Spring 1-1-2016

Stochastic Space-Time Modeling for Agricultural Decision Support in the Argentine Pampas

Andrew Paul Verdin
University of Colorado at Boulder, andrew.verdin@colorado.edu

Stochastic space-time modeling for agricultural decision support in the Argentine Pampas

by

Andrew P. Verdin

B.A., University of Minnesota, 2008

M.S., University of Colorado, 2013

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

Doctor of Philosophy

Department of Civil, Environmental, and Architectural Engineering

2016

This thesis entitled:
Stochastic space-time modeling for agricultural decision support in the Argentine Pampas
written by Andrew P. Verdin
has been approved for the Department of Civil, Environmental, and Architectural Engineering

_____
Prof. Balaji Rajagopalan

_____
Prof. Guillermo Podestá

_____
Prof. William Kleiber

_____
Prof. Ben Livneh

_____
Prof. Joseph Kasprzyk

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.

Verdin, Andrew P. (Ph.D., Civil Engineering)

Stochastic space-time modeling for agricultural decision support in the Argentine Pampas

Thesis directed by Prof. Balaji Rajagopalan

This dissertation presents three statistical models and applies them to the predominantly rain-fed Argentine Pampas, one of the most productive agricultural regions in the world. The Argentine Pampas experienced an upward trend in annual precipitation since the 1960s; global soybean prices surged shortly thereafter, which provided an optimal combination of climate, economics, and technology, and motivated vast agricultural expansion to semi-arid regions. Annual precipitation totals have declined since the turn of the century, which begs the question: "Are the existing agricultural production systems viable in a drier future?" Stochastic weather generators have long been used to produce synthetic daily weather series to drive process based models, which in turn are used to assess likely impacts on climate-sensitive sectors of society, and to evaluate the outcomes of alternative adaptive actions. Unfortunately, many traditional approaches of stochastic weather generation are limited in their ability to generate space-time weather (i.e., at unobserved locations), or values outside the range of the historical record, which is particularly important for climate change applications in rural agricultural regions, such as the Argentine Pampas.

To this end, we developed a coupled stochastic weather generator (GLMGEN), which takes advantage of the flexibility of generalized linear models (GLMs) to model skewed and discrete variables (i.e., precipitation intensity and occurrence, respectively). Spatial process models estimate the GLM parameters in space to simulate at arbitrary locations, such as on a regular grid. Subsequent application of GLMGEN within a nonstationary context, such as climate change studies, is presented for the Salado *A* sub-basin of the

Argentine Pampas. The inclusion of large-scale climate indices as covariates enables the simulation of daily weather ensembles that exhibit the traits and trends of seasonal forecasts and climate model projections. Regional climate model, experiment RCP8.5, and two IRI seasonal forecasts are used to condition the output of GLMGEN, thus translating this coarse scale climate information into more salient information for decision makers. In addition, we present a Bayesian stochastic weather generator (BayGEN), which quantifies and preserves the uncertainty associated with all model parameters. Uncertainty will subsequently propagate to synthetic daily weather ensembles and their respective uses, such as to drive crop simulation and hydrologic models, properly quantifying risk for decision making and climate change adaptation strategies. Direct comparison of BayGEN with GLMGEN will illustrate the benefit of propagating this uncertainty to simulation space. Finally, a statistical space-time hierarchical metamodel for monthly actual evapotranspiration (ET) and monthly water table depth (WTD) was developed as a complementary tool for near real-time decision support. In the first level of hierarchy, ET is modeled as a function of climate and land use decision variables; the second level models WTD as a function of climate and predicted ET. The metamodel was conditioned on and validated by a calibrated hydrologic model (i.e., MIKE-SHE) for the Salado A sub-basin, and is shown to adequately capture the dominant mechanisms of spatial and temporal variability. Use of the metamodel with output from a weather generator, as well as with ensembles of different land uses, can identify regions of high risk by producing distributions of WTD and its response to climate and land use change scenarios.

Dedication

For Carsen.

Acknowledgements

A big thank you to everyone who has helped me in one way or another over the years. Balaji Rajagopalan and William Kleiber for being so positive and encouraging; Guillermo Podestá for writing the NSF grant that funded this research, and for always asking for clarification; Federico Bert for being amazingly patient with my incessant requests to run (and re-run) DSSAT; Angel Menéndez, Pablo Garcia, Santiago Rovere, and the rest of the Buenos Aires team. Ben Livneh and Joe Kasprzyk for agreeing to be on my committee, and for always asking the tough questions.

The biggest thank you must go to Carsen, my future wife, for not only being extraordinarily patient with me, but also for her help with maintaining my sanity in high stress times.

Contents

Chapter

1 Introduction 1

1.1 Background 1

1.2 Overview of the Pampas 4

1.3 Motivation 6

1.4 Dissertation outline 6

2 Coupled stochastic weather generation using spatial and generalized linear models 8

2.1 Introduction 8

2.2 Stochastic model 12

2.2.1 Estimation and modeling choices 14

2.3 Stochastic weather simulation in the Pampas 16

2.4 Discussion 23

3 A conditional stochastic weather generator for seasonal to multi-decadal simulations 25

3.1 Introduction 25

3.2 Study region and data 29

3.3 Methodology 31

3.3.1 Model structure 31

3.3.2 Significance testing . . . 35
3.4 Results from application in the Salado *A* sub-basin . . . 36
3.4.1 Covariate selection . . . 36
3.4.2 Validation . . . 38
3.4.3 Seasonal forecasts . . . 40
3.4.4 Multi-decadal projections . . . 45
3.5 Summary and future work . . . 47

4 BayGEN: a Bayesian space-time stochastic weather generator 52
4.1 Introduction . . . 52
4.2 BayGEN . . . 56
4.2.1 Model definition . . . 56
4.2.2 Likelihood computation . . . 60
4.2.3 Implementation . . . 60
4.3 Study region and data . . . 61
4.4 Results . . . 62
4.4.1 Model fit . . . 62
4.4.2 Posterior distribution . . . 63
4.4.3 Multi-site weather simulation . . . 64
4.4.4 Gridded simulation . . . 68
4.4.5 Coupling with DSSAT . . . 70
4.5 Summary and discussion . . . 73

5 A statistical metamodel for monthly groundwater fluctuations 75
5.1 Introduction . . . 75
5.2 Study region and data . . . 77
5.3 MIKE-SHE . . . 78
5.4 The metamodel . . . 79

5.4.1 Hierarchy 80
5.4.2 Covariates 80
5.4.3 Simulation 81
5.5 Results 82
5.5.1 Simulation of historic period 82
5.5.2 Performance during wet and dry months 83
5.5.3 Use of the metamodel in a rudimentary decision mode 84
5.6 Summary and discussion 86

6 Conclusion 88
6.1 Summary 88
6.2 Discussion 92
6.3 Future work 93

Bibliography 95

Tables

Table

3.1 Differences between the AIC for occurrence, amounts, minimum temperature, and maximum temperature models of the original and modified weather generators (positivity implies a decrease). 36

3.2 Kolmogorov-Smirnov test comparing the original and modified weather generator output. P-values lower than 0.05 indicate the output from original and modified generators come from different distributions. 43

Figures

Figure

1.1 Schematic illustrating empirical relationship between water table depth and crop yield in floodplain agriculture. 4

2.1 Study region geography with elevation (in meters), black dots are observation station locations; red dot represents Santiago del Estero. 17

2.2 a-b) Densities for simulated local wet and dry spells at Santiago del Estero, c-d) regional dry and wet spells; observed densities shown in red. 18

2.3 Dry spells validation for Pergamino, a station not included in the model. Boxplots show the number of dry spells equal to (left) or greater than (right) ten days for the 100 realizations. Red dots represent the number of these dry spells for the historical data. 19

2.4 Precipitation occurrence coefficient for a) minimum temperature and b) maximum temperature. 20

2.5 Observed versus simulated pairwise correlations for minimum and maximum temperatures. 21

2.6 a-b) Densities for simulated local cold and hot spells at Santiago del Estero, c-d) regional cold and hot spells. Observed density shown in red. 22

2.7 Variograms for a) observed and b) simulated daily January maximum temperatures. 23

3.1 Study region: weather stations shown as dots, numbers correspond to Table 1. The Salado *A* sub-basin is outlined. Three stations withheld in spatial validation shown as triangles. 30

3.2 First principal components of OND precipitation, minimum temperature, and maximum temperature, scaled and shown as points, and domain-averaged and scaled OND precipitation (top), minimum temperature (middle), and maximum temperature (bottom), shown as lines. 37

3.3 Spatial validation: (a-b) OND 1961-2013 observed versus ensemble mean simulated probability of precipitation occurrence, (c-d) total precipitation, (e-f) mean maximum temperature, and (g-h) mean minimum temperature, for the three withheld stations. Top row (a,c,e,g) corresponds to simulations from the original generator and bottom row (b,d,f,h) is for simulations from the modified generator. 39

3.4 Temporal validation: JFM 2001 – OND 2013 observed minus ensemble mean simulated (a-b) seasonal total precipitation, (c-d) mean maximum temperature, and (e-f) mean minimum temperature. Left panels (a,c,e) are for the original weather generator and right panels (b,d,f) for the modified weather generator. 40

3.5 Top panels: Kernel density estimates of PDF of domain-averaged seasonal precipitation, maximum temperature and minimum temperature from 100 simulated weather scenarios from the modified (blue) and original (red) weather generators (OND 2010 and OND 2012 denoted as solid and dashed lines, respectively), along with the climatological PDF (dotted black line). Observed values are shown as vertical lines. Bottom panels: Sampled seasonal precipitation and temperatures from the categorical probabilistic forecasts with the domain-averaged values generated from the two weather generators – modified (blue) and original (red). 42

3.6 OND 2010 differences in ensemble mean of seasonal (a) total precipitation (mm season-1), (b) mean maximum temperature (deg C), and (c) mean minimum temperature. Differences calculated as original minus modified generators. Salado *A* sub-basin is outlined. 44

3.7 OND 2010 differences in 95% ensemble spread for (a) seasonal total precipitation (mm season-1), (b) seasonal mean maximum temperature (deg C), and (c) seasonal mean minimum temperature. Differences calculated as original minus modified generators. Salado A sub-basin is outlined. . . . 45

3.8 JFM 2015 – OND 2050 projected minus ensemble mean (a-b) simulated seasonal total precipitation, (c-d) mean maximum temperature, and (e-f) mean minimum temperature. 48

3.9 OND 2015-2050 differences in ensemble mean of seasonal (a) total precipitation (mm season-1), (b) mean maximum temperature (deg C), and (c) mean minimum temperature. Differences calculated as original minus modified generators. Salado A sub-basin is outlined. 49

4.1 Salado *A* sub-basin location relative to South America; station locations shown as black dots, grid cell locations shown as grey dots, Salado *A* sub-basin is outlined. Junín shown as large black triangle. 61

4.2 Posterior distributions (density plots) and MLE (vertical line) of the intercept term for the precipitation occurrence latent Gaussian process. Note location 13 is not shown, but is consistent with these findings. 63

4.3 4,000 simulations at Junín: (a-c) Climatological mean of daily precipitation, maximum temperature, and minimum temperature; (d-f) monthly standard deviation of daily precipitation, maximum temperature, and minimum temperature. BayGEN shown as grey boxplots; GLMGEN shown as white boxplots. 65

4.4 Q-Q plots of BayGEN daily domain temperature extrema computed as maximum or minimum of daily values at all locations: (a) domain maximum of maximum temperatures, (b) domain minimum of maximum temperatures, (c) domain minimum of minimum temperatures, and (d) domain maximum of minimum temperatures, units are degrees Celsius. 67
4.5 Same as Figure 4.4, but from GLMGEN simulations. 68
4.6 Top: Average summer growing season extreme daily precipitation (mm/day) at the stations (boxplots), as simulated by BayGEN, ordered by the average observed extreme daily precipitation (solid line). Bottom: Same as top but from GLMGEN simulations. 69
4.7 BayGEN: (a-c) Ensemble mean precipitation, maximum temperature, and minimum temperature, respectively, for the OND season. (d-f) Corresponding 95% ensemble spread. Units for precipitation are millimeters; units for temperature are degrees Celsius. 70
4.8 Same as Figure 4.7 but from GLMGEN simulations. 71
4.9 Cumulative density functions of summer growing season (October – March) (a) total precipitation, and (b) soybean yields at Junín, based on 100 simulations of daily weather sequences for a two-year period. Simulations from BayGEN are shown as a solid line, and GLMGEN as a dashed line. Break-even production risk is shown as dots on the distributions of soybean yield. 73

5.1 Schematic illustrating empirical relationship between WTD and crop yield in floodplain agriculture. 76
5.2 Study region: weather stations shown as black dots; MIKE-SHE grid shown as grey dots; well locations shown as red triangle. Salado *A* sub-basin shown as outline. 78

5.3 Basin average WTD for metamodel ensembles (boxplots, full range of ensemble) and MIKE-SHE (solid line) for the historic period (January 1961–December 2013). 82

5.4 RMSE (meters) and % Bias between MIKE-SHE and the metamodel ensemble mean water table levels for the historic period (January 1961–December 2013). 83

5.5 (a-c) Dec 2000: MIKE-SHE, metamodel ensemble mean, and difference; (d-f) Feb 2012: MIKE-SHE, metamodel ensemble mean, and difference. Difference calculated as MIKE-SHE minus metamodel ensemble mean. Units are meters. 84

5.6 Dec 2013: (a-c) Water table depth as simulated by MIKE-SHE, metamodel ensemble mean, and difference between MIKE-SHE and metamodel ensemble mean; (d-f) 97.5% percentile of metamodel ensemble, 2.5% percentile of metamodel ensemble, and 95% ensemble spread, respectively. Units are meters. 85

Chapter 1

Introduction

1.1 Background

Climate variability is one of the dominant driving forces of uncertainty and risk in climate sensitive sectors around the world. Such risk is particularly prevalent in agricultural systems, due to their artificiality, which "makes them less flexible, and therefore more vulnerable to climatic change than the naturally occurring species of the ecosystem within which they fit..." (Oram, 1985). The overarching message that Oram is portraying in his manuscript – which is equally, if not more, relevant today – is that agricultural systems are a disturbance of the natural system by the human system. It should be no surprise, then, that there exist unrealistic expectations for crops to produce optimal yields in non-native environments, regardless of whether or not the current season's conditions are ideal. Such disturbances to the natural system can compound the negative effects normally imposed by climate variability and contribute a further element of unpredictability to how the natural system will respond to climate variability in the long run. This is especially true in developing nations, where agriculture is predominantly rain-fed – i.e., irrigation is limited. To this end, a better understanding of the interactions between agricultural systems, uncertainty in future climate, and land use changes must be identified and studied on seasonal to multi-decadal scales.

Process based models (e.g., crop simulation models, hydrological models) can be useful tools to assess likely impacts on climate-sensitive sectors of society, and to eval-

uate the outcomes of alternative adaptive actions (Ferreyra et al., 2001b; Berger, 2001; Berger et al., 2006; Happe et al., 2008; Freeman et al., 2009; Schreinemachers and Berger, 2011; Bert et al., 2006, 2007, 2014). These models typically require daily weather data. Although historical daily weather can be used, obtaining long-term daily weather is laborious and costly at best and, in some cases, impossible. Typically, historical observations have missing data that are not accepted by process based models. Similarly, point measurements may not represent the true spatial variability of a nonstationary natural process (e.g., daily precipitation). Most importantly, observed sequences provide a solution based on only one realization (i.e., instance) of the weather process (Richardson, 1981). Stochastic weather generators have long been used for risk assessment and adaptation, as they can provide a rich variety (i.e., ensemble) of plausible climatic scenarios. In this, an ensemble of climate scenarios can be used as input data to a process based model, which will produce a distribution of system variables upon which decisions can be made.

Over the years there have been numerous contributions to the field of weather generation – for an historic overview of daily weather simulation methods, see Wilks and Wilby (1999); chapters 2-4 of this dissertation also provide complete literature reviews of stochastic weather generation. Traditional weather generators, stemming from Richardson (1981), model precipitation occurrence as a chain-dependent process (Katz, 1977). Precipitation intensity and temperature are parameterized using probability distributions and linear time series models, respectively. This approach is effective in capturing climatological variability and linear relationships between variables but fails to capture extreme weather. Nonparametric weather generators have an improved ability to capture nonlinearities between variables and sites. Included in this subclass are the K-nearest neighbor (K-NN) bootstrap resampling method (Brandsma and Buishand, 1998; Rajagopalan and Lall, 1999; Buishand and Brandsma, 2001; Beersma and Buishand, 2003; Yates et al., 2003; Sharif and Burn, 2007) and kernel density based estimators (Rajagopalan et al., 1997b; Harrold et al., 2003; Mehrotra and Sharma, 2007). These methods are simple

and powerful; however, their main drawback is that they cannot generate values outside the range of historical data. More importantly, it is not easy to generate weather sequences at locations other than those with historical observations. Generalized linear models (GLMs) are able to straightforwardly model non-normal data through a suite of link functions, thus can be used to model and simulate daily weather sequences (Furrer and Katz, 2007; Kim et al., 2012; Yan et al., 2002; Yang et al., 2005; Chandler, 2005). The use of spatial process models on GLM regression coefficients enables the simulation of weather trajectories at unobserved locations (Kleiber et al., 2012, 2013; Verdin et al., 2015b, 2016).

Information regarding the uncertainty in future climate may be of interest to a farmer or decision maker. Probabilistic seasonal forecasts may be of use for short-term planning; climate model output for mid-term projections (i.e., 20-40 years from present) may assist in sustainability, viability, or long-term planning. Unfortunately, in developing nations many farmers may not have access to, or may not understand the implications of, the seasonal climate forecast information provided by a number of research teams (Goddard et al., 2003; Saha et al., 2006). The lack of such information, or the inability to interpret the coarse scale of climate forecasts, can lead them to rely on traditional practices. That is, farmers tend to make conservative decisions that generally fail to capitalize on beneficial conditions or buffer against negative effects (Jones et al., 2000; Hansen, 2002; Meinke and Stone, 2005). However, access to seasonal climate forecast information alone will likely fail to persuade adaptation in traditional agricultural practices. To support public and private adaptation and mitigation responses, climate forecast information must be relevant to the needs of the decision makers (Cash et al., 2003). For example, potential outcomes of adaptation actions are more informative for stakeholders and decision makers than are raw climate forecasts. It has been shown that stochastic weather generators can be used as statistical downscaling tools (Apipattanavis et al., 2010; Verdin et al., 2016) and, when used with process based models, effectively translate climate forecast information into

distributions of outcomes for risk assessment and management (Hansen et al., 2006).

1.2 Overview of the Pampas

The Pampas are an agriculturally productive region in southeast South America known for growing soybean, cereal, maize, and wheat. The terrain of the Pampas can best be described as vast and flat, thus there is little subsurface lateral flow. Water inputs to the system (i.e., precipitation) in general have only one method of exiting the system (i.e., evapotranspiration). Due to its flat topography and lack of drainage, there exists a strong coupling between human systems (i.e., agriculture) and natural systems (i.e., groundwater table depth, climate). For instance, an optimal water table depth can augment crop water supply in times of drought, but a shallow water table can drown the roots, and a deep water table can starve the roots, both of which lead to crop failure (see Figure 1.1).

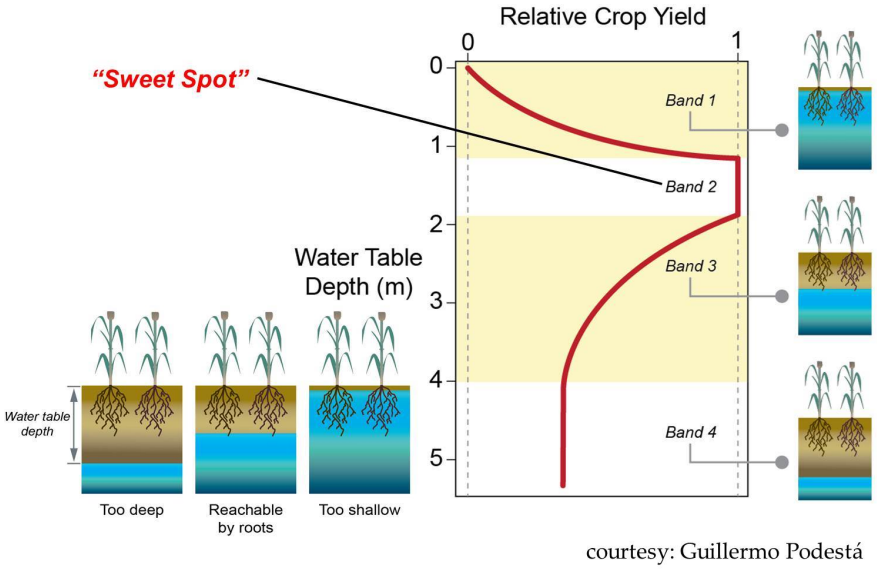

Figure 1.1: Schematic illustrating empirical relationship between water table depth and crop yield in floodplain agriculture.

The Pampas experience climate in pronounced epochal periods: alternating floods

and droughts that displace populations and disrupt productive activities and livelihoods for extended periods. Floods were frequent during the late 19th and early 20th centuries. In contrast, extensive droughts were more frequent during the drier 1930s – 1950s (Herzer, 2003; Seager et al., 2010), which happened to coincide with the "dust bowl" of the northern hemisphere. Climate then reverted to a wet epoch, and, partly in response to significant positive increases in both annual and extreme precipitation that began in the 1960s, severe floods plagued the Pampas in 1980, 1991-93, and 2000-01 (Herzer, 2003). Floods in the western half of the Pampas between 1997 and 2003 left 27% of the landscape under water, halved grain production, damaged infrastructure and soil quality, and transformed the few remaining natural areas (Viglizzo et al., 2009). In contrast, an almost unprecedented drought in 2008 (Skansi et al., 2009) decreased soybean and wheat production in the region by about 30% and 50% respectively.

The positive trend in precipitation of the second half of the 20th century, coupled with favorable economic conditions and technological advances, resulted in large scale land use changes, for instance an expansion of agricultural infrastructure to the semi-arid regions of the Pampas. Where agricultural infrastructure had previously been established, this land use change took the form of continuous cropping practices replacing the grain-pasture rotations. The most notable change in land use came as a result of the global market, local climate, and lower costs to the farmer favoring soybean production, which has since displaced other native crops, pastures for grazing, and forests. As this positive trend in precipitation has abated, public concern has shifted to whether existing agriculture infrastructure in the semi-arid Pampas will remain viable in the future. Such concerns can be addressed using the tools presented in this dissertation – specifically that of chapter 3, which illustrates how to simulate weather ensembles that exhibit specific traits or trends.

1.3 Motivation

Historically, researchers, agronomists, and decision makers in the Pampas have relied mainly on crop simulation models to assess the viability of agricultural infrastructure, and to analyze alternative adaptive actions to climate variability. Stochastic weather generators have long been used to simulate ensembles of future weather trajectories for use in agricultural risk analyses, resulting in ensembles of decision variables. However, climate in the second half of the 20^{th} century was extraordinarily wet, which contributed in part to an increasing water table depth (i.e., approaching the surface). These wetter conditions, along with large scale land use change from perennial pasture to seasonal agriculture, lead to the discovery of a strong coupling between water table depth and relative crop yield. Regions where the water table depth increased drastically can be identified as risk-prone and, therefore, not conducive to agricultural production. Such risk would not be quantified if decisions were made solely on crop simulation models. Thus it is now common for agronomists in the Pampas to work with hydrologists, in order to run distributed hydrologic models for the purpose of identifying risk-prone regions (i.e., shallow or deep water table). Distributed hydrologic models require input sequences of daily weather data on a regular grid, which can be difficult to produce with traditional weather generators. This research is motivated by the need to produce ensembles of daily weather sequences at arbitrary locations, such as on a regular grid, for use in such distributed models.

1.4 Dissertation outline

To assist in agricultural decision making on seasonal to multi-decadal scales in the Argentine Pampas, this dissertation has three broad objectives: (i) to develop a suite of space-time stochastic weather generators to simulate ensembles of daily weather for use with crop simulation and hydrologic models; (ii) to illustrate the efficacy of condition-

ing weather generators on coarse scale climate covariates to assist in seasonal to multi-decadal downscaling; and (iii) to develop a statistical metamodel for estimating monthly evapotranspiration and water table flux, as calculated by the hydrologic model MIKE-SHE.

The following chapters of this dissertation will be presented as standalone, self-contained manuscripts, and are organized as follows: Chapter 2 describes a generalized linear model (GLM) based stochastic weather generator, which is coupled with spatial models to enable simulation at unobserved locations. Chapter 3 demonstrates an application of the GLM weather generator to enable conditional weather generation for seasonal to multi-decadal time scales. In this, probabilistic seasonal forecasts and regional climate model runs are used to condition the output of the GLM weather generator for a network of stations in and around the Salado *A* sub-basin of the Argentine Pampas. However, traditional methods of stochastic weather generation do not effectively quantify nor propagate uncertainty from observed data to parameter space, and subsequently to the simulated weather sequences. To this end, Chapter 4 presents BayGEN, a Bayesian space-time stochastic weather generator, which will provide a means of quantifying the uncertainty associated with weather data and model parameter estimation – values that are commonly underestimated in traditional methods. Finally, distributed hydrologic models, such as MIKE-SHE, can be computationally intensive, especially when considering an ensemble of future weather trajectories. Therefore, Chapter 5 presents a statistical metamodel for estimating monthly actual evapotranspiration and water table depth. Use of this metamodel with simulated weather sequences from weather generators can provide a more thorough risk analysis on seasonal to multi-decadal scales.

Chapter 2

Coupled stochastic weather generation using spatial and generalized linear models

This chapter is published in Stochastic Environmental Research and Risk Assessment with the following citation:
Verdin, A., B. Rajagopalan, W. Kleiber, R. Katz (2015), Coupled stochastic weather generation using spatial and generalized linear models, *Stochastic Environmental Research and Risk Assessment*, 29(2), pp 347–356.
The final publication is available at Springer via http://dx.doi.org/10.1007/s00477-014-0911-6

2.1 Introduction

Risk-based approaches are widely used in natural resources management such as water, land, crop, and ecology. Process-based models of these resources are driven with ensembles of input sequences, which are typically daily weather, resulting in ensembles of system variables and their probability density functions that provide estimates of risk that are useful for decision making. Historic data is often limited in space and time hence the risk estimates based solely on them do not accurately reflect the underlying variability. Therefore, robust generation of weather sequences that capture the underlying variability is essential. Generating random weather sequences that are statistically consistent with historical observations is known as stochastic weather generation.

Crop models for agriculture planning, hydrologic models for generating streamflow

needed for water resources management, and erosion models for land erosion management (Wallis and Griffiths, 1997; Richardson, 1981; Richardson and Wright, 1984; Wilks, 1998; Wilks and Wilby, 1999; Friend et al., 1997) have motivated the development of stochastic weather generators over the years. Traditional weather generators at a single location model the precipitation occurrence as a Markov Chain (Richardson, 1981; Katz, 1977; Stern and Coe, 1984; Woolhiser, 1992) or within a Poisson process framework (Foufoula-Georgiou and Georgakakos, 1991; Furrer and Katz, 2008). The daily rainfall amounts are modeled by fitting a gamma density function (Katz, 1977; Buishand, 1978; Yang et al., 2005; Furrer and Katz, 2007). These models are traditionally estimated for each month or shorter to capture the seasonality. Conditioned on the rainfall state, temperatures are then simulated using autoregressive models (Richardson, 1981).

Multi-site extensions of single site weather generators can be unwieldy with large number of parameters to capture the statistics at each site and their spatial correlation (Mehrotra et al., 2006), more so with a large number of locations. Wilks (1998) proposed a multi-site precipitation model with two state Markov chain and mixed exponential distribution coupled with spatially correlated transformed normal variables to enable capturing the spatial correlation in precipitation. Later, Wilks (1999) extended this to multiple variables (temperature, solar radiation). Variations of this have been used in subsequent multi-site rainfall generators (Srikanthan and Pegram, 2009; Brissette et al., 2007) and weather generators (Qian et al., 2002; Baigorria and Jones, 2010b; Khalili et al., 2009). Markov chains and direct acyclic graphs have been developed and proposed for stochastic multisite rainfall simulation (Kim et al., 2008) as promising and less complex alternatives for space-time simulation. Bayesian hierarchical models for spatial rainfall have been developed in recent years (Lima and Lall, 2009) which have the ability to provide robust estimation of uncertainties and are proving to be an attractive alternative with increase in computational power. Along this same line, Fassò and Finazzi (2011) offer a state of the art approach to space-time modeling using recent maximum likelihood ad-

vances based on the EM algorithm. For modeling sites with heavy tailed precipitation distributions, a generalized Pareto distribution (Lennartsson et al., 2008) or a stretched exponential distribution (Furrer and Katz, 2008) have been shown to be good alternatives to the more traditional methods.

Generalized linear models (GLMs), can greatly reduce the modeling effort of weather generators besides enabling the modeling of non-normal variables and being parsimonious (McCullagh and Nelder, 1989). Herein, a GLM model (probit regression) is adopted for precipitation occurrence with a suite of covariates enabling the spatial modeling of occurrence with a single model – unlike a number of Markov chain models. A separate GLM is fitted to precipitation intensity, often using a Gamma distribution and appropriate link function to capture non-Gaussian features. Early use of GLM for weather generation was by Stern and Coe (1984) with subsequent work by Yang et al. (2005) and Chandler (2005). Furrer and Katz (2007) developed this framework to include climate variables such as El Nino Southern Oscillation index for a location in the Pampas region of Argentina. Other methods to incorporate large scale climate information in weather generators include modeling the underlying climate process using a Hidden Markov Model and then conditionally generating stochastic weather sequences (Hauser and Demirov, 2013). A limited extension of the GLM approach with a Poisson cluster model to multi-site precipitation generation was proposed by Wheater et al. (2005).

Semi-parametric approaches have been developed that resample historical data using an empirical distribution function – with precipitation occurrence modeled as a wet and dry process and seasonality addressed using Fourier components (Racsko et al., 1991; Semenov and Barrow, 1997). These are relatively easy to implement and are widely used in climate change studies, especially in Europe (Calanca and Semenov, 2013; Semenov et al., 2013). Multi-site extensions and enhancements to generate extremes have also been proposed in the above references and in (Semenov, 2008). These weather generators have been shown to capture extreme events well over a region in New Zealand (Hashmi et al.,

2011). In general, weather generators have difficulty capturing the properties of extreme events well – a formal way to enable this using extreme value distributions is proposed in Furrer and Katz (2008).

Nonparametric weather generators which make no assumption of the underlying distribution of the process and are data-driven have gained prominence in recent decades. Kernel density based generators of precipitation (Lall and Sharma, 1996; Harrold et al., 2003; Mehrotra and Sharma, 2007) and other variables (Rajagopalan et al., 1997b,a) have been shown to perform very well at capturing non-normal and nonlinear features. Kernel methods perform poor in high dimensions. To alleviate this, K-nearest neighbor (K-NN) time series bootstrap (Lall and Sharma, 1996) based weather generators were developed (Rajagopalan and Lall, 1999). In this, K-NNs of a weather vector on a current day is obtained from historical days within a window of the current day, and one of the neighbors (i.e., one of the historical days) is resampled with a weighted metric. The historical weather on the following day of the resampled neighbor becomes the simulated weather for the subsequent day. This is akin to resampling from a nonparametric estimation of the local conditional probability density function. For multi-site generation this is done on the site-averaged time series and the weather vector at all the locations of the selected day is taken to obtain multisite simulation. As can be seen this is easy to implement and robust in capturing non-Gaussian features. This has been extended to multisite and also has been conditioned on large scale climate information, climate forecasts, climate change projections, etc. (Yates et al., 2003; Apipattanavis et al., 2007; Buishand and Brandsma, 2001; Beersma and Buishand, 2003; Sharif and Burn, 2007). Recently, Caraway et al. (2014) modified this approach for multi-site weather simulation by incorporating a cluster analysis wherein the sites are clustered and a single site weather generator is applied to each cluster average. This modeling approach shows good performance in mountainous terrain.

One of the major drawbacks with the weather generators described above is their

relative inability to generate weather sequences at any arbitrary locations, other than the locations with data. This is quite important for running hydrology, crop, and ecology models which require weather sequences on a grid. It is in this context that the GLM-based methods offer a parsimonious and robust approach. Kleiber et al. (2012) extended this with latent Gaussian processes to model spatial occurrence and amounts of rainfall over the state of Iowa, US and the Pampas region of Argentina, and to temperature in complex terrain (Kleiber et al., 2013). Motivated by this drawback, this chapter presents the development of a GLM-based spatial weather generator, which combines the precipitation and temperature generator of Kleiber et al. (2012) and Kleiber et al. (2013) and demonstrates it for application to the Pampas region.

2.2 Stochastic model

A basic full stochastic weather generator requires simultaneous simulation of minimum and maximum temperature, as well as precipitation, including both occurrence and intensity. The idea behind our approach is to condition the bivariate temperature process on precipitation occurrence. Although there is clearly a physical relationship between temperature and precipitation, precipitation largely occurs due to large scale atmospheric movement, while surface temperatures are highly controlled by local climate factors and by whether or not precipitation occurs. By maintaining a generalized linear modeling framework, it is straightforward to condition temperature simulations on precipitation occurrence, thus allowing for distinct precipitation stochastic models to be used.

We follow the framework proposed by Kleiber et al. (2013) in focusing on the two components of local climate and weather. Local climate refers to the average behavior of a weather variable across time and space, while the weather component yields variability and individual realizations that deviate from climatology. For minimum and maximum temperatures at location $\boldsymbol{s} \in \mathbb{R}^2$ and day t, $Z_N(\boldsymbol{s}, t)$ and $Z_X(\boldsymbol{s}, t)$, respectively, the follow-

ing decomposition may be used,

$$Z_N(\boldsymbol{s},t) = \boldsymbol{\beta}_N(\boldsymbol{s})'\mathbf{X}_N(\boldsymbol{s},t) + W_N(\boldsymbol{s},t) \tag{2.1}$$

$$Z_X(\boldsymbol{s},t) = \boldsymbol{\beta}_X(\boldsymbol{s})'\mathbf{X}_X(\boldsymbol{s},t) + W_X(\boldsymbol{s},t). \tag{2.2}$$

The first component is a local regression on some covariate vector $\mathbf{X}_i(\boldsymbol{s},t)$, while the weather component (denoted by W for weather) generates variability and spatial correlation via a multivariate normal Gaussian process. In our experience, temperature persistence is most straightforwardly accounted for by autoregressive terms in the mean function, and the weather component can then be viewed as temporally independent. It is worthwhile to note that the use of Gaussian models for temperature is justified (see Kleiber et al. (2013)). Transformed variables can be used to produce stochastic realizations without assuming a Gaussian distribution, but these realizations have shown results that are consistent with this study.

The local climate component is a spatially varying coefficient model, where $\boldsymbol{\beta}_i(\boldsymbol{s}) = (\beta_{0i}(\boldsymbol{s}), \beta_{1i}(\boldsymbol{s}), \ldots, \beta_{pi}(\boldsymbol{s}))'$, for $i = N, X$ determines the influence of each covariate on temperature at a given location. For example, the intercept term $\beta_{0i}(\boldsymbol{s})$ accounts for the fact that, typically, temperatures at higher elevations tend to be lower than those at lower elevations or near oceans or seas. In our experience, it is useful and appropriate to include autoregressive terms in the covariate vectors $\mathbf{X}_i(\boldsymbol{s},t) = (X_{0i}(\boldsymbol{s},t), \ldots, X_{pi}(\boldsymbol{s},t))'$.

Estimation of the coefficients $\boldsymbol{\beta}_i(\boldsymbol{s})$ rely on observations at a network of locations $\boldsymbol{s} = \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ over a time period $t = 1, \ldots, T$ (note that an incomplete historical record does not affect estimation). A Bayesian approach would be to impose a prior distribution on the coefficients, viewing them as spatial processes, which will be visited later in this dissertation. Below, the coefficients are allowed to vary with location, but a stochastic representation is suppressed.

The precipitation process is broken into two components, the occurrence at location $\boldsymbol{s}$ on day t, $O(\boldsymbol{s},t)$, and the intensity or amount, $A(\boldsymbol{s},t)$, given that there is some precipi-

tation. In particular, this follows Kleiber et al. (2012) in modeling the occurrence process as a probit process,

$$O(\boldsymbol{s}, t) = \mathbb{1}_{[W_O(\boldsymbol{s},t)\geq 0]} \tag{2.3}$$

where the latent process $W_O(\boldsymbol{s}, t)$ is Gaussian. If the latent process is positive, it rains at location $\boldsymbol{s}$, whereas if the process is negative, it does not rain. The latent process is given a mean function that is a regression on some covariates, $\boldsymbol{\beta}_O(\boldsymbol{s})'\mathbf{X}_O(\boldsymbol{s}, t)$, with spatially varying coefficients as in the temperature model. Realizations are spatially correlated by imposing a nontrivial covariance structure for $W_O(\boldsymbol{s}, t)$. Kleiber et al. (2012) used an exponential covariance function to model spatial correlation, for example. Briefly, the precipitation intensity process follows the same approach as Kleiber et al. (2012). In particular, the intensity at a particular location and time is modeled as a gamma random variable, whose scale and shape parameters vary with location and time. Simulations are spatially correlated by imposing a zero-mean Gaussian process $W_A(\boldsymbol{s}, t)$ with covariance function $C_A(\boldsymbol{h}, t)$, such that

$$A(\boldsymbol{s}, t) = G^{-1}_{\boldsymbol{s},t}(\Phi(W_A(\boldsymbol{s}, t))) \tag{2.4}$$

where $G_{\boldsymbol{s},t}$ is the cumulative distribution function (CDF) of the gamma distribution at site $\boldsymbol{s}$ and time t, and Φ is the CDF of a standard normal. This transformation approach is called a spatially varying anamorphosis function (Chilès and Delfiner, 1999), which retains the gamma distribution at individual locations but allows for spatial correlation between locations. This model is not explored in detail here, acknowledging that other precipitation models can be swapped in easily.

2.2.1 Estimation and modeling choices

A conditional approach to estimation is taken by first gathering local estimates $\hat{\beta}_{ji}(\boldsymbol{s})$ by ordinary least squares at each observation location for both minimum and maximum

temperature. Spatial covariance often exhibits seasonal patterns, where, for example, temperatures tend to exhibit greater variability in summer than in winter. Additionally, length scale of spatial correlation can also vary across time. To account for these nonstationarities, the spatial covariance structures for $W_i(\boldsymbol{s}, t)$ are estimated on a monthly basis, using the empirical covariance matrix of the residuals of the observation network. For each day t and spatial location s, we define the residuals as $W_i(\boldsymbol{s}, t) = Z_i(\boldsymbol{s}, t)\beta_i(\boldsymbol{s})^T X(\boldsymbol{s}, t)$, where $\beta_i(\boldsymbol{s})$ is the least squares estimate. These residuals are then assumed to be realizations from the W_i process, and from these values the empirical covariance matrices are formed. Estimation for precipitation occurrence follows a similar strategy; first we estimate local mean coefficients $\hat{\beta}_{jO}(\boldsymbol{s})$ by probit regression, using all historical occurrence observations. The spatial structure for the latent process is estimated as the empirical correlation based on the probit model errors at all network stations, using occurrences as observations separately for each month.

For this coupled weather generator, a multivariate autoregressive structure is chosen for the temperature process, conditional on precipitation occurrence. In particular, the covariates for the temperature process are

$$\mathbf{X}_N(\boldsymbol{s}, t) = (1, \cos(2\pi t/365), \sin(2\pi t/365), r(t), Z_N(\boldsymbol{s}, t-1), Z_X(\boldsymbol{s}, t-1), O(\boldsymbol{s}, t))'. \quad (2.5)$$

The first three entries are an intercept and two harmonics to account for seasonal trends; $r(t)$ is a linear drift between -1 and 1 (for numerical stability), which is included to control for temperature trends over the period of our data set; the latter three entries imply a trivariate autoregressive structure. Note that temperature is conditioned on the coincidental occurrence; in practice this usually implies cooler temperatures on rainy days and warmer temperature on dry days. The same covariates are used for the maximum temperature process. The precipitation occurrence process is given the following covariates,

$$\mathbf{X}_O(\boldsymbol{s}, t) = (1, \cos(2\pi t/365), \sin(2\pi t/365), O(\boldsymbol{s}, t-1))', \quad (2.6)$$

where now precipitation uses a single autoregressive model. Note the linear drift is not included in the precipitation model because that process tends to exhibit epoch-like traits rather than linear trends.

Simulation can proceed by simulating an entire trajectory of precipitation occurrence, with amounts if required for scientific purposes. Conditional on this realization, an initial temperature is chosen (e.g., the average of that calendar day's observations), and daily realizations are then available by simulating the weather component as samples from multivariate normal distributions, and adding the weather to the local climate.

2.3 Stochastic weather simulation in the Pampas

To illustrate the performance and capability of the proposed coupled model, a dataset of minimum temperature, maximum temperature and precipitation for a network of 19 locations in the Pampas region of Argentina is considered (see Figure 2.1). The Pampas region covers much of northeastern Argentina, all of Uruguay, and very little of southeastern Brazil – covering more than 750,000 km^2 – and is of utmost agricultural importance for much of the South American continent. With global food prices on the rise, the ability to quantify and forecast climate variability for the purposes of climate change impact assessment in the region is absolutely necessary. Observations are available over approximately an 80-year period, although the longest station record is from 1908 to 2010. Spatial precipitation simulation was previously explored by Kleiber et al. (2012) on this dataset, but these temperature observations were not considered.

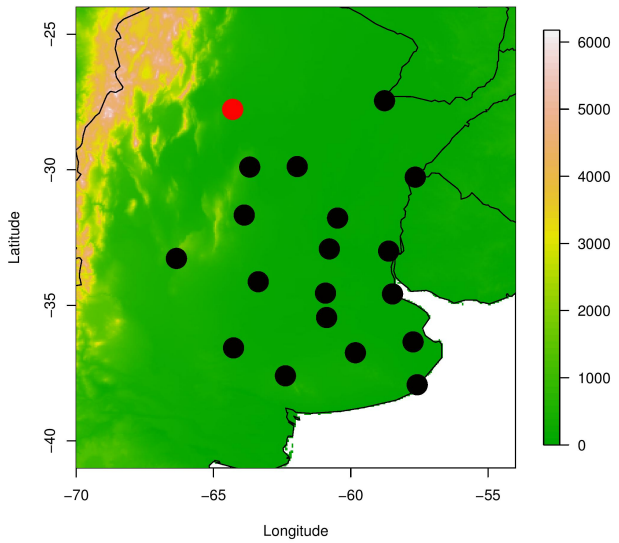

Figure 2.1: Study region geography with elevation (in meters), black dots are observation station locations; red dot represents Santiago del Estero.

The model setup as outlined in Section 2.2 is adopted, and local regression coefficients are estimated (linear regression for temperature, probit regression for precipitation occurrence). Conditional on these estimates, the spatial covariance structure is estimated empirically. To begin, precipitation occurrence is simulated for each day over the 19 locations throughout the Pampas. 100 trajectories of occurrence are simulated independently, thus producing an ensemble of daily precipitation patterns. To consistently compare simulations to observations, it is necessitated that output from the coupled weather generator be masked to match the pattern of missing values from the observed precipitation time series.

Validation of the precipitation occurrence model is carried out through spells analysis – local and regional wet and dry spells. A regional dry spell occurs when all 19 locations report no rain – occurrence at any location breaks the regional dry spell. Figure 2.2 shows the density of spells from the 100 trajectories as well as that from the observed

time series.

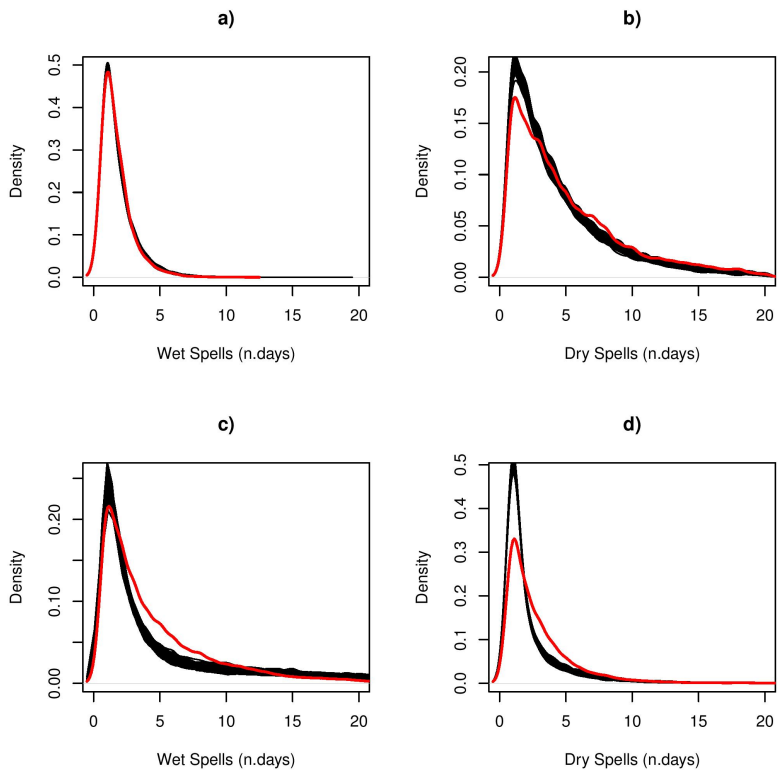

Figure 2.2: a-b) Densities for simulated local wet and dry spells at Santiago del Estero, c-d) regional dry and wet spells; observed densities shown in red.

As can be seen in Figure 2.2, both wet and dry spells are reproduced with good skill by the trajectories at Santiago del Estero. Simulated wet spells are very nearly perfect, while there is slight discrepancy in the density occurrence of dry spells – simulations are producing systematically shorter dry spells than the observations. Due to this underrepresentation of local dry spells, there is even greater discrepancy between observed and simulated regional dry spells. The trajectories imply that regional dry spells are very rarely longer than one day, while observations show they will likely last at least three days. However, the frequency of longer domain dry spells is adequately reproduced, particularly above spells of eight or more days. To illustrate the validity of this model in reproducing the frequency of long dry spells, two dry spells analysis were carried out for

an independent station. Weather for a station not included in the model was simulated – thus validating the ability of this model to simulate weather at any arbitrary location – and the ability to reproduce long dry spells is analyzed. These long dry spells are crucial to capture for impact assessment planning. As can be seen in Figure 2.3, the model is quite impressive in its ability to reproduce the frequency and longevity of dry spells, especially considering this station was not included in the model fit.

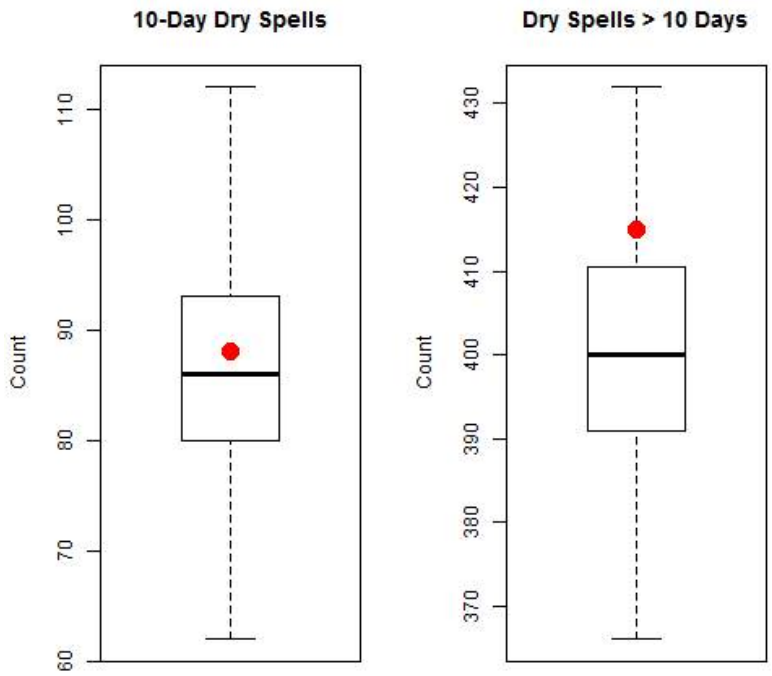

Figure 2.3: Dry spells validation for Pergamino, a station not included in the model. Boxplots show the number of dry spells equal to (left) or greater than (right) ten days for the 100 realizations. Red dots represent the number of these dry spells for the historical data.

Allowing the coupled relationship between temperature and precipitation to vary with location is important over large domains, such as in the Pampas. Figure 2.4a illustrates the relationship that precipitation occurrence has with minimum temperature. It can be seen there is little spatial structure relating these processes, implying that precipi-

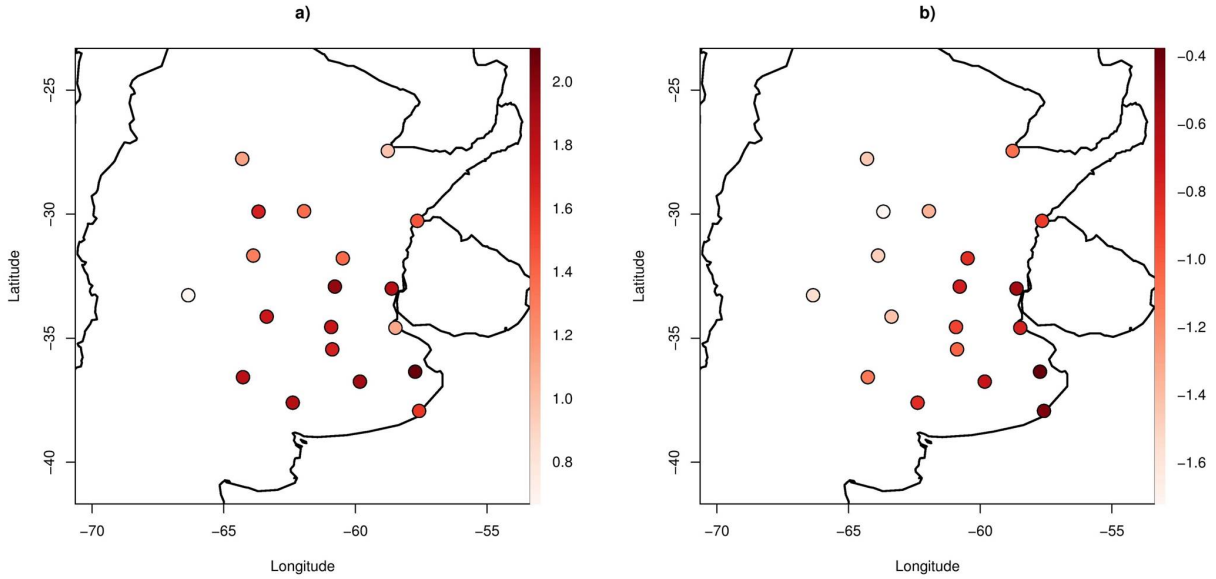

Figure 2.4: Precipitation occurrence coefficient for a) minimum temperature and b) maximum temperature.

tation occurrence is not the only contributing factor in simulating minimum temperatures in the region. Note that the coefficient on occurrence is generally positive, indicating that the presence of precipitation implies that minimum temperatures tend to be between $1^\circ - 2^\circ$ C warmer. Conversely, Figure 2.4b shows that there is a much stronger spatial structure relating precipitation occurrence and maximum temperature. Indeed, inland maximum temperatures tend to be reduced with the presence of precipitation, while the maximum temperature at locations near the ocean are less affected by precipitation.

Capturing the spatial coherence of daily weather patterns is of utmost importance in producing realistic weather generator output. To this end, pairwise correlations are considered for minimum and maximum temperatures at all 19 locations, producing $\frac{19x18}{2} = 171$ pairs, as can be seen in Figure 2.5. These pairwise correlations are somewhat consistent between the simulations and historical observations, although there is evidence of slightly reduced model correlation, on the order of 5%.

Figure 2.5 illustrates that minimum and maximum temperatures are positively spa-

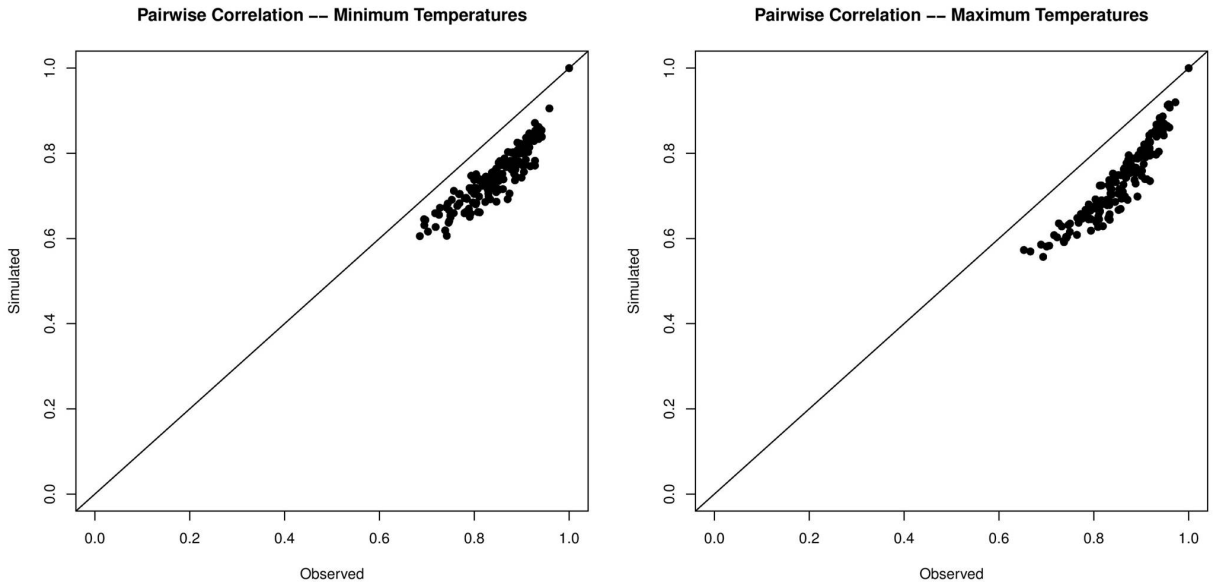

Figure 2.5: Observed versus simulated pairwise correlations for minimum and maximum temperatures.

tially correlated, in that neighboring locations have similar daily temperature patterns. It follows to analyze the output from the coupled temperature models further, thus assessing its ability to reproduce cold and hot spells. A cold spell is defined as the number of consecutive days that the minimum temperature at a location is less than 5° C. Similarly, a hot spell is defined as the number of consecutive days that the maximum temperature at a location exceeds 30° C. A regional cold spell occurs when the minimum temperatures at all 19 locations are less than 5° C. For a regional hot spell, the maximum temperatures at all 19 locations must exceed 30° C. For consistency, and because simulated precipitation occurrence is used as input in the minimum and maximum temperature models, local cold and hot spells are analyzed for Santiago del Estero, Argentina. Figure 2.6a-b shows that the 100 trajectories reproduce the density of cold and hot spells with good skill. The timing and peak of simulated cold spells are nearly perfect, while those of hot spells show slight discrepancy with respect to observations. In Figure 2.6c-d, it can be seen there is

not systematic underrepresentation of local hot spells – as was the case for local dry spells at this location – as the simulated and observed densities of regional cold and hot spells are reproduced with very good skill.

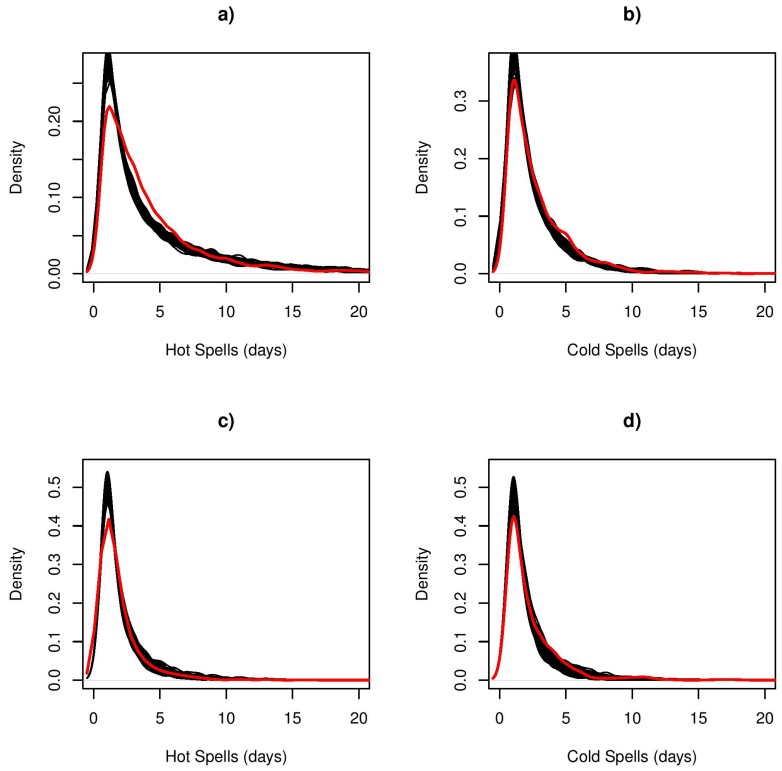

Figure 2.6: a-b) Densities for simulated local cold and hot spells at Santiago del Estero, c-d) regional cold and hot spells. Observed density shown in red.

The space-time aspect of this stochastic weather generator may be examined by obtaining the covariance of daily weather on a monthly scale, resulting in an ensemble of empirical variograms per month. It is assumed the monthly covariance is isotropic, such that weather patterns vary on the same scale in all directions to a given lag. For a given month, the daily weather is used to obtain an empirical variogram; the process is repeated for all days within the given month with non-missing data, and the ensemble of empirical variograms may be visualized as boxplots. Ensemble variograms of daily maximum temperature for January are shown in Figure 2.7. Note that Figure 2.7a shows the ensemble

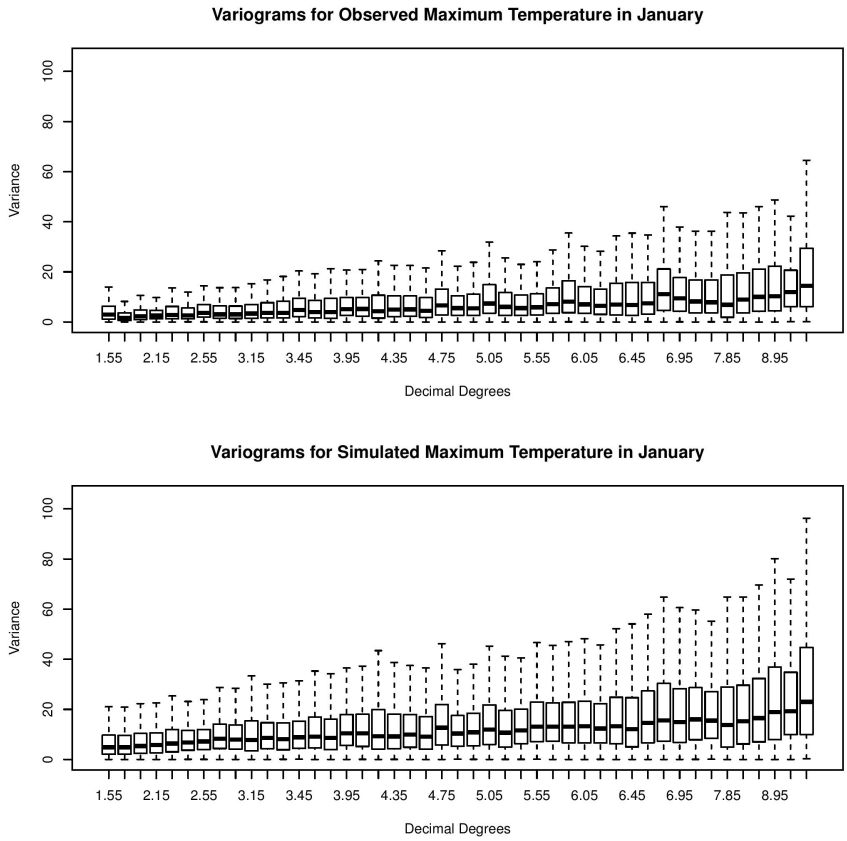

Figure 2.7: Variograms for a) observed and b) simulated daily January maximum temperatures.

of empirical variograms for observed maximum temperature, while Figure 2.7b shows that for a randomly-selected trajectory. It can be seen that the spatial correlation structure based on model realizations are similar to those observed in the historical data, indicating that the model adequately captures the spatial behavior as well as the local behavior of temperature.

2.4 Discussion

A conditional approach to daily space-time stochastic weather simulation has been introduced, wherein temperature is conditioned on precipitation occurrence. Although

this idea has previously been explored, the model is endowed with additional flexibility by allowing model coefficients to vary by location within a local climate framework. Simulations are correlated via Gaussian process residual terms, yielding spatially and temporally consistent realizations.

The focus of validation has been a spells analysis for model output assessment, as the purpose of this research is provide a stochastic weather generator that generates daily weather ensembles where the minimum and maximum temperatures are conditioned on precipitation occurrence. The precipitation intensities have not been validated nor reported on because they are produced from the same technique as in Kleiber et al. (2012). Validation of the model's ability to simulate weather at any arbitrary location was validated by producing statistically consistent simulations at an independent station. Not only were short spells reproduced with near perfection, the longevity and frequency of dry spells were maintained throughout all 100 realizations. The ability to reproduce long dry spells is crucial for impact assessment planning.

Chapter 3

A conditional stochastic weather generator for seasonal to multi-decadal simulations

This chapter is published in the Journal of Hydrology with the following citation: Verdin, A., B. Rajagopalan, W. Kleiber, G. Podestá, and F. Bert (2016), A conditional stochastic weather generator for seasonal to multi-decadal simulations, *Journal of Hydrology*. doi:10.1016/j.jhydrol.2015.12.036

3.1 Introduction

Scientific and technological advances, together with awareness of the importance of climate on human endeavors, are creating increased worldwide demand for climate information. Fortunately, our ability to monitor and predict variations in climate has increased substantially (Barnston et al., 2010; Stockdale et al., 2010). A number of groups now forecast climate conditions a few seasons ahead (Goddard et al., 2003; Saha et al., 2006). Emerging developments may enable climate projections 10 to 20 years into the future, a scale intermediate between seasonal forecasts and manmade climate change projections (Haines et al., 2009; Hurrel et al., 2009; Meehl et al., 2009). These advances, however, must be matched by a better understanding of how science can inform climate-resilient planning and development (Stainforth et al., 2007).

To support public and private adaptation and mitigation responses, climate information must be credible, legitimate and, especially, salient – e.g., relevant to the needs of decision makers (Cash et al., 2003). Needs include not only predictions or projections

(Bray and von Storch, 2009) of regional climate: *potential outcomes of adaptation actions are probably more relevant to stakeholders than raw climate information.* Thus, an enhanced capacity is needed to "translate" climate information into distributions of outcomes for risk assessment and management (Hansen et al., 2006).

Process models (e.g., crop biophysical models, hydrological models) can be useful tools to assess likely impacts on climate-sensitive sectors of society, and to evaluate the outcomes of alternative adaptive actions (Ferreyra et al., 2001b; Berger, 2001; Berger et al., 2006; Happe et al., 2008; Freeman et al., 2009; Schreinemachers and Berger, 2011; Bert et al., 2006, 2007, 2014). These models, however, typically require daily weather data. Although historical daily weather can be used, getting long-term daily weather is laborious and costly at best and, in some cases, impossible. Typically, historical observations have missing data that are not accepted by impact models. Similarly, point measurements may not represent the true spatial variability of a nonstationary natural process (e.g., daily precipitation). Most importantly, observed sequences provide a solution based on only one realization of the weather process (Richardson, 1981).

The use of seasonal forecasts of regional climate and its impacts can help decision-makers to lessen the adverse effects of unfavorable conditions or, alternatively, to capitalize on favorable conditions. Nevertheless, a major obstacle to broader use of seasonal climate forecasts is their coarse spatial and temporal resolution. Similarly, 10 to 20 year projections of regional climate conditions have been identified as important to infrastructure planners, water resource managers, and many others (Hurrel et al., 2009). Unfortunately, projections of regional monthly precipitation and temperature from climate models not only are coarse in space and time – as seasonal forecasts – but also involve considerable uncertainty, which requires exploration of the impacts of alternative, plausible trajectories. Stochastic weather generators have long been used for risk assessment and adaptation, as they can provide a rich variety of plausible climatic scenarios. Moreover, weather generators can produce spatially consistent sequences that can be used to

downscale larger-scale scenarios.

Traditional weather generators (stemming from Richardson (1981)) model precipitation occurrence as a chain-dependent process (Katz, 1977) and thus are capable of generating physically realistic prolonged wet and dry spells. The remaining weather variables (e.g., precipitation intensity and temperature) are parameterized using probability distributions (for precipitation intensity) and linear time series models (for temperature), which capture historical climatological variability and linear relationships between variables but fail to capture extremes (e.g., extreme drought or flooding). In order to capture the variability of weather attributes in any specific season, the simulations need to be conditioned on appropriate covariates. One approach is to estimate the parameters of the generator conditionally by considering ENSO (El Niño Southern Oscillation; Trenberth and Stepaniak (2001)) phase, or any other teleconnection to a region's climate, which enables simulation of skillful sequences (Grondona et al., 2000; Ferreyra et al., 2001b; Wilby et al., 2002; Meza, 2005; Katz et al., 2002). Wilks (2008) illustrated the capability of interpolating weather generator parameters to arbitrary locations (e.g., on a grid) using local weighted regressions; Wilks (2009b) subsequently offered a method to synchronize gridded synthetic weather sequences on observed weather data. Approaches to producing weather sequences that deviate from climatology have included the implementation of seasonal correction factors, perturbation of parameters or input data, and spectral approaches (Caron et al., 2008; Kilsby et al., 2007; Hansen and Mavromatis, 2001; Schoof et al., 2005; Qian et al., 2010).

Nonparametric weather generators have an improved ability to capture nonlinearities between variables and sites. Included in this subclass are the K-nearest neighbor (K-NN) bootstrap resampling method (Brandsma and Buishand, 1998; Rajagopalan and Lall, 1999; Buishand and Brandsma, 2001; Beersma and Buishand, 2003; Yates et al., 2003; Sharif and Burn, 2007) and kernel density based estimators (Rajagopalan et al., 1997b; Harrold et al., 2003; Mehrotra and Sharma, 2007). Caraway et al. (2014) first applied a

clustering algorithm to identify regions of similar climatology before applying the K-NN approach, which has shown good performance in regions of complex terrain. Apipattanavis et al. (2010) modified the K-NN approach to create a semi-parametric weather generator that better captures the duration of wet and dry spells via Markov chain modeling. Modifications of the K-NN based weather generator to incorporate seasonal precipitation forecasts (Apipattanavis et al., 2010) and multi-decadal projections (Podestá et al., 2009) have also been proposed. In these situations, the resampling is weighted to reflect the projected distribution of regional climate conditions. These methods are simple and powerful; however, their main drawback is that they cannot generate values outside the range of historical data. More importantly, it is not easy to generate weather sequences at locations other than those with historical observations.

Pioneered by Stern and Coe (1984), generalized linear models (GLMs) are able to straightforwardly model non-normal data through a suite of link functions. Relevant to this research, GLMs can be used to model and simulate daily weather sequences, and have paved the way for generating space-time weather sequences at any desired location (Kleiber et al., 2012, 2013; Furrer and Katz, 2007; Kim et al., 2012; Yan et al., 2002; Yang et al., 2005; Chandler, 2005; Verdin et al., 2015b). Recently Verdin et al. (2015b) incorporated these developments into a robust space-time weather generator and demonstrated its capability to generate realistic weather sequences at arbitrary locations in the Pampas of Argentina – also the region targeted by this paper. The GLM framework offers several advantages – mainly they reduce the effort in modeling non-normal variables and are parsimonious (McCullagh and Nelder, 1989), especially for discrete and skewed variables (e.g., precipitation occurrence and intensity, respectively). Coupled with spatial processes, GLMs can generate sequences at any spatial resolution – which is important for resource management. Furthermore, covariates such as ENSO information, seasonal climate forecasts, and annual cycles can easily be incorporated in the GLMs to refine or narrow the distribution of expected values (Chandler and Wheater, 2002; Wheater et al.,

2005; Furrer and Katz, 2007; Kim et al., 2012).

As motivated earlier in this section, skillful and realistic sequences of daily weather in any given season are essential for efficient planning and management of agricultural resources. One method of obtaining such sequences requires generating space-time weather sequences that are consistent with, and conditioned on, coarse climate information from seasonal to decadal time scales. To this end, here we propose a modification to the stochastic weather generator presented in Verdin et al. (2015b) to include the coarse scale information as covariates. We refer to the weather generator of Verdin et al. (2015b) as "original"; that of this research will be called the "modified" weather generator. The paper is organized as follows: the study region and data are described in section 3.2; section 3.3 contains a brief summary of the modified methodology. In section 3.4 we discuss the results, and in section 3.5 we conclude with a summary of the research and future work.

3.2 Study region and data

Application of this methodology is focused on a network of seventeen weather stations located in and around the Salado *A* sub-basin of the Pampas of Argentina (see Figure 3.1). The Salado is part of the large Río de la Plata basin (Herzer, 2003). Note the study region differs from that of Verdin et al. (2015b).

The *A* sub-basin is an agriculturally productive sub-basin within the Salado River basin where maize, soybean, and wheat are grown. The Salado basin has very flat topography and a poorly developed and disintegrated drainage system. The western sub-basin (Salado *A*) includes mega-parabolic dunes separated by depressions that constrain evacuation of surface water (Aragón et al., 2011; Viglizzo et al., 2009, 1997). Since colonial times, the Salado has shown alternating floods and droughts that displace populations and disrupt productive activities and livelihoods for extended periods. Floods were frequent during the late 19^{th} and early 20^{th} centuries, a relatively wet epoch. In contrast, extensive

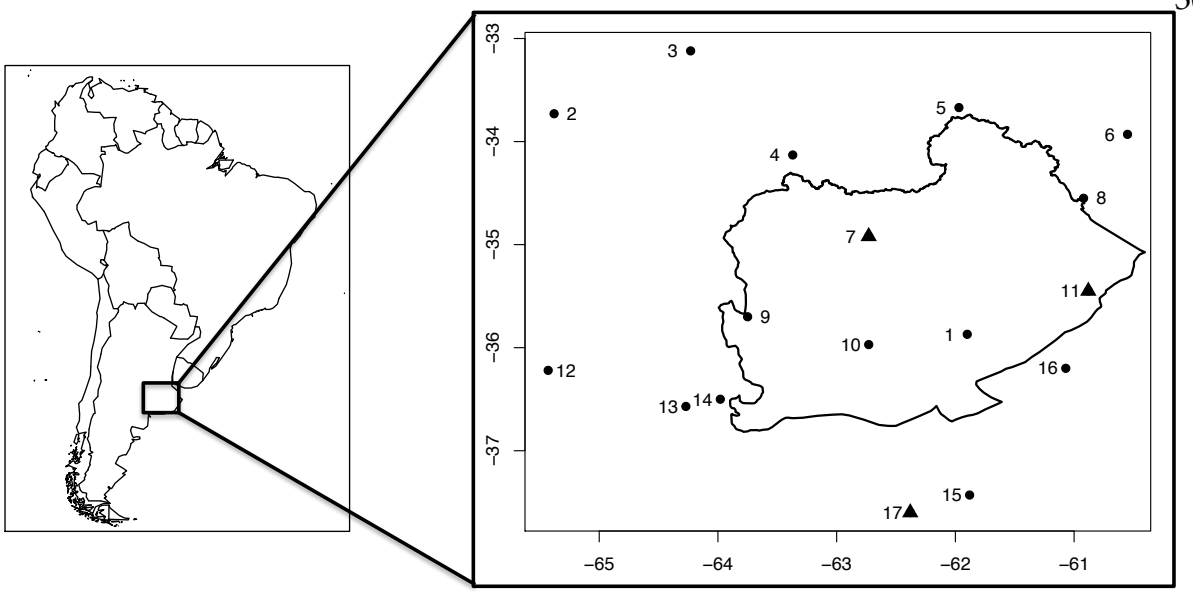

Figure 3.1: Study region: weather stations shown as dots, numbers correspond to Table 1. The Salado *A* sub-basin is outlined. Three stations withheld in spatial validation shown as triangles.

droughts were more frequent during the drier 1930s–1950s (Herzer, 2003; Seager et al., 2010). Partly in response to rain increases since the 1970s, severe floods have occurred in the Salado Basin in 1980, 1991–93, and 2000–01 (Herzer, 2003). Floods in the western half of the Pampas between 1997 and 2003 left 27% of the landscape under water, halved grain production, damaged infrastructure and soil quality, and transformed the few remaining natural areas (Viglizzo et al., 2009). In contrast, an almost unprecedented drought in 2008 (Skansi et al., 2009) decreased soybean and wheat production in the region by about 30% and 50% respectively.

We apply the proposed methodology on a sub-basin scale to illustrate its ability in downscaling coarse seasonal (multi-decadal) forecasts (projections) to local daily weather patterns while maintaining physically realistic climatic characteristics. As agriculture in the Pampas is entirely rainfed, it is of interest to provide a robust risk assessment for crop yields in this region.

During the last half of the 20th century, the study region experienced one of the most significant positive trends in annual precipitation amounts in the world (Giorgi, 2002). This overall increase in precipitation partly contributed to immense agricultural expansion to the semi-arid regions of the western Pampas (Bert et al., 2014). Since the turn of the 21st century, however, observed conditions suggest a significant decrease in regional annual precipitation, which begs the question: "Are the existing agricultural production systems viable in a drier future?" Analysis of a system's response to an ensemble of possible futures that exhibit significant fluctuations in annual precipitation on the multi-decadal scale is of utmost importance for production risk analysis in climatically marginal regions such as the western Pampas.

Daily time series of precipitation, minimum temperature, and maximum temperature are available for a network of 17 weather stations from 1 January 1961 to near present (in this research we use data up to 31 December 2013). This data was collected and organized by associates at the Servicio Meteorológico Nacional (National Meteorological Service) of Buenos Aires, Argentina, and extensive quality control was carried out to ensure its validity. While there is a significant longitudinal gradient in precipitation and temperature (800 mm/year precipitation and 24° C maximum temperature in the west, 1000 mm/year precipitation and 20° C maximum temperature in east), the climatic tendencies (e.g., trends) are similar between all weather stations, thus the *A* sub-basin serves as an optimal test bed for this methodology.

3.3 Methodology

3.3.1 Model structure

We follow the model structure defined in Verdin et al. (2015b), a summary of which is provided below. In describing this methodology, we also develop modifications to improve flexibility by producing conditional weather sequences driven by seasonal fore-

casts, multi-decadal projections, climate drivers or variables, or any other relevant information introduced as time series of covariates. It should be noted that in equations 3.2, 3.4, 3.7, and 3.8, the ellipses denote any number of relevant covariates the user wishes to include, such as seasonal characteristics (e.g., mean temperature or total precipitation), large-scale climate modes (e.g., El Niño-Southern Oscillation, Pacific Decadal Oscillation, Atlantic Multi-decadal Oscillation), or any other climatic variables. Here we propose to use seasonal spatial average precipitation and temperatures as covariates. These additional covariates are calculated from the gauge data. It is acknowledged that a possible scale mismatch exists between the domain average calculated from 17 stations and the true domain average. However, the network of stations is evenly spaced throughout the domain, thus it is fair to assume the stations adequately represent the true domain average.

In the weather generator described here we define two explicit components of daily weather patterns: local climate and daily variability (as suggested by Kleiber et al. (2013)). Local climate represents the expected value of a given meteorological process largely due to seasonal cycle; daily variability provides perturbations to local climate due to weather. Precipitation is considered the primary variable in that occurrence of precipitation tends to modify maximum and minimum temperatures on that day (e.g., due to cloud cover and latent heat transfer). Minimum and maximum temperatures are therefore conditional on precipitation occurrence; precipitation intensities are modeled and simulated independently from occurrence.

In this research, precipitation occurrence and intensity (e.g., amounts), and minimum and maximum temperatures at location $s \in \mathbb{R}^2$ for day $t = 1, \ldots, T$, where T is the number of days in the observational record, are denoted as $O(s,t)$, $A(s,t)$, $Z_N(s,t)$, and $Z_X(s,t)$, respectively. As in Verdin et al. (2015b), occurrence is modeled as a probit process driven by a latent Gaussian process $W_O(s,t)$ via:

$$O(s,t) = \mathbb{1}_{W_O(s,t)\geq 0} \tag{3.1}$$

If $W_O(s,t)$ is positive, this is indicative of rain on day t at location s and is assigned the value 1; if the latent Gaussian process is negative or equal to zero, day t at location s is dry and is assigned the value 0. The mean function of the latent Gaussian process is simply a regression on covariates that are appropriate for the domain of interest. Similar to Verdin et al. (2015b), this regression has covariates

$$X_O(s,t) = (1, O(s,t-1), \cos(2\pi t/365), \sin(2\pi t/365), \mathrm{ST}(t), \ldots), \tag{3.2}$$

which are the intercept term, the previous day's occurrence, two harmonic terms to account for seasonality, and the domain-averaged seasonal total precipitation. The key modification to this regression is the seasonal total precipitation covariate, denoted by $\mathrm{ST}(t)$. In practice this covariate is divided into four distinct covariates relating to each season; covariates are set to zero for times not included in their respective season. To maintain spatial correlations of precipitation occurrence in the domain, an explicit correlation function is defined for $W_O(s,t)$. A correlation function is used instead of a covariance function because probit regression has variance unity by definition. Precipitation amounts at any individual location are modeled as a Gamma random variable as in Kleiber et al. (2012) as follows:

$$A(s,t) = G_{s,t}^{-1}\big(\Phi(W_A(s,t))\big), \tag{3.3}$$

where $G_{s,t}^{-1}$ is the quantile function (e.g., inverse cumulative distribution function) of the Gamma distribution at location s and day t, and Φ is the cumulative distribution function of a standard normal. The simulated rainfall values maintain spatial correlation by applying a spatially varying copula function to the zero-mean Gaussian process $W_A(s,t)$ with correlation function $C_A(h,t)$ (Chilès and Delfiner, 1999). The shape parameter varies

with space, such that each location has its own distinct value; the scale parameter varies with both space and time – its time dependence is based on the seasonal characteristics of precipitation, which are generally captured by annual harmonics. Similar to the occurrence process, the Gamma model parameters are informed by a set of covariates, including the areal seasonal total precipitation covariate as in the occurrence model, as follows:

$$X_A(s,t) = (1, \cos(2\pi t/365), \sin(2\pi t/365), \mathrm{ST}(t), \ldots) \tag{3.4}$$

Following Verdin et al. (2015b), the minimum and maximum temperatures, $Z_N(s,t)$ and $Z_X(s,t)$, respectively, at location s and day t are decomposed as follows:

$$Z_N(s,t) = \boldsymbol{\beta}_N(s)' X_N(s,t) + W_N(s,t) \tag{3.5}$$

$$Z_X(s,t) = \boldsymbol{\beta}_X(s)' X_X(s,t) + W_X(s,t) \tag{3.6}$$

In each equation, the product on the right side of the equality is a regression on some covariates, $X_N(s,t)$ and $X_X(s,t)$ for minimum and maximum temperatures, respectively; these products represent the average behavior of temperatures over the observational period. In this, the key modification to the weather generator of the previous chapter is the inclusion of areal seasonal mean minimum ($\mathrm{SMN}(t)$) and maximum ($\mathrm{SMX}(t)$) temperature covariates, which are included in both temperature models, as follows,

$$\begin{aligned} X_N(s,t) = (1, \cos(2\pi t/365), \sin(2\pi t/365), r(t), Z_N(s,t-1), \\ Z_X(s,t-1), O(s,t), \mathrm{SMN}(t), \mathrm{SMX}(t), \ldots) \end{aligned} \tag{3.7}$$

$$\begin{aligned} X_X(s,t) = (1, \cos(2\pi t/365), \sin(2\pi t/365), r(t), Z_N(s,t-1), \\ Z_X(s,t-1), O(s,t), \mathrm{SMN}(t), \mathrm{SMX}(t), \ldots), \end{aligned} \tag{3.8}$$

which are the intercept term, two harmonic terms to account for seasonality, $r(t)$, which is a linear drift ranging from -1 to 1 to account for temperature trends over the observational period, the previous day's minimum and maximum temperatures, the current day's precipitation occurrence, and the seasonal mean minimum and mean maximum temperatures, respectively. Daily variability is denoted as $W_N(s,t)$ and $W_X(s,t)$ for minimum and maximum temperatures, respectively, and maintains spatial correlation by realizations from a mean zero Gaussian process with an empirical covariance structure defined by the residuals of the local regressions. Kleiber et al. (2013) found that the Gaussian assumption for minimum and maximum temperature models was appropriate. The above are GLMs and are fitted hierarchically – we refer the reader to Verdin et al. (2015b) and Kleiber et al. (2012, 2013) for details on implementation.

It should be noted that the additional covariates are applied only to the local climate component, and not the daily weather component. The daily weather component is by definition random, temporally independent noise (see Kleiber et al. (2012) for validation of this assumption), thus the daily weather component is not conditional on the additional covariates, rather it is conditioned only by the calendar date – there are distinct correlation (and covariance) matrices for each month.

3.3.2 Significance testing

The inclusion of seasonal covariates could lead to a reduction in the significance of the harmonic covariates. For all 17 stations the seasonal covariates are highly significant, indicated by the respective p-values of their regression coefficients. For many of the stations both cosine and sine covariates remain highly significant, however, at few stations the sine covariate loses significance. The Akaike information criterion (AIC) of the modified models at each location for each climate variable are consistently lower than those for the original models that do not contain the seasonal covariates, implying that the modified weather generator more adequately describes the modeled processes. Table

3.1 reports the change in AIC value (original minus modified models – positivity implies a decrease) for all 17 stations, for the four variables that make up the weather generators.

Station	1	2	3	4	5	6	7	8	9
OCC	112	92	66	98	56	122	92	112	106
AMT	103	29	20	89	17	63	74	92	97
MIN	732	439	420	559	157	489	340	661	625
MAX	240	186	165	254	39	144	196	165	237
Station	10	11	12	13	14	15	16	17	
OCC	89	103	48	134	114	98	127	98	
AMT	82	87	66	50	63	54	66	58	
MIN	560	540	258	603	495	547	877	403	
MAX	176	181	104	231	264	244	194	219	

Table 3.1: Differences between the AIC for occurrence, amounts, minimum temperature, and maximum temperature models of the original and modified weather generators (positivity implies a decrease).

3.4 Results from application in the Salado *A* sub-basin

3.4.1 Covariate selection

We apply the methodology as described in the previous section to the network of 17 stations in and around the Salado *A* sub-basin of the Argentine Pampas (see Figure 3.1). Given the relative homogeneity of the sub-basin area and scale at which seasonal climate forecasts are available, three domain-averaged covariates are proposed: seasonal total precipitation, seasonal mean minimum temperature, and seasonal mean maximum temperature. The growing season for summer crops in the Salado *A* sub-basin begins in October with harvest coming in late March or April, therefore we focus on the OND season. Seasons are defined as January-March (JFM), April-June (AMJ), July-September (JAS), and October-December (OND).

The first principal component of OND seasonal total precipitation explains 47% of the total variance; those of OND seasonal average minimum and maximum tempera-

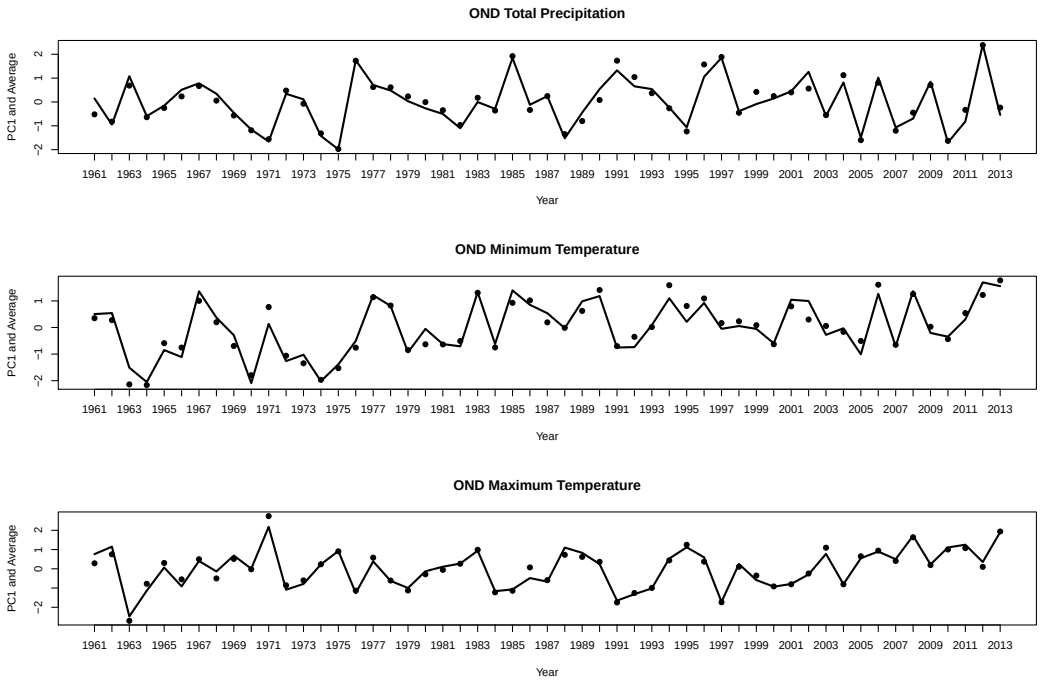

Figure 3.2: First principal components of OND precipitation, minimum temperature, and maximum temperature, scaled and shown as points, and domain-averaged and scaled OND precipitation (top), minimum temperature (middle), and maximum temperature (bottom), shown as lines.

tures explain 71% and 77% of the total variance, respectively. The magnitudes of these first principal components are nearly constant across space, which further justifies the use of domain-averaged information. Figure 3.2 shows the first principal component of the three variables along with the domain-averaged time series, the behavior of which are well described by their first principal components. Thus the four GLMs as described in the previous section were fitted with the additional covariates described above. These covariates were found to be highly significant at all the locations (e.g., regression assigns all additional covariates p-values < 0.001). Other seasons show similar results (not shown).

3.4.2 Validation

To assess the efficacy of the additional covariates, we employ both the original and modified weather generators in spatial and temporal validations, described in the following subsections.

3.4.2.1 Spatial validation

To assess the spatial performance of the modified weather generator, three stations were withheld from the model fitting process – these withheld stations are identified in Figure 3.1. Spatial process models were used to estimate the model parameters at the withheld locations, and 100 realizations over the 53-year observational period were produced using the estimated parameters. Figure 3.3 shows the relationship between the observed and ensemble mean OND probability of occurrence, seasonal total rainfall, mean maximum temperature, and mean minimum temperature for each of the three stations as produced by the original (top row) and modified (bottom row) weather generators. Simulations from the original weather generator show no relationship with the observations; this is to be expected, as only harmonic and autoregressive covariates are considered. Conversely, simulations from the modified generator capture the observations strongly, due to the inclusion of domain-averaged seasonal covariates. Similar results were seen for other seasons (figures not shown).

3.4.2.2 Temporal validation

It is also worthwhile to investigate the temporal performance of the weather generator to validate its use for seasonal forecasts, multi-decadal projections, and climate change scenarios. To this end, we fitted the original and modified weather generators on historic data for the calibration period: 1 January 1961 – 31 December 2000. Then 100 realizations were generated for the validation period: 1 January 2001 – 31 December 2013. Figure 3.4

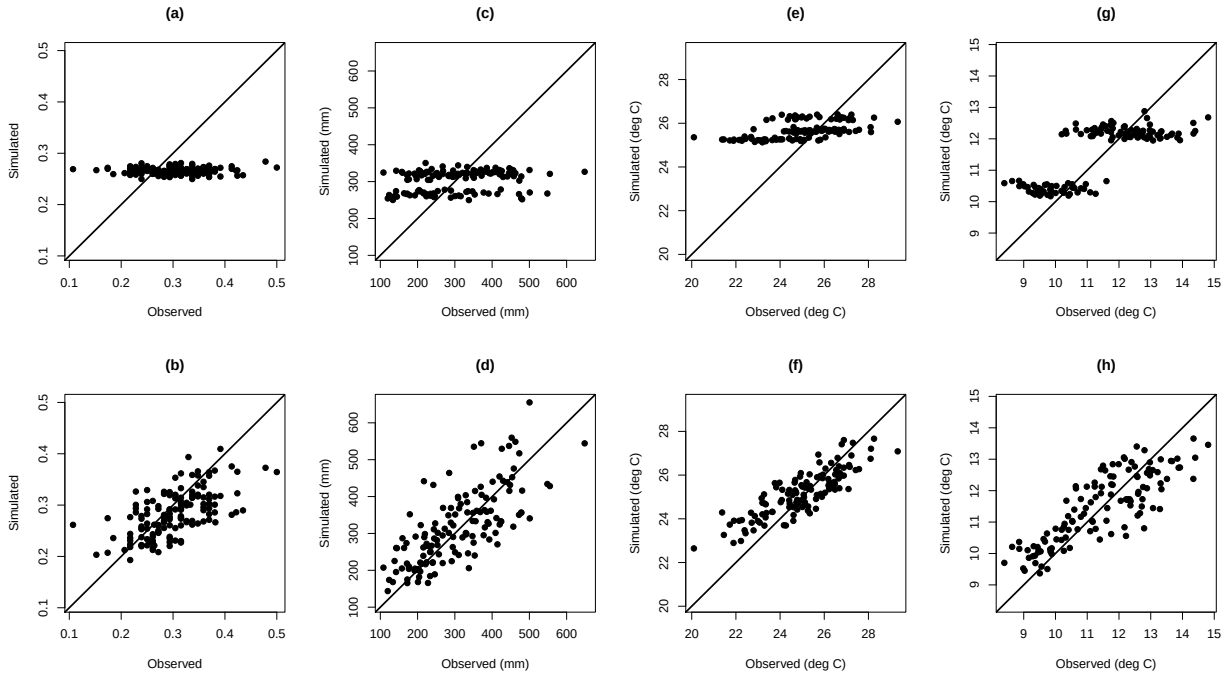

Figure 3.3: Spatial validation: (a-b) OND 1961-2013 observed versus ensemble mean simulated probability of precipitation occurrence, (c-d) total precipitation, (e-f) mean maximum temperature, and (g-h) mean minimum temperature, for the three withheld stations. Top row (a,c,e,g) corresponds to simulations from the original generator and bottom row (b,d,f,h) is for simulations from the modified generator.

shows the difference between observed and ensemble mean simulated domain-averaged seasonal total precipitation, mean maximum temperature, and mean minimum temperature: a "perfect fit" would show a horizontal line of ordinate zero. Root mean square error (RMSE) is calculated between simulated and observed seasonal values for all 100 realizations. The RMSE is greatly reduced by including the domain-averaged seasonal covariates in the validation period. For seasonal total precipitation the RMSE is reduced from 77 ($\pm$ 2.4) mm to 21 ($\pm$ 3.6) mm; for seasonal mean maximum temperature the RMSE is reduced from 1.05 ($\pm$ 0.05)$^\circ$ C to 0.37 ($\pm$ 0.05)$^\circ$ C; and for seasonal mean minimum temperature the RMSE is reduced from 0.99 ($\pm$ 0.04)$^\circ$ C to 0.37 ($\pm$ 0.04)$^\circ$ C.

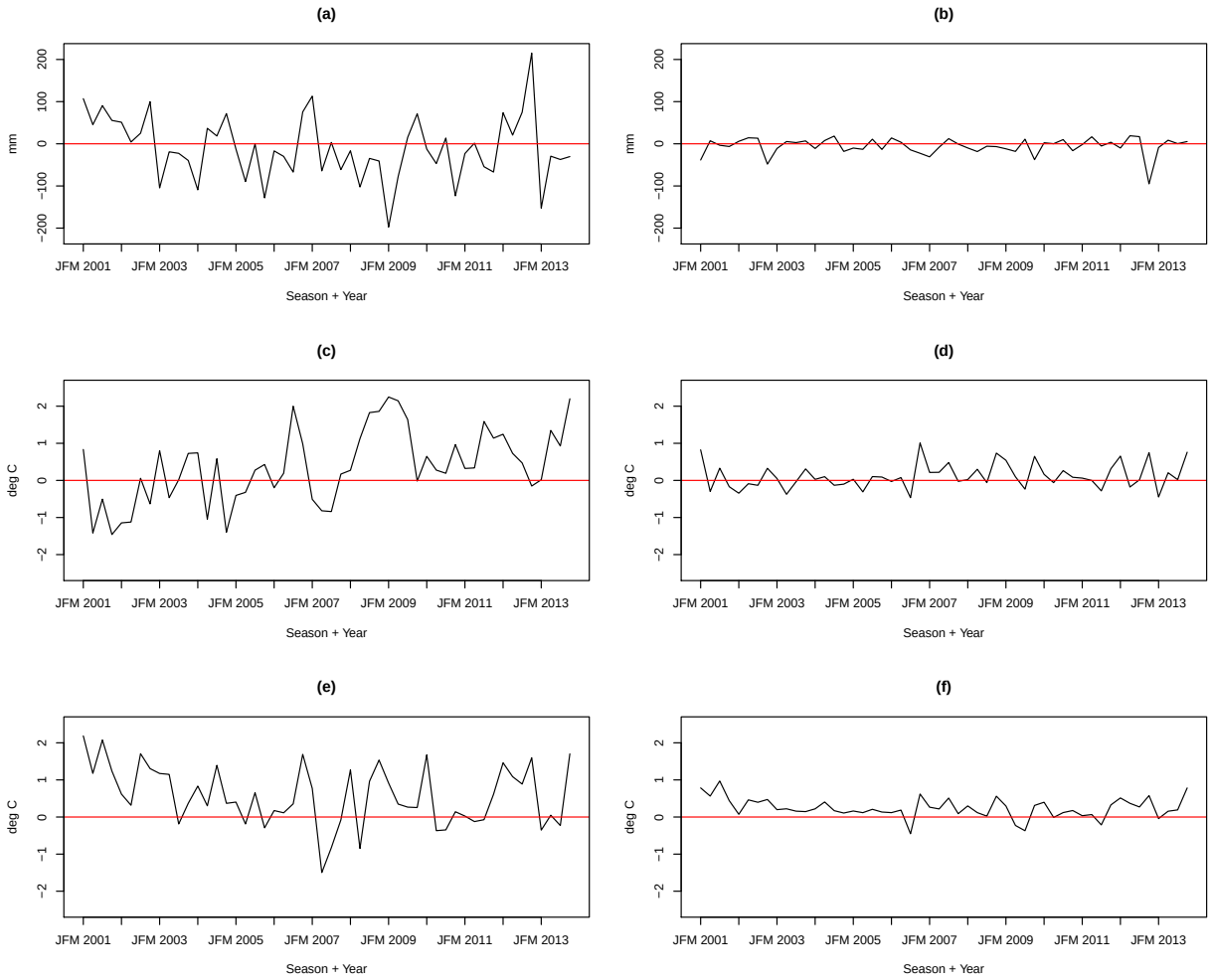

Figure 3.4: Temporal validation: JFM 2001 – OND 2013 observed minus ensemble mean simulated (a-b) seasonal total precipitation, (c-d) mean maximum temperature, and (e-f) mean minimum temperature. Left panels (a,c,e) are for the original weather generator and right panels (b,d,f) for the modified weather generator.

3.4.3 Seasonal forecasts

Often times, seasonal climate forecasts are provided as probabilities of precipitation and temperature being within different ranges (e.g., terciles) for a large region – this is a common format for presenting uncertain climate information. Among other agencies around the world, the International Research Institute for Climate and Society (IRI, www.iri.columbia.edu) provides seasonal (three-month) probabilistic forecasts with one

to four months' lead-time. The IRI presents these forecasts in terms of A:N:B likelihoods, where "A" is above-normal, "N" is near-normal, and "B" is below-normal. The three categories span an equal range and are defined with respect to climatological terciles (e.g., 33^{rd} and 67^{th} percentiles). For example, a 15:35:50 precipitation forecast implies there is a 15% chance of experiencing above-normal conditions, a 35% chance of experiencing near-normal conditions, and a 50% chance of experiencing below-normal precipitation in the upcoming season.

Agricultural decisions in the Salado A sub-basin are typically made before the beginning of the summer growing season (1 October) every year, thus we focus on OND seasonal forecasts. The OND season is also a critical period in terms of crop yield generation, and has shown tendencies towards skillful climate predictions, in part due to significant ENSO signals (Grimm et al., 2000; Montecinos et al., 2000; Ropelewski and Halpert, 1987; Barros and Silvestri, 2002; Boulanger et al., 2005; Grimm et al., 1998; Ropelewski and Bell, 2008; Grimm, 2011; Barreiro, 2010). We select IRI forecasts for OND 2010 (a dry and hot forecast; e.g., 15:35:50 for precipitation, 40:35:25 for temperature) and OND 2012 (a wet and hot forecast; e.g., 40:35:25 for both precipitation and temperature) as case studies for this methodology, issued on 1 September 2010 and 1 September 2012, respectively.

To generate space-time weather sequences for the two OND seasons from the modified generator, ensembles of domain-averaged seasonal precipitation and temperature are needed to use as covariates. To this end, 100 observed OND domain-averaged values of precipitation and maximum and minimum temperature are sampled with replacement. This is accomplished by first categorizing the observed domain-averaged seasonal weather as above-, near-, or below-normal based on the empirical terciles, then assigning the categorical forecasts as probabilities (e.g., 15:35:50 and 40:35:25 for precipitation and temperature, respectively) to the values in each category and resampling with these assigned weights as the probability metric. For instance, there is a 15% chance of sampling

an above-normal precipitation value; there are 35% and 50% chances of sampling near-normal and below-normal precipitation values, respectively. The result of this resampling scheme is 100 values that are used as covariates to drive the modified weather generator 100 separate times. The output of these 100 independent runs is essentially a downscaled ensemble of weather patterns that exhibit the traits of the seasonal forecasts.

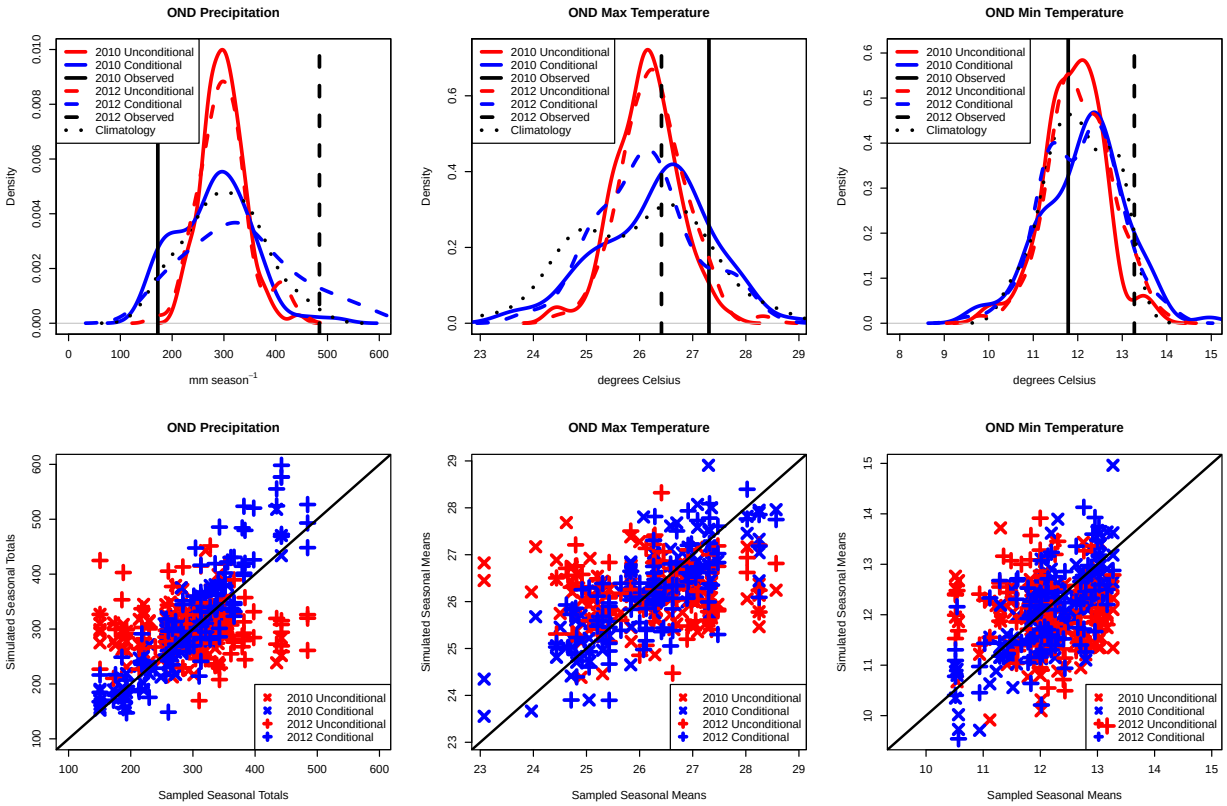

Figure 3.5: Top panels: Kernel density estimates of PDF of domain-averaged seasonal precipitation, maximum temperature and minimum temperature from 100 simulated weather scenarios from the modified (blue) and original (red) weather generators (OND 2010 and OND 2012 denoted as solid and dashed lines, respectively), along with the climatological PDF (dotted black line). Observed values are shown as vertical lines. Bottom panels: Sampled seasonal precipitation and temperatures from the categorical probabilistic forecasts with the domain-averaged values generated from the two weather generators – modified (blue) and original (red).

The top row of Figure 3.5 shows the probability density functions (PDFs) of domain-averaged OND precipitation and temperatures from the original and modified weather

generators, the PDF of OND climatology, and the observed values of OND 2010 and 2012. The precipitation PDFs from the modified generator have shifted towards the observed values in both 2010 and 2012. This shift is indicative not only of forecast skill, but also the effectiveness of the modified generator in simulating scenarios representative of the forecasts. Mean maximum (minimum) temperature during OND 2010 (2012) was observed to be above-normal, and the PDF from the modified generator gives greater probability to above-average temperatures than that of the original generator. OND 2012 (2010) experienced near-normal maximum (minimum) temperatures, so the original generator gives highest probability to observed conditions. However, the range of possible scenarios offered by the original generator is limited and will give near-zero probability to above- and below-normal conditions, which for planning purposes can be misleading. The domain-averaged seasonal totals of precipitation and seasonal averages of temperatures that are generated from the two weather generators are plotted with the observed in the bottom row of Figure 3.5.

Table 3.2 reports p-values from Kolmogorov-Smirnov tests comparing the distributions of original and modified generator output. The differences between the original and modified distributions for OND 2010 weather scenarios and OND 2012 precipitation are significant at the 95% level; maximum and minimum temperature scenarios for OND 2012 are not significantly different, indicating the covariate values sampled from the IRI probabilistic forecast (thus the scenarios produced by the modified weather generator) do not deviate significantly from climatology.

2010			2012		
Precip	Max Temp	Min Temp	Precip	Max Temp	Min Temp
0.0039	0.0014	0.0243	$<$0.001	0.0541	0.5806

Table 3.2: Kolmogorov-Smirnov test comparing the original and modified weather generator output. P-values lower than 0.05 indicate the output from original and modified generators come from different distributions.

Weather simulations on a regular grid are of particular interest, as they are used to drive hydrologic and agriculture models for agricultural planning to mitigate crop failure. To simulate daily weather on a grid, the β coefficients for each covariate of the weather generator models are estimated in space from their respective spatial models to the desired spatial resolution (5 km x 5 km). These gridded coefficients are then used to obtain the mean function, and the daily weather processes are simulated via mean zero Gaussian random fields.

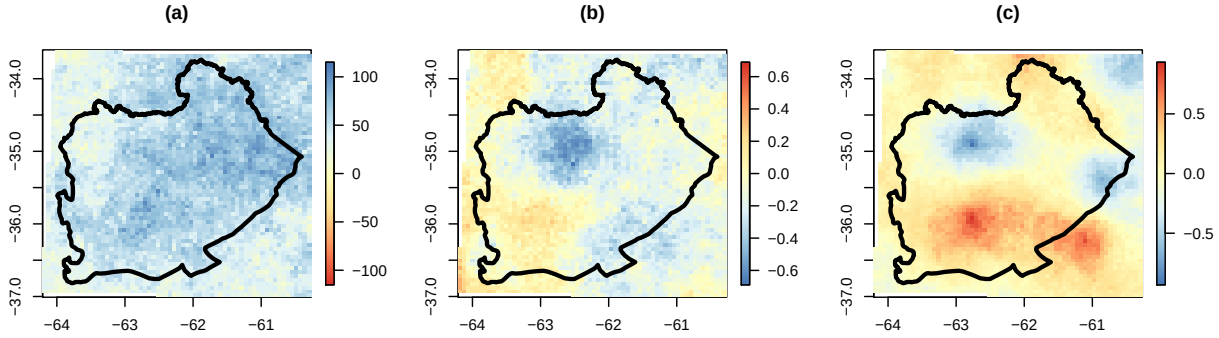

Figure 3.6: OND 2010 differences in ensemble mean of seasonal (a) total precipitation (mm season-1), (b) mean maximum temperature (deg C), and (c) mean minimum temperature. Differences calculated as original minus modified generators. Salado *A* sub-basin is outlined.

Figure 3.6 shows the difference between the ensemble mean of gridded seasonal total precipitation, mean maximum temperature, and mean minimum temperature for OND 2010 from the original and modified weather generators. The modified generator simulates a drier and hotter domain than the original generator, which is consistent with the seasonal forecast. Notably, the modified weather generator simulates a cooling for mean minimum temperature in the southern part of the sub-basin, which is inconsistent with the seasonal forecast.

The differences between the 95% ensemble spread (97.5^{th} percentile minus 2.5^{th} percentile) produced by the original and modified weather generators are shown in Figure

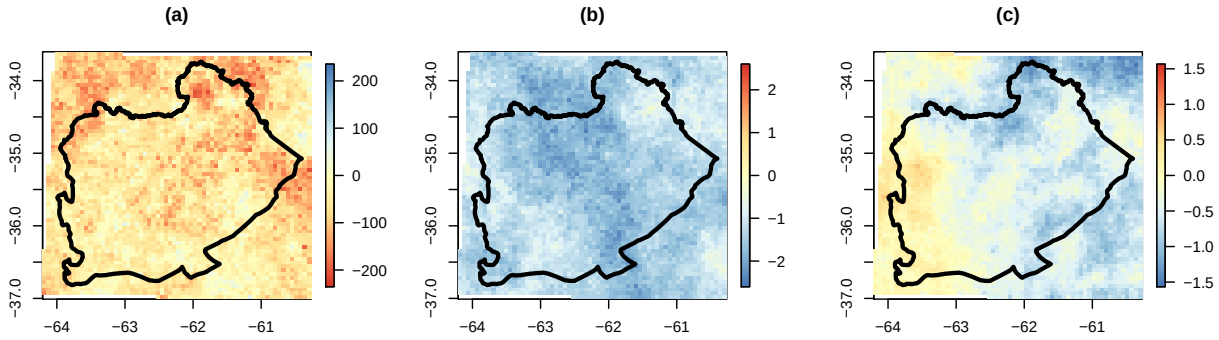

Figure 3.7: OND 2010 differences in 95% ensemble spread for (a) seasonal total precipitation (mm season-1), (b) seasonal mean maximum temperature (deg C), and (c) seasonal mean minimum temperature. Differences calculated as original minus modified generators. Salado A sub-basin is outlined.

3.7. As can be seen, the ensemble spread difference for seasonal total precipitation is mostly red and yellow, while those for mean maximum and minimum temperature are mostly blue and yellow, which illustrates that the modified generator produced wider ensemble spreads than the original generator. The uncertainty in the probabilistic seasonal climate forecast is propagated to the conditional weather generator, resulting in a wider distribution than that of the original generator. Similar findings can be seen for OND 2012 (figures not shown).

3.4.4 Multi-decadal projections

Multi-decadal projections are useful in a number of applications, including environmental impact studies, agricultural decision-making, and water resources management, to name a few. In agriculture, multi-decadal projections help in making informed investment decisions (e.g., whether or not to buy a farm in a climatically marginal area, invest in irrigation, etc.). Specifically, the climate of the Pampas has shown significant decadal variability, and since the 1970s has exhibited a steady increase in both annual

and extreme precipitation. This trend in precipitation has in part promoted significant expansion of agricultural area to climatically marginal regions of the Pampas. Given the uncertainty of future climate, coupled with a known decadal variability, it is unclear if existing agricultural systems may remain viable if climate reverts to a drier epoch.

Specific to this research, future climate scenarios can be used to drive hydrologic and crop simulation models, thus providing an assessment of the viability of existing agricultural production systems in climatically marginal regions of the Salado *A* sub-basin. However, future climate projections from climate models are generally of coarse spatial (e.g., on a grid) and temporal (e.g., monthly) resolutions, and therefore cannot provide reliable projections of weather at the local scale. These monthly and consequently seasonal projections can be incorporated into the modified generator and thus enable the generation of daily weather sequences conditioned on the projections at any desired location – both monitored and unmonitored – in the study region.

To this end, we explored the ability of the modified generator to downscale medium-term projections in the Salado *A* sub-basin. A regional climate model projection, experiment RCP8.5, was obtained for the period 1 January 2015 to 31 December 2050 (a 36-year projection), produced using the CORDEX-CMIP5 regional climate model (Consortium, 2014) and bias-corrected (McGinnis et al., 2015) using the CLARIS-LPB dataset (Penalba et al., 2014). This projection focuses on South America and is gridded at 0.44 degrees. No notable long-term trends in annual precipitation totals are projected, but the magnitudes are significantly lower than seen in the historic record; both maximum and minimum annual average temperatures show positive trends, and are projected to increase by approximately 1° C by the year 2050. Seasonal values of areal precipitation and temperature for the Salado *A* sub-basin were computed to use as covariates to drive the modified weather generator. Only the grids that cover the Salado *A* sub-basin and the 17 station data are considered when computing domain-averaged precipitation totals and temperature means. 100 realizations of daily weather sequences were simulated using both the

original and modified weather generators.

Figure 3.8 shows seasonal residuals (projected minus simulated) of the ensemble mean of the original and modified weather generator simulations for the 36-year projection period. As was seen previously, the original generator shows much larger and more variable residuals than does the modified generator. Consistent with the temporal validation, the RMSE is greatly reduced by including the domain-averaged seasonal covariates in the models. RMSE for seasonal total precipitation is reduced from 90 mm to 14 mm; RMSE for seasonal mean maximum temperature is reduced from 1.09° C to 0.48° C; RMSE for seasonal mean minimum temperature is reduced from 1.43° C to 0.23° C. There is a slight warm bias in seasonal mean maximum temperature simulated by the modified generator as compared to that projected by the RCM.

To illustrate the spatial ability of the weather generators, 100 realizations of daily weather sequences are simulated for the period 2015-2050 conditioned on the projected seasonal characteristics. Figure 3.9 shows the difference in ensemble mean OND total precipitation, maximum temperature, and minimum temperature as simulated by the original and modified generators. Consistent with the climate model trends, the modified generator simulates a drier and hotter future across the domain.

3.5 Summary and future work

We have proposed and validated the use of a parametric stochastic weather generator in a nonstationary context, such as climate change impact studies, with application in the Salado *A* sub-basin of the Argentine Pampas. This region was selected due to its status as one of the most productive agricultural regions in South America, and its strong climatic variability that is experienced at multiple time scales. Agriculture in the Pampas is predominantly rainfed, thus high quality seasonal forecast information could greatly impact the outcome (e.g., crop yield, risk of failure) of a growing season. The modified

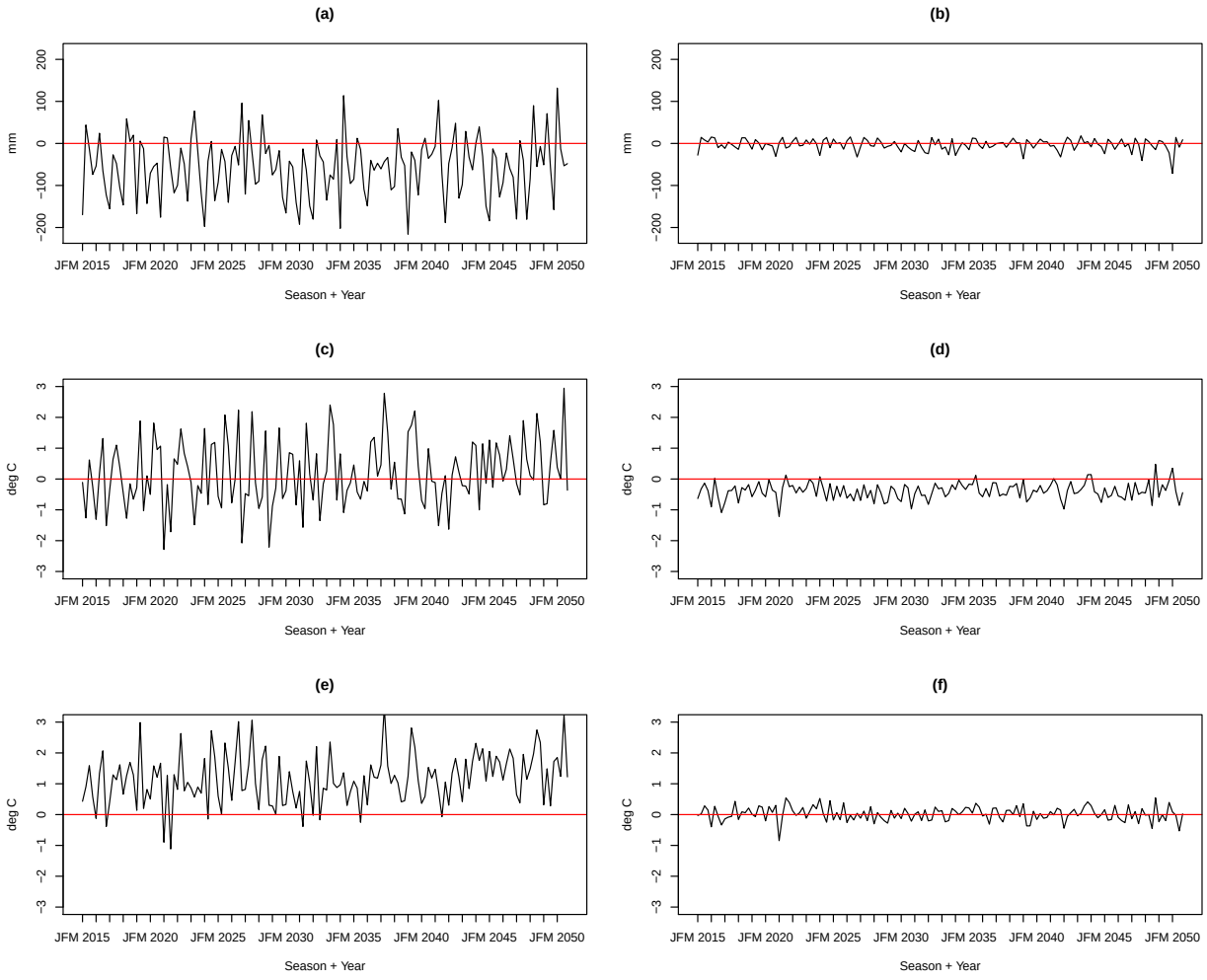

Figure 3.8: JFM 2015 – OND 2050 projected minus ensemble mean (a-b) simulated seasonal total precipitation, (c-d) mean maximum temperature, and (e-f) mean minimum temperature.

weather generator presented in this research has flexibility in its GLM framework such that any number of covariates can be included in the model fit, effectively conditioning the weather generator to produce downscaled weather sequences.

For example, in this research we used areal average seasonal total precipitation and mean minimum and maximum temperatures as additional covariates, which were shown to be highly significant in the model fit. The use of areal averages was justified via principal component analysis; for non-homogeneous or mountainous regions, con-

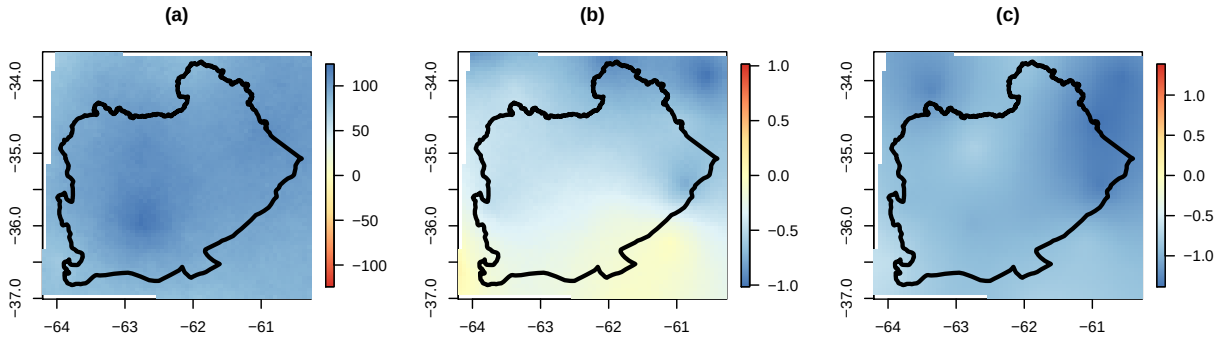

Figure 3.9: OND 2015-2050 differences in ensemble mean of seasonal (a) total precipitation (mm season-1), (b) mean maximum temperature (deg C), and (c) mean minimum temperature. Differences calculated as original minus modified generators. Salado A sub-basin is outlined.

sider site-specific averages or a clustering algorithm. The coarse information provided by these additional covariates successfully trickled from seasonal (regional) down to daily (local) scales, such that wet (dry) days are more prevalent during seasons with above-normal (below-normal) seasonal total precipitation. It is with the conditioned output of the weather generator that research teams may provide a more robust estimate of production risk for a region, by running the daily weather sequences through process based (e.g., crop simulation, hydrologic) models. The output of process based models may be interpreted and provided to a farmer or decision maker, who then will have seasonal forecast information that is relevant to the decisions they must make (e.g., probability of not meeting a crop yield goal, where and when to plant a certain crop) as opposed to spatially coarse probabilistic statements as are typically reported. Similarly, multi-decadal projection information can be used to generate conditional weather sequences to assist in assessing the viability of existing agricultural infrastructure in climatically marginal regions. In this, a regional climatic trend may be extracted and used to produce conditional weather sequences, which may be used to drive any relevant process based models.

The output of the modified weather generator presented in this manuscript has been

validated by direct comparison to the original weather generator of Verdin et al. (2015b). It has been shown that using simple covariates such as domain-averaged seasonal total (mean) precipitation (temperatures) improves the skill of the generator in producing daily weather sequences that exhibit the traits (and trends) of a seasonal forecast or multi-decadal projection. In representing domain-averaged behavior for the temporal validation period (2001-2013), this modification to the weather generator reduced RMSE values from 77 mm to 21 mm for precipitation, 1.05° C to 0.37° C for maximum temperature, and 0.99° C to 0.37° C for minimum temperature. Similarly, the modified generator faithfully reproduced the trends and variability of historic precipitation and temperature at individual sites, while the original generator replicates the expected behavior of (e.g., climatology) of each season with little to no interannual variability.

In generating sequences consistent with a seasonal forecast, the Kolmogorov-Smirnov tests suggest the output from original and modified weather generators exhibit significantly different traits with 95% confidence, unless the seasonal forecast is similar to climatology. The modified weather generator was shown to produce PDFs that better represent the range of possible futures, while the PDFs from the original weather generator give near-zero probability to the upper and lower terciles (e.g., wet (hot) and dry (cold) conditions). On the multi-decadal scale, the modified weather generator is flexible in its ability to capture the considerable interannual and decadal variability prevalent in the projected precipitation totals, as well as the increase in both minimum and maximum temperatures.

Application of this methodology to other areas is called for. However, careful attention must be paid to the spatial and temporal climatic variability in the region of interest. Local climate, regional teleconnections, and global climate drivers should be identified for optimal skill in downscaling seasonal forecasts and multi-decadal projections. Principal component analysis on seasonal attributes such as seasonal total precipitation and mean temperatures can help decide if domain-averaged, clustered, or site-specific covari-

ates should be considered. However, the model setup, as defined in Verdin et al. (2015b), should be considered a baseline model for use in any basin – the additional covariates as described in this manuscript need be fine tuned to successfully generate skillful weather scenarios.

One shortcoming to the weather generator presented in this research is that it uses only precipitation and temperature covariates to condition the weather generator. There has been great progress in the identification of teleconnections and climate drivers for regions around the world – the Pampas are no exception. While it was mentioned in this manuscript that such teleconnections could be used as covariates to condition the weather generator output, this approach was not investigated. A second shortcoming to this methodology is that the uncertainties associated with the parameters of the weather generator are not propagated to the simulations, as the maximum likelihood estimates of the parameters are kept fixed. As a result, the variability of simulations can be underestimated. Bayesian methods that explicitly quantify the parameter uncertainties are attractive options.

The methodology of this weather generator is inherently hierarchical, thus considering the use of a Bayesian hierarchical framework is a natural extension to this problem. In a Bayesian context, the parameters are treated as random variables and are sampled from appropriate distributions (typically via Markov chain Monte Carlo) based on likelihood acceptance criteria, which results in posterior distributions of all model parameters. These posterior distributions better represent the uncertainty involved in traditional parameter estimation techniques, and when used in a weather generation framework will provide a more realistic range of uncertainty in synthetic weather sequences.

Chapter 4

BayGEN: a Bayesian space-time stochastic weather generator

4.1 Introduction

Time series of weather variables can assist in agricultural and water resources planning and decision-making via crop simulation, hydrologic, and other process-based models (Ferreyra et al., 2001a; Berger, 2000; Berger et al., 2006; Happe et al., 2008; Freeman et al., 2009; Schreinemachers and Berger, 2011; Bert et al., 2006, 2007, 2014). However, the historical record can be limited in that it covers too short of a period or it can be incomplete or inaccurate due to missing data and human error. Many process-based models require complete sequences of input data (i.e., missing data are not allowed), so the problem of "filling in the blanks" arises. If the problem requires the use of distributed hydrologic models, the user is often required to input daily weather sequences on a regular grid, which is difficult, if not impossible, to collect for even moderately sized basins. Stochastic weather generators have long been used as a way to produce daily weather sequences for use in such applications, as they provide a means to simulate a plausible range of climatic scenarios while representing the true underlying variability of the natural processes being modeled. For agricultural planning on seasonal to multi-decadal scales, weather generators can be used to simulate input data consistent with forecasted scenarios (Apipattanavis et al., 2010; Verdin et al., 2016), which can be used to drive crop simulation models to assess how such scenarios may impact crop yields. A more complete history on stochastic weather generation can be found in Wilks and Wilby (1999)

and Verdin et al. (2015b); a brief summary is provided below.

Long held in regard as the original stochastic weather generator, the methodology of Richardson (1981) has paved the way for other approaches to producing synthetic daily weather sequences. These traditional single-site weather generators model precipitation occurrence as a chain-dependent process (Katz, 1977), thus they are able to capture the length and frequency of wet and dry spells. Precipitation amounts are modeled using probability distributions (e.g., Gamma, Lognormal, etc.). Minimum and maximum temperatures are modeled using autoregressive (AR) time series models. There have been several multi-site extensions to traditional weather generators (Wilks, 1998; Wilks and Wilby, 1999; Qian et al., 2002; Baigorria and Jones, 2010a; Khalili et al., 2009). However, parametric weather generators can quickly become unwieldy when applied for multi-site simulation, as a large number of parameters is needed to capture the spatial correlation and statistics (Mehrotra et al., 2006). Recent works (Wilks, 2008, 2009a; Kleiber et al., 2012, 2013; Verdin et al., 2015b) use spatial process models to interpolate weather generator parameters to enable the simulation of gridded daily weather.

Nonparametric weather generators using K-nearest neighbor time series bootstrapping, proposed for single-site (Rajagopalan and Lall, 1999), and subsequently extended to multi-site (Buishand and Brandsma, 2001; Beersma and Buishand, 2003; Yates et al., 2003; Sharif and Burn, 2007; Apipattanavis et al., 2010) are robust in their ability to capture nonlinearities in both space and time, and can easily capture the spatial correlation and statistics for multi-site simulation. However, they have two main limitations – (i) due to bootstrap resampling of the historical record, values outside the historical observations are not possible, and (ii) they cannot easily simulate sequences at unobserved locations.

Generalized linear models (GLMs) can greatly reduce the effort of modeling non-normal variables through a suite of link functions, and can more easily simulate at arbitrary locations when coupled with spatial models (Verdin et al., 2015b). Early use of GLMs for weather generation was by Stern and Coe (1984), with subsequent work by

Yang et al. (2005) and Chandler (2005). GLMs can incorporate any number of covariates to facilitate the simulation of weather sequences under various conditions such as wet or dry years. Such covariates have included large scale climate drivers (i.e., El Niño Southern Oscillation, North Atlantic Oscillation, etc.) and total seasonal precipitation and mean temperature trajectories, to name a few (Furrer and Katz, 2008; Hauser and Demirov, 2013; Verdin et al., 2016).

GLM-based space-time stochastic weather generators have become more popular recently due to their ability to generate weather sequences at any desired spatial resolution and conditioned on large scale climate drivers. However, they do not incorporate model parameter uncertainty in their weather simulations. Specifically, point estimates of model parameters based on maximum likelihood are fixed for all ensemble members, and the variability between ensemble members is obtained solely from the residual terms. Lack of treatment of parameter uncertainty is a major shortcoming and it limits the variability of the simulated weather sequences, which, from a planning perspective, can lead to overconfident estimates of system variables by not representing the full range of possible future scenarios. Bayesian methods are known for their ability to capture uncertainty by providing the full distribution of model parameters. A Bayesian stochastic weather generator would be advantageous in that parameter uncertainty would propagate to the simulated weather sequences, and consequently to the process based model output, which in turn would provide better estimates of uncertainty and risk.

Bayesian methods are widely used to quantify uncertainty in a variety of models and applications. They are computationally intensive, but with increasing computation power, Bayesian methods are becoming more attractive. One of the more common Bayesian methods is Bayesian model averaging, first introduced by Leamer (1978), and used in a wide array of disciplines – including hydrology, climate science, ecology, economics, medicine, politics, among others – for combining information from multiple sources (e.g., multiple model outputs), making skillful predictions, calibrating fore-

casts, risk assessment, and quantifying model uncertainty (Raftery et al., 2005; Sloughter et al., 2007; Duan et al., 2007; Wintle et al., 2003; Wright, 2008; Volinsky et al., 1997; Vialle-font et al., 2001; Trujillo-Barreto et al., 2004; Vrugt and Robinson, 2007; Montgomery and Nyhan, 2010). Bayesian hierarchical models have been developed for a variety of applications – coupling with stochastic weather generators for enhanced decision making (Pezzulli et al., 2006; Hashmi et al., 2009); spatial and temporal modeling of precipitation extremes (Cooley et al., 2007; Cooley and Sain, 2010; Reich and Shaby, 2012); daily precipitation modeling (Smith and Robinson, 1997; Lima and Lall, 2009); analysis of precipitation and temperature trends (Tebaldi et al., 2004; Tebaldi and Sansó, 2009); and Bayesian kriging (Omre, 1987; Handcock and Stein, 1993; Cui et al., 1995; Sahu and Mardia, 2005; Aelion et al., 2009; Jin et al., 2014; Verdin et al., 2015a). A spatially-consistent Bayesian precipitation state generator was developed by Cano et al. (2004), though a Bayesian weather generator was yet to be developed.

With the motivation to fully quantify and incorporate parameter uncertainties in daily weather simulation, we developed a GLM-based Bayesian space-time stochastic weather generator, hereafter referred to as BayGEN. In this, posterior distributions of the GLM model parameters are obtained from the Bayesian framework; consequently these parametric uncertainties are propagated to the daily weather sequences – described in the following sections.

The BayGEN model is presented in Section 4.2, followed by a description of the study region and data in Section 4.3. We discuss the results in Section 4.4, and Section 4.5 concludes this manuscript with a summary of the research and future work.

4.2 BayGEN

4.2.1 Model definition

There are two layers in the BayGEN model. The first layer (i.e., data layer) explicitly models the data, in this case the observed precipitation occurrence and amount, and maximum and minimum temperatures at each location in the network. The second layer (i.e., process layer) models the latent process that drives the precipitation occurrence model, the covariates for all variables in the data layer, and the prior distributions on the model parameters.

Let $O(\boldsymbol{s}, t)$, $A(\boldsymbol{s}, t)$, $Z_X(\boldsymbol{s}, t)$, and $Z_N(\boldsymbol{s}, t)$ denote precipitation occurrence and amount, and maximum and minimum temperature, respectively, at location $\boldsymbol{s}$ and time t. The BayGEN model structure is motivated by a GLM framework similar to Verdin et al. (2015b). In the GLM weather generator (GLMGEN), at each location, precipitation occurrence is modeled using probit regression; precipitation amounts are modeled using Gamma distributions; minimum and maximum temperatures are modeled using linear regression. The coefficients of the GLMs are modeled as spatial Gaussian processes (GP) to enable the simulation of daily weather at arbitrary locations. We translate this GLM approach into a Bayesian hierarchical framework, which is defined as follows:

Data layer:

$$O(\boldsymbol{s},t) = \mathbb{1}_{[W_O(\boldsymbol{s},t)\geq 0]} \tag{4.1}$$

$$A(\boldsymbol{s},t) \sim Gamma(\alpha_A(\boldsymbol{s}), \alpha_A(\boldsymbol{s})/\mu_A(\boldsymbol{s},t)) \tag{4.2}$$

$$Z_N(\boldsymbol{s},t) \sim GP(\mu_N(\boldsymbol{s},t), C_N(t)) \tag{4.3}$$

$$Z_X(\boldsymbol{s},t) \sim GP(\mu_X(\boldsymbol{s},t), C_X(t)) \tag{4.4}$$

Process layer:

$$W_O(\boldsymbol{s},t) \sim N(\mu_O(\boldsymbol{s},t), \sigma_O^2) \tag{4.5}$$

$$\mu_A(\boldsymbol{s},t) = exp(\mathbf{X}_A(\boldsymbol{s},t)'\boldsymbol{\beta}_A(\boldsymbol{s})) \tag{4.6}$$

$$\mu_i(\boldsymbol{s},t) = \mathbf{X}_i(\boldsymbol{s},t)'\boldsymbol{\beta}_i(\boldsymbol{s}) \qquad \text{for } i = O,N,X \tag{4.7}$$

$$\boldsymbol{\beta}_j(\boldsymbol{s}) \sim GP(\hat{\boldsymbol{\beta}}_j(\boldsymbol{s}), C_{\boldsymbol{\beta}_j}) \qquad \text{for } j = O,A,N,X \tag{4.8}$$

$$\alpha_A(\boldsymbol{s}) \sim GP(\hat{\alpha}_A(\boldsymbol{s}), C_{\alpha_A}) \tag{4.9}$$

4.2.1.1 Precipitation occurrence

Precipitation occurrence, $O(\boldsymbol{s},t)$, which is a binary process where success (i.e., $O(\boldsymbol{s},t) = 1$) is defined as precipitation amounts exceeding 0.1 mm, is modeled using probit regression, denoted by $\mathbb{1}$, an indicator function with latent Gaussian process $W_O(\boldsymbol{s},t)$, which is assumed to be a realization from a normal distribution with mean $\mu_O(\boldsymbol{s},t)$ and variance unity (i.e., $\sigma_O^2 = 1$). If $W_O(\boldsymbol{s},t) \geq 0$, it is indicative of precipitation occurrence; if $W_O(\boldsymbol{s},t) < 0$, it implies precipitation has not occurred. The mean of the latent Gaussian process $\mu_O(\boldsymbol{s},t)$ is defined by a local regression on some covariates,

$$\mathbf{X}_O(\boldsymbol{s},t) = (1, O(\boldsymbol{s},t-1), \cos(2\pi t/365), \sin(2\pi t/365)). \tag{4.10}$$

The regression parameters $\boldsymbol{\beta}_O(\boldsymbol{s})$ are spatially correlated via Gaussian process prior distributions with spatially-varying mean $\hat{\boldsymbol{\beta}}_O(\boldsymbol{s})$ centered at the maximum likelihood estimates (MLE) and exponential covariance functions $C_{\boldsymbol{\beta}_O}$.

To simulate spatially continuous precipitation using probit regression, a correlation matrix is needed to drive random samples from a mean-zero multivariate normal distribution. A correlation matrix is used in place of a covariance matrix because probit regression has variance unity. It has been shown that multivariate probit regression is possible, however, for this application, multivariate probit regression would necessitate the inclusion of approximately 500,000 additional parameters, which quickly makes the model unwieldy and impossible to run in reasonable time. To this end, after the model is fitted, the probit regression residuals are calculated for each sample from the posterior distribution and for each location, and the spatial correlation matrices are calculated using a frequentist method. There are distinct correlation matrices, $C_O(t)$, for each calendar month to account for seasonality.

4.2.1.2 Precipitation amount

Precipitation amount is modeled as a Gamma random variable with spatially-varying shape $\alpha_A(\boldsymbol{s})$, to account for the climatological gradients, and spatially- and temporally-varying scale $\alpha_A(\boldsymbol{s})/\mu_A(\boldsymbol{s},t)$, to account for seasonality. Spatial dependence is captured by modeling $\alpha_A(\boldsymbol{s})$ as a realization from a Gaussian process with mean vector centered at the MLE $\hat{\alpha}_A$ and covariance function C_{α_A}. The $\boldsymbol{\beta}_A$ parameters are also modeled as realizations from Gaussian processes centered at the MLE $\hat{\boldsymbol{\beta}}_A$ with covariance functions $C_{\boldsymbol{\beta}_A}$. The mean function $\mu_A(\boldsymbol{s},t)$ is exponentiated to replicate a gamma GLM. The covariate vector,

$$\mathbf{X}_A(\boldsymbol{s},t) = (1, \cos(2\pi t/365), \mathrm{MT}(t)), \tag{4.11}$$

includes a seasonality covariate (cosine) to capture the generic seasonal cycle, and a scaled monthly average precipitation covariate (MT) to capture climatology. In simulation, a spatially-varying copula function (Chilès and Delfiner, 1999) is used to produce spatially consistent amounts, which is essentially a transform of the mean-zero Gaussian process

used to generate simulated occurrences (with correlation function $C_O(t)$); this approach avoids disagreement between the amounts and occurrence models.

4.2.1.3 Maximum and minimum temperature

As in Richardson (1981), we condition the maximum and minimum temperature models on precipitation occurrence. Maximum temperature $Z_X(\boldsymbol{s}, t)$ is assumed to be a realization from a Gaussian process with mean defined as a regression on some covariates,

$$\mathbf{X}_X(\boldsymbol{s}, t) = (1, Z_X(\boldsymbol{s}, t-1), \cos(2\pi t/365), \sin(2\pi t/365), O(\boldsymbol{s}, t)), \tag{4.12}$$

and monthly covariance functions $C_X(t)$. Minimum temperature $Z_N(\boldsymbol{s}, t)$ has a similar structure, such that it is assumed to be a realization from a Gaussian process with mean defined as a regression on some covariates,

$$\mathbf{X}_N(\boldsymbol{s}, t) = (1, Z_N(\boldsymbol{s}, t-1), \cos(2\pi t/365), \sin(2\pi t/365), O(\boldsymbol{s}, t)), \tag{4.13}$$

and monthly covariance functions $C_N(t)$. Note that both maximum and minimum temperatures on day t depend on precipitation occurrence on day t and maximum and minimum temperatures on day $t-1$. The $\boldsymbol{\beta}_X$ and $\boldsymbol{\beta}_N$ parameters are assumed to be realizations from Gaussian processes with means centered on the MLE $\hat{\boldsymbol{\beta}}_X$ and $\hat{\boldsymbol{\beta}}_N$, and covariance functions $C_{\boldsymbol{\beta}_X}$ and $C_{\boldsymbol{\beta}_N}$.

4.2.1.4 Spatial models

To enable simulation at any arbitrary location, regression coefficients $\boldsymbol{\beta}_i$, Gamma model shape parameter α_A, and maximum and minimum temperature residuals W_X and W_N, are spatially correlated via Gaussian processes with exponential covariance functions $C_{\boldsymbol{\beta}_i}$, C_{α_A}, $C_N(t)$, and $C_X(t)$, respectively, for $i = O, A, X, N$. We seek the posterior distribution for marginal variance σ_j^2 and effective range τ_j^2 for $j = O, A, X, N, \alpha_A, W_X, W_N$. For computational stability, we fix a finitely small nugget value of 0.001 for all spatial

processes. As mentioned in Section 4.2.1.1, the residual processes for precipitation occurrence have monthly correlation functions $C_O(t)$, and are estimated in post-processing.

4.2.2 Likelihood computation

We define θ as a vector of all model parameters,

$$\theta = (\boldsymbol{\beta}_O, \boldsymbol{\beta}_A, \boldsymbol{\beta}_X, \boldsymbol{\beta}_N, C_X(t), C_N(t), C_{\boldsymbol{\beta}_O}, C_{\boldsymbol{\beta}_A}, C_{\boldsymbol{\beta}_X}, C_{\boldsymbol{\beta}_N}, \alpha_A, C_{\alpha_A}), \tag{4.14}$$

and $\boldsymbol{Y}$ as the data,

$$\boldsymbol{Y} = (O(\boldsymbol{s}, t), A(\boldsymbol{s}, t), Z_X(\boldsymbol{s}, t), Z_N(\boldsymbol{s}, t)). \tag{4.15}$$

We seek $p(\theta|\boldsymbol{Y})$, the conditional distribution of the parameters given the data, which is also known as the posterior distribution. Using Bayes' rule, the posterior distribution is expressed as

$$p(\theta|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\theta)p(\theta)}{p(\boldsymbol{Y})} \propto p(\boldsymbol{Y}|\theta)p(\theta), \tag{4.16}$$

where $p(\boldsymbol{Y}|\theta)$ is the likelihood function of the data $\boldsymbol{Y}$ given the parameters θ, and $p(\theta)$ is the prior joint distribution of the parameters θ. Priors are defined as normal distributions centered on the maximum likelihood estimates with very large standard deviation (i.e., 100). Parameters with non-negative support are endowed with truncated normal distributions.

4.2.3 Implementation

We use the `R` (R Core Team, 2014) package *rstan* (Carpenter, 2015; Stan Development Team, 2015a,b) for model development, which employs the No-U Turn sampler, an adaptive form of Hamiltonian Monte Carlo sampling (see Hoffman and Gelman (2014)). Existing software typically uses Markov chain Monte Carlo sampling, which, while effective, can take a long time to converge with its "random walk" behavior. We have chosen the *rstan* package also because it provides an `R` interface to the Stan modeling language, which has an active users group mailing list for real-time troubleshooting.

4.3 Study region and data

Application of BayGEN is focused on the Salado *A* sub-basin (hereafter "the *A*") of the Argentine Pampas (see Figure 4.1). The climatology of the region exhibits longitudinal gradients, with higher temperatures and lower rainfall in the west; lower temperatures and higher rainfall in the east. The *A* is unique in that it is located where the climatological gradient is the greatest and is one of the most agriculturally productive regions of the Pampas. For a detailed description of the *A* and its hydroclimatology, see Verdin et al. (2016).

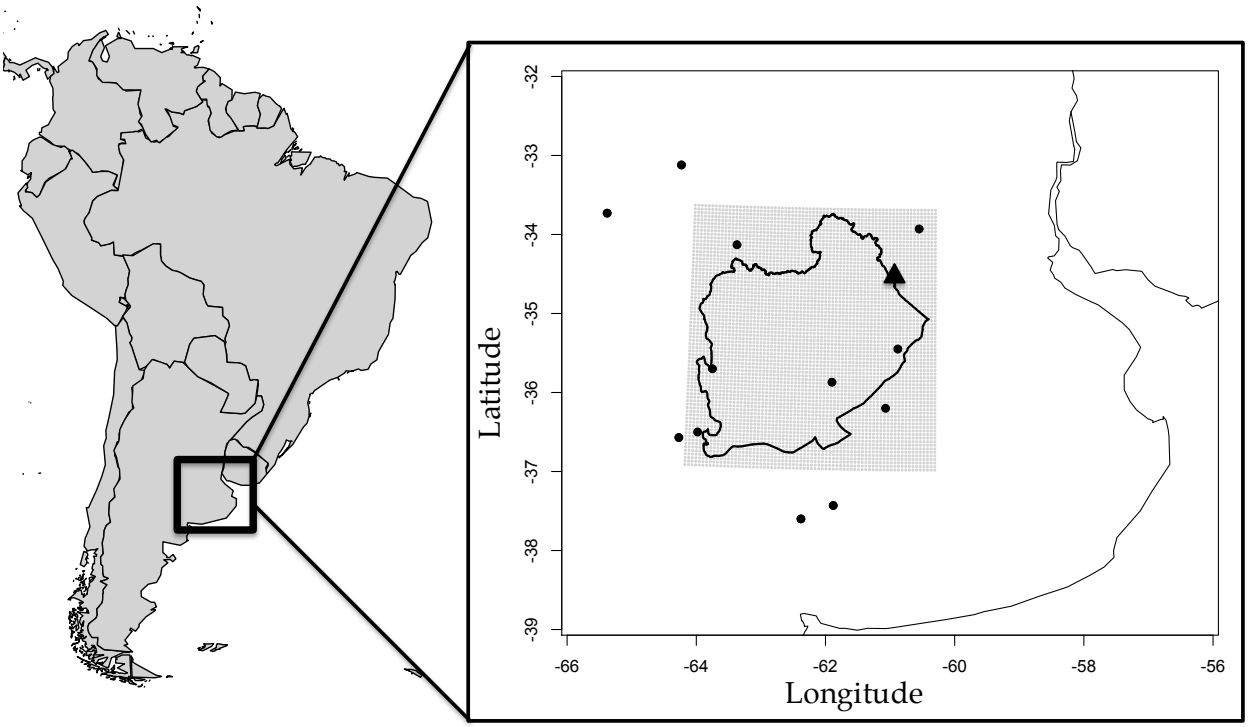

Figure 4.1: Salado *A* sub-basin location relative to South America; station locations shown as black dots, grid cell locations shown as grey dots, Salado *A* sub-basin is outlined. Junín shown as large black triangle.

Daily time series for three weather variables – precipitation, minimum temperature, and maximum temperature – are available for a network of seventeen weather stations located in and around the *A* for the period January 1961 – December 2013 (53 years).

Because four stations have substantial missing data ($> 25\%$), they were removed from the network. The remaining thirteen stations (see Figure 4.1) were used in the analyses. The data were collected and organized by collaborators at the Servicio Meteorológico Nacional (National Meteorological Service) of Argentina.

4.4 Results

4.4.1 Model fit

We employ BayGEN in eight parallel chains (i.e., independent sampling routines) of 1,000 iterations, the first 500 of which are discarded as "burn-in" (i.e., warmup), which results in 4,000 posterior samples of θ. Each chain uses the No-U Turn sampler to generate a sequence of plausible parameter values from the model. When the values from each chain are taken together we are given a full representation of the range of uncertainty in the parameter estimates. In *rstan*, the data are first transformed to the unconstrained scale, and the initialization values are random samples from a Uniform distribution spanning -2 to $+2$. It is advantageous to run multiple chains in that it is easier to diagnose convergence. Numerical convergence is monitored by a weighted combination of the between- and within-sequence variances, as follows:

$$\hat{R} = \sqrt{\frac{\hat{var}^{+}(\theta|y)}{W}} \tag{4.17}$$

where

$$\hat{var}^{+}(\theta|y) = \frac{n-1}{n}W + \frac{1}{n}B. \tag{4.18}$$

In Equations 4.17 and 4.18, θ is a vector of the parameters of interest, y is the response data, W is the within-sequence variance, B is the between-sequence variance, and n is the length of the simulated sequences (after burn-in). $\hat{R}$ is the factor by which the scale of the current distribution for θ might be reduced if the simulations were continued in the limit $n \to \infty$. A value of $\hat{R} < 1.1$ implies convergence (Gelman et al., 2004). Every

parameter in this model has a value $\hat{R} < 1.1$ (most are $\hat{R} = 1$, which is optimal), thus have all converged within the allotted burn-in period. Visual inspection of traceplots (i.e., the evolution of the sampled values of θ) was also used to assess convergence (figures not shown), which confirms what is reported by $\hat{R}$.

4.4.2 Posterior distribution

The model parameter's posterior distribution was compared to its MLE for consistency. Figure 4.2 shows the posterior density for the intercept term of the precipitation occurrence latent Gaussian processes and the MLEs. The MLE are close to the mode of the distributions, except for stations 6 and 9. Furthermore, the distributions are wide, which is indicative of the uncertainty that is not quantified by using the point estimates of MLE. Posterior distributions of other model parameters exhibit similar relationships with their respective MLE, and capture the parameter uncertainty (figures not shown).

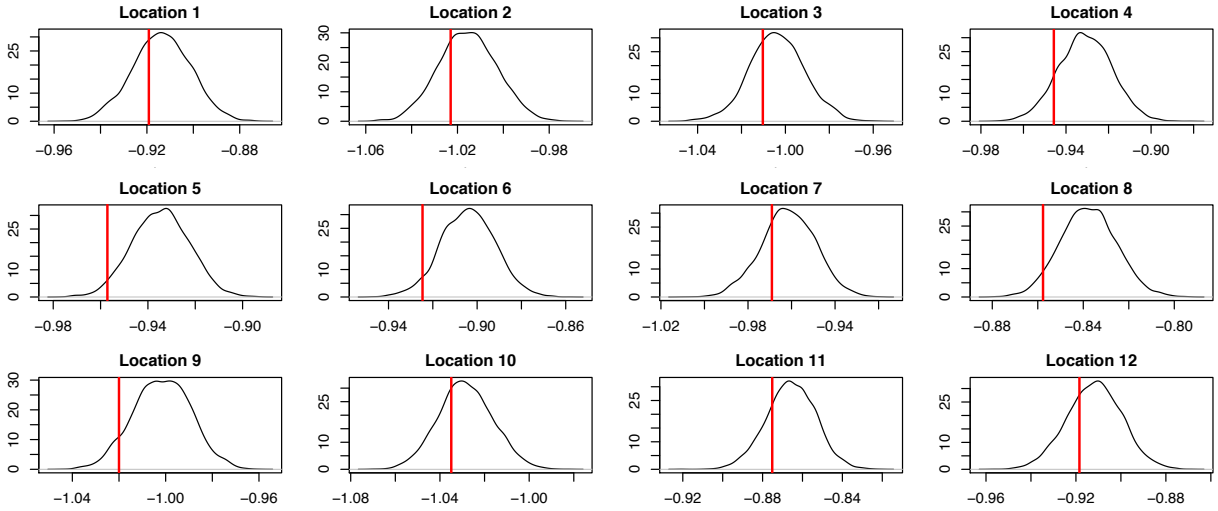

Figure 4.2: Posterior distributions (density plots) and MLE (vertical line) of the intercept term for the precipitation occurrence latent Gaussian process. Note location 13 is not shown, but is consistent with these findings.

4.4.3 Multi-site weather simulation

BayGEN was applied in simulation mode to simultaneously generate daily weather sequences at the network of thirteen stations for the 53-year observational period. As there are 4,000 samples in the posterior distribution, we generated an ensemble of 4,000 daily weather sequences, each using a parameter vector from the posterior distribution. For each parameter sample from the posterior distribution, the daily precipitation occurrence, amounts, maximum and minimum temperatures are generated following the equations and methods described in Section 4.2. A suite of monthly and spatial statistics is computed and compared with the observed; further validation is by comparison to the GLM-based frequentist weather generator (GLMGEN), developed by Verdin et al. (2015b), which is structurally similar and uses the same covariates.

4.4.3.1 Model validation

To validate BayGEN, we simulated 4,000 daily weather sequences of the same length as the historic period, computed a suite of statistics and compared them to their observed values. We also compared the performance with GLMGEN. The following subsections contain comprehensive validations of the BayGEN model.

4.4.3.2 Basic statistics

We first investigated the ability of BayGEN to reproduce the climatological properties of the A by comparing a suite of monthly statistics with the historical statistics, and those from the GLMGEN simulations. Figure 4.3 shows boxplots of daily mean for (a) rainfall, (b) maximum temperature, and (c) minimum temperature at Junín (see Figure 4.1 for the location of this station); (d), (e), and (f) show the corresponding standard deviations for each month. The observed values are shown for comparison. It can be seen that the means and standard deviations are well captured by the simulations as they all

fall within the interquartile range of the boxplots. This is comparable to the performance of GLMGEN, which is shown as white boxplots in Figure 4.3. Considering that BayGEN and GLMGEN are structurally similar, we found comparable performance between them in their ability to capture the basic statistics at all other locations (figures not shown).

We expect the BayGEN simulations to exhibit increased variability, as the parameter uncertainty is propagated to the simulations. This can be seen in the means and standard deviations of precipitation during the summer season of October – March (Figure 4.3a,d). However, minimum and maximum temperatures show similar range of variability, which is likely due to the fact that the temperatures over the A are homogeneous and less variable due to its flat topography.

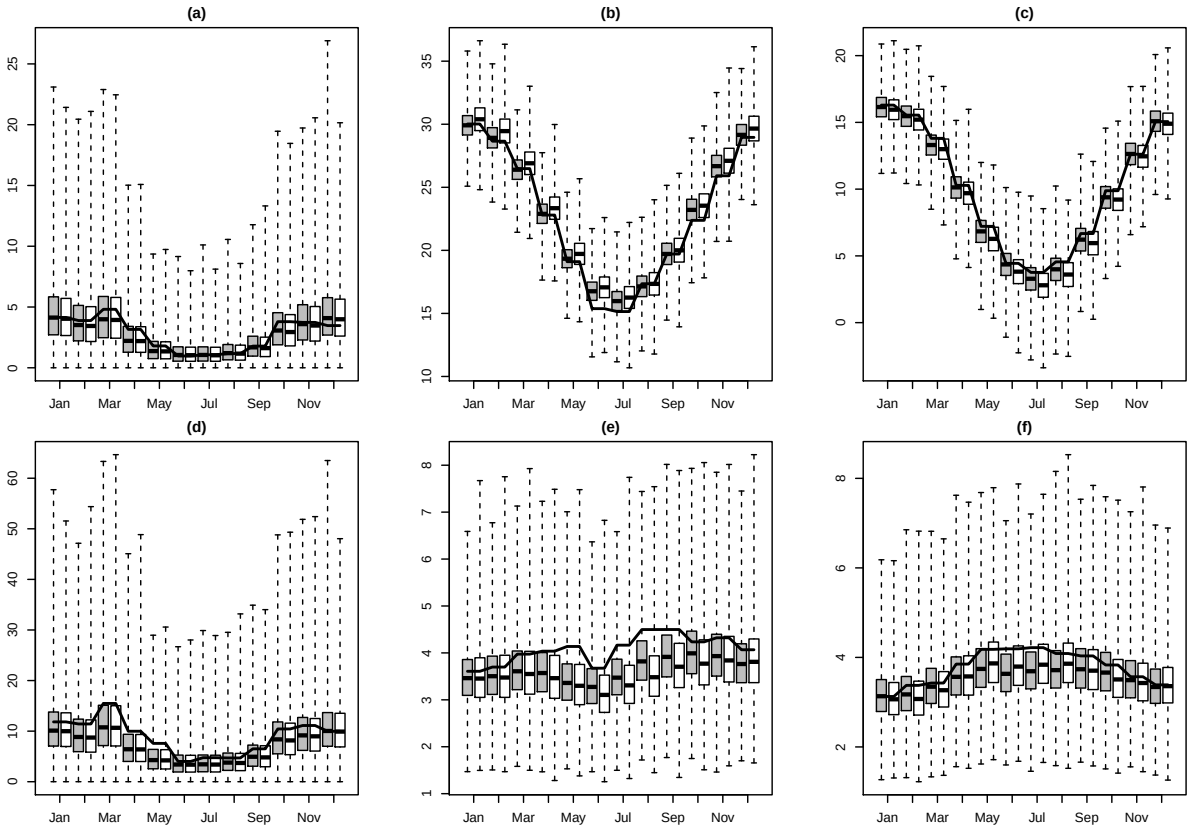

Figure 4.3: 4,000 simulations at Junín: (a-c) Climatological mean of daily precipitation, maximum temperature, and minimum temperature; (d-f) monthly standard deviation of daily precipitation, maximum temperature, and minimum temperature. BayGEN shown as grey boxplots; GLMGEN shown as white boxplots.

4.4.3.3 Extremes

For crop and hydrologic modeling it is important to assess the capability of BayGEN to reproduce precipitation and temperature extremes. Figure 4.4 shows the Q-Q plots, the quantiles of the simulations plotted against those of the observations, of the domain daily extrema of temperature simulated by BayGEN. The simulations are using a random parameter vector from the posterior distribution. The quantiles of the simulated and observed daily extreme temperatures lie on the 1:1 line, except for a slight deviation in the maxima of maximum temperatures (Figure 4.4a) and maxima of minimum temperatures (Figure 4.4d) in the lower tails. All samples from the posterior exhibit similar ability in capturing the domain daily extrema (figures not shown). Considering that we do not explicitly account for temperature extremes in the BayGEN model, the simulations perform well in capturing the spatial extremes. For comparison, we show the Q-Q plots of the domain daily extrema of temperature from GLMGEN in Figure 4.5. There is good agreement between the simulated and observed, although there is a consistent underestimation of the minima of maximum temperatures (Figure 4.5b) and overestimation of the maxima of maximum (Figure 4.5a) and maxima of minimum (Figure 4.5d) temperatures.

For the daily maximum precipitation, the summer season (October – March) is considered, as the majority of crop production (i.e., maize, soybean) occurs during this time. At each station, the simulated daily maximum precipitation for the season are calculated for each year, and the average over all years are calculated for each of the 4,000 simulations, resulting in one value for each simulation. In Figure 4.6, we show boxplots of average daily summer maximum precipitation, as simulated by BayGEN from the posterior samples at each location, with the observed values as a solid line. For comparison we also show the average daily maximum precipitation from GLMGEN. For visual purposes, the locations are ordered from lowest to highest observed daily maximum precipitation (i.e., the x-axis is arbitrary). It can be seen that both the BayGEN and GLMGEN simu-

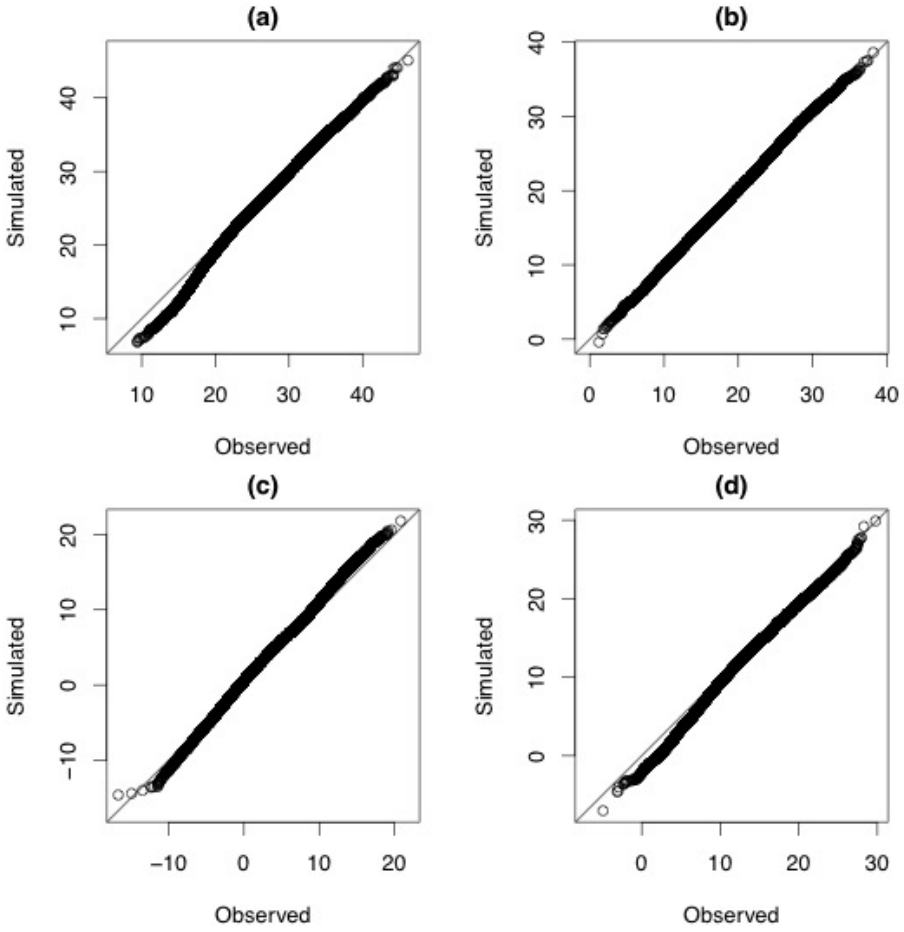

Figure 4.4: Q-Q plots of BayGEN daily domain temperature extrema computed as maximum or minimum of daily values at all locations: (a) domain maximum of maximum temperatures, (b) domain minimum of maximum temperatures, (c) domain minimum of minimum temperatures, and (d) domain maximum of minimum temperatures, units are degrees Celsius.

lations capture the seasonal maximum at each location well, as the observed values are within the interquartile range of the boxplots. However, the BayGEN simulations tend to show a wider range of values, indicating that the parameter uncertainty captured by BayGEN is propagated to the variability in the simulations.

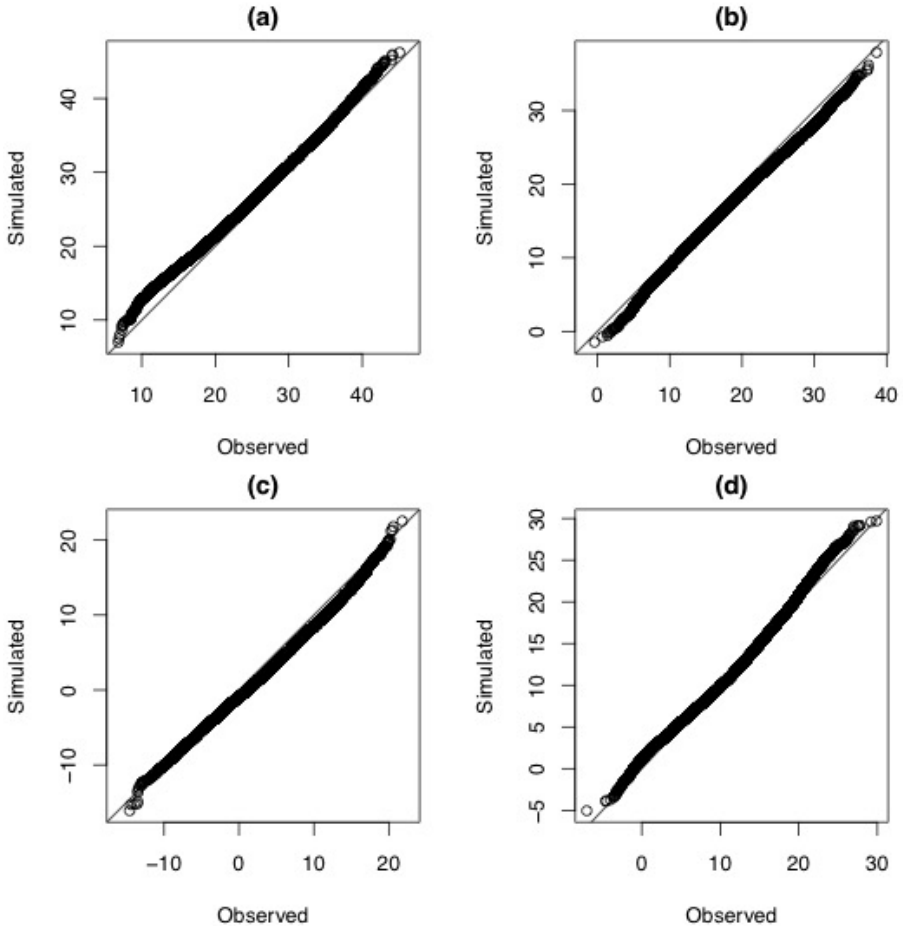

Figure 4.5: Same as Figure 4.4, but from GLMGEN simulations.

4.4.4 Gridded simulation

As mentioned in Section 4.2.1.4, the BayGEN framework enables the modeling of the parameters at any arbitrary location via spatial process models. Thus simulation of weather sequences at any location involves first obtaining the model parameters from the spatial processes, and consequently using them to simulate on a 5km grid. We obtained 4,000 weather generator parameters at each location from the posterior distribution samples. Because gridded sequences are computationally intensive, we randomly sampled 100 parameters from the BayGEN posterior distribution of model parameters to simulate 100 daily weather sequences for the OND season. We aggregated the daily weather to

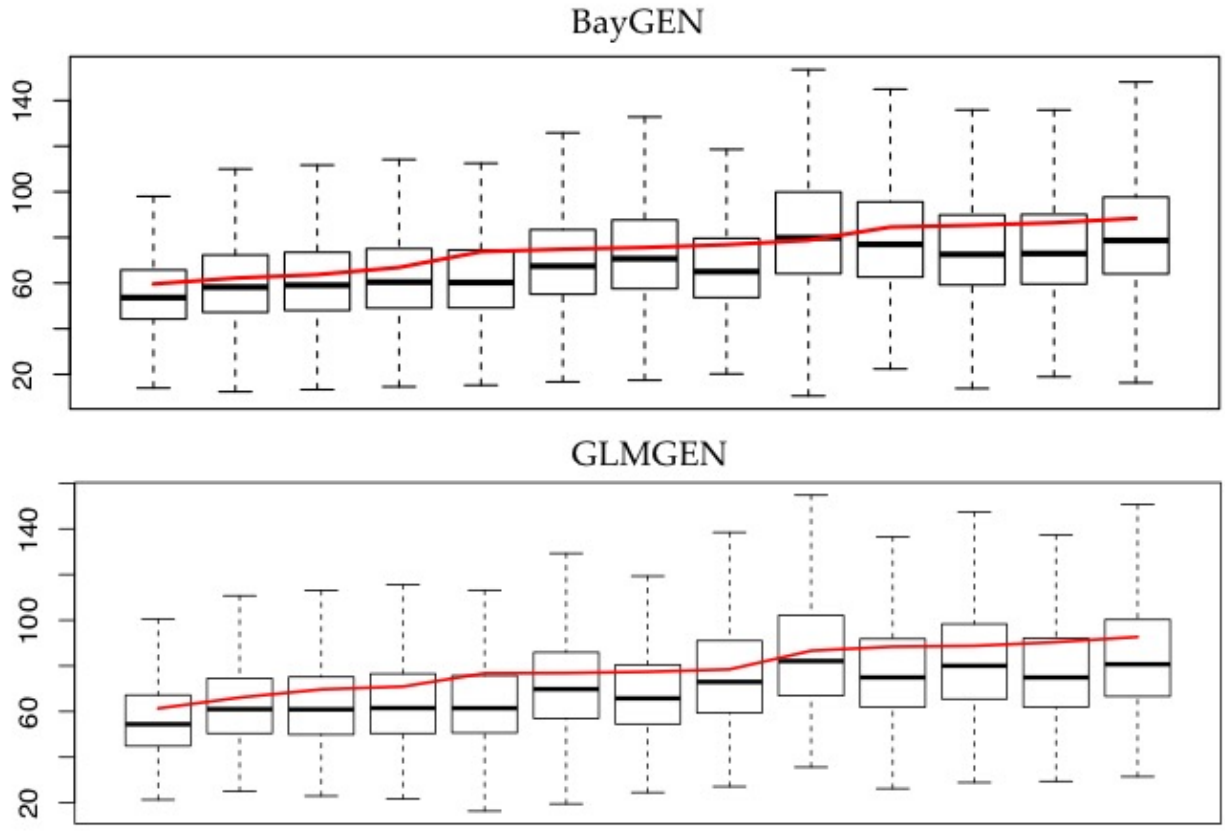

Figure 4.6: Top: Average summer growing season extreme daily precipitation (mm/day) at the stations (boxplots), as simulated by BayGEN, ordered by the average observed extreme daily precipitation (solid line). Bottom: Same as top but from GLMGEN simulations.

the seasonal scale – seasonal total precipitation and seasonal average temperatures. The ensemble mean and the 95% ensemble spread (i.e., the 97.5^{th} percentile minus the 2.5^{th} percentile) are calculated and shown in Figure 4.7. The BayGEN, as expected, exhibits a wide ensemble spread, which illustrates the fact that model parameter uncertainty is effectively propagated to the simulations. We simulated 100 realizations using GLMGEN and show them in Figure 4.8. The ensemble mean of GLMGEN (Figure 4.8a-c) is comparable to that from BayGEN (Figure 4.7a-c). However, the 95% ensemble spread of precipitation from GLMGEN (Figure 4.8d) is smaller than that of BayGEN (Figure 4.7d) – the 95% ensemble spreads for maximum and minimum temperatures are comparable

between BayGEN (Figure 4.7e-f) and GLMGEN (Figure 4.8e-f), which is consistent with the findings from Figure 4.3. Although the residual process provides the majority of variability of the weather sequences, the posterior distribution of model parameters enhances this variability.

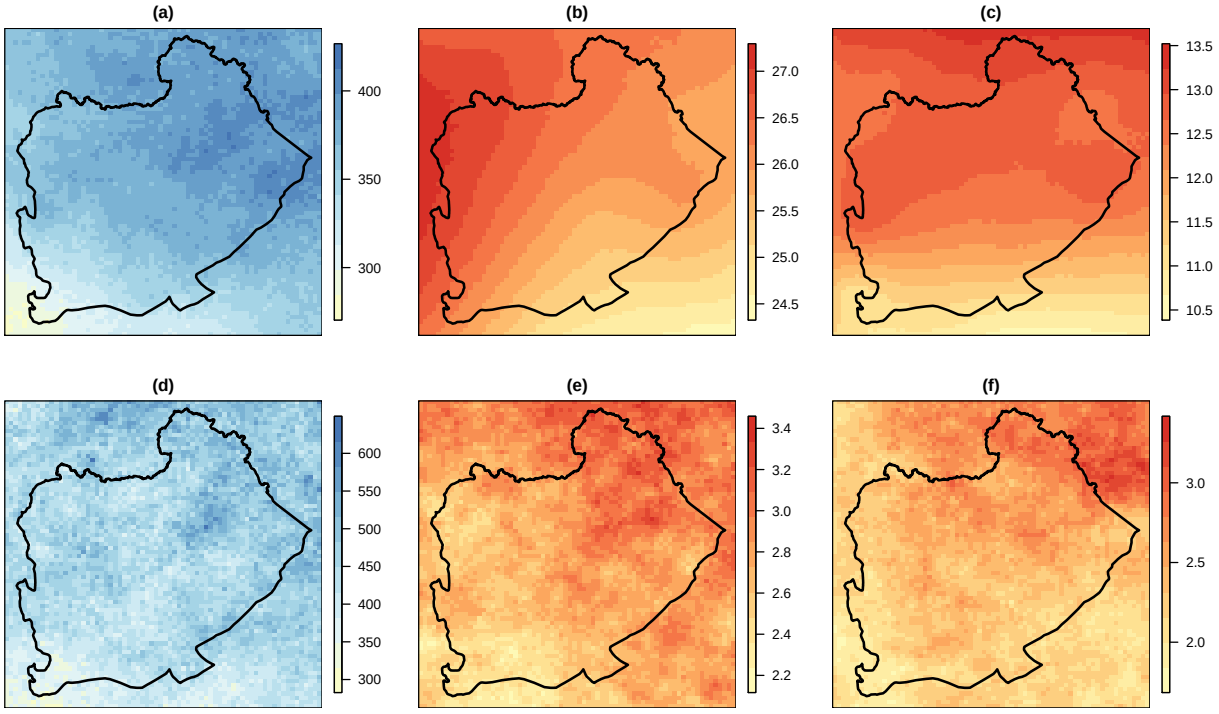

Figure 4.7: BayGEN: (a-c) Ensemble mean precipitation, maximum temperature, and minimum temperature, respectively, for the OND season. (d-f) Corresponding 95% ensemble spread. Units for precipitation are millimeters; units for temperature are degrees Celsius.

4.4.5 Coupling with DSSAT

The Decision Support System for Agrotechnology Transfer (DSSAT; Jones et al. (2003)) is a software package consisting of many crop simulation models. DSSAT can be used to assist in the analysis of complex alternative decisions in agriculture management and adaptation. For a selected crop, the agronomic management (e.g., planting date, fertilization rate, etc.) and land use and soil type (values for many parameters that describe

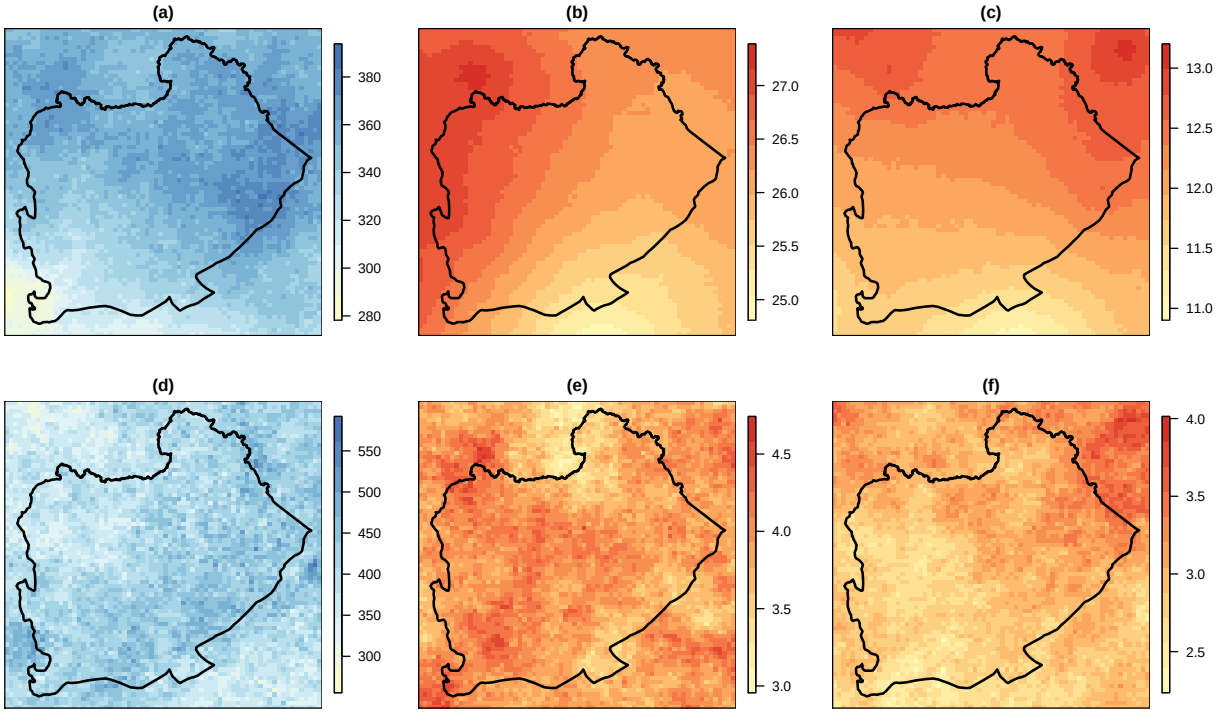

Figure 4.8: Same as Figure 4.7 but from GLMGEN simulations.

the soil of the agricultural plot of interest) are prescribed. Daily weather – minimum and maximum temperatures, precipitation, and solar radiation – are required inputs to DSSAT; optional inputs include dew point and wind speed. Outputs include crop yield (at the seasonal scale), biomass, flowering and harvest date, soil water content for each layer, and root depth, among many others. DSSAT simulation models have been calibrated and tested in the Pampas for soybean, maize, and wheat (Meira et al., 1999; Mercau et al., 2007; Bert et al., 2006, 2007; Ferreyra et al., 2001a; Apipattanavis et al., 2010; Podestá et al., 2009), which are commonly grown crops. Currently, DSSAT simulates crop development and growth for one point (i.e., plot), however, there are at present some tools that facilitate running a gridded version of DSSAT (McNider et al., 2011; Elliott et al., 2014).

BayGEN models precipitation and temperature, but in its current form does not model additional weather variables, (e.g., solar radiation, wind speed, potential evapotranspiration, etc.). However, BayGEN can easily be modified to model any relevant ad-

ditional variables. Here we estimated solar radiation using a modified Bristow-Campbell method (Bristow and Campbell, 1984). Bristow and Campbell (1984) suggest using the mean of minimum temperature for days t and $t+1$ to help reduce the effect of large-scale hot or cold air masses moving through the domain. The modification is that we consider the range in daily temperature extremes (ΔT) to be calculated as simply the maximum temperature minus the minimum temperature for day t. However, the daily temperature data used in this research is aggregated from the hourly scale, thus it is fair to consider only temperatures from day t when calculating ΔT.

We simulated daily weather sequences for two years using 100 randomly selected parameter samples from the posterior distribution. Using a calibrated DSSAT soybean simulation model, 100 soybean yields were computed for the summer growing season at Junín (October – March). Junín is an agriculturally productive region in the northeastern corner of the A, and has long been a test bed for agriculture and climate research in the Argentine Pampas. We also simulated 100 soybean yields using weather simulations from GLMGEN for comparison. Figure 4.9 shows the cumulative density function (CDF) of the ensembles of summer precipitation and soybean yields from BayGEN and GLMGEN simulations. The CDF of seasonal precipitation from BayGEN and GLMGEN are similar, however, BayGEN shows a wider range, as indicated by a more moderate slope. The BayGEN simulations result in lower yield (Figure 4.9b), due to increased variability compared to GLMGEN, which has consistently higher yields. The nonlinear relationship between seasonal weather and soybean yield is quantified by a comparison of the "break-even" production risk, shown as dots on the CDF curves in Figure 4.9b. The risk of not meeting break-even soybean production, as estimated by GLMGEN simulations, is 16%; that from BayGEN simulations is 31%.

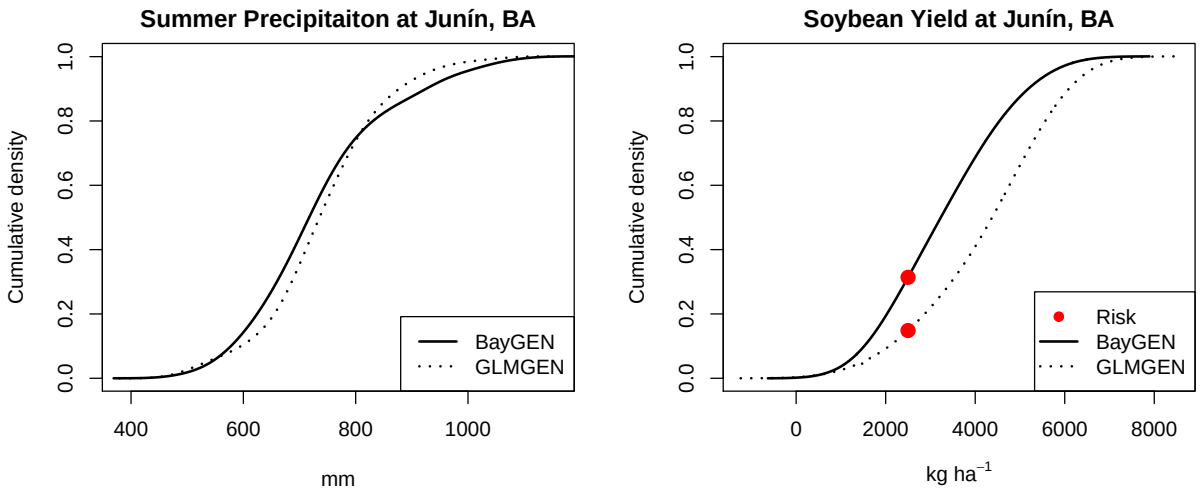

Figure 4.9: Cumulative density functions of summer growing season (October – March) (a) total precipitation, and (b) soybean yields at Junín, based on 100 simulations of daily weather sequences for a two-year period. Simulations from BayGEN are shown as a solid line, and GLMGEN as a dashed line. Break-even production risk is shown as dots on the distributions of soybean yield.

4.5 Summary and discussion

We developed a novel Bayesian hierarchical space-time stochastic weather generator, BayGEN. In the data layer of the hierarchy, precipitation occurrence at each site is modeled using probit regression, the latent process of which is assumed to be a realization from a Gaussian process; precipitation amounts are modeled as Gamma random variables; and minimum and maximum temperatures are modeled as realizations from Gaussian processes. In the process layer, the model parameters of the data layer are assumed to be spatially distributed as Gaussian processes – consequently enabling the simulation of daily weather at arbitrary locations or on a regular grid. Via the posterior distribution of the model parameters, the associated uncertainty is propagated to the weather simulations, which is an important feature that makes BayGEN unique compared to traditional weather generators.

We demonstrated the utility of BayGEN with application to daily weather generation in the Salado *A* sub-basin of the Argentine Pampas. Daily weather ensembles, each 53

years long (the same length as the observational record), were generated for each of the 4,000 posterior parameter samples, thus producing a rich variety of weather sequences that incorporate the parameter uncertainty. The BayGEN daily weather ensembles captured the basic statistics and domain extremes of the observations very well, and is similar in performance to the simulations from a GLM based weather generator, GLMGEN. The BayGEN simulations at each location, and also over the domain in a gridded simulation mode, provided wider ranges of simulations than did the GLMGEN simulations, except when compared to temperature extremes, which is due to the homogeneity of temperatures in the region. The enhanced variability of daily precipitation was also seen to propagate effectively through to crop yield simulations.

The BayGEN, while effective in quantifying parameter uncertainty, comes with a high computational cost, especially when applying to larger domains. Also, larger domains will involve a significant amount of additional parameters. Even in this application, due to the above reasons, spatial correlation of latent residuals in the probit regression for precipitation occurrence is modeled during post-processing using empirical correlation matrices for each month. Given the performance of BayGEN in comparison to GLMGEN, BayGEN offers a robust and attractive alternative to traditional stochastic weather generation. Development of BayGEN also creates a new direction in stochastic weather generation, which has been limited in modeling parameter uncertainty. Extending BayGEN to include seasonal covariates, to simulate daily weather ensembles consistent with seasonal forecasts and multi-decadal projections, is possible with a method similar to that of Verdin et al. (2016). Teleconnections to the climate of a region, such as the El Niño Southern Oscillation, may also be included to enable seasonal forecasting.

Chapter 5

A statistical metamodel for monthly groundwater fluctuations

5.1 Introduction

Climate in the Argentine Pampas is experienced in distinct wet and dry regimes (Minetti et al., 2003). The second half of the 20th century was remarkably wet, considering that during this time the region also experienced one of the most significant increases in annual precipitation in the world (Giorgi, 2002). This substantial increase in precipitation, coupled with advances in technology and a favorable global economy, led to an expansion of agricultural infrastructure to the semi-arid regions of the Pampas, effectively increasing the area of rain-fed agriculture (Podestá et al., 2009). Large scale land use change, such as from perennial pasture to seasonal agriculture, in part contributed to a decrease in annual evapotranspiration (e.g., due to fallow land after harvest) and, consequently, to a rising water table. A shallow water table – and the relatively flat topography of the Pampas – increases the risk of large scale persistent flooding events, and can dramatically reduce crop yields. The capstone of this upward trend in both water table depth (WTD) and precipitation hit during a six-year period from 1997–2003, when a significant portion of the Pampas was flooded, crop yields were significantly reduced, infrastructure was damaged, and soil quality was diminished (Viglizzo et al., 2009).

Empirical relationships between seasonal WTD and relative crop yield have been identified for floodplain regions such as the Pampas (Nosetto et al., 2009). The discovery of this relationship, and subsequently the ideal WTD (i.e. the "sweet spot" as shown

in Figure 5.1), has motivated the need to adapt agricultural management and mitigation practices with a coupled crop and hydrologic model strategy for more complete agricultural risk analyses.

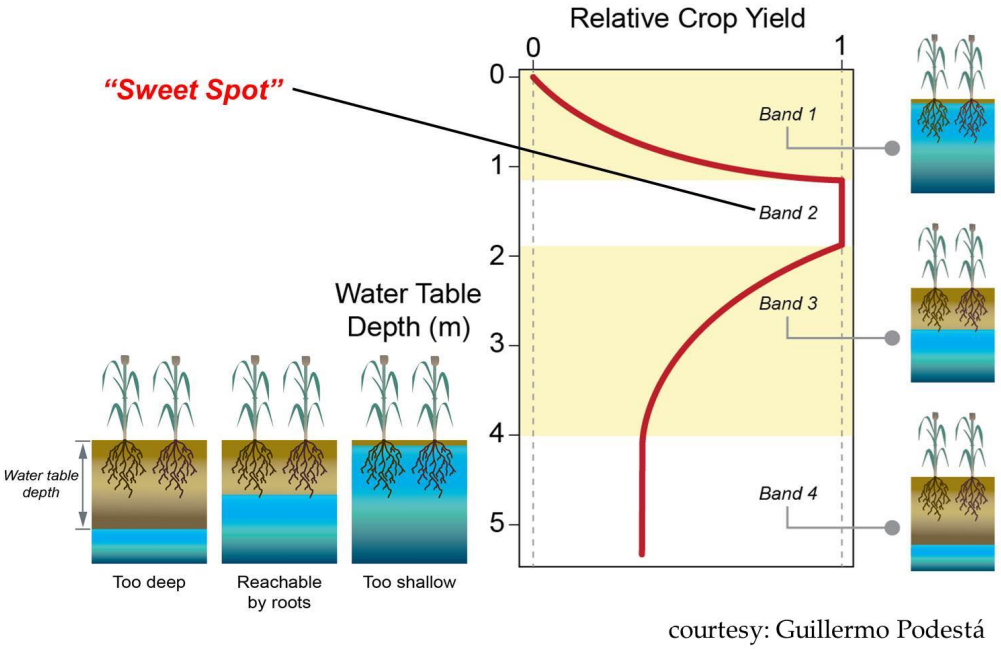

Figure 5.1: Schematic illustrating empirical relationship between WTD and crop yield in floodplain agriculture.

Agricultural risk analyses in the Pampas involve considering a number of agronomic management strategies (e.g., cropping pattern, type of cultivar, etc.) to be run through a crop simulation model, such as those included in DSSAT (Jones et al., 2003), which is driven by ensemble weather scenarios (Bert et al., 2006, 2007, 2014). Recognizing the importance of WTD on crop yields, recent efforts include coupling crop simulation models with hydrologic models, such as MIKE-SHE (Refshaard et al., 1995). Typically, the crop yields estimated from the DSSAT are corrected using empirical relationships between WTD and relative crop yield.

Modeling and simulating spatially consistent WTD at fine resolution to enable cropping decisions at the farm scale is challenging, as it requires gridded inputs of weather and land use data, coupled with a computationally intensive physically based hydrologic

model (i.e., MIKE-SHE). The cause of such computational expense is found in modeling the physical processes such as evapotranspiration, infiltration, and WTD fluctuations at fine spatial and temporal scales. Furthermore, running ensembles to assess the performance of various management strategies under a range of weather scenarios to obtain risk estimates for decision making is prohibitive, which motivates the need for an efficient modeling approach. To this end, we propose the development of a statistical metamodel for monthly fluctuations of WTD, which incorporates the physical mechanisms of WTD variability and provides computational efficiency.

The study region and the data are described in Section 5.2, followed by a brief description of the MIKE-SHE model in Section 5.3. The description of the metamodel is included in Section 5.4, and the results are presented and discussed in Section 5.5. We conclude with a summary and discussion.

5.2 Study region and data

Figure 5.2 shows the outline of the Salado *A* sub-basin, the weather stations, groundwater well locations, and the 5km x 5km resolution grid cells at which the MIKE-SHE model is run. Observed daily weather at the network of 17 weather stations is available for the historic period (1 January 1961–31 December 2013). The daily weather was interpolated onto a 5km x 5km grid using the space-time weather generator of Verdin et al. (2015b), which is referred to as the "pseudo-historic" sequence. A MIKE-SHE model, calibrated for the Salado *A* sub-basin, was run using the pseudo-historic weather sequence, along with historic land uses, soil type data, and limited groundwater well observations. The MIKE-SHE model produced daily actual evapotranspiration (ET) and WTD on the 5km x 5km grid, which we saved to use as response variables in the metamodel.

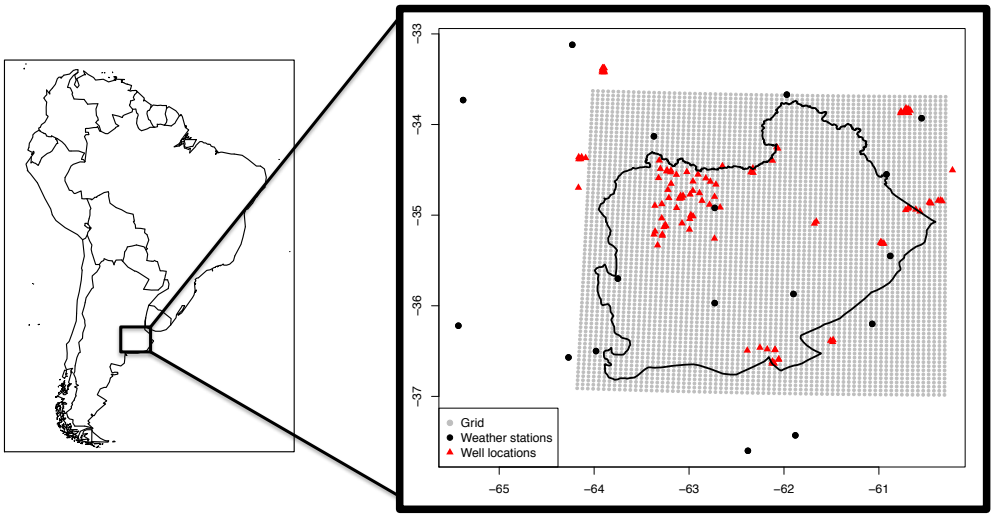

Figure 5.2: Study region: weather stations shown as black dots; MIKE-SHE grid shown as grey dots; well locations shown as red triangle. Salado *A* sub-basin shown as outline.

5.3 MIKE-SHE

Although this research does not focus on developing a MIKE-SHE model, we provide a brief description for completeness. MIKE-SHE is a spatially-distributed physically based hydrologic model that couples the surface and groundwater flow. The model divides the region of interest into equally sized grid cells, which are characterized by land use data; it is possible to include fractional land use data. At each grid point, for each time step (which is dynamic, but typically sub-daily), energy and mass balance equations are solved separately, which provide estimates of actual ET, infiltration, surface runoff, WTD, and other relevant variables.

Required inputs at each grid point include daily precipitation and potential evapotranspiration, soil type data, and land use data (e.g., leaf-area index, root depth). River networks may also be incorporated. Outputs are WTD, actual ET, overland flow, unsaturated zone flow, water content in the unsaturated zone, and groundwater flow, among others. Actual ET is modeled using the methodology of Kristensen and Jensen (1975); vadose zone flow is modeled using Richards (1931) equation. MIKE-SHE is widely used

around the world for hydrologic and hydraulic applications (Thompson et al., 2004; Singh et al., 1999; Liu et al., 2007; Graham and Butts, 2005; Im et al., 2009; Doummar et al., 2012; Zhang et al., 2015; Sandu and Virsta, 2015; Ma et al., 2016; Moussoulis et al., 2016; Larsen et al., 2016), and has been previously calibrated and validated for the Pampas, and specifically for the *A* sub-basin (García et al., 2014; Laxagüe et al., 2014; García et al., 2016). MIKE-SHE has also been used to study a proposed channelization plan (UTN-FRA, 2007) for the Salado Basin (Menéndez and Badano, 2010; Badano et al., 2008; Badano, 2010; Re et al., 2008).

It is acknowledged that MIKE-SHE is similar to other distributed hydrologic models such VIC (Liang et al., 1994). It is assumed that the framework of this metamodel may be applied to hydrologic models other than MIKE-SHE. We use the MIKE-SHE because it has been previously developed and calibrated for our study region.

5.4 The metamodel

A hierarchical metamodel using traditional linear models and spatial process models is proposed. In the first level of hierarchy, we model the monthly actual ET as a function of monthly precipitation, maximum and minimum temperatures, potential evapotranspiration, and land use. In the second level, monthly WTD is modeled as a function of the previous month's WTD, the current month's precipitation, maximum and minimum temperatures, and the estimated actual ET from the first level. We fit both models at representative "knot" locations, which can be thought of as a network of observations. The regression coefficients for both models are estimated at the spatial resolution of the MIKE-SHE using spatial process models, which enables the estimation of actual ET and the simulation of WTD for the entire *A* sub-basin. Land use is available at the county level (as percentages spanning [0,1]), and are constant for each year. The model formulation is described below.

5.4.1 Hierarchy

The response variables of interest are monthly evapotranspiration and water table depth, which we define as $Y_{ET}(\boldsymbol{s}, t)$ and $Y_{WT}(\boldsymbol{s}, t)$ for location $\boldsymbol{s}$ and time t. The hierarchical metamodel is defined as follows:

$$Y_{ET}(\boldsymbol{s}, t) = \mathbf{X}_{ET}(\boldsymbol{s}, t)' \boldsymbol{\beta}_{ET}(\boldsymbol{s}) \tag{5.1}$$

$$Y_{WT}(\boldsymbol{s}, t) = \mathbf{X}_{WT}(\boldsymbol{s}, t)' \boldsymbol{\beta}_{WT}(\boldsymbol{s}) + \epsilon(\boldsymbol{s}, t) \tag{5.2}$$

$$\epsilon(\boldsymbol{s}, t) \sim GP(0, C(t)) \tag{5.3}$$

$$s = 1, 2, ..., S \tag{5.4}$$

$$t = 1, 2, ..., T, \tag{5.5}$$

where S represents the number of knot locations, and T represents the number of months in the record. We neglect the error term for ET because the estimates of $Y_{ET}(\boldsymbol{s}, t)$ are used as covariates to model WTD; effectively, we are combining the error into one term. The error $\epsilon(\boldsymbol{s}, t)$ is modeled as a stochastic Gaussian process with monthly covariance functions $C(t)$ to account for seasonality.

5.4.2 Covariates

Monthly average precipitation, maximum temperature, minimum temperature, and potential evapotranspiration are denoted as $X_P(\boldsymbol{s}, t)$, $X_X(\boldsymbol{s}, t)$, $X_N(\boldsymbol{s}, t)$, and $X_E(\boldsymbol{s}, t)$, respectively. There are six land use types – pasture, soybean, wheat, soybean/winter wheat, sunflower, and maize, which are denoted as $X_{L1}(\boldsymbol{s}, t)$, $X_{L2}(\boldsymbol{s}, t)$, $X_{L3}(\boldsymbol{s}, t)$, $X_{L4}(\boldsymbol{s}, t)$, $X_{L5}(\boldsymbol{s}, t)$, and $X_{L6}(\boldsymbol{s}, t)$, respectively. These are parameterized as fractional area of a county, with range [0,1], and are constant for each calendar year. The covariate vectors are defined as follows:

$$\mathbf{X}_{ET}(\boldsymbol{s},t) = (X_P(\boldsymbol{s},t), X_X(\boldsymbol{s},t), X_N(\boldsymbol{s},t), X_E(\boldsymbol{s},t), X_{L1}(\boldsymbol{s},t), X_{L2}(\boldsymbol{s},t), X_{L3}(\boldsymbol{s},t), X_{L4}(\boldsymbol{s},t), X_{L5}(\boldsymbol{s},t), X_{L6}(\boldsymbol{s},t)) \tag{5.6}$$

$$\mathbf{X}_{WT}(\boldsymbol{s},t) = (Y_{WT}(\boldsymbol{s},t-1), X_P(\boldsymbol{s},t), X_X(\boldsymbol{s},t), X_N(\boldsymbol{s},t), Y_{ET}(\boldsymbol{s},t)) \tag{5.7}$$

5.4.3 Simulation

In simulation, the gridded coefficients for the first level of hierarchy are used with the weather and land use data to estimate the actual ET. Given a properly defined initial water table depth, simulation of Y_{WT} can proceed using gridded coefficients from the second level of hierarchy. Due to the stochasticity involved in the local regression residuals, mean zero Gaussian processes are simulated using the monthly covariance functions. An ensemble of water table depths will provide a means for quantifying risk of saturation or depletion, which can be used to make more informed agricultural management decisions.

To summarize the simulation process, the following steps are used:

(i) Regression coefficients are estimated at the desired spatial resolution using spatial process models.

(ii) Daily weather is obtained for the period of interest (e.g., simulations from a stochastic weather generator) and is aggregated to monthly scale.

(iii) With a defined land use pattern and the monthly weather, actual ET is estimated from the gridded coefficients from the first level of hierarchy.

(iv) Gridded coefficients from the second level of hierarchy are used with the estimated actual ET and monthly weather to simulate WTD.

(v) Steps (i)-(iv) provide the mean estimate. To account for local noise, mean zero Gaussian processes are simulated using monthly covariance functions (Equation 5.3), and added to the mean estimate to obtain an ensemble of water table depths.

5.5 Results

5.5.1 Simulation of historic period

We evaluate the model by simulating an ensemble of WTD on the MIKE-SHE grid for the historic period using the gridded coefficients, pseudo-historic weather, and land use data as forcings. An ensemble of 100 WTD trajectories is simulated for the period January 1961–December 2013. We first analyze the ability of the metamodel to reproduce the temporal variability of WTD in Figure 5.3. The dominant mechanisms of WTD variability as simulated by MIKE-SHE (solid red line) are well reproduced by the metamodel ensemble, which is shown as a time series of boxplots.

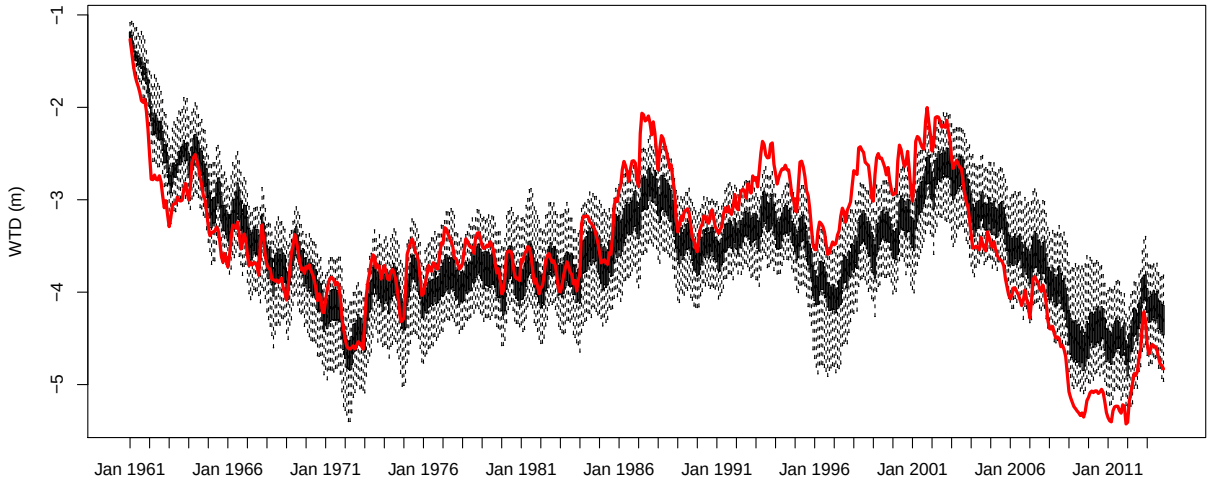

Figure 5.3: Basin average WTD for metamodel ensembles (boxplots, full range of ensemble) and MIKE-SHE (solid line) for the historic period (January 1961–December 2013).

Spatial maps of summary statistics (i.e., root mean square error [RMSE] and percent bias [% Bias]) between the metamodel ensemble mean and the MIKE-SHE for each grid cell are shown in Figure 5.4. For most of the sub-basin the errors are quite small except for a region in the northwestern corner where the disagreement between the metamodel and

MIKE-SHE are relatively large, as indicated by RMSE. We hypothesize that this is likely caused by the assumption of homogeneous soil type. However, the relative lack of bias in the region implies the model adequately represents the WTD fluctuations for the historic period.

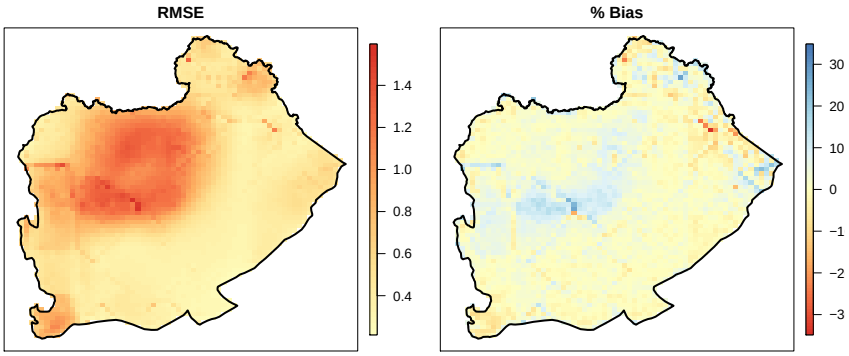

Figure 5.4: RMSE (meters) and % Bias between MIKE-SHE and the metamodel ensemble mean water table levels for the historic period (January 1961–December 2013).

5.5.2 Performance during wet and dry months

Times of known saturation (i.e., flooding) and depletion (i.e., drought) were selected as representative wet and dry months. We select December 2000 as the wet month, as this was a time when a significant portion of the Pampas landscape was flooded (Viglizzo et al., 2009); February 2012 is selected as the dry month, as the northwestern sub-basin is at its lowest level for the historic period.

Figure 5.5 shows the WTD as simulated by MIKE-SHE, the metamodel ensemble mean, and their difference; the top row is for December 2000, and the bottom row is for February 2012. It can be seen that the metamodel estimates show good agreement with the MIKE-SHE simulation. However, the estimates in the northwestern part of the sub-basin show increased error, which is consistent with the findings in Figure 5.4. Overall, the metamodel captures the dominant mechanisms of WTD variability in the sub-basin well. The differences in the northwestern region can be considered negligible given the

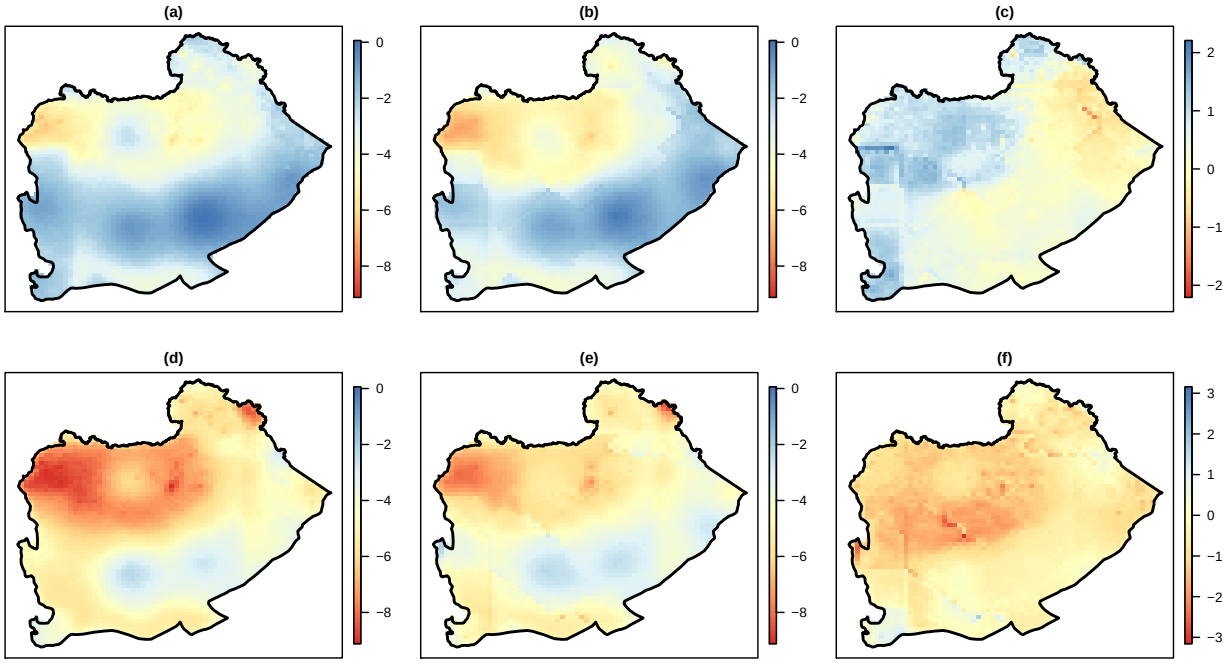

Figure 5.5: (a-c) Dec 2000: MIKE-SHE, metamodel ensemble mean, and difference; (d-f) Feb 2012: MIKE-SHE, metamodel ensemble mean, and difference. Difference calculated as MIKE-SHE minus metamodel ensemble mean. Units are meters.

magnitude of depletion seen for this month, which is more than 9 meters below the surface at its lowest.

5.5.3 Use of the metamodel in a rudimentary decision mode

It is of interest to assess the capability of the metamodel for use in decision mode, such as for seasonal planning. In this, an ensemble of WTD is simulated based on seasonal climate forecasts and various cropping patterns. First, probabilistic seasonal climate forecasts are downscaled to produce ensembles of daily weather at the spatial resolution of the metamodel using the stochastic weather generator approach of Verdin et al. (2016). The resulting daily weather ensemble is aggregated to the monthly scale and is used to drive the metamodel, which produces an ensemble of WTD. We demonstrate this for the seasonal climate forecast for October – December (OND) 2013, issued

in September 2013 by the International Research Institute for Climate and Society (IRI, www.iri.columbia.edu). The forecasted probabilities (in A:N:B format) for this season were reported as 35:40:25 for precipitation and 20:35:45 for temperature – for more information on the A:N:B format of these forecasts, see Verdin et al. (2016) and the IRI website.

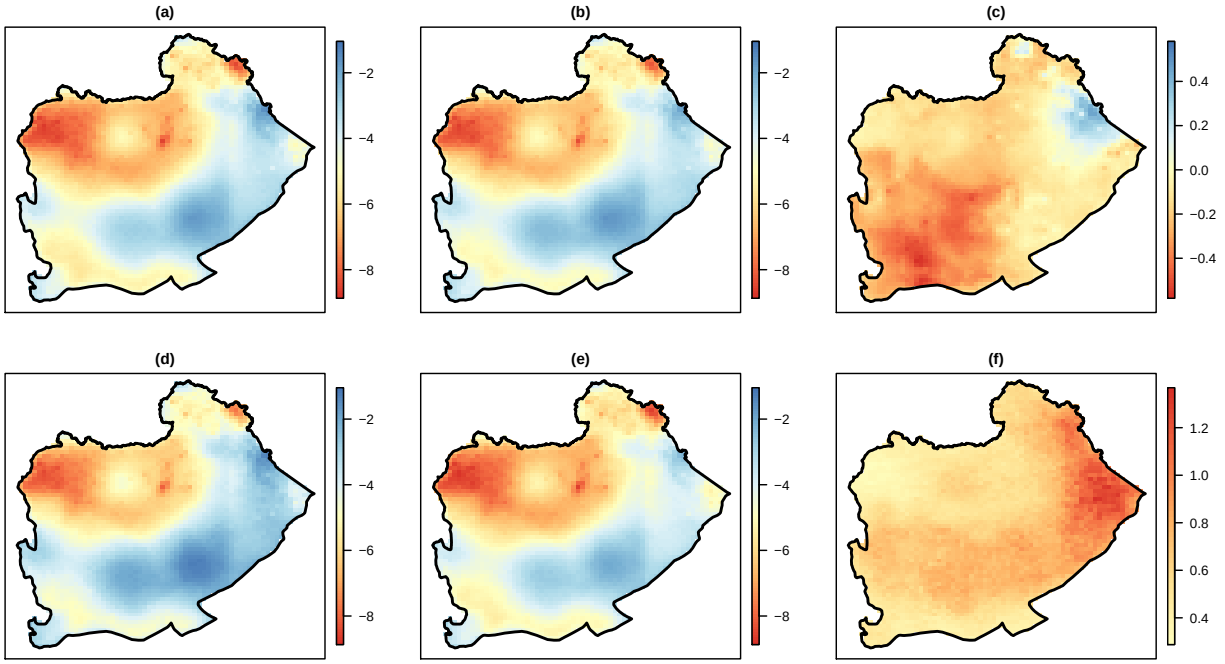

Figure 5.6: Dec 2013: (a-c) Water table depth as simulated by MIKE-SHE, metamodel ensemble mean, and difference between MIKE-SHE and metamodel ensemble mean; (d-f) 97.5% percentile of metamodel ensemble, 2.5% percentile of metamodel ensemble, and 95% ensemble spread, respectively. Units are meters.

For December 2013, Figure 5.6 shows the WTD as simulated by MIKE-SHE, the metamodel ensemble mean, and the difference (MIKE-SHE minus ensemble mean) in the top row; the bottom row shows the 97.5th percentile of the metamodel ensemble, the 2.5th percentile of the metamodel ensemble, and the 95% ensemble spread (i.e., 97.5th percentile minus 2.5th percentile). The ensemble mean of the metamodel closely resembles the estimates from the MIKE-SHE, and the difference between them is very small, with errors no greater than 0.5 meters. The 95% ensemble spread provides the uncertainty in WTD for the upcoming season, which can be useful for quantifying the risk of flooding

or depletion.

5.6 Summary and discussion

We have developed, validated, and illustrated the use of a statistical metamodel for monthly water table depth (WTD) in the Salado *A* sub-basin of the Argentine Pampas. The metamodel was created as a complementary tool to the daily MIKE-SHE hydrologic model, which was previously calibrated and validated for this region by colleagues in Buenos Aires. The MIKE-SHE model was calibrated on gridded historic (i.e., pseudo-historic) daily weather for the period 1 January 1961–31 December 2013. Estimates of actual evapotranspiration (ET) and WTD from the pseudo-historic simulation were saved and aggregated to monthly scale to be used as response variables in the metamodel.

In the first level of hierarchy, local regressions of ET as a function of climate and land use are fitted at equally-spaced knot locations to reduce computational intensity – akin to a well-distributed network of observation locations. The regression coefficients at the knot locations are estimated on the MIKE-SHE grid using spatial process models to enable prediction of ET given future climate and land use. The second level of hierarchy fits local regressions of WTD as a function of climate and the estimated ET from the first level; similarly, the coefficients are obtained at the knot locations and subsequently estimated on the MIKE-SHE grid using spatial process models. The error for the ET model is lumped with that for the WTD model, because estimated ET is used as a covariate in the WTD model. The residuals from the WTD model are used to estimate spatial process parameters (i.e., marginal variance and effective range; nugget is fixed at zero) for twelve distinct mean zero Gaussian processes corresponding to each calendar month. Due to the stochastic properties of Gaussian random fields, we simulate an ensemble of WTD in order to obtain the full distribution of future WTD.

Validation of the metamodel was by direct comparison to the MIKE-SHE model. For

the historic period, much of the sub-basin exhibited adequate summary statistics, with RMSE being less than 0.5 meters and % Bias near 0%. Additional validation was carried out by isolating times of saturation and of depletion (i.e., December 2000 and February 2012, respectively), and ensuring that the metamodel adequately replicated the spatial properties of the MIKE-SHE model (i.e., saturation or depletion in the same parts of the sub-basin). For both wet and dry times, it was shown that the metamodel captured the dominant mechanisms of spatial and temporal variability during these high-risk times. We then illustrated the use of the metamodel as a rudimentary decision support tool, wherein a stochastic weather generator (Verdin et al., 2016) was used to simulate a gridded ensemble of future daily weather trajectories, conditioned on the probabilistic IRI seasonal forecast for OND 2013. This gridded ensemble is aggregated to monthly and used as the weather input data to the metamodel, which results in a distribution of future WTD corresponding to the seasonal forecast.

One shortcoming of this methodology is the inherent assumption of homogeneous soil type throughout the study region. It was seen that there is increased error and bias in the northwestern region of the A sub-basin, which suggests that this region experienced WTD depletion rate that was much higher than the rest of the sub-basin, and that this rate of depletion is likely due to differences in soil type and porosity. It follows that explicitly modeling soil type and porosity would be advantageous, and would add flexibility to the metamodel. Another drawback to the model is that we do not consider leaf-area index (LAI) or root depth (RD) in the land use data, nor do we consider the annual cycle that different land use will have on the estimates of ET. To address this problem, we suggest replacing the six land use covariates with two covariates – weighted averages of LAI and RD. There are empirical data upon which we can derive the annual cycle of biomass generation for each land use type. Given the proportion of land use for a given county, we can calculate the weighted average, which would better reflect the seasonal impact of agriculture on evapotranspiration.

Chapter 6

Conclusion

6.1 Summary

This dissertation has presented a suite of statistical models for enhanced agricultural decision support. Included in this suite are two stochastic weather generators (SWGs) and one statistical-hydrologic metamodel. The SWGs were designed to simulate weather at arbitrary locations, such as on a regular grid, which is necessary for distributed hydrologic modeling. Some hydrologic models may come with the limitation of being computationally expensive and having hard limits to the number of processes that can be run simultaneously. Motivated by this shortcoming, the metamodel is a hierarchical space-time model for estimating actual evapotranspiration (ET) and, consequently, water table depth (WTD) on the monthly scale. The response variables are ET and WTD, as estimated by a calibrated MIKE-SHE model for the Salado A sub-basin (hereafter the A) – the same study region as for the SWGs . While the tools presented in this research have been calibrated and validated on the sub-basin scale, it is assumed that they are portable to any reasonably sized region. When considering application of these tools in a new study region, careful attention must be paid to the spatial and temporal characteristics of weather and climate. Model selection should be carried out such that the covariates are relevant and statistically significant.

The second chapter of this dissertation provides the framework for a space-time stochastic weather generator. The use of generalized linear models (GLMs) simplifies the

effort in modeling non-normal variables through a suite of link functions. Spatial process models are used on the regression coefficients and their residuals to enable simulation at unobserved locations, such as on a regular grid. Precipitation occurrence and amounts are modeled separately: occurrence is modeled using probit regression, and amounts are modeled as Gamma random variables. The benefit of probit regression is the ability to use linear regression principles to model and simulate binary data through an indicator function: if the latent Gaussian process is greater than or equal to zero, this indicates occurrence, and vice versa. These latent Gaussian processes are autoregressive time series models, which are fitted at each observation location; the maximum likelihood estimates (MLE) of the regression coefficients and the residuals are saved for simulation. Minimum and maximum temperatures are modeled using traditional linear regression. At each location, separate autoregressive time series models are fitted to minimum and maximum temperatures; their regression coefficients are saved for simulation. In simulation, the coefficients from the precipitation occurrence, maximum temperature, and minimum temperature models, along with their respective covariates, are used to reproduce the mean functions; deviations from the mean functions are driven by random samples from mean zero multivariate normal distributions with covariance functions defined by the spatial correlation and covariances of the precipitation occurrence and temperature model residuals, respectively. The multivariate normal random samples which drive precipitation occurrence are transformed using a copula function to generate spatially correlated precipitation amounts. For gridded simulation, the spatial process models are used to estimate the regression coefficients at the desired spatial resolution to produce the mean function; residuals are simulated via Gaussian random fields with spatial process parameters defined from the model residuals.

Application of the stochastic weather generator within a nonstationary context, such as conditional simulation on seasonal to multi-decadal scales, is presented in the third chapter. In this, a principal component analysis was carried out on seasonal total pre-

cipitation (ST), mean maximum temperature (SMX), and mean minimum temperature (SMN) at each location, to identify the dominant source of variability within these variables. It was confirmed that the first principal component for each climate variable is the domain-average of that climatic process. Furthermore, the first principal components for ST, SMX, and SMN, explain 47%, 77%, and 71% of the total variances, respectively. Using this knowledge as motivation, these additional covariates (ST, SMX, SMN) were included in the weather generator's model fit, and were all statistically significant at the 99% level. It was shown that the inclusion of these additional covariates was effective in conditioning the output of the weather generator, reproducing the spatial and temporal characteristics at the coarse scales such as seasonal and regional, and fine scales, such as daily and local. Translation of probabilistic seasonal forecasts, such as those provided by IRI, involves classifying observed ST, SMX, and SMN as above-, near-, or below-normal, the thresholds of which are defined by empirical terciles. Using the seasonal forecast probabilities as weights, ensembles of ST, SMX, and SMN are bootstrapped (i.e., sampled with replacement); these ensembles are then used to drive the weather generator, thus translating the relatively uninformative probabilistic forecasts to ensembles of weather trajectories. It was further illustrated that the weather generator can be effective in downscaling regional climate model output, in this case the CORDEX-CMIP5 regional climate model for near-term projection (2015-2050). The grid cells (each at 0.44x0.44° resolution) surrounding the A were used to calculate values of ST, SMX, and SMN for the projection period, which were then used to drive the weather generator, which was skillful in capturing the positive trend in temperature and the epochal traits of precipitation. It was acknowledged that any number of covariates may be included in the weather generator – including teleconnections such as ENSO or other relevant climate drivers – which could further improve the skill on seasonal to multi-decadal scales.

The fourth chapter of this dissertation presents BayGEN, a Bayesian space-time stochastic weather generator. BayGEN model structure is based on the GLM weather

generator from the second chapter, with the added benefit of a Bayesian framework. The main benefit of BayGEN is the quantification and subsequent propagation of uncertainty, which is commonly ignored in traditional parametric weather generators. Another advantage to using a Bayesian framework is the ability to spatially correlate regression coefficients via spatial process parameters (i.e., nugget, sill, range), which enables weather generation at arbitrary (e.g., gridded) locations. The benefit of BayGEN was illustrated through direct comparison to the weather generator of the second chapter (hereafter GLMGEN). Ensembles of daily weather trajectories were simulated by BayGEN and GLMGEN, which were then used to drive the DSSAT crop simulation model for an arbitrary growing season at a location in the northeastern corner of the *A*. It was shown that the uncertainty is preserved by BayGEN, which propagates not only to daily weather sequences, but also to simulated crop yield quantities. The enhanced uncertainty reflected in BayGEN ensemble scenarios provide the decision maker with a better estimate of risk for seasonal crop planning.

A space-time metamodel to predict ET and, consequently, WTD in the Salado *A* sub-basin is detailed in the fifth chapter of this dissertation. The objective of this tool is to mimic the MIKE-SHE hydrologic model, which is routinely used by colleagues in Buenos Aires, Argentina. Although MIKE-SHE is run on the daily scale, the metamodel was developed at monthly resolution. Therefore, the metamodel is not promoted as a replacement for the MIKE-SHE; rather, it is a complementary tool for near real-time decision making. Ensembles of future weather trajectories, as produced by GLMGEN or BayGEN, can be used to drive this model for the purpose of identifying regions of high risk (e.g., fully saturated soils, very deep water table, etc.). The framework is stochastic, in the sense that randomness is needed to effectively reproduce the magnitude and direction of WTD fluctuations. The metamodel was validated by direct comparison to the MIKE-SHE. For the historic period (January 1961–December 2013), the metamodel was shown to capture the basin average WTD well. Summary statistics show good agreement be-

tween the metamodel ensemble mean and the MIKE-SHE for most of the A, with a region of increased error in the northwestern sub-basin. Further validation involved identification of wet and dry months, and comparing the snapshot of MIKE-SHE to the metamodel ensemble mean. The spatial distribution of risk (i.e., saturation or depletion) for these months was adequately reproduced by the metamodel. Additionally, we illustrated the use of the metamodel as a rudimentary decision tool for seasonal planning. The seasonal forecast downscaling method from chapter three was used to generate an ensemble of gridded daily weather for the IRI forecast for OND 2013. This gridded weather was aggregated to monthly and used as weather input data to the metamodel. The resulting distribution of WTD reflects the uncertainty of the seasonal forecast, and consequently identify high-risk regions.

6.2 Discussion

Overall, GLMGEN requires very little computing power, and is unique in its ability to simulate daily weather at arbitrary locations. Furthermore, its flexible framework enables the simulation of daily weather that exhibits desired traits or trends. Its near real-time capability makes it useful for both climate variability and land use change studies.

BayGEN was developed to address the fact that GLMGEN ignores the cascade of uncertainty from parameter space to decision space. Unfortunately, even with modern computing power, a minimum of one week (i.e., seven days) is required to fit the BayGEN model on 53 years of data for 13 stations. Furthermore, the inclusion of missing data can dramatically increase this computing time. Therefore, application of BayGEN is ideal for situations where time is not an issue.

The metamodel for monthly WTD and ET is run on the monthly scale, thus inherently will have a much faster run time than the daily MIKE-SHE hydrologic model. Additionally, a physically based model can be far too cumbersome for agricultural planning

purposes where WTD is the only variable of interest. Therefore, the metamodel is best used as a complementary tool for agricultural planning on the seasonal to multi-decadal scales, where the user may wish to run the metamodel for hundreds or thousands of alternate land use or climatic scenarios and analyze the WTD response.

This research has been motivated by the need to better understand the impacts of climate variability and land use change on agricultural production, especially in the semi-arid Pampas. The suite of tools presented in this dissertation can provide agronomists, farmers, and decision makers with powerful statistical tools for informed decision making and risk analysis for alternative adaptation strategies.

6.3 Future work

Several extensions and enhancements of this research are possible. A few are proposed below.

Foremost is the development of GLMGEN as a contributed package for the `R` statistical library on the Comprehensive R Archive Network (CRAN) for easy installation. The generalization of GLMGEN to an `R` package would enable easier and wider use of the model by researchers and resource managers with limited statistical knowledge or coding experience. As a contributed package to the CRAN, the framework of GLMGEN will be defined to enable both conditional simulation (i.e., Chapter 3) and unconditional simulation (i.e., Chapter 2). Since this will be developed for `R`, which is an open source project, other researchers can improve upon the code as well.

Conditional weather simulation (i.e., Chapter 3) on the seasonal scale is based on categorical probabilistic forecasts. However, large scale climate indices from the ocean-atmospheric system can be identified as indicative of the future state of seasonal precipitation and temperature. Therefore, it would be a significant improvement to include such large scale climate indices as covariates in the GLMGEN model, which would help to

exploit climate teleconnections that might not be captured in the categorical forecasts.

The BayGEN model can be improved significantly, by incorporating seasonal climate covariates, and modifications to explicitly capture the extremes of all weather variables and the spatial covariance of precipitation occurrence. Efforts to improve the computational efficiency should also be considered, as the model takes numerous days to fit.

Covariates for land use within the metamodel in Chapter 5 can be improved. Currently, there are six separate land use covariates for each land use type, and are defined as a proportion (i.e., with range [0,1]) of land allocated for each type (e.g., pasture, soybean, maize, etc.). These land use proportions are reported as constants for a calendar year, therefore the covariates remain fixed for each year (i.e., there is no annual cycle). The main limitation of defining land use covariates in this way is that biomass generation (i.e., leaf area index [LAI] and root depth [RD]) is not considered. There are clear temporal signatures of LAI and RD as crops develop, which will significantly modulate ET. The temporal signatures for each crop can be obtained from DSSAT crop simulation models, by averaging the LAI and RD for each crop type on each day of the year. Consequently, the six land use covariates can be replaced with two covariates: weighted averages of LAI and RD. The weights for each LAI and RD are obtained from the proportion of land use allocated for each land use type. This modification will reduce the number of covariates and make the model parsimonious.

Bibliography

Aelion, C., H. Davis, Y. Liu, A. Lawson, and S. McDermott. Validation of bayesian kriging of arsenic, chromium, lead, and mercury surface soil concentrations based on internode sampling. Environmental Science & Technology, 43(12):4432–4438, 2009.

Apipattanavis, S., G. Podestá, B. Rajagopalan, and R. W. Katz. A semiparametric multivariate and multisite weather generator. Water Resources Research, 43, 2007. doi: 10.1029/2006WR005714.

Apipattanavis, Somkiat, Federico Bert, Guillermo Podestá, and Balaji Rajagopalan. Linking weather generators and crop models for assessment of climate forecast outcomes. Agricultural and Forest Meteorology, 150:166–174, 2010.

Aragón, R., E.G. Jobbágy, and E.F. Viglizzo. Surface and groundwater dynamics in the sedimentary plains of the western pampas (argentina). Ecohydrology, 4:433–447, 2011.

Badano, N. Modelación hidrológica integrada en Grandes Cuencas de Baja Pendiente con Énfasis en la Evaluación de Inundaciones. PhD thesis, Facultad de Ingeniería, Universidad de Buenos Aires, 2010.

Badano, N., E. Lecertúa, M. Re, F. Re, and A.N. Menéndez. Modelación hidrológica integrada superficial-subterránea de una cuenca de llanura extensa. In XXIII Congreso Latinoamericano de Hidráulica, Cartagena de Indias, Colombia, 2008.

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

Baigorria, Guillermo A and James W Jones. Gist: A stochastic model for generating spatially and temporally correlated daily rainfall data. Journal of Climate, 23(22), 2010b.

Barnston, A.G., S. Li, S.J. Mason, D.G. DeWitt, L. Goddard, and X. Gong. Verification of the first 11 years of iri's seasonal climate forecasts. Journal of Applied Meteorology and Climatology, 49(3):493–520, 2010.

Barreiro, Marcelo. Influence of enso and the south atlantic ocean on climate predictability over southeastern south america. Climate Dynamics, 35(7):1493–1508, 2010.

Barros, Vicente R. and Gabriel E. Silvestri. The relation between sea surface temperature at the subtropical south-central pacific and precipitation in southeastern south america. Journal of Climate, 15(1):251–267, 2002.

Beersma, Jules J and T Adri Buishand. Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. Climate Research, 25(2):121–133, 2003.

Berger, T. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes, and policy analysis. XXIV International Conference of Agricultural Economics (IAAE), 25:245–260, 2001.

Berger, T., P. Schreinemachers, and J. Woelcke. Multi-agent simulation for the targeting of development policies in less-favored areas. Agricultural Systems, 88:28–43, 2006.

Berger, Thomas. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, rsource use changes, and policy analysis. XXIV International Conference of Agricultural Economics (IAAE), 25(0169):245–260, 2000.

Bert, F., E. Satorre, F. Toranzo, and G. Podestá. Climatic information and decision-making in maize crop production systems of the argentinean pampas. Agricultural Systems, 88 (2-3):180–204, 2006.

Bert, F., C. Laciana, G. Podestá, E. Satorre, and A. Menéndez. Sensitivity of ceres-maize simulated yields to uncertainty in soil properties and daily solar radiation. Agricultural Systems, 94(2):141–150, 2007.

Bert, F., S. Rovere, C. Macal, M. North, and G. Podestá. Lessons from a comprehensive validation of an agent based-model : The experience of the pampas model of argentinean agricultural systems. Ecological Modelling, 273:284–298, 2014.

Boulanger, Jean-Philippe, Julie Leloup, Olga Penalba, Matilde Rusticucci, Florence Lafon, and Walter Vargas. Observed precipitation in the paraná-plata hydrological basin: long-term trends, extreme conditions and enso teleconnections. Climate Dynamics, 24 (4):393–413, 2005.

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

Bray, D. and H. Storchvon . "prediction" or "projection"? Science Communication, 30(4): 534–543, 2009.

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

Bristow, Keith L. and Gaylon S. Campbell. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agricultural and Forest Meteorology, 31:159–166, 1984.

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

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

Calanca, P. and M. A. Semenov. Local-scale climate scenarios for impact studies and risk assessments: integration of early 21st century ENSEMBLES projections into the ELPIS database. Theoretical and Applied Climatology, 113:445–455, 2013.

Cano, Rafael, Carmen Sordo, and José M. Gutiérrez. Applications of bayesian networks in meteorology. In Gámez, José A., Serafín Moral, and Antonio Salmerón, editors, Advances in Bayesian Networks, volume 146 of Studies in Fuzziness and Soft Computing, pages 309–328. Springer Berlin Heidelberg, 2004.

Caraway, N. M., J. L. McCreight, and B. Rajagopalan. Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. Journal of Hydrology, 508:197–213, 2014.

Caron, Annie, Robert Leconte, and François Brissette. An improved stochastic weather generator for hydrological impact studies. Canadian Water Resources Journal, 33(3): 233–256, 2008.

Carpenter, B. Stan: A probabilistic programming language. Journal of Statistical Software, 2015.

Cash, D.W., W.C. Clark, F. Alcock, N.M. Dickson, N. Eckley, D.H. Guston, J. Jäger, and R.B. Mitchell. Knowledge systems for sustainable development. Proceedings of the National Academy of Sciences of the United States of America, 100(14):8086–8091, 2003.

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

Chandler, Richard E. and Howard S. Wheater. Analysis of rainfall variability using generalized linear models: A case study from the west of ireland. Water Resources Research, 38(10):1–11, 2002.

Chilès, J. P. and P. Delfiner. Geostatistics: Modeling Spatial Uncertainty. New York: Wiley, 1999.

Consortium, EC-Earth. Ec-earth model output prepared for cmip5 rcp85, served by esgf. World Data Center for Climate. CERA-DB "IHECr8", 2014.

Cooley, D. and S. Sain. Spatial hierarchical modeling of precipitation extremes from a regional climate model. Journal of Agricultural, Biological, and Environmental Statistics, 15(3):381–402, 2010.

Cooley, D., D. Nychka, and P. Naveau. Bayesian spatial modeling of extreme precipitation return levels. Journal of the American Statistical Association, 102:824–840, 2007.

Cui, H., A. Stein, and D. Myers. Extension of spatial information, bayesian kriging and updating of prior variogram parameters. Environmetrics, 6:373–384, 1995.

Doummar, Joanna, Martin Sauter, and Tobias Geyer. Simulation of flow processes in a large scale karst system with an integrated catchment model (mike she) - identification of relevant parameters influencing spring discharge. Journal of Hydrology, 426-427: 112–123, 2012.

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Research, 30:1371–1386, 2007.

Elliott, Joshua, David Kelly, James Chryssanthacopoulos, Michael Glotter, Kanika Jhunjhnuwala, Neil Best, Michael Wilde, and Ian Foster. The parallel system for integrating impact models and sectors (psims). Environmental Modelling & Software, 62:509–516, 2014.

Fassò, Alessandro and Francesco Finazzi. Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. Environmetrics, 22(6):735–748, September 2011. ISSN 11804009. doi: 10.1002/env.1123. URL http://doi.wiley.com/10.1002/env.1123.

Ferreyra, R. A., Guillermo P. Podestá, Carlos D. Messina, David Letson, Julio Dardanelli, Edgardo Guevara, and Santiago Meira. A linked-modeling framework to estimate maize production risk associated with ENSO-related climate variability in Argentina. Agricultural and Forest Meteorology, 107(3):177–192, April 2001a. ISSN 01681923. doi: 10.1016/S0168-1923(00)00240-9. URL http://linkinghub.elsevier.com/retrieve/pii/S0168192300002409.

Ferreyra, R. Andrés, Guillermo P. Podestá, Carlos D. Messina, David Letson, Julio Dardanelli, Edgardo Guevara, and Santiago Meira. A linked-modeling framework to estimate maize production risk associated with enso-related climate variability in argentina. Agricultural and Forest Meteorology, 107:177–192, 2001b.

Foufoula-Georgiou, Efi and Konstantine P Georgakakos. Hydrologic advances in space-time precipitation modeling and forecasting. In Recent advances in the modeling of hydrologic systems, pages 47–65. Springer, 1991.

Freeman, T., J. Nolan, and R. Schoney. An agent-based simulation model of structural change in canadian prairie agriculture, 1960-2000. Canadian Journal of Agricultural Economics-Revue Canadienne D Agroeconomie, 57:537–554, 2009.

Friend, A. D., A. K. Stevens, R. G. Knox, and M. G. R. Cannell. A process-based terrestrial biosphere model of ecosystem dynamics. Ecological Modelling, 95:249–287, 1997.

Furrer, E. M. and R. W. Katz. Generalized linear modeling approach to stochastic weather generators. Climate Research, 34:129–144, 2007.

Furrer, E. M. and R. W. Katz. Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resources Research, 44, 2008. doi: 10.1029/2008WR007316.

García, P.E., A.N. Menéndez, J. Lopez Laxagüe, F. Bert, G. Podestá, and E. Jobaggy. Caracterización de efectos espaciales en la napa debido a distintos usos del suelo en cuencas de llanura. In 2do Congreso Internacional de Hidrología de Llanuras, Santa Fe - Argentina, Septiembre 2014.

García, P.E., A.N. Menéndez, J. Lopez Laxagüe, F. Bert, G. Podestá, E. Jobaggy, and P. Arora. Land use as a possible strategy for managing water table depth in flat basins with shallow groundwater. Journal of Agricultural Water Management, (In Review), 2016.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, 2004.

Giorgi, F. Variability and trends of sub-continental scale surface climate in the twentieth century. part i: observations. Climate Dynamis, 18:675–691, 2002.

Goddard, L., A.G. Barnston, and S.J. Mason. Evaluation of the iri's "net assessment seasonal climate forecasts: 1997-2001. Bulletin of the American Meteorological Society, 84 (12):1761–1781, 2003.

Graham, Douglas N. and Michael B. Butts. Watershed Models, chapter Flexible, integrated watershed modelling. CRC Press, 2005.

Grimm, Alice M. Interannual climate variability in south america: impacts on seasonal precipitation, extreme events, and possible effects of climate change. Stochastic Environmental Research and Risk Assessment, 25(4):537–554, 2011.

Grimm, Alice M., Simone E. T. Ferraz, and Júlio Gomes. Precipitation anomalies in southern brazil associated with el niño and la niña events. Journal of Climate, 11(11):2863–2880, 1998.

Grimm, Alice M., Vicente R. Barros, and Moira E. Doyle. Climate variability in southern south america associated with el niño and la niña events. Journal of Climate, 13:35–58, 2000.

Grondona, M. O., G. P. Podestá, M. Bidegain, M. Marino, and H. Hordij. A stochastic precipitation generator conditioned on ENSO phase: A case study in southeastern South America. Journal of Climate, 13:2973–2986, 2000.

Haines, K., L. Hermanson, C. Liu, D. Putt, R. Sutton, A. Iwi, and D. Smith. Decadal climate prediction (project gcep). Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 367(1890):925–937, 2009.

Handcock, M. S. and M. L. Stein. A Bayesian analysis of kriging. Technometrics, 35: 403–410, 1993.

Hansen, J. Realizing the potential benefits of climate prediction to agriculture: issues, approaches, challenges. Agricultural Systems, 74(3):309–330, 2002.

Hansen, James W. and Theodoros Mavromatis. Correcting low-frequency variability bias in stochastic weather generators. Agricultural and Forest Meteorology, 109:297–310, 2001.

Hansen, J.W., A. Challinor, A. Ines, T. Wheeler, and V. Moron. Translating climate forecasts into agricultural terms: advances and challenges. Climate Research, 33(1):27–41, 2006.

Happe, K., A. Balmann, K. Kellermann, and C. Sahrbacher. Does structure matter? the impact of switching the agricultural policy regime on farm structures. Journal of Economic Behavior and Organization, 67:431–444, 2008.

Harrold, Timothy I, Ashish Sharma, and Simon J Sheather. A nonparametric model for stochastic generation of daily rainfall amounts. Water resources research, 39(12), 2003.

Hashmi, M., A. Shamseldin, and B. Melville. Statistical downscaling of precipitation: state-of-the-art and application of bayesian multi-model approach for uncertainty assessment. Hydrology and Earth System Sciences Discussions, 6:6535–6579, 2009.

Hashmi, Muhammad Zia, Asaad Y Shamseldin, and Bruce W Melville. Comparison of sdsm and lars-wg for simulation and downscaling of extreme precipitation events in a watershed. Stochastic Environmental Research and Risk Assessment, 25(4):475–484, 2011.

Hauser, Tristan and Entcho Demirov. Development of a stochastic weather generator for the sub-polar north atlantic. Stochastic Environmental Research and Risk Assessment, 27(7):1533–1551, 2013.

Herzer, H. Building Safer Cities: The Future of Disasater Risk. Disaster Risk Management Series, chapter Flooding in the Pampean Region of Argentina: The Salado Basin, pages 137–147. The World Bank, Disaster Management Facility, Washington, D.C., 2003.

Hoffman, M.D. and A. Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research, 2014.

Hurrel, J.W., T. Delworth, G. Danabasoglu, H. Drange, S. Griffies, N. Holbrook, B. Kirtman, N. Keenlyside, M. Latif, J. Marotzke, G.A. Meehl, T. Palmer, H. Pohlmann, T. Rosati, R. Seager, D. Smith, R. Sutton, A. Timmermann, K.E. Trenberth, and J. Tribbia. Decadal climate prediction: Opportunities and challenges. In Lido, Venice, editor, OceanObs'09, 2009.

Im, Sangjun, Hyeonjun Kim, Chulgyum Kim, and Cheolhee Jang. Assessing the impacts of land use changes on watershed hydrology using mike she. Environmental Geology, 57:231–239, 2009.

Jin, B., Y. Wu, B. Miao, X. Wang, and P. Guo. Bayesian spatiotemporal modeling for blending in situ observations with satellite precipitation estimates. Journal of Geophysical Research: Atmosphers, pages 1806–1819, 2014.

Jones, J., J. Hansen, F. Royce, and C.D. Messina. Potential benefits of climate forecasting to agriculture. Agriculture, Ecosystems and Environment, 82(1-3):169–184, 2000.

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie. The dssat cropping system model. European Journal of Agronomy, 18:235–265, 2003.

Katz, R. W. Precipitation as a chain-dependent process. Journal of Applied Meteorology, 16:671–676, 1977.

Katz, Richard W., Marc B. Parlange, and Philippe Naveau. Statistics of extremes in hydrology. Advances in Water Resources, 25:1287–1304, 2002.

Khalili, Malika, François Brissette, and Robert Leconte. Stochastic multi-site generation of daily weather data. Stochastic environmental research and risk assessment, 23(6): 837–849, 2009.

Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby. A daily weather generator for use in climate change studies. Environmental Modelling and Software, 22:1705–1719, 2007.

Kim, Tae-woong, Hosung Ahn, Gunhui Chung, and Chulsang Yoo. Stochastic multisite generation of daily rainfall occurrence in south florida. Stochastic Environmental Research and Risk Assessment, 22(6):705–717, 2008.

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

Kleiber, W., R. W. Katz, and B. Rajagopalan. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. Water Resources Research, 48, 2012. doi: 10.1029/2011WR011105.

Kleiber, W., R. W. Katz, and B. Rajagopalan. Daily minimum and maximum temperature simulation over complex terrain. Annals of Applied Statistics, 7:588–612, 2013.

Kristensen, K.J. and S.E. Jensen. A model for estimating actual evapotranspiration from potential evapotranspiration. Hydrology Research, 6(3):170–188, 1975.

Lall, U. and A. Sharma. A nearest neighbor bootstrap for resampling hydrological time series. Water Resources Research, 32:679–693, 1996. doi: 10.1029/95WR02966.

Larsen, M.A.D., S.H. Rasmussen, M. Drews, M.B. Butts, J.H. Christensen, and J.C. Refsgaard. Assessing the influence of groundwater and land surface scheme in the modelling of land surface-atmosphere feedbacks over the fife area in kansas, usa. Environmental Earth Sciences, 75(130), 2016.

Laxagüe, J. Lopez, P.E. García, A.N. Menéndez, and F. Bert. Influencia sobre el nivel freático en zonas de llanura debido al efecto del cambio en el uso del suelo y los condicionantes climáticos. In 2do Encuentro de Investigadores en Formación en Recursos Hídricos IFRH, Ezeiza, Buenos Aires, Argentina, Octubre 2014.

Leamer, Edward E. Specification searches. New York: Wiley, 1978.

Lennartsson, Jan, Anastassia Baxevani, and Deliang Chen. Modelling precipitation in Sweden using multiple step markov chains and a composite model. Journal of Hydrology, 363(1-4):42–59, December 2008. ISSN 00221694. doi: 10.1016/j.jhydrol.2008.10.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0022169408004848.

Liang, Xu, Dennis P. Lettenmaier, Eric F. Wood, and Stephen J. Burges. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres, 99(D7):14415–14428, July 1994.

Lima, C. H. R. and U. Lall. Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak and end. Water Resources Research, 45, 2009. doi: 10.1029/2008WR007485.

Liu, Hai-Long, Xi Chen, An-Ming Bao, and Ling Wang. Investigation of groundwater response to overland flow and topography using a coupled mike she/mike 11 modeling system for an arid watershed. Journal of Hydrology, 347:448–459, 2007.

Ma, Liang, Chunguang He, Hongfeng Bian, and Lianxi Sheng. Mike she modeling of ecohydrological processes: Merits, applications, and challenges. Ecological Engineering, In press, 2016.

McCullagh, P. and J. A. Nelder. Generalized Linear Models. Chapman and Hall, London, 1989.

McGinnis, S., D. Nychka, and L.O. Mearns. A new distribution mapping technique for bias correction of climate model output. In Lakshmanan, V., E. Gilleland, A. McGovern, and M. Tingley, editors, Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the Fourth International Workshop on Climate Informatics. Springer, 2015.

McNider, Richard T., John R. Christy, Don Moss, Kevin Doty, and Cameron Handyside. A real-time gridded crop model for assessing spatial drought stress on crops in the southeastern united states. Journal of Applied Meteorology and Climatology, 50(7):1459–1475, 2011.

Meehl, G.A., L. Goddard, J. Murphy, R.J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M.A. Giorgetta, E. Hawkins A.M. Greene, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale. Decadal prediction: Can it be skillful? Bulletin of the American Meteorological Society, 90(10):1467–1485, 2009.

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

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

Meinke, H. and R.C. Stone. Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Climatic Change, 70:221–253, 2005.

Meira, S., H. Gaigorri, E. Guevara, and M. Maturano. Calibration of soybean cultivars for the soygro model in two environments of argentina. In Global Soy Forum, pages 4–7, August 1999.

Menéndez, Angel N. and Nicolás D. Badano. Integrated hydrological modelling to assess flood and drought risk under climate and land use change. In 2nd International Interdisciplinary Conference on Predictions for Hydrology, Ecology and Water Resources Management, HydroPredict, 2010.

Mercau, J.L., J.L. Dardanelli, D.J. Collino, J.M. Andriani, A. Irigoyen, and E.H. Satorre. Predicting on-farm soybean yields in the pampas using cropgro-soybean. Field Crops Research, 100(2-3):200–209, 2007.

Meza, Francisco J. Variability of reference evapotranspiration and water demands. association to enso in the maipo river basin, chile. Global and Planetary Change, 47:212–220, 2005.

Minetti, J.L., W.M. Vargas, A.G. Poblete, L.R. Acuña, and G. Casagrande. Non-linear trends and low frequency oscillations in annual precipitation over argentina and chile, 1931-1999. Atmósfera, 16:119–135, 2003.

Montecinos, Aldo, Alvaro Díaz, and Patricio Aceituno. Seasonal diagnostic and predictability of rainfall in subtropical south america based on tropical pacific sst. Journal of Climate, 13(4):746–758, 2000.

Montgomery, Jacob M and Brendan Nyhan. Bayesian model averaging: Theoretical developments and practical applications. Political Analysis, 18(2):245–270, 2010.

Moussoulis, Elias, Ierotheos Zacharias, and Nikolaos P. Nikolaidis. Combined hydrological, rainfall-runoff, hydraulic and sediment transport modeling in upper acheloos river catchment. Desalination and Water Treatment, 57(25):11540–11549, 2016.

Nosetto, M.D., E.G. Jobbágy, R.B. Jackson, and G.A. Sznaider. Reciprocal influence of crops and shallow ground water in sandy landscapes of the inland pampas. Field Crops Research, 113:138–148, 2009.

Omre, H. Bayesian kriging - merging observations and qualified guesses in kriging. Mathematical Geology, 19(1):25–39, 1987.

Oram, P.A. Sensitivity of agricultural production to climatic change. Climatic Change, 7 (1):129–152, 1985.

Penalba, O.C., J.A. Rivera, and V.C. Pántano. The claris lpb database: constructing a long-term daily hydro-meteorological dataset for la plata basin, southern south america. Geoscience Data Journal, 1:20–29, 2014.

Pezzulli, S., P. Frederic, S. Majithia, S. Sabbagh, E. Black, R. Sutton, and D. Stephenson. The seasonal forecast of electricity demand: A hierarchical bayesian model with climatological weather generator. Applied Stochastic Models in Business and Industry, 22: 113–125, 2006.

Podestá, G., F. Bert, B. Rajagopalan, S. Apipattanavis, C. Laciana, E. Weber, W. Easterling, R. Katz, D. Letson, and A. Menendez. Decadal climate variability in the argentine pampas: regional impacts of plausible climate scenarios on agricultural systems. Climate Research, 40:199–210, 2009.

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

Qian, Budong, Samuel Gameda, Reinder Jongde , Pete Falloon, and Jemma Gornall. Comparing scenarios of canadian daily climate extremes derived using a weather generator. Climate Research, 41:131–149, 2010.

R Core Team, . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2014.

Racsko, P., L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. Ecological Modelling, 57:27–41, 1991.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review, 133:1155–1174, 2005.

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

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

Rajagopalan, Balaji, Upmanu Lall, and David G Tarboton. Evaluation of kernel density estimation methods for daily precipitation resampling. Stochastic Hydrology and Hydraulics, 11(6):523–547, 1997b.

Re, M., N. Badano, E. Lecertúa, F. Re, and A.N. Menéndez. Modelación hidrológica integrada de una cuenca de llanura extensa. In XVII Congreso sobre Métodos Numéricos y sus Aplicaciones, ENIEF'2008, San Luis, Argentina, 2008.

Refshaard, J.C., B. Storm, and V.P. Singh. Mike she. Computer models of watershed hydrology, pages 809–846, 1995.

Reich, B. and B. Shaby. A hierarchical max-stable spatial model for extreme precipitation. Annals of Applied Statistics, 6(4):1430–1451, 2012.

Richards, Lorenzo Adolph. Capillary conduction of liquids through porous mediums. Journal of Applied Physics, 1(5):318–333, 1931.

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

Richardson, Clarence W and David A Wright. Wgen: A model for generating daily weather variables. ARS (USA), 1984.

Ropelewski, C.F. and M.A. Bell. Shifts in the statistics of daily rainfall in south america conditional on enso phase. Journal of Climate, 21(5):849–865, 2008.

Ropelewski, C.F. and M.S. Halpert. Global and regional scale precipitation patterns associated with the el niño/southern oscillation. Monthly Weather Review, 115(8):1606–1626, 1987.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhiang, H.M. Van Doolden , H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie. The ncep climate forecast system. Journal of Climate, 19(15):3483–3517, 2006.

Sahu, S. and K. Mardia. A bayesian kriged kalman model for short-term forecasting of air pollution levels. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(1):223–244, 2005.

Sandu, Mirela-Alina and Ana Virsta. Applicability of mike she to simulate hydrology in argesel river catchment. Agriculture and Agricultural Science Procedia, 6:517–524, 2015.

Schoof, J.T., A. Arguez, J. Brolley, and J.J. O'Brien. A new weather generator based on spectral properties of surface air temperatures. Agricultural and Forest Meteorology, 135:241–251, 2005.

Schreinemachers, P. and T. Berger. An agent-based simulation model of human-environment interactions in agricultural systems. Environmental Modelling and Software, 26(7):845–859, 2011.

Seager, R., N. Naik, W.E. Baethgen, A.W. Robertson, Y. Kushnir, J. Nakamura, and S. Jurburg. Tropical oceanic causes of interannual to multidecadal precipitaiton variability in southeast south america over the past century. Journal of Climate, 23:5517–5539, 2010.

Semenov, M. A. Simulation of extreme weather events by a stochastic weather generator. Climate Research, 35:203–212, 2008.

Semenov, M. A. and E. M. Barrow. Use of a stochastic weather generator in the development of climate change scenarios. Climatic Change, 35:397–414, 1997.

Semenov, M. A., S. Pilkington-Bennett, and P. Calanca. Validation of ELPIS 1980–2010 baseline scenarios using the European Climate Assessment observed dataset. Climate Research, 51:1–9, 2013.

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

Singh, R., K. Subramanian, and J.C. Refsgaard. Hydrological modelling of a small watershed using mike she for irrigation planning. Agricultural Water Management, 41: 149–166, 1999.

Skansi, M.M., S.E. Nuñez, G.P. Podestá, V.H., and N. Garay. La sequía del año 2008 en la región húmeda argentina descripta a través del indice de precipitación estandarizado. CONGREMET, 2009.

Sloughter, J. M. L., A. E. Raftery, T. Gneiting, and C. Fraley. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Monthly Weather Review, 135(9):3209–3220, 2007.

Smith, R.L. and P.J. Robinson. A bayesian approach to the modeling of spatial-temporal precipitation data. In Gatsonis, Constantine, James S. Hodges, Robert E. Kass, Robert McCulloch, Peter Rossi, and Nozer D. Singpurwalla, editors, Case Studies in Bayesian Statistics, volume 121 of Lecture Notes in Statistics, pages 237–269. Springer New York, 1997.

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

Stainforth, D.A., M.R. Allen, E.R. Tredger, and L.A. Smith. Confidence, uncertainty and decision-support relevance in climate predictions. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 365(1857):2145–2161, 2007.

Stan Development Team, . Stan: A c++ library for probability and sampling, version 2.8.0. http://mc-stan.org, 2015a.

Stan Development Team, . Stan modeling language user's guide and reference manual, version 2.8.0. http://mc-stan.org, 2015b.

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

Stockdale, T.N., O. Alves, G. Boer, M. Deque, Y. Ding, A. Kumar, K. Kumar, W. Landman, S. Mason, P. Nobre, A. Scaife, O. Tomoaki, and W.T. Kun. Understanding and predicting seasonal-to-interannual climate variability - the producer perspective. Procedia Environmental Sciences, 1:55–80, 2010.

Tebaldi, C. and B. Sansó. Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach. Journal of the Royal Statistical Society (Series A), 172:83–106, 2009.

Tebaldi, C., L. Mearns, D. Nychka, and R. Smith. Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. Geophysical Research Letters, 31, 2004. doi: 10.1029/2004GL021276.

Thompson, J.R., H. Refstrup Sørenson, H. Gavin, and A. Refsgaard. Application of the coupled mike she/mike 11 modelling system to a lowland wet grassland in southeast england. Journal of Hydrology, 293:151–179, 2004.

Trenberth, K.E. and D.P. Stepaniak. Indices of el niño evolution. Journal of Climate, 14: 1697–1701, 2001.

Trujillo-Barreto, Nelson J, Eduardo Aubert-Vázquez, and Pedro A Valdés-Sossa. Bayesian model averaging in eeg/meg imaging. NeuroImage, 21(4):1300–1319, 2004.

UTN-FRA, . Plan de desarrollo integral del río salado: Estudio de impacto ambiental, social y territorial. Technical report, DiPSOH y MAA, Buenos Aires, Argentina, 2007.

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, and Chris Funk. A Bayesian kriging approach for blending satellite and ground precipitation observations. Water Resources Research, 51:1–14, 2015a. ISSN 00221694. doi: 10.1016/0022-1694(68)90080-2.

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, and Richard W. Katz. Coupled stochastic weather generation using spatial and generalized linear models. Stochastic Environmental Research and Risk Assessment, 29(2):347–356, 2015b.

Verdin, Andrew, Balaji Rajagopalan, William Kleiber, Guillermo Podestá, and Federico Bert. A conditional stochastic weather generator for seasonal to multi-decadal simulations. Journal of Hydrology, 2016.

Viallefont, Valerie, Adrian E Raftery, and Sylvia richardson. Variable selection and bayesian model averaging in case-control studies. Statistics in medicine, 20(21):3215–3230, 2001.

Viglizzo, E.F., Z.E. Roberto, F. Lertora, E.L. Gay, and J. Bernardos. Climate and land-use change in field-crop ecosystems of argentina. Agriculture, Ecosystems and Environment, 66(1):61–70, 1997.

Viglizzo, E.F., E.G. Jobbágy, F.C. Frank, R. Aragón, L. Orode , and V. Salvador. The dynamics of cultivation and floods in arable lands of central argentina. Hydrology and Earth System Sciences, 13:491–502, 2009.

Volinsky, Chris T, David Madigan, Adrian E Raftery, and Richard A Kronmal. Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(4):433–448, 1997.

Vrugt, Jasper A and Bruce A Robinson. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and bayesian model averaging. Water Resources Research, 43(1), 2007.

Wallis, T. W. and J. F. Griffiths. Simulated meteorological input for agricultural models. Agricultural and Forest Meteorology, 88:241–258, 1997.

Wheater, HS, RE Chandler, CJ Onof, VS Isham, E Bellone, C Yang, D Lekkas, G Lourmas, and M-L Segond. Spatial-temporal rainfall modelling for flood risk estimation. Stochastic Environmental Research and Risk Assessment, 19(6):403–416, 2005.

Wilby, R.L., D. Conway, and P.D. Jones. Prospects for downscaling seasonal precipitation variability using conditioned weather generator parameters. Hydrological Processes, 16:1215–1234, 2002.

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

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

Wilks, D. S. High-resolution spatial interpolation of weather generator parameters using local weighted regressions. Agricultural and Forest Meteorology, 148:111–120, 2008.

Wilks, D. S. Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteorological Applications, 16(3):361–368, 2009a.

Wilks, D. S. A gridded multisite weather generator and synchronization to observed weather data. Water Resources Research, 45, 2009b. doi: 10.1029/2009WR007902.

Wilks, D. S. and R. L. Wilby. The weather generation game: a review of stochastic weather models. Progress in Physical Geography, 23(3):329–357, 1999. doi: 10.1177/030913339902300302.

Wintle, Brendan A, Michel A McCarthy, Chris T Volinsky, and Rodney P Kavanagh. The use of bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology, 17(6):1579–1590, 2003.

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

Wright, Jonathan H. Bayesian model averaging and exchange rate forecasts. Journal of Econometrics, 146(2):329–341, 2008.

Yan, Zhongwei, Steven Bate, Richard E. Chandler, Valerie Isham, and Howard Wheater. An analysis of daily maximum wind speed in northwestern europe using generalized linear models. Journal of Climate, 15:2073–2088, 2002.

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater. Spatial-temporal rainfall simulation using generalized linear models. Water Resources Research, 41, 2005. doi: 10.1029/2004WR003739.

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

Zhang, Donghua, Henrik Madsen, Marc E. Ridler, Jens C. Refsgaard, and Karsten H. Jensen. Impact of uncertainty description on assimilating hydraulic head in the mike she distributed hydrological model. Advances in Water Resources, 86:400–413, 2015.