Stochastic Weather Generator Based Ensemble Streamflow Forecasting

by

Nina Marie Caraway

B.E., Vanderbilt University, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master's of Science

Department of Civil Engineering

2012

This thesis entitled:
Stochastic Weather Generator Based Ensemble Streamflow Forecasting
written by Nina Marie Caraway
has been approved for the Department of Civil Engineering

Balaji Rajagopalan

Edith Zagona

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Caraway, Nina Marie (M.S., Civil Engineering)

Stochastic Weather Generator Based Ensemble Streamflow Forecasting

Thesis directed by Prof. Balaji Rajagopalan

On a seasonal time scale, forecast centers of National Weather Service produce streamflow forecasts via a method called Ensemble Streamflow Prediction (ESP). In conjunction with the physically-based Sacramento Soil Moisture Accounting model (SAC-SMA), ESP uses historical weather sequences for the forecasting period starting from model's current initial conditions, to produce ensemble streamflow. There are two major drawbacks of this method — (i) the ensembles are limited to the length of historical record thereby producing limited variability and (ii) incorporating seasonal climate forecasts such as El Niño Southern Oscillation (ENSO) is done by selecting a subset of historical sequences which further reduces the variability of streamflow forecasts. The need for alleviating these drawbacks motivates the proposed research. To this end, this research effort has two components (i) an improved multi-site stochastic weather generator and (ii) coupling it to the SAC-SMA model for ensemble streamflow forecasting.

We enhanced the traditional K-nearest neighbor semi-parametric stochastic weather generator (SWG). In SWG the daily precipitation state (wet or dry) is modeled as a Markov Chain and the weather vector on a given day is simulated conditioned on the previous day's precipitation state and weather vector and current day's precipitation state. A K-nearest neighbor resampling approach is used to simulate from the conditional probability density function. Our improvements to this stochastic generator include (i) clustering the locations into climatologically homogeneous regions and applying the weather generator separately for each region and jointly to better capture the spatial heterogeneity and, (ii) modifying the resampling approach to incorporate probabilistic seasonal climate forecast. We tested this enhanced weather generator by applying it to daily weather sequences at 66 locations in the San Juan River Basin. The proposed method generates a rich variety of weather sequences capturing the distributional properties at all the locations and the

spatial dependence. It also simulates consistent weather sequences conditioned on seasonal climate forecasts.

The multi-site stochastic weather generator was coupled with the SAC-SMA model (WG-ESP) within NWS's new Community Hydrologic Prediction System (CHPS) to produce ensemble streamflow forecast. Spring season ensemble forecasts at several lead times from Nov through Apr for the period 1981-2010 were made from WG-ESP and the traditional ESP for the San Juan River Basin. We show that the weather generator based ensemble produces a rich variability in the flows including extremes and a higher skill at long lead times. Especially, skill in wet year forecast was found to be higher than dry years.

The flexible and robust framework provides many opportunities to further improve the ESP system in enabling increased skills at longer lead times that will be of immense help to water resources managers.

Dedication

To all my loved ones.

Acknowledgements

Firstly I would like to thank my advisor Balaji for his guidance on this project and for exposing me to the fascinating world of data analysis. I would like to thank my second member Andy for all of his invaluable help in sending data, explaining NWS methodologies, and helping the set-up and subsequent running of CHPS. Also, I would like to thank Edie for serving as the third member of my committee and for hosting me and providing computational resources for me at CADSWES.

A special thanks goes to Cameron for helping me develop my programming skills and for dealing with my now painfully-dumb sounding coding questions. Jason gets a thanks for converting the weather generator codes from FORTRAN to R. Thank you James for helping me set up an account on your group's NASA machine. Additionally, thanks for introducing me to a bevy of R techniques, for all of your help developing cluster methods for the weather generator, and for helping write the weather generator paper.

Thank you Lisa for being a great office mate and thanks to all in my research group for our incredibly-helpful meetings (and happy hours).

Finally, I would like to thank Kevin and others at the CBRFC for their help and NOAA for their funding of this project.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 Background

Balancing competing demands of various water users in the western United States has proven to be an ever-increasing challenge for water managers. Not only does the region have projected population and economic growth, but the 2000's drought,[1] climate variability, and climate change further complicate planning and management. McCabe and Wolock (2009) (and references therein), among others, have shown trends in warming since 1980 and general decrease in spring snowpack throughout the western U.S. Rajagopalan et al. (2009) highlight the risk of reservoir depletion given population growth and climate change, and suggest that flexibility be added to water management practices. Studies referenced therein also predict that average annual flow will decline as a result of climate change.

A majority of streamflow in the western U.S. originates from snowmelt. Whereas April 1st snow water equivalent (SWE) as a predictor provides skillful predictions of late spring/early summer runoff, many critical management decisions are made months beforehand, even in November or earlier. In the seven-state Colorado River Basin (CRB), the Colorado Basin River Forecast Center (CBRFC) is the primary official provider of streamflow forecasts to water managing agencies such as the U.S. Bureau of Reclamation and others. Currently the CBRFC and the Natural Resource Conservation Service (NRCS) work together to create water supply outlooks in the CRB. Forecasts are generated by each agency and subjectively combined into a joint, official-forecast (Hartmann

[1] `http://www.usbr.gov/uc/feature/drought.html`

et al., 2002). NRCS uses a principle components regression (PCR) technique that primarily relies on current snowpack and proxies of soil moisture such as antecedent streamflow and autumn precipitation (Garen, 1992). At the seasonal time scale, CBRFC implements two techniques: Statistical Water Supply (SWS), a regression-based method that relates observed data (e.g., snow, streamflow, precipitation) with future streamflow, and the model-based Ensemble Streamflow Prediction (ESP). In conjunction with a physically based watershed model, ESP relies historical meteorological data as possible representations of the future; each historical year is then used to simulate a streamflow trace (Day, 1985).

Ensemble forecasts have gained momentum in preference over deterministic forecasts, as probabilistic forecasts have been found to be more "appropriate and articulate" (Pagano and Garen, 2005) and offer more skill and relative economic value than deterministic forecasts (Roulin, 2007; Boucher et al., 2012). Ensemble forecasting is especially promising as the U.S. Bureau of Reclamation has developed the Mid-Term Operations Model (MTOM) (outlined in Grantz (2011) and described in detail in Bracken (2011)). MTOM is an objective, ensemble-based operations model where reservoir operations planning is engaged in a probabilistic mode. Having started running in experimental mode in 2010, it is a upgrade of their "24-month study", which helps anticipate monthly inflow volumes to the major Reclamation-operated reservoirs in the Colorado Basin.

The ESP methodology has shortcomings in that the ensembles created are hindered by limited historical data, which becomes even more limited with the addition of climatological forecasts (forecast based on region). Improvements over ESP have been developed for short-to-medium range (days-to-weeks) forecasting with the incorporation of ensemble weather forecasts (see Cloke and Pappenberger (2009) for a review). As these weather forecasts come from numerical weather models, their weather-scale reliability deteriorates after a medium-range, while their climate-scale signals may still reduce uncertainty in ESP.

To improve upon ESP limitations, we propose the incorporation of climate-scale probabilistic precipitation and temperature forecasts via the use of a hybrid nonparametric weather generator to create a variety of weather sequences that are more skillful and comprehensive than those found

in a historical climatology. The approach is a multi-site, multi-variable generator based on Apipattanavis et al. (2007) and furhter developed as described in Chapter 2. Precipitation occurrence is modeled with a two-state Markov chain and weather variables are selected using a k-nearest neighbor (k-NN) resampling algorithm. The resulting weather sequences are run through a watershed model using the ESP framework to produce streamflow forecast ensembles (Chapter 3). This weather generator can produce weather sequences that are unconditional or else conditioned on climate forecasts from an arbitrary source. Unlike ESP, there is no limit to the number of traces (i.e., size of ensemble) than can be generated to produce streamflow forecasts.

1.2 Study area and context

Our study was performed on the San Juan River Basin. With a drainage area of approximately 25,000 miles^2, the San Juan River is the second largest tributary of the Colorado River and runs a distance of 355 miles. The river flows from Colorado into New Mexico, then west to the Colorado River and Lake Powell in Utah. Having an area near in size to West Virginia, drainage areas are 39, 23, 20, and 17 percents in New Mexico, Colorado, Arizona, and Utah, respectively. The San Juan river basin includes a wide range of elevations, from roughly 4000 ft before confluence with the Colorado River, to above 14,000 ft in the San Juan Mountains. Climate zones range from desert plateau to mountain forests. Winter snow and rain due to frontal storms and modest rains from convective storms in summer are the main moisture input to the basin. Precipitation can vary from above 60 inches annually in the mountain peaks, to below 10 inches in the desert plateau[2] .

Completed in 1962, the Navajo Reservoir dramatically altered the natural hydrograph of the San Juan River. When filled, it occupies 15,610 acres, with a total capacity of 1.7 million acre-feet (MaF) and an active capacity of 1.0 MaF.[3] Major tributaries above the dam are the Navajo, Piedra, and Los Pinos Rivers. Of these only the Los Pinos is dammed (for agricultural purposes). The Animas River is the major tributary below the Navajo Dam and is free flowing. A majority

[2] http://www.usbr.gov/uc/wcao/rm/sjrip/
[3] http://www.usbr.gov/uc/wcao/water/ps/navajo.html

of the reservoir inflow occurs during the April-July runoff when an average of 0.66 MaF enter the reservoir. Before damn construction, the gage at Bluff saw approximately 72% of the total annual discharge occurring during that period (USBR, 2008).

Protecting natural ecology plays an important role in managing the San Juan. Tourism is especially important to the basin economy as many anglers visit each year for the famous abundance of rainbow trout. Fish such as the Colorado pikeminnow *(Ptychocheilus lucius)* and the razorback sucker *(Xyrauchen texanus)* declined to endangered levels in the years after the Navajo Reservoir was completed. Based on the recommended flows from joint research study of multiple agencies (Holden, 1999), spring releases from the Navajo Reservoir must be a minimum of 250 cfs and maximum controlled releases must be about 5000 cfs. Reservoir management must also take into account other needs such as storing water for consumptive use, irrigation, flood control, and generation of hydroelectric power. Finally, management of the San Juan River and the Navajo Reservoir must meet agreed-upon flows as defined in the Upper Colorado River Basin Compact and the Colorado River Compact.

1.3 Thesis outline

This thesis is written in manuscript form for the middle two chapters, which are self contained sections in a format that is acceptable for submission to an academic journal. After this introductory chapter, Chapter 2 describes the multisite stochastic weather generator. Then, Chapter 3 presents the linkage of the weather generator with a physical model to produce streamflow forecasts. Chapter 4 provides overall conclusions and discussion for future work.

Chapter 2

Multisite Stochastic Weather Generation Using Cluster Analysis and K-nearest neighbor Time Series Resampling

2.1 Introduction

Generation of synthetic weather sequences has been a topic of great interest in recent decades. Since historical data is limited, these sequences are needed to drive process models of hydrology, agriculture, erosion, ecology, construction delay, etc. (Wallis and Griffiths, 1997; Friend et al., 1997; Eberle et al., 2002; Mountain and Jones, 2006; Leander et al., 2006, 2007; Caron et al., 2008), to provide robust estimates of risks of decision variables, to enable better management of resources. The generation is based on stochastic models fit to historic data hence commonly referred to as stochastic weather generation. There is a rich literature on stochastic weather generators and the traditional generators trace their origin to the Weather Generator Model (WGEN) of Richardson (1981); Richardson and Wright (1984). In this, precipitation occurrence is modeled using Markov chains (Richardson, 1981; Katz, 1977; Stern and Coe, 1984; Woolhiser, 1992) or as Poisson process (Foufoula-Georgiou and Georgakakos, 1991; Furrer and Katz, 2008) and the amounts using probability density functions, such as two-parameter gamma (Katz, 1977; Buishand, 1978; Yang et al., 2005; Furrer and Katz, 2007). Bivariate autoregressive models of first order lag are fit to model maximum and minimum temperatures (Richardson, 1981). The models are fit for each month to capture the seasonality. This method is also referred as parametric weather generators, given the number of parameters of the various components. Generalized Linear Model (GLM) based weather generators offer an alternative parametric approach to modeling daily weather (Chandler

and Wheater, 2002; Chandler, 2005; Furrer and Katz, 2007; Yang et al., 2005). In this the precipitation and temperature are modeled as a series of GLMs with several covariates to capture seasonality, lagged dependence etc. The flexible nature of GLMs enable the modeling of binary (precipitation state) and continuous (precipitation intensity, maximum and minimum temperatures) variables with mixed covariates. We refer the reader to Wilks and Wilby (1999) for a comprehensive review of traditional parametric stochastic weather generators.

The above WGEN-based weather generators can be easily fit to daily weather at single locations, but many applications, such as hydrologic, require daily weather a multiple locations simultaneously. However, extending parametric models to multiple sites is not trivial. A major disadvantage is that model parameters grow exponentially with number of locations and spatial dependency also needs to be captured (Smith, 1994; Mehrotra et al., 2006). One such design (Wilks, 1998) involved a two-state, first-order Markov chain for precipitation occurrence and a mixed exponential distribution for precipitation generation. Serially independent but spatially correlated transformed normal variables enabled multisite generation; this was extended to additional weather variables in Wilks (1999). Many subsequent parametric multisite rainfall generators have been adaptations of the Wilks (1998) technique (Mehrotra et al., 2006; Brissette et al., 2007; Srikanthan and Pegram, 2009) with further additional variations for temperature simulation (Qian et al., 2002; Baigorria and Jones, 2010). Other multivariable methods involve disaggregating to individual locations from a regionally developed model like a statistical downscaling model (Segond et al., 2006; Mezghani and Hingray, 2009). Spatial models for rainfall occurrence and amounts using GLM for individual sites and Latent Gaussian process to spatially interpolate the GLM parameters and thus generate precipitation process in space, were developed by Kleiber et al. (2012). This approach has the ability to incorporate maximum and minimum temperature to result in a parsimonious spatial weather generator.

Nonparametric weather generators are an attractive alternative. Being data-driven, they can capture deviations from standard probability distributions, as well as nonlinearities between variables. Past methods include kernel density estimators (Rajagopalan et al., 1997; Harrold et al.,

2003; Mehrotra and Sharma, 2007) and k-nearest neighbor (K-NN) bootstrapping (Brandsma and Buishand, 1998; Rajagopalan and Lall, 1999; Buishand and Brandsma, 2001; Yates et al., 2003; Beersma and Buishand, 2003; Sharif and Burn, 2007). The K-NN approach is increasing in popularity due to its ease of implementation and effectiveness. In this, k-nearest neighbors are identified to the weather vector on a current day 't', from the historical data. Then one of these days is resampled using a weight metric that gives most weight to the nearest neighbor and least to the farthest. The uniformly distributed random number then resamples from these weighted days, simulating weather on day '$t + 1$'. This is akin to simulating from the conditional probability density function (PDF) $f(x_t \mid x_{t-1})$ with the PDF estimated locally in phase space. This method was first introduced by Lall and Sharma (1996) and adopted for weather generation by Rajagopalan and Lall (1999) and applied to different situations such as generating weather sequences conditioned on climate change projections, for hydrologic forecasting etc. This approach was modified by Apipattanavis et al. (2007) to a hybrid nonparametric model, where precipitation occurrence is modeled with a two-state Markov chain and weather variables are selected using a K-NN resampling algorithm. They also extended this to simulating weather sequences at multiple locations by applying the K-NN bootstrap weather generator on the daily average weather time series over these locations. In this domain-aggregated, 'da', approach, the weather at all the locations is simulated by resampling from the historical record at all locations on the same day, thereby maintaining spatial correlations. Apipattanavis et al. (2007) demonstrated this with application to four locations in a climatologically homogeneous region in northern Argentina.

Our motivation in this research comes from the need to generate daily weather sequences at a large number of locations sprinkled over a heterogeneous watershed, to subsequently drive a hydrologic modeling system to produce streamflow forecasts. The methodology of Apipattanavis et al. (2007), based on resampling from a spatially averaged daily weather time series, may not adequately capture the spatial nonhomogeneity. We propose a new adaptation to this approach where the sites are (i) first clustered into homogeneous sub-regions based on historical seasonal precipitation (or any other suite of attributes); (ii) a Markov chain is fit to cluster-averaged pre-

cipitation time series over the joint two-state, three-cluster ($2^3 = 8$ state) system to capture the spatial correlation in the precipitation occurrence between the clusters; (iii) the K-NN bootstrap is then applied to generate daily weather sequences conditioned on the precipitation state, for each cluster. Thus, generating daily weather sequences at all the desired locations.

This weather generator has an additional feature where it can serve as a downscaling link between probabilistic climate forecasts and hydrologic modeling. We investigate the proposed methodology in the context of large-scale seasonal precipitation forecasts. With this option, the conditioned generated weather sequences reflect prediction of wet or dry climate, which will then result in wet or dry streamflow forecasts.

The paper is organized as follows. The application region and context are described along with the data sets. The methodology is then described with the implementation. After presenting the results, we conclude with discussion of the methodology and potential applications and improvements.

2.2 Study Region, Application Context and Data

With a drainage area of approximately 25,000 sq. miles, the San Juan River is the second largest tributary of the Colorado River. Its area, nearly that of West Virginia, is approximately split between New Mexico (39%), Colorado (23%), Arizona (20%), and Utah (17%). The San Juan river basin includes a wide range of elevations, from roughly 4000-14,000 ft above sea level, and climate zones, including desert and forest. DJF snow and rain due to frontal storms and modest rains from convective storms in summer are the main moisture input to the basin. Precipitation can vary from above 60 inches annually in the mountain peaks, to below 10 inches in the desert plateau[1] . Prior to construction of the Navajo Dam, flows were snowmelt dominated, but with reservoir operations, MAM runoff is stored and released during summer and later months (Holden, 1999).

[1] http://www.usbr.gov/uc/wcao/rm/sjrip/

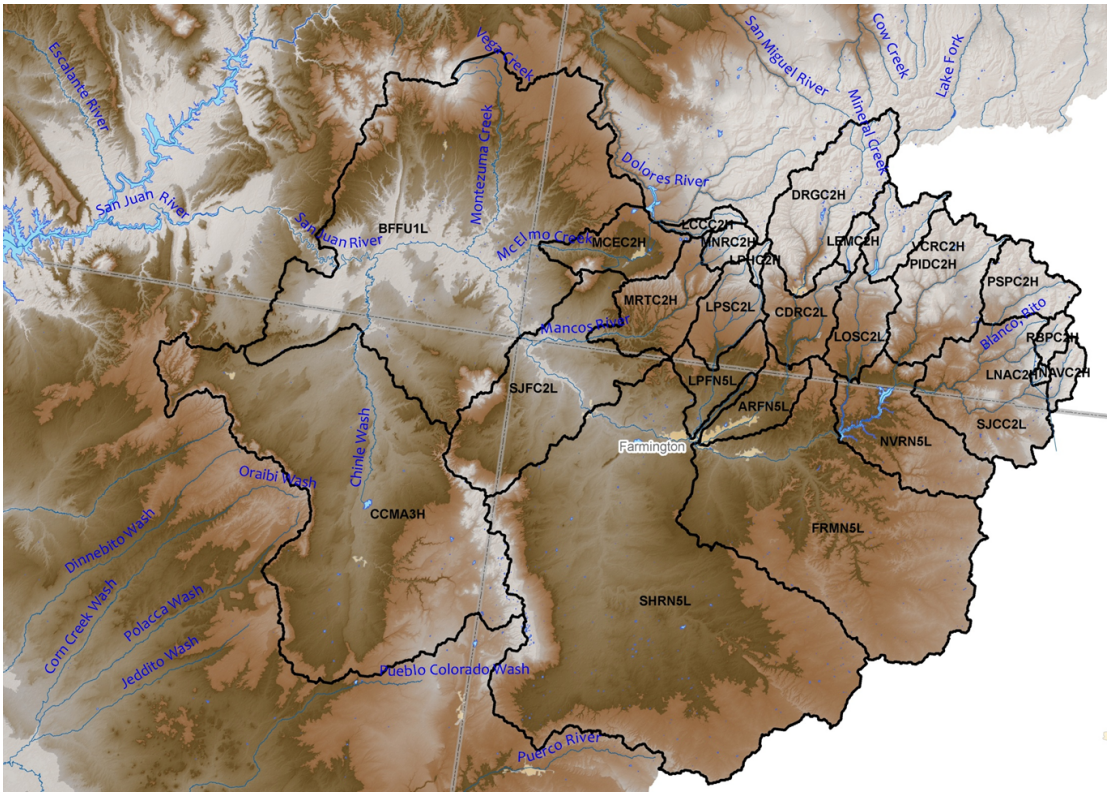


Figure 2.1: San Juan Watershed: the four corners of Utah, Colorado, New Mexico, and Arizona is just above the center of the figure

The Colorado Basin River Forecasting Center (CBRFC) provides ensemble seasonal streamflow forecasts at several lead times at multiple locations on the San Juan River. They use the historic daily weather sequences to drive a physically based watershed model (Sacramento Soil Moisture model, SAC-SMA). This generates as many ensemble members for the current forecast season as the number of times the season appears in the historical record (Day, 1985). The historic weather sequence thus provides a very limited set of ensembles. Ensembles are further reduced when a seasonal forecast condition (e.g. "wet") is imposed. Our proposed stochastic weather generator aims to solve this problem by simulating a rich variety of weather sequences over the study domain.

The CBRFC divides the San Juan river basin in to 24 sub-basins (Figure 3.4) which are further divided by elevation bands into two to three zones each, resulting in a total of 66 spatial

zones, or subcatchments. Based on observed historical data at weather stations scattered across the basin, the CBRFC has created mean areal precipitation (MAP) and mean areal temperature (MAT) for each zone (NWSRFS, NWSRFS) at 6-hourly time steps. These start in October of 1980 through September of 2010. From the 29 complete years, we calculate daily precipitation totals as well as daily minimum and maximum temperatures which serve as inputs to our stochastic weather generator. We generate 100 ensemble members of daily weather sequences for all 66 zones over the 29 year period. These weather ensembles will be applied in future studies to drive a watershed model and to robustly estimate streamflow probabilities at various points in the basin as well as management risks. In this study, we evaluate our weather generator approach against the observed, seasonal statistics of daily MAP and MAT timeseries.

2.3 Methodology

2.3.1 Cluster Analysis

To parse spatial inhomogeneity of weather over our domain, we employ K-means cluster analysis (Everitt, 1979)), with clustering on seasonal precipitation totals. The objective is to classify $M = 66$ points in $N = 29$ years dimensions into k clusters such that the within sum of squares (WSS) over all clusters is minimized without over-fitting the clustering (i.e., using too high a k).

The Hartigan-Wong approach algorithm is employed (Hartigan and Wong, 1979). Because initial centroids are selected at random, clustering is repeated a 50 times to investigate uncertainty in WSS for each choice of k. Our 'kink plot' (Figure 2.2) shows WWS and its 50% (boxes) and 90% (whiskers) uncertainty over 50 clustering trials as a function of k, the number of clusters. The kink, or decrease in the reduction of WSS with increasing number of clusters, indicates the optimal number of clusters (e.g. Hastie et al. (2009)).

2.3.2 Spatial Precipitation Occurrence Model

The domain-aggregate ('da') weather generation approach of Apipattanavis et al. (2007) models temporal precipitation occurrence via a 2-state Markov chain with states "wet" and "dry." The region is wet (dry) if its domain average precipitation is above (below) 0.1 inches. The 2^2 elements $(p_{dd}, p_{dw}, p_{wd}, p_{ww})$ of the transition probability matrix (TPM) are estimated using maximum likelihood. In the domain aggregate approach, Markov transition probabilities are calculated on domain averaged weather at each time step. We compute the Markov transition probabilities for each month separately within each season.

Described in the following section, the Markov modeled state conditions selection of a historical weather observation under K-NN resampling. The problem with this method is that daily precipitation and temperature over multiple locations become more heterogeneous as the domain size grows. The use of domain average precipitation becomes inappropriate for estimating the domain-wide precipitation state and thus for both calculation of state transition probabilities and for selecting nearest neighbors from the historical record. Domain-averaged weather is also used in the K-NN resampling scheme described below. By clustering simulations by total seasonal precipitation within the domain, we aim to resolve the large-scale, spatial heterogeneity problem of the 'da' approach.

After identifying k clusters based on seasonal precipitation totals within the spatial domain, the simplest approach is to directly model the clusters individually and independently using the 'da' approach in each cluster separately. We term this approach 'ca' for cluster-aggregate. Because our application is hydrologic response, with future consideration of the basin outlet, 'ca' has the obvious shortcoming that the upstream weather inputs are not coordinated.

We explore two solutions to this problem. First, in order to coordinate (or correlate) basin-wide response, we propose generating the precipitation state over the full domain as in the 'da' approach and then using this global state within each cluster to condition selection of observed weather from the within-cluster historical record. Deemed 'caShared', this approach may suffer

from the obvious deficiency that gross heterogeneities over the domain at the cluster level will not be appropriately modeled. For example, if it tends to rain in one cluster while others are dry, then this situation will be undersimulated.

Second, we take a more nuanced approach to the problem of spatial coordination of cluster states by modeling Markov transition probabilities between all possible states of the three cluster system. If N_s is number of precipitation states and k the number of clusters, then say for $k = 3$ clusters there are 8 possible states.

$$N_s^k = 8 = ddd, ddw, dwd, dww, wdd, wdw, wwd, www \tag{2.1}$$

We term this approach 'caJoint'. The probability of transitioning to any joint state can be computed using cluster-average precipitation and the established threshold of 0.1 inches to determine "wet". Calculating the joint transition probability begins with cluster-average precipitation and the previous, two-state threshold of 0.1 inches to distinguish wet from dry. We compute the Markov transition probabilities for each month separately based on the historical transitions between the resulting 8 states of the system. For our simple, k-cluster system, the transition probability matrix is 8 x 8 when $k = 3$ (though all transitions may not actually occur in the data). Note that even for a two state, four cluster system the number of distinct transition probabilities is already 16 x 16. One may employ any number of states and clusters, but the complexity can rapidly increase when modeling the joint states of the system as proposed here.

2.3.3 K-NN Resampling

For all of the above approaches ('da', 'ca', 'caShared', and 'caJoint'), daily weather sequences at all locations are generated using the algorithm of Apipattanavis et al. (2007). The basic idea is, for a given day t, to sample areal-averaged weather vectors from the conditional PDF, $f(x_t \mid DOY, x_{t-1}, S_{t-1}, S_t)$, of areal-averaged weather vectors, x_t, given the day of year, DOY, yesterday's simulated weather, x_{t-1}, yesterday's precipitation state, S_{t-1}, and today's precipitation state, S_t. Once an areal-averaged weather vector is selected, the actual observed weather on that same day

at each location in the region is used in the simulation. The cluster-based approaches have three regions, and so three potentially different days will be selected from the record, corresponding to each region, to simulate the locations in each.

Given an initial area-average weather vector, the historical areal-average weather vectors, and the previously generated sequence of areal-averaged precipitation states for some area, the following algorithm applies to all approaches ('da' and cluster based):

(1) A 7-day window centered on the day of year for time t in the simulated sequence filters out the remainder of the historical record ($\mid DOY$).

(2) Days within this window with matching, areal-averaged state transitions to simulated day $t-1$ are extracted ($\mid S_{t-1}, S_t$).

(3) Each areal-averaged weather variable within the selected days is scaled by the reciprocal of its historical standard deviation on day $t-1$ to provide equal weight over all variables in the next step.

(4) Euclidean distances are calculated between all candidate weather vectors over the region and the previously simulated weather vector on day $t-1$.

(5) The most similar (closest) weather vectors are limited to a neighborhood of size $K = \sqrt{N}$, where N is the length of the dataset (Lall and Sharma, 1996).

(6) These K nearest neighbors are assigned a probability using a discrete decreasing kernel

$$p(i) = \frac{1/j}{\sum_{j=1}^{k} 1/j} \tag{2.2}$$

(Lall and Sharma, 1996; Rajagopalan and Lall, 1999), where $p(i)$ is the probability that the ith neighbor will be selected, thus giving the kth neighbor the lowest probability.

(7) A neighbor (i.e., a historical day) is randomly resampled according to these weights.

(8) Its *successive* day in the historical record is selected as the weather on day t and represents all locations in the modeled area.

This algorithm is repeated for each day in the desired simulation period. Table 2.1 demonstrates the selection of nearest neighbors for a window centered on December 9th.

2.3.3.1 Conditional Resampling

Synthetic weather sequences are often required that are consistent with large-scale, seasonal climate forecasts. We incorporate the seasonal precipitation forecasts issued by the International Research Institute for Climate and Society[2] into our methodology. These give the probability of each seasonal precipitation tercile, above:normal:below (ANB). The approaches of Apipattanavis et al. (2007) and Yates et al. (2003) to conditional resampling based on climate forecasts is to modify the weights in step 6 of the K-NN resampling algorithm. Because we found this approach to only mildly modify the outcome, we propose a new strategy here.

(1) Calculate seasonal precipitation totals in each year and associate it with each day in a given season.

(2) Follow the K-NN algorithm through step 2: find all historical days in the 7-day moving window which match the simulated transitions, which we will call T.

(3) Now the neighborhood of size $K = \sqrt{N}$ is no longer determined by euclidean distances (replacing steps 3 through 5).

(4) Then for a wet (dry) ANB

   (a) Sort the matching days, T, based on decreasing (increasing) seasonal totals.

   (b) The nearest (farthest) neighbors are $A \times K$ of head (tail) of T.

   (c) The farthest (nearest) neighbors are $B \times K$ of tail (head) of T.

   (d) What remains is filled by $N \times K/2$ on either side of the median seasonal total of T.

(5) Resume step 6 from the algorithm above.

[2] `http://iri.columbia.edu/climate/forecast/net_asmt/`

This method ensures preferential treatment of seasons that have desirable characteristics. Since daily precipitation intensities are not necessarily sorted in order of highest or lowest, resampling is not too severely modified and a variety of weather scenarios is still maintained. Table 2.2 demonstrates how the K neighbors are determined for an ANB of 65:25:10. With $K = 14$, $0.65 \times K = 9$, thus the neighborhood from 1 to 9 is filled from the 9 highest totals according to the sorted T. As $0.1 \times K = 1$, the last member of the neighborhood is filled by the day corresponding to the lowest value in sorted T. The remaining values are filled by days falling on either side of the median. In the unconditional simulation example (Table 2.1), the weather on December 12, 2007 was determined as the 8th nearest neighbor. For the wet simulation (Table 2.2), it became the nearest neighbor because of its corresponding seasonal totals, and thus is more likely to be sampled.

Table 2.1: K-nearest neighbors for ordinal day of December 9th: unconditional simulation

	year	month	day	p	tmin	tmax	seq	state	trans	total
1	2000	12	9	0.12	-3.63	2.36	7278	w	w2d	6.58
2	1996	12	11	0.39	-1.20	3.58	5820	w	w2d	8.80
3	1991	12	11	0.71	-7.10	1.05	3995	w	w2d	5.37
4	1997	12	8	0.38	-6.76	-0.85	6182	w	w2d	5.02
5	1993	12	12	0.44	-6.49	0.02	4726	w	w2d	4.91
6	1996	12	7	0.14	-9.89	3.46	5816	w	w2d	8.80
7	1998	12	6	0.15	-11.62	-4.44	6545	w	w2d	3.08
8	2007	12	11	0.37	-7.12	-1.18	9835	w	w2d	14.77
9	2006	12	11	0.24	-8.11	-0.17	9470	w	w2d	5.19
10	2003	12	8	0.30	-5.75	0.91	8372	w	w2d	6.71
11	1982	12	9	0.51	-4.41	2.84	708	w	w2d	6.33
12	1982	12	11	0.11	-3.94	1.00	710	w	w2d	6.33
13	1985	12	10	0.21	-14.11	-4.86	1804	w	w2d	4.29
14	1984	12	8	0.43	-7.64	3.11	1437	w	w2d	7.76

Table 2.2: Same as Table 2.1, but for wet simulation with ANB of 65:25:10

	year	month	day	p	tmin	tmax	seq	state	trans	total
1	2007	12	11	0.37	-7.12	-1.18	9835	w	w2d	14.77
2	1996	12	7	0.14	-9.89	3.46	5816	w	w2d	8.80
3	1996	12	11	0.39	-1.20	3.58	5820	w	w2d	8.80
4	2008	12	8	0.26	-3.05	1.79	10197	w	w2d	8.68
5	1984	12	8	0.43	-7.64	3.11	1437	w	w2d	7.76
6	1994	12	6	0.88	-2.51	2.06	5085	w	w2d	7.00
7	2003	12	8	0.30	-5.75	0.91	8372	w	w2d	6.71
8	1986	12	6	0.70	-3.75	3.06	2165	w	w2d	6.68
9	2000	12	9	0.12	-3.63	2.36	7278	w	w2d	6.58
10	1982	12	9	0.51	-4.41	2.84	708	w	w2d	6.33
11	1982	12	11	0.11	-3.94	1.00	710	w	w2d	6.33
12	2009	12	8	0.83	-13.20	-3.25	10562	w	w2d	6.28
13	1991	12	11	0.71	-7.10	1.05	3995	w	w2d	5.37
14	2001	12	11	0.11	-10.68	-1.42	7645	w	w2d	1.91

2.4 Results

Cluster analysis was performed on seasonal precipitation totals, separately for each three month season, Dec–Feb (DJF), Mar–May (MAM), Jun–Aug (JJA) and Sep–Nov (SON). Based on the kink plot for DJF (Figure 2.2) three clusters were chosen to be the optimal number, similarly for other seasons. Boxplots of elevations of the locations falling in each cluster are shown in Figures 2.3 and 2.4 — it can be seen that the clusters fall quite well along the elevations. In that, all the high elevation locations are together in cluster 'a', the middle elevations in cluster 'b' and the lower in 'c'. This is consistent in that the DJF precipitation has a distinct elevation signal, with higher elevation regions receiving more snow and vice-versa. This is found to be the case in all the seasons

except for summer where the precipitation is less organized by elevation, as would be the case with convective nature of precipitation in this season. The number of locations in each cluster in each season is shown in Table 2.3. The spatial state simulation and K-NN resampling are applied to locations in each cluster to generate daily weather sequences.

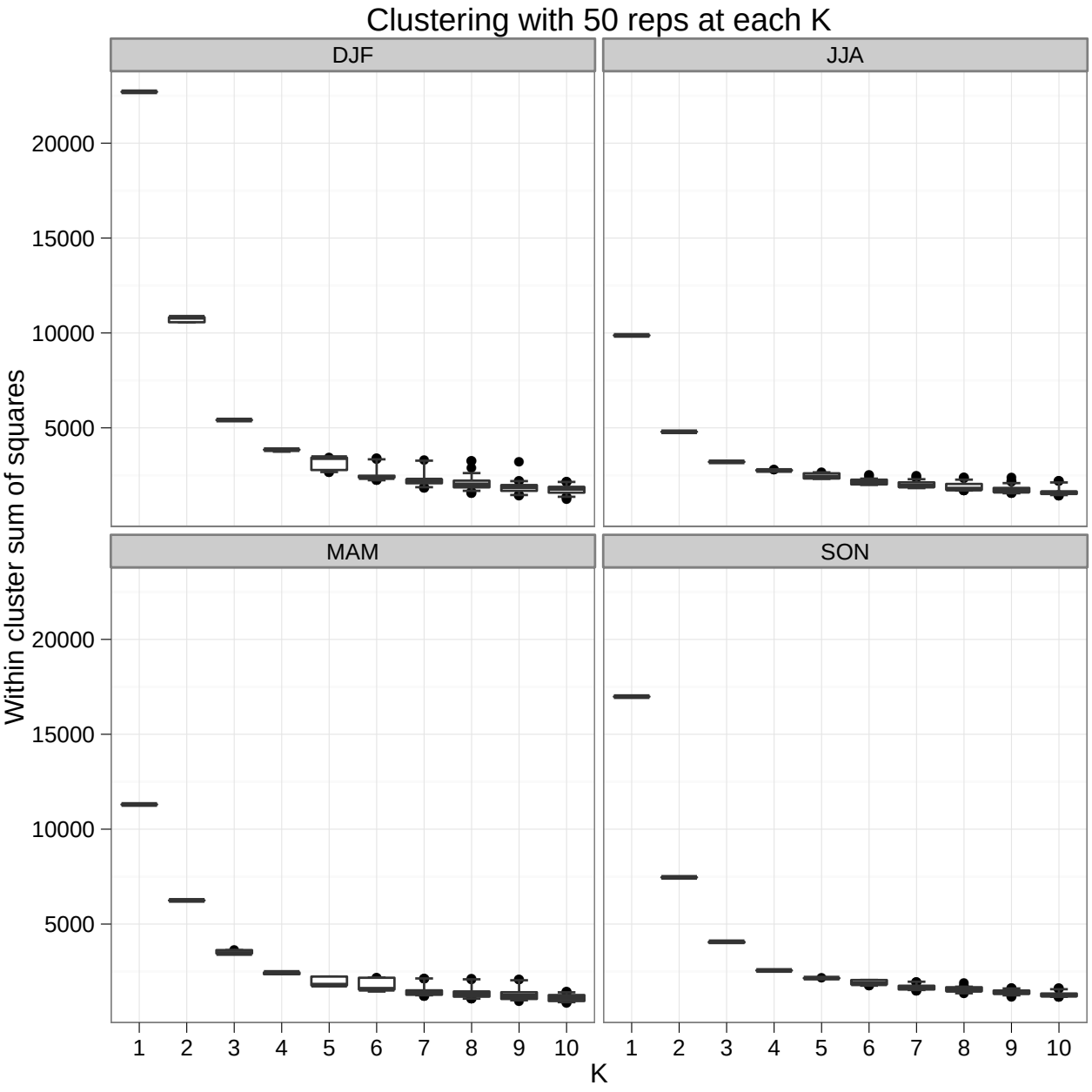


Figure 2.2: Uncertainty in the within sum of squares (WSS) as a function of **k**th for 50 independent clusterings of total seasonal precipitation over 66 stations in 29 years.

2.4.1 Clusters

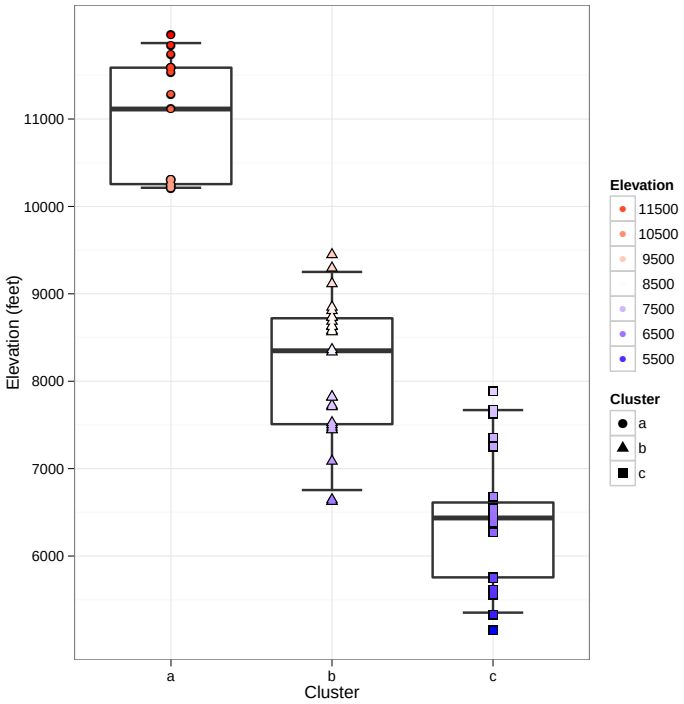
Classification of ‘a’, ‘b’, or ‘c’ correspond to cluster groupings with highest, middle, and lowest median elevations, respectively in each season. Assignment of subcatchments within groups will change with each analysis due to the heuristic nature of the algorithm. However, as very little change in WSS is seen in Figure 2.2 for $k = 3$, the results are assumed as stable.

Table 2.3: Number of stations falling within each cluster

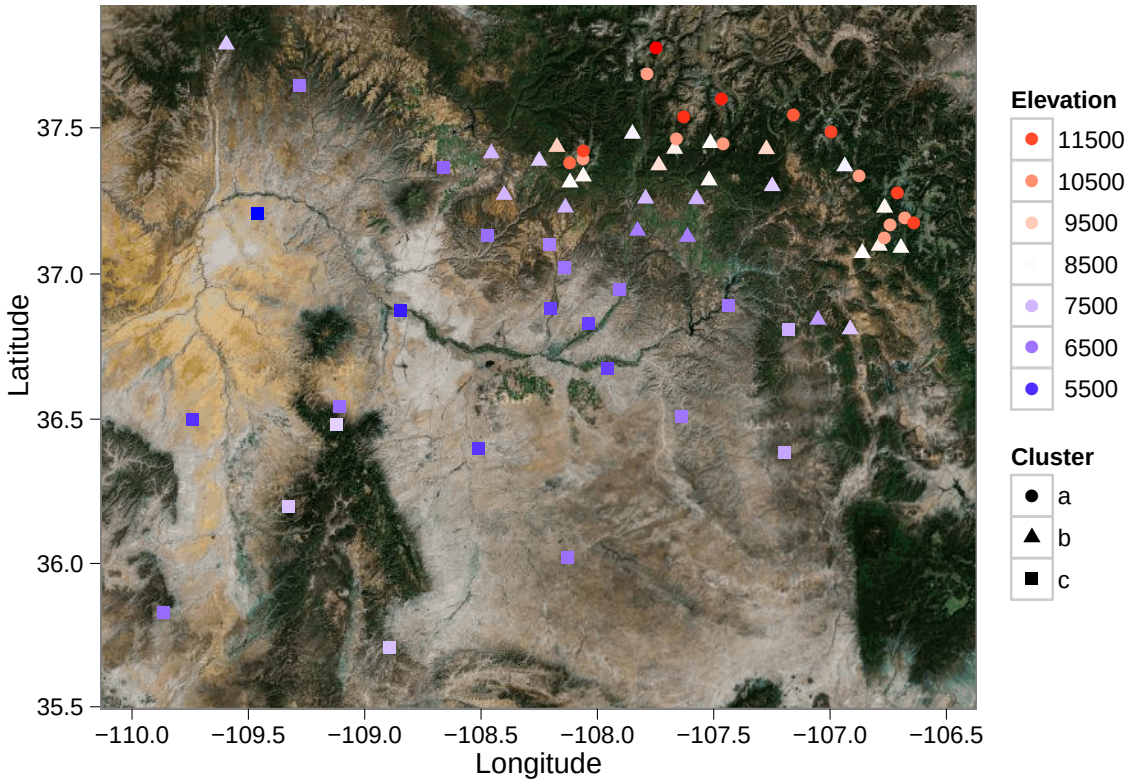
	a	b	c
DJF	17	26	23
MAM	19	25	22
JJA	22	22	22
SON	17	26	23

Boxplots in Figures 2.3a and 2.4a depict elevations of subcatchment centroids organized by cluster groups. Though clustering was performed on precipitation, this shows there is a heavy correlation with elevation as well. DJF is more organized by elevation while summer has more influence from spatial proximity. Figures 2.3b and 2.4b show cluster ‘a’ predominantly falls within forested, mountainous areas while ‘c’ falls in low-lying arid regions. Cluster ‘b’ has a mix of characteristics from the other two groups. Cluster medians change appreciable with season. Cluster ‘c’ is the most stable of the groupings—its member population remains relatively consistent between the seasons and its median elevation stays near 6500 ft. Conversely ‘a’ and ‘b’ have more variable member numbers and the median elevations can change up to 1000 ft between seasons.

Figure 2.5 displays between-cluster correlations in the observed record and illuminates the motivation behind the ‘caShared’ and ‘caJoint’ methodologies. From these figures it is evident that each cluster cannot be assumed as independent from others, thus they should not be simulated independently, as in the ‘ca’ approach shown below.

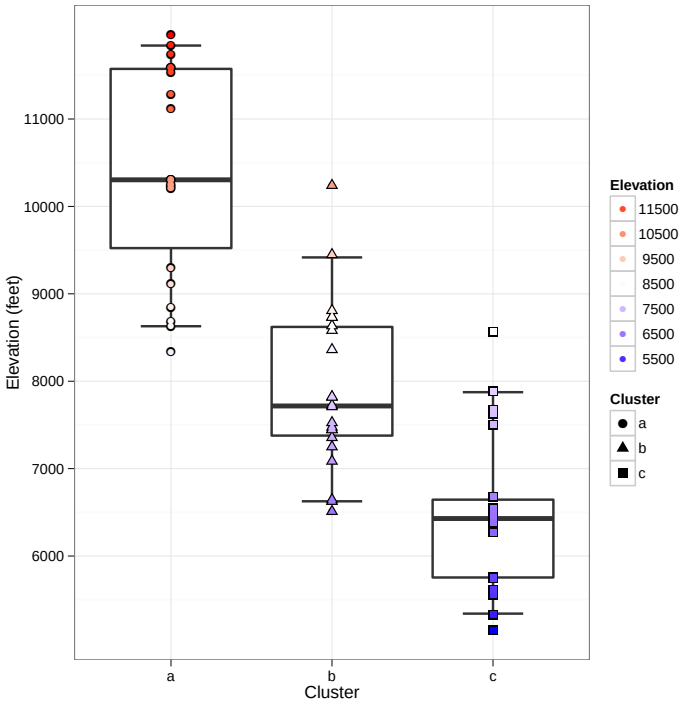


(a)

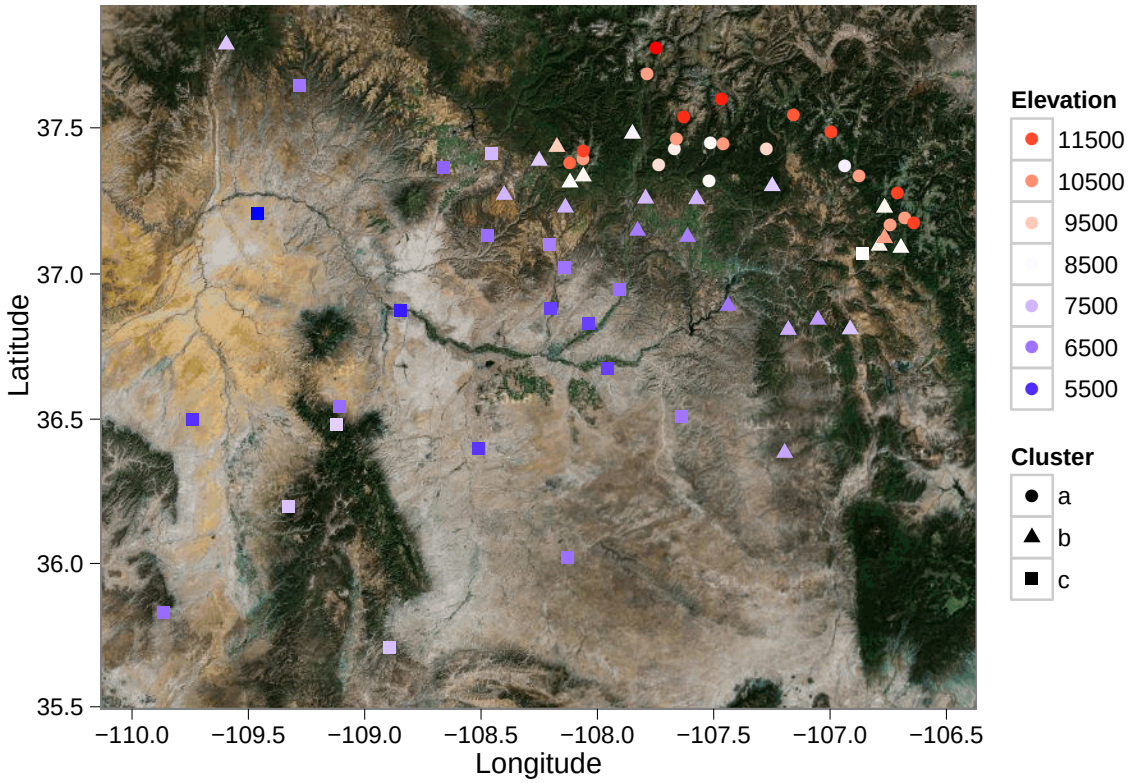


(b)

Figure 2.3: DJF geographical clustering of observation locations by seasonal total precipitation

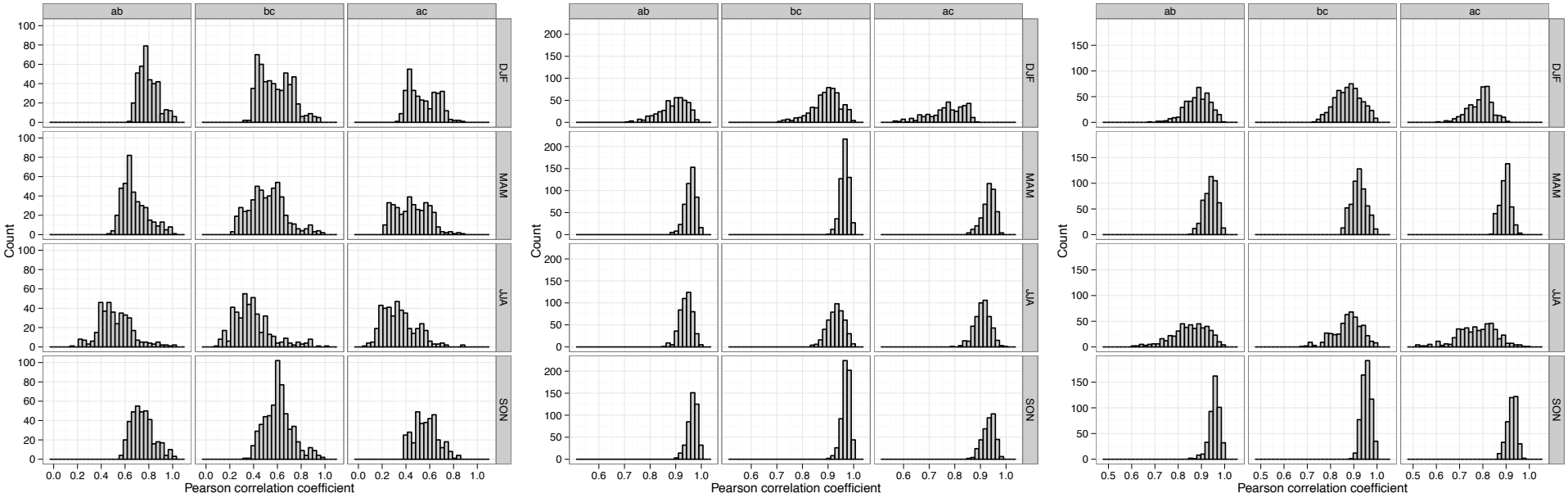


(a)



(b)

Figure 2.4: Same as Figure 2.3 for JJA



(a) Precipitation (b) Max temperature (c) Min temperature

Figure 2.5: Histograms of between-cluster correlations in the observed data for all stations in each pair of clusters

2.4.2 Unconditional Simulation

We first evaluate characteristics of the four methodologies against the observed statistics of the seasonal precipitation. Figure 2.6 compares probability density functions (PDF) of basin total seasonal precipitation for the four methods along with the historical PDF. With a lack of inter-cluster dependency in the 'ca' method, its PDF tends towards the mean with a thin tail behavior in all seasons. In DJF and MAM, the other three methods are similar to each other and closer to the historical distribution. All four seasons reinforce the conclusion drawn from Figure 2.5 that independence cannot be assumed between clusters.

Figure 2.7 shows simulated versus observed average number of wet day occurrences per season. It is evident that forcing the same precipitation state across the clusters in the 'caShared' method oversimulates the number of wet days in the lower elevation clusters and undersimulates in the higher elevation for all seasons. Thus 'caShared' does not appear like a viable option. The other three methods are near-similar in their performance and have R^2 values of 0.99 with the observed.

Figure 2.8 divides Figure 2.6 into within-cluster total seasonal precipitation, showing the PDFs of seasonal precipitation for each cluster. The cluster division, 'ca' now has a strong performance as it simulates each cluster independently thus reproducing individual cluster distributions, whereas it performed poorly on the basin-wide precipitation (Figure 2.6). The 'caShared' method seems to oversimulate in the lower elevation clusters (b and c) and undersimulates in the high elevation cluster (a), similar to the behavior in Figure 2.7. With the exception of summer, 'da' and 'caJoint' exhibit good performances in terms of capturing the observed PDF.

Figure 2.9 shows the PDF of correlation between simulated and observed seasonal precipitation and temperature between the clusters. The historical PDF aligns almost exactly with 'da' because this method produces simultaneous simulation of all locations, hence their correlations. As emphasized above, 'ca' behaves poorly because each cluster is simulated independently of the other, so the between-cluster correlations are not captured. Here, 'caShared' does well because between cluster correlation is enhanced by having a shared state, but we have established the side effects

of this assumption in within-cluster simulation. While 'caJoint' is performing better than 'ca', the between-cluster correlations are still poorly represented.

Figure 2.10 shows the root mean square error (RMSE) of seasonal precipitation over all simulations for 'da' and 'caJoint' for the three variables. Except for SON, 'caJoint' shows smaller RMSE and less variability (tighter boxes) compared to 'da', suggesting that it better captures the observed variability. Conversely, 'da' has lower errors for temperature and lower median RMSE, again excepting SON. This is not unexpected as temperature was not included in the clustering process.

Lag-1 daily autocorrelation of the observed and simulated at all locations for all the methods and variables are shown in Figure 2.11. A best fit line for each method is shown along a 1:1 line. Lag-1 autocorrelations are notoriously difficult to simulate with nonparametric weather generators (Buishand and Brandsma, 2001; Yates et al., 2003; Apipattanavis et al., 2007). All methods, as evident from their lines of regression, suffer from notable under-simulation.

We now compare simulated and historical statistics at three sites (i.e., subcatchments) in the domain for purposes of assessing single-site behavior. The representative sites are: drgc2hu (upper zone near Durango, CO and Animas River), pidc2hl (lower zone near Piedra River and Arboles, CO), and bffu1ll (lower zone near San Juan River and Bluff, UT)[3] . Drgc2hu has the highest elevation in the basin, pidc2hl the median, and bffu1ll the lowest. Each falls within the same cluster (a, b, and c, respectively) for all four seasons.

Differences between simulated and observed values for basic distributional statistics are shown in Figures 2.12, 2.13, and 2.14 for the three variables. The vertical black lines mark zero around which the simulations should ideally be centered. Both methods capture the basic statistics of precipitation (Figure 2.12) quite well, but there is a tendency to slightly undersimulate the mean and standard deviation, especially for 'da'. This reduction in the simulated variance, known as overdispersion, is a common phenomenon in weather generation unless additional measures are

[3] NWSRFS handler id: first five characters are the sub-basin name, sixth character describes whether or not it is a headwater basin and seventh denotes elevation zone

taken to provoke variability (Kim et al., 2012). Unusually, ‘caJoint‘ over-simulates mean and standard deviation for fall, while ‘da’ over-simulates for bffu1ll DJF. In general, ‘da’ performs better in capturing the temperature distributions (Figures 2.13 and 2.14).

Figures 2.15 and 2.16 demonstrate the cumulative distribution function (CDF) of temperature spells below and above selected thresholds, for the three locations and for winter and summer, respectively. Both methods perform similarly but ‘caJoint’ appears to be slightly closer to the observed CDF. The first order Markov chain has noticeable effect on the probabilities — smaller spell lengths are simulated with much higher probability than what happened historically. A higher order Markov chain might help in improving this.

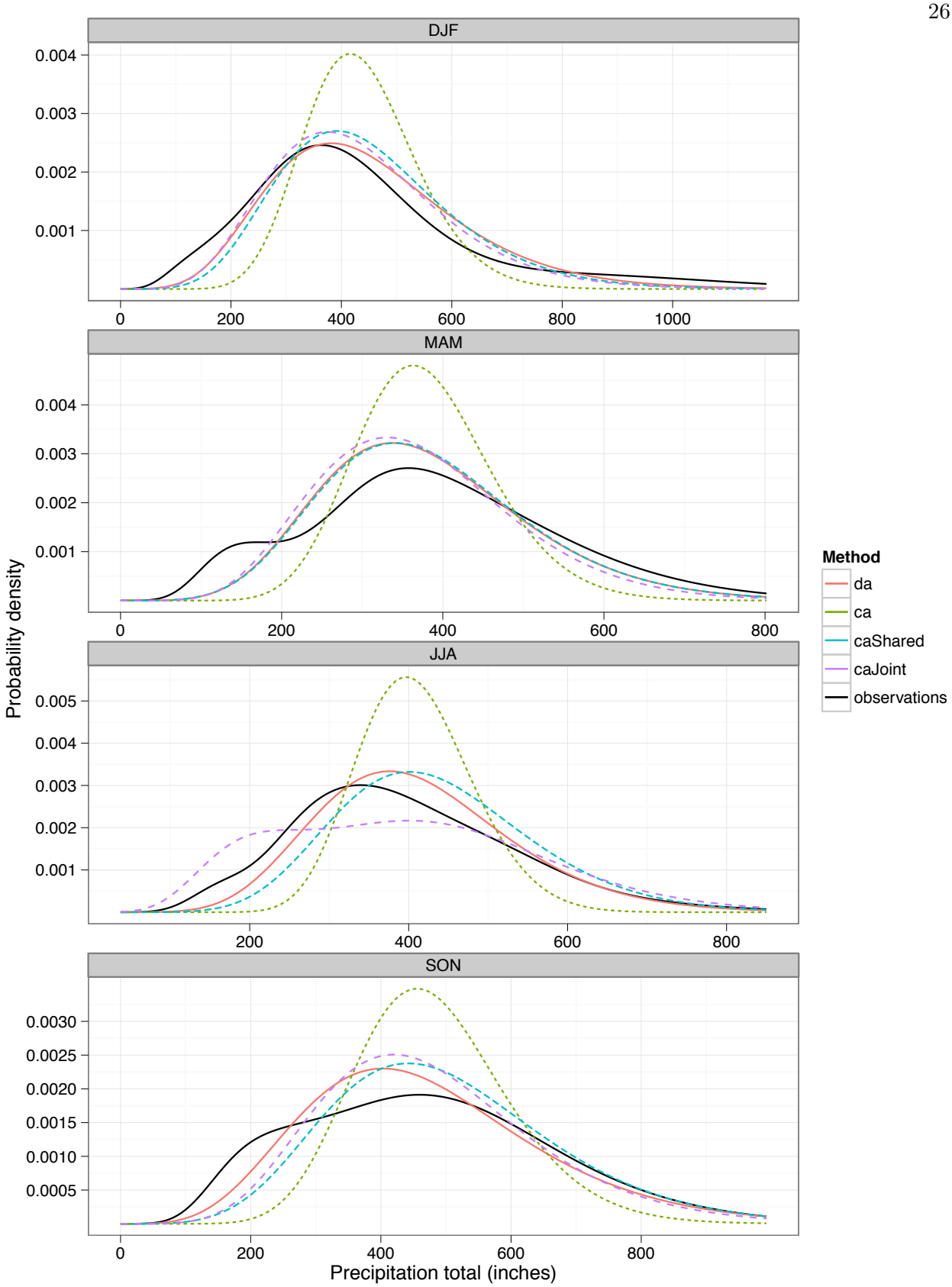
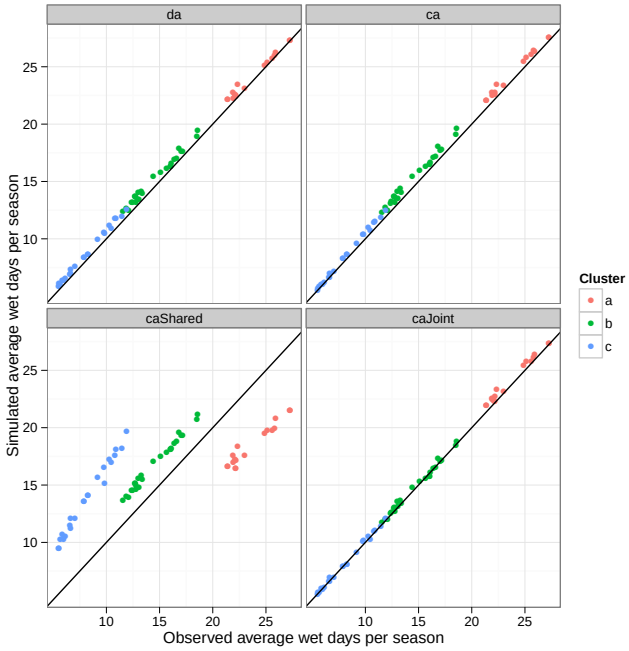
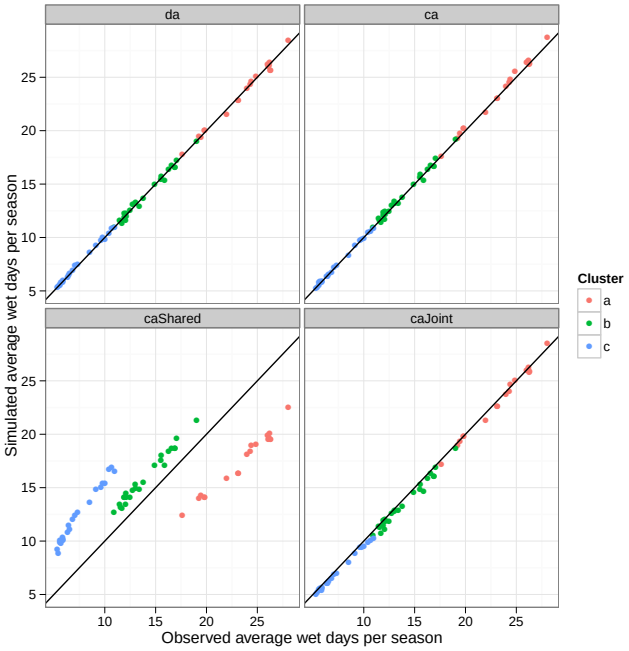


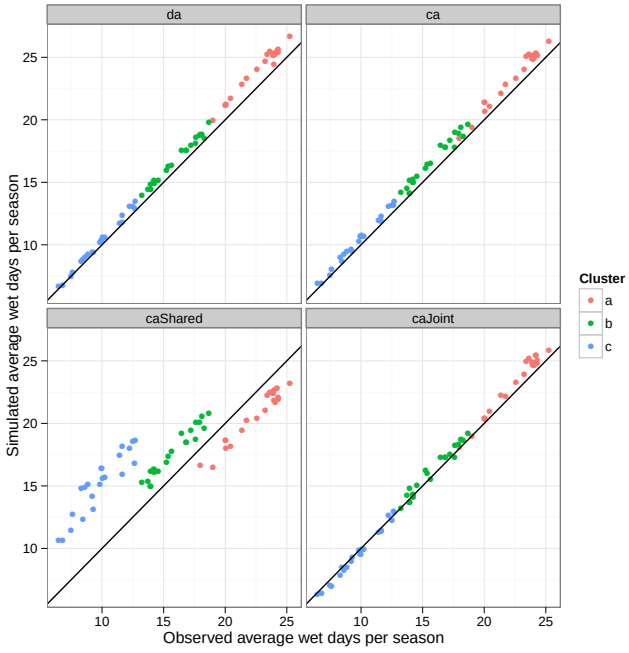
Figure 2.6: Smoothed probability densities of basin-total seasonal precipitation for all methods



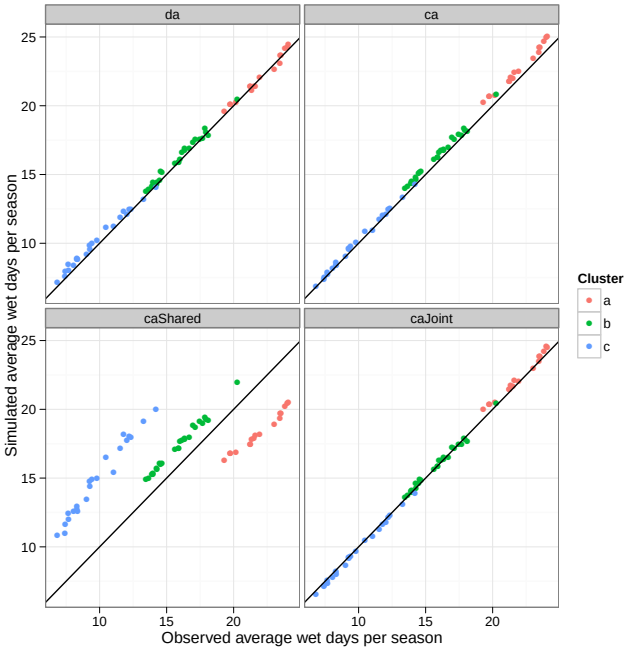
(a) DJF



(b) MAM



(c) JJA



(d) SON

Figure 2.7: Average number of wet day occurrences per season plotted against observed for each approach

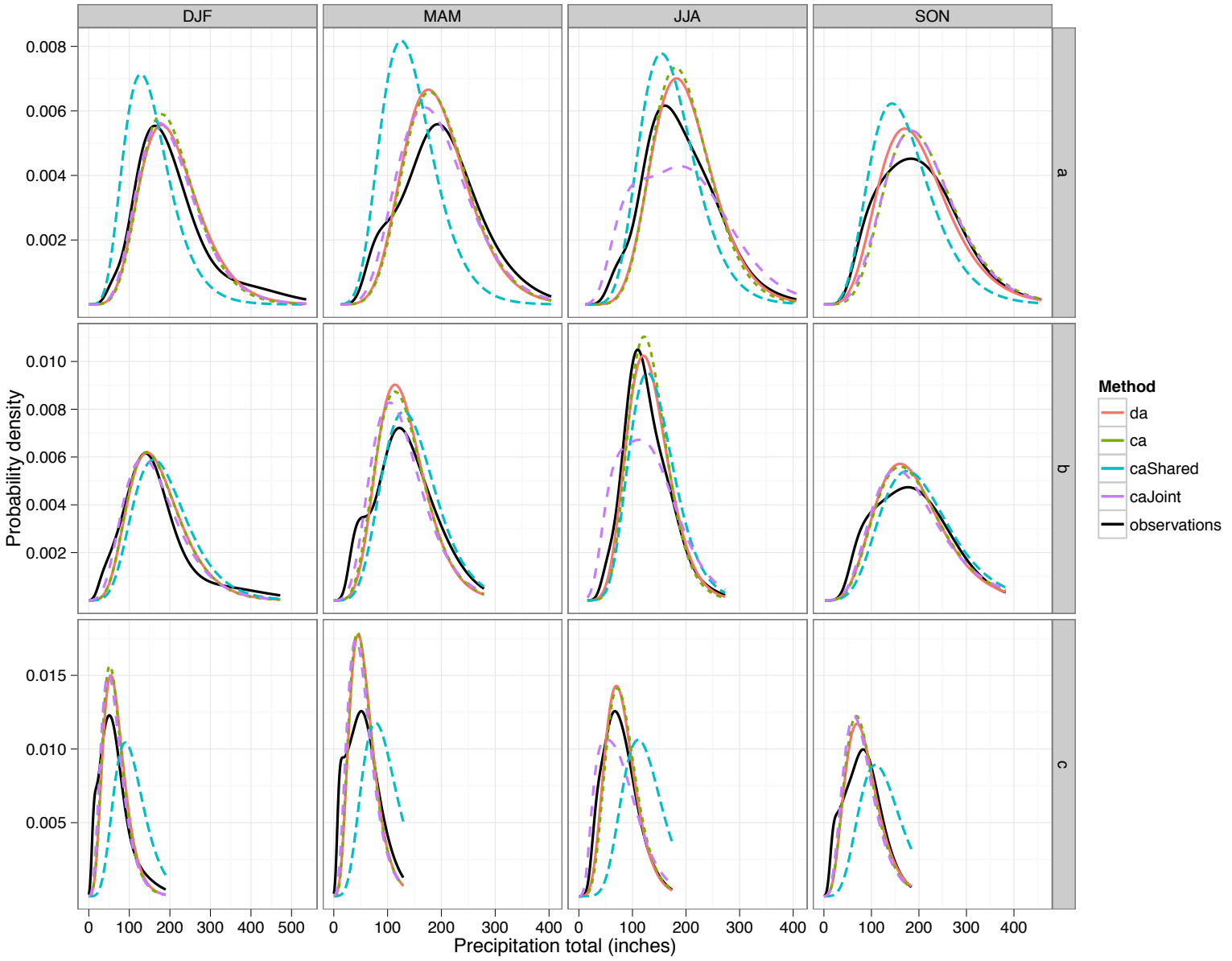
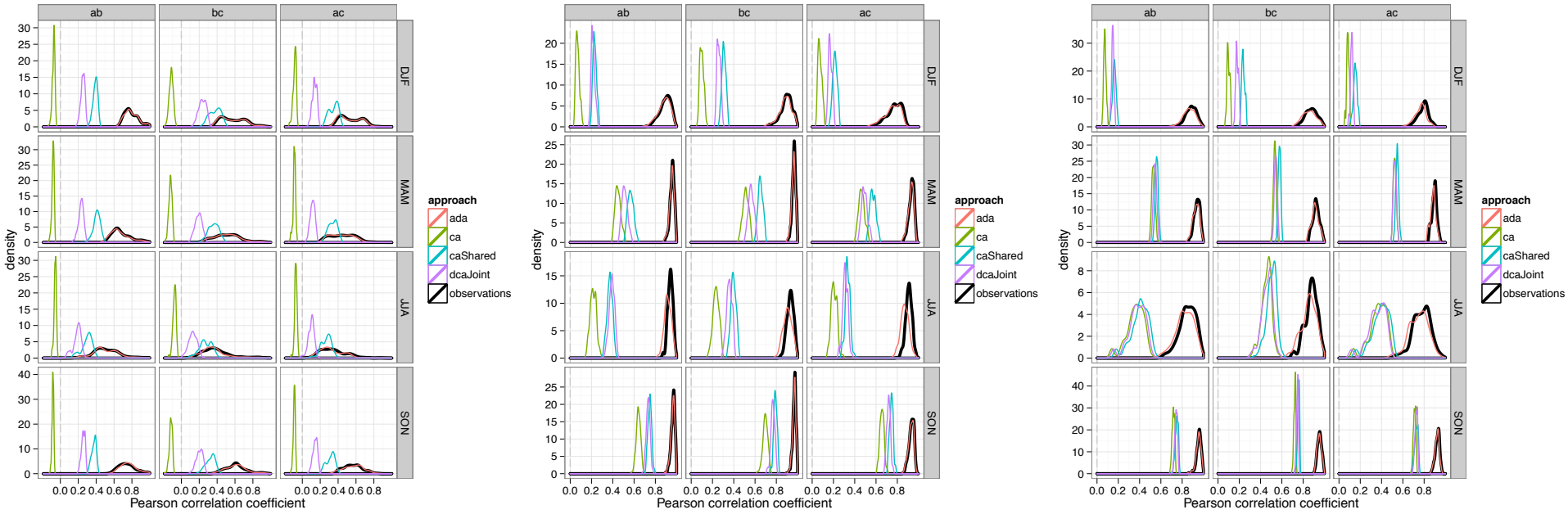
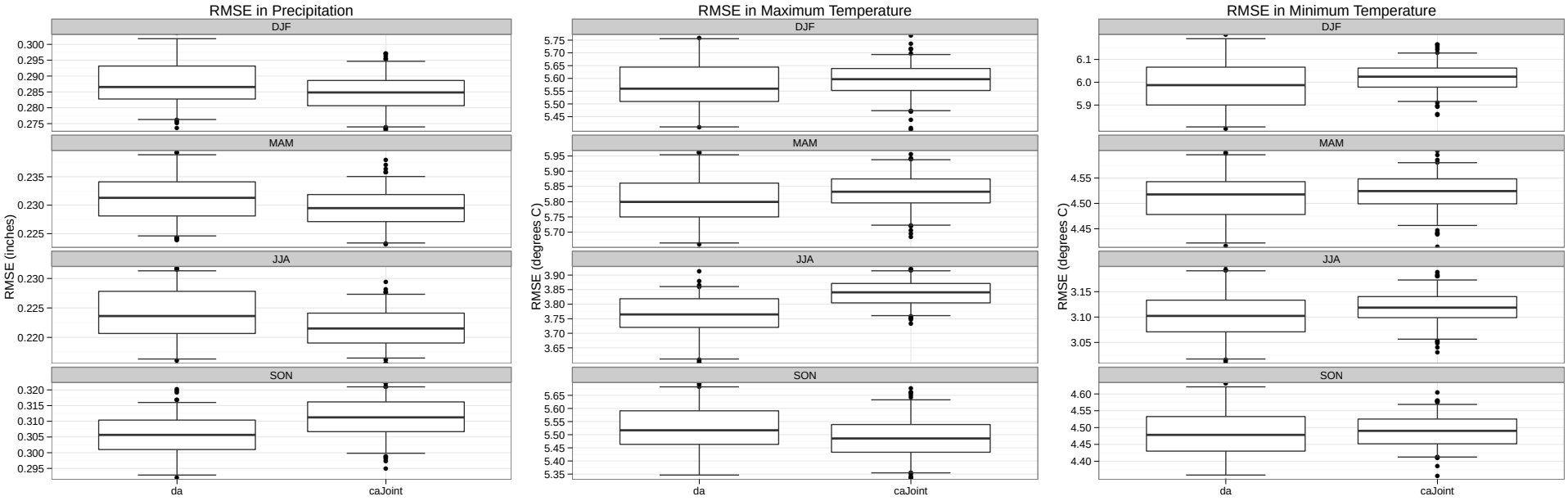


Figure 2.8: Smoothed probability densities of cluster-total seasonal precipitation for all methods



(a) Precipitation (b) Max Temperature (c) Min Temperature

Figure 2.9: Simulated and observed distributions of between-cluster correlations



(a) Precipitation (b) Maximum Temperature (c) Minumum Temperature

Figure 2.10: Root mean square error over all locations and simulations

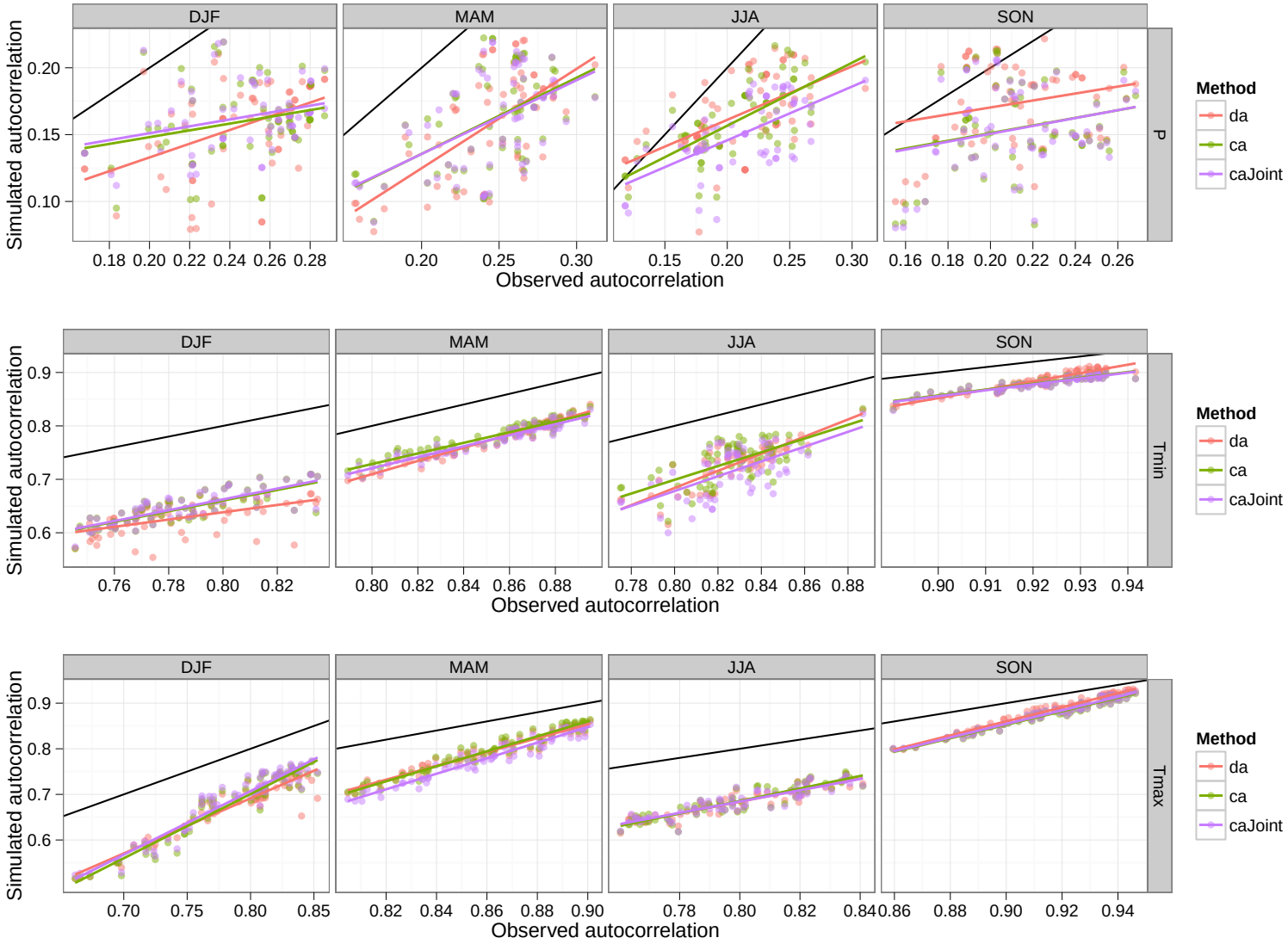
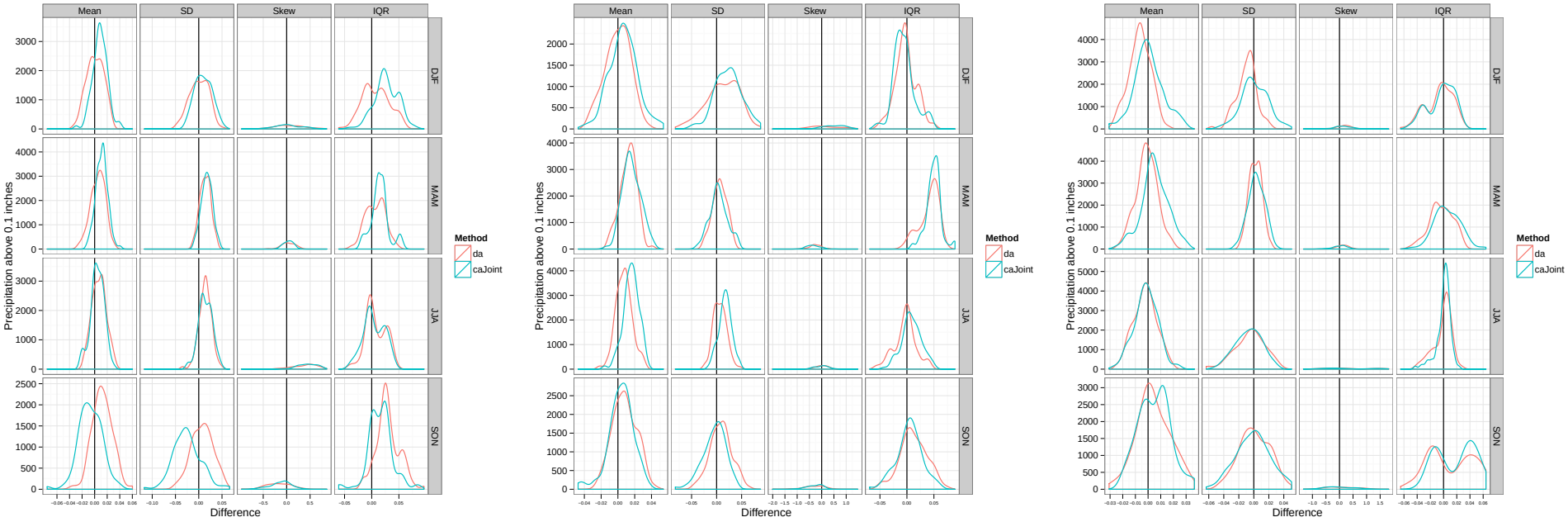
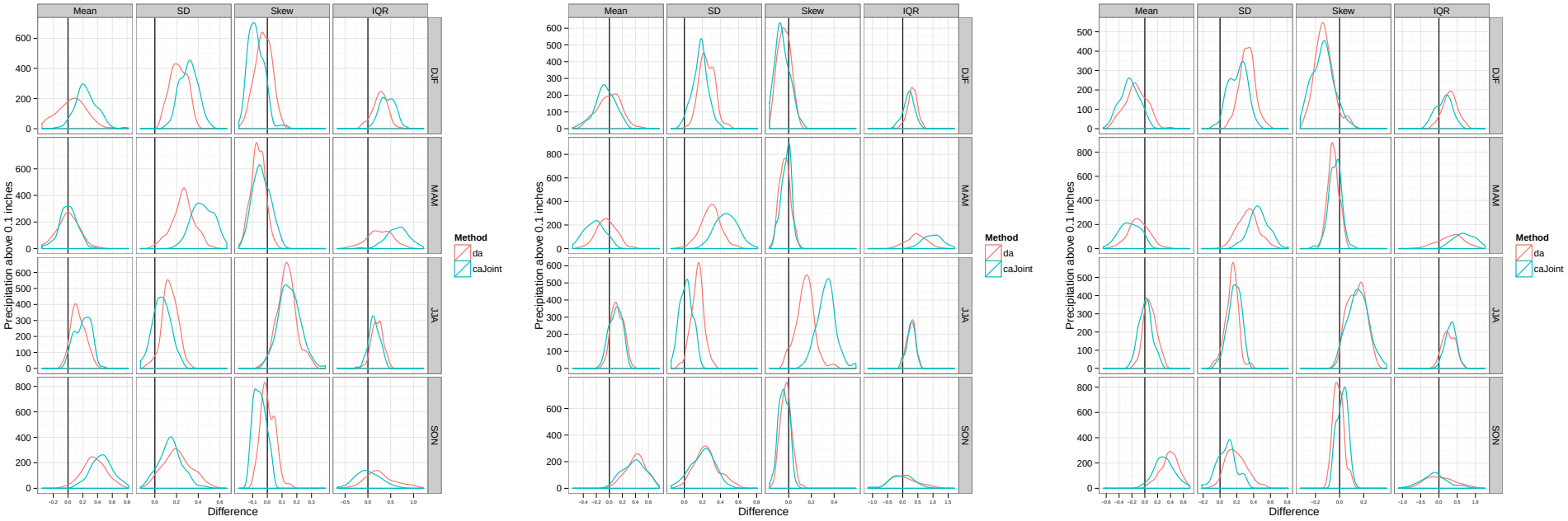


Figure 2.11: Simulated versus observed lag-1 autocorrelations



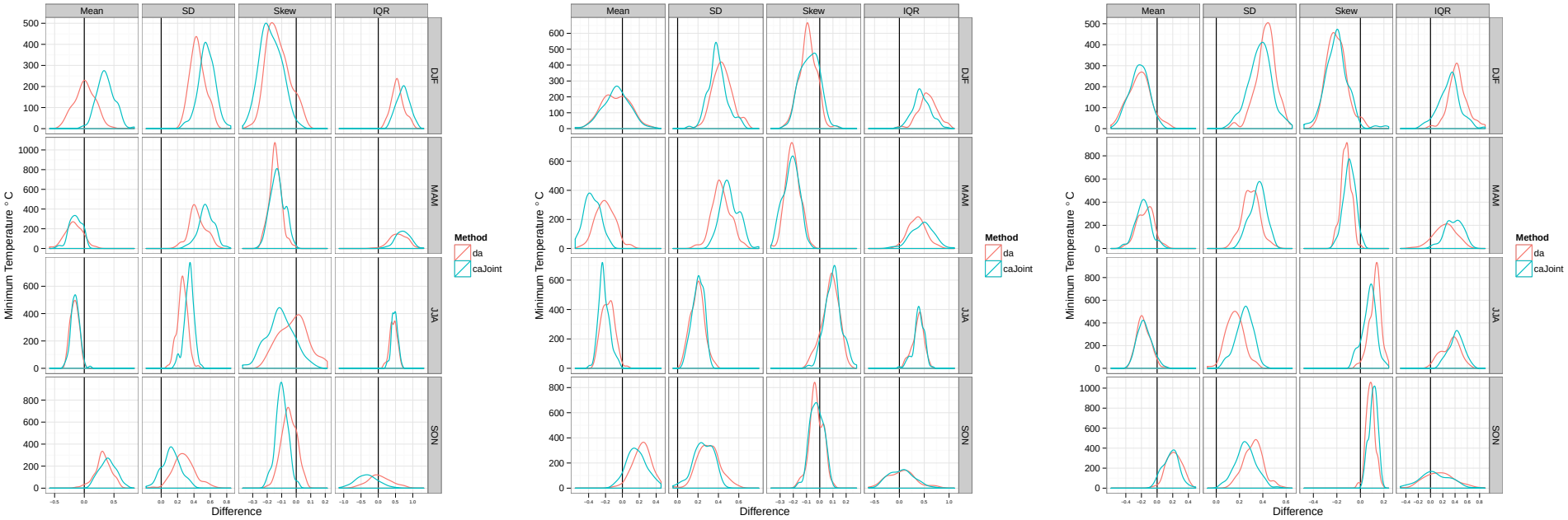
(a) drgc2hu (b) pidc2hl (c) bffu1ll

Figure 2.12: Distribution (over all simulations) of differences between observed and simulated statistics of precipitation



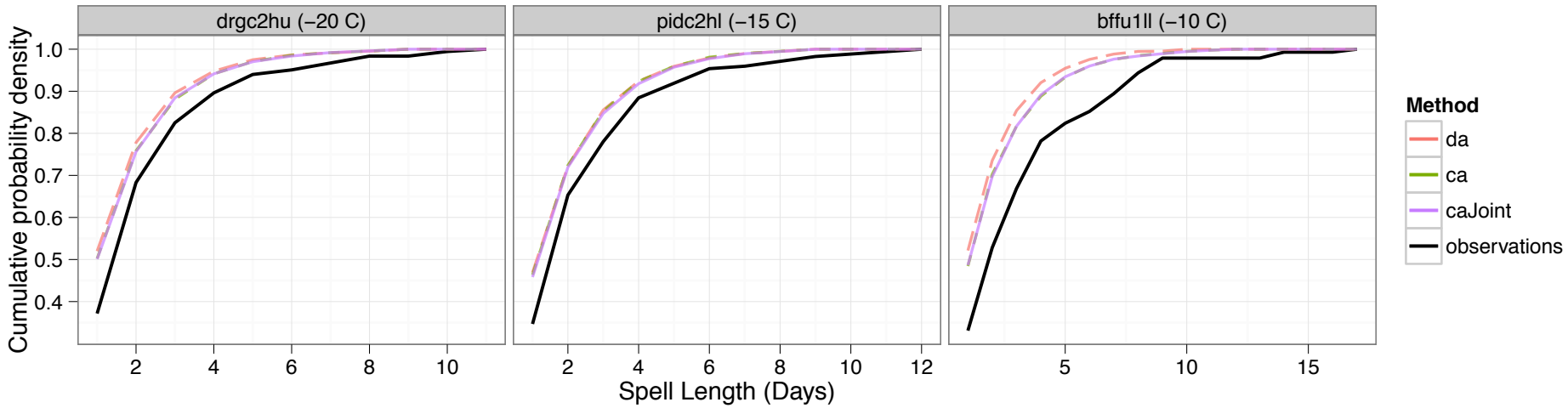
(a) drgc2hu (b) pidc2hl (c) bffu1ll

Figure 2.13: Same as Figure 2.12 for max temperature

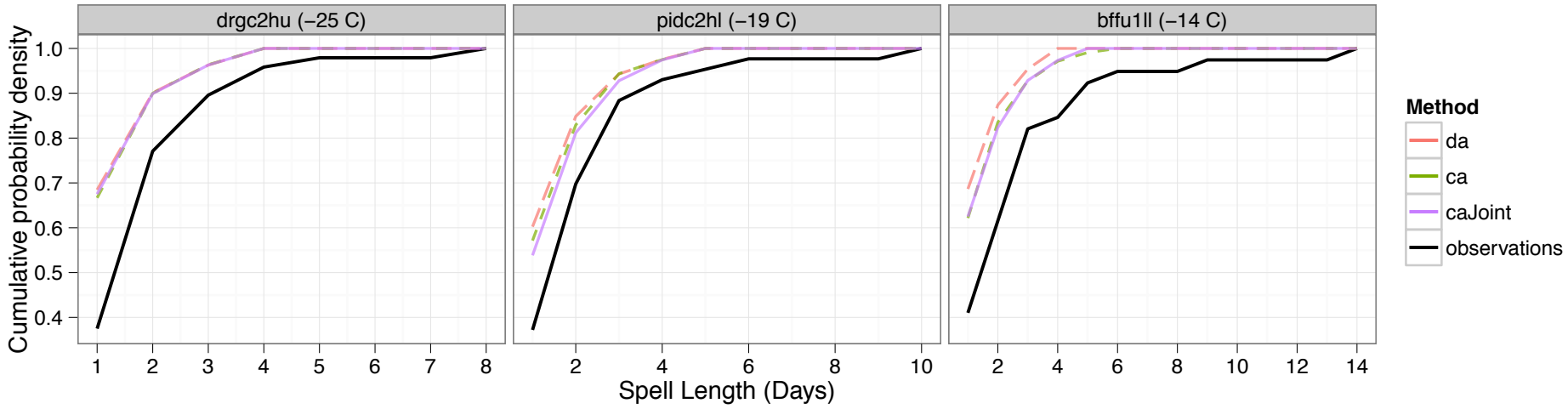


(a) drgc2hu (b) pidc2hl (c) bffu1ll

Figure 2.14: Same as Figure 2.12 for min temperature

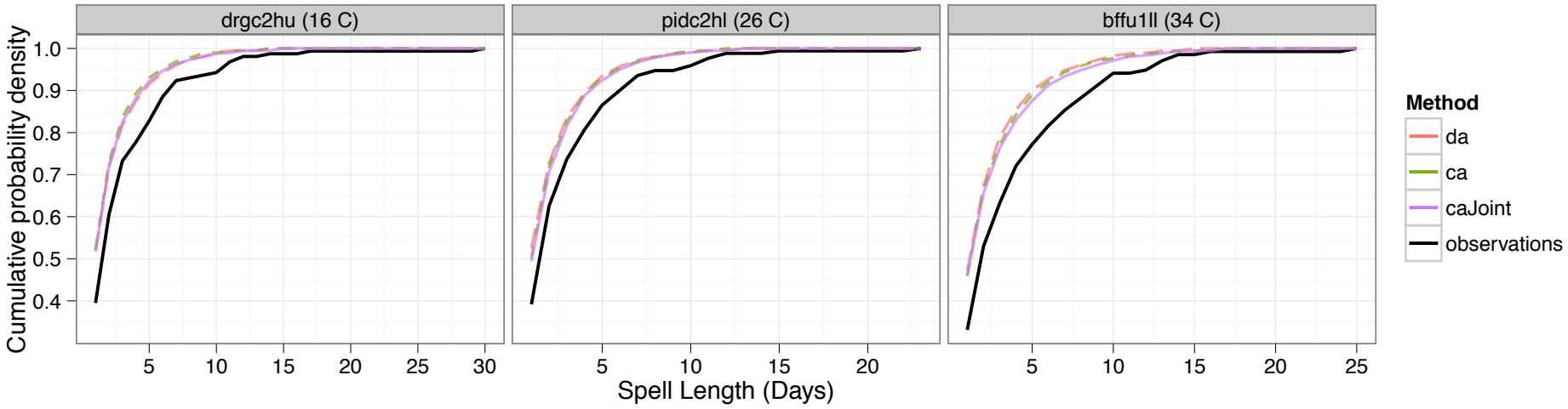


(a) DJF spells below 5th percentile

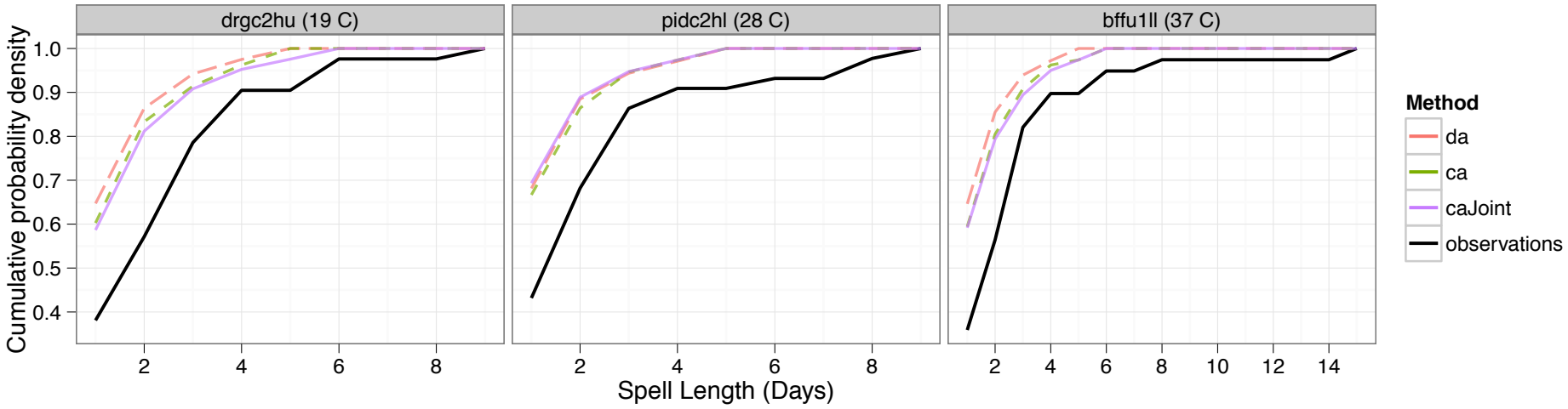


(b) DJF spells below 1st percentile

Figure 2.15: DJF cumulative probability density of temperature spells at three individual sites. Spell thresholds corresponding to quantile are shown with the location name in the individual titles. The median simulation at each site is shown.



(a) JJA spells above 95th percentile



(b) JJA spells above 99th percentile

Figure 2.16: Same as Figure 2.15 for JJA

2.4.3 Conditional Simulation

We chose two representative, somewhat strong, probabilistic climate forecast for the purposes of demonstration—dry simulation refers to ANB forecast of 10:25:65 and wet of 65:25:10. Figures 2.17, 2.18, 2.19, and 2.20 show the CDFs of seasonal precipitation for the three clusters and three forecast cases (dry, normal, wet), along with that of the observed. For DJF (Figure 2.17), the non-'caShared' methods perform well in that the CDFs shift in the appropriate direction relative to the observed for the three forecast cases. However, the CDFs show bias at the lower and higher precipitation values. The wet simulation CDF has a much more noticeable shift than the dry, though at the extremes the probabilities are only slightly different than historical. MAM (Figure 2.18) and SON (Figure 2.20) have similar patterns to DJF, though in SON the normal simulation no longer has a dry bias for high totals. The 'caShared' has a wet bias in all cases, especially for the lower elevation cluster. This is largely due to the fact that the precipitation state is shared and dominated by wet locations in the higher elevation cluster. For these three seasons across all clusters, it appears 'da' would be best for wet simulations and 'caJoint' for dry. In the following section, more detailed examination is given to the behavior seen here.

To understand the shifts in probability of exceedance we chose two thresholds, 50th and 75th percentiles of the seasonal total precipitation, and computed the exceedance probabilities and are shown as boxplots in Figures 2.21, 2.22, 2.23, and 2.24 for the four seasons, three forecast cases and clusters. The motivation behind this statistic is that shifts in precipitation risk will translate to shifts in hydrologic risk. The red triangles denote climatological exceedance probabilities. For example, one would desire higher probabilities of exceeding a threshold for a wet simulation. Figure 2.21a (DJF) shows changes in boxplots with respect to historical for exceedances above the 50th percentile—in dry cases the boxplots are shifted lower indicating a lower probability of exceedance and vice-versa for wet case. Both methods show consistent shifts but for normal case the 'caJoint' does better. Similar to the earlier CDFs, 'da' has a stronger shift for wet and 'caJoint' for dry. Similar observations can be noted for the other three seasons (Figures 2.22–2.24). The 'da' seems

to shift the probabilities strongly for wet simulations and caJoint for the dry, with JJA being the main exception. This is consistent with what we noticed in the unconditional case as well. in that, 'da' being the average of the entire domain is driven more by wet locations located mainly in the upper elevation cluster, while 'caJoint' attempts to separate the inhomogeneities in precipitation state, allowing the dry areas to be modeled as dry and the wet areas as wet.

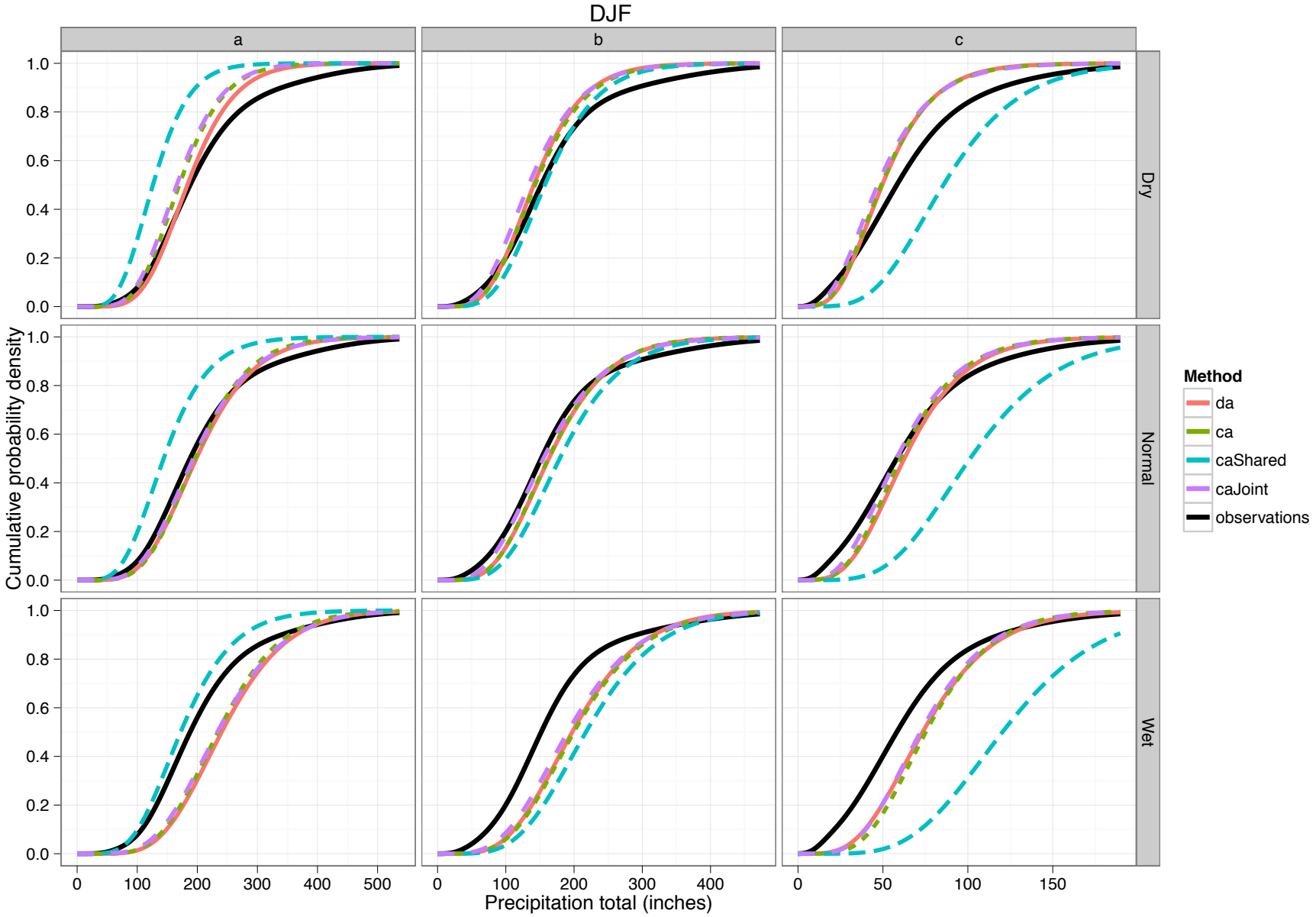


Figure 2.17: Cumulative probabilities of conditional, within-cluster seasonal precipitation totals for DJF, with black being historical

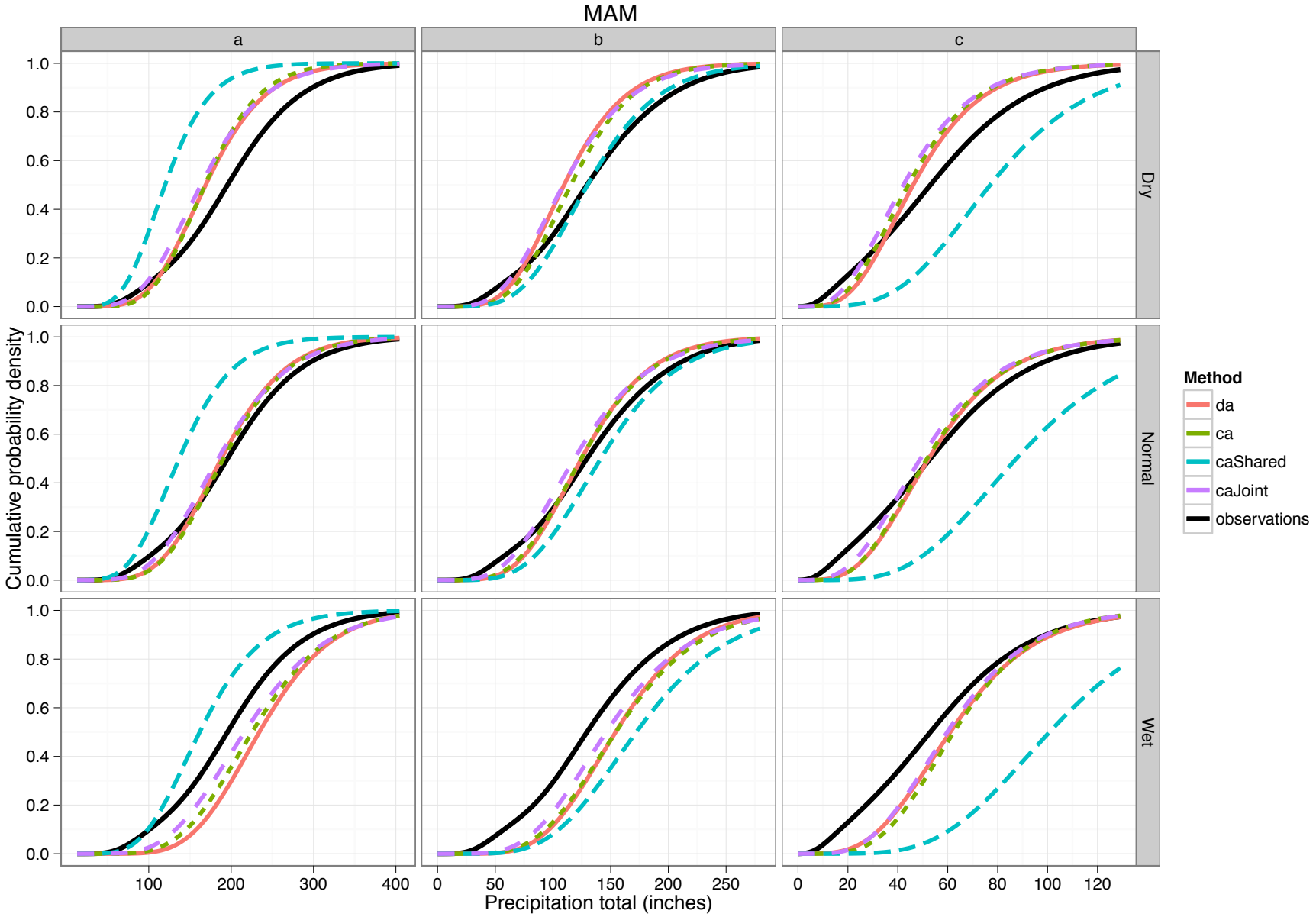


Figure 2.18: Same as Figure 2.17 for MAM

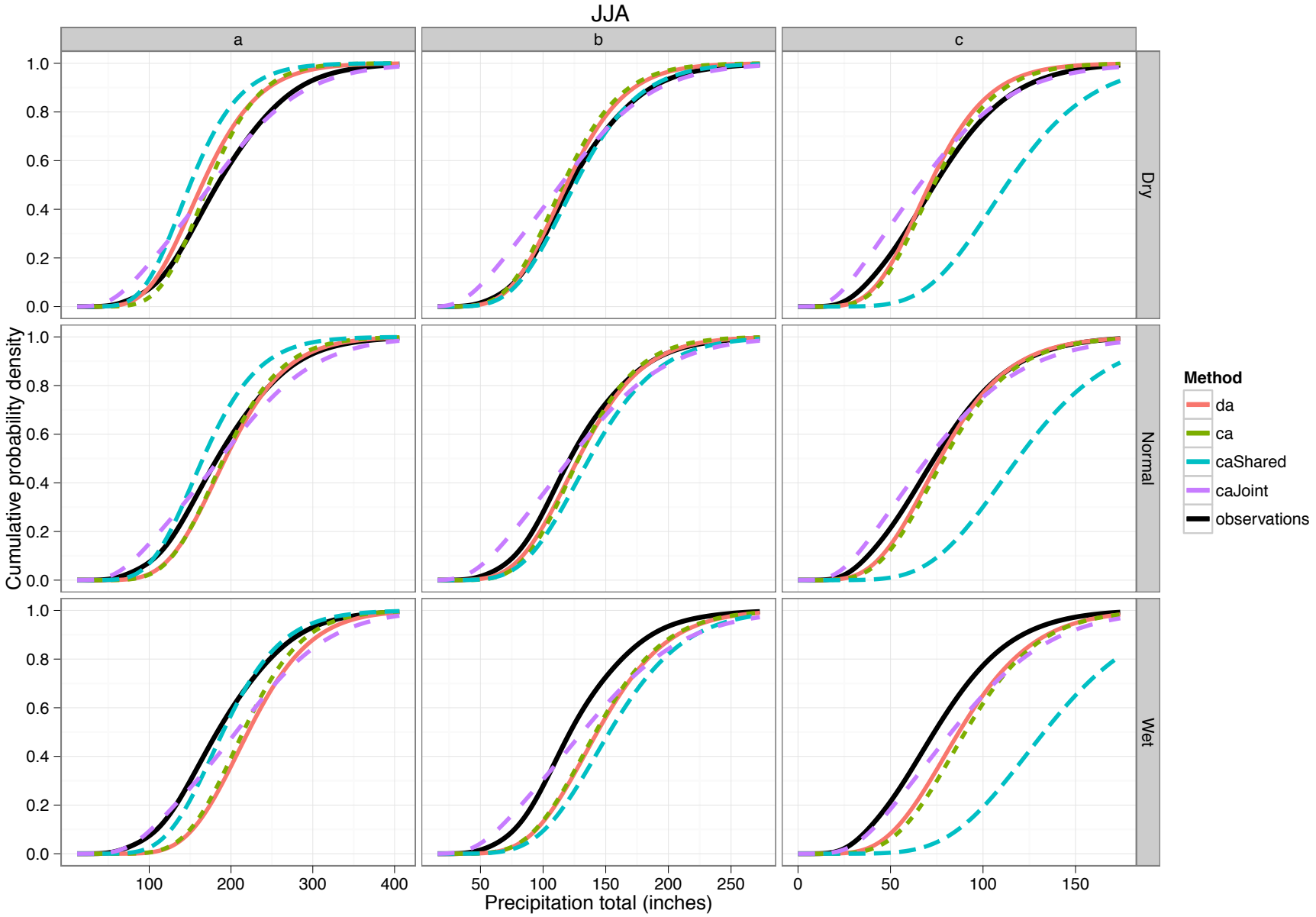


Figure 2.19: Same as Figure 2.17 for JJA

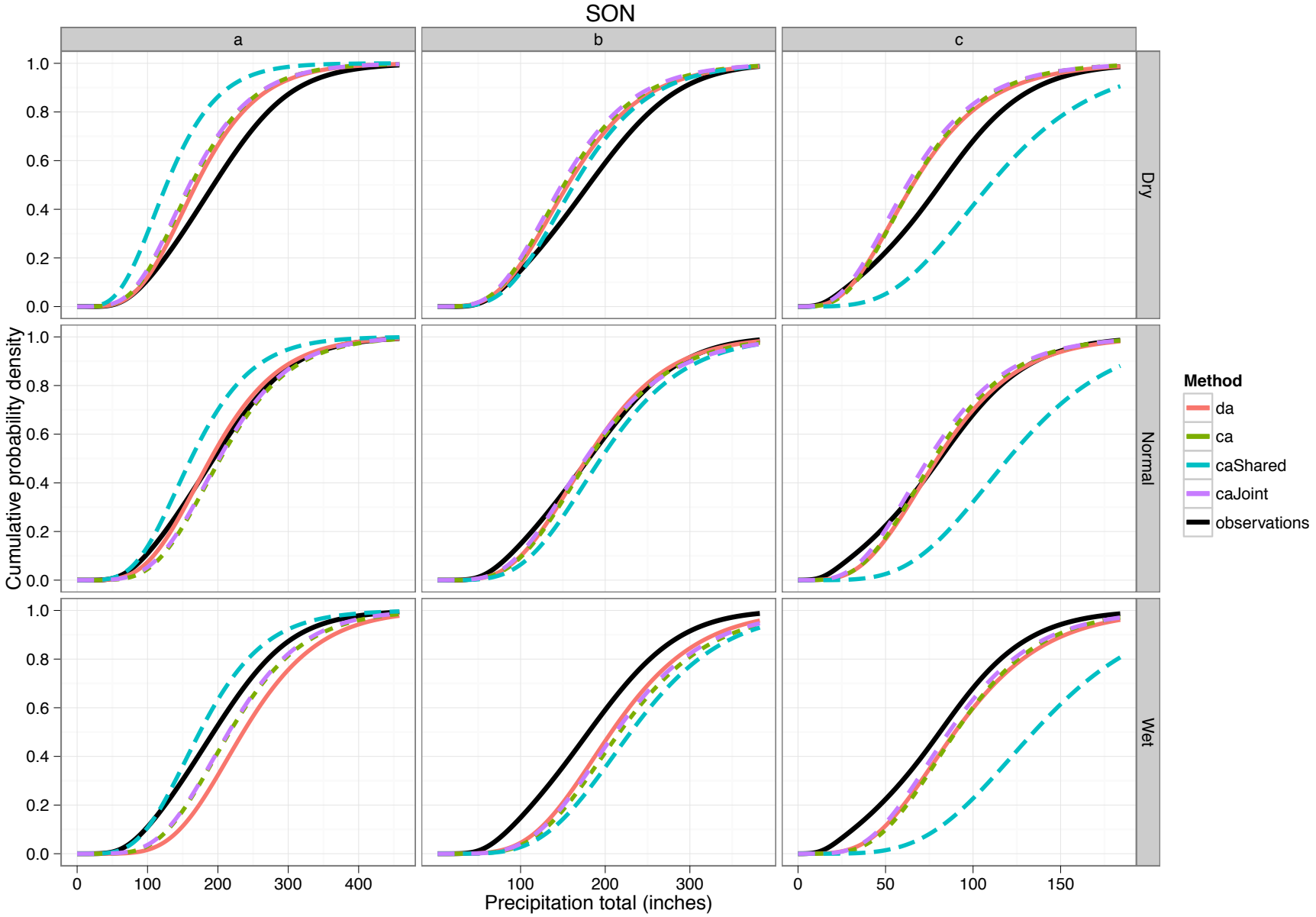
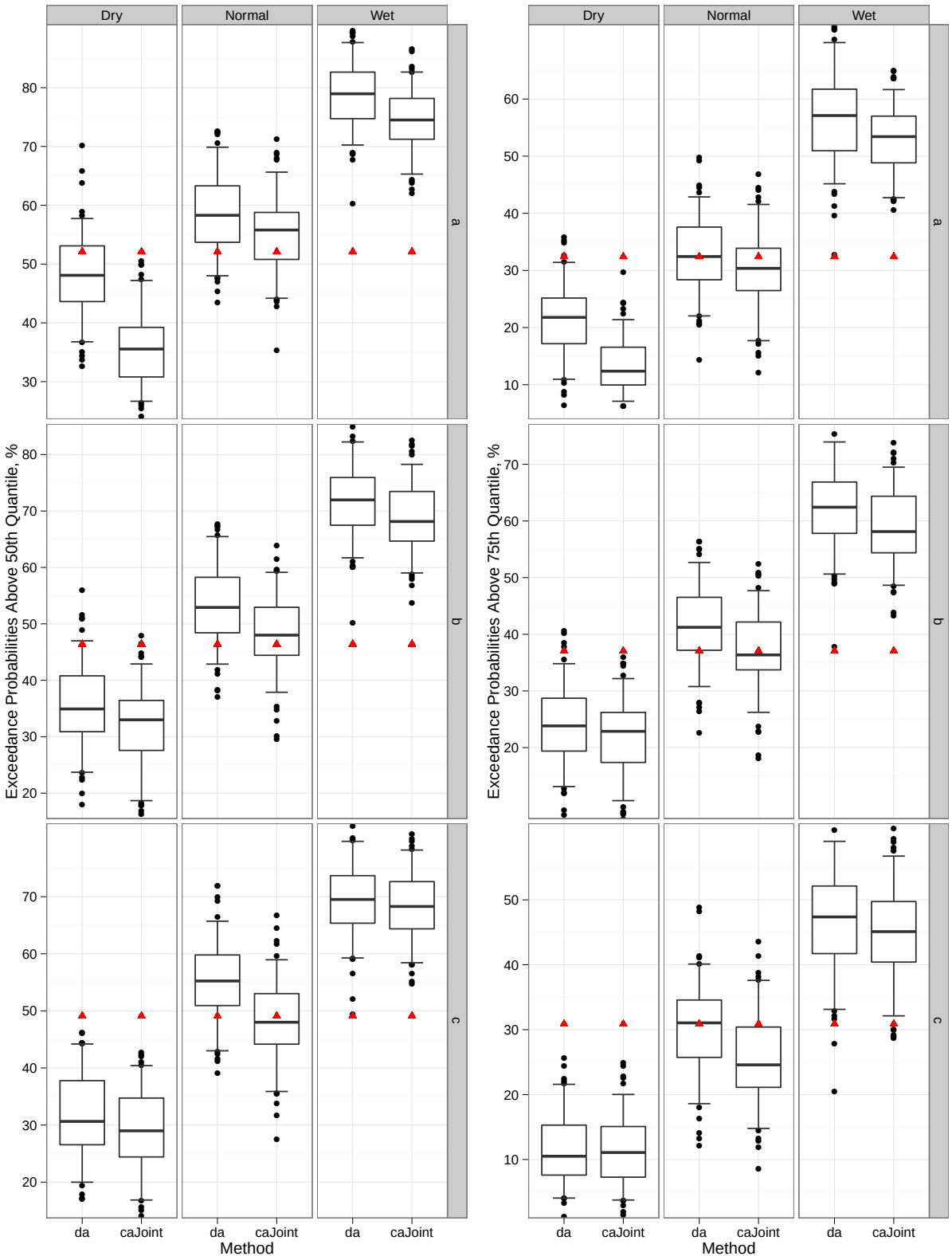


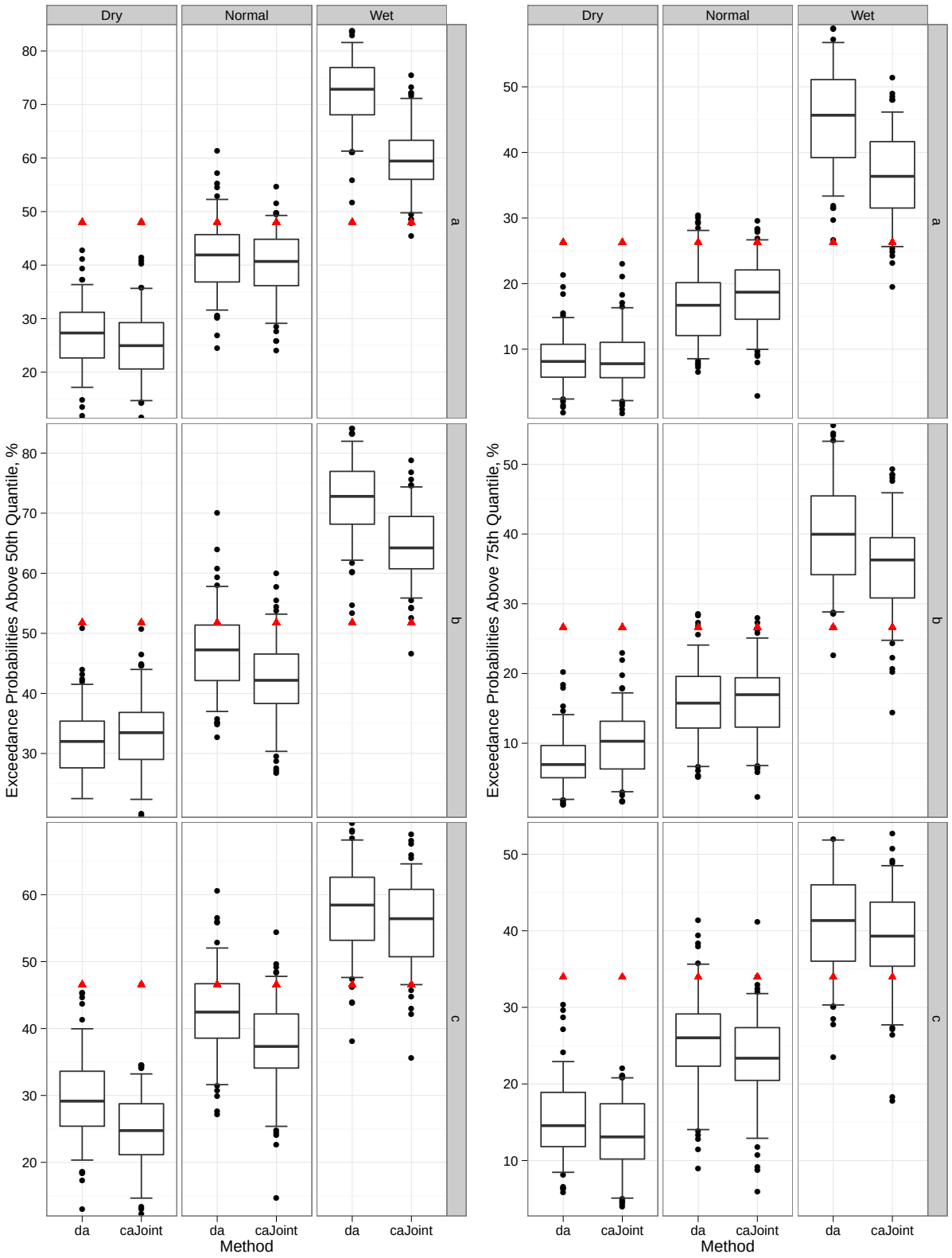
Figure 2.20: Same as Figure 2.17 for SON



(a) 50th quantile

(b) 75th quantile

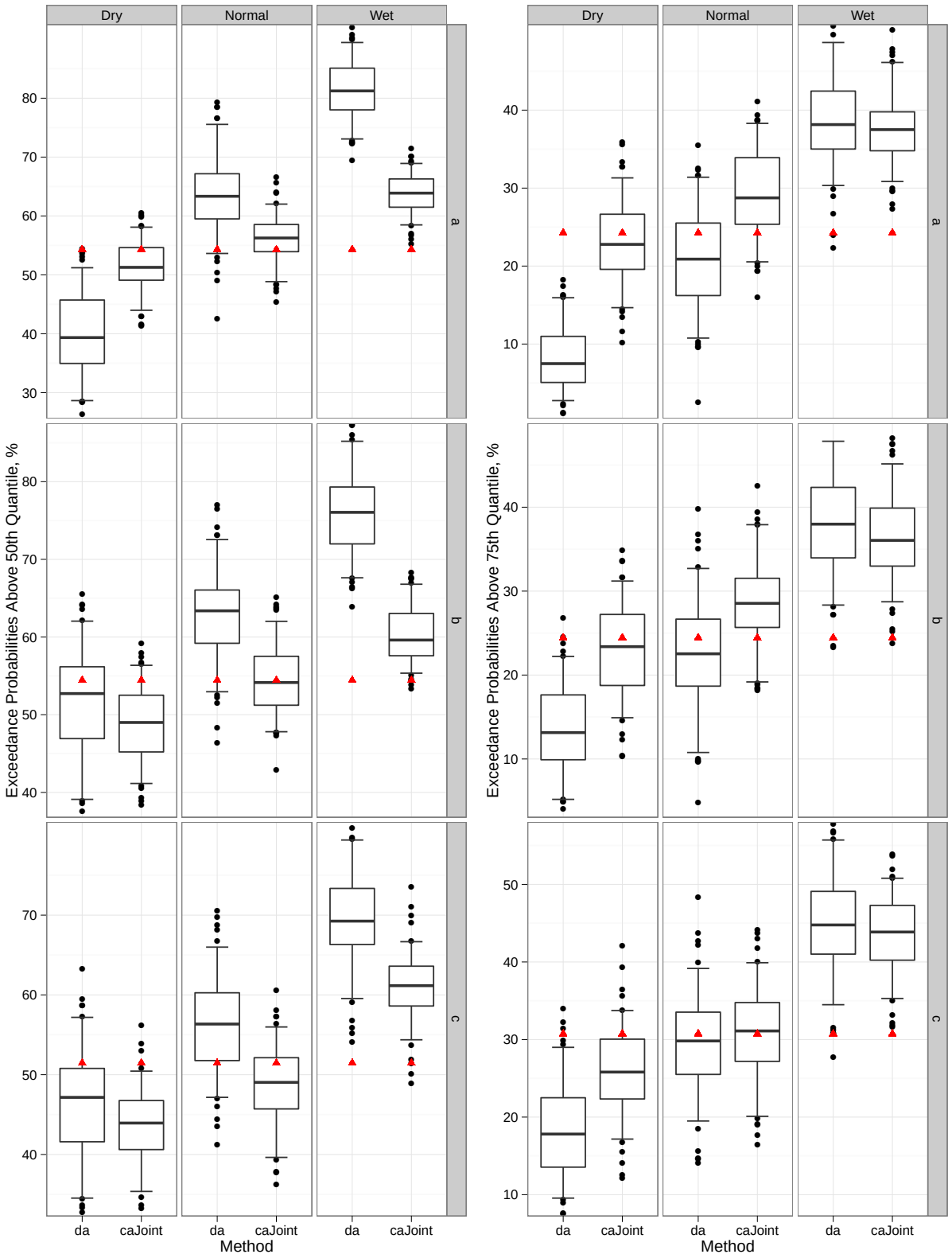
Figure 2.21: Shifts in exceedance risk for conditional within-cluster DJF seasonal totals



(a) 50th quantile

(b) 75th quantile

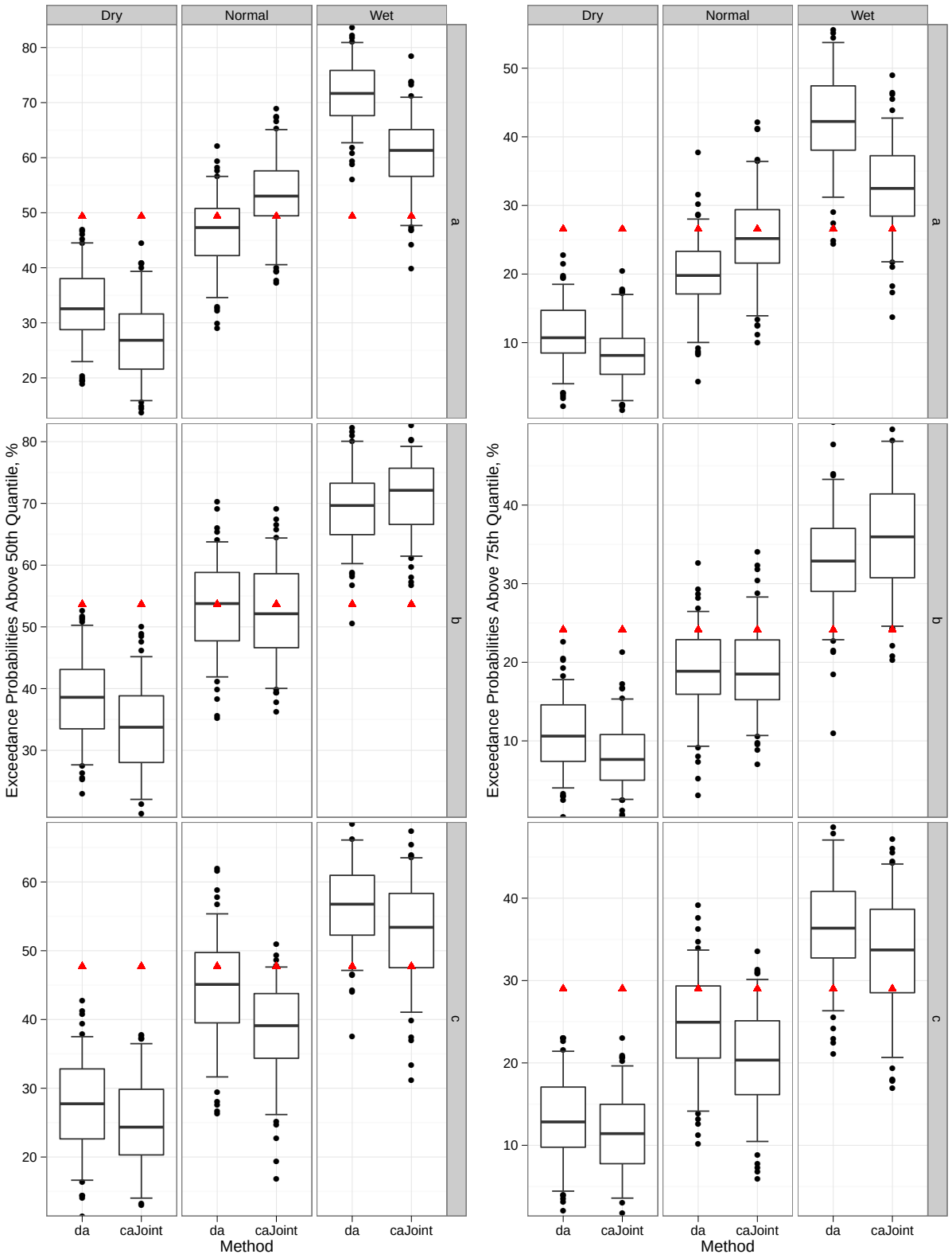
Figure 2.22: Same as Figure 2.21 for MAM



(a) 50th quantile

(b) 75th quantile

Figure 2.23: Same as Figure 2.21 for JJA



(a) 50th quantile

(b) 75th quantile

Figure 2.24: Same as Figure 2.21 for SON

2.5 Summary and Discussion

We developed a multisite and multivariable stochastic weather and demonstrated its application to the San Juan Basin. Four multisite methods were investigated: the standard domain-aggregate method, 'da', where weather is simulated (resampled) across all locations simultaneously, and three clustering-based approaches ('ca', 'caShared', and 'caJoint'). Clustering was performed on seasonal precipitation totals for the four seasons (DJF, MAM, JJA, SON). The 66 subcatchments of the San Juan were then divided into three groupings based on efficient minimization ('kink') of within sum of squares. The first approach, called cluster-aggregate ('ca'), applies the 'da' method to each cluster independently. Second, 'caShared', calculates the precipitation state as 'da' but then forces the same precipitation state for all clusters. The third method, 'caJoint', considers the precipitation state with each cluster and coordinates all possible states of the three cluster system using Markov transition probabilities.

Multiple statistics proved inadequacies of the 'ca' method as the complete independence assumption proved detrimental to between-cluster behavior. With 'caShared', within-cluster attributes suffered with the forced precipitation state calculated at the domain level. The 'caJoint' method was competitive with 'da' in certain situations. The 'caJoint' approach is most applicable in winter when spatial variance of precipitation is most reduced and precipitation has more spatial structure. In summer, 'caJoint' is worse than 'da' as there is little spatial precipitation structure and clustering reduces spatial variance much less than in winter. Both approaches can be considered viable options and may be dependent on the desired use of simulations. For instance, 'da' performs better for capturing distributional statistics of climatology, but 'caJoint' is better for temperature spells. All the statistics investigated have important roles in hydrologic modeling.

In addition, we modified the k-NN resampling component to allow for probabilistic seasonal forecasts. The proposed conditional methodology adds preferential treatment to seasons with desirable characteristics and will tend towards resampling days with low-frequency patterns. Unconditional simulations should be used for seasons where variability is not expected. Some figures

of seasonal precipitation totals displayed a tendency towards underprediction of tail values and overprediction of low-end values. The 'da' method proved the best at capturing heavy tails of distributions and would be the most suitable for pairing with a wet climate forecast-based simulation. On the other end, 'caJoint' appears to be the most appropriate for dry climate forecast pairing, excepting summer.

A complete comparison between 'da' and 'caJoint' would involve juxtaposing their results after linkage with a physical model to produce streamflow as they could have significant differences in forecasting streamflow. Currently, the NWS's ESP is limited by sequences that have only occurred in climatology. With the 'da' method, the entire basin is forced to behave the same due to the simultaneously resampling. The 'caJoint' approach offers more heterogeneity, but it can only be speculated as to whether this would offer an increase or decrease (or no change) in skill.

As for future modification, clustering can be performed in different ways. For example, daily could be the temporal scale instead of seasonal and if scaled, temperature or 500 mb heights could be added to the clustering matrix. For this watershed, elevation was heavily correlated with precipitation, where lower elevations corresponded to arid regions and higher to mountains. In other regions, say if precipitation had more of a spatial correlation, then may need more careful distinction of whether precipitation is frozen or liquid (Mezghani and Hingray, 2009). Such a distinction would be aided by the inclusion of temperature in the clustering matrix. Furthermore, cluster analysis was performed after post-processing of observed point values into areal-averaged values. This smoothing of likely heterogeneous station data may inadvertently promote 'da' as satisfactory over a clustering approach.

There is also likely a relationship between the joint state of the three cluster system and the magnitude of the precipitation in each cluster. This conditional probability could be introduced to further nudge correlation between clusters as well as simulation of extreme events on a basin wide scale.

It is also possible the ad hoc assignment of 3 clusters for each season may not be appropriate. Cluster analysis provides a means for dividing stations into natural groupings. If there are k^* "true"

groupings, then for $k < k^*$ the algorithm will not allocate observations in the same naturally occurring group to different estimated clusters. For $k > k^*$, one of the "true" groups must be partitioned into subgroups (Hastie et al., 2009). If clustering were to employed on an operational scale, an method would be needed to minimize consequences of subjectivity. The **Gap statistic** automates locating the "kink" by finding the largest gap between two curves: expected and observed log of dissimilarity within each cluster (Tibshirani et al., 2001). Given the variable behavior of the elevation boxplots in Figures 2.3a and 2.4a and calculated statistics, 'c' possibly could be a "true" group while 'a' and 'b' is the result of unnecessary splitting.

A common criticism of k-NN resampling is that values not seen in history are not produced. Leader & Buishand 2009 developed k-NN method that produces values beyond the range of history. AR1 processes yield residuals associated with historical observations, which can then be resampled. (Sharif and Burn, 2007) allows K-NN resampling with perturbation of the historic data via the addition of a random component to the individual resampled data points. Furrer and Katz (2008) have some interesting discussion on hybrid weather generators where an extreme value distribution can be incorporated to improve tail behavior. This would prove challenging for multiple sites, but parameters could possibly be estimated using the smoothing and interpolation methods presented in Kleiber et al. (2012).

Chapter 3

Advancing Ensemble Streamflow Prediction with Stochastic Meteorological Forcings for Hydrologic Modeling

3.1 Introduction

The National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) provides river forecasts to support the protection of life and property during flood events, enhancement of economic interests associated with water and climate, and basic hydrologic information ranging from economic to environmental needs. These services are accomplished through 13 U.S. River Forecast Centers (RFCs), where forecast services are modernized by the Advanced Hydrologic Prediction System (AHPS) program. In the seven-state Colorado River Basin (CRB), the Colorado Basin River Forecast Center (CBRFC) is the primary official provider of streamflow forecasts to the U.S. Bureau of Reclamation and others. Currently the CBRFC and the Natural Resource Conservation Service (NRCS) work together to create seasonal water supply outlooks in the CRB. Forecasts are generated by each agency and subjectively combined into a joint, official-forecast (Hartmann et al., 2002). NRCS uses a principle components regression (PCR) technique that primarily relies on current snowpack and proxies of soil moisture such as antecedent streamflow and autumn precipitation (Garen, 1992). At the seasonal time scale, CBRFC implements two techniques: Statistical Water Supply (SWS), a regression-based method that relates observed data (e.g., snow, streamflow, precipitation) with future streamflow, and the model-based Ensemble Streamflow Prediction (ESP). In conJunection with a physically based watershed model, ESP relies historical meteorological data as possible representations of the future; each historical year is then

used to simulate a streamflow trace (Day, 1985).

The structure of the AHPS program streamlines the issuance of 30- to 90-day probabilistic hydrologic forecasts. An official forecast describes a probability distribution using the 10th, 50th, and 90th percentiles. Ensemble forecasts have gained momentum in preference over deterministic forecasts, as probabilistic forecasts have been found to be more "appropriate and articulate" (Pagano and Garen, 2005) and offer more skill and relative economic value than deterministic forecasts (Roulin, 2007; Boucher et al., 2012). Ensemble forecasting is especially promising as the U.S. Bureau of Reclamation has developed the Mid-Term Operations Model (MTOM) (outlined in Grantz (2011)). MTOM is an objective, ensemble-based operations model where reservoir operations planning is engaged in a probabilistic mode. Having started running in experimental mode in 2010, it is a upgrade of their "24-month study", which helps anticipate monthly inflow volumes to the major Reclamation-operated reservoirs in the Colorado Basin. More robust operations management and planning not only aids Reclamation, but can further affect water and energy utilities such as Denver Water or PacificCorps (Werner, 2011b).

Being largely snowmelt-driven, western U.S.streamflow forecasts have a particular predicament—the most skillful predictions come from April 1st snow water equivalent (SWE) but many critical management decisions are made months beforehand during fall and early winter. Promising work has been developed for forecasting at long lead times before snowpack has accumulated. Statistical forecast models (Grantz et al., 2005, 2007; Regonda et al., 2006; Bracken et al., 2010) have shown skill improvement for early winter forecasting using large scale climate predictors. Despite evidence that large-scale climate variability have important consequences to western hydroclimatology, such information is only used minimally in operational forecasts (Hartmann et al., 2002) or there is no structured framework for incorporating such climate information into water-resources decision making on a basin scale (Grantz et al., 2007).

As ESP is already embedded in the NWS architecture, modifying rather than replacing ESP has been of great interest in recent years and a variety of techniques have been developed. Variations included calibrating models linked to ESP based on observed flow errors (Wood and

Schaake, 2007), or a reverse-ESP with an ensemble of initial conditions (Wood and Lettenmaier, 2008). Incorporating climate signals allows for weighting of ensemble members, ranging from using Niño-3.4 indices (Werner et al., 2004) to including multiple indices (Najafi et al., 2012). However, ESP methodology has shortcomings in that the ensembles created are hindered by limited historical data, which becomes even more limited with the addition of climatological forecasts. Improvements over ESP have been developed for short-to-medium (days-to-weeks) range forecasting with the use of ensemble weather forecasts (see Cloke and Pappenberger (2009) for a review). As these weather forecasts come from numerical weather models, reliability deteriorates after a medium-range, while their climate-scale signals may still reduce uncertainty in ESP. Thus ESP still has the advantage of providing uncertainty in long-range (monthly-to-seasonal) future climate.

We propose applying a weather generator as a downscaling link between climate forecasts and NWS models, thus preserving hydrology-relevant structures of inputs. This weather generator is a hybrid nonparametric statistical model that will create a rich variety of weather sequences beyond that found in climatology. This is a multi-site, multi-variable generator based on Apipattanavis et al. (2007) and developed in Chapter 2. Precipitation occurrence is modeled with a two-state Markov chain and weather variables are selected using a k-nearest neighbor (k-NN) resampling algorithm. Then, these weather sequences can be run through the watershed model using the ESP framework. Nonparametric methods lend themselves well to hydrologic modeling as corresponding sub-daily timesteps are associated with resampled days. Additionally, this model does not require extensive parameter estimation and is thus easily adapted to different watersheds (see Chapter 2 for summary of inflexibility with parametric weather generator models). Preferential resampling can also be performed based on seasonal climatological forecasts. Apipattanavis et al. (2010) successfully implemented their weather generator with a crop model to simulate yields. They assessed likely outcomes and production risks for seasonal forecasts of dry and wet climate. While linking downscaling models with hydrologic models requires a weather generation step, this linkage is still in a development process. Weather generator-produced streamflow has been developed for watersheds in various countries (Canada, France, England, Belgium, Korea, Iran) (Mountain and

Jones, 2006; Leander et al., 2006; Leander and Buishand, 2009; Khalili et al., 2011; Eum et al., 2010; Khazaei et al., 2011), but none for the CRB. This study goes beyond the cited work as a) our methodology produces streamflow forecasts rather than only simulations and b) their work was on very small watersheds that are dwarfed by those found in the CRB.

The paper is organized as follows. Section 3.2 outlines current and proposed methodology. Section 3.3 describes the application region and dataset. Section 3.4 presents and compares weather generator-based forecasts with those from regular ESP. Section 3.5 concludes with summary and discussion.

3.2 Proposed Framework

3.2.1 Current Methodology

Forecasters currently use components of the NWS River Forecast System (RFS), including the temperature-index snow model, Snow17 (Anderson, 1973) and the conceptual Sacramento Soil Moisture Accounting (SAC-SMA) (Burnash et al., 1973) model to simulate streamflow. The model creates flow by distributing moisture through a soil mantle divided into two zones (upper and lower), each with tension and free water components (Burnash et al., 1973; Burnash, 1995). Tension water consists of water so tightly bound to soil molecules that it can only be removed via evapotranspiration. Free water is not bound to soil particles and can freely move amidst the upper and lower zones. About twenty parameters determine direct runoff, surface runoff, interflow, and base flow in and between these zones. Figure 3.1 summarizes the moisture model. Calibration includes historical rainfall, temperature, and streamflow, where parameters are tweaked such that model output matches observed historical streamflow (Larson, 2002). The SNOW-17 model (Anderson, 1973) simulates snow accumulation and melt, and additional routing algorithms (e.g., Lag and K, Muskingum) distribute streamflow through the channel network. These models are calibrated to reproduce observed daily flow flow over historical period (in this case, 1981-2010).

The motivation for ESP derives from attempting to capture a more physical basis for esti-

mating future hydrology. ESP assumes historical meteorological data are representative of possible future conditions, and uses historic sequences as input to the hydrologic models. Using SAC-SMA within the ESP framework produces an ensemble of deterministic predictions by using current conditions generated by the snow and soil moisture models to define catchment states at the start of the forecast. After initializing forecasts with these model states, distinct hydrologic traces can be generated using meteorological inputs from each year of the historical record. The ensemble product is a range of streamflow values with an associated probability. Since ESP is not a stochastic model, only model forcings uncertainties are captured in the probabilities. The ESP representation of forecast uncertainty is problematic due to ignoring model calibration and data, which leads to the tendency to produce overconfident forecasts with narrow forecast distributions (Wood and Schaake, 2007).

Recently, the NWS forecast system has been updated to be contained within the Community Hydrologic Prediction System (CHPS) software infrastructure. CHPS is the NWS implementation of Deltares's Flood Early Warning System (FEWS). Figure 3.2 simplifies the ESP progression which occurs inside CHPS. Climatological weather sequences serve a dual role in forcing the model to produce watershed states (e.g., snow depth, soil moisture, etc) that initialize each forecast, and serving as meteorological input ensembles during the forecast period. For example, for forecasting the 2004 runoff season on March 1st, 2004, the model is forced with 6-hourly 2004 historical weather up to March 1st. Then, however many N years in the forcing record will provide N traces.

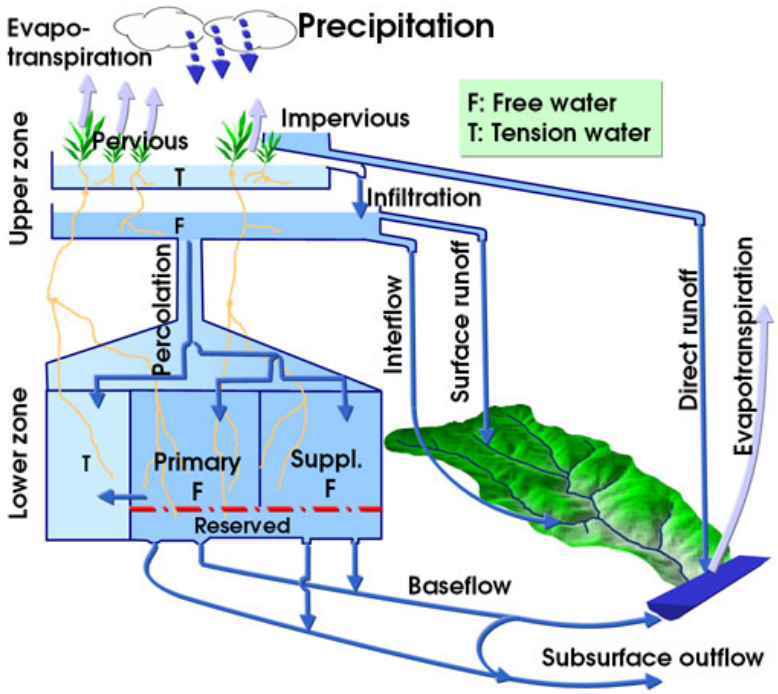


Figure 3.1: Sacremento soil moisture accounting model schematic (Werner, 2011a)

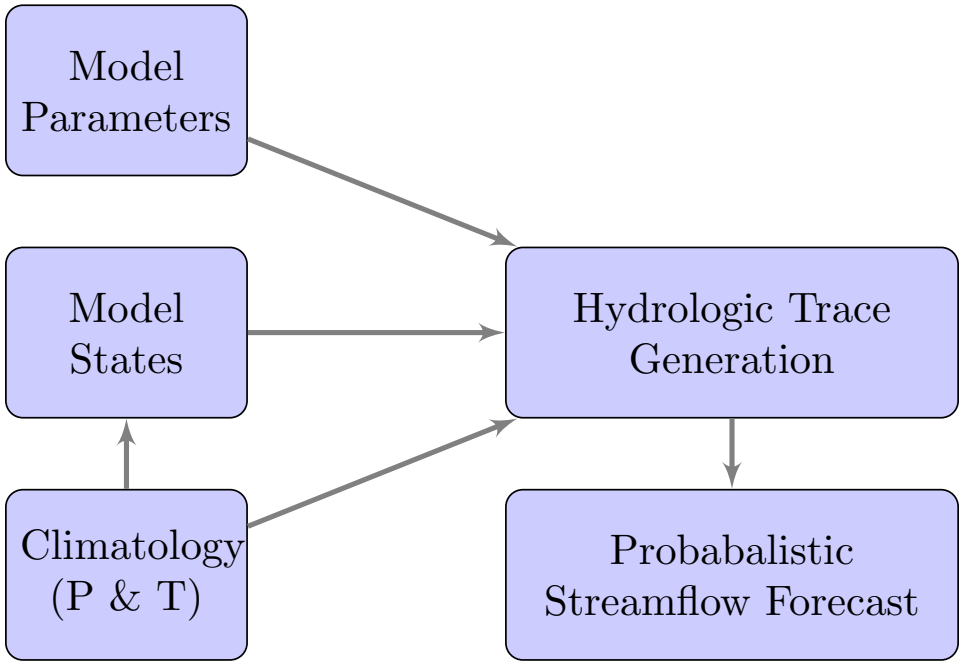


Figure 3.2: ESP flow chart

3.2.2 Proposed Improvement

Figure 3.3 outlines the change in ESP with weather generator results; model states will remain determined by climatology (i.e., forcing record) but the traces themselves come from weather generator ensembles. In the example from earlier, a forecast starting March 1st, 2004 will still be initialized by 2004 conditions, but the weather sequences are no longer restricted by the n years of the historic record. As the weather sequences are formatted similarly to climatology, this methodology offers ease of implementation over pre- and post-processing methods for ESP ensemble members. In later sections, these results will be referred to as 'WG ESP'.

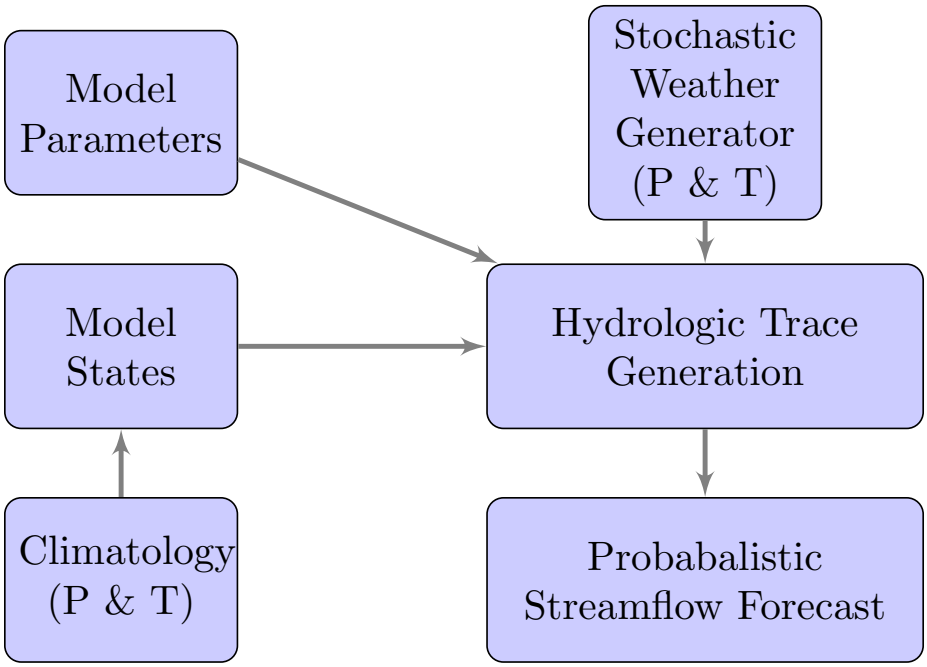


Figure 3.3: ESP flow chart with weather generator input

3.2.2.1 Weather Generation

We refer to Chapter 2 for a detailed explanation of the weather generator. To summarize, for a given day t, to sample areal-averaged weather vectors from the conditional PDF, $f(x_t \mid DOY, x_{t-1}, S_{t-1}, S_t)$, of areal-averaged weather vectors, x_t, given the day of year, DOY, yesterday's

simulated weather, x_{t-1}, yesterday's precipitation state, S_{t-1} and today's precipitation state. Then, k-NN resampling occurs based on this conditional PDF. Regarding the multisite methods explored in Chapter 2, we chose to first proceed with the 'da' method. Of the three cluster based methods explored, only 'caJoint' proved competitive to 'da' on a basin-wide scale. The 'da' method is also the simplest to implement computationally.

Synthetic weather sequences are often required that are consistent with large-scale, seasonal climate forecasts. We incorporate the seasonal precipitation forecasts issued by the International Research Institute for Climate and Society[1] into our methodology. These forecasts are expressed as the probability of each seasonal precipitation tercile, above:normal:below (ANB). Then, the k-NN resampling will undergo preferential resampling based on these probabilities.

3.3 Application Region and Data

3.3.1 Basin Characteristics

With a drainage area of approximately 25,000 miles^2, the San Juan River is the second largest tributary of the Colorado River and runs a distance of 355 miles. Having an area near in size to West Virginia, drainage areas are about 39, 23, 20, and 17 percents in New Mexico, Colorado, Arizona, and Utah, respectively. The San Juan river basin includes a wide range of elevations, from roughly 4000-14,000 ft above sea level, and climate zones, including desert and forest. Winter snow and rain due to frontal storms and modest rains from convective storms in summer are the main moisture input to the basin. Precipitation can vary from above 60 inches annually in the mountain peaks, to below 10 inches in the desert plateau[2] .

The San Juan Basin contains the Navajo Reservoir, which when filled occupies 15,610 acres, with a total capacity of 1.7 million acre-feet (MaF) and an active capacity of 1.0 MaF.[3] Major tributaries above the dam are the Navajo, Piedra, and Los Pinos Rivers. Of these only the Los Pinos is dammed (for agricultural purposes). The Animas River is the major tributary below the Navajo

[1] http://iri.columbia.edu/climate/forecast/net_asmt/
[2] http://www.usbr.gov/uc/wcao/rm/sjrip/
[3] http://www.usbr.gov/uc/wcao/water/ps/navajo.html

Dam and is free flowing. A majority of the reservoir inflow occurs during the April-July runoff when an average of 0.66 MaF enter the reservoir (USBR, 2008). Operations of the reservoir have been heavily influenced in the past decade by recommended flow for endangered fish populations (Holden, 1999), and the basin overall is managed through a complex set of federal and state laws and river compacts.

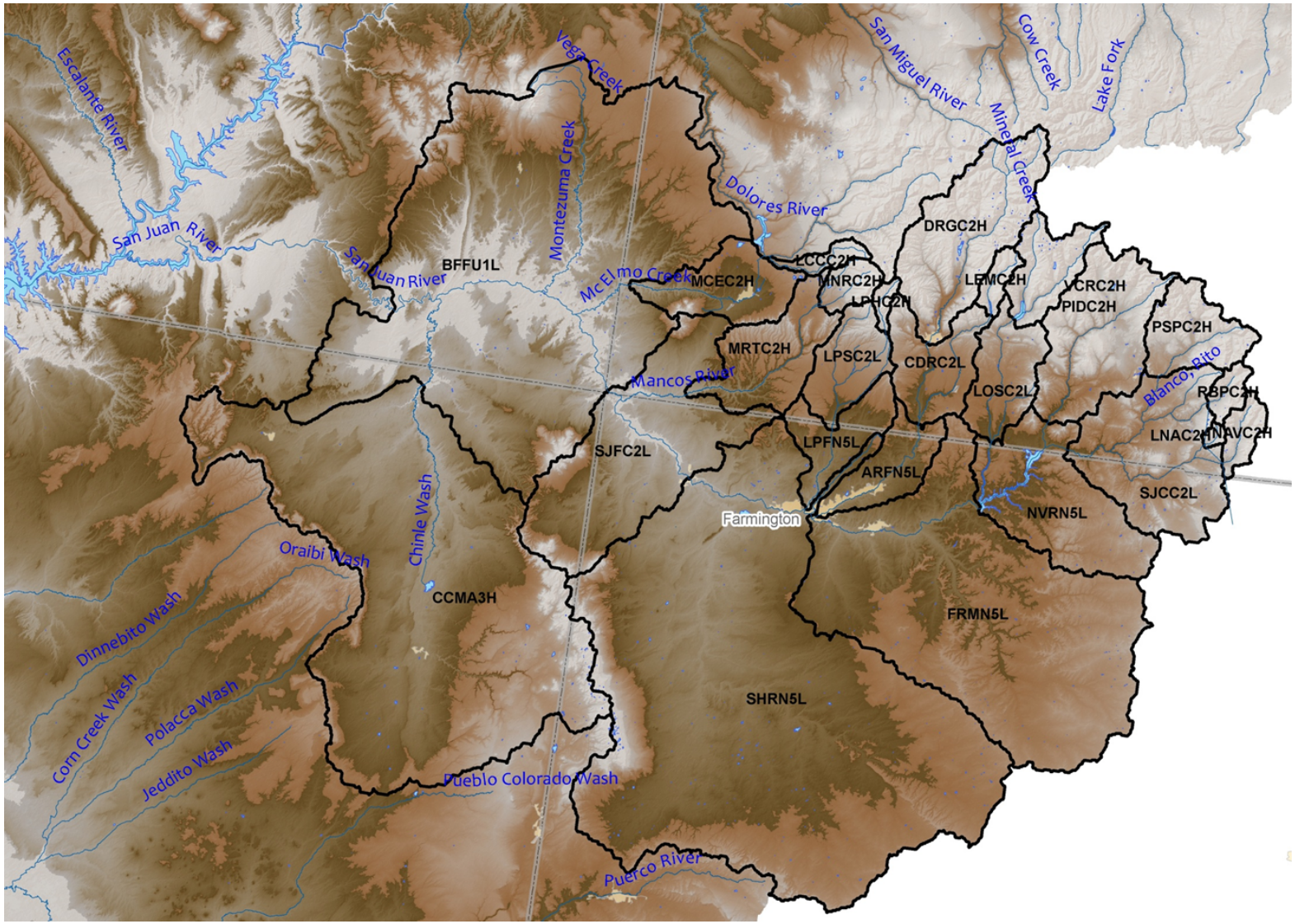


Figure 3.4: San Juan Watershed

3.3.2 Data

3.3.2.1 Weather Variables

The CBRFC divides the San Juan river basin in to 24 sub-basins (Figure 3.4) which are further divided by elevation bands into two to three zones each, resulting in a total of 66 spatial zones. Based on observed historical data at weather stations scattered across the basin, the CBRFC has created mean areal precipitation (MAP) and mean areal temperature (MAT) for each zone (NWSRFS, NWSRFS) at 6-hourly time steps. These start in October of 1980 through September of 2010. From the 29 complete years, we calculate daily precipitation totals as well as daily minimum and maximum temperatures which serve as inputs to our stochastic weather generator. As the weather generator samples daily values, daily precipitation was defined as the sum of the four values per day, and minimum and maximum temperatures were defined as the minimum and maximum, respectively. After running daily weather generation, the corresponding 6-hourly MAP and MAT are pulled from the daily index for input into the hydrologic model.

3.3.2.2 Streamflow

Historical monthly streamflow values were extracted from `http://wateroutlook.nwrfc.noaa.gov/`. Four representative gages were selected for analysis. BFFU1 (near Bluff, UT) is the last gage on the main stem of the San Juan before entering the Colorado River. DRGC2 (on Animas River, near Durango, CO) is a headwater gage in the San Juan mountains. NVRN5 is the gage upstream of the Navajo Reservoir. FRMN5 (near Farmington, NM) is on the San Juan downstream of the Animas confluence. Their locations are displayed in Table 3.1. Figure 3.5 has historical volumes for April–July runoff season for these gages. Of the four, only DRGC2 has not been adjusted to account for human activities (no need being a headwater region).[4] These gages exhibit high yearly variability with respect to their averages and also appear

[4] Monthly values have Standard Hydrometeorological Exchange Format (SHEF) codes corresponding to the type of data. DRGC2 is QCMRZZZ (collected flow), BFFU1 and FRMN5 are QCMPAZZ (adjusted monthly streamflow, likely from QADJUST program), and NVRN5 is QCMPBZZ (adjusted monthly streamflow, external agency provided or calculation). Further details concerning SHEF can be found here: `http://www.nws.noaa.gov/directives/sym/`

correlated with each other.

Table 3.1: Characteristics of representative locations (gages)

Gage	Lat	Long	Elevation (ft)
BFFU1	37.1469	-109.8642	4048
FRMN5	36.7228	-108.2250	5230
NVRN5	36.8244	-107.5919	5775
DRGC2	37.2792	-107.8797	6502

3.4 Results

As described in Section 3.2.1, hydrologic state information is needed before running ESP. As our study regenerates historical forecasts from 1981–2010 (a practice termed "reforecasting"), a retrospective continuous historical simulation ("hindcasting") is performed to generate watershed initial conditions (states) that can be used to initialize the reforecast ensembles. To evaluate the performance of the synthetic climatological forcings from a hydrologic simulation perspective, the reforecast process was run both for ordinary ESP (as described in Figure 3.2) and for the WG ESP. In contrast to other techniques where only preceding years are used for each year (Bracken et al., 2010), the entire period of record was used as traces for each year (i.e. 1984 and later years are still used as traces for the 1984 reforecast). Both keeping and leaving out the forecast year is explored later. Typically in operational forecasting, model states can be subject to manual adjustments to reflect current observations or short-term forecasts. Additionally, short term weather forecasts (1–5 day quantitative precipitation forecast (QPF) and 1–10 temperature forecasts) are occasionally used. Records of these adjustments do not exist for the entire period of study, and are thus not included in our reforecasts (as established in Franz et al. (2003)).

The period of interest in the RFC runoff forecasts, termed 'water-supply forecasts', is the late

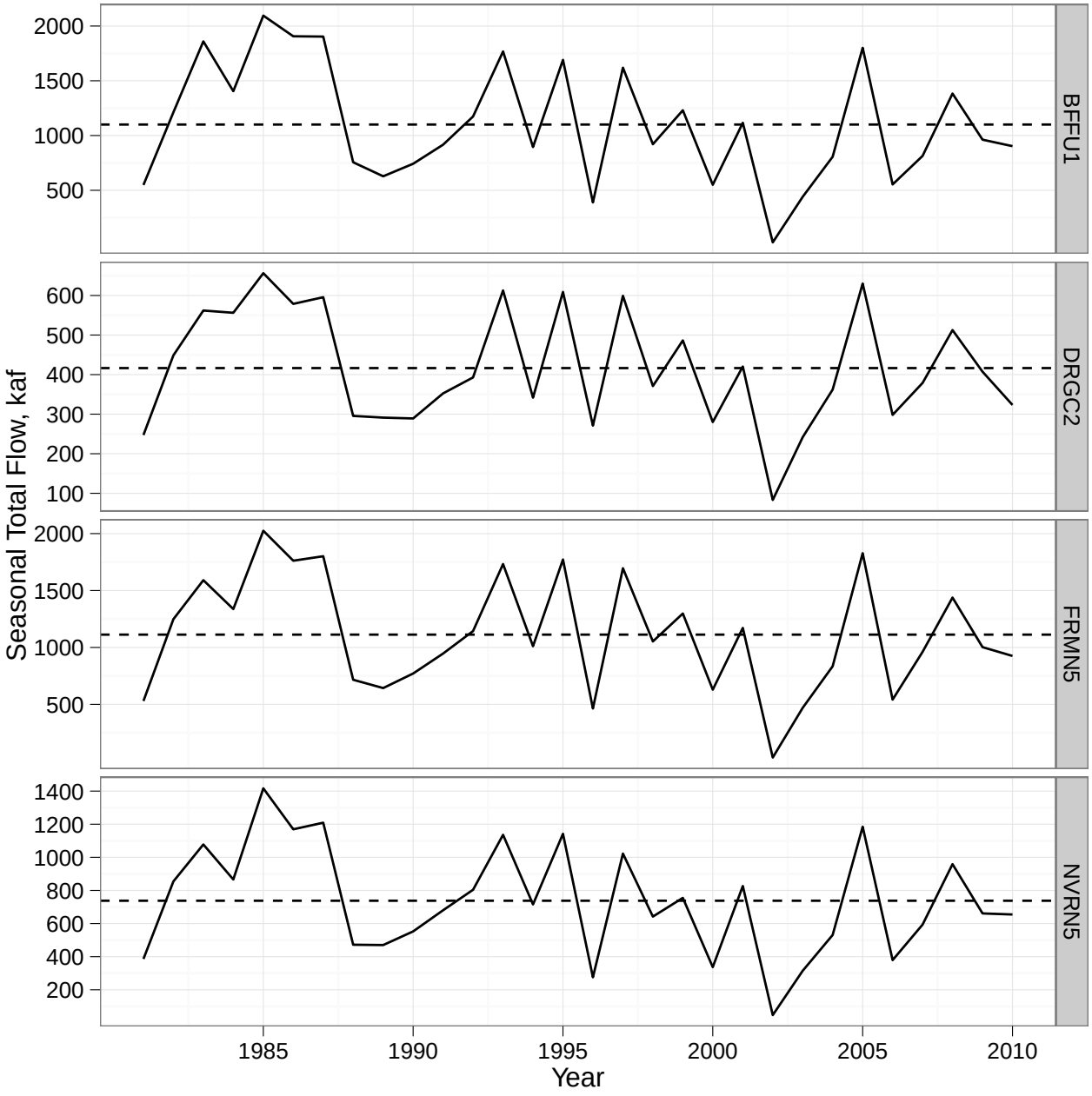


Figure 3.5: Historical April–July runoff volumes for representative locations with averages shown by dotted lines

spring and summer (particularly April–July) with issuance at monthly intervals from November 1 to April 1. April to July period runoff forecasts are displayed below while April to May and June to July sub-seasons appear in Appendix A. Unconditional forecast results (i.e., using climatology) were generated for the entire 1981–2010 period, and conditional forecasts were generated for one wet and one dry year. In addition to the verification described below, other measures of analysis include examining distributional statistics.

3.4.1 Forecast Skill Evaluation

We assessed the ensemble forecasts using a suite of performance metrics. The following techniques are as defined by Wilks (1995).

3.4.1.1 Rank Probability Skill Score

Ranked probability skill score (RPSS) evaluates model skill in capturing categorical probabilities relative to climatology. As generating ensemble forecasts can be described as a PDF, RPSS is an appropriate skill measure by its evaluation in probabilistic terms. Forecasts are divided into multiple categories — here we chose three (terciles, with boundaries at 33rd percentile and 66th percentile). The categorical probability forecast is obtained as the proportion of ensemble members falling in each category. The climatology forecast is the proportion of historical observations falling in each category; as we are using terciles, the climatological probability is one third. Flows are divided into k (here 3) categories for which the proportion of ensembles falling in each category constitutes the forecast probability in a given year, $(p_1, p_2, \ldots, p_k)$. The vector $(d_1, d_2, \ldots, d_k)$ designates whether the observation falls in the kth category ($d = 1$) or not ($d = 0$). The rank probability score (RPS) is defined as:

$$RPS = \sum_{i=1}^{k} \left[\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} d_j \right)^2 \right] \tag{3.1}$$

Then, RPSS is:

$$RPSS = 1 - \frac{RPS(forecast)}{RPS(climatology)} \tag{3.2}$$

RPSS ranges from negative infinity to one. Negative values denote a forecast accuracy worse than climatology; likewise positive is better than climatology and one is a perfect score.

3.4.1.2 Reliability Diagram

Reliability diagrams graphically represent the performance of probability forecasts, consisting of a plot of observed relative frequency, $\bar{o}_i$, as a function of forecast probability, y_i. Reliability can be though of as how well the *a priori* predicted probability forecast of an event coincides with the *a posteriori* observed frequency of the event. Curve of reliability diagrams conveys what the observed frequency was each time a given probability was forecasted and ideally lies along a 1:1 diagonal. Plot construction is as follows:

(1) Decide discrete number, I, of forecast values y_i (i.e., bins)

(2) Let N_i be the number of times each forecast y_i is used in the collection of forecasts being verified. The total number of forecast/event pairs is the sum of these subsample sizes (i.e., bin the data):

$$n = \sum_{i=1}^{I} N_i \tag{3.3}$$

(3) For each subsample delineated by the I allowable forecast values there is a relative frequency of occurrence of the forecast system. Computing the observed subsample relative frequency is,

$$\bar{o}_i = \frac{1}{N_i} \sum_{k \in N_i} o_k \tag{3.4}$$

where $o_k = 1$ if the event occurs for the kth forecast/event pair and $o_k = 0$ otherwise (i.e., computing conditional frequency in each bin).

(4) Plot observed frequency $\bar{o}_i$ versus forecast probability y_i

These plots are sensitive to the length of the observed record. Typically bin sizes are split by deciles, but as the record size of 30 years is small for this type of metric, bin size was increased to tercile. Reliability is ascertained by predicted probabilities agreeing with the observed frequencies. If the plotted values ($\bar{o}_i$ as a function of y_i) fall above the 1:1 line left of $y = 0.5$ and below to the right, it is considered an overconfident model. The opposite would be an underconfident model. Horizontal trends mean minimal resolution. Points falling entirely above or below indicate underforecasting and overforecasting bias, respectively. Erratic points can signify lack of skill or limitations of small sample size. Missing points mean it was a forecast of a rare event.

3.4.1.3 Significance Testing

The two metrics above are applied to each ESP technique separately, thus the results must be juxtaposed to allow for comparison. Significance testing allows for a more quantitative comparison of the two methods. In the case of testing two samples of forecast volumes, if the underlying distribution can be assumed as Gaussian, then significance can be tested with a *Student's t* test (referred to as '*t* test' henceforth). As there is no prior knowledge of the sample means, this would require a two-sided test where the null hypothesis is that the true difference in means is zero. In the event normality cannot be assumed, the nonparametric version is the Wilcoxon-Mann-Whitney rank-sum test. The motivation behind this technique is that outliers that would adversely affect the *t* test would have little or no influence. The nonparametric analog of mean is *location*, which is a function of rank within n observations. Therefore, the null hypothesis is that both samples have the same location.

Both tests are two-sided, assume unequal variance, and a test level, α, of 5% is applied. Both are also unpaired as WG ESP is a larger sample size than ESP by several factors (depending on the type of simulation, as clarified in later sections). A p-value of 0.05 or lower indicates rejection of the null hypothesis with 95% confidence.

3.4.2 Unconditional Forecasts

The first part of the evaluation was a simple visual comparison between reforecasts using ESP with observed climatology (denoted 'ESP') and those using ESP with the unconditional weather generator output ('WG ESP'). Examples are shown for BFFU1 in Figure 3.6 and DRGC2 in Figure 3.7. ESP comprises 30 traces (based on observed 6-hour precipitation and temperature sequences from 1981–2010) whereas WG ESP comprises 90 traces that are synthetically generated based on the statistical characteristics of the ESP climatology. In Figures 3.6 and 3.7, the red points show the observed refer to actual historical runoff volumes. Though only two gages are shown, two general characteristics from the visual comparisons emerge:

(1) For both methods (ESP and WG ESP) the boxplots tend to become tighter as initialization dates progress from November 1 to April 1. This streamflow prediction phenomenon is well-known in the western U.S., where the seasonality of the hydrologic cycle causes initial watershed conditions (snow and soil moisture) to dominate the forecast signal as the seasonal snow accumulation peak and dry summer season approach.

(2) At longer lead times, WG ESP produces more interannual variability in forecast anomalies and wider forecast spread, hence greater extremes.

These two characteristics will prove a reoccurring theme in the following results. This second characteristic becomes especially pertinent in the discussion of forecast reliability.

The next part of the analysis compares the first three moments (mean, standard deviation, and skew) of the WG ESP vs ESP ensemble forecasts. Figures 3.8 and 3.9 show April to July runoff moments for two of the representative gages. The three moments form the columns, while first of month leads times (November 1st to April 1st) form the rows. Results for the other two gages in are shown in Figures A.2 and A.3, as well as sub-seasons for the two gages displayed in this section (Figures A.4–A.7). Forecast ensemble means were nearly equivalent, with virtually 1:1 ratios, and standard deviations were also highly correlated. The WG ESP ensemble skews, however, do not perform as well in reproducing the ESP statistics. WG ESP skews were almost always positive, and

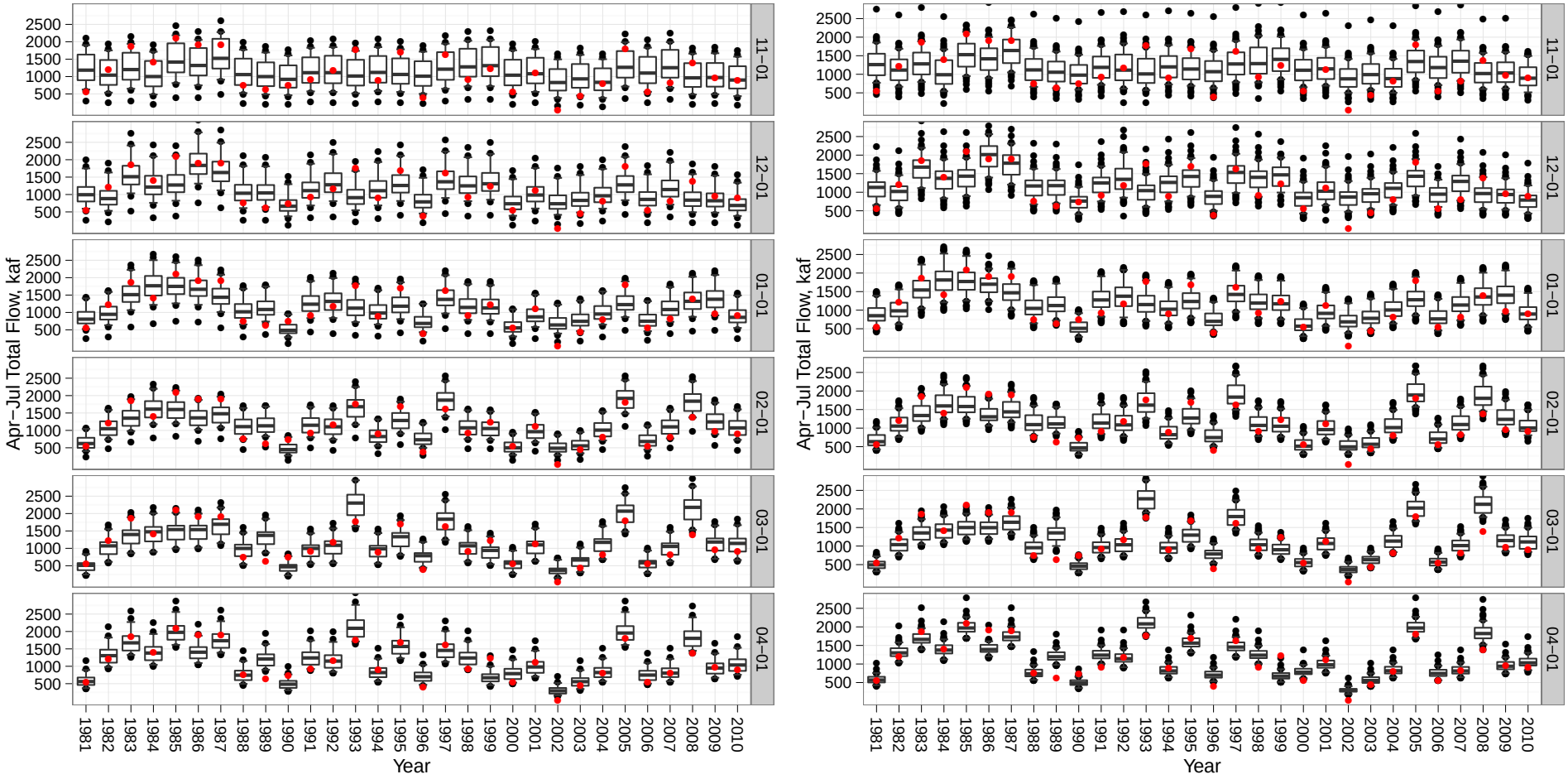
at long lead times tended to be larger than ESP skews. Occasionally ESP produces a negatively skewed ensemble. For BFFU1, skews would better fit a linear regression for June to July than April to May and in general for April to May this linear regression fit deteriorates as lead time decreases. DRGC exhibits similar results, and the standard deviation of the DRGC2 WG ESP forecasts have less agreement with ESP forecasts for April to May as well.

To compare the skill of the ensemble forecasts, yearly rank probability sum scores (RPSS) of predicted seasonal runoff were calculated for different combinations of lead time and ESP type for different gages, and their median and mean values are shown in Figure 3.10. As explained in Section 3.4.1.1, positive values represent skill over climatology. Unexpectedly, the median RPSS appears to be best for both methods on March 1st, not April 1st (when snowpack is more complete), though for average RPSS, March and April have very similar values. For medians, WG ESP outperforms standard ESP at later lead times while WG ESP consistently outperforms ESP on average values. Yearly RPSS differences are shown in Figure 3.11, where positive values denote improvement using WG ESP and the reverse for negative values. Changes mostly are of small magnitudes where peaks and valleys roughly correspond to abnormal years seen in Figure 3.5. The largest positive changes are seen at longer lead times, and largest negative for shorter, but the latter has smaller magnitudes of change than the former. Further investigation of RPSS can be seen in Figures A.12 and A.13. While not a strict linear trend, there is a pattern of higher RPSS values for high flow years.

While RPSS describes overall skill, reliability plots delve into more specific behavior and provide additional insight. Figures 3.12, 3.13, and 3.14 show reliability diagrams for the runoff season above the 10th, 50th, and 90th percentiles. Section 3.4.1.2 explains how to interpret this diagrams. Because of the limited historical record, bin sizes were reduced to terciles. Even with this conservative adjustment, often there are missing points. Across q10 and q90 thresholds, WG ESP has less missing points than standard ESP. This ties into the higher skew and increased variability noted at the beginning of this section. For q10, both methods are most accurate for higher probabilities, with middle probabilities being the most irregular and often lacking resolution. Lines for q50 have more resolution than q10, as well as there being less missing values. WG ESP

has a few occurrences of erratic behavior for mid-probabilities at late lead times. Both methods tend to overforecast, but ESP is slightly more underconfident. Both methods in general lack skill for q90, with a few exceptions (mostly in the lowest bin).

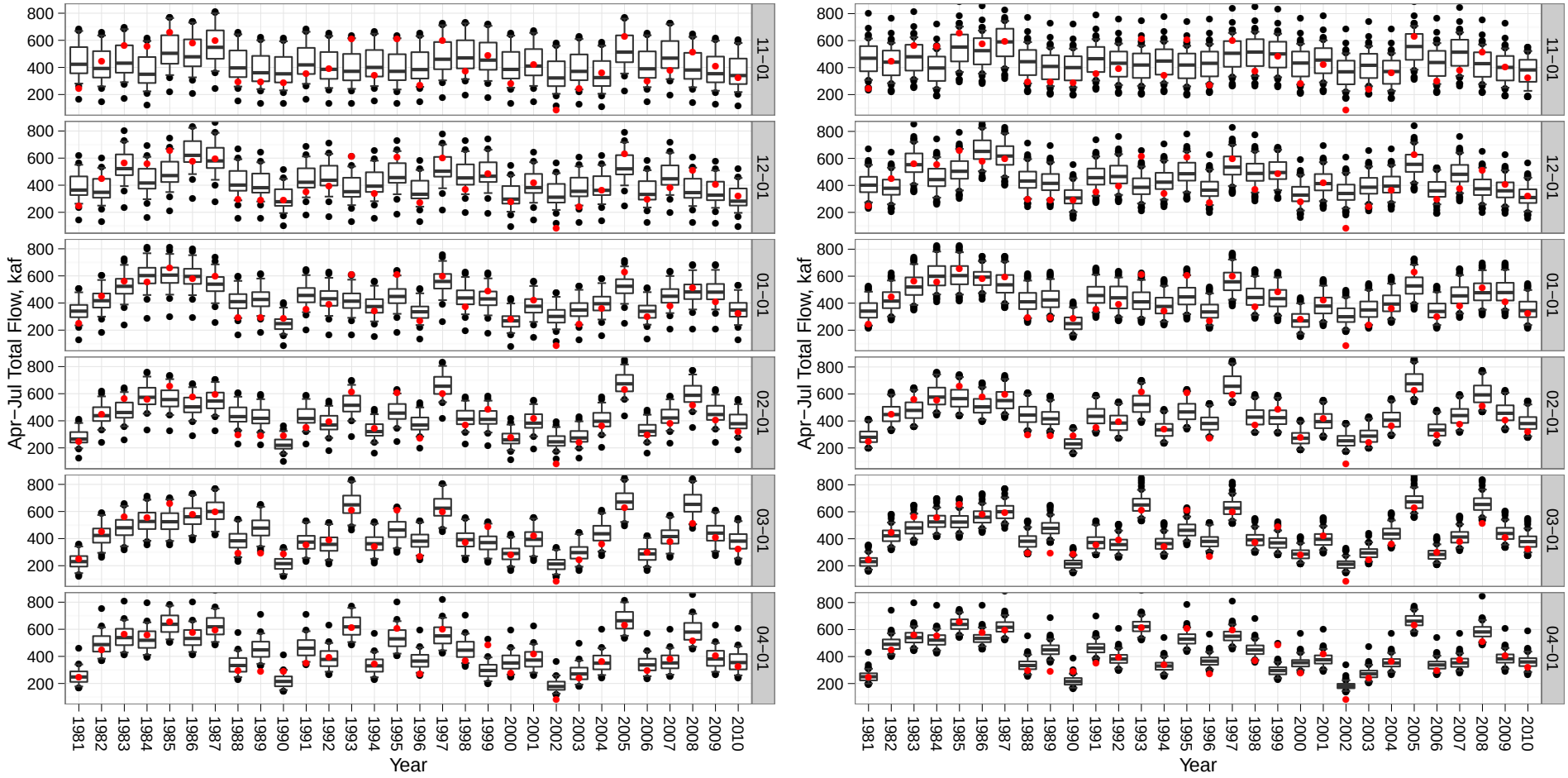
As outlined in Section 3.4.1.3, significance tests can indicate if two samples are statistically different. A quantile-quantile (QQ) plot (Figure A.21) indicates normality can be assumed for unconditional forecasts, though skewness appears in extreme values. Table 3.2 shows p-values from a two-sided t-test. November and December (except BFFU1 in December) are well below 0.05, thus is a confident rejection of the null hypothesis that ESP and WG ESP are the same. January and beyond have non-significant values (thus accepting the null) that become larger with each later month. This table supports the more qualitative assessments thus far — at long leads WG ESP produces variability unseen in ESP, but as lead times become shorter, the physical model forces convergence to similar values.



(a) 1981-2010 climatology as traces

(b) WG output as traces: 90 traces

Figure 3.6: BFFU1 ESP April to July reforecasts for Novemberto Aprillead times



(a) 1981-2010 climatology as traces

(b) WG output as traces: 90 traces

Figure 3.7: DRGC2 ESP April to July reforecasts for Novemberto Aprillead times

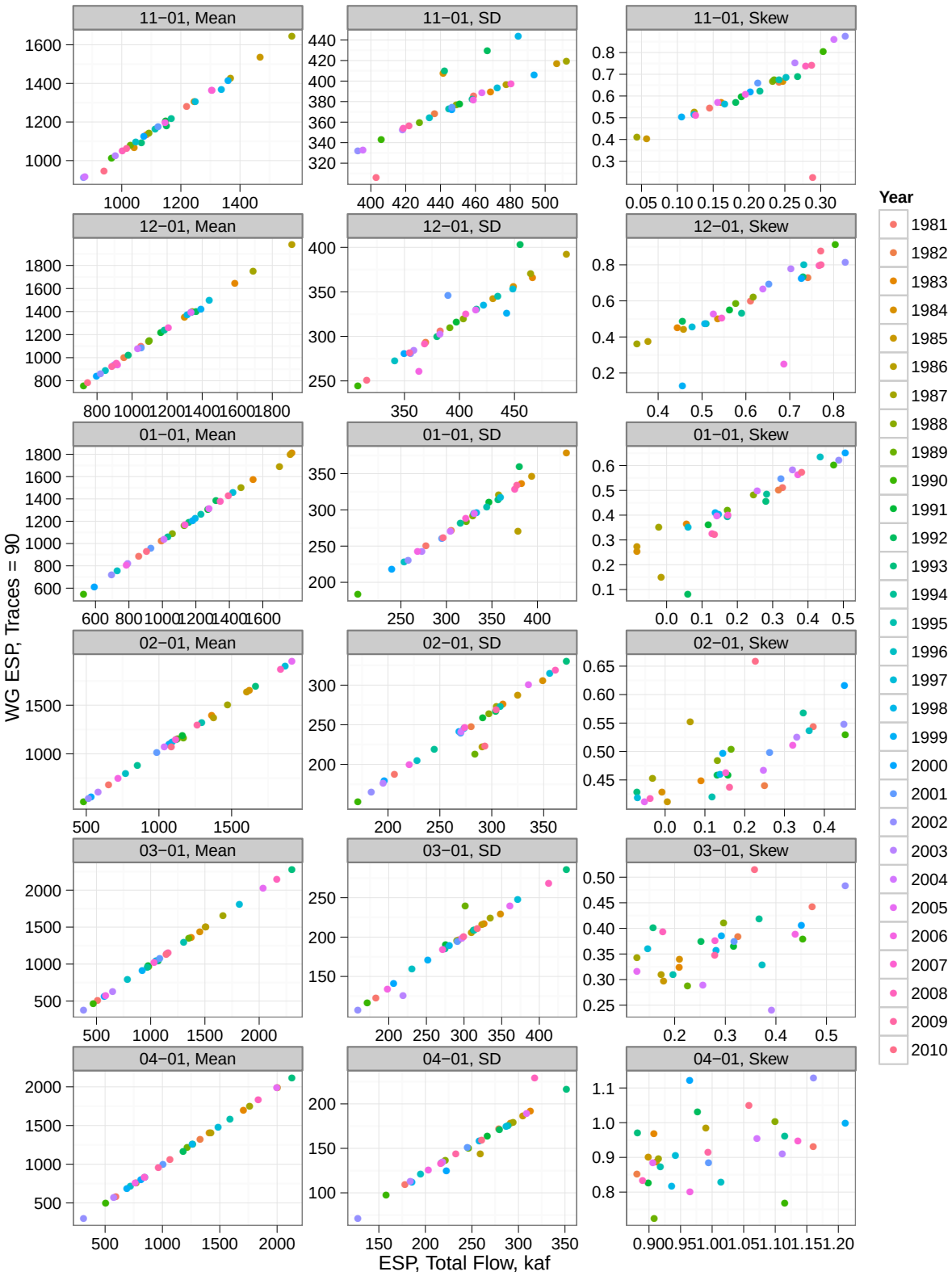


Figure 3.8: April to July BFFU1 WG ESP vs ESP moments

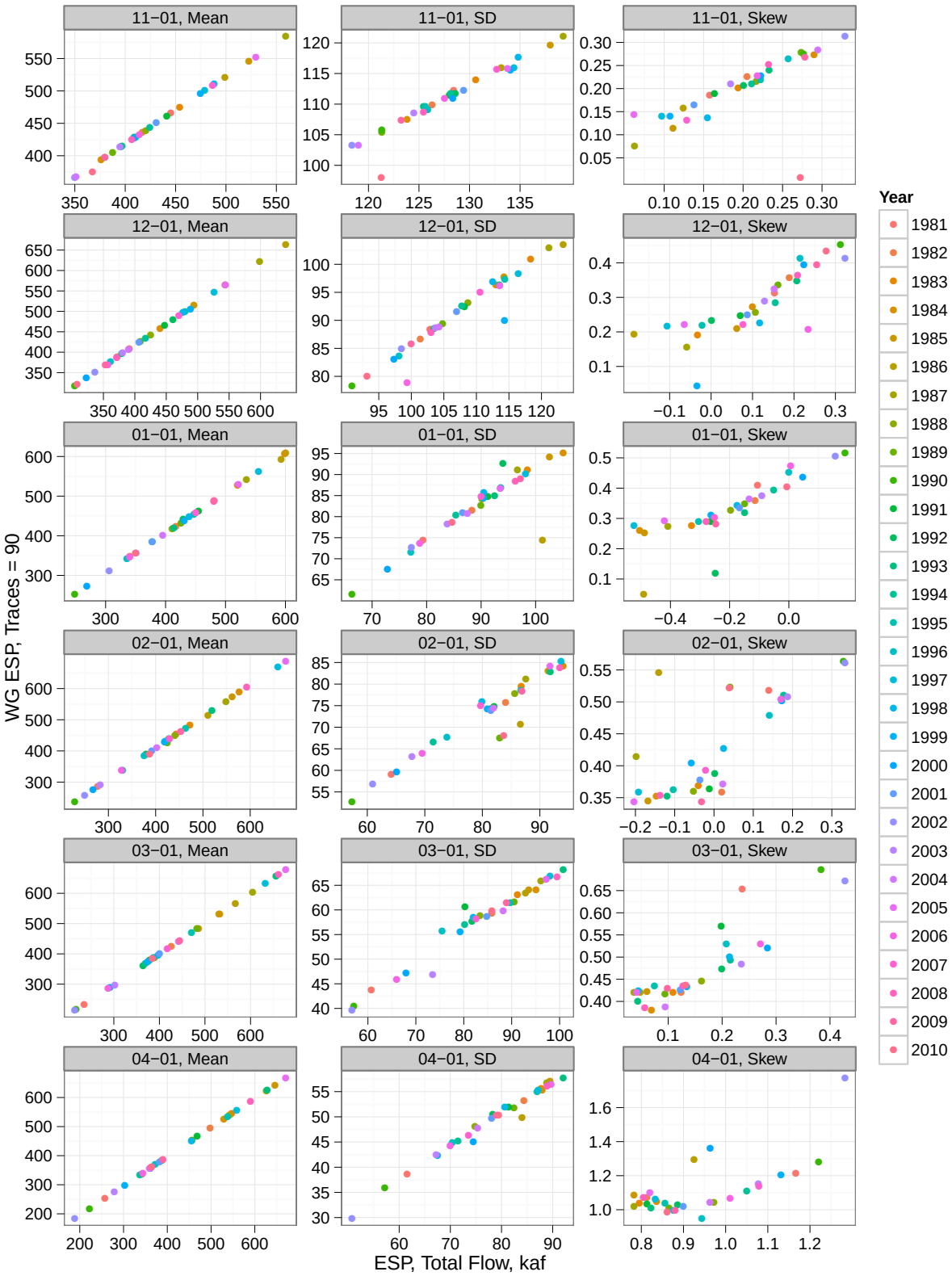


Figure 3.9: April to July DRGC2 WG ESP vs ESP moments

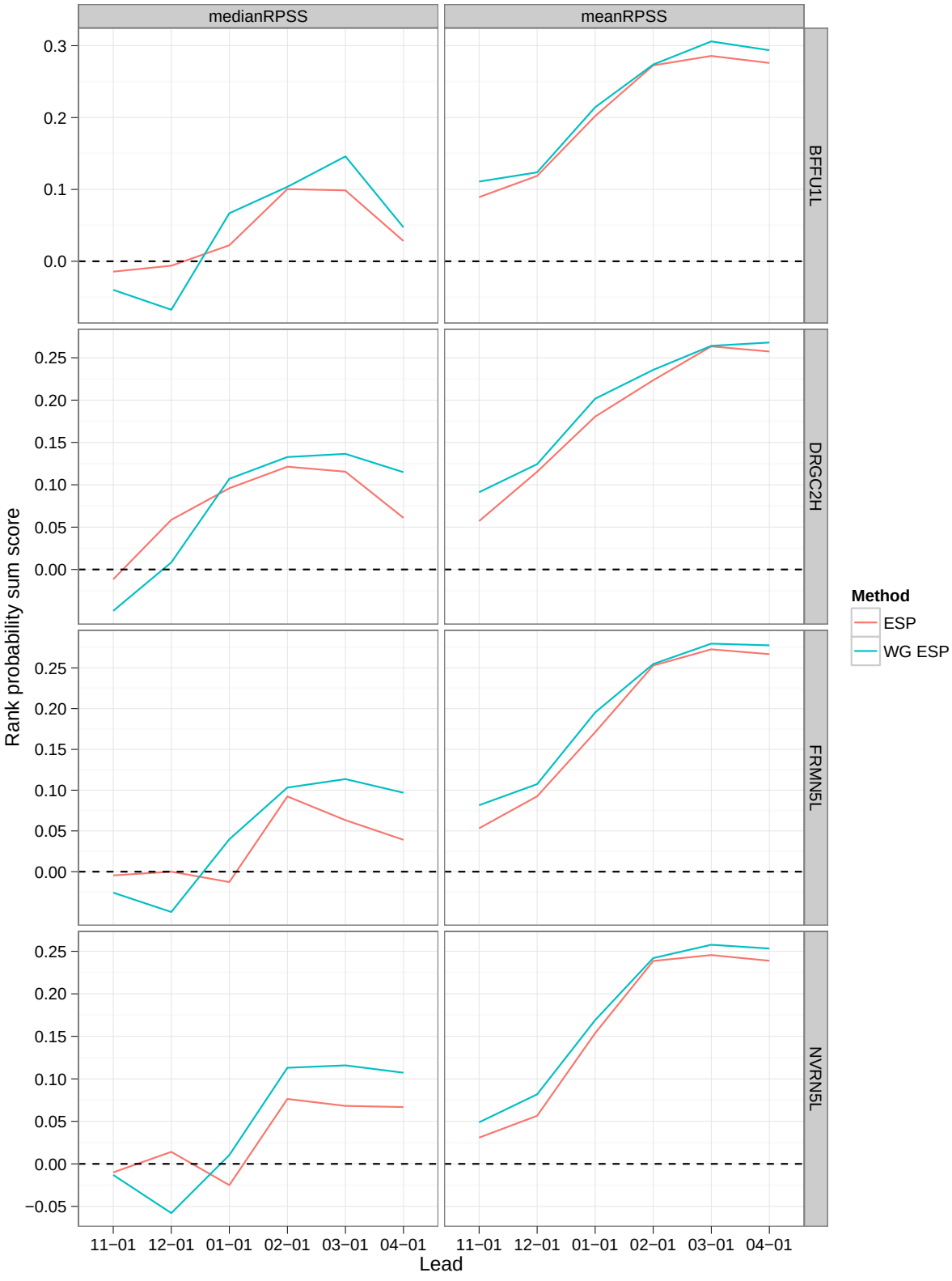


Figure 3.10: April to July median and mean RPSS

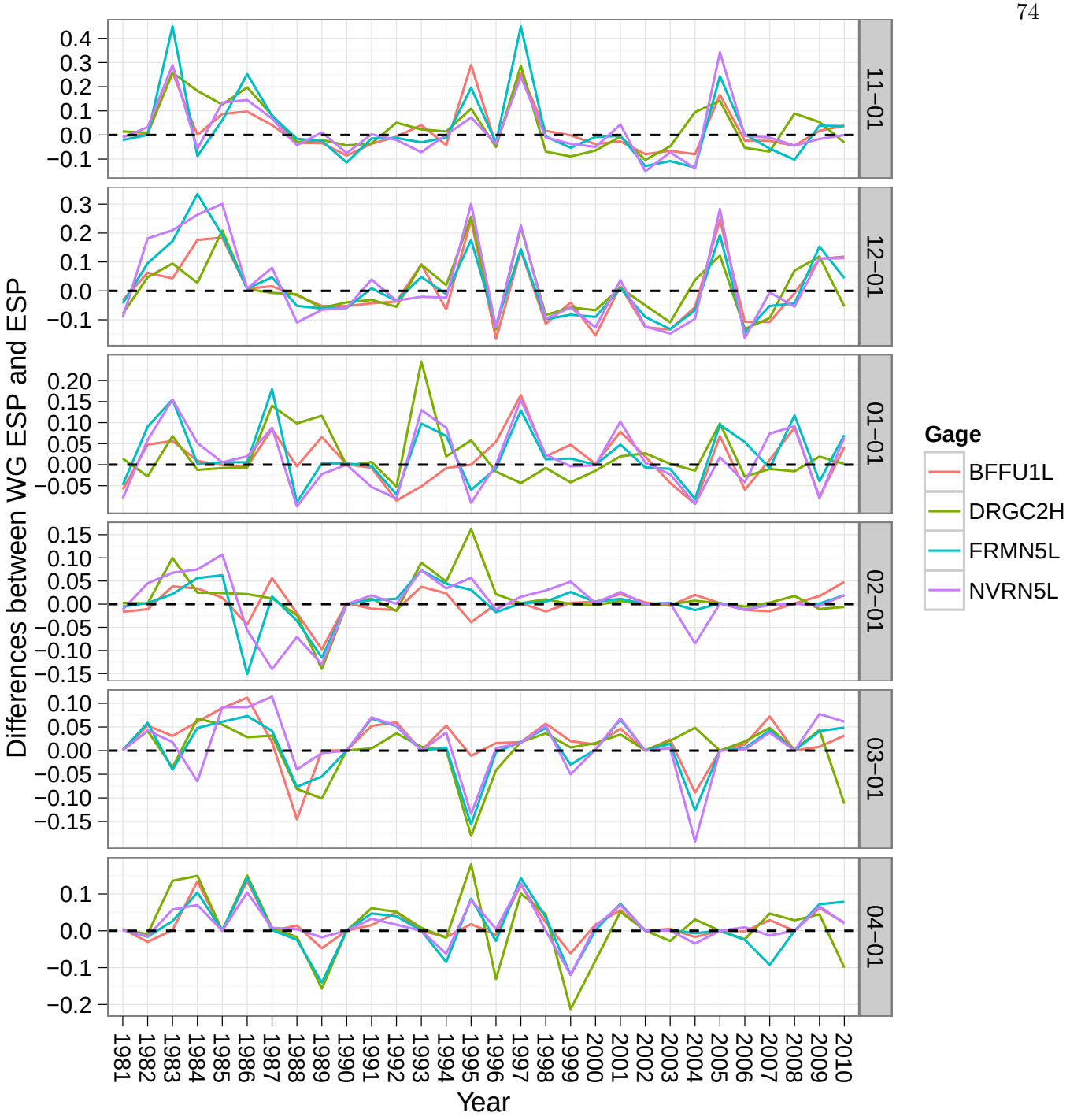


Figure 3.11: Yearly RPSS differences: positive values demonstrate improvement over ESP with WG ESP and vice versa for negative

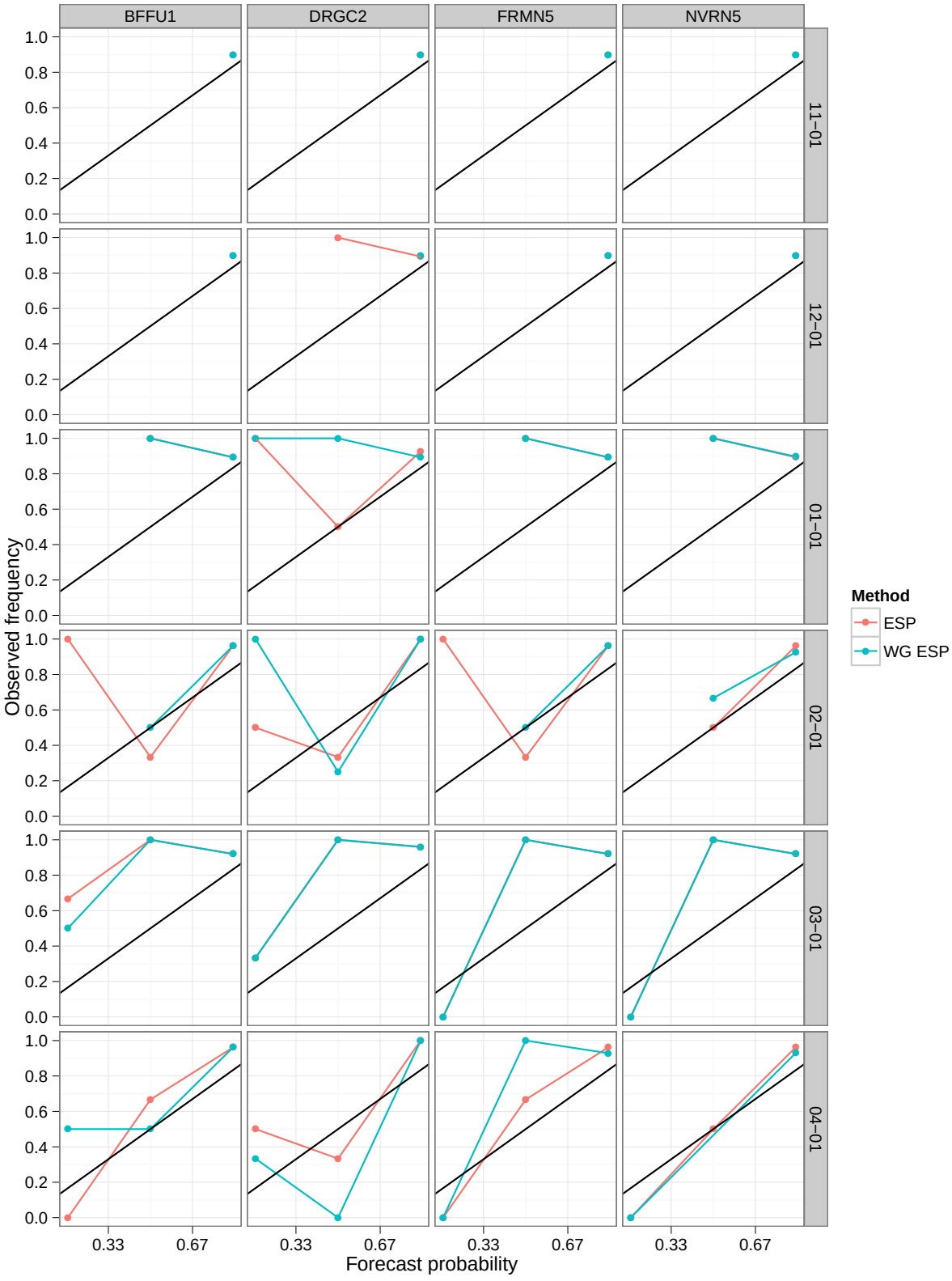


Figure 3.12: Reliability diagram of April to July runoff above 10th percentile

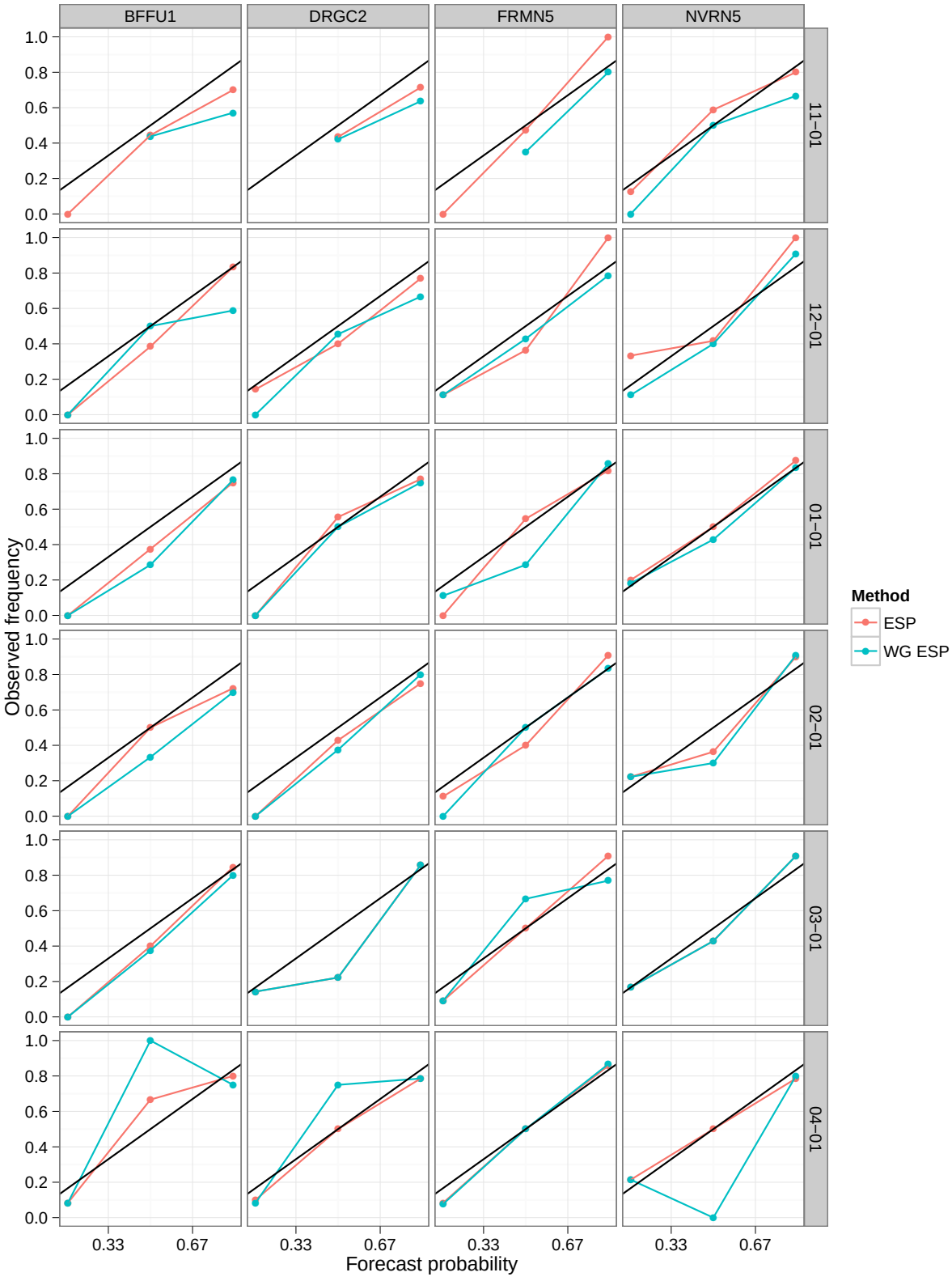


Figure 3.13: Reliability diagram of April to July runoff above 50th percentile

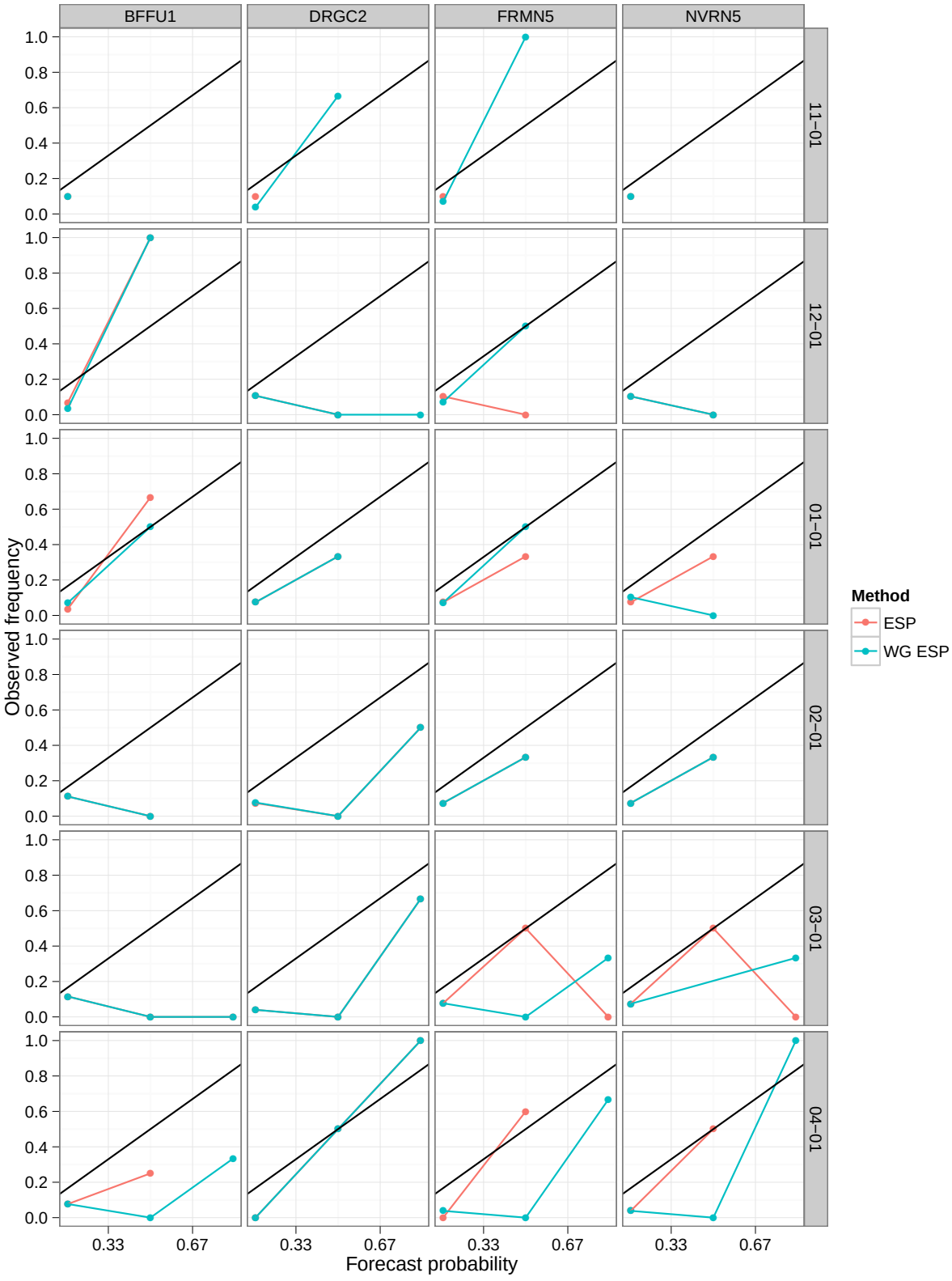


Figure 3.14: Reliability diagram of April to July runoff above 90th percentile

Table 3.2: P-values from two-sided t-tests of April to July runoff

	Gage	Lead	p.value
1	BFFU1L	11-01	0.0069
2	DRGC2H	11-01	0.0002
3	FRMN5L	11-01	0.0022
4	NVRN5L	11-01	0.0014
5	BFFU1L	12-01	0.0117
6	DRGC2H	12-01	0.0005
7	FRMN5L	12-01	0.0039
8	NVRN5L	12-01	0.0040
9	BFFU1L	01-01	0.0991
10	DRGC2H	01-01	0.1751
11	FRMN5L	01-01	0.0615
12	NVRN5L	01-01	0.0149
13	BFFU1L	02-01	0.1883
14	DRGC2H	02-01	0.0932
15	FRMN5L	02-01	0.1414
16	NVRN5L	02-01	0.1002
17	BFFU1L	03-01	0.6046
18	DRGC2H	03-01	0.8749
19	FRMN5L	03-01	0.6845
20	NVRN5L	03-01	0.7740
21	BFFU1L	04-01	0.6657
22	DRGC2H	04-01	0.5279
23	FRMN5L	04-01	0.6823
24	NVRN5L	04-01	0.8904

3.4.3 Conditional Forecasts

The prior sections demonstrated that the WG ESP was able to generate weather sequences that were sufficiently consistent with an observed meteorological climatology that they could be used to force a hydrologic model and achieve streamflow outputs that were consistent with model outputs driven by the original climatology. The weather generation approach provides an important capability, that of linking climate expectations to weather data at arbitrary space and time scales, and in particular at scales of interest for driving a hydrologic model. The remainder of this thesis illustrates the potential translation of climate forecasts (i.e., differing from the observed climatology) into weather inputs to a hydrology model and consequently to streamflow outputs. Due to data processing constraints, the approach could only be applied to a few years of forecasts, thus the results do not form the basis for a comprehensive skill assessment, but rather a demonstration that shows the potential value of the approach in the hydrologic forecasting context.

For hydrologic simulations conditioned on climate forecasts, we selected past climate forecasts based on observed hydrologic outcomes. Past climate forecasts were selected from the IRI website (discussed in Section 2.3.3.1). Options for wet and dry years were limited to the late 1990's and the 2000's, the period covered by the IRI forecasts. Figure 3.5 was used to determine what constituted a wet or dry year, and 2005 was selected for wet conditions and 2006 for dry. Sub-season flows (Figure A.1) for these years depict April–May being much higher than June–July for 2006. For 2005, June-July has more of a deficit with respect to the average than April–May. The IRI probability forecasts were rather conservative, being either 40:35:25 for the wet winter or 25:35:40 for the dry winter (forecast and verification maps shown by Figures B.1 and B.2 in Appendix B). Additionally, these forecasts are issued in three month segments, so overlap will occur for longer seasons. 2005 and 2006 were selected as well for having little change between different forecast periods. To start with, the weather generator was run with these probabilities for November to March. The corresponding changes in distributional statistics for the weather variables are shown by Figures B.3 and B.4.

Figures 3.15 and 3.16 have probability density functions (PDFs) for the two selected con-

ditional years. We added an additional PDF where the year in particular being forecasted is not included as a trace. In Appendix A, Figures A.28 and A.29 more clearly show the difference caused by leaving the year out (they are Figures 3.15 and 3.16 not categorized by lead time). At long lead times for 2005, WG ESP's distribution falls closer to the observed value. As the lead time becomes shorter, ESP "catches up" and all methods become aligned with observed outcomes. With 2006, again WG ESP stands out at long lead times (DRGC2 in particular), and all methods fall closer to the observed line at short leads. However, with a dry year there is a distinct lower bound in predicted flow. With a wet year, a forecast can benefit by shifting the entire distribution across a relatively larger range, but for a dry year, increased skill must come from skewing the distribution toward high probabilities of low flow values and making the upper distribution tail as thin as possible. In 2005, by late winter the area under the PDF is evenly split by the observed value. In 2006, while the bell of the PDF becomes split, splitting of the density is more variable. Interestingly, March appears to have more skillful PDFs than April. In dry years, it is not unusual for snowpack to have started melting before April 1st, diminishing the value of SWE as a predictor as the snowpack declines.

Figure 3.17 has RPSS for the two years. Now, unconditional WG ESP from Section 3.4.2 is included in the event of a bias and WG ESP conditioned on climate forecasts is now referred to as 'condWG ESP'. The unconditional RPSS for these two years highlighted in Figures A.12 and A.13 offer additional insight into the inherent behavior before conditioning. In 2005, condWG ESP consistently out-performs ESP until February, after which both have similar skill. WG ESP falls in between the two, demonstrating that skill exists without climate conditioning. In 2006, there is somewhat contradictory behavior compared to the wet year PDFs. The condWG ESP had closer peaks to the observed line at long leads. Here, ESP performs better at long leads, and January and later condWG ESPs overtake it. WG ESP tends to have the worst performance, indicating that condWG ESP may perform better if it did not have to overcome an inherent bias. The tail behavior in the PDFs likely plays a role in this behavior and may be the reason both methods have worse skill than climatology for all months. Further analysis could include decomposing RPSS into

components by delving into bias, resolution, or uncertainty.

QQ plots of the conditional years (Figures A.32 and A.35) have non-Gaussian behavior and thus motivate non-parametric significance testing (discussed in Section 3.4.1.3). Table 3.3 shows p-values from Wilcoxon rank sum test for the two years. Tables are quantitative summary of what is seen in previous figures. Fall and early winter forecasting for a wet year has significant differences between ESP and WG ESP. For the dry year, while the values are low early on, they are not significantly low.

Figures 3.18 (2005) and 3.19 (2006) show exceedance probabilities above the 10th, 50th, and 90th percentiles. A forecast would be considered skillful if the exceedance probabilities are appropriately shifted with respect to the quantiles. For instance for q10, a wet forecast should have probabilities above 90%, and likewise below 90% for dry. In 2005, this metric echoes the conclusions from above where condWG ESP has the best performance, aided by the already-skillful WG ESP. In 2006, condWG ESP mostly performs better than the ESP, with exceptions occurring in November/December for q10 and q90. Ordinary WG ESP clearly performs the worst for q10, but has more mixed behavior for q50 and q90.

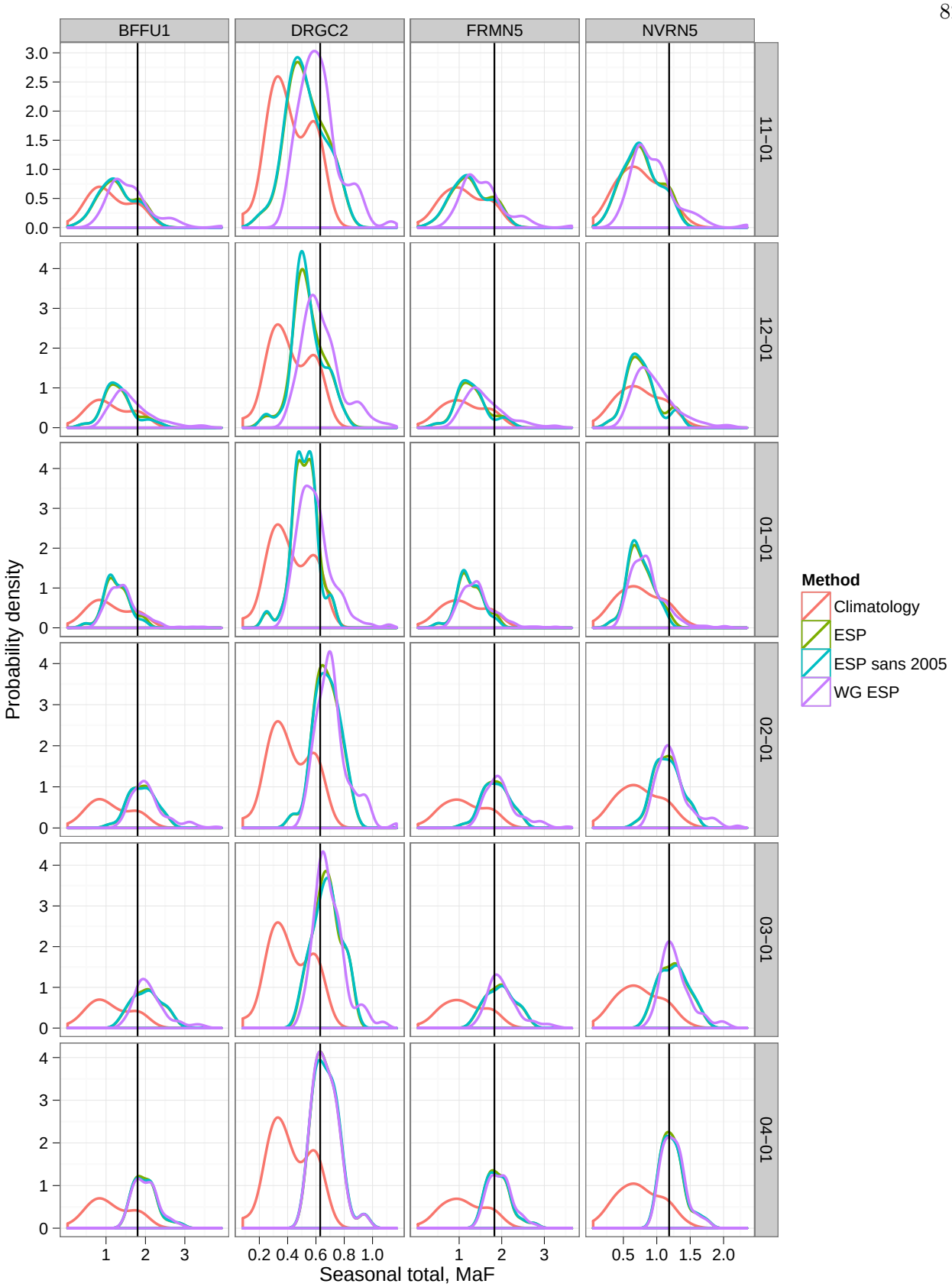


Figure 3.15: PDFs for 2005 runoff with vertical line showing 2005 value

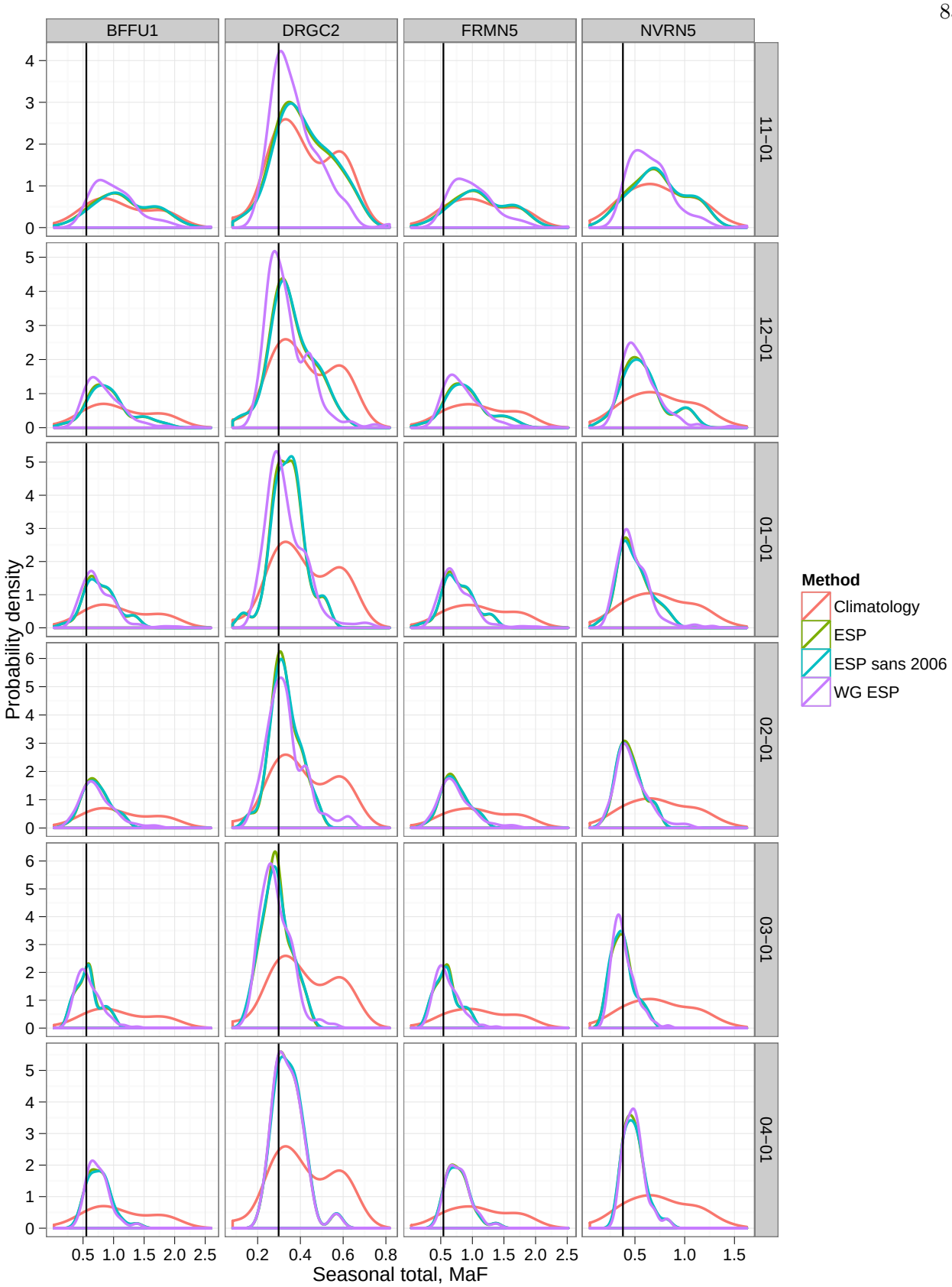
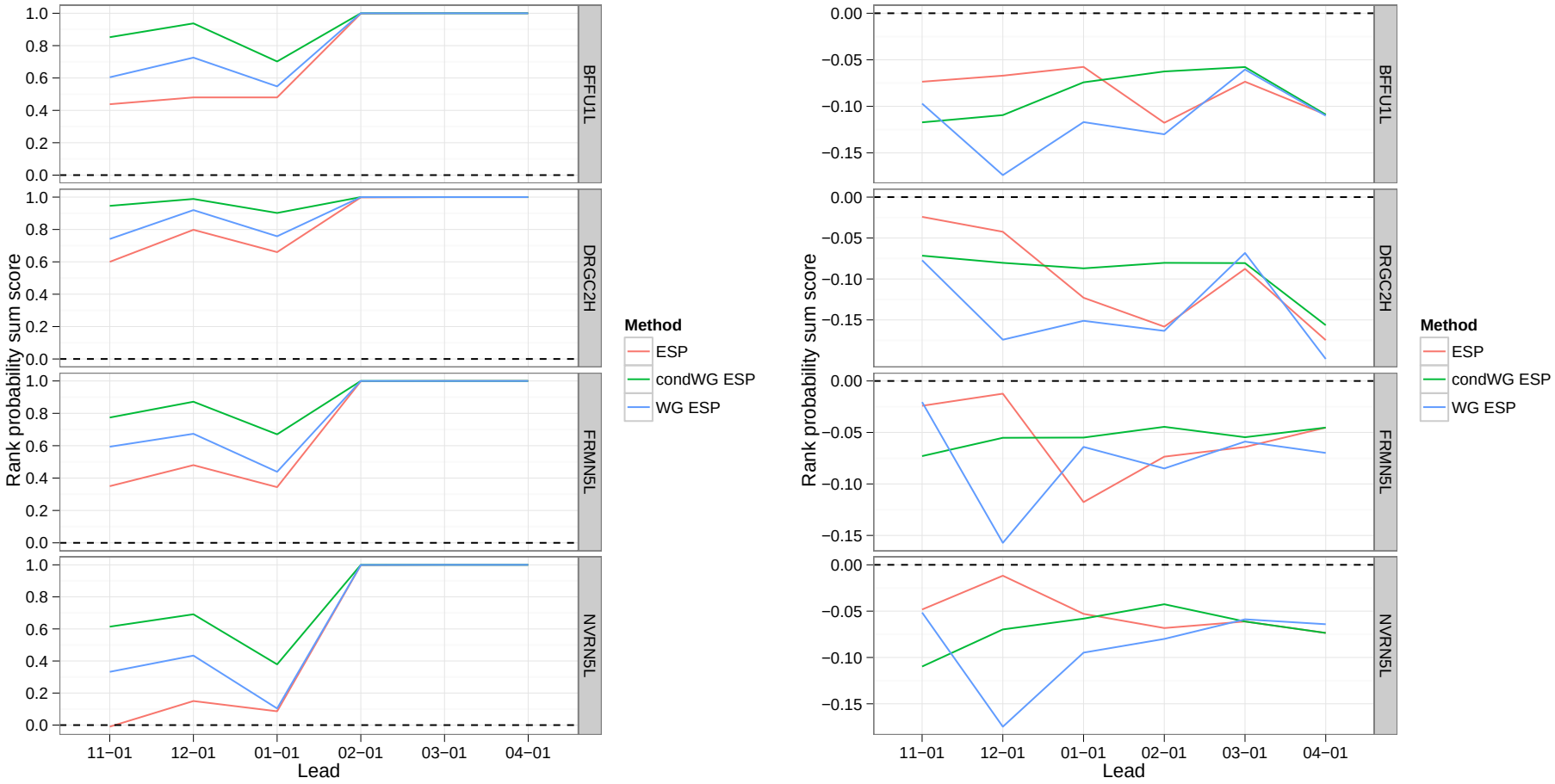


Figure 3.16: Same as Figure 3.15, but for 2006



(a) 2005

(b) 2006

Figure 3.17: RPSS for the conditional years

Table 3.3: P-values from Wilcoxon rank sum test for 2005 and 2006 April to July runoff

	Gage	Lead	2005	2006
1	BFFU1L	11-01	0.004	0.133
2	DRGC2H	11-01	0.001	0.079
3	FRMN5L	11-01	0.004	0.130
4	NVRN5L	11-01	0.005	0.222
5	BFFU1L	12-01	0.001	0.272
6	DRGC2H	12-01	0.000	0.110
7	FRMN5L	12-01	0.001	0.324
8	NVRN5L	12-01	0.002	0.629
9	BFFU1L	01-01	0.052	0.436
10	DRGC2H	01-01	0.015	0.171
11	FRMN5L	01-01	0.055	0.482
12	NVRN5L	01-01	0.109	0.710
13	BFFU1L	02-01	0.285	0.772
14	DRGC2H	02-01	0.292	0.619
15	FRMN5L	02-01	0.271	0.825
16	NVRN5L	02-01	0.257	0.937
17	BFFU1L	03-01	0.796	0.819
18	DRGC2H	03-01	0.946	0.540
19	FRMN5L	03-01	0.819	0.883
20	NVRN5L	03-01	0.790	0.879
21	BFFU1L	04-01	0.849	0.849
22	DRGC2H	04-01	0.760	0.775
23	FRMN5L	04-01	0.876	0.879
24	NVRN5L	04-01	0.849	0.879

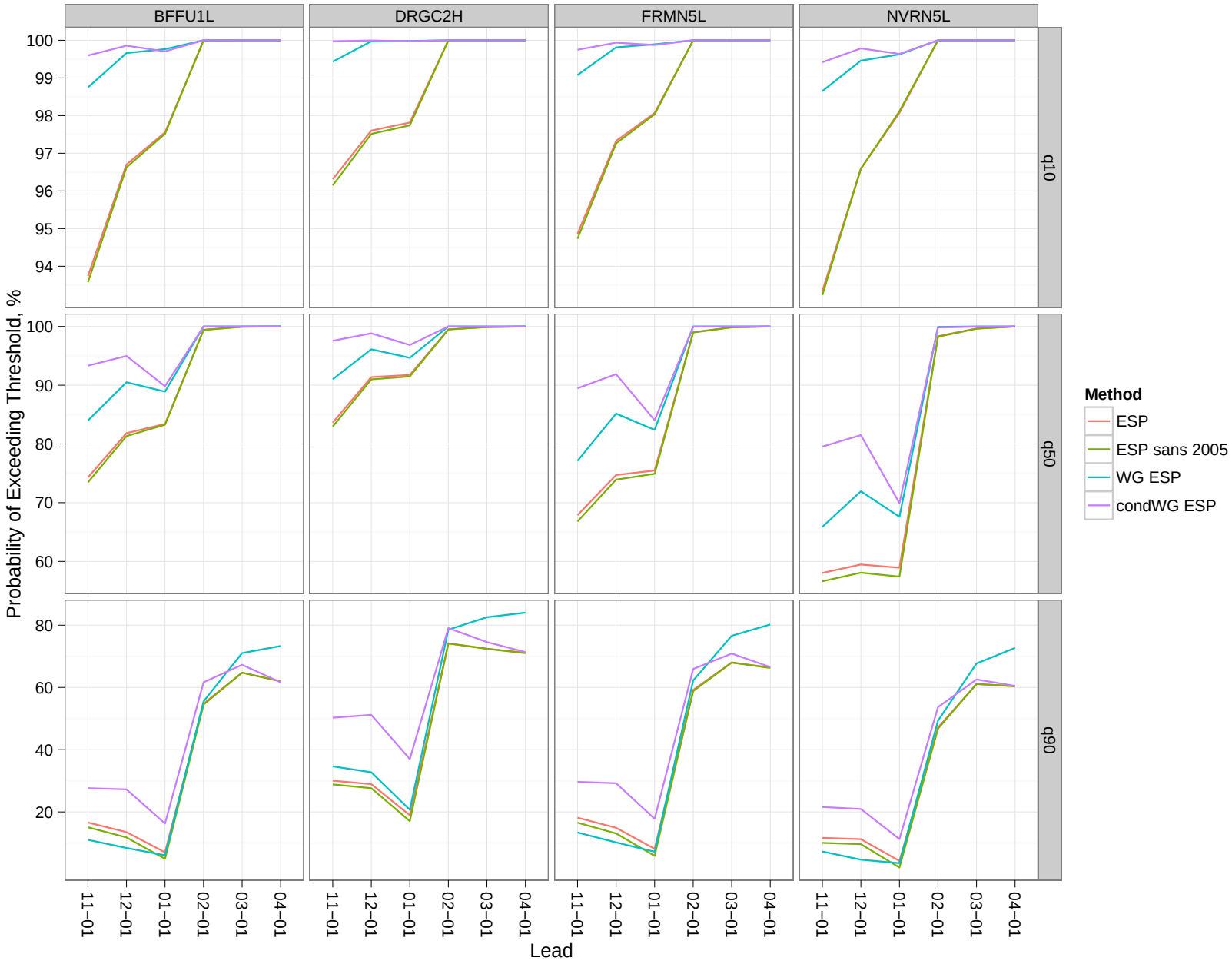


Figure 3.18: Shifts in 2005 exceedance probabilities

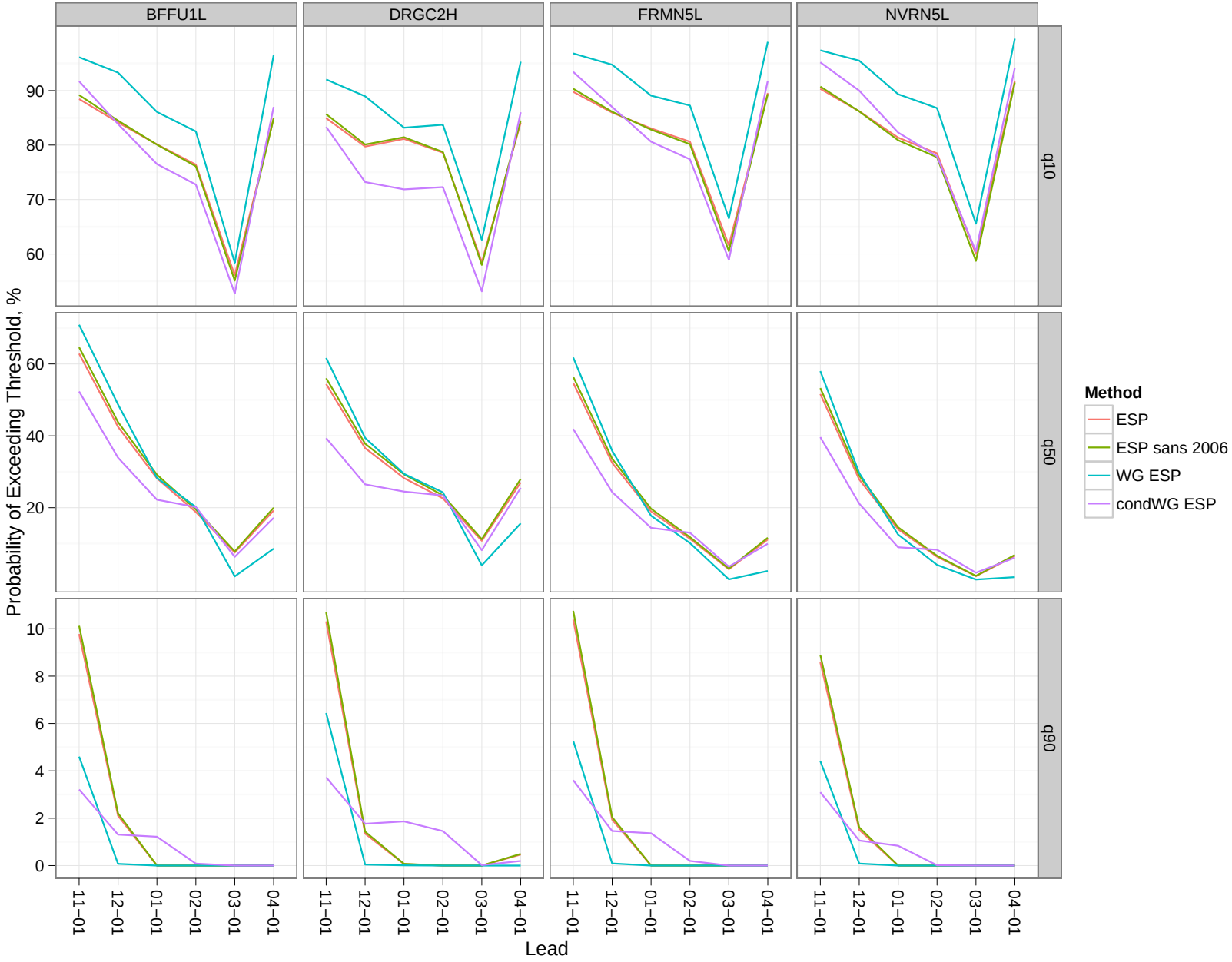


Figure 3.19: Same as Figure 3.18, but for 2006

3.5 Summary and Discussion

We have developed an integrated framework which offers significant potential to enhance streamflow forecasting. Our stochastic weather generator proves a flexible alternative to simply using the observed meteorological climatology as the basis for ensemble traces, as is the case with traditional ESP. This framework behaves similarly to ESP and fits well inside the NWS architecture, but also offers the capability of expanding the ensemble size, adding stochasticity and incorporating climate forecast conditioning. For the San Juan Basin, we performed reforecasts of April to July runoff for 1981 to 2010 and compared traditional ESP with our proposed WG ESP. For unconditional reforecasts, WG ESP adds higher variability and more extremes, especially at November/December lead times. As expected, both ESP and unconditional WG ESP better predict observed streamflow values, with lesser bias and reduced spread, as lead time shortened. This convergence can be attributed to the increasing influence of moisture conditions in the watershed and is noted in forecasting literature (Franz et al., 2003; Grantz et al., 2005; Wood and Lettenmaier, 2008). We also found that April 1st forecasts tended to have the best skill, but not always compared to March 1st. Franz et al. (2003) found forecast reliability peaks at March 15th then appears to become worse, suggesting that sub-monthly lead times would help resolve the skill fluctuations temporally, beyond the findings of this study.

ESP has limitations in that ensemble members must be pre- or post-processed to account for climate forecasts and despite this is still limited to the length and values of the historical record. For unconditional reforecasts WG ESP had 90 traces, which was expanded to 150 for conditional reforecasts, but could easily have been expanded to a larger ensemble size. When using IRI climate forecasts to condition the weather generator, we found that the representative wet year, 2005, was better predicted with conditional WG ESP than with ESP. Performance varied for the dry year of 2006 and differences between ESP and WG ESP were not found to be significant. The absolute differences in flow that separate a dry year from a normal one are small than those that separate a wet year from a normal one, thus the dry year predictions may be more difficult to predict (given

other influences on streamflow, such as the timing and interaction of weather sequences) than wet years. Climatology is a skillful competitor, hence the overall negative RPSS for both methods in 2006. The IRI forecast for 2006 was also very conservative. It is possible more severe ANB would be warranted to see a true shift in the distribution. Skill in wet forecasts have more importance management-wise, especially in arid regions, on seasonal time scales. As discussed in Grantz (2006), operations on the Pecos River did not change between normal and dry years. However, with skillful wet forecasts, additional runoff can be managed appropriately for uses such as agriculture, and unwanted spills to Texas could be minimized.

Additional insights could be found from analyzing daily flow. From such a timeseries, one could identify patterns of timing and peak for seasonal (or annual) hydrographs. Other variables of interest include reservoir values, such as storage, or physical conditions, such as soil moisture. As mentioned earlier, we performed forecasts using only the 'da' methodology for the weather generator. Exploring forecasts derived from the 'caJoint' approach may offer a new variety in hydrologic information.

Future meteorology provides the primary source of uncertainty in ESP forecasts at long lead times (Day et al., 1992). Possible developments could include exploring the uncertainty in initial conditions as Wood and Lettenmaier (2008) suggest a trade-off exists between initial conditions versus improved climate forecast accuracy. Linking these two could provide an "ensembles of ensembles".

Conditional streamflow forecasts may require more elaborate approaches than conditioning winter precipitation on climate forecasts. Pagano and Garen (2005), among others, suggest there is not a predictable signal between ENSO and the CRB above Lake Powell. For example, El Niño favors cold April-Junee conditions, thus a forecast based on snowpack alone would underestimate observed flows. This is attributed to increased runoff efficiency during cold springs. Wood and Werner (2011) also discuss effects of precipitation anomalies in winter and temperature anomalies in spring. Developments for the weather generator will include conditional temperature simulations for the purpose of modeling spring/early summer. As outlined previously, it currently performs

preferential resampling based on ranked seasonal totals. For temperature, the same can be performed for ranked seasonal temperature averages. Conditional temperature simulations may be of particular significance as the western U.S. has seen trends in warming and early timing of peak snowmelt since the 1980's (see work of McCabe and Wolock (2009) and sources referenced therein).

Lastly, future work can involve the combination of multiple models to improve forecast skill. A multi-model combination can be developed using a Bayesian framework, and could include forecasts such as from Bracken et al. (2010). Combining statistical and physical models for streamflow prediction may hold great promise for water resources management.

Chapter 4

Conclusion

4.1 Summary and Conclusions

The research presented in this thesis demonstrated the applicability of a stochastic weather generation approach in producing weather sequences specific to 66 subcatchments in the San Juan Basin. We explored four multisite resampling methods. The first, domain-aggregate (da), resamples all locations simultaneously. The diverse nature of the basin provided a motivation for testing three clustering-based approaches in which a clustering of subcatchments was performed on seasonal precipitation totals for the purpose of deriving generation parameters that are responsive to spatio-temporal climate variability. Of these, the method 'caJoint', compares favorably to the 'da' method, which used Markov transition probabilities to model wet/dry states over the three cluster system. Both 'da' and 'caJoint' reasonably capture distributional and higher-order climatological statistics. Testing wet and dry probabilistic 3-category above-normal-below (ANB) forecasts yielded expected shifts in distributions.

The 'da'-produced weather generator results were used to provide forecast weather sequences to the hydrologic SAC-SMA model in an ESP framework within CHPS. Forecast validation involved running reforecasts of the 1981-2010 period. We found that the WG ESP created heightened variability at long lead times and produce more extreme values. We saw this manifest in various measures, such as reliability plots and dissimilarity of skew between WG ESP and ESP. At shorter lead times, we confirmed the common wisdom that initial conditions (e.g., soil moisture, snowpack) influence forecasts more than future climatic uncertainty, causing smaller differences between WG

ESP and ESP.

By matching IRI climate forecasts with historical flow years, we selected 2005 and 2006 as wet and dry years, respectively. The IRI forecast probabilities for precipitation were then used to condition the weather generator. We found that assigning a conservative wet forecast yields significant streamflow forecast skill, especially at long lead times (e.g., November/December initializations for April–July runoff predictions). However, the conservative dry forecast did not lead to significant improvements in skill relative to ESP. The absolute differences in flow that separate a dry year from a normal one are small than those that separate a wet year from a normal one, thus the dry year predictions may be more difficult to predict. Additionally, the ANB forecast may not be strong enough to cause a significant shift in precipitation and the resulting runoff. As the overlap between IRI forecasts and the ESP forcing record was limited to about a decade, this part of the study was limited to a two year sample. While this is not enough to demonstrate real "skill", this work is promising as a downscaling link between climate forecasts and streamflow forecasts. The weather generator can make use of any probabilistic climate forecast, thus is a general method for use by the community.

4.2 Recommendations for Future Work

Proponents of parametric weather generators typically criticize resampling-based ones for not producing values unseen in the sampling record. While our weather generator does not produce new precipitation and temperature, it does produce an unconstrained variety in sequences and spells. Leander and Buishand (2009) created a two-stage resampling algorithm to ensure that new values could be produced and ultimately found it did not affect the range of streamflow when applied to a hydrologic model. Extreme totals were found to be caused from a sequence of moderately large daily precipitation values rather than one large daily event. However, it may be worth investigating hybrid methods that integrate semi-parametric sampling methods with the use of an extreme value distribution to extrapolate beyond historical weather ranges.

Improvement in streamflow forecasts can directly aid operational and planning decisions in de-

cision support systems (DSSs). DSSs are crucial for systems with complex operations and multiple objectives, which easily applies to reservoirs such as Navajo. Elsewhere, Grantz (2006) demonstrated skillful wet forecasts can benefit operations on the Pecos River by reducing unwanted spills to Texas. There is great potential for assessing ensemble forecast-influenced reservoir operations. For short-term ensemble forecasts (using weather forecasts) Boucher et al. (2012) demonstrated gains in terms of economic value, such as electricity production, reduction of spillage, and prevention of inundations. Similar work for long-term forecasts holds great promise for not only the San Juan, but watersheds within and outside of the Colorado River Basin.

Bibliography

Anderson, E. A. (1973). National Weather Service River Forecast System—Snow Accumulation and Ablation Model. Technical report, NOAA Tech. Memo. NWS Hydro-17, U.S. National Weather Service.

Apipattanavis, S., F. Bert, G. Podestá, and B. Rajagopalan (2010). Linking weather generators and crop models for assessment of climate forecast outcomes. Agricultural and Forest Meteorology 150(2), 166–174.

Apipattanavis, S., G. Podesta, and B. Rajagopalan (2007). A semiparametric multivariate and multisite weather generator. Water Resources Research 43(11), W11401.

Baigorria, G. a. and J. W. Jones (2010). Gist: A stochastic model for generating spatially and temporally correlated daily rainfall data. Journal of Climate 23(22), 5990–6008.

Beersma, J. J. and T. A. Buishand (2003). Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. Climate Research 25, 121–133.

Boucher, M.-a., D. Tremblay, L. Delorme, L. Perreault, and F. Anctil (2012, January). Hydro-economic assessment of hydrological forecasting systems. Journal of Hydrology 416-417, 133–144.

Bracken, C., B. Rajagopalan, and J. Prairie (2010). A multisite seasonal ensemble streamflow forecasting technique. Water Resources Research 46(3), W03532.

Bracken, C. W. (2011). Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management. Ph. D. thesis, University of Colorado at Boulder.

Brandsma, T. and A. T. Buishand (1998). Simulation of extreme precipitation in the rhine basin by nearest-neighbor resampling. Hydrology & Earth System Sciences 2(2-3), 195–209.

Brissette, F. P., M. Khalili, and R. Leconte (2007). Efficient stochastic generation of multi-site synthetic precipitation data. Journal of Hydrology 345(3-4), 121–133.

Buishand, T. (1978). Some remarks on the use of daily rainfall models. Journal of Hydrology 36(3-4), 295–308.

Buishand, T. A. and T. Brandsma (2001). Multisite simulation of daily precipitation and temperature in the rhine basin by nearest-neighbor resampling distribution. Water Resources Research 37(11), 2761–2776.

Burnash, R. J. C. (1995). The NWS River Forecast System - Catchment modeling. In V. P. Singh (Ed.), Computer Models of Watershed Hydrology, pp. 311–366. Water Resources Publications.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973). A generalized streamflow simulation system—Conceptual modeling for digital computers. Technical report, Joint Federal and State River Forecast Center, U.S. National Weather Service and California Department of Water Resources.

Caron, A., R. Leconte, and F. Brissette (2008). An improved stochastic weather generator for hydrological impact studies. Canadian Water Resources Journal 33(3), 233–255.

Chandler, R. E. (2005). On the use of generalized linear models for interpreting climate variability. Environmetrics 16(7), 699–715.

Chandler, R. E. and H. S. Wheater (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of ireland. Water Resources Research 38(10), 674– 689.

Cloke, H. L. and F. Pappenberger (2009). Ensemble flood forecasting: A review. Journal of Hydrology 375(3-4), 613–626.

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. Journal of Water Resources Planning and Management 111(2), 157–170.

Day, G. N., L. E. Brazil, C. S. McCarthy, and D. P. Laurine (1992). Verification of the national weather service extended streamflow prediction procedure. In Proc. AWRA 28th Annual Conf. and Symp., Reno, NV, pp. 163–172. American Water Resources Association.

Eberle, M., H. Buiteveld, J. Beersma, P. Krahe, and K. Wilke (2002). Estimation of extreme floods in the river rhine basin by combining precipitation-runoff modelling and a rainfall generator. In M. Spreafico and R. Weingartner (Eds.), Proceedings International Conference on Flood Estimation, Berne, Switzerland, pp. 459–467. CHR Report 11-17, Lelystad, CHR Secretariat, the Netherlands.

Eum, H. I., S. P. Simonovic, and Y. O. Kim (2010). Climate Change Impact Assessment Using K-Nearest Neighbor Weather Generator: Case Study of the Nakdong River Basin in Korea. Journal of Hydrologic Engineering 15(10), 772–785.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. Biometrics 35(1), 169.

Foufoula-Georgiou, E. and K. P. Georgakakos (1991). Recent Advances in the Modeling of Hydrologic Systems, Chapter Hydrologic advances in space-time precipitation modeling and forecasting, pp. 47–65. New York: Springer.

Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales (2003). Verification of national weather service ensemble streamflow predictions for water supply forecasting in the colorado river basin. 4, 1105–1118.

Friend, A. D., A. K. Stevens, R. G. Knox, and M. G. R. Cannell (1997). A process-based terrestrial biosphere model of ecosystem dynamics. Ecological Modelling 95, 249–87.

Furrer, E. M. and R. W. Katz (2007). Generalized linear modeling approach to stochastic weather generators. Climate Dynamics 34(2), 129–144.

Furrer, E. M. and R. W. Katz (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resources Research 44(12), W12439.

Garen, D. C. (1992). Improved techniques in regression-based streamflow volume forecasting. Journal of Water Resources Planning and Management 118, 654–670.

Grantz, K. (2011, March 21–22). Reclamation mid-term operational modeling. In Seasonal to Year-Two Colorado River Streamflow Prediction Workshop, Salt Lake City, Utah. Colorado Basin River Forecast Center.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. Water Resources Research 41, W10410.

Grantz, K., B. Rajagopalan, E. Zagona, and M. Clark (2007). Water Management Applications of Climate-Based Hydrologic Forecasts : Case Study of the Truckee-Carson River Basin. Journal of Water Resources Planning and Management 133(4), 339–350.

Grantz, K. A. (2006). Interannual variability of North American Monsoon hydroclimate and application to water management in the Pecos River Basin. Ph. D. thesis, University of Colorado.

Harrold, T. I., A. Sharma, and S. J. Sheather (2003). A nonparametric model for stochastic generation of daily rainfall amounts. Water Resources Research 39(12), 1–12.

Hartigan, J. A. and M. A. Wong (1979). A k-means clustering algorithm. Applied Statistics 28(1), 100–108.

Hartmann, H. C., R. Bales, and S. Sorooshian (2002). Weather, climate, and hydrologic forecasting for the US Southwest : a survey. Climate Research 21, 239–258.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction in Statistics (2nd ed.). Springer Series in Statistics.

Holden, P. B. (1999). Flow Recommendations for the San Juan River. Technical Report May, The San Juan River Basin Recovery Implementation Program. U.S. Fish and Wildlife Service, Albuquerque, NM.

Katz, R. W. (1977). Precipitation as a chain-dependent process. Journal of Applied Meteorlogy 16(7), 671–676.

Khalili, M., F. Brissette, and R. Leconte (2011, April). Effectiveness of multi-site weather generator for hydrological modeling. Journal of the American Water Resources Association 47(2), 303–314.

Khazaei, M. R., B. Zahabiyoun, and B. Saghafian (2011, August). Assessment of climate change impact on floods using weather generator and continuous rainfall-runoff model. International Journal of Climatology, n/a–n/a.

Kim, Y., R. Katz, B. Rajagopalan, G. Podestá, and E. Furrer (2012, May). Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach. Climate Research 53(1), 13–24.

Kleiber, W., R. W. Katz, and B. Rajagopalan (2012). Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. Water Resources Research 48(1), W01523.

Lall, U. and A. Sharma (1996). A nearest neighbor bootstrap for time series resampling. Water Resources Research 32(3), 679–693.

Larson, L. (2002). National Weather Service River Forecast System (NWSRFS). In V. P. Singh and D. K. Frevert (Eds.), Mathematical Models of Small Watershed Hydrology and Applications, pp. 657–706. Littleton, CO: Water Resources Publications, LLC.

Leander, R. and T. A. Buishand (2009, August). A daily weather generator based on a two-stage resampling algorithm. Journal of Hydrology 374(3-4), 185–195.

Leander, R., T. A. Buishand, P. Aalders, and M. de Wit (2006). Estimation of extreme floods of the river Meuse using a stochastic weather generator and a rainfall-runoff model. Hydrological Sciences Journal 50(6), 1089–1103.

Leander, R., T. A. Buishand, B. J. van den Hurk, and M. J. de Wit (2007). Estimated changes in flood quantiles of the river meuse from resampling of regional climate model output. Journal of Hydrology 351(3-4), 331–343.

McCabe, G. J. and D. M. Wolock (2009, October). Recent Declines in Western U.S. Snowpack in the Context of Twentieth-Century Climate Variability. Earth Interactions 13(12), 1–15.

Mehrotra, R. and A. Sharma (2007). A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. Journal of Hydrology 335(1-2), 180–193.

Mehrotra, R., R. Srikanthan, and A. Sharma (2006). A comparison of three stochastic multi-site precipitation occurrence generators. Journal of Hydrology 331(1-2), 280–292.

Mezghani, A. and B. Hingray (2009). A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the upper rhone river basin. Journal of Hydrology 377(3-4), 245–260.

Mountain, N. and J. Jones (2006). Reconstructing extreme flows using an airflow index-based stochastic weather generator and a hydrological simulation model. Catena 66(1-2), 120–134.

Najafi, M. R., H. Moradkhani, and T. Piechota (2012, April). Ensemble streamflow prediction: Climate signal weighting methods vs. climate forecast system reanalysis. Journal of Hydrology.

NWSRFS. NWSRFS User Manual Documentation.

Pagano, T. C. and D. C. Garen (2005). Integration of Climate Information and Forecasts into Western US Water Supply Forecasts. In J. D. Garbrecht and T. C. Piechota (Eds.), Climate Variations, Climate Change, and Water Resources Engineering. ASCE.

Qian, B., J. Corte-Real, and H. Xu (2002). Multisite stochastic weather models for impact studies. International Journal of Climatology 22(11), 1377–1397.

Rajagopalan, B. and U. Lall (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables. Water Resources Research 35(10), 3089–3101.

Rajagopalan, B., U. Lall, D. G. Tarboton, and D. S. Bowles (1997). Multivariate nonparametric resampling scheme for generation of daily weather variables. Stochastic Hydrology and Hydraulics 11(1), 523–547.

Rajagopalan, B., K. Nowak, J. Prairie, M. Hoerling, B. Harding, J. Barsugli, A. Ray, and B. Udall (2009, August). Water supply risk on the Colorado River: Can management mitigate? Water Resources Research 45(8), W08201.

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resources Research 42(9), 9404.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research 17(1), 182–190.

Richardson, C. W. and D. A. Wright (1984). WGEN : A Model for Generating Daily Weather Variables. Technical report, U.S. Department of Agricultural Research Service, Washington, DC.

Roulin, E. (2007). Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrology and Earth System Sciences 11(2), 725–737.

Segond, M. L., C. Onof, and H. S. Wheater (2006). Spatial-temporal disaggregation of daily rainfall from a generalized linear model. Journal of Hydrology 331, 674–689.

Sharif, M. and D. H. Burn (2007). Improved k-nearest neighbor weather generating model. Journal of Hydrologic Engineering 12(1), 42–51.

Smith, R. E. (1994). Statistics for the Environment 2: Water Related Issues, Chapter Spatial modelling of rainfall data, pp. 19–41. New York: John Wiley.

Srikanthan, R. S. and G. G. Pegram (2009). A nested multisite daily rainfall stochastic generation model. Journal of Hydrology 371(1-4), 142–153.

Stern, R. D. and R. Coe (1984). A model fitting analysis of daily rainfall data. Journal of the Royal Statistical Society 147(1), 1–34.

Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society - Series B: Statistical Methodology 63(2), 411–423.

USBR (2008, June). Resource Management Plan Navajo Reservoir Area. Technical Report June, U.S. Department of the Interior, Bureau of Reclamation.

Wallis, T. W. and J. F. Griffiths (1997). Simulated meteorological input for agricultural models. Agricultural and Forest Meteorology 88(1-4), 241–258.

Werner, K. (2011a, March 21). Current CBRFC Water Supply Forecast Methodology. In Seasonal to Year-Two Colorado River Streamflow Prediction Workshop, Salt Lake City, Utah. U.S. National Oceanic and Atmospheric Administration.

Werner, K. (2011b, August 12). Noaa's colorado basin river forecast center:"climate services on the colorado river: Capabilities, gaps, and chasms". In Climate Test Bed Joint Seminar Series, Camp Springs, Maryland. U.S. National Oceanic and Atmospheric Administration.

Werner, K., D. Brandon, M. Clark, and S. Gangopadhyay (2004). Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts. Journal of Hydrometeorology 5(6), 1076–1090.

Wilks, D. S. (1995). Statistical Methods in the Atmospheric Sciences. Academic Press, New York.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. Journal of Hydrology 210, 178–191.

Wilks, D. S. (1999). Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. Agricultural and Forest Meteorology 96(1-3), 85–101.

Wilks, D. S. and R. L. Wilby (1999). The weather generation game: a review of stochastic weather models. Progress in Physical Geography 23(3), 329–357.

Wood, A. and K. Werner (2011, October). Development of a Seasonal Climate and Streamflow Forecasting Testbed for the Colorado River Basin. In 36th Climate Diagnostic and Prediction Workshop, Fort Worth, TX, pp. 101–105. NOAA NWS.

Wood, A. W. and D. P. Lettenmaier (2008, July). An ensemble approach for attribution of hydrologic prediction uncertainty. Geophysical Research Letters 35, L14401.

Wood, A. W. and J. C. Schaake (2007, February). Correcting errors in streamflow forecast ensemble mean and spread. Journal of Hydrometeorology 9(1), 132–148.

Woolhiser, D. A. (1992). Modeling daily precipitation: Progress and problems. Statistics in the Environmental and Earth Sciences, 71–89.

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater (2005). Spatial-temporal rainfall simulation using generalized linear models. Water Resources Research 41(11), W11415.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resources Research 39(7), 1199.

Appendix A

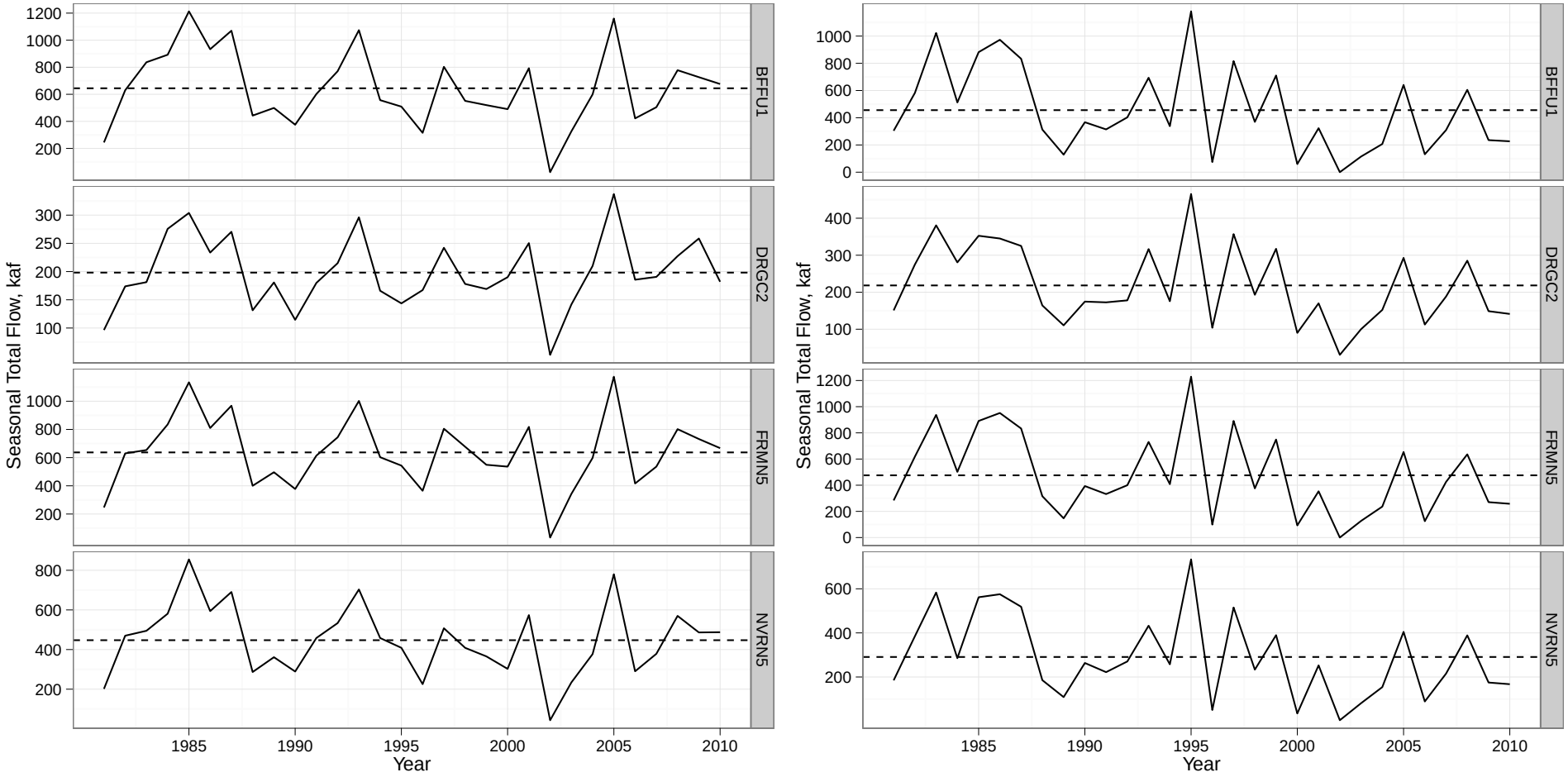
Additional CHPS Figures

This section contains April–May and June–July sub-season runoff figures corresponding to their April–July counterparts in Chapter 3, as well as additional figures that were mentioned but not displayed.

A.1 Unconditional

Figure A.14 has correlations between historical runoff and ensemble means and medians. Both methods yield near-identical results. Oddly, correlations are smaller for March than for April — opposite of the RPSS trend. In Figure A.8, median RPSS is largely negative, median negative for long lead times. These improve considerably in Figure A.9. Figures A.12 and A.13 attempt to discern a pattern in values of RPSS with respect to magnitude of runoff. While there is not a strict linear trend, there appears a type of regime where once a certain magnitude is reached, RPSS indicates higher skill for both methods. In the differences, however, there is not as strong of a pattern. Both of these figures contain the unconditional values of the select years (2005 and 2006), contributing additional insight into the trends seen in the conditional figures. Figure A.15 shows WG ESP underforecasts at shorter lead times, but ESP tends to lack points. Figure A.16 has WG ESP only falling in the highest bin for long leads (again ESP does not show up). For shorter leads, but have poor resolution in the middle bin. Both methods offer skill in Figure A.17. Figure A.18 is also mostly skillful, but sometimes lacks skill for the middle bin. Figures A.19 and A.20 have similar conclusions to 3.14, though June to July has more missing points for ESP. Comparing

Figure A.22 with Figure A.23, the WG ESP for late runoff has more skew. Outliers in the late runoff may be adversely affecting the test as p-values are smaller in Table A.1 than Table A.2



(a) April to May

(b) June to July

Figure A.1: Historical runoff volumes for sub-seasons

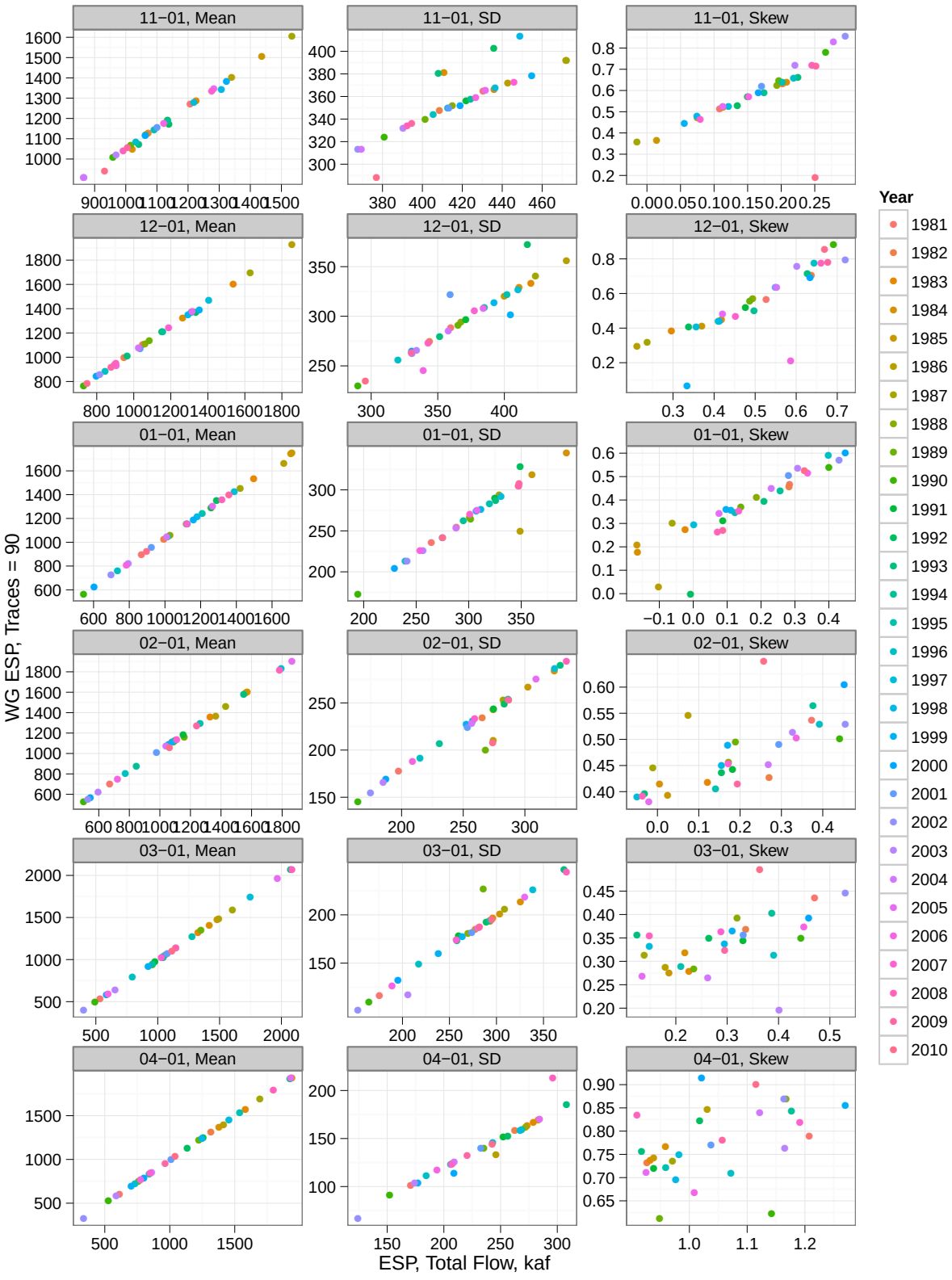


Figure A.2: April to July FRMN5 WG ESP vs ESP moments

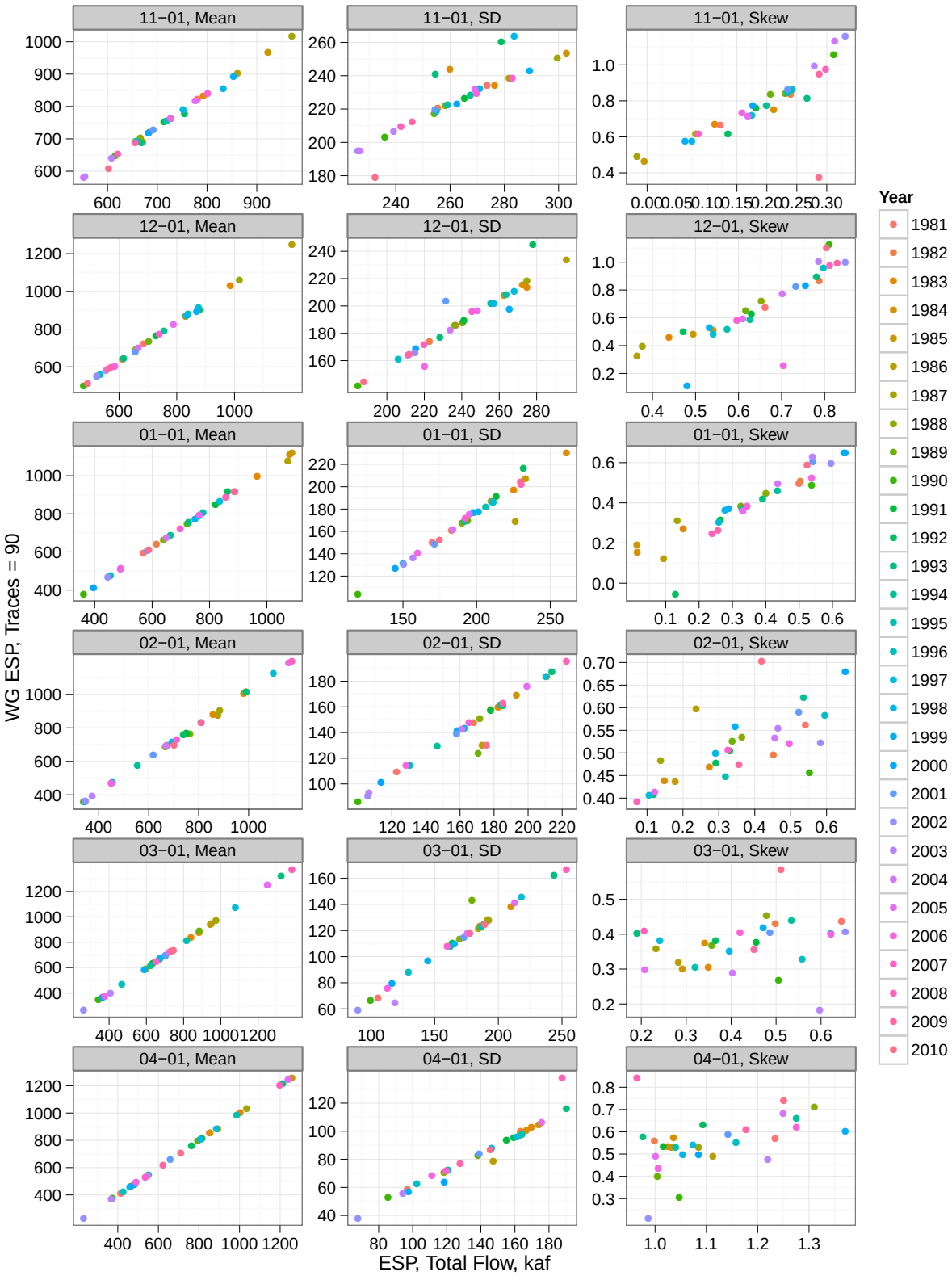


Figure A.3: April to July NVRN5L WG ESP vs ESP moments

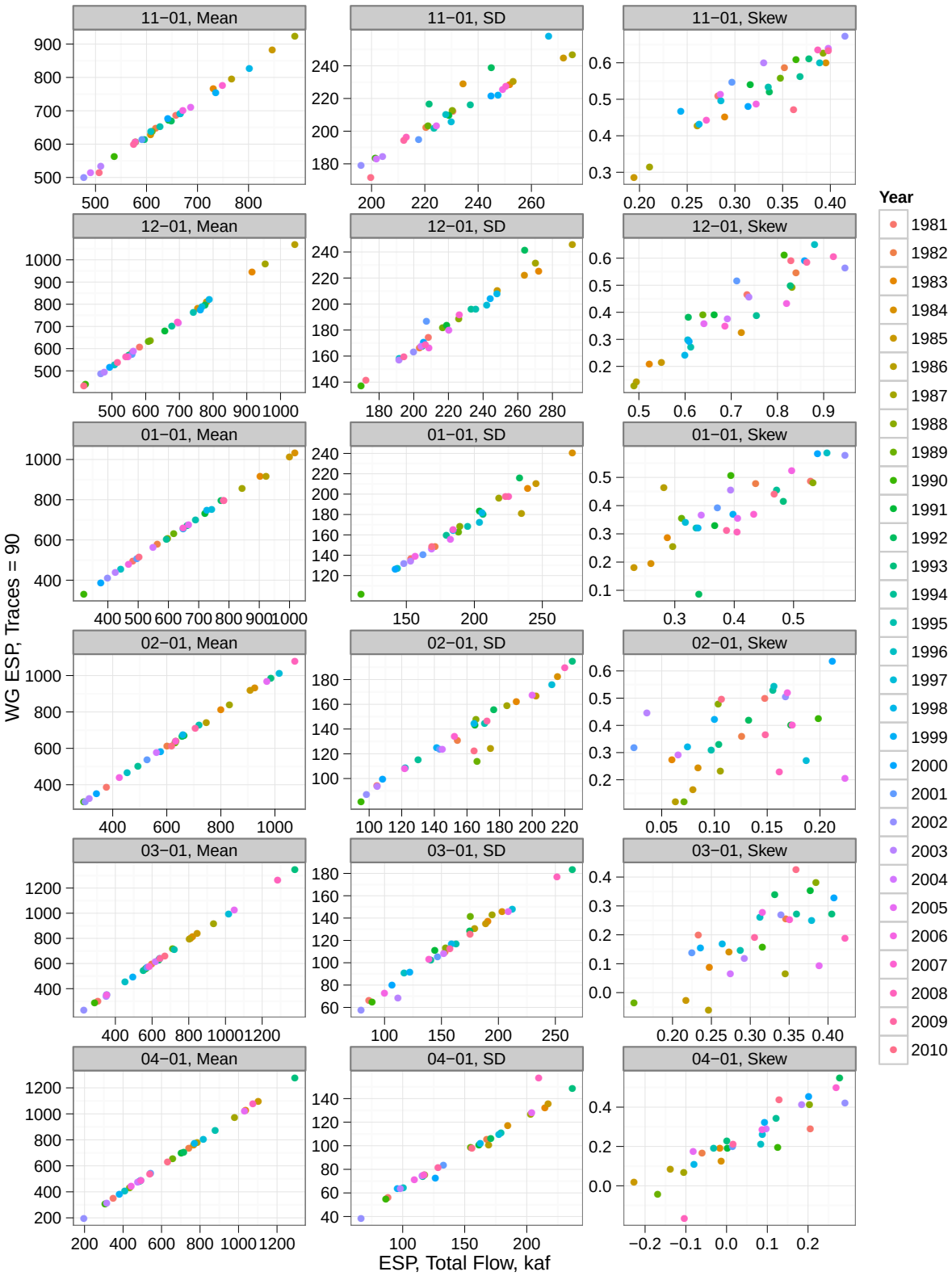


Figure A.4: April to May BFFU1 WG ESP vs ESP moments

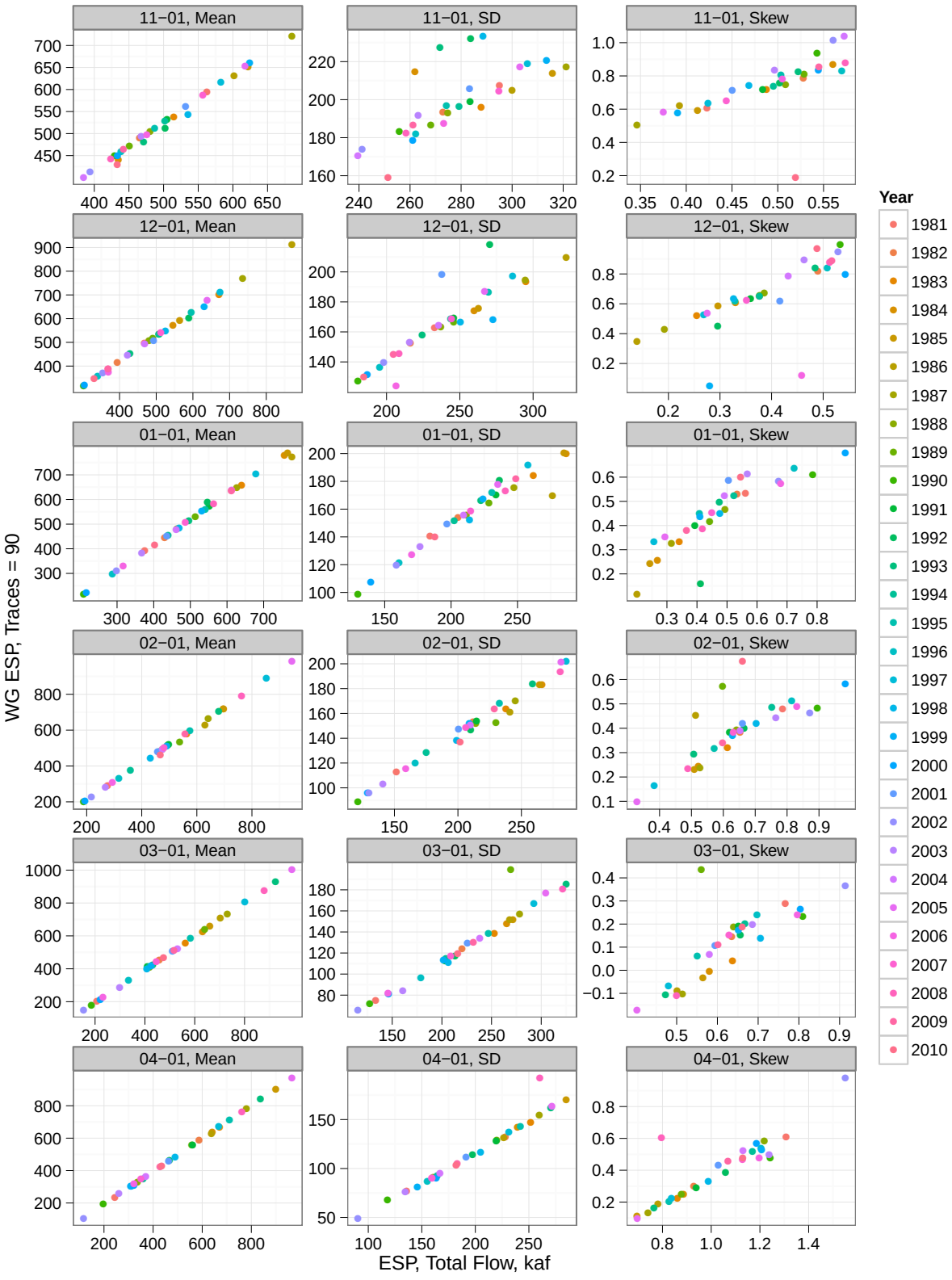


Figure A.5: June to July BFFU1 WG ESP vs ESP moments

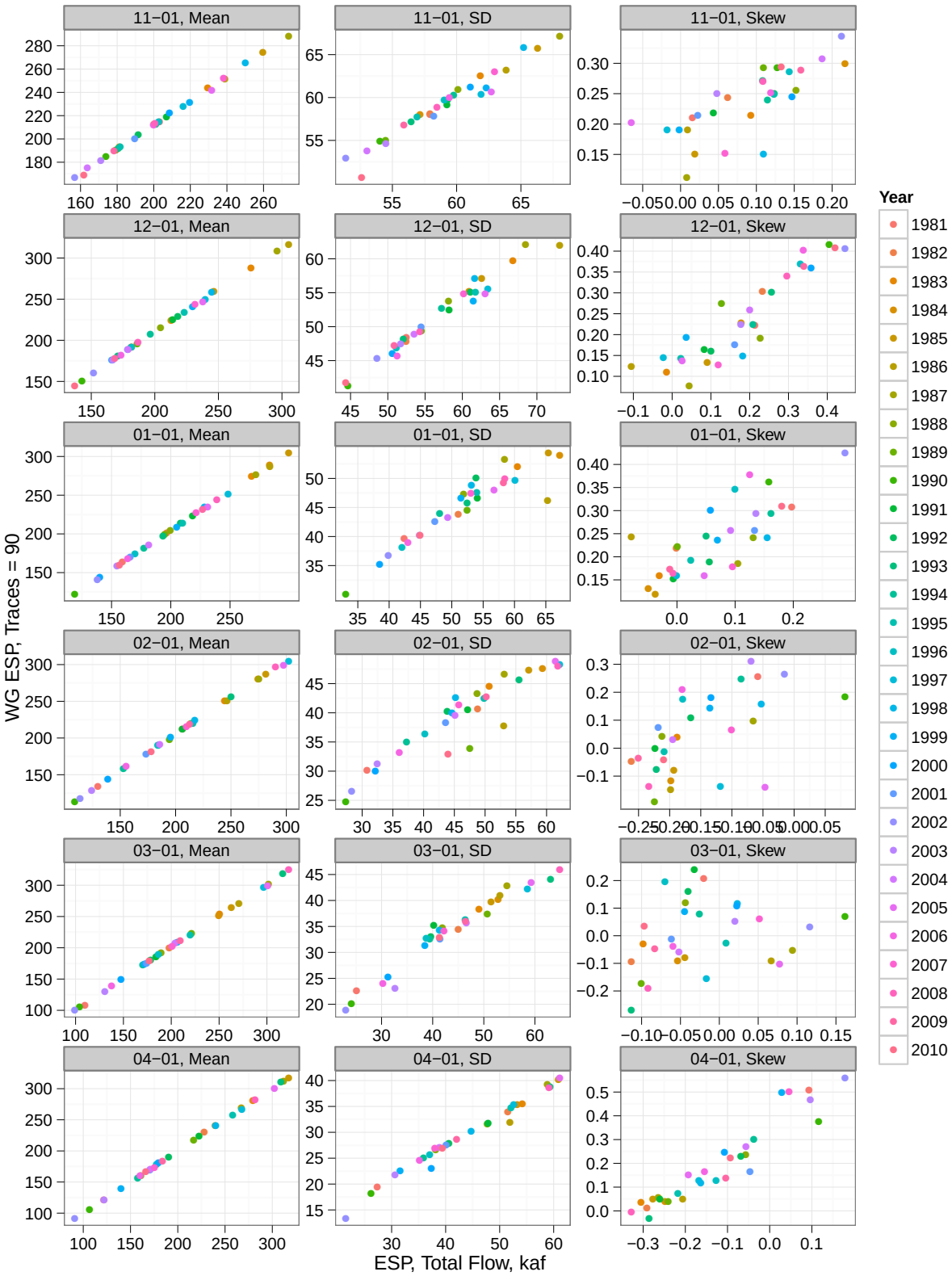


Figure A.6: April to May DRGC2 WG ESP vs ESP moments

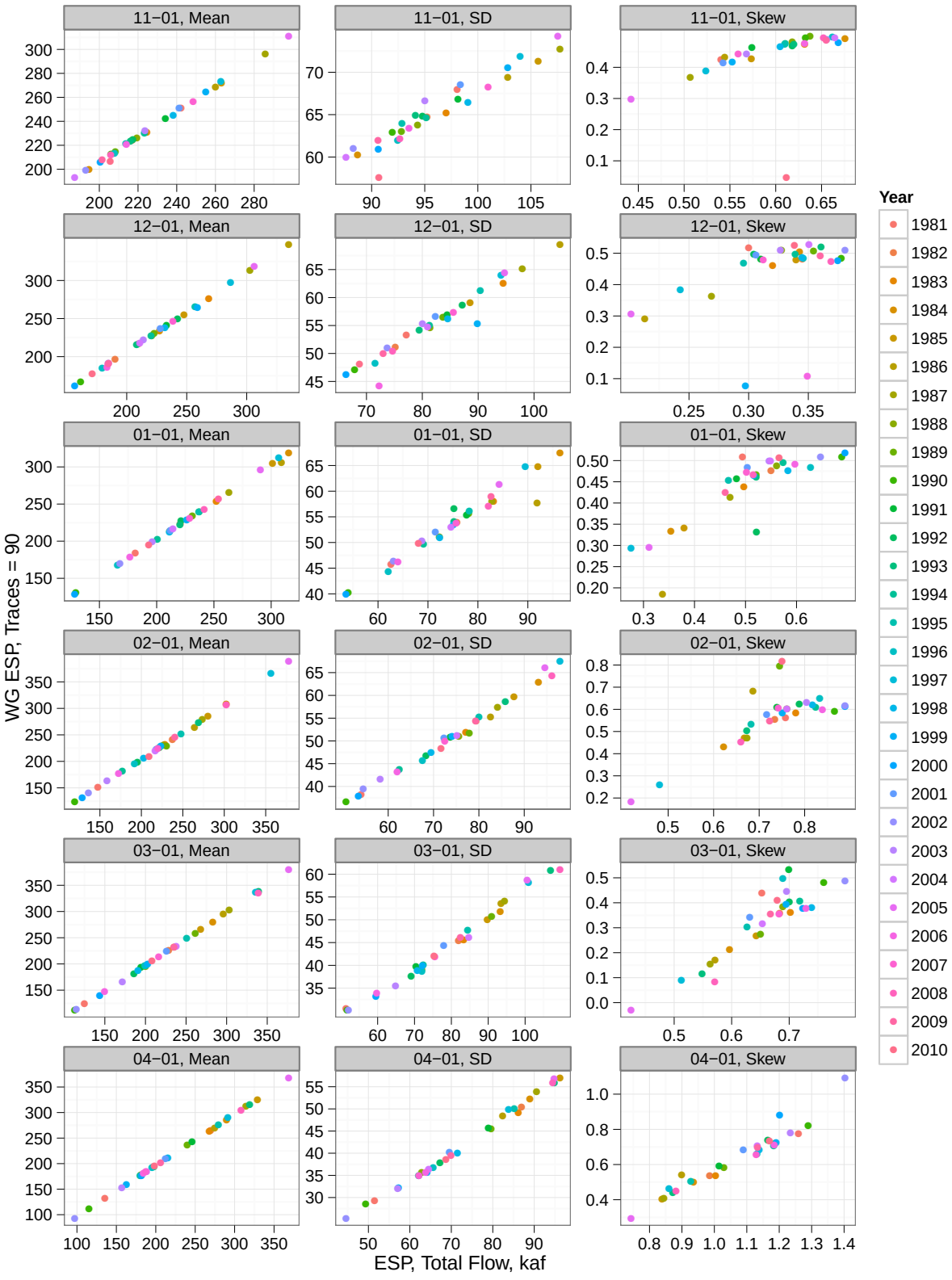


Figure A.7: June to July DRGC2 WG ESP vs ESP moments

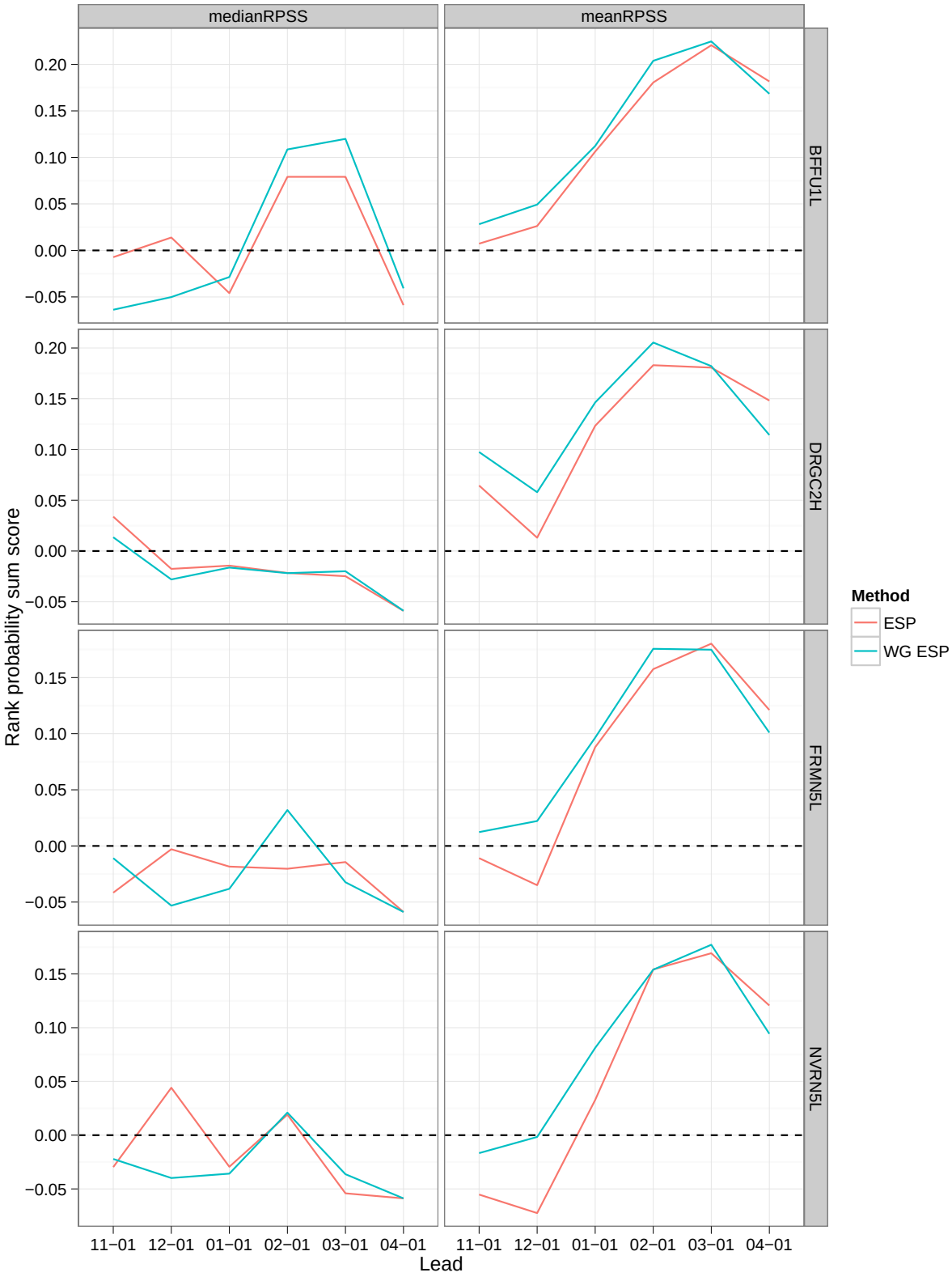


Figure A.8: April to May median and mean RPSS

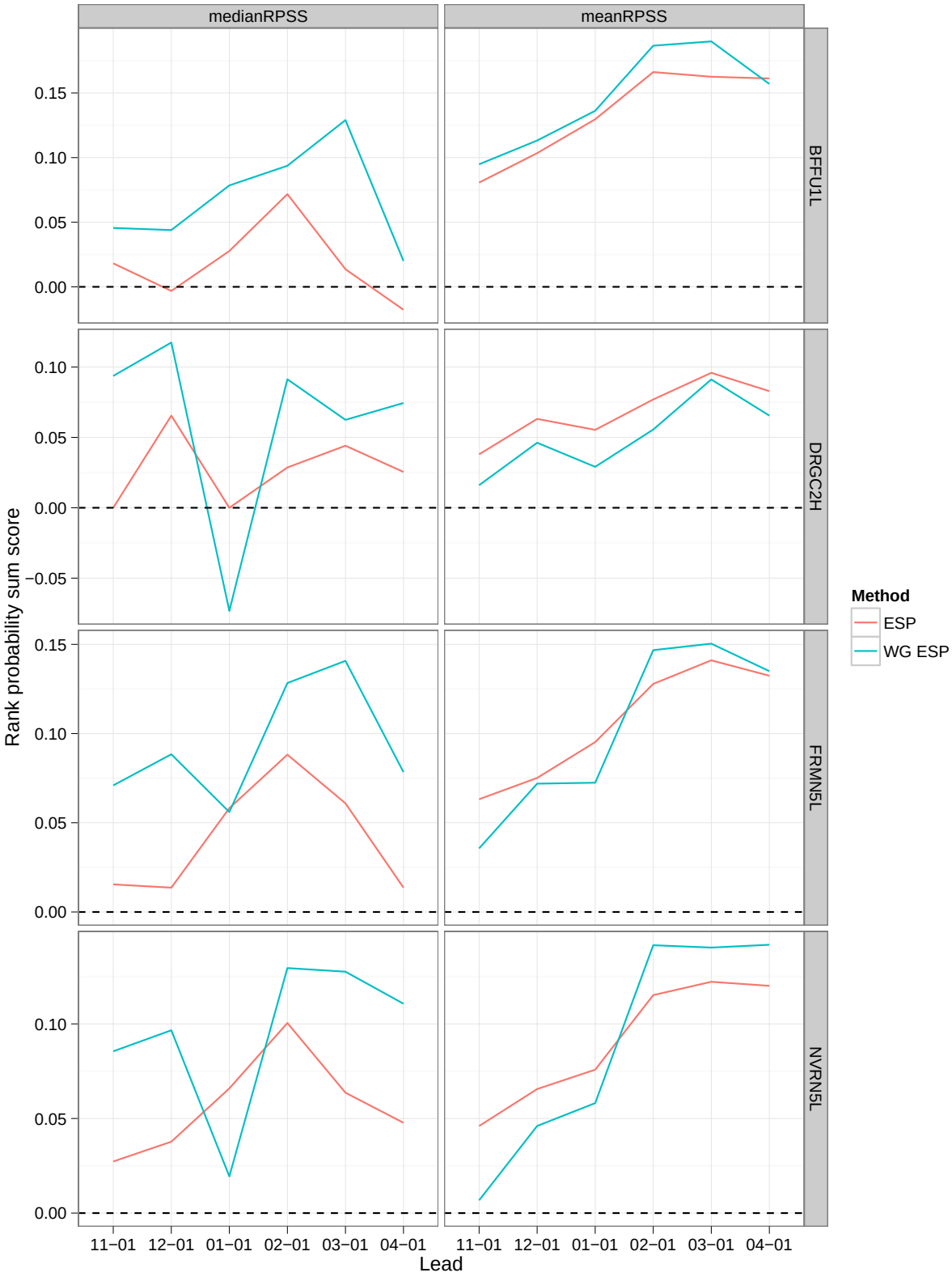


Figure A.9: June to July median and mean RPSS

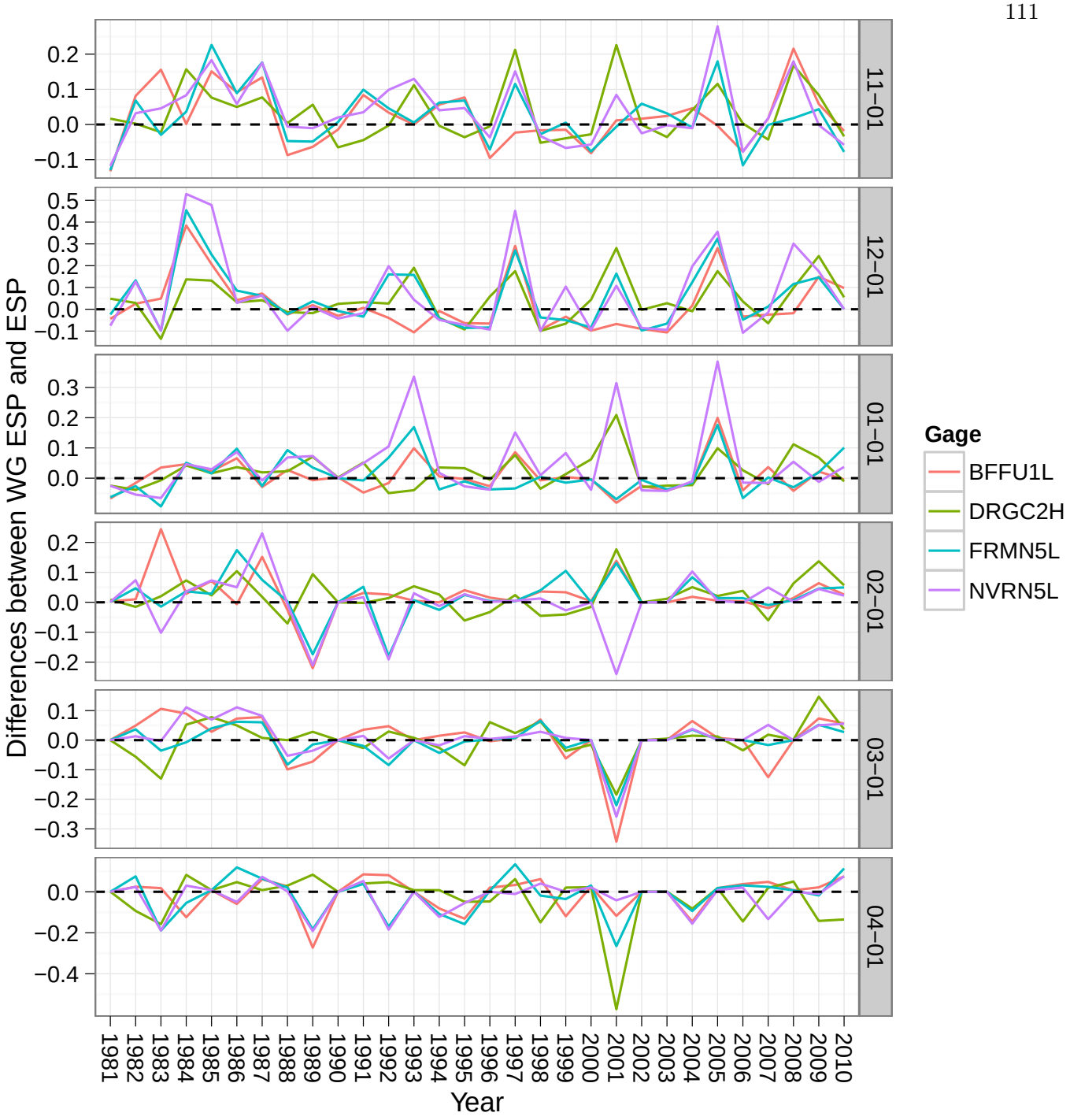


Figure A.10: April to May yearly RPSS differences (WG ESP - ESP)

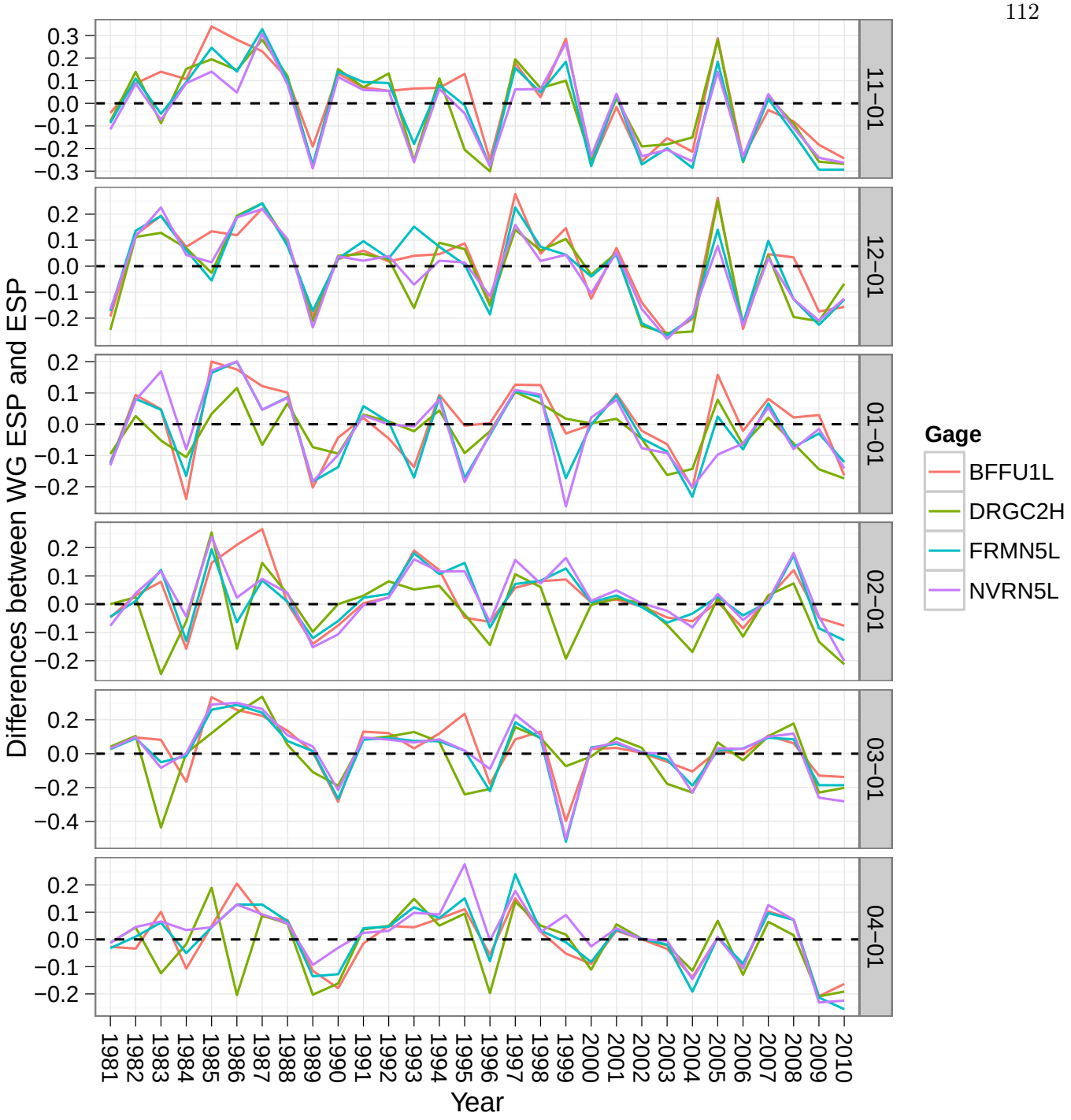


Figure A.11: June to July yearly RPSS differences (WG ESP - ESP)

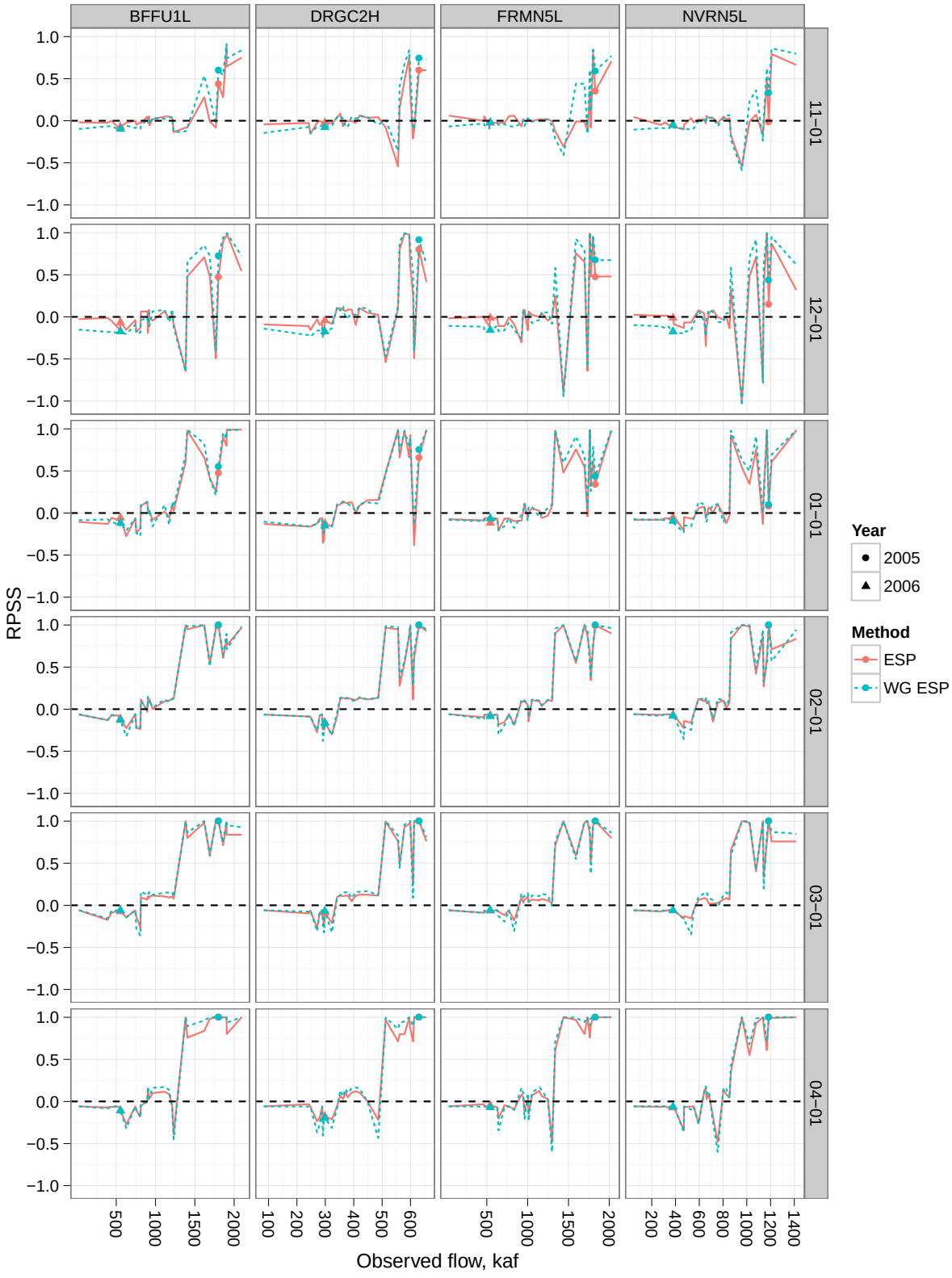


Figure A.12: RPSS of the two methods against observed April–July runoff, with unconditional RPSS of the select years highlighted

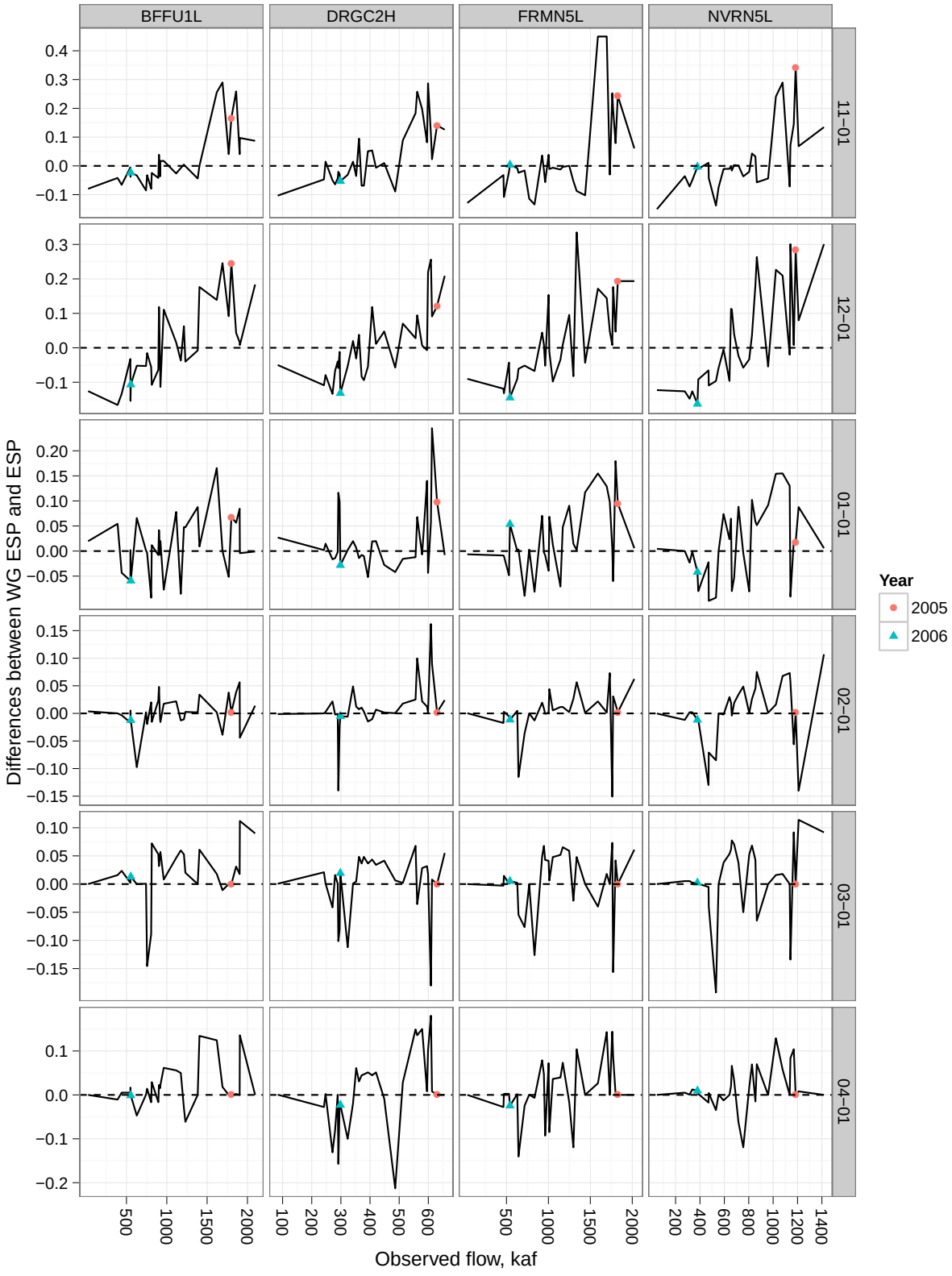


Figure A.13: RPSS differences of the two methods (WG ESP - ESP) against observed April–July runoff, with unconditional RPSS of the select years highlighted

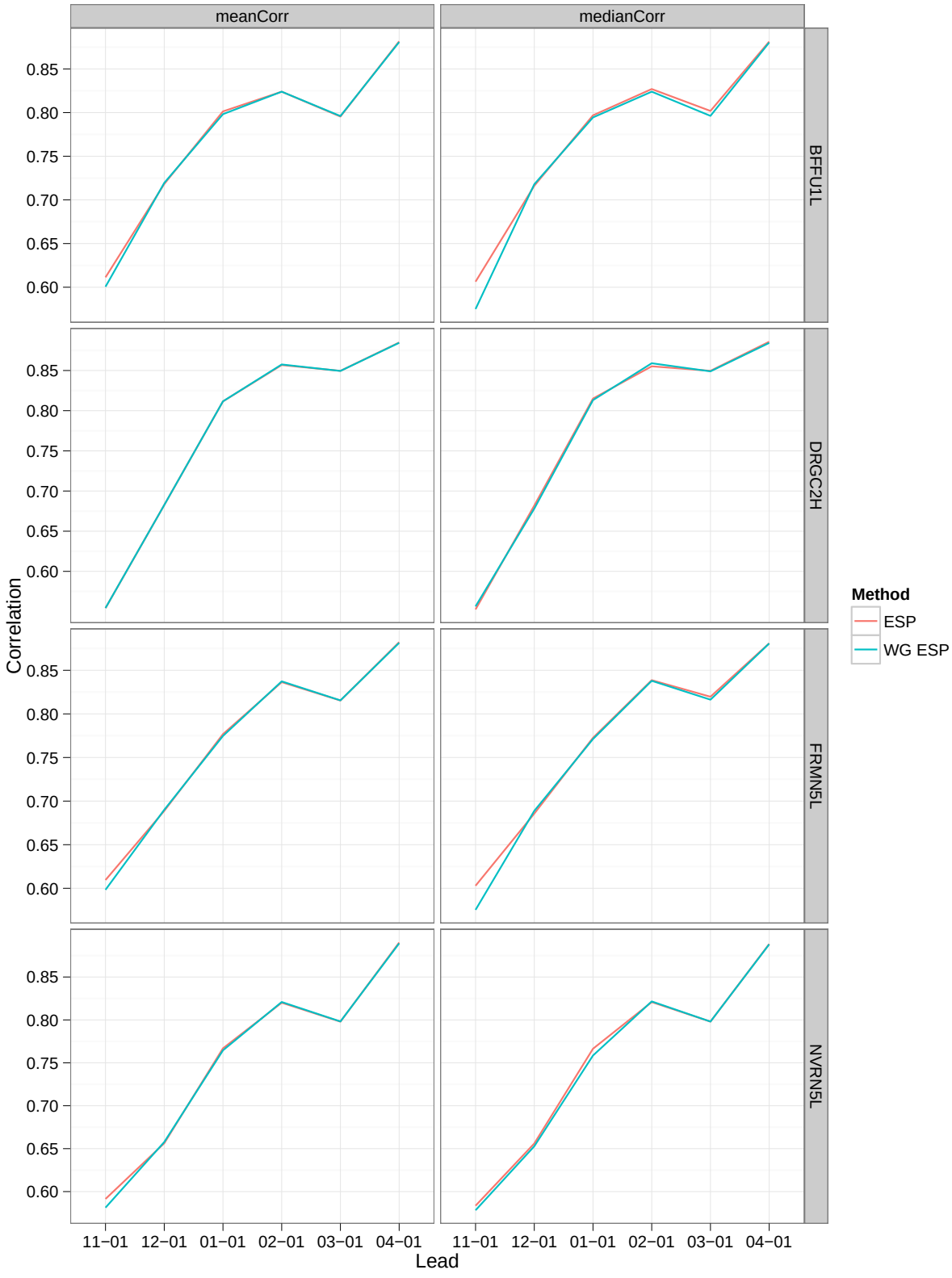


Figure A.14: Correlations between ensemble mean and median values with historical April to July runoff

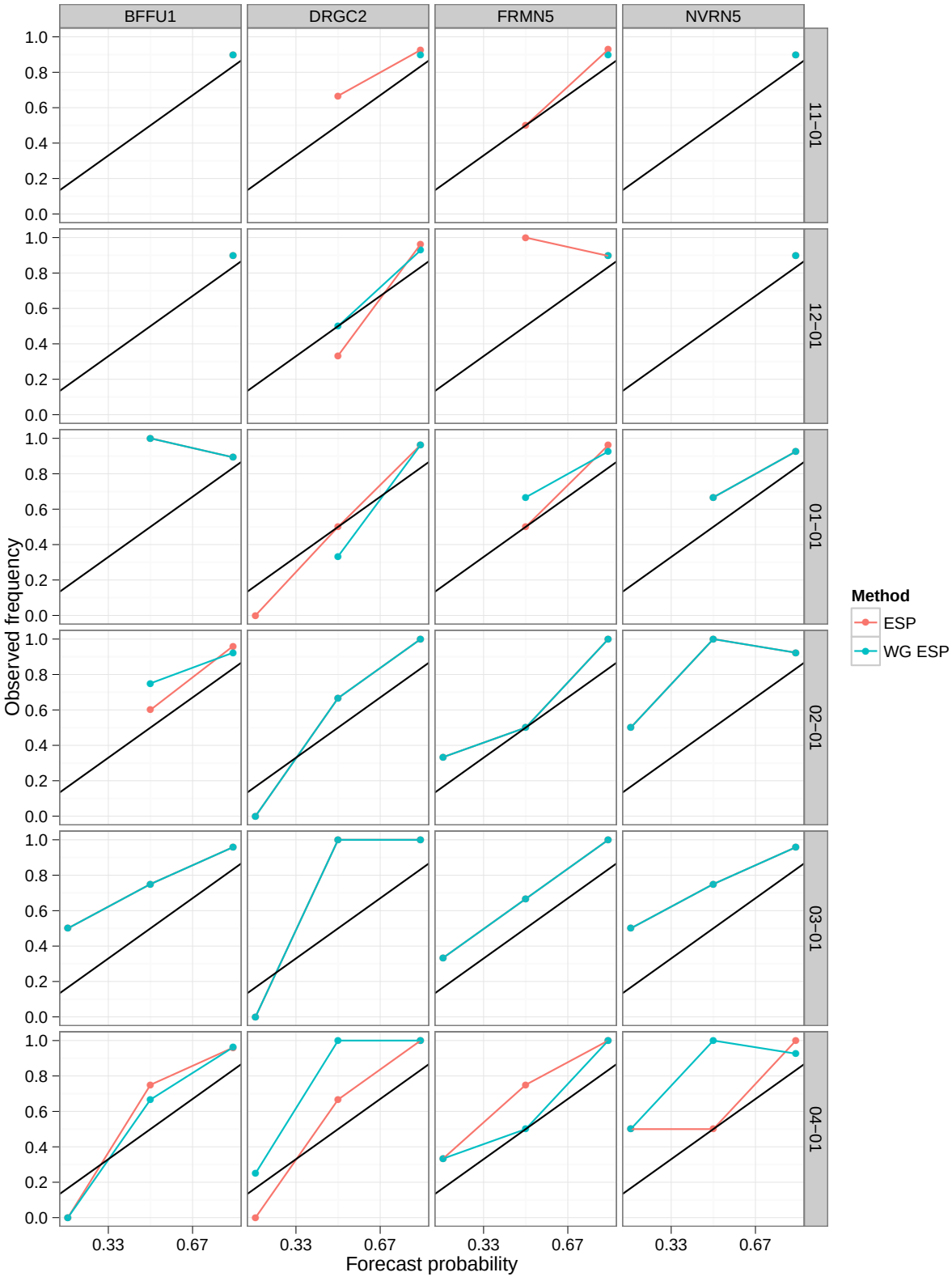


Figure A.15: Reliability diagram of April to May runoff above 10th percentile

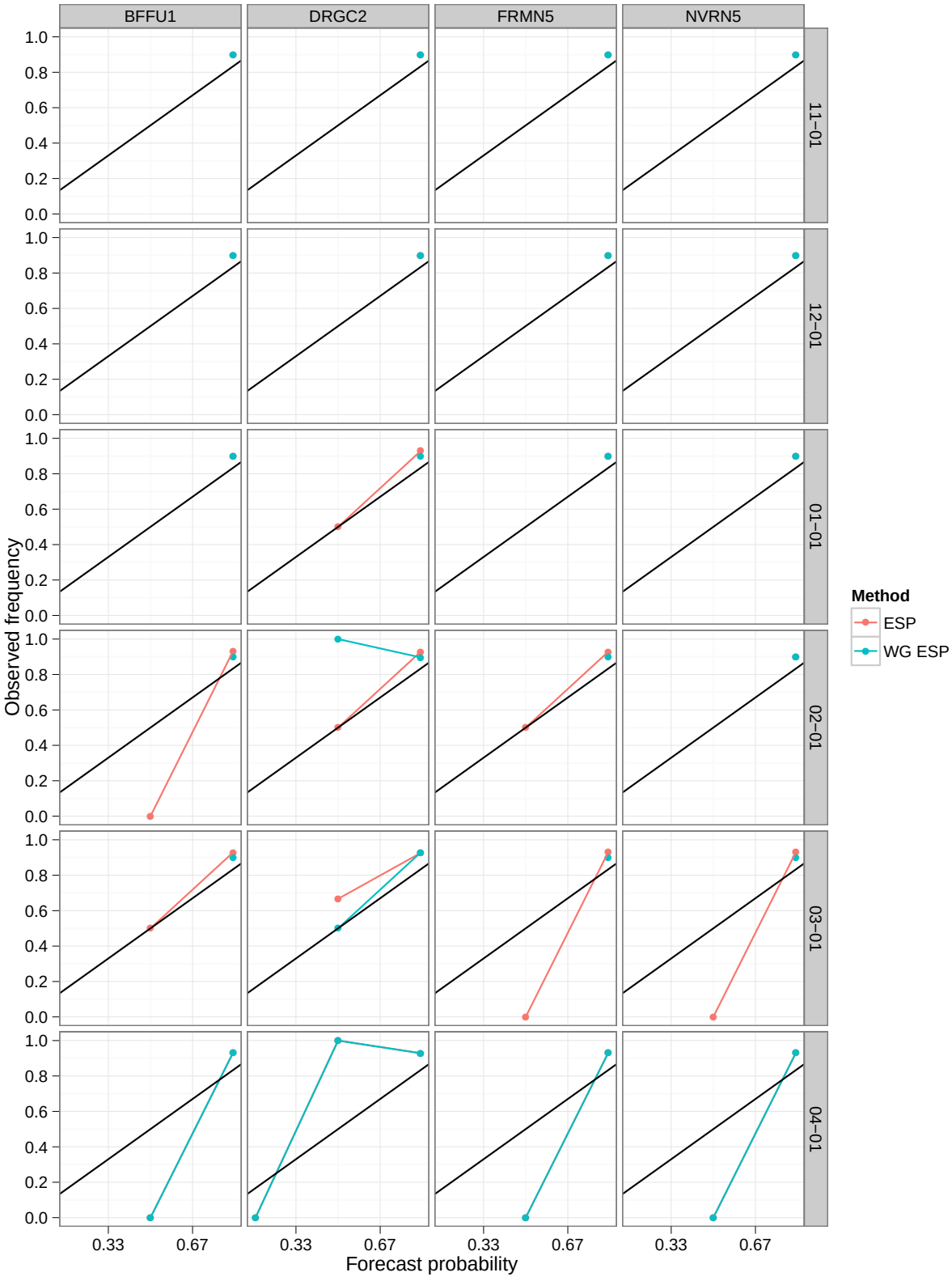


Figure A.16: Reliability diagram of June to July runoff above 10th percentile

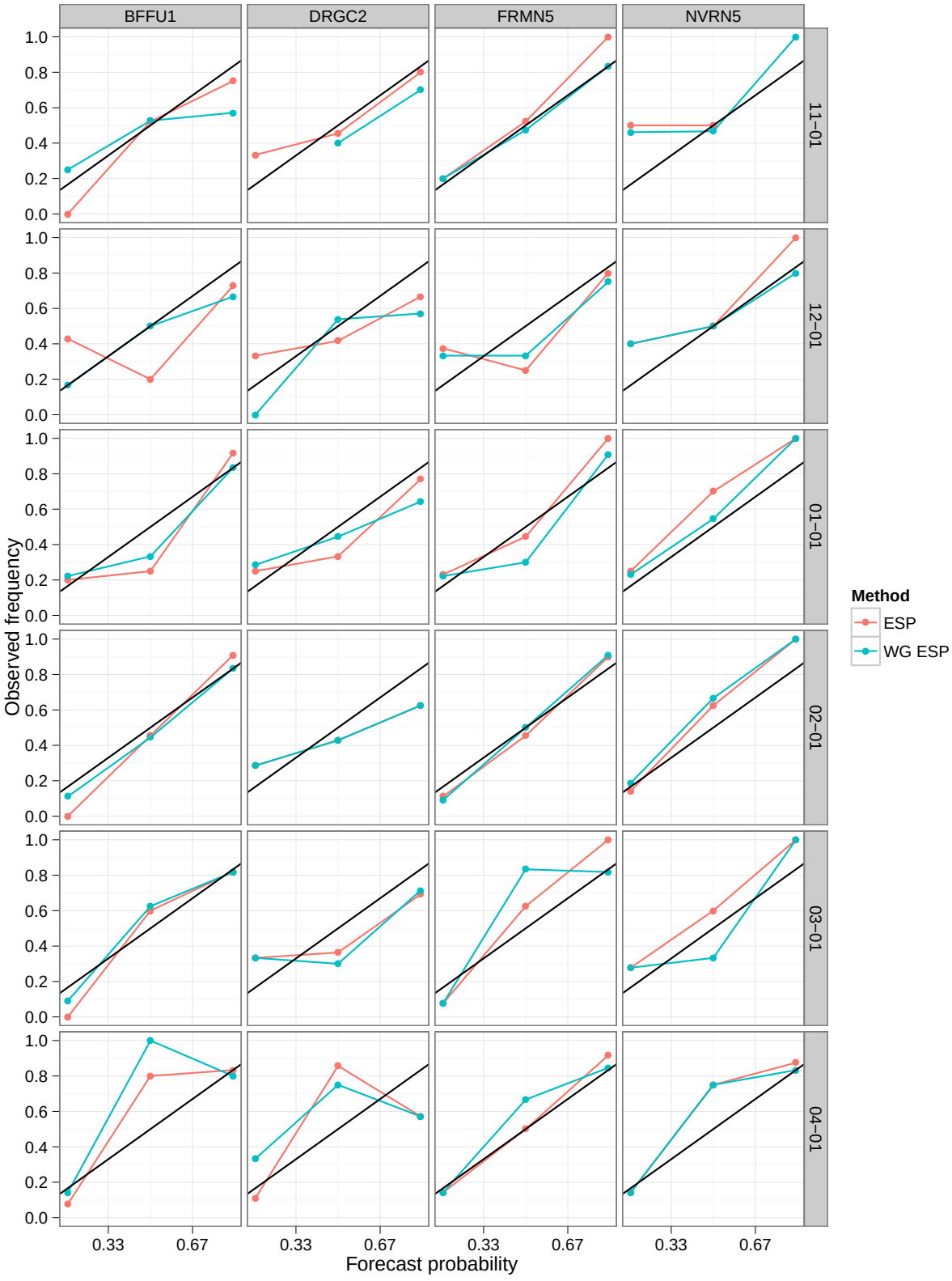


Figure A.17: Reliability diagram of April to May runoff above 50th percentile

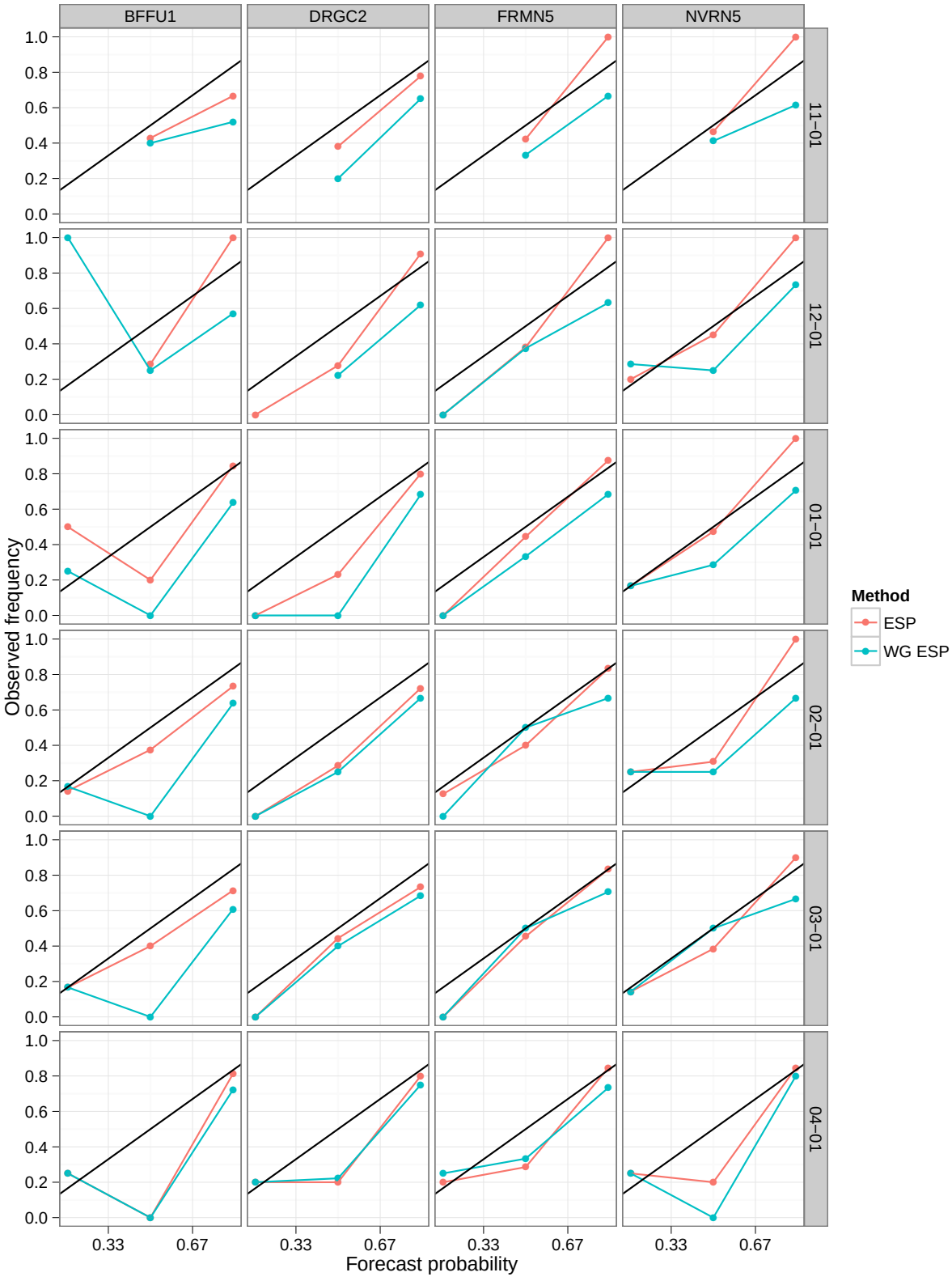


Figure A.18: Reliability diagram of June to July runoff above 50th percentile

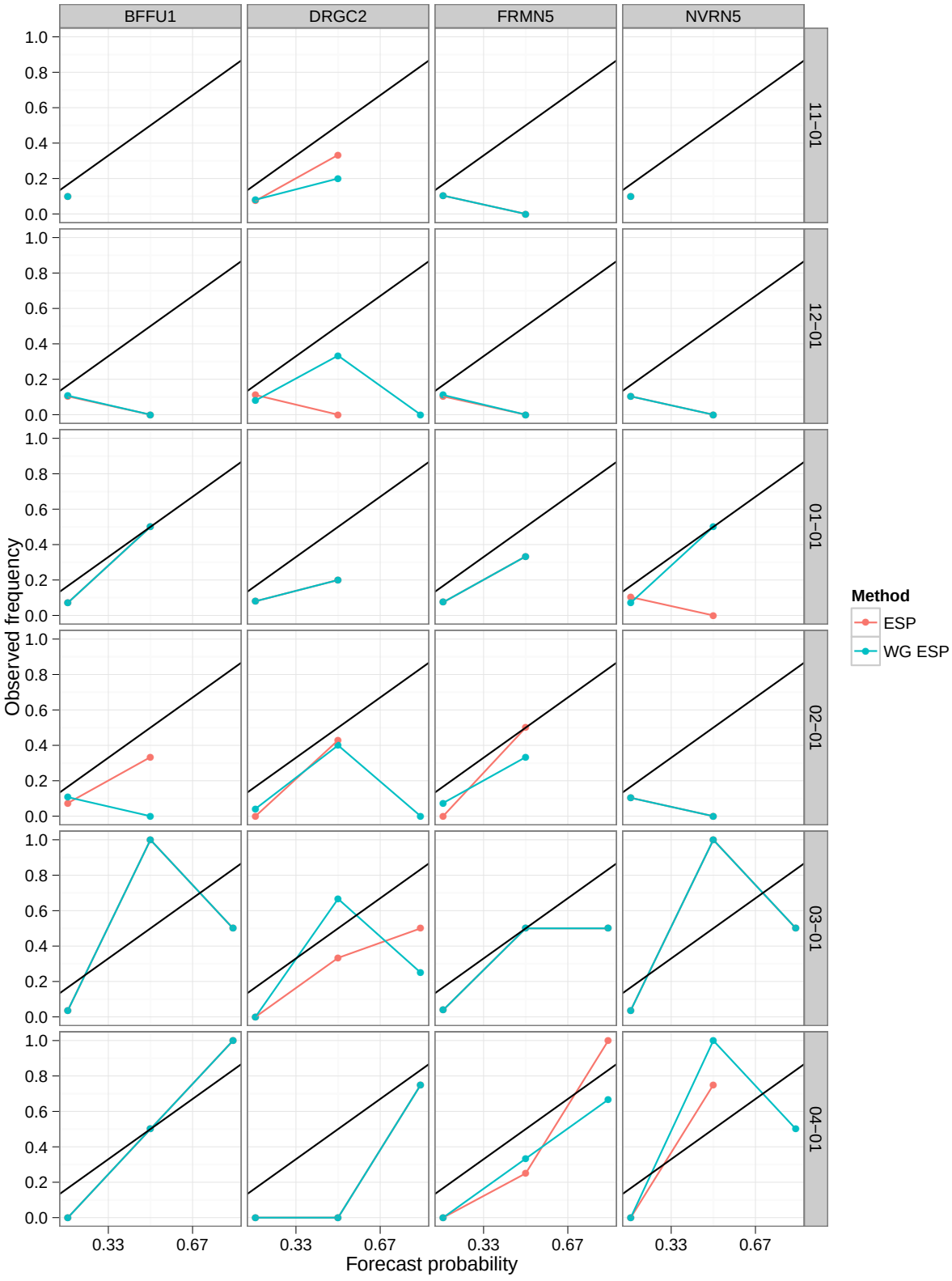


Figure A.19: Reliability diagram of April to May runoff above 90th percentile

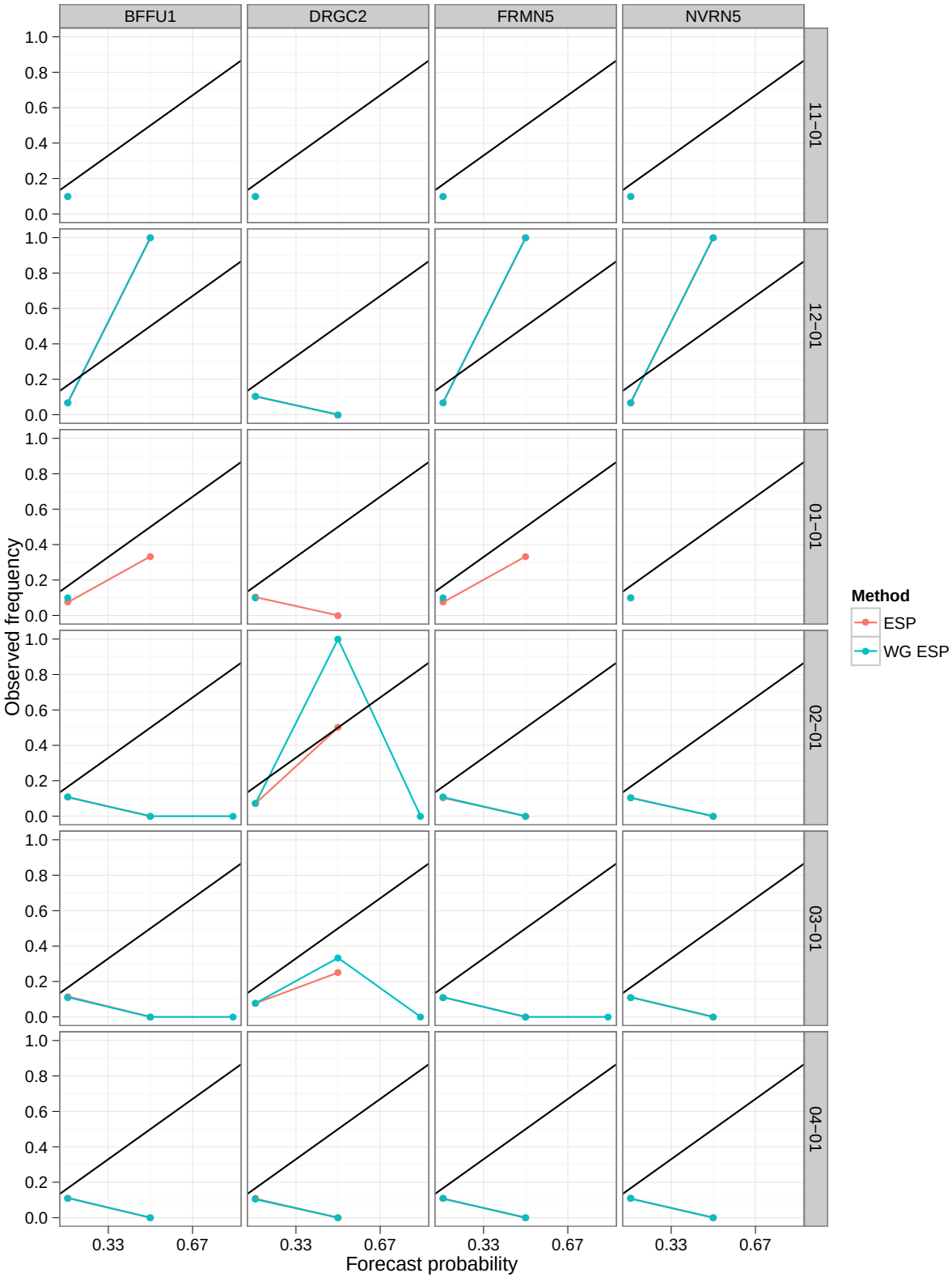


Figure A.20: Reliability diagram of June to July runoff above 90th percentile

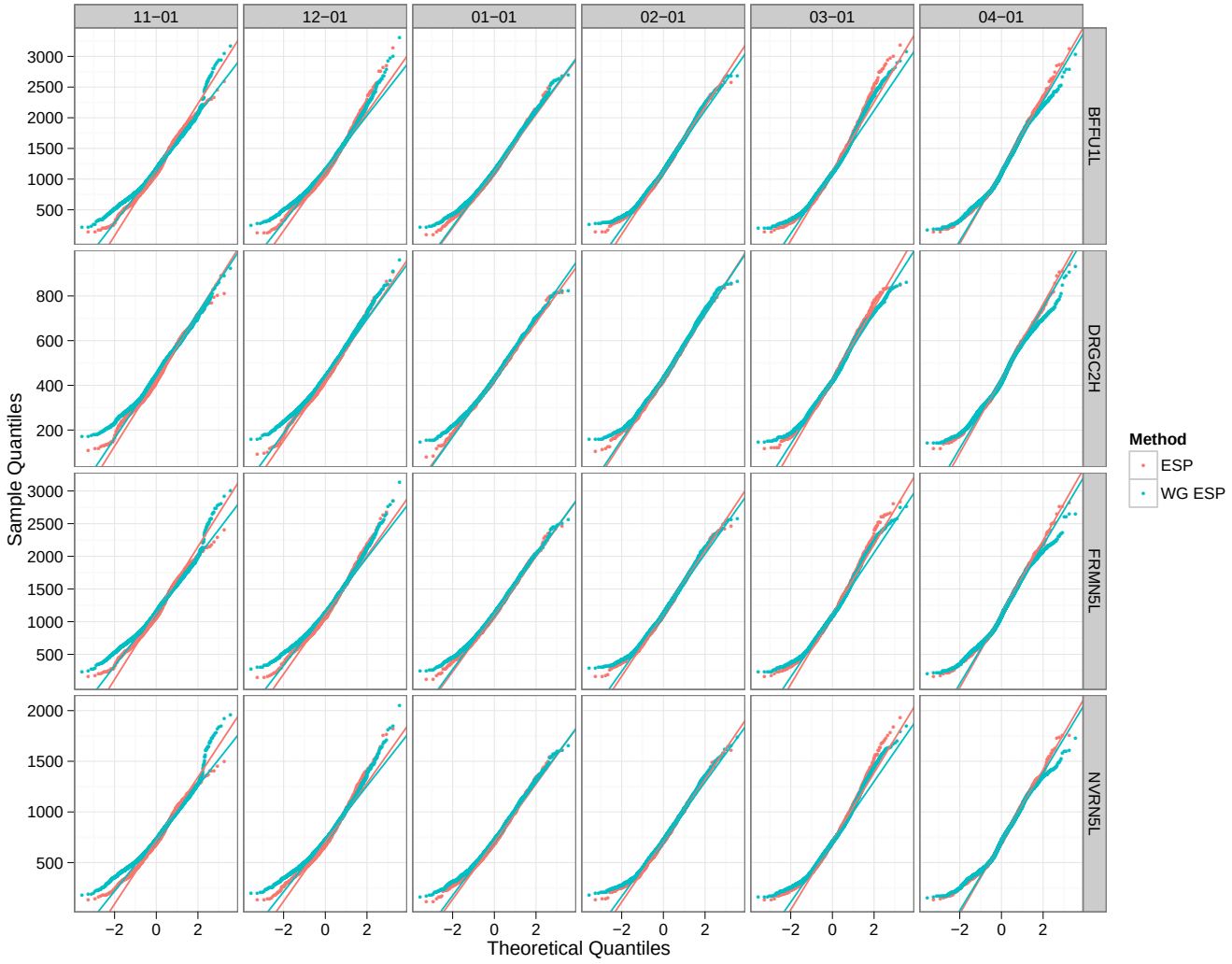


Figure A.21: April to July quantile-quantile plot for testing normality

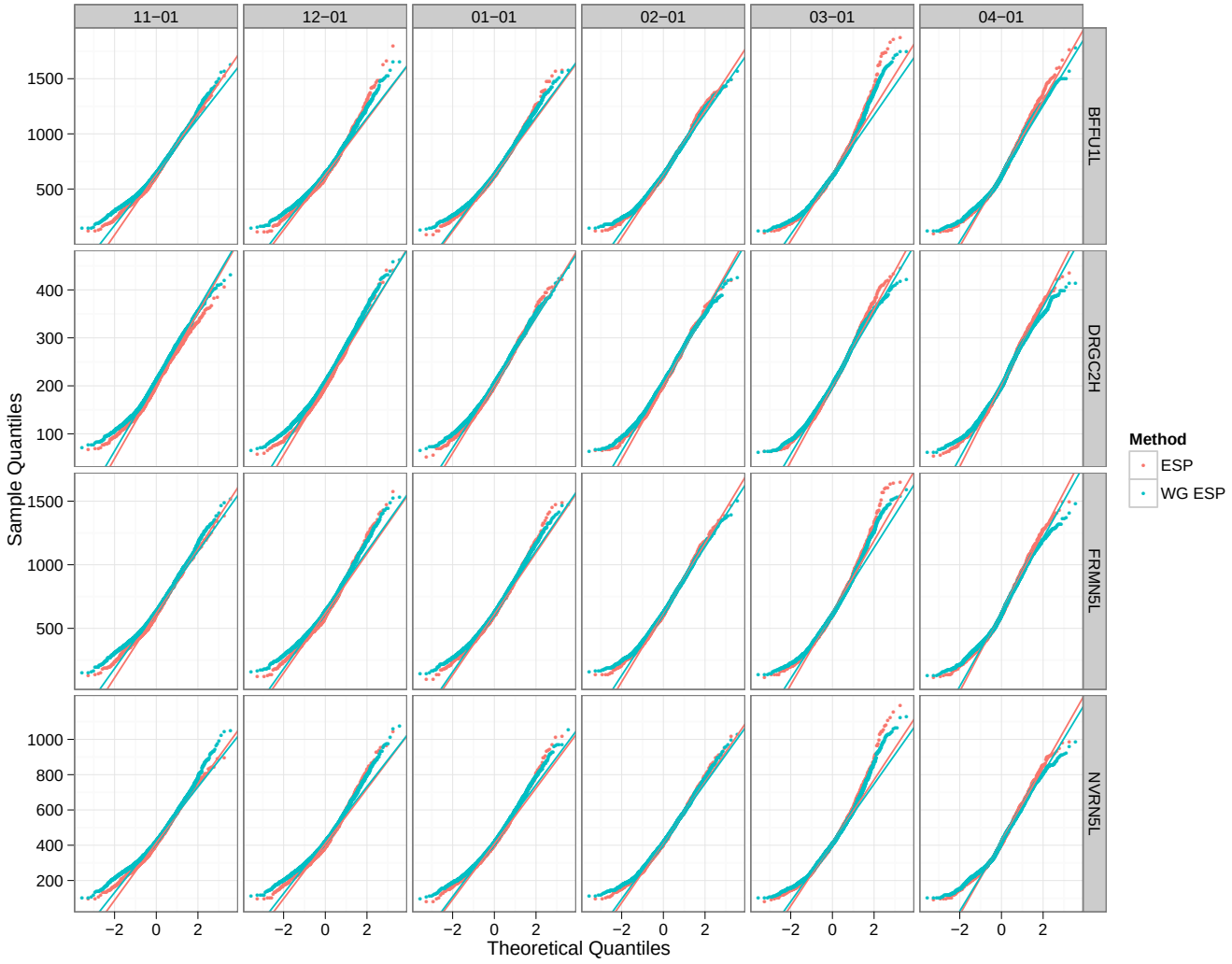


Figure A.22: Same as Figure A.21 for April to May

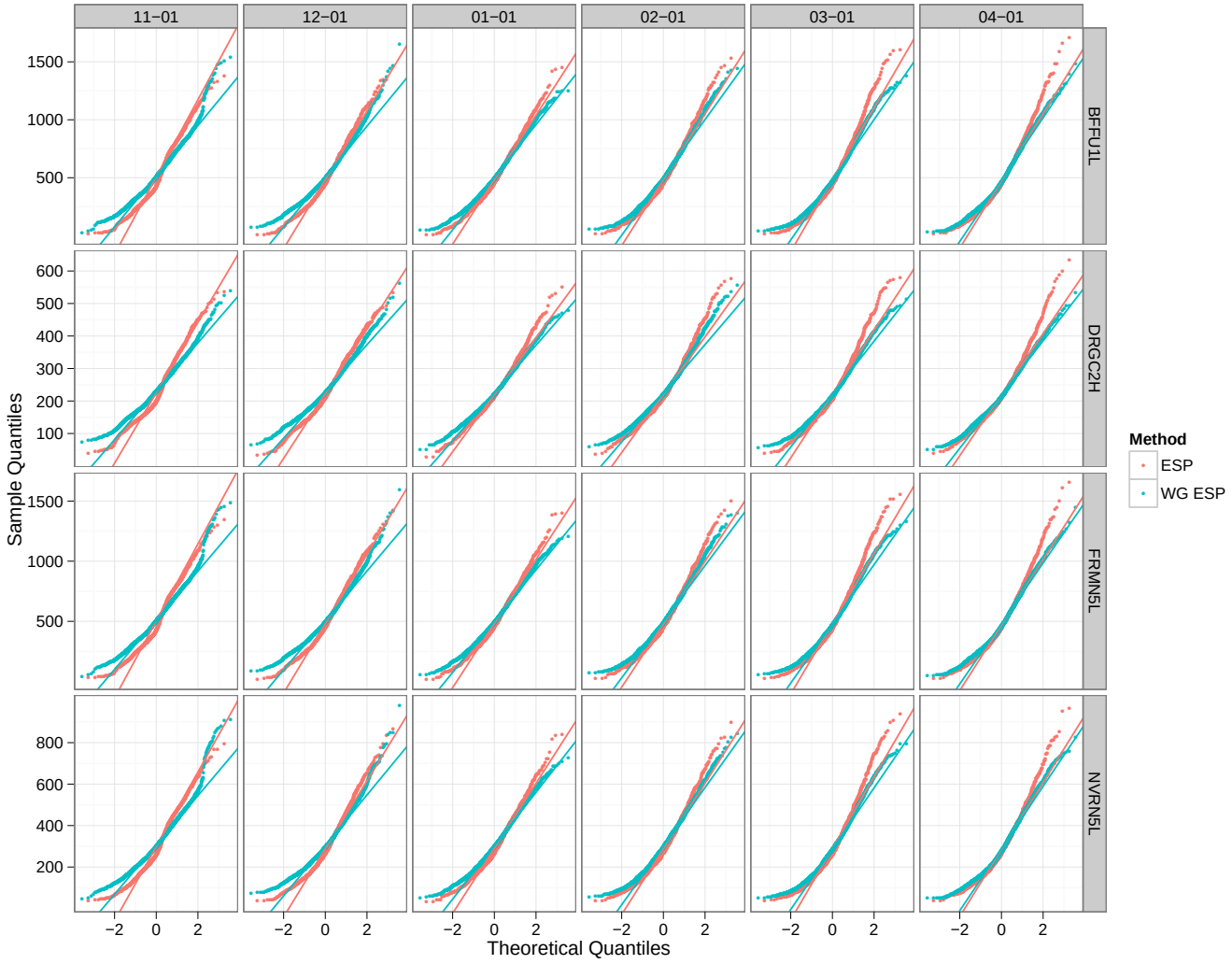


Figure A.23: Same as Figure A.21 for June to July

Table A.1: P-values from two-sided t-tests of April to May runoff

	Gage	Lead	p.value
1	BFFU1L	11-01	0.0068
2	DRGC2H	11-01	0.0000
3	FRMN5L	11-01	0.0006
4	NVRN5L	11-01	0.0007
5	BFFU1L	12-01	0.0304
6	DRGC2H	12-01	0.0001
7	FRMN5L	12-01	0.0046
8	NVRN5L	12-01	0.0070
9	BFFU1L	01-01	0.2486
10	DRGC2H	01-01	0.1034
11	FRMN5L	01-01	0.1066
12	NVRN5L	01-01	0.0452
13	BFFU1L	02-01	0.4928
14	DRGC2H	02-01	0.0702
15	FRMN5L	02-01	0.2983
16	NVRN5L	02-01	0.3188
17	BFFU1L	03-01	0.5253
18	DRGC2H	03-01	0.6375
19	FRMN5L	03-01	0.7315
20	NVRN5L	03-01	0.6674
21	BFFU1L	04-01	0.6889
22	DRGC2H	04-01	0.9870
23	FRMN5L	04-01	0.8013
24	NVRN5L	04-01	0.8155

Table A.2: P-values from two-sided t-tests of June to July runoff

	Gage	Lead	p.value
1	BFFU1L	11-01	0.0311
2	DRGC2H	11-01	0.0383
3	FRMN5L	11-01	0.0363
4	NVRN5L	11-01	0.0173
5	BFFU1L	12-01	0.0168
6	DRGC2H	12-01	0.0299
7	FRMN5L	12-01	0.0185
8	NVRN5L	12-01	0.0097
9	BFFU1L	01-01	0.0749
10	DRGC2H	01-01	0.4747
11	FRMN5L	01-01	0.0919
12	NVRN5L	01-01	0.0154
13	BFFU1L	02-01	0.0967
14	DRGC2H	02-01	0.2585
15	FRMN5L	02-01	0.1165
16	NVRN5L	02-01	0.0428
17	BFFU1L	03-01	0.7657
18	DRGC2H	03-01	0.5508
19	FRMN5L	03-01	0.6961
20	NVRN5L	03-01	0.9404
21	BFFU1L	04-01	0.7031
22	DRGC2H	04-01	0.3163
23	FRMN5L	04-01	0.6236
24	NVRN5L	04-01	0.9995

A.2 Conditional

Similar conclusions can be drawn from the sub-season PDFs as discussed in Section 3.4.3. Figures A.28 and A.29 demonstrate that leaving out the forecasted year decreases the peak probabilities. This difference is stronger for 2005 than 2006. As established in Section 3.4.3, the nomenclature now changes for comparison purposes where 'WG ESP' refers to unconditional and 'condWG ESP' refers to conditional (for figures, not tables). Unlike 2005 RPSS for April to July (Figure 3.17a), Figure A.30a shows ESP having some negative values (but not condWG ESP). June to July bounces back to the April to July pattern. Conversely, ESP (but not condWG ESP) has brief instances of positive RPSS in A.30b. The slope and values of sub-season quantiles in Figures A.33 to A.37 may have to do with the differences in magnitudes in Figure A.1. Comparing Table A.3 and A.4, June to July has larger p-values for both years. Figures A.38 and A.40 do not appear to offer additional insight to their April to July counterpart. June to July 2006 (Figure A.41) shows WG ESP having the worst performace for three gages for q10 and for all leads except Nov for q90.

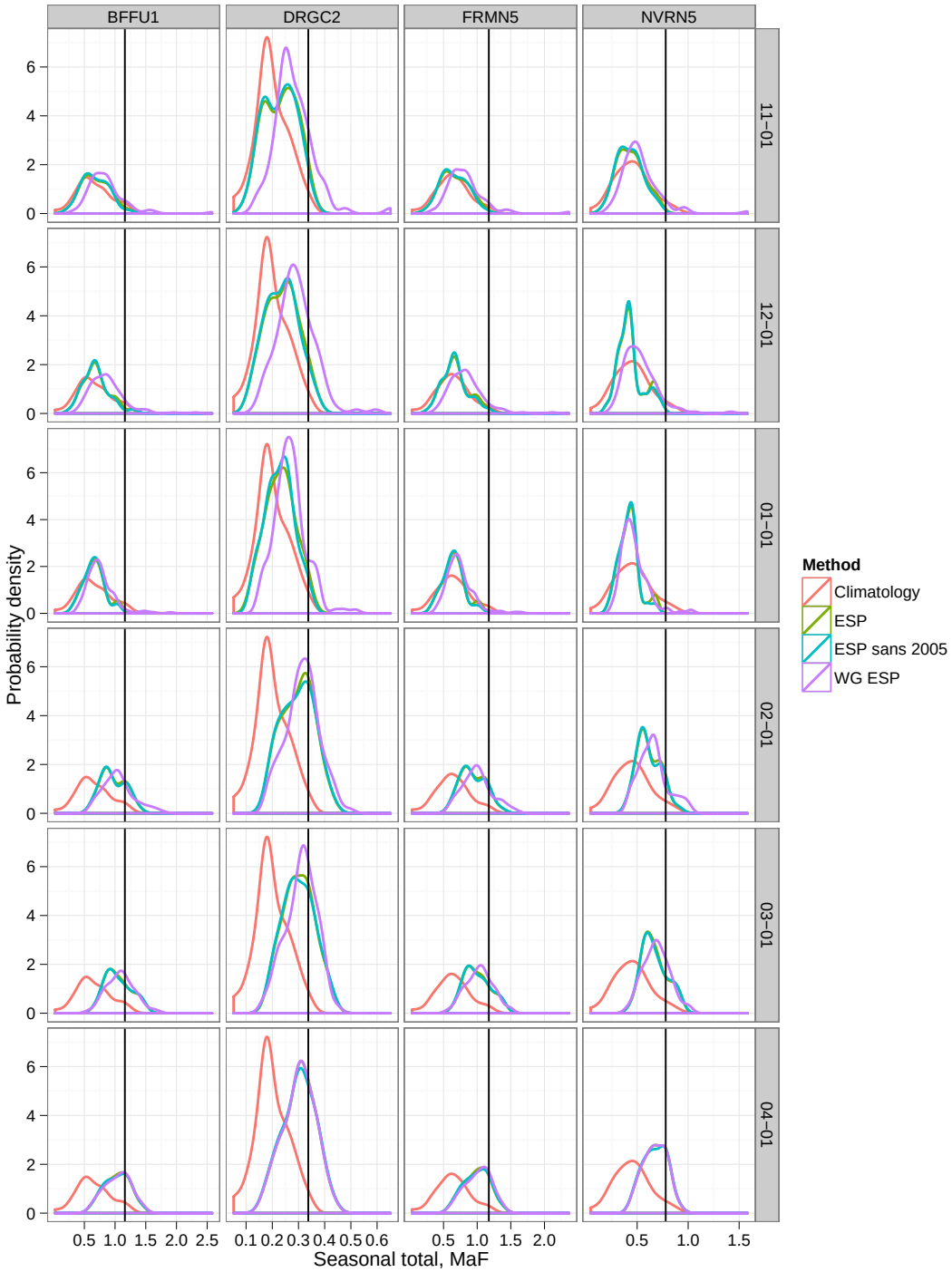


Figure A.24: April to May PDFs for 2005 runoff with vertical line showing 2005 value

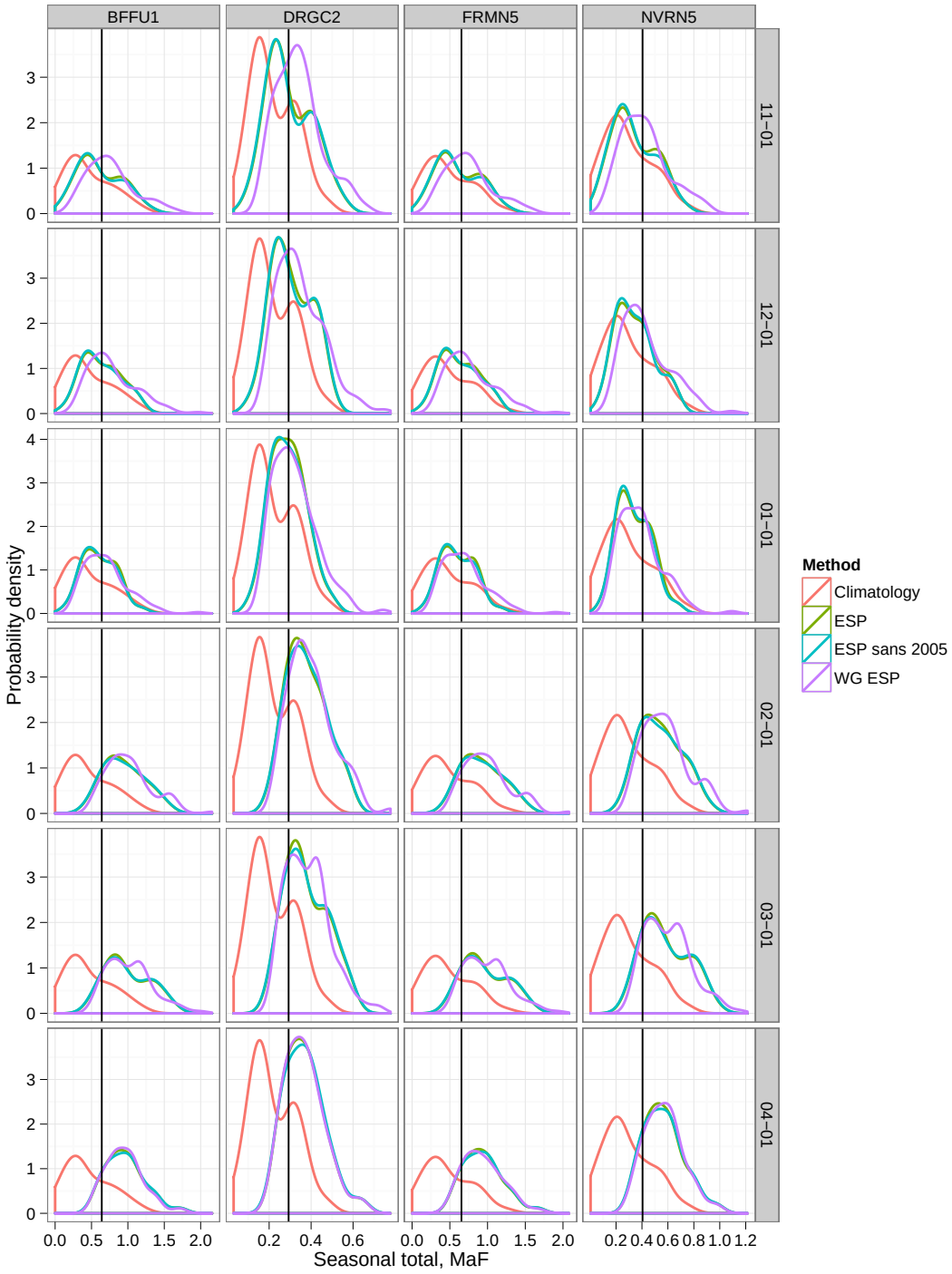


Figure A.25: June to July PDFs for 2005 runoff with vertical line showing 2005 value

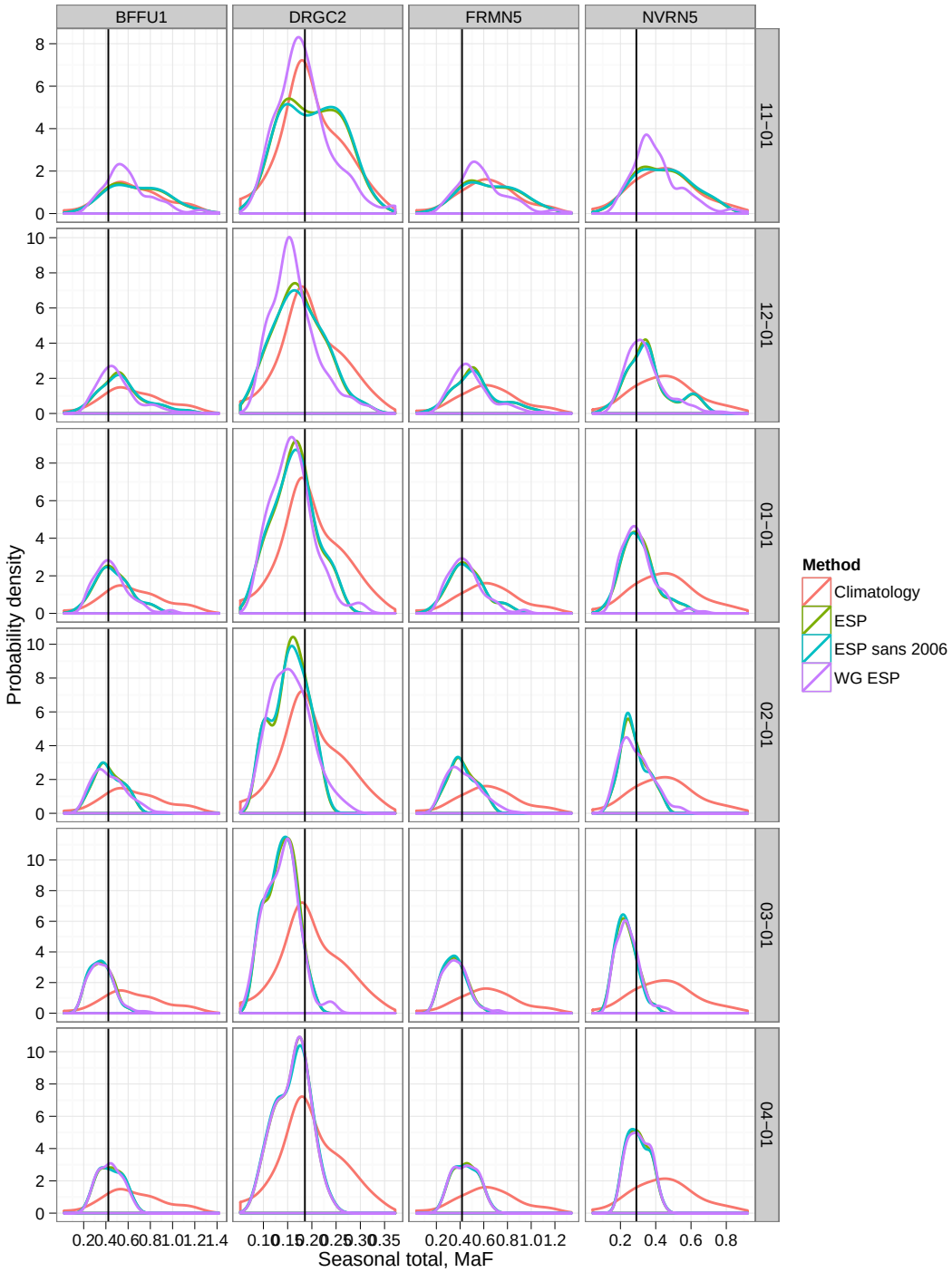


Figure A.26: Same as Figure A.24, but for 2006

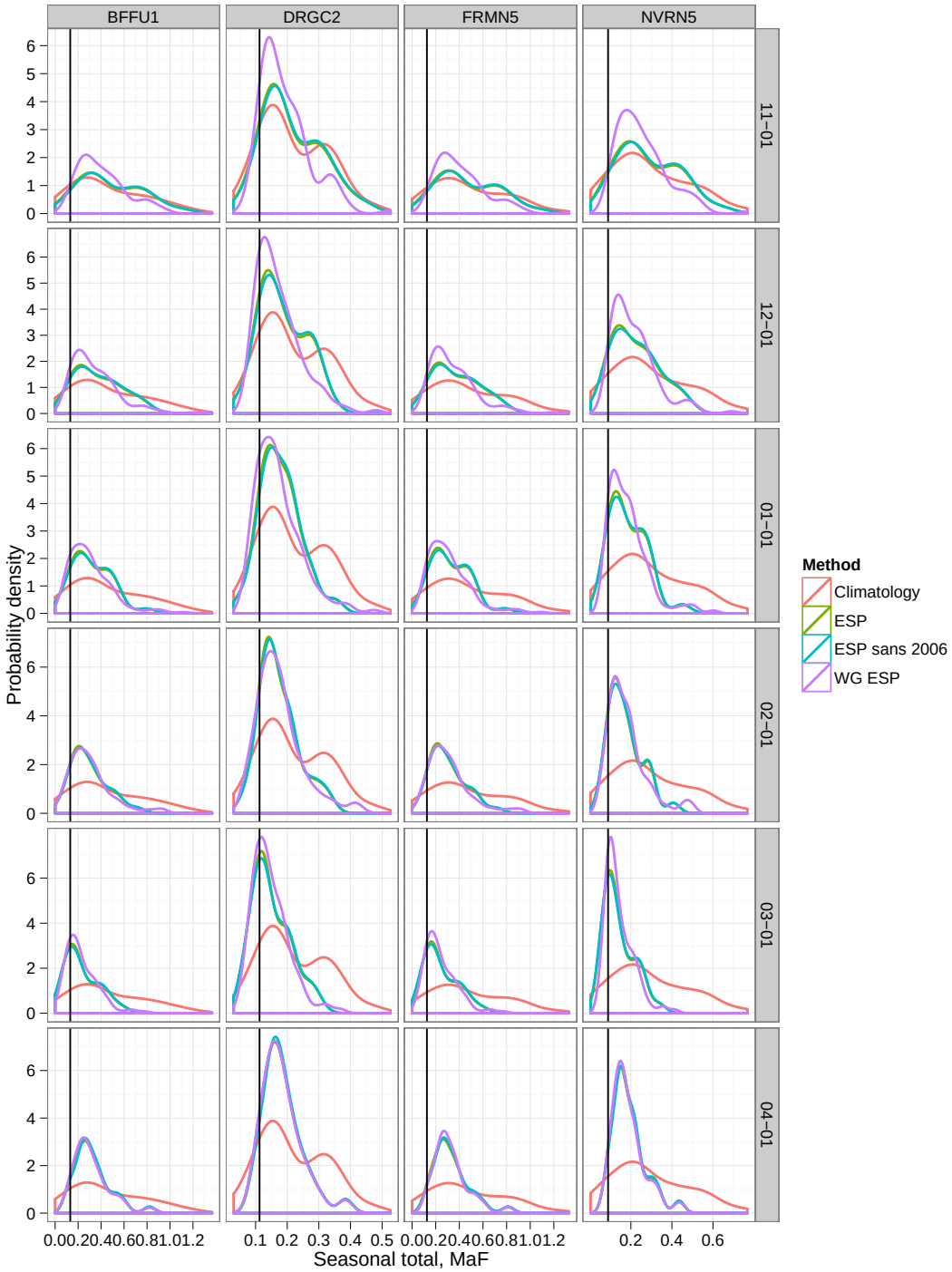


Figure A.27: Same as Figure A.25, but for 2006

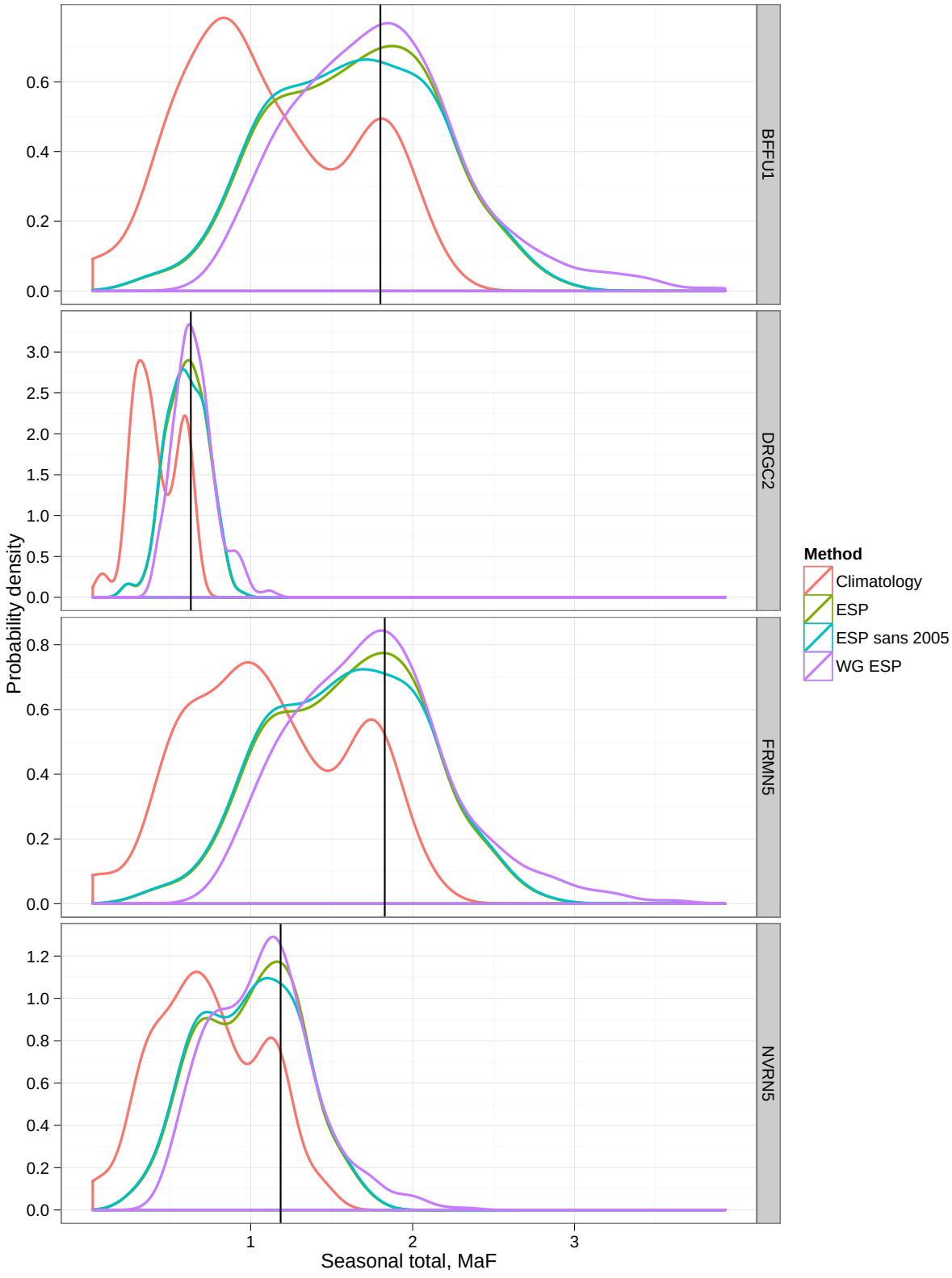


Figure A.28: April to July PDFs for 2005 runoff, not separated by lead time

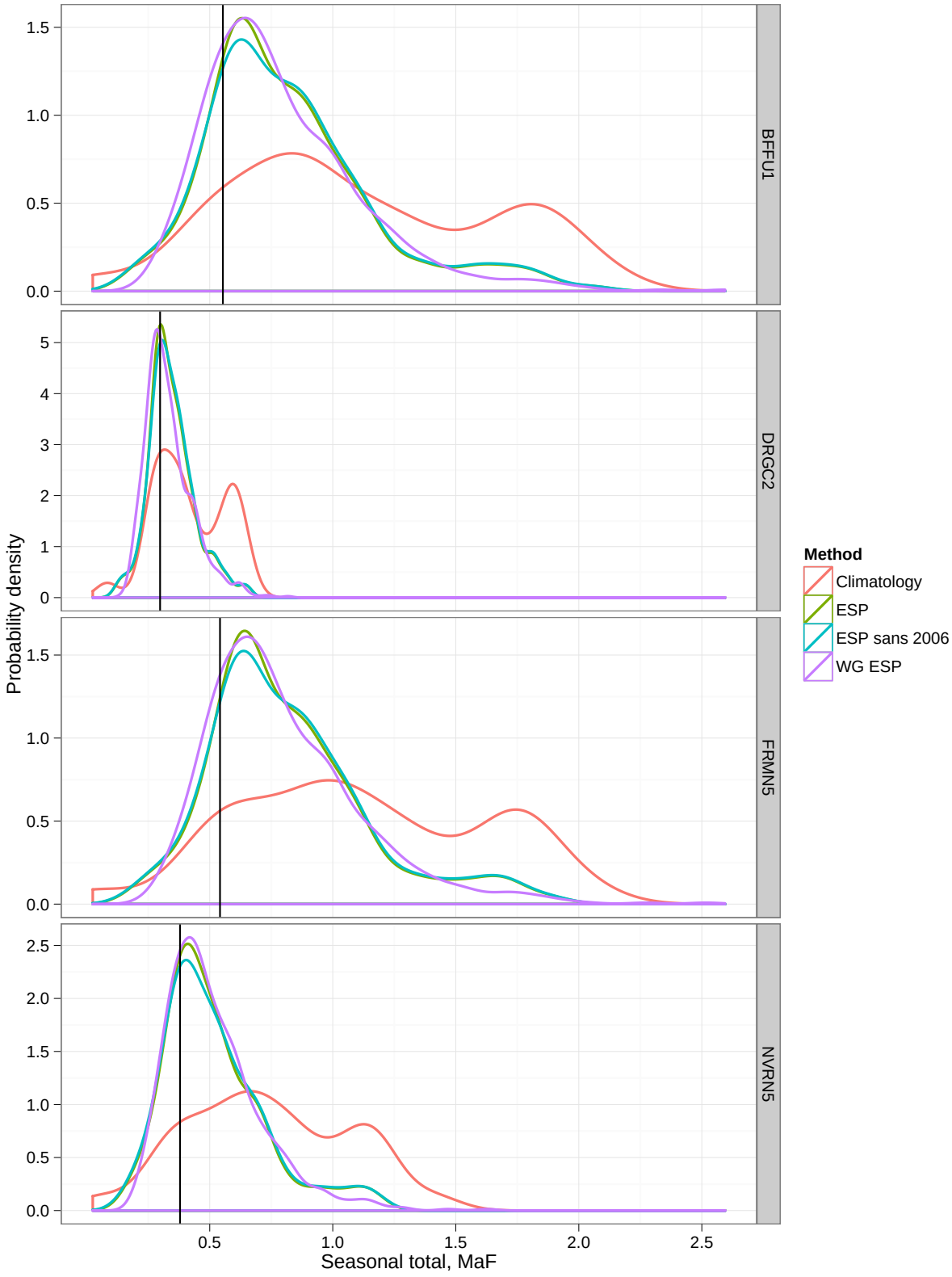
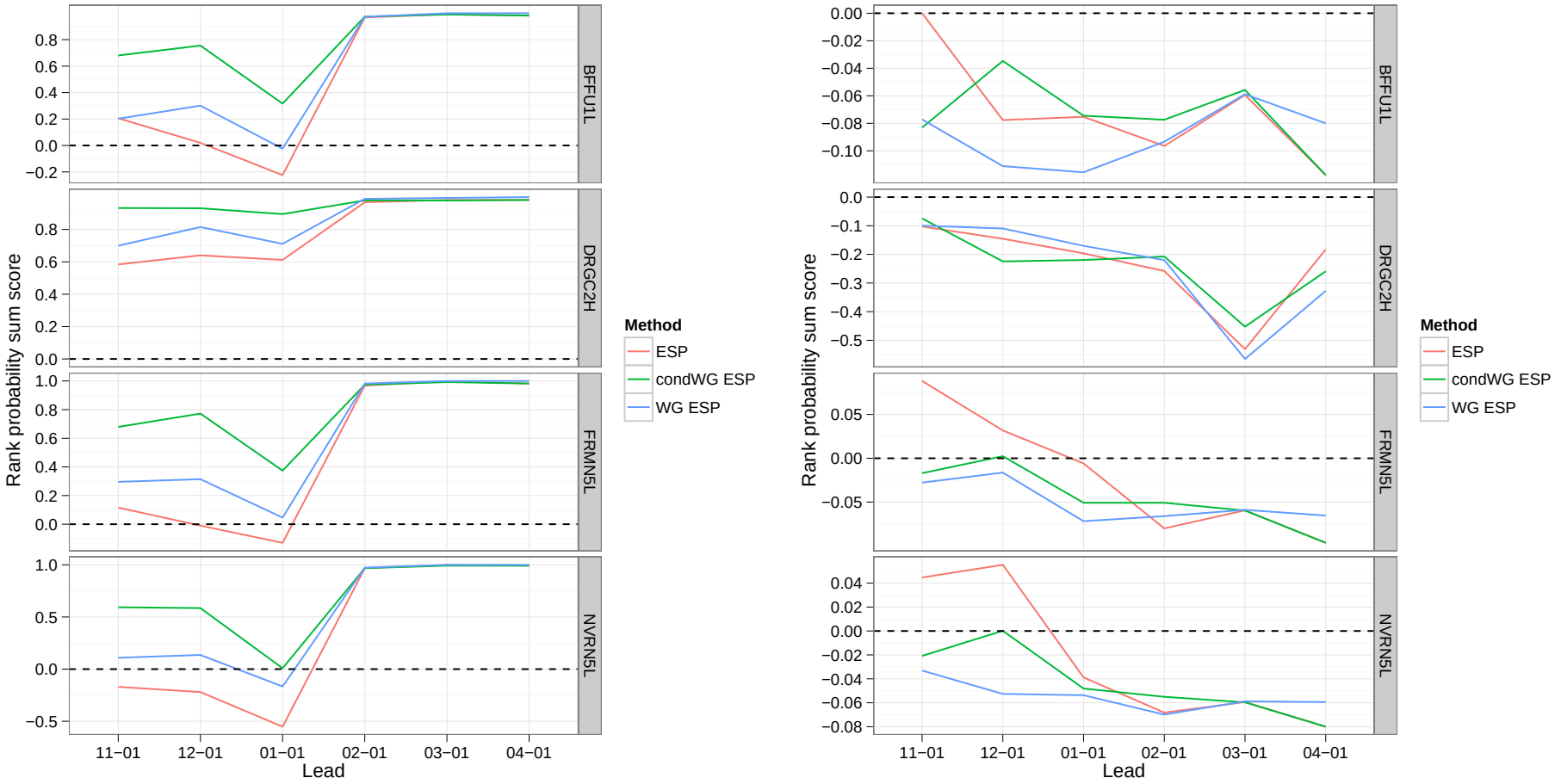


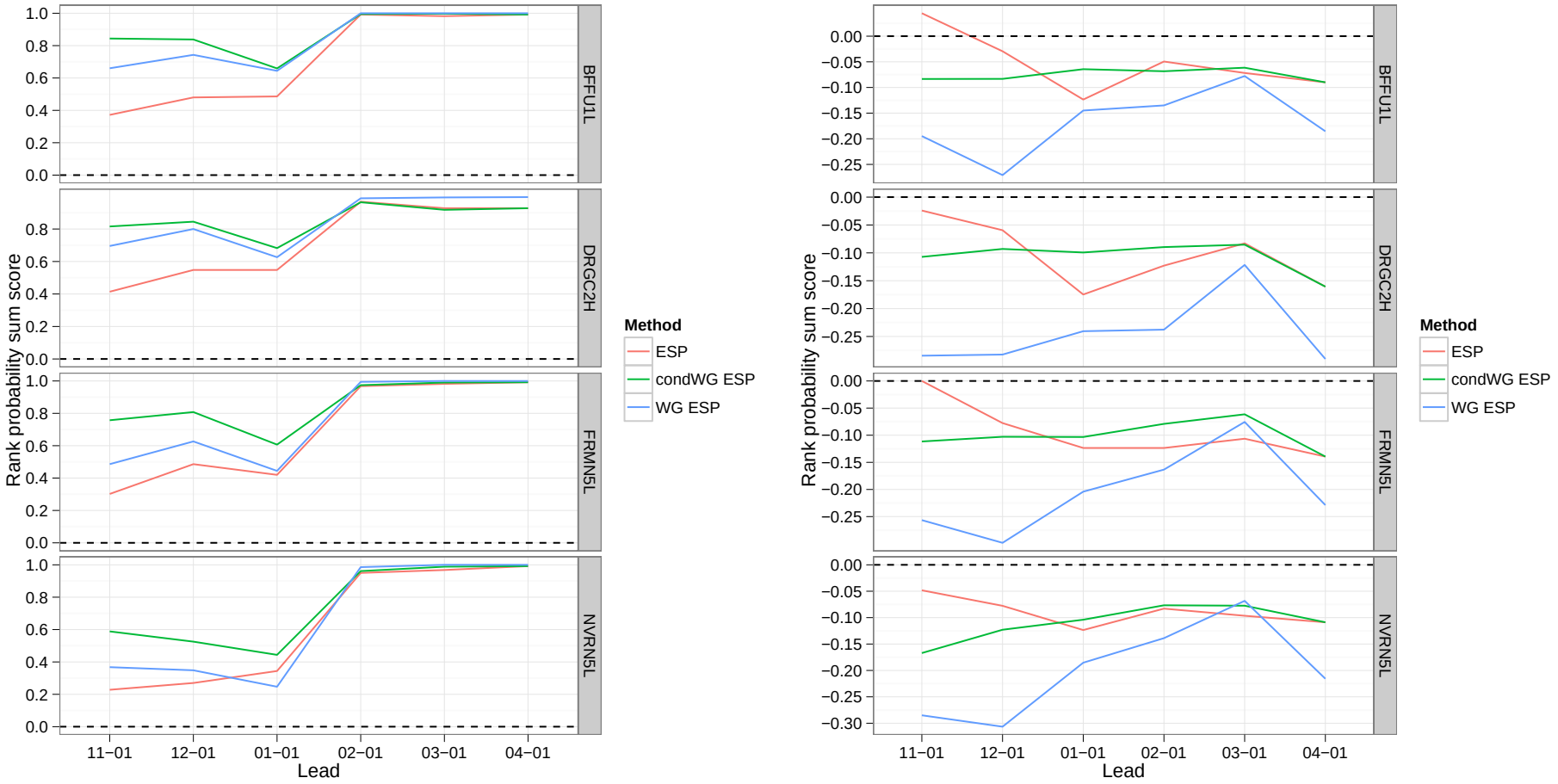
Figure A.29: Same as Figure A.28, but for 2006



(a) 2005

(b) 2006

Figure A.30: April to May RPSS for the conditional years



(a) 2005

(b) 2006

Figure A.31: June to July RPSS for the conditional years

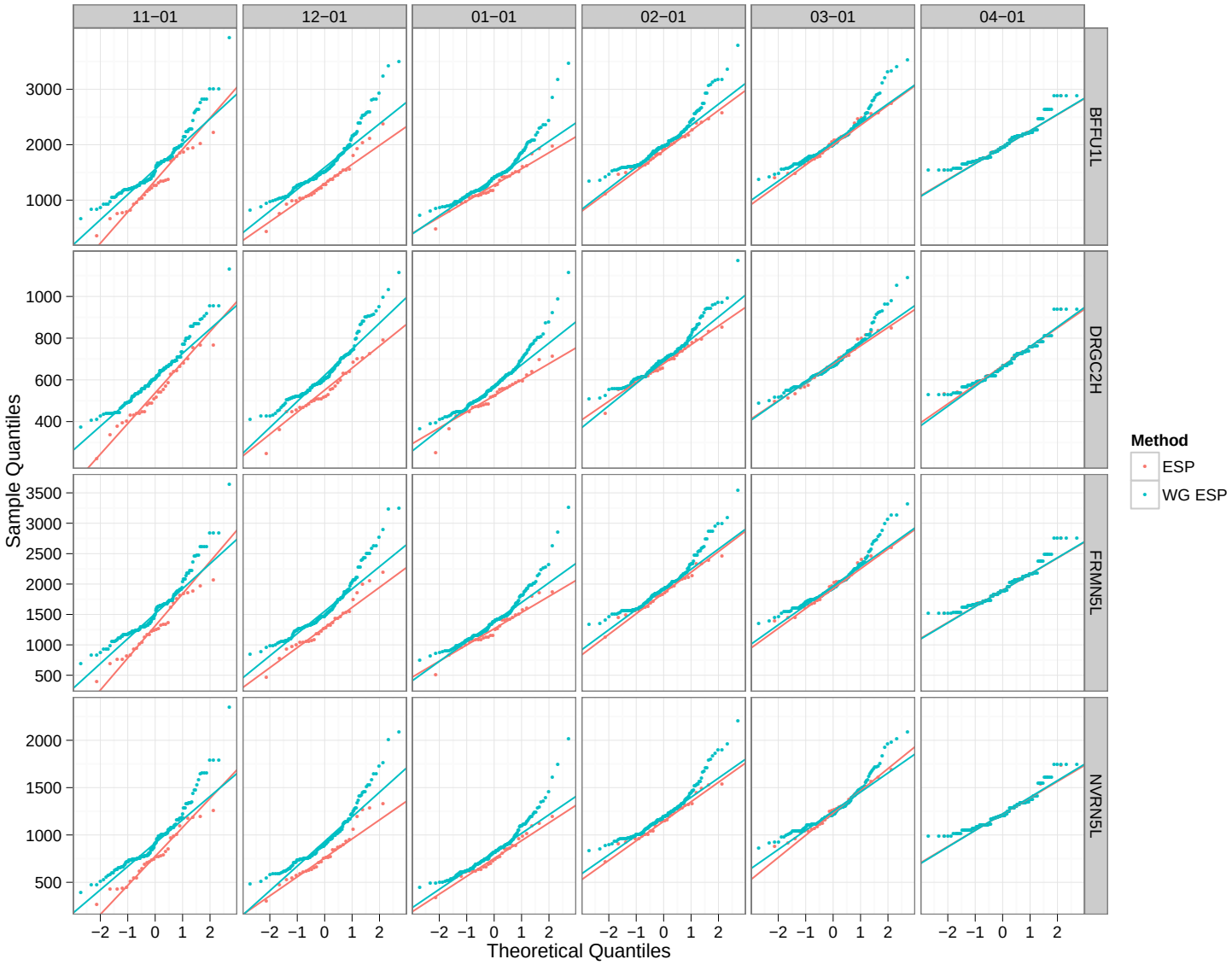


Figure A.32: April to July 2005 QQ

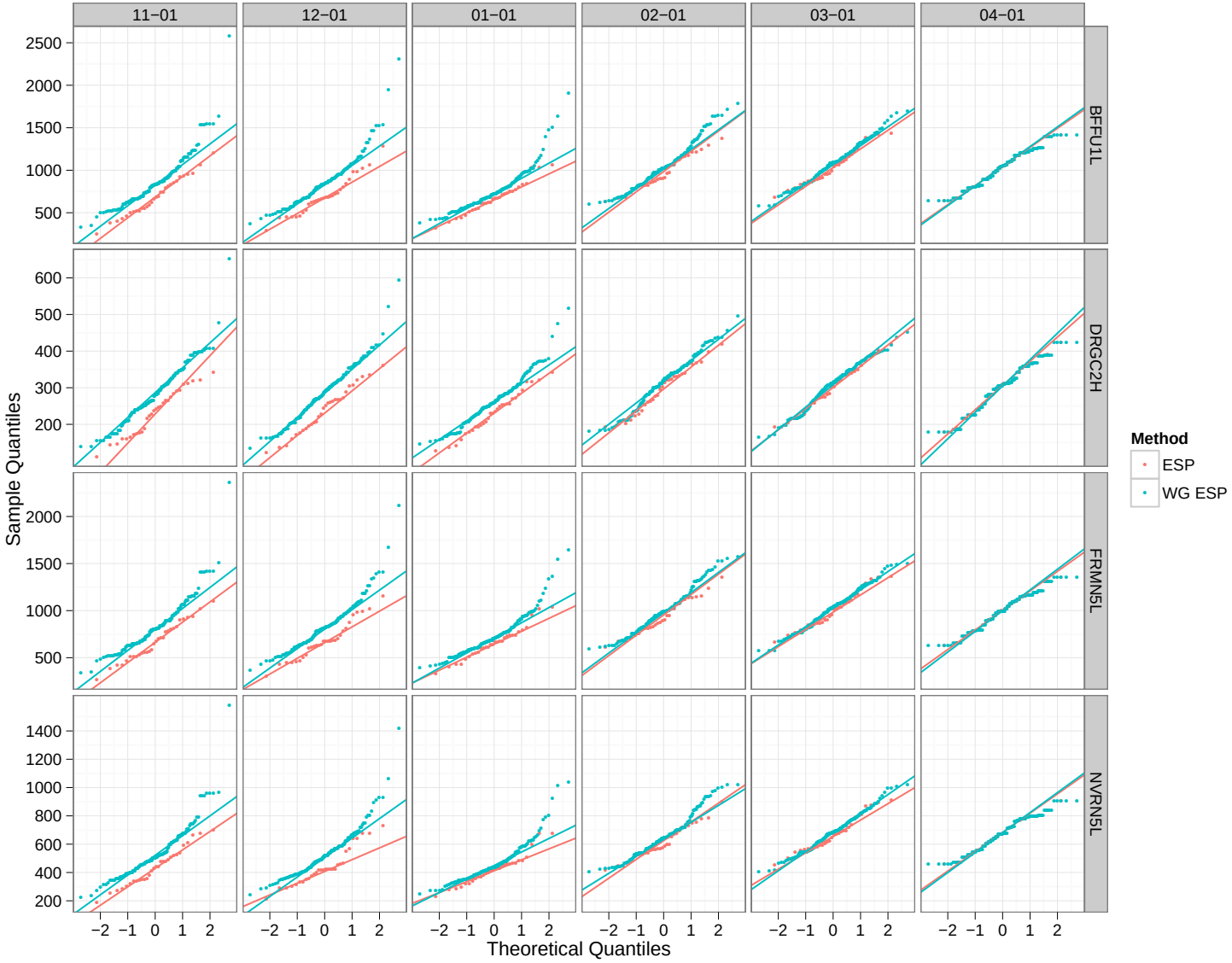


Figure A.33: April to May 2005 QQ

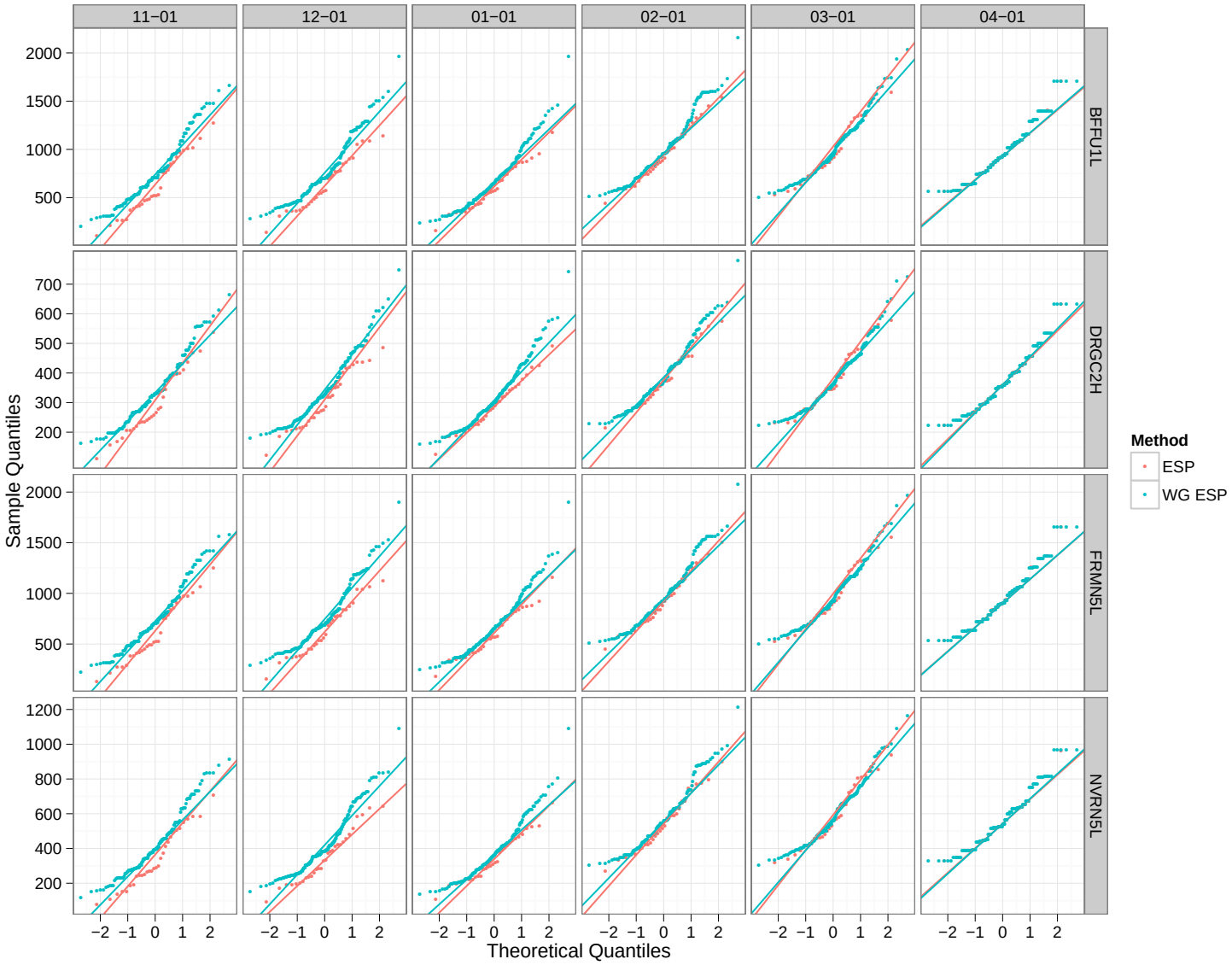


Figure A.34: June to July 2005 QQ

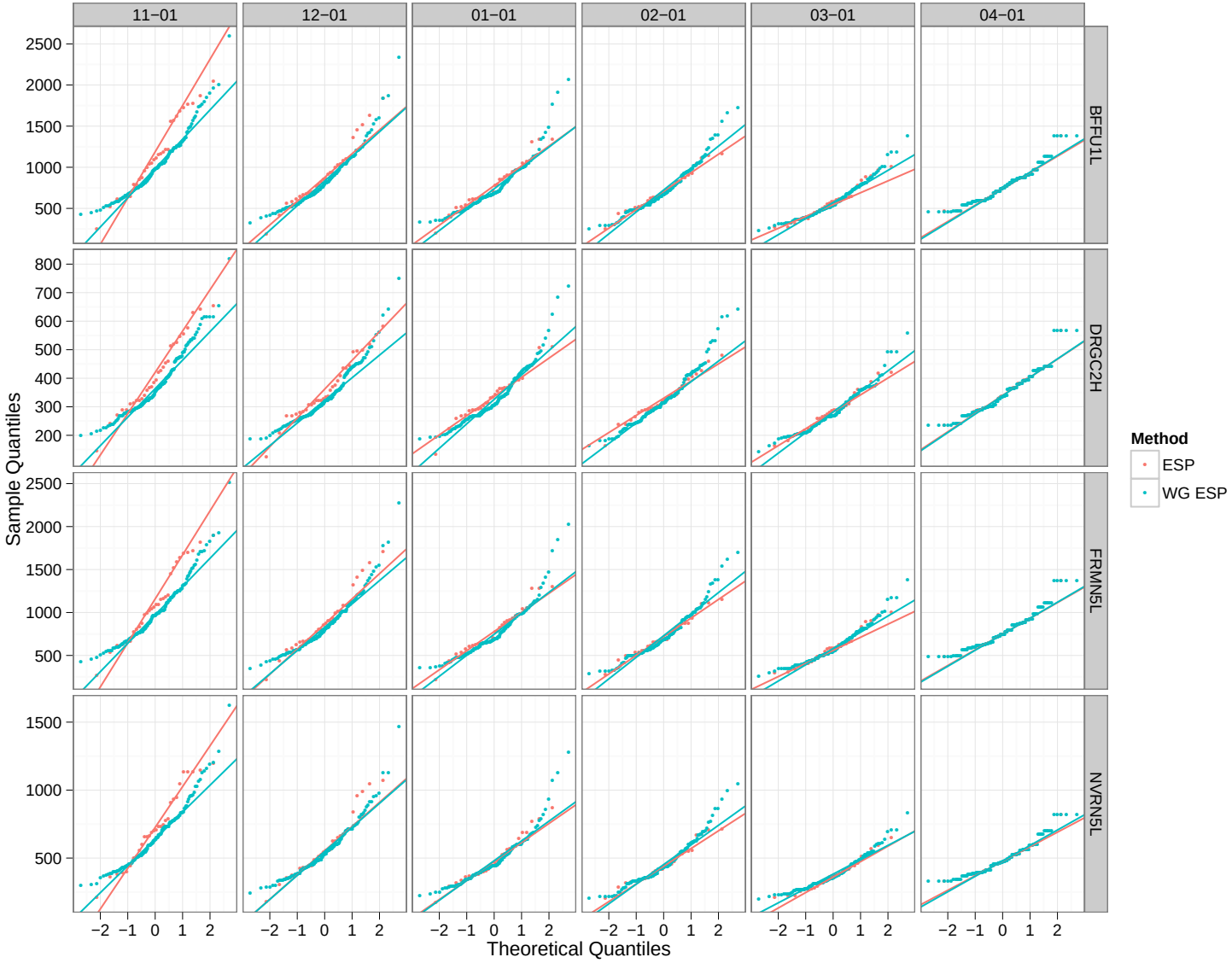


Figure A.35: April to July 2006 QQ

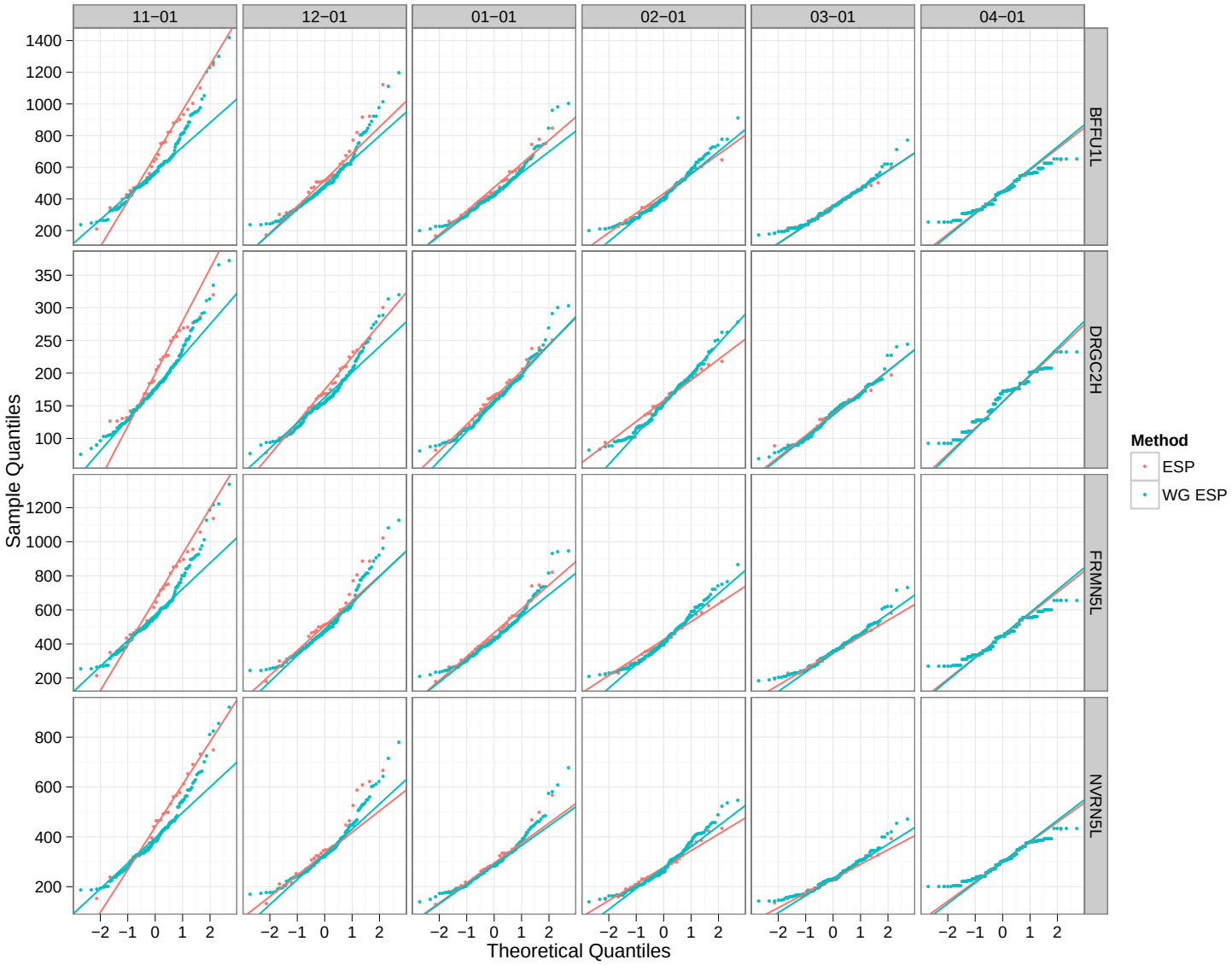


Figure A.36: April to May 2006 QQ

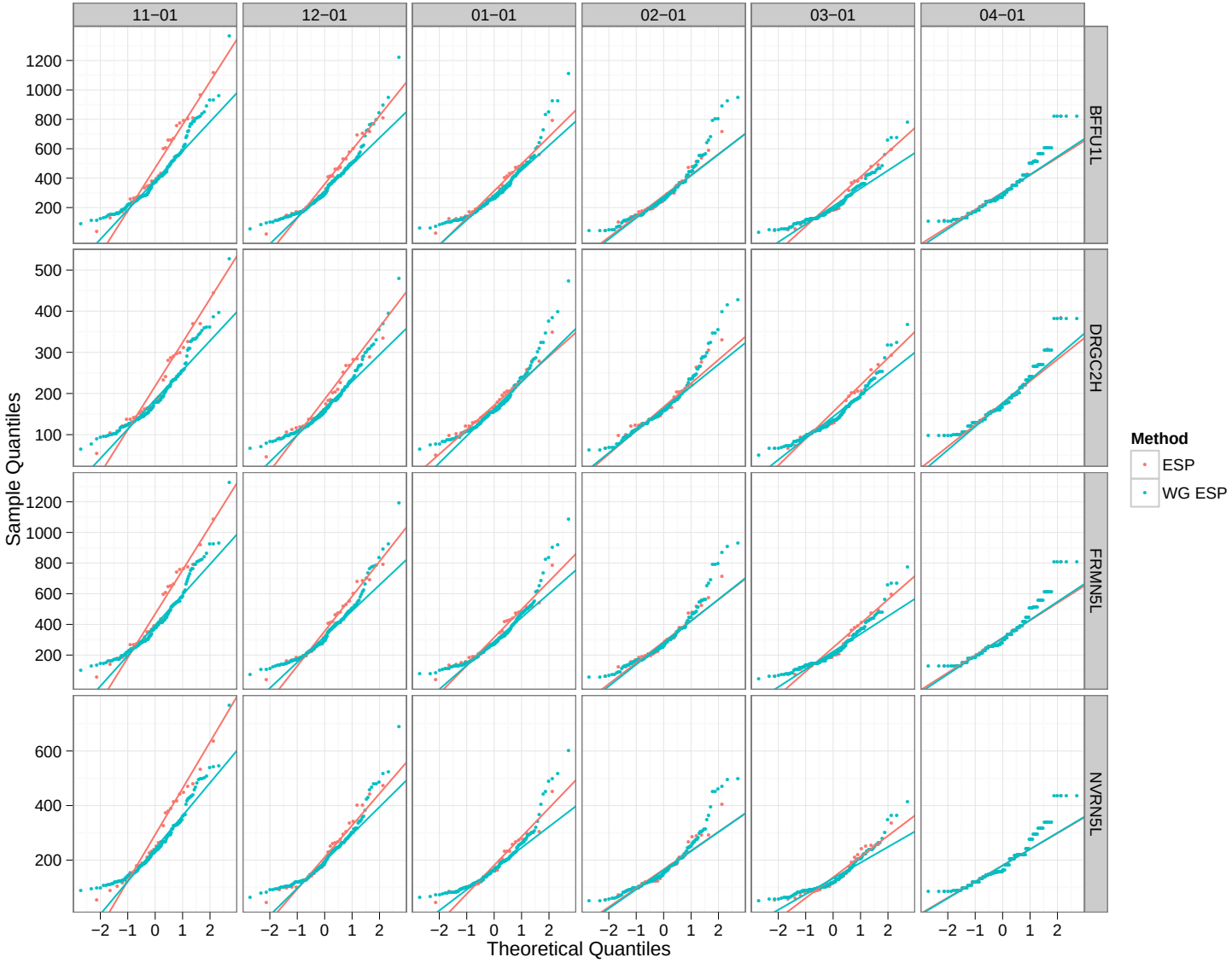


Figure A.37: June to July 2006 QQ

Table A.3: P-values from Wilcoxon rank sum test for 2005 and 2006 April to May runoff

	Gage	Lead	2005	2006
1	BFFU1L	11-01	0.002	0.150
2	DRGC2H	11-01	0.001	0.146
3	FRMN5L	11-01	0.002	0.200
4	NVRN5L	11-01	0.002	0.227
5	BFFU1L	12-01	0.001	0.211
6	DRGC2H	12-01	0.001	0.224
7	FRMN5L	12-01	0.001	0.288
8	NVRN5L	12-01	0.001	0.454
9	BFFU1L	01-01	0.054	0.409
10	DRGC2H	01-01	0.011	0.384
11	FRMN5L	01-01	0.049	0.447
12	NVRN5L	01-01	0.111	0.678
13	BFFU1L	02-01	0.189	0.781
14	DRGC2H	02-01	0.220	0.781
15	FRMN5L	02-01	0.196	0.892
16	NVRN5L	02-01	0.205	0.922
17	BFFU1L	03-01	0.476	0.971
18	DRGC2H	03-01	0.491	0.840
19	FRMN5L	03-01	0.479	0.931
20	NVRN5L	03-01	0.508	0.683
21	BFFU1L	04-01	0.925	0.879
22	DRGC2H	04-01	0.804	0.775
23	FRMN5L	04-01	0.986	0.925
24	NVRN5L	04-01	0.887	0.971

Table A.4: P-values from Wilcoxon rank sum test for 2005 and 2006 June to July runoff

	Gage	Lead	2005	2006
1	BFFU1L	11-01	0.020	0.291
2	DRGC2H	11-01	0.044	0.240
3	FRMN5L	11-01	0.026	0.316
4	NVRN5L	11-01	0.018	0.414
5	BFFU1L	12-01	0.039	0.452
6	DRGC2H	12-01	0.071	0.341
7	FRMN5L	12-01	0.037	0.475
8	NVRN5L	12-01	0.024	0.759
9	BFFU1L	01-01	0.256	0.534
10	DRGC2H	01-01	0.243	0.438
11	FRMN5L	01-01	0.267	0.537
12	NVRN5L	01-01	0.284	0.753
13	BFFU1L	02-01	0.465	0.925
14	DRGC2H	02-01	0.556	0.793
15	FRMN5L	02-01	0.501	0.904
16	NVRN5L	02-01	0.505	0.898
17	BFFU1L	03-01	0.861	1.000
18	DRGC2H	03-01	0.968	0.728
19	FRMN5L	03-01	0.828	0.998
20	NVRN5L	03-01	0.813	0.734
21	BFFU1L	04-01	0.834	0.790
22	DRGC2H	04-01	0.760	0.790
23	FRMN5L	04-01	0.849	0.804
24	NVRN5L	04-01	0.895	0.917

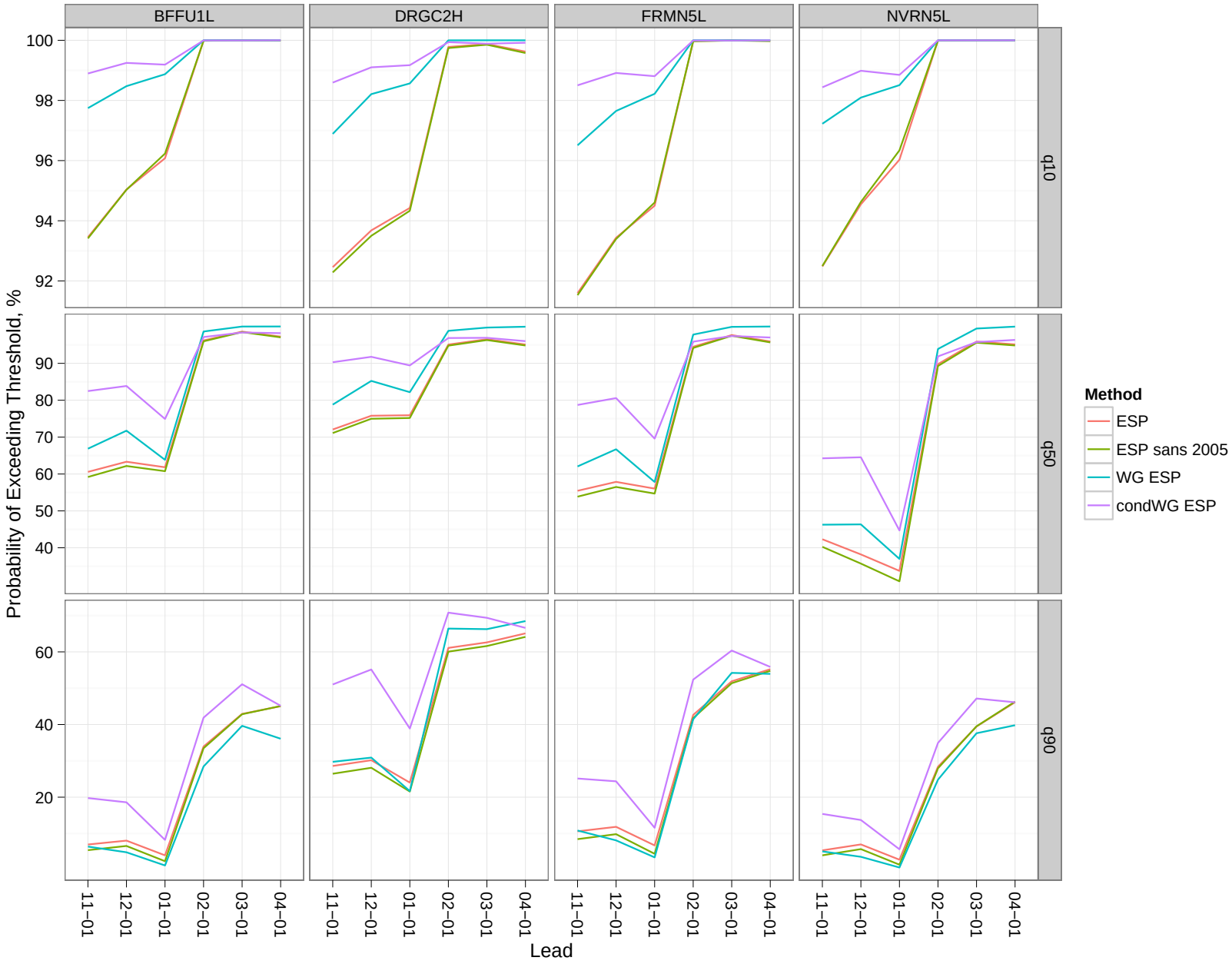


Figure A.38: Shifts in 2005 exceedance probabilities (April to May)

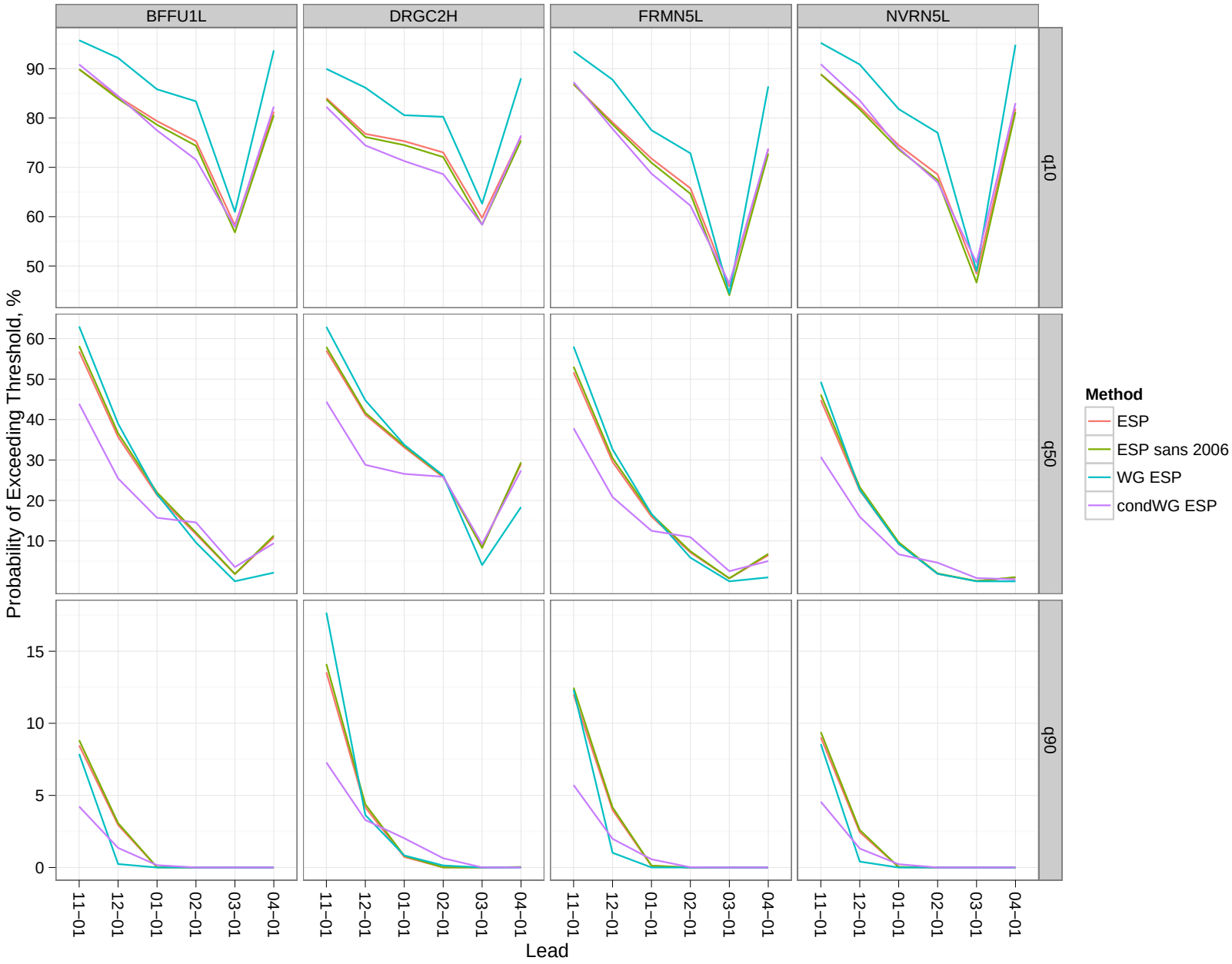


Figure A.39: Same as Figure A.38, but for 2006

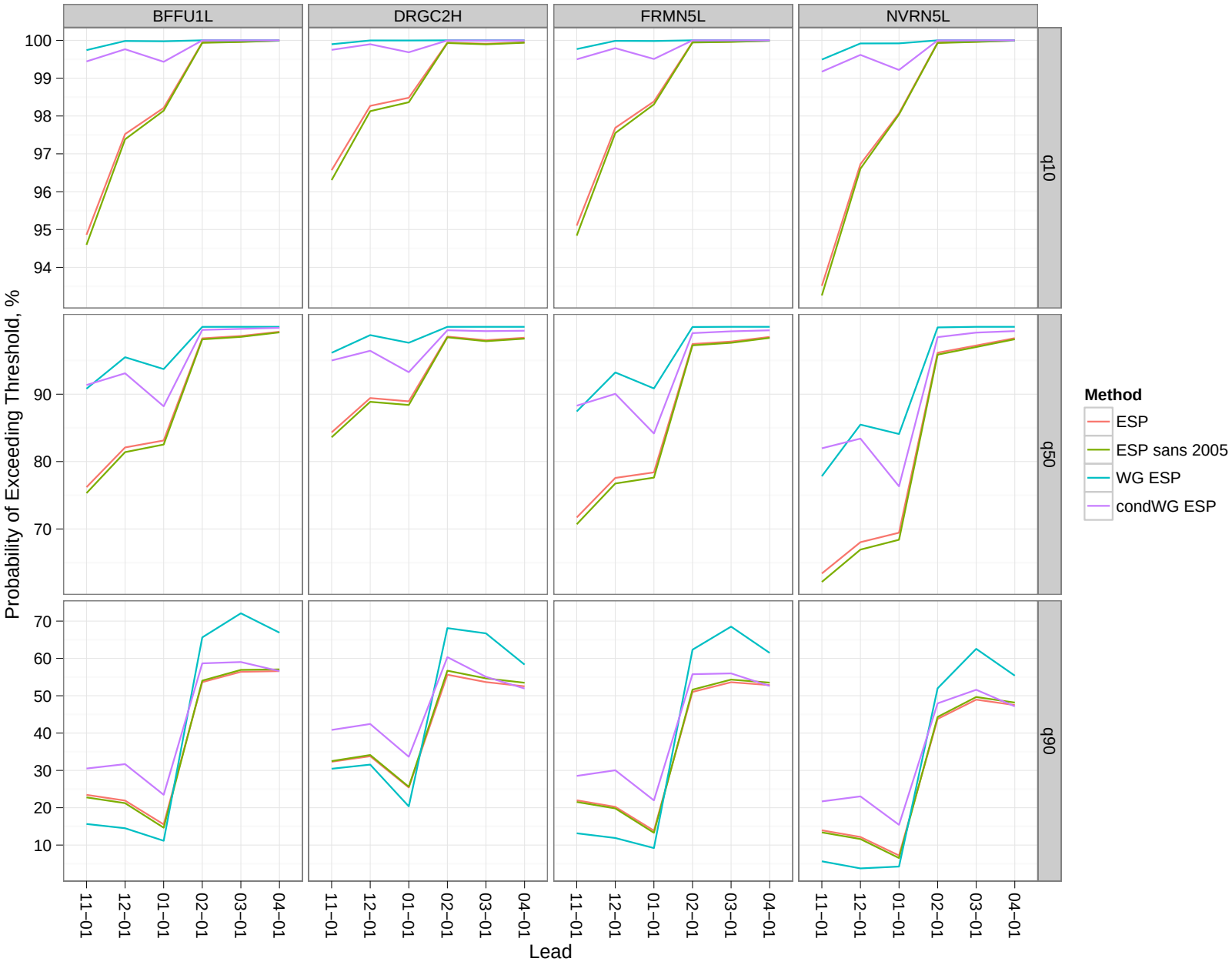


Figure A.40: Shifts in 2005 exceedance probabilities (June to July)

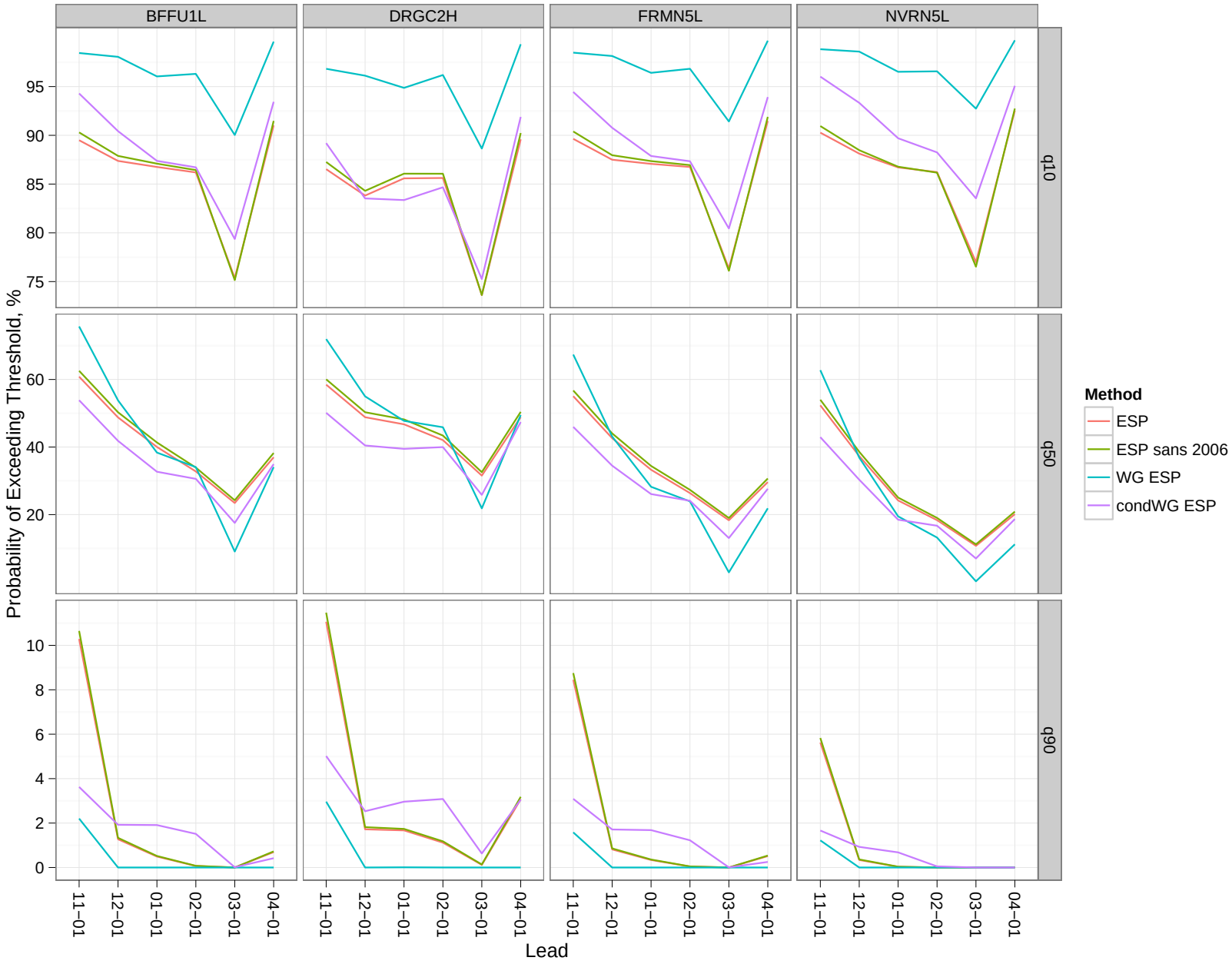


Figure A.41: Same as Figure A.40, but for 2006

Appendix B

IRI Forecasts and Resulting Weather Generator Output

This section displays the IRI climate forecasts used for weather generator, as well as basic distributional statistics of the resulting weather sequences.

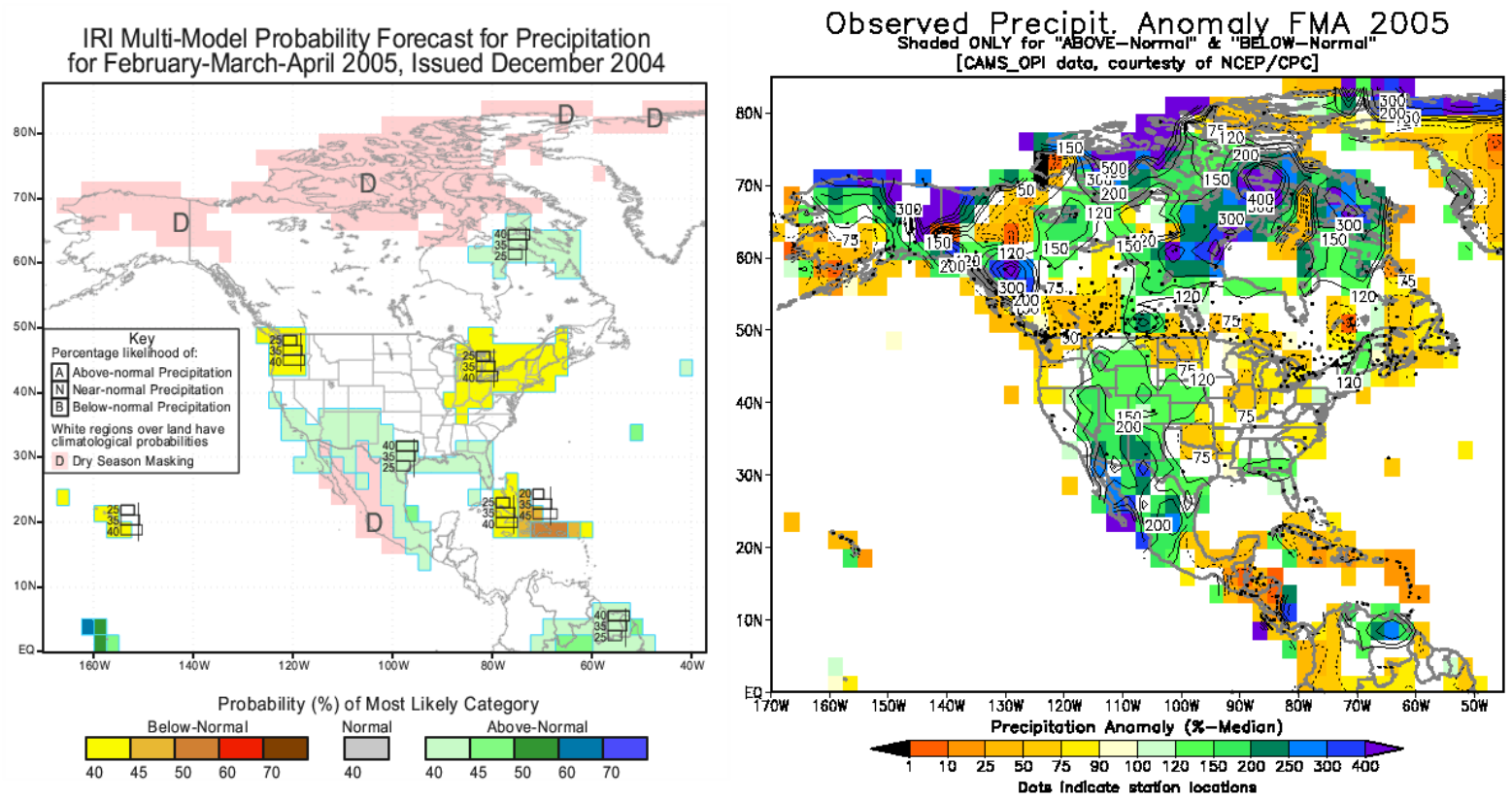


Figure B.1: 2005 IRI winter precipitation forecast

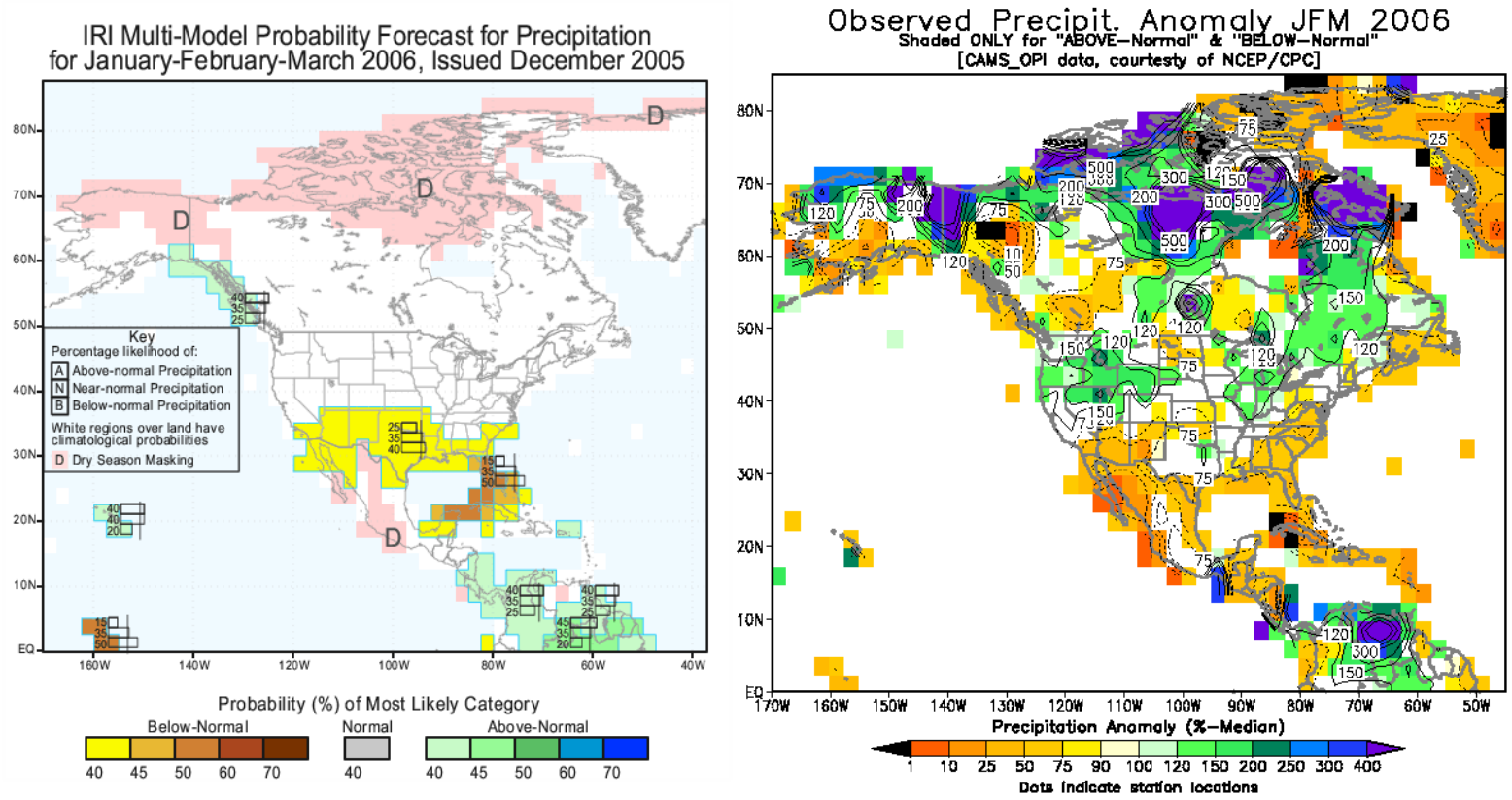
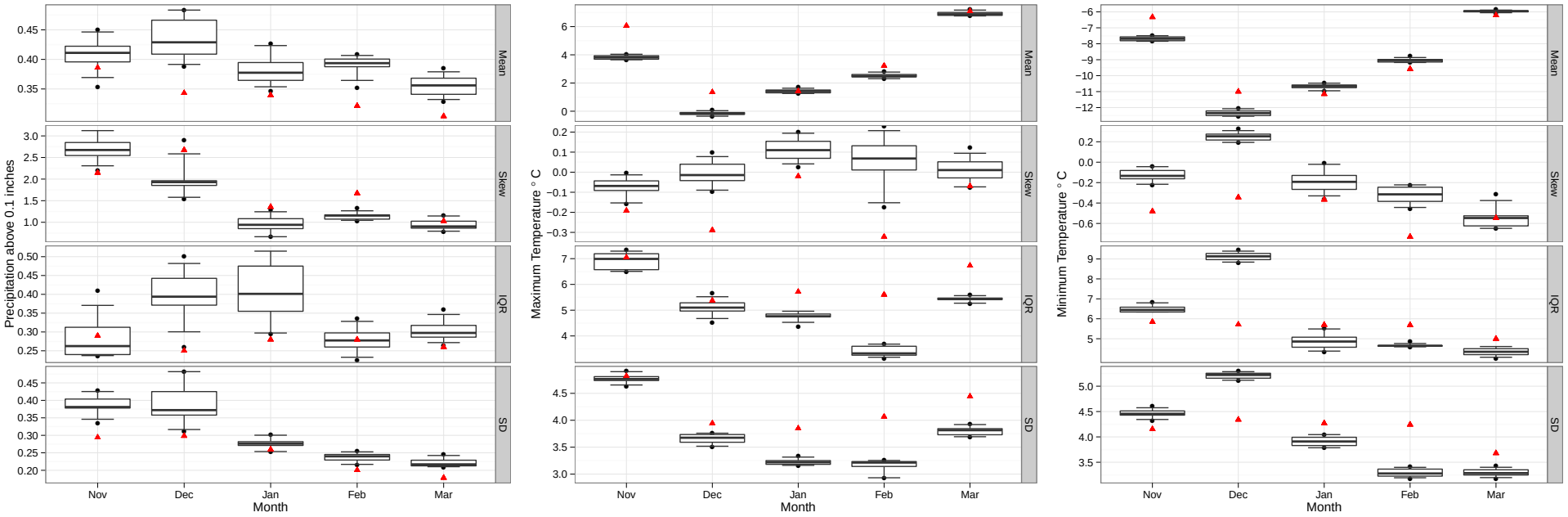
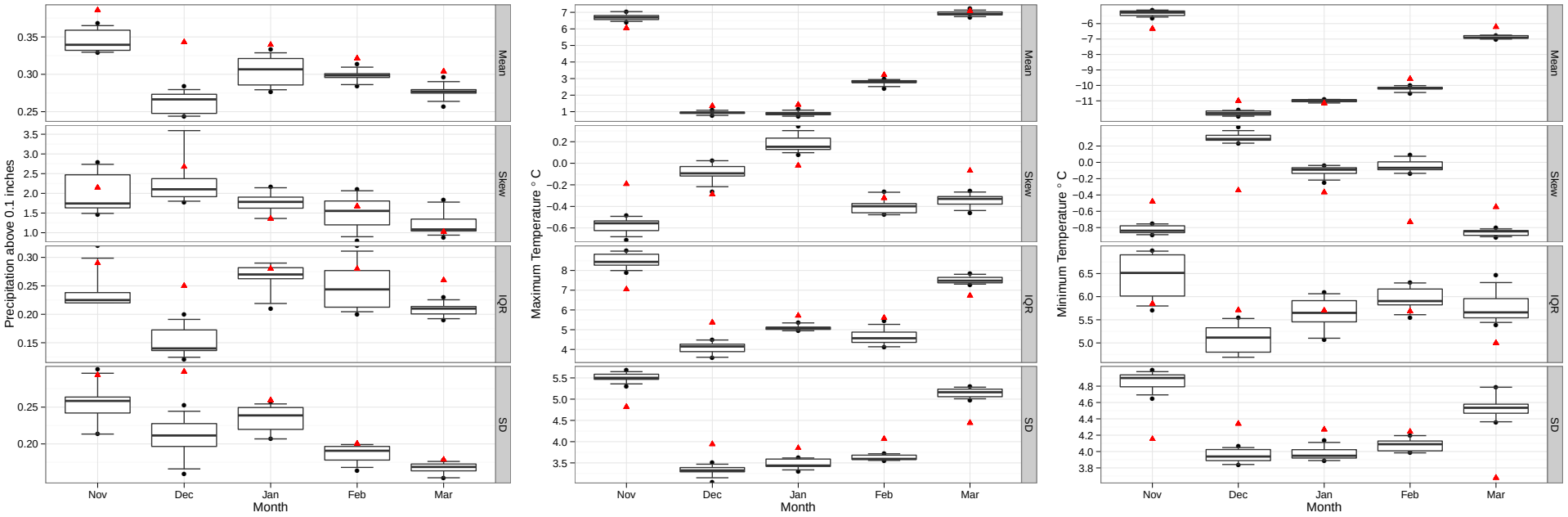


Figure B.2: 2006 IRI winter precipitation forecast



(a) Precipitation (b) Max temperature (c) Min temperature

Figure B.3: Distributional statistics of daily weather variables for ANB of 40:35:25



(a) Precipitation (b) Max temperature (c) Min temperature

Figure B.4: Distributional statistics of daily weather variables for ANB of 25:35:40