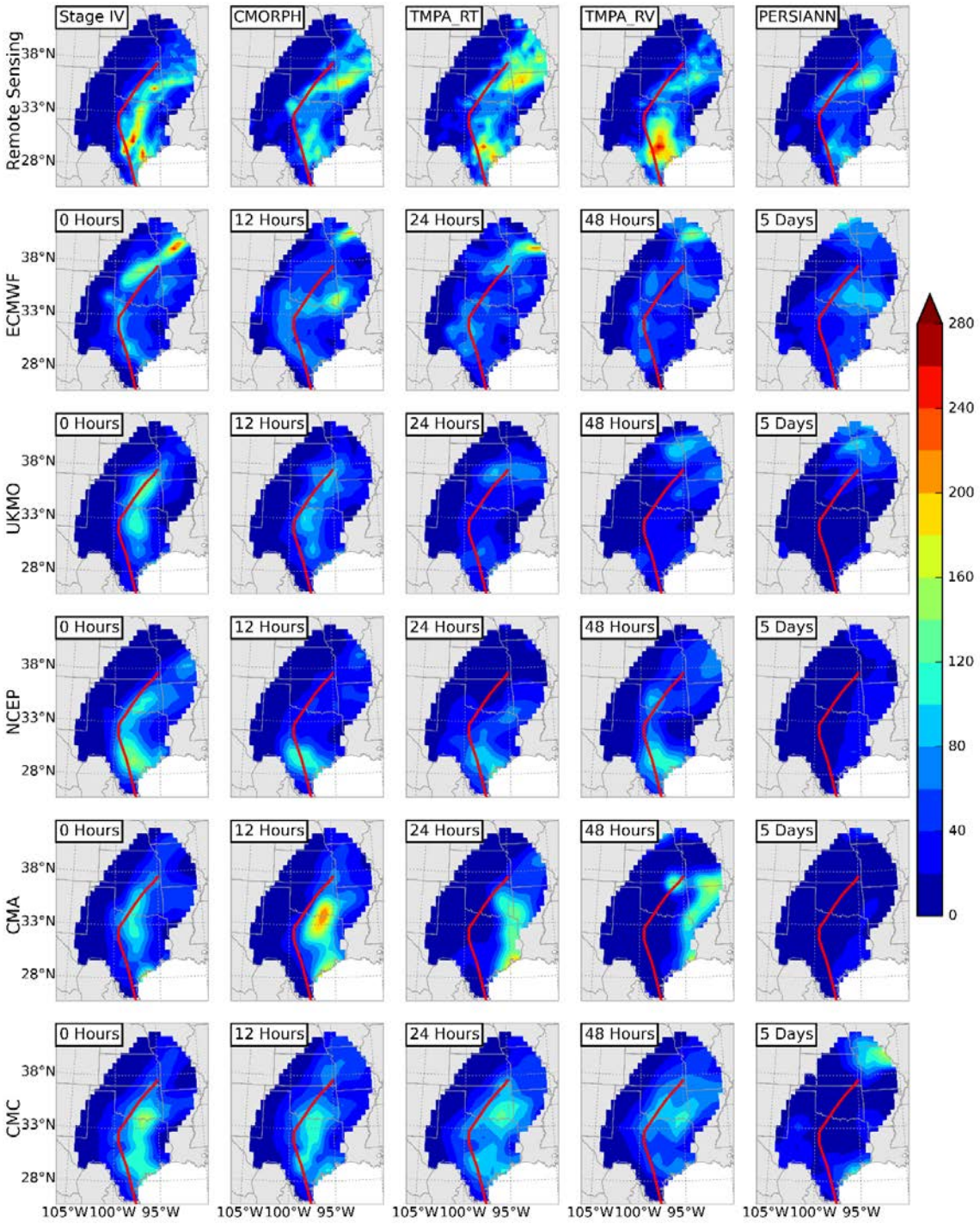

Figure B-24: Rainfall accumulation (in mm) (in mm) for Tropical Storm Hermine (September 6-11 2010) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

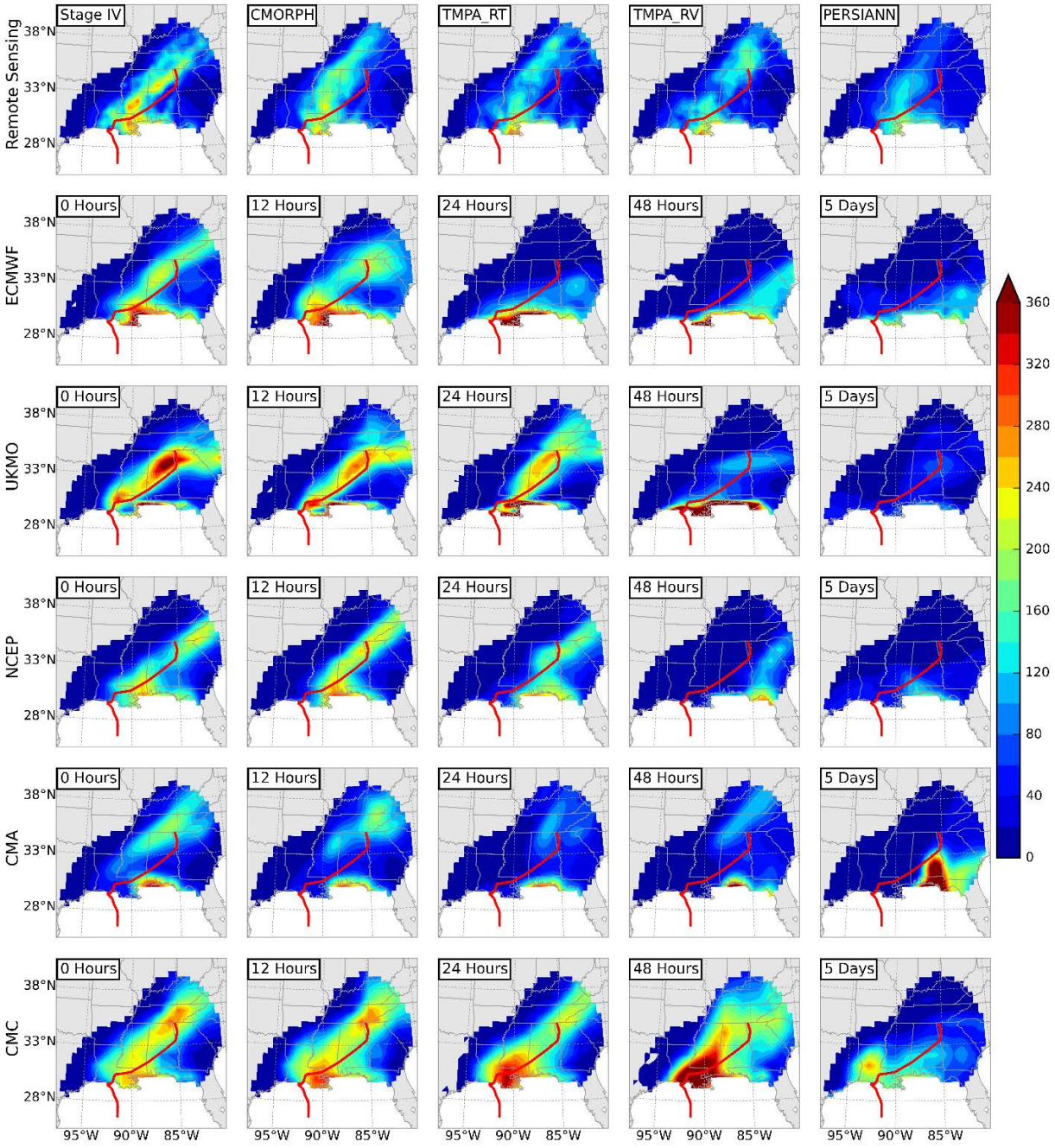

Figure B-25: Rainfall accumulation (in mm) (in mm) for Tropical Storm Lee (September 3-8 2011) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

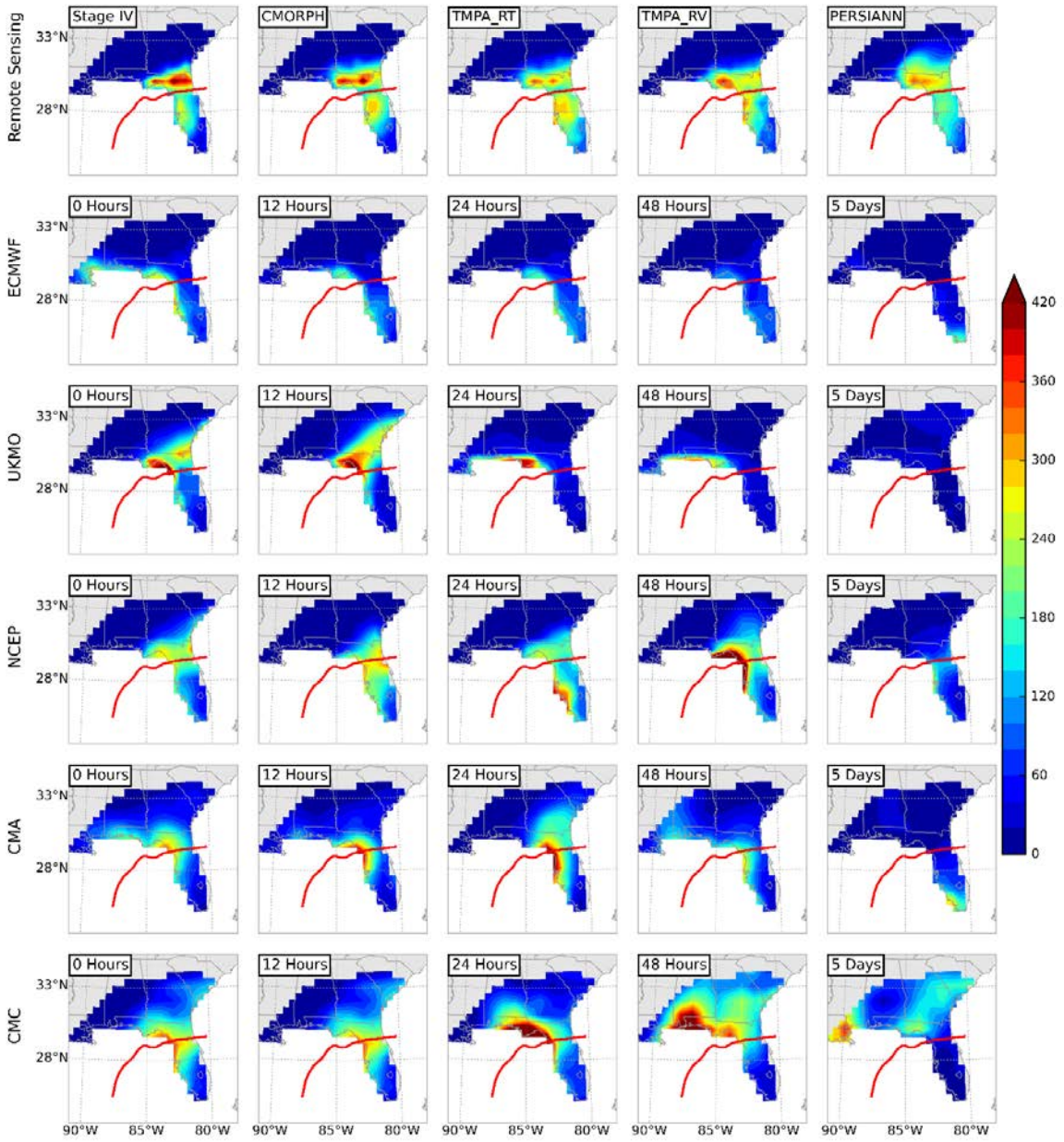

Figure B-26: Rainfall accumulation (in mm) (in mm) for Tropical Storm Debby (June 25-28 2012) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

.

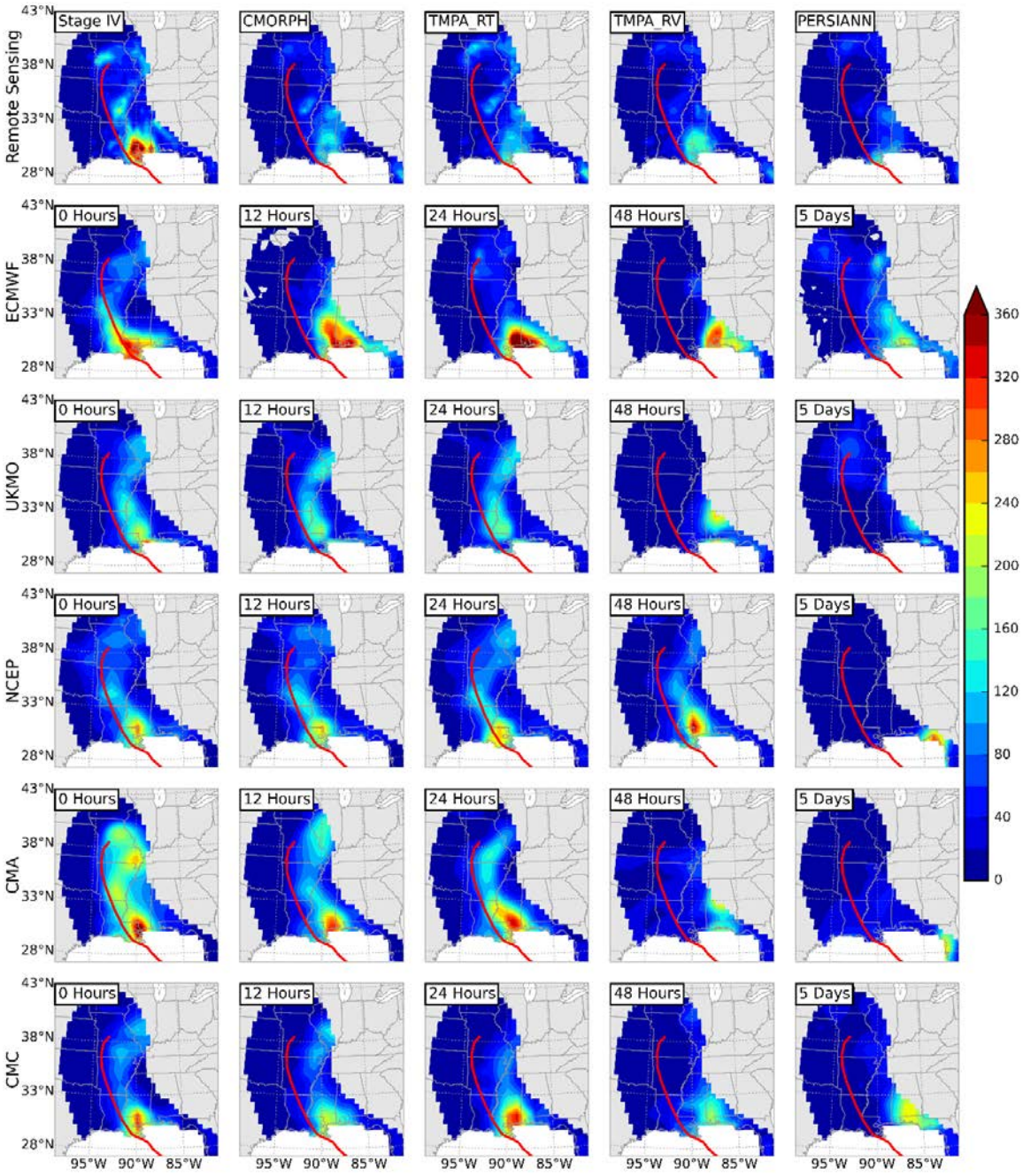

Figure B-27: Rainfall accumulation (in mm) (in mm) for Hurricane Isaac (August 28-September 2 2012) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

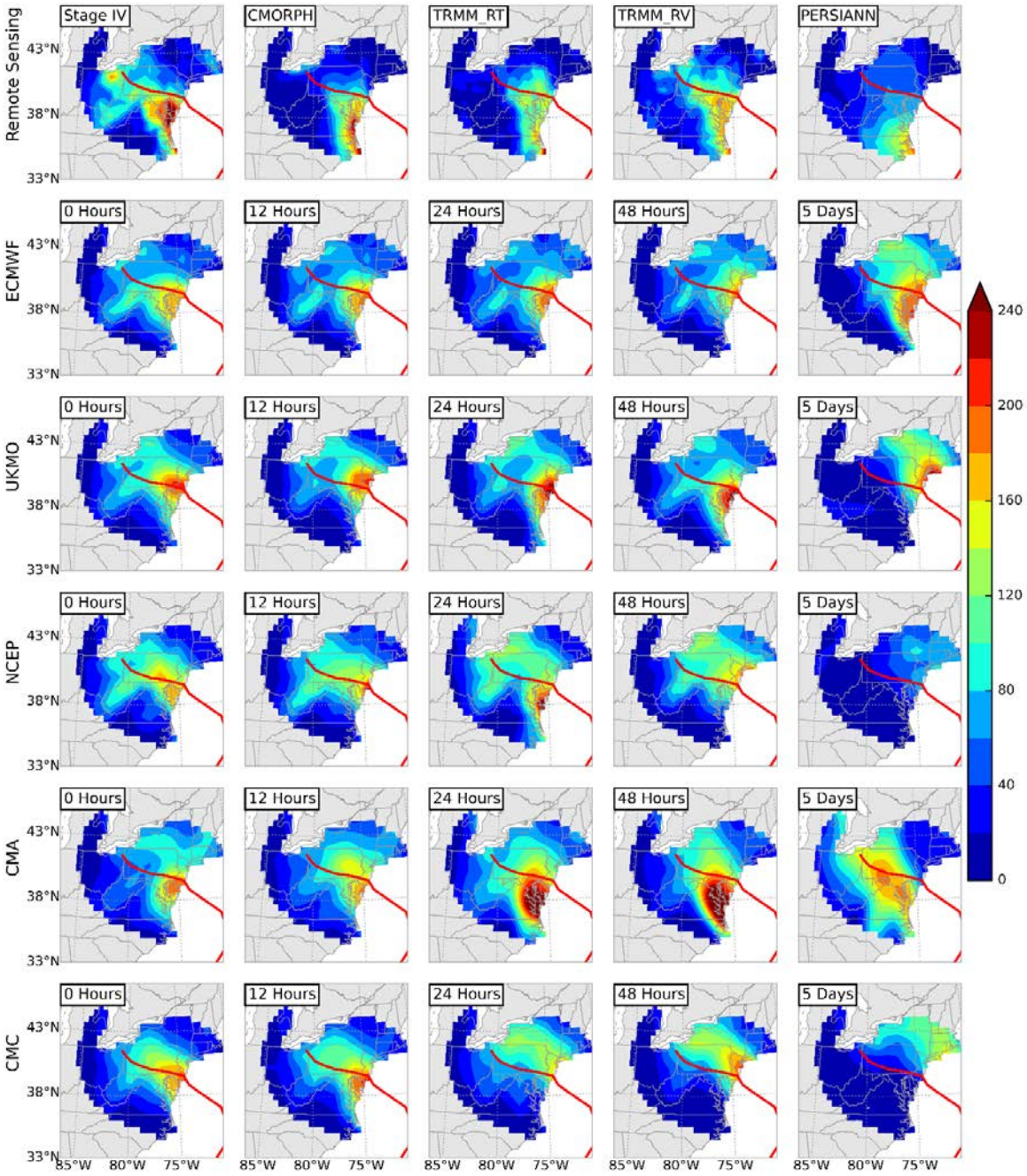

Figure B-28: Rainfall accumulation (in mm) (in mm) for Hurricane Sandy (October 28-31 2012) based on remote sensing (top row) and NWP models (from second to last rows). For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

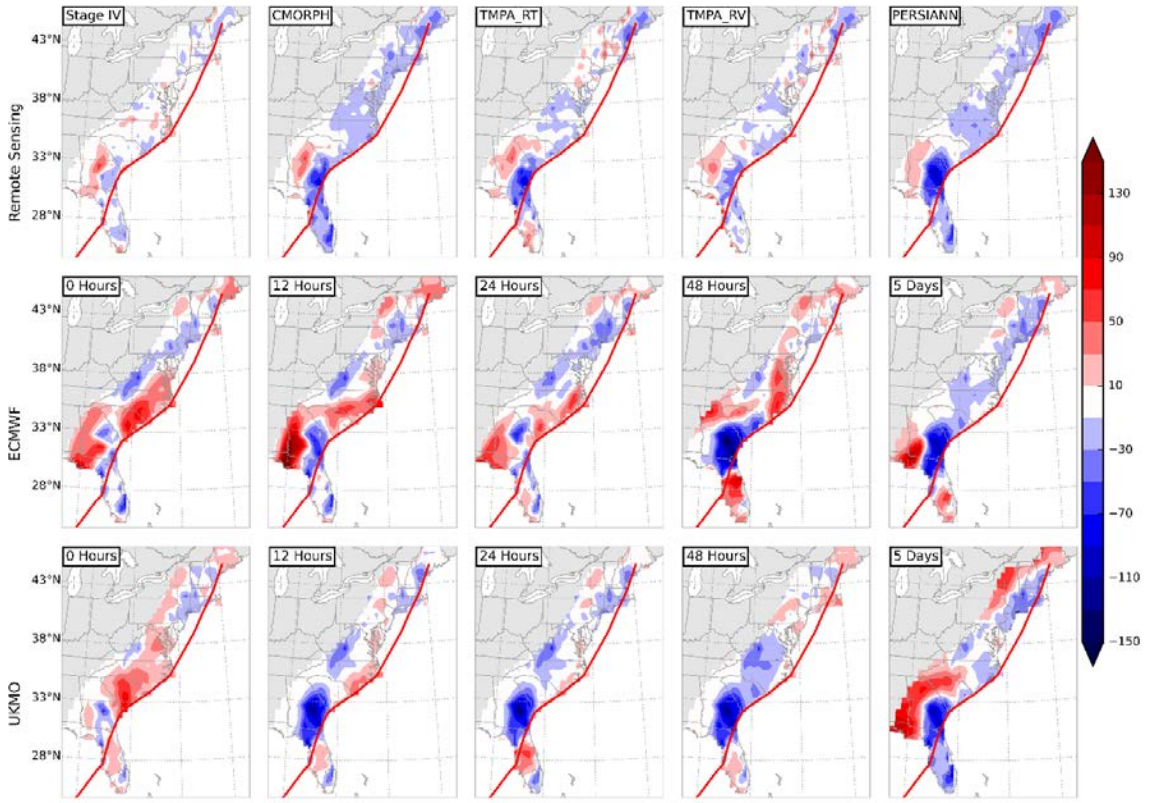

Figure B-29: Rainfall errors (in mm) for Tropical Storm Barry (June 2-6 2007) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

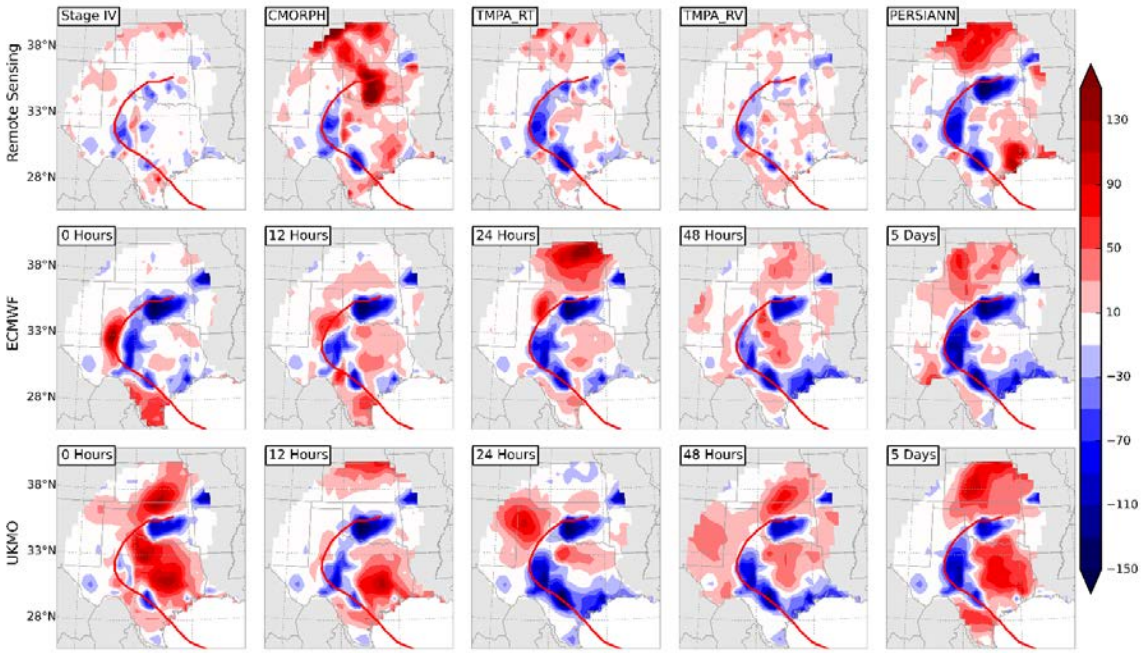

Figure B-30: Rainfall errors (in mm) for Tropical Storm Erin (August 16-21 2007) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

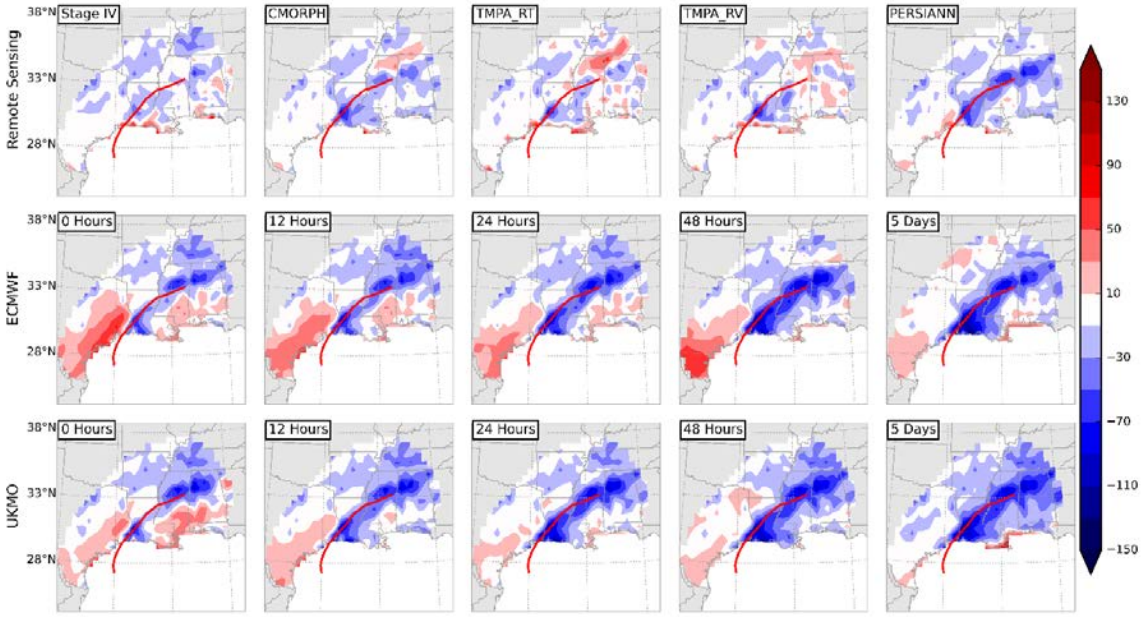

Figure B-31: Rainfall errors (in mm) for Hurricane Humberto (September 11-15 2007) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track. Forecasts for NCEP, CMA and CMC were not available.

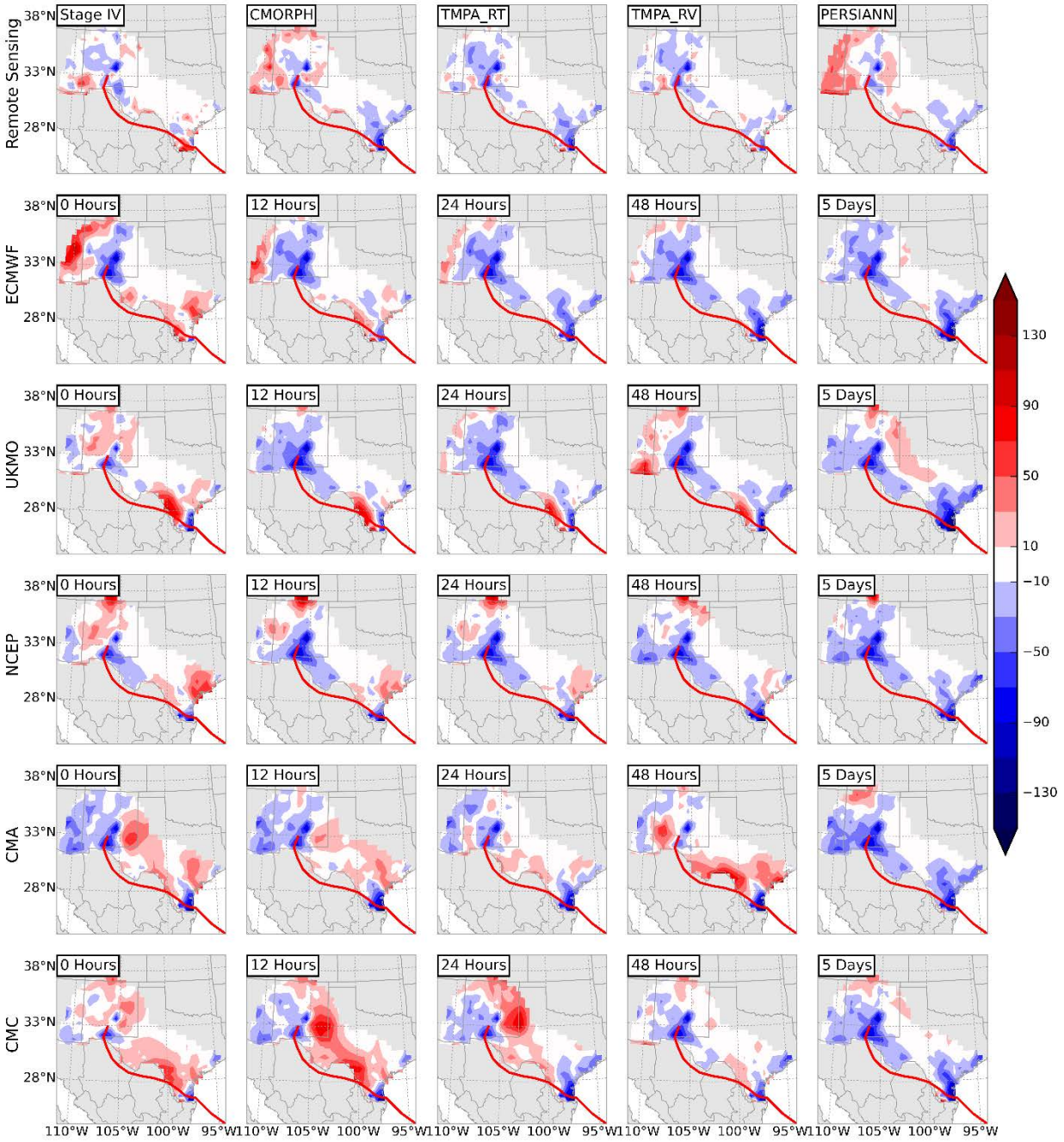

Figure B-32: Rainfall errors (in mm) for Hurricane Dolly (July 23-28 2008) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

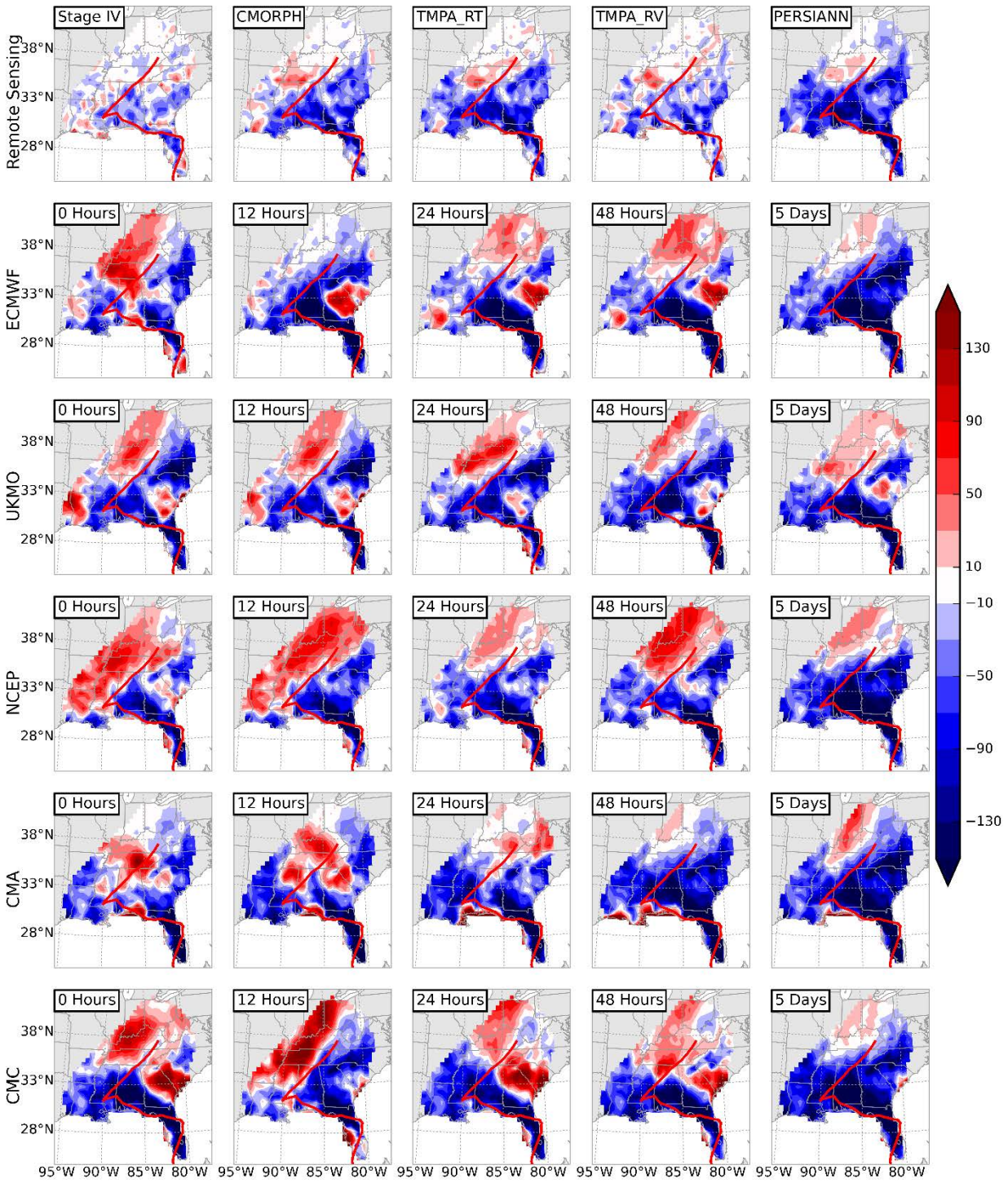

Figure B-33: Rainfall errors (in mm) for Tropical Storm Fay (August 19-28 2008) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

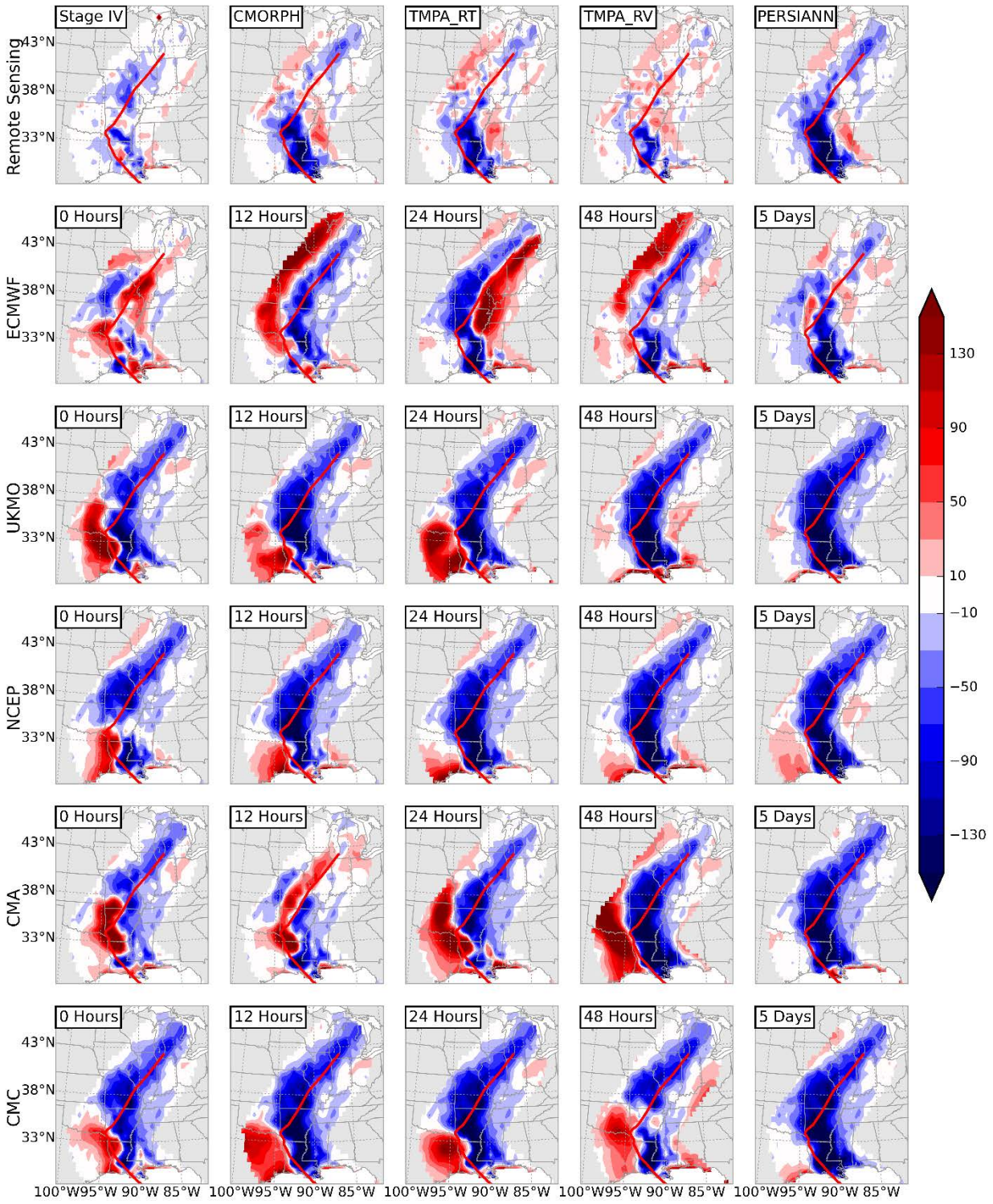

Figure B-34: Rainfall errors (in mm) for Hurricane Gustav (August 31- September 5 2008) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

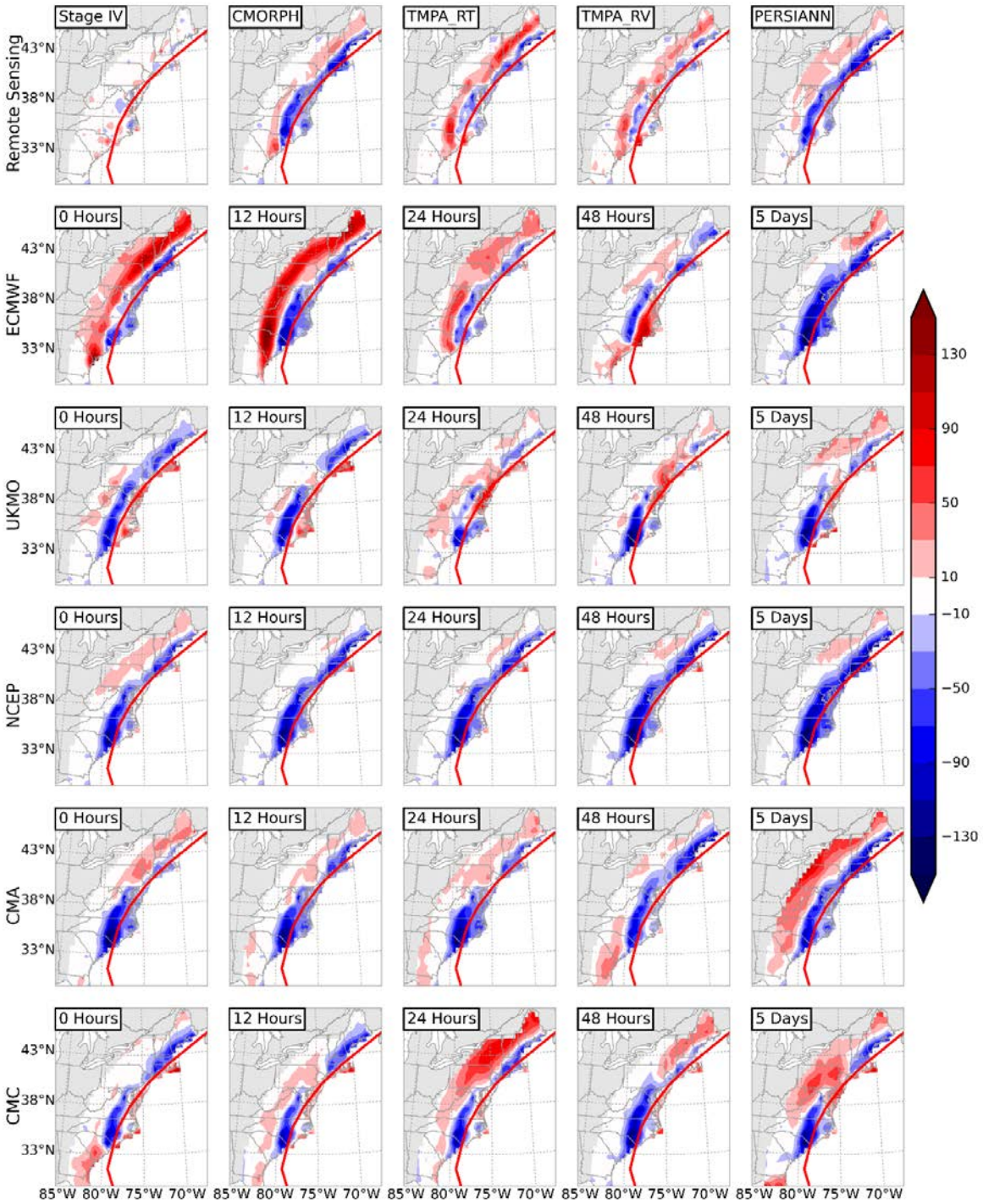

Figure B-35: Rainfall errors (in mm) for Tropical Storm Hanna (September 5-8 2008 based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

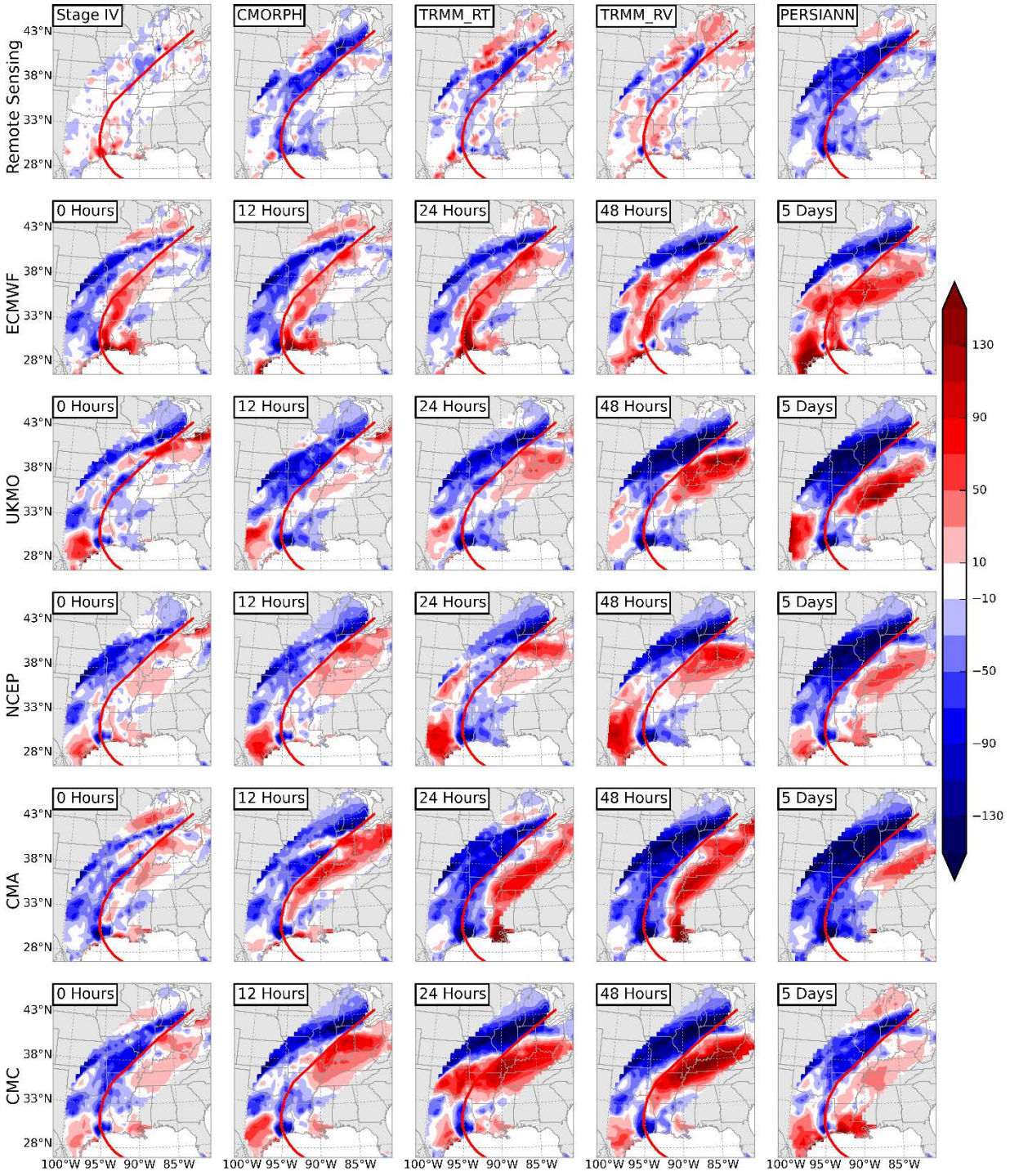

Figure B-36: Rainfall errors (in mm) for Hurricane Ike (September 12-15 2008) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

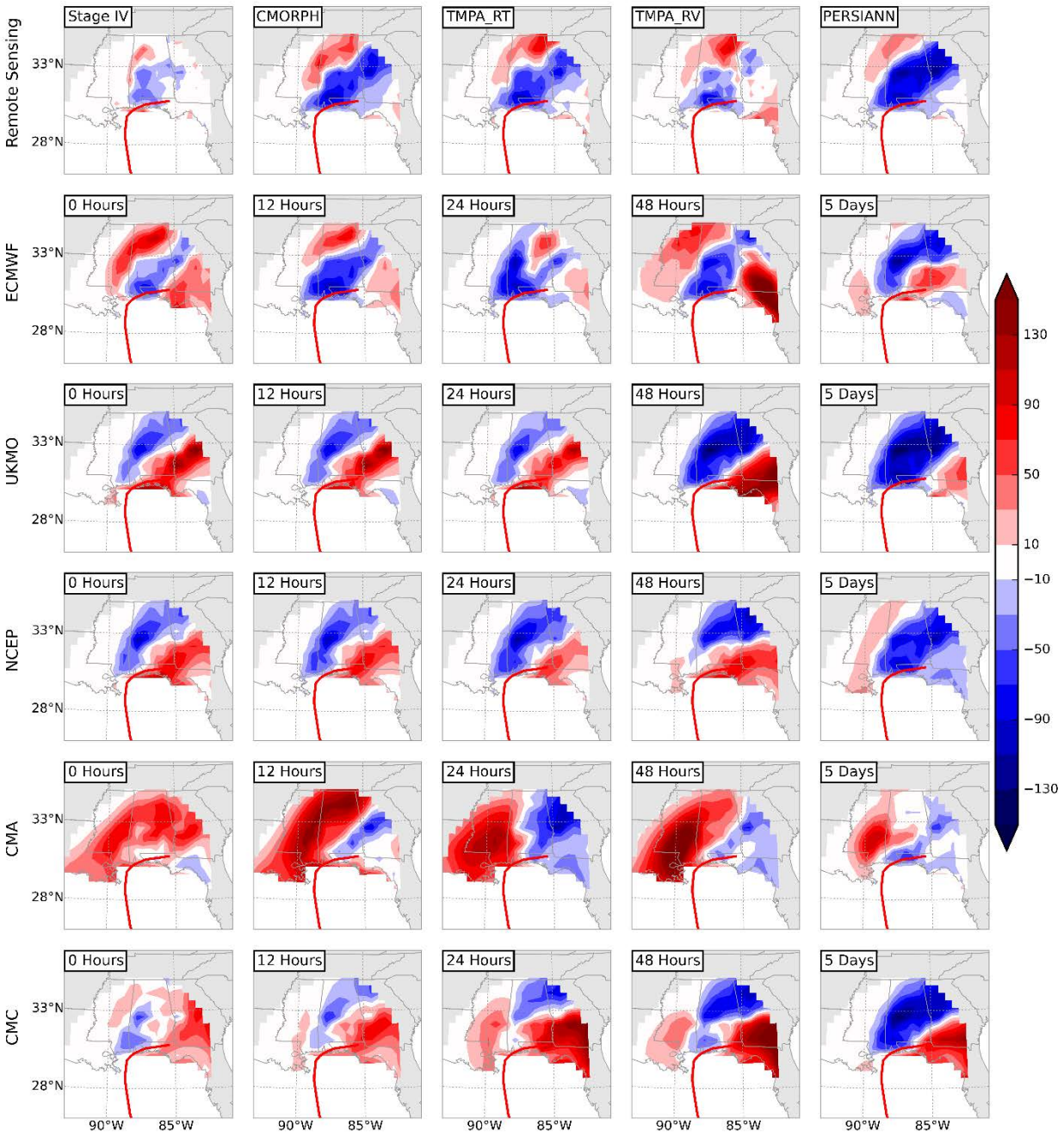

Figure B-37: Rainfall errors (in mm) for Hurricane Ida (November 10-12 2009) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

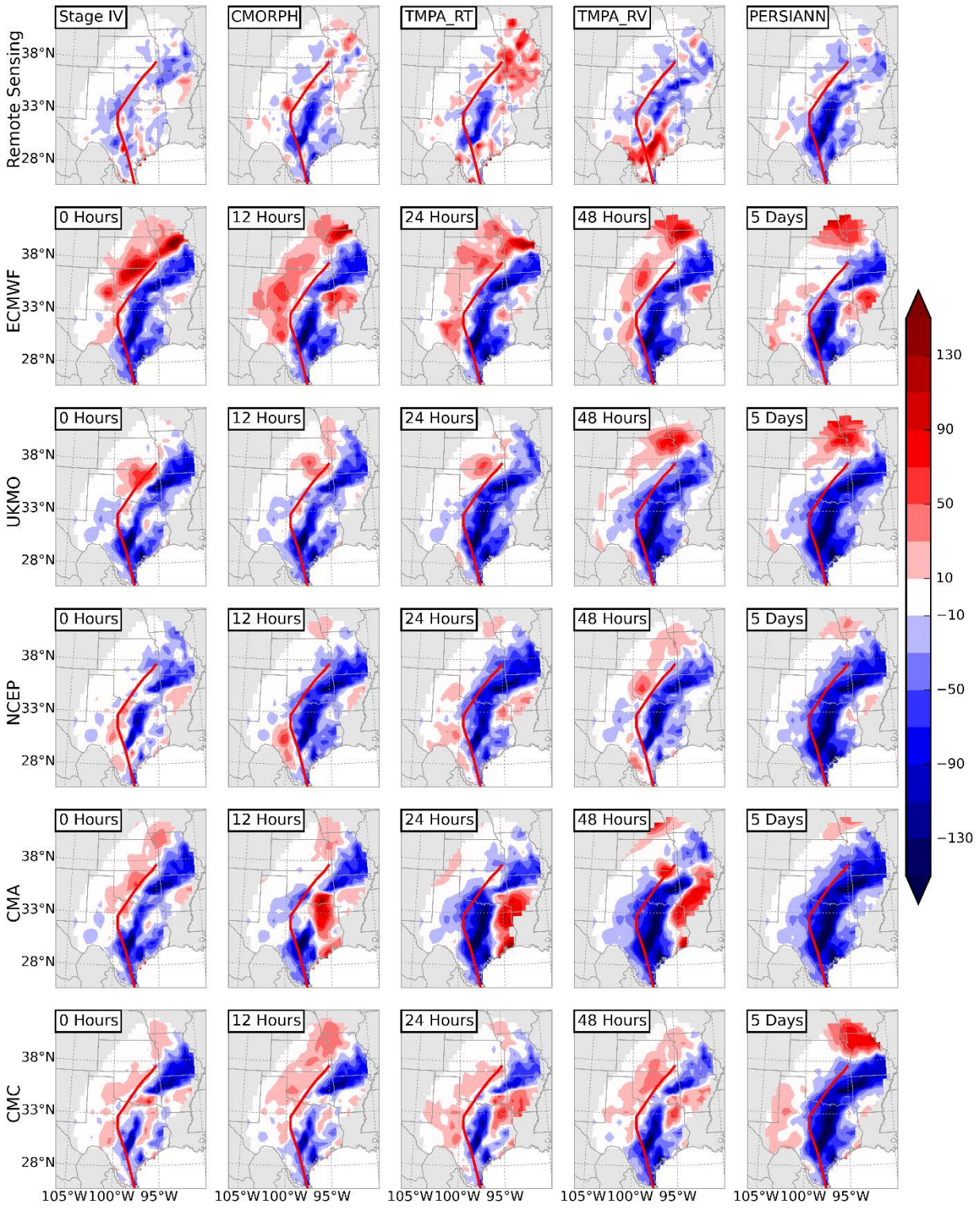

Figure B-38: Rainfall errors (in mm) for Tropical Storm Hermine (September 6-11 2010) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

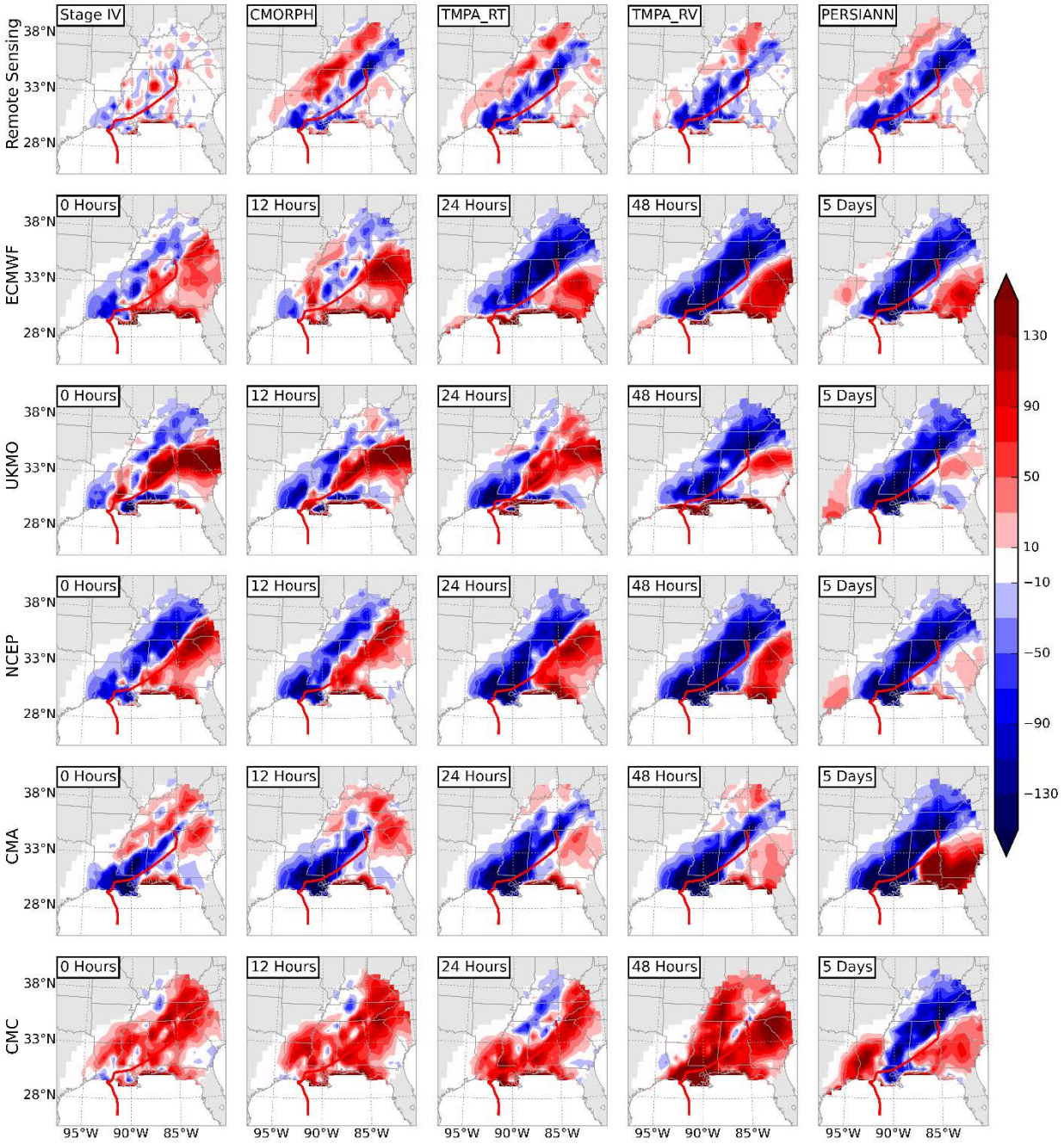

Figure B-39: Rainfall errors (in mm) for Tropical Storm Lee (September 3-8 2011) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

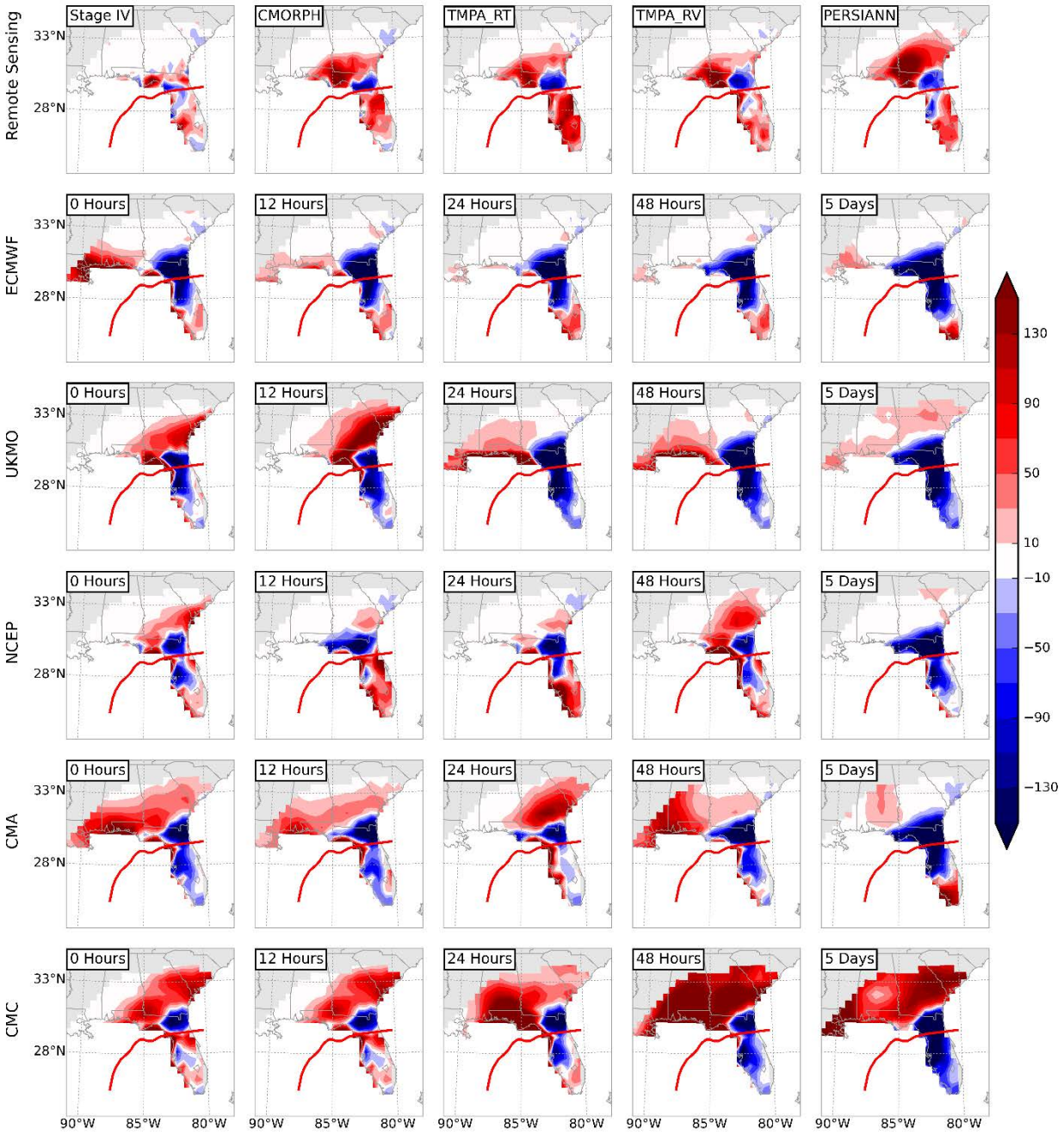

Figure B-40: Rainfall errors (in mm) for Tropical Storm Debby (June 25-28 2012) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

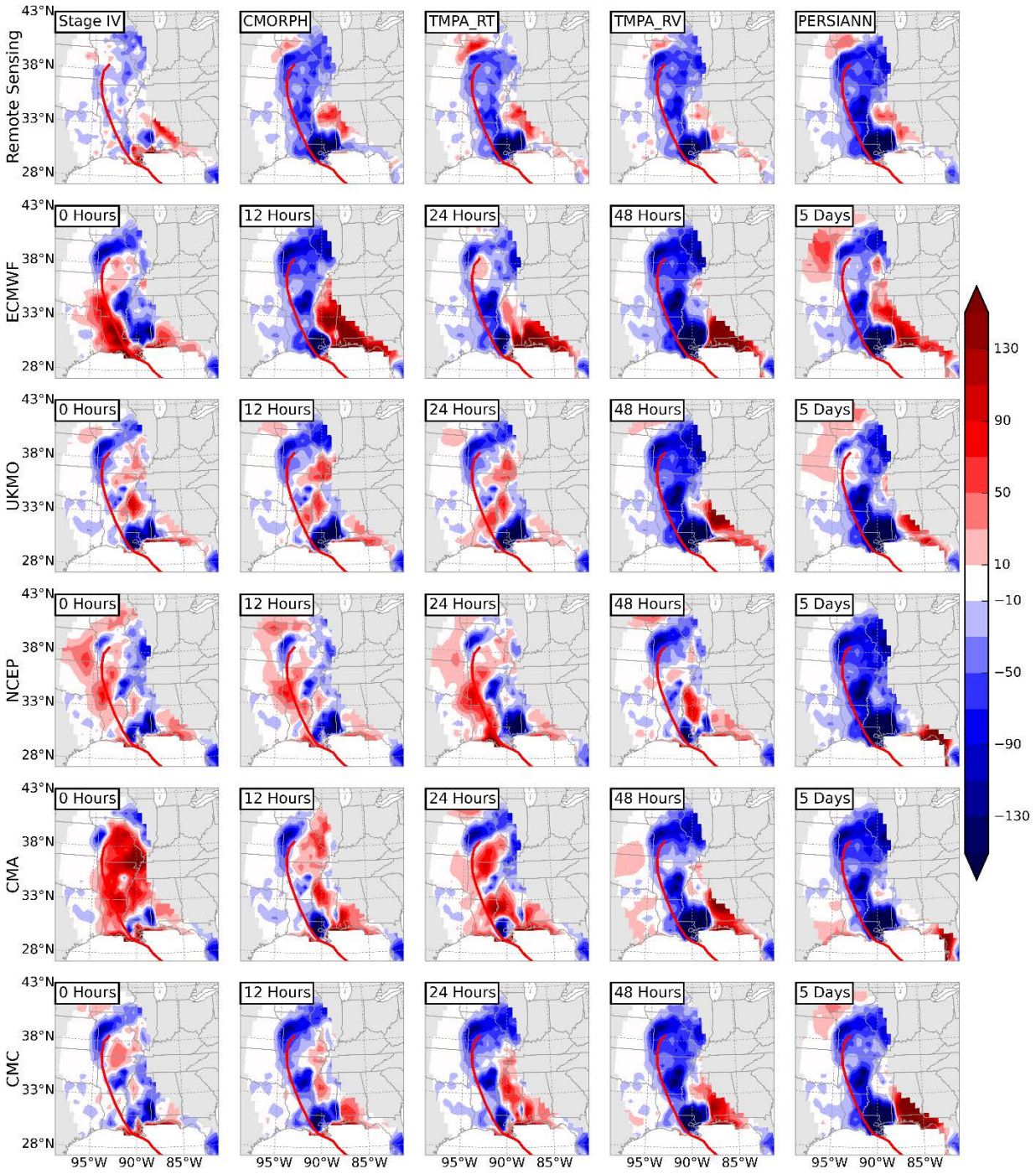

Figure B-41: Rainfall errors (in mm) for Hurricane Isaac (August 28- September 2 2012) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

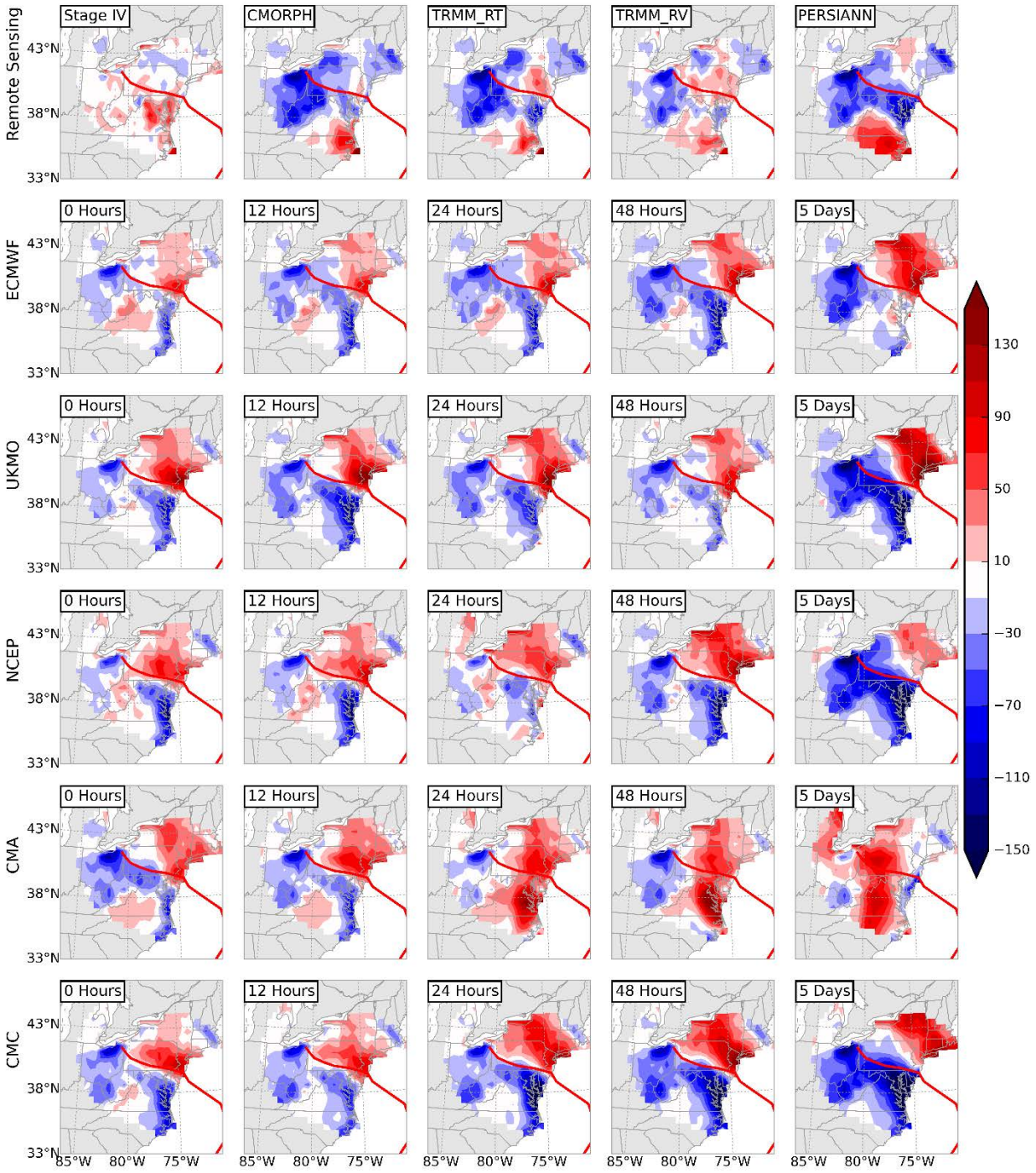

Figure B-42: Rainfall errors (in mm) for Hurricane Sandy (October 28-31 2012) based on remote sensing (top row) and NWP models (second to last rows) The errors are computed with respect to the CPC data. For the NWP models, lead-times increases left to right from 0 hour to 5 days. The red lines represent the storm track.

REFERENCES

Avila, L. A., and J. Cangialosi, 2011: Tropical Cyclone Report: Hurricane Irene. *National Hurricane Center*.

Barlow, M., 2011: Influence of hurricane-related activity on North American extreme precipitation. *Geophysical Research Letters*, **38**.

Bell, G. D., and M. Chelliah, 2006: Leading tropical modes associated with interannual and multidecadal fluctuations in North Atlantic hurricane activity. *Journal of Climate*, **19,** 590-612.

Bister, M., and K. A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, **65,** 233-240.

Blake, E. S., C. Landsea, and E. J. Gibney, 2007: *The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts).* NOAA/National Weather Service, National Centers for Environmental Prediction, National Hurricane Center.

Blake, E. S., C. W. Landsea, E. J. Gibney, and N. Asheville, 2011: The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)

Bougeault, P., and Coauthors, 2010: The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, **91,** 1059-1072.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **134,** 703-722.

Buizza, R., J. R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (variable resolution ensemble prediction system). *Quarterly Journal of the Royal Meteorological Society*, **133,** 681-695.

Camargo, S. J., and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Weather and Forecasting*, **24,** 472-491.

Camargo, S. J., A. G. Barnston, P. J. Klotzbach, and C. W. Landsea, 2007: Seasonal tropical cyclone forecasts. *Wmo Bulletin*, **56,** 297.

Camp, J., and Coauthors, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*.

Changnon, S. A., 2008: Assessment of flood losses in the United States. *Journal of Contemporary Water Research & Education*, **138,** 38-44.

Czajkowski, J., G. Villarini, E. Michel-Kerjan, and J. A. Smith, 2013: Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. *Environmental Research Letters*, **8,** 044056.

Czajkowski, J., and J. Done, 2014: As the wind blows? Understanding hurricane damages at the local level through a case study analysis. *Weather, climate, and society*, **6,** 202-217.

Elsberry, R. L., 2002: Predicting hurricane landfall precipitation: Optimistic and pessimistic views from the symposium on precipitation extremes. *Bulletin of the American Meteorological Society*, **83,** 1333-1339.

Elsner, J. B., and C. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Weather and Forecasting*, **8,** 345-351.

Elsner, J. B., and A. B. Kara, 1999: *Hurricanes of the North Atlantic: Climate and society.* Oxford University Press.

Elsner, J. B., and T. H. Jagger, 2006: Prediction Models for Annual U.S. Hurricane Counts. *Journal of Climate*, **19,** 2935-2952.

Elsner, J. B., X. Niu, and T. H. Jagger, 2004: Detecting shifts in hurricane rates using a Markov chain Monte Carlo approach. *Journal of Climate*, **17,** 2652-2666.

Emanuel, K, 2007: Environmental factors affecting tropical cyclone power dissipation. *Journal of Climate*, **20,** 5497-5509.

Gray, W. M., 1984a: Atlantic Seasonal Hurricane Frequency. Part I: El Niño and 30 mb Quasi-Biennial Oscillation Influences. *Monthly Weather Review*, **112,** 1649-1668.

——, 1984b: Atlantic Seasonal Hurricane Frequency. Part II: Forecasting its Variability. *Monthly Weather Review*, **112,** 1669-1683.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry, 1992: Predicting Atlantic Seasonal Hurricane Activity 6–11 Months in Advance. *Weather and Forecasting*, **7,** 440-455.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept. *Tellus A*, **57,** 219-233.

Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Weather and Forecasting*, **28,** 1423-1445.

Higgins, R., W. Shi, E. Yarosh, and R. Joyce, 2000: Improved US precipitation quality control system and analysis. NCEP/Climate Prediction Center ATLAS No. 7. *Camp Springs, Maryland. available at http://www. cpc. ncep. noaa. gov/research_papers/ncep_cpc_atlas/7/index. html*.

Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *Journal of Climate*, **10,** 1769-1786.

Houtekamer, P., H. L. Mitchell, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, **137,** 2126-2143.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM multi-satellite precipitation analysis (TMPA). *Satellite rainfall applications for surface hydrology*, Springer, 3-22.

Jiang, H., and E. J. Zipser, 2010: Contribution of tropical cyclones to the global precipitation from eight seasons of TRMM data: Regional, seasonal, and interannual variations. *Journal of climate*, **23,** 1526-1543.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 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,** 487-503.

Kam, J., J. Sheffield, X. Yuan, and E. F. Wood, 2013: The influence of Atlantic tropical cyclones on drought over the eastern United States (1980–2007). *Journal of Climate*, **26,** 3067-3086.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, **95,** 585-601.

Klotzbach, P. J., 2014: Prediction of seasonal Atlantic basin accumulated cyclone energy from 1 July. *Weather and Forecasting*, **29,** 115-121.

Kunkel, K. E., D. R. Easterling, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2010: Recent increases in US heavy precipitation associated with tropical cyclones. *Geophysical Research Letters*, **37**.

Laio, F., and S. Tamea, 2007: Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences Discussions*, **11,** 1267-1277.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Monthly Weather Review*, **141,** 3576-3592.

Latif, M., N. Keenlyside, and J. Bader, 2007: Tropical sea surface temperature, vertical wind shear, and hurricane development. *Geophysical Research Letters*, **34**.

Lavers, D. A., and G. Villarini, 2013: Were global numerical weather prediction systems capable of forecasting the extreme Colorado rainfall of 9–16 September 2013? *Geophysical Research Letters*, **40,** 6405-6410.

Lehmiller, G., T. Kimberlain, and J. Elsner, 1997: Seasonal prediction models for North Atlantic basin hurricane location. *Monthly Weather Review*, **125,** 1780-1791.

Lin, N., J. A. Smith, G. Villarini, T. P. Marchok, and M. L. Baeck, 2010: Modeling extreme rainfall, winds, and surge from Hurricane Isabel (2003). *Weather and Forecasting*, **25,** 1342-1361.

Lin, N., P. Lane, K. A. Emanuel, R. M. Sullivan, and J. P. Donnelly, 2014: Heightened hurricane surge risk in northwest Florida revealed from climatological-hydrodynamic modeling and paleorecord reconstruction. *Journal of Geophysical Research: Atmospheres*, **119,** 8606-8623.

Lin, Y., and K. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints, 19th Conf. on Hydrology, San Diego, CA, Am. Meteorol. Soc., 1.2.

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

Marchok, T., R. Rogers, and R. Tuleya, 2007: Validation schemes for tropical cyclone quantitative precipitation forecasts: Evaluation of operational models for US landfalling cases. *Weather and forecasting*, **22,** 726-746.

Mason, S. J., and O. Baddour, 2008: *Statistical Modelling. In Seasonal Climate: Forecasting and Managing Risk.* Vol. 82, Springer Science & Business Media, 163-201 pp.

Maxwell, J. T., P. T. Soulé, J. T. Ortegren, and P. A. Knapp, 2012: Drought-busting tropical cyclones in the southeastern Atlantic United States: 1950–2008. *Annals of the Association of American Geographers*, **102,** 259-275.

Maxwell, J. T., J. T. Ortegren, P. A. Knapp, and P. T. Soulé, 2013: Tropical cyclones and drought amelioration in the gulf and Southeastern coastal United States. *Journal of Climate*, **26,** 8440-8452.

McCallum, B. E., J. A. Painter, and E. R. Frantz, 2012: Monitoring inland storm tide and flooding from Hurricane Irene along the Atlantic Coast of the United States, August 20112331-1258.

Mohanty, U., K. K. Osuri, R. Nadimpalli, and S. Gopalakrishnan, 2014: Uncertainty in rainfall prediction of land-falling tropical cyclones over India: Impact of data assimilation. *3rd International Workshop on Tropical Cyclone Landfall Processes (IWTCLP-III), Jeju*, 8-10.

Murakami, H., G. Villarini, G. A. Vecchi, W. Zhang, and R. Gudgel, 2016: Statistical-dynamical seasonal forecast of North Atlantic and US Landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Monthly Weather Review*.

Palmer, T. N., F. J. Doblas-Reyes, R. Hagedorn, and A. Weisheimer, 2005: Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360,** 1991-1998.

Pielke Jr, R. A., and C. N. Landsea, 1999: La nina, el nino and atlantic hurricane damages in the united states. *Bulletin of the American Meteorological Society*, **80,** 2027-2033.

Ramsay, H. A., and A. H. Sobel, 2011: Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *Journal of Climate*, **24,** 183-193.

Rappaport, E. N., 2000: Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bulletin of the American Meteorological Society*, **81,** 2065-2073.

——, 2014: Fatalities in the United States from Atlantic tropical cyclones: new data and interpretation. *Bulletin of the American Meteorological Society*, **95,** 341-346.

Rayner, N. A., and Coauthors, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **108**.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **126,** 649-667.

Rigby, R. A., and D. M. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54,** 507-554.

——, 2007: Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54,** 507-554.

Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Monthly Weather Review*, **132,** 2732-2744.

Smith, A. B., and R. W. Katz, 2013: US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural hazards*, **67,** 387-410.

Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skilful multi-year predictions of Atlantic hurricane frequency. *Nature geoscience*, **3,** 846-849.

Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, **81,** 2035-2046.

Swanson, K. L., 2007: Impact of scaling behavior on tropical cyclone intensities. *Geophysical Research Letters*, **34**.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **365,** 2053-2075.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, **125,** 3297-3319.

Vecchi, G. A., and G. Villarini, 2014: Next Season's Hurricanes. *Science*, **343,** 618-619.

Vecchi, G. A., K. L. Swanson, and B. J. Soden, 2008: Whither hurricane activity. *Science*, **322,** 687-689.

Vecchi, G. A., and Coauthors, 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity. *Monthly Weather Review*, **139,** 1070-1082.

Vecchi, G. A., and Coauthors, 2013: Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations. *Journal of Climate*, **26,** 5337-5357.

Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, **27,** 7994-8016.

Villarini, G., and J. A. Smith, 2010: Flood peak distributions for the eastern United States. *Water Resources Research*, **46**.

Villarini, G., and G. A. Vecchi, 2012: North Atlantic Power Dissipation Index (PDI) and Accumulated Cyclone Energy (ACE): Statistical modeling and sensitivity to sea surface temperature changes. *Journal of Climate*, **25,** 625-637.

——, 2013: Multiseason Lead Forecast of the North Atlantic Power Dissipation Index (PDI) and Accumulated Cyclone Energy (ACE). *Journal of Climate*, **26,** 3631-3643.

Villarini, G., and J. A. Smith, 2013: Flooding in Texas: Examination of temporal changes and impacts of tropical cyclones. *JAWRA Journal of the American Water Resources Association*, **49,** 825-837.

Villarini, G., G. A. Vecchi, and J. A. Smith, 2010: Modeling the Dependence of Tropical Storm Counts in the North Atlantic Basin on Climate Indices. *Monthly Weather Review*, **138,** 2681-2705.

Villarini, G., R. Goska, J. A. Smith, and G. A. Vecchi, 2014: North Atlantic tropical cyclones and US flooding. *Bulletin of the American Meteorological Society*, **95,** 1381-1388.

Villarini, G., J. A. Smith, M. L. Baeck, T. Marchok, and G. A. Vecchi, 2011: Characterization of rainfall distribution and flooding associated with US landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research: Atmospheres (1984–2012)*, **116**.

Vitart, F., and J. Anderson, 2001: Sensitivity of Atlantic tropical storm frequency to ENSO and interdecadal variability of SSTs in an ensemble of AGCM integrations. *Journal of climate*, **14,** 533-545.

Wehner, M. F., R. L. Smith, G. Bala, and P. Duffy, 2010: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate dynamics*, **34,** 241-247.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences.* Vol. 100, Academic press.

World Meteorological Organization, W., 2008: Report from expert meeting to evaluate skill of tropical cyclone seasonal forecasts. *World Meteorological Organization Tech. Doc*, **1455,** 27.

Yun, W. T., L. Stefanova, A. K. Mitra, T. S. V. V. Kumar, W. Dewar, and T. N. Krishnamurti, 2005: A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus A*, **57,** 280-289.

Zhao, M., I. M. Held, and G. A. Vecchi, 2010: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Monthly Weather Review*, **138,** 3858-3868.

Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *Journal of Climate*, **22,** 6653-6678.

Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Advances in atmospheric sciences*, **22,** 781-788.