

ABSTRACT

Flexible Spatial Interpolation and Uncertainty Quantification: With Applications in
Radar Rainfall Estimation

RJ Waken, Ph.D.

Chairperson: Joon J. Song, Ph.D.

In quantitative precipitation estimation, prediction and uncertainty quantification are difficult due to the errors in the available data sources. Weather radars are used to predict precipitation with high spatial and temporal resolution, but do not measure ground level rainfall intensity, which is the quantity of interest. To account for the error resulting from the use of a proxy variable, predictions are calibrated to ground level measurements of the rainfall intensity rate with spatial prediction methods. For prediction at a specific location, kriging is a simple and popular spatial prediction method, but suffers from several shortcomings. In particular, prediction is quite unstable and fails when sample sizes are small and the error normality assumption necessary for uncertainty quantification with kriging predictors may not hold in real data sets. In this dissertation, we propose two flexible and efficient deterministic spatial predictors, with several advantages over kriging. We then further propose a robust data fusion uncertainty quantification scheme to produce gridded prediction output with stochastic errors. These methods are illustrated with radar rainfall data.

Flexible Spatial Interpolation and Uncertainty Quantification: With Applications in
Radar Rainfall Estimation

by

RJ Waken, B.S.

A Dissertation

Approved by the Department of Statistical Science

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Joon J. Song, Ph.D., Chairperson

James D. Stamey, Ph.D.

Jack D. Tubbs, Ph.D.

Dean M. Young, Ph.D.

Joe C. Yelderian, Ph.D.

Accepted by the Graduate School
December 2016

J. Larry Lyon, Ph.D., Dean

Copyright © 2016 by RJ Waken

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	ix
DEDICATION	x
1 Introduction	1
1.1 Motivating Example	1
1.2 Plan of the Dissertation	3
2 Literature Review	4
2.1 Current Methodology	5
2.1.1 Inverse Distance Weighting	6
2.1.2 Kriging	6
2.1.3 Regression-based Inverse Distance Weighting (RIDW)	8
3 RIDW for Spatial Interpolation with Spatio-temporal Data	11
3.1 Introduction	11
3.2 stRIDW	12
3.2.1 Distance Function	13
3.2.2 Uncertainty and Interval Estimate	14
3.3 Application to Radar Rainfall Estimation	15
3.3.1 Results	19

3.4	Discussion	22
4	RIDW for Anisotropic Spatial Interpolation	24
4.1	Introduction	24
4.2	Methodology	25
4.2.1	Anisotropy in Kriging Estimators	25
4.2.2	Anisotropic Weight Function for IDW	26
4.2.3	Point Estimate	27
4.2.4	Uncertainty and Interval Estimate	27
4.2.5	Anisotropy Parameter Selection	28
4.3	Application to Rainfall Data	29
4.3.1	Results	32
4.4	Conclusion	36
5	Stochastic Uncertainty Attachment for Gridded Deterministic Estimates	37
5.1	Introduction	37
5.2	An Approach to Uncertainty Quantification for Deterministic Outputs	40
5.2.1	Model Specification	40
5.2.2	Modeling Semi-continuous Processes	43
5.2.3	Prior Parameterization for the Degrees of Freedom Parameter .	44
5.3	Data Analysis	45
5.3.1	Results	45
5.4	Conclusion	50
6	Conclusion	51
A	Cross Validation through Variance Estimators for RIDW, stRIDW, and dRIDW	54

B	Hourly Prediction Maps for RIDW and stRIDW	56
C	Full Conditional Distributions	65
C.1	Full Conditionals	66
3.1.1	Degrees of Freedom Parameter ν	66
3.1.2	NN Model	67
3.1.3	NT Model	68
3.1.4	TN Model	68
3.1.5	TT Model	69
D	Uncertainty Quantification Maps	70
	BIBLIOGRAPHY	75

LIST OF FIGURES

1.1	Location of the 185 rain gauges in the Mount Bisl Region, South Korea, produced using the <code>ggmap</code> (Kahle and Wickham, 2013) package for the R programming language.	2
2.1	A visual representation of an empirical semivariogram (denoted by points), and parametric variogram fit (denoted by the solid line).	8
3.1	A scatter plot of rain gauge and radar rainfall measurements, pooled across all hours.	16
3.2	Boxplots of hourly rain gauge and radar data, one rain gauge observation (84mm) omitted in hour 23.	17
3.3	Histogram of pooled residuals resulting from a simple linear regression.	17
3.4	Empirical variogram (points) and parametric variogram fit (lines) by hour.	18
3.5	Residual errors for KED, RIDW, and stRIDW by hour.	20
3.6	Boxplots of CV estimated coverage probabilities by estimator.	21
4.1	Rain gauge locations reporting, hour 20 2012-07-14, created with <code>ggmap</code> (Kahle and Wickham, 2013).	30
4.2	Rain gauge measurements by radar estimates.	31
4.3	Residuals resulting from simple linear regression between the rainfall measurements and radar estimates.	31
4.4	Directional empirical variogram.	32
4.5	Box plots of $CVSE_i$ (on square root scale) plotted by estimator. An “X” indicates that this prediction was not captured in the prediction interval.	34
4.6	$\sqrt{CVSE_i}$ at each location for RIDW, dRIDW, and KED.	34
4.7	Rain gauge measurements captured in prediction intervals, by site.	35
4.8	95% PIL for RIDW and dRIDW.	35

5.1	Validation data (black points) and RIDW grid cell rainfall predictions across study area, hour 12 2012-07-06.	46
5.2	Residual error resulting from deterministic prediction of rainfall in grid cells with validation sites, 2012-07-06. Data from hour 12 is highlighted in blue.	46
5.3	Histogram of errors resulting from deterministic prediction of rainfall in grid cells with validation sites, hour 12 2012-07-06.	47
5.4	Boxplots of the validation site error $V(s_j) - \tilde{V}(s_j)$	48
B.1	RIDW predictions, hours 1-6.	57
B.2	RIDW predictions, hours 7-12.	58
B.3	RIDW predictions, hours 13-18.	59
B.4	RIDW predictions, hours 19-24.	60
B.5	stRIDW predictions, hours 1-6.	61
B.6	stRIDW predictions, hours 7-12.	62
B.7	stRIDW predictions, hours 13-18.	63
B.8	stRIDW predictions, hours 19-24.	64
D.1	Predictions in grid cells by model.	71
D.2	Predictions interval width in grid cells, by model.	72
D.3	$ R(A_i) - \tilde{R}(A_i) $ in grid cells, by model.	73
D.4	Proportion of zeros resulting from point mass scheme in grid cells, by model.	74

LIST OF TABLES

3.1	Descriptive statistics of the simple linear regression residuals by hour. . .	19
3.2	Summary of CV estimated prediction interval lengths for KED, RIDW, and stRIDW.	23
4.1	Comparison of $CVSE_i$ error mean and quantiles for RIDW, dRIDW, and KED.	33
4.2	Comparison of PIL mean and quantiles for RIDW, dRIDW, and KED. .	36
5.1	Posterior parameter mean and 95% credible intervals, by model.	48
5.2	Comparison of (5.9)	49

DEDICATION

To my wonderful family that supported me through my studies, and my friends
that kept me motivated.

CHAPTER ONE

Introduction

In this dissertation, we consider two deterministic spatial predictors and a robust spatial fully Bayesian data fusion model. The deterministic predictors combine mean modeling and flexible spatial interpolation, and are used for prediction in spatial analyses. The Bayesian robust data fusion model incorporates a validation data set as well as a set of areal deterministic predictions, and produces areal predictions with stochastic uncertainty. We begin with a motivating example.

1.1 Motivating Example

In radar rainfall estimation, radar-based rainfall intensity rates need to be calibrated to ground level rainfall intensity rates to insure accurate precipitation prediction. Further, uncertainty quantification is desired, but is not produced by the deterministic algorithm that generates the radar-based rainfall intensity rates. The study area is covered by the radar circle centered at Mount Bisl located near Daegu in the southeastern region of South Korea. The rain gauge network (map of all locations given in Figure 1.1) consists of 185 tipping bucket rain gauges, which record rainfall intensity in 0.5 mm/hour resolution and measure zero rainfall for any amount less than 0.5 mm. Chapters Three and Five of this dissertation will focus on the case of 2012-07-06, in which 1514 total rain gauge rainfall intensity measurements are recorded across 24 hours with hourly sample sizes ranging from 4 to 168. Specific focus is put on hour 12 in Chapter Five. Chapter Four covers the case of hour 20, 2012-07-14.



Figure 1.1: Location of the 185 rain gauges in the Mount Bisul Region, South Korea, produced using the `ggmap` (Kahle and Wickham, 2013) package for the R programming language.

The radar data are collected from S-band dual polarization radar over the 150 km range with a gate spacing of 125 m, a 1° resolution beam, and 6 elevation angles from -0.5° to 1.6° every 2.5 minutes. The radar rainfall intensity rates are calculated by using the measured reflectivity from the lowest elevation that is not affected by beam blockage and non-meteorological echo. The radar rainfall intensity rates from reflectivity are estimated with the following equation,

$$R(Z_H) = 0.017Z_H^{0.71}, \quad (1.1)$$

where Z_H is radar reflectivity. The calculated radar-based rainfall intensity rates, in mm/hour, are converted to 1×1 km resolution.

1.2 Plan of the Dissertation

A brief literature review is given in Chapter Two. In Chapter Three, we discuss incorporating previous time information into spatial interpolation with mean modeling for the radar calibration problem described. In Chapter Four, we propose a deterministic predictor for anisotropic spatial interpolation combined with mean modeling for the radar calibration problem. In Chapter Five, we implement a robust Bayesian data fusion model to attach uncertainty to gridded deterministic predictions. Chapter Six follows with a conclusion and discussion.

CHAPTER TWO

Literature Review

Quantitative precipitation estimation (QPE) is applicable in a wide array of fields, including agriculture, hydrology, and atmospheric sciences. Rain gauges and weather radar are two common instruments used to measure precipitation. Rain gauges measure proximate surface precipitation at a sparse set of locations, while weather radar provides less accurate precipitation estimates at a high temporal and spatial resolution.

Radar estimated rainfall rates are computed from radar reflectivity data using a radar reflectivity-rainfall rate relationship, which, in the present case, is given by (1.1). Radar measurements can be generated in real-time from either a single radar or a network of radars. This, combined with the high spatial resolution of predictions provided by radar measurement, makes radar based prediction more convenient than rainfall gauge measurement. However, purely radar-based rainfall prediction suffers from several types of errors such as radar mis-calibration (Vivekanandan et al., 2003), (Ryzhkov et al., 2005), (Kwon et al., 2015), non-meteorological echo (Cho et al., 2006), variation of drop size distribution (Lee, 2006), and radar data quality issues (Fulton et al., 1998), (Maddox et al., 2002). Neural network methods (Xia and Chandrasekar, 1997), Bayesian methods (Seo and Smith, 1991), and merging methods with rain gauges (Goudenhooff and Delobbe, 2009) have been used to overcome these problems. It has been shown that, even in the presence of a dense grid of rain gauges in the area of interest, incorporation of radar and rain gauge information in predicting rainfall values is more effective than simply interpolating

between rainfall measurements from rain gauges (Song et al., 2015), (Sun et al., 2000), (Sempere-Torres et al., 1999).

Given the geospatial nature of rain gauge measurement data, spatial interpolation methods are often used in QPE. Inverse distance weighting (IDW) and kriging are commonly employed for the purpose. IDW is a popular interpolator due to its simple and fast implementation, but suffers from poor prediction performance and does not allow stochastic uncertainty quantification. Kriging is able to quantify uncertainty as a stochastic predictor and create interval estimates, but requires a restrictive set of assumptions. It is expected to quantify uncertainty poorly and produce inaccurate interval estimates when those assumptions are not met (Cressie, 1993), or in cases where small sample sizes are present (Pebesma, 2000). Further, kriging methods require that the spatial structure of the underlying spatial process is presumed known, which is quite susceptible to misspecification. Joseph and Kang (2012) introduced regression-based inverse distance weighting (RIDW) as an alternative to kriging that combines mean modeling through regression and residual adjustment through IDW.

2.1 Current Methodology

IDW and kriging (Cressie, 1993) are commonly applied to rainfall prediction in order to account for the spatial structure assumed in the data. Kriging methods are stochastic in nature, while IDW is deterministic (Isaaks and Srivastava, 1989). Variants of kriging methods include ordinary kriging, kriging with external drift, and others (Pebesma, 2000). Here, we provide a brief review of IDW, kriging, and RIDW.

2.1.1 Inverse Distance Weighting

The IDW predictor for a quantity $y_{i'}$ at a new location i' is given by

$$\hat{y}_{i'} = \mathbf{v}_{i'}^T \mathbf{y}, \quad (2.1)$$

where

$$\mathbf{v}_{i'} = v_1, v_2, \dots, v_n, \quad (2.2)$$

$v_i = w_i / \sum_{i=1}^n w_i$, \mathbf{y} is the vector of observations, and w_i is a function of the inverse distance between locations i and i' that weights the information contributed by each location. It is common to use $w_i = 1/(d_i)^p$, where d_i denotes the distance between locations i and point i' and $p > 0$. Larger values of p give less weight to distant observations, and $p = 1$ gives a “cone-like” surface. As a deterministic method, IDW requires no assumptions and is distribution-free.

2.1.2 Kriging

Variogram estimation is an integral aspect of kriging because it characterizes the spatial structure of the underlying spatial process through the semivariance $\gamma(h)$, given by

$$2\gamma(h) = \text{Var}(\mu(y_i) - \mu(y_{i+h})),$$

where h is the spatial lag and $\mu(y_i)$ denotes the process mean at location i . As distances are continuous in nature, multiple point pairs exactly h units apart are not possible. Instead, the variogram is typically estimated by averaging over the N_k point pairs in the k th binned distance interval h_k created across a set of distance intervals,

$$2\hat{\gamma}(h_k) = \frac{1}{N_k} \sum_{j=1}^{N_k} (\mu(y_{i_j}) - \mu(y_{i_j+h_k}))^2, \quad (2.3)$$

for all distance intervals h_k , where $y_{i+h_k} \in h_k$ with respect to the location of y_i . This is the empirical variogram. If the number of point pairs N_k in some interval h_k is small, the empirical variogram is less reliable in that region.

After computing the empirical variogram, a theoretical variogram model is fit to (2.3), typically through least squares or maximum likelihood procedures, along with “eye-balling” fitting methods through plotting. For further information on variogram analysis, see (Cressie, 1993). In this paper, variogram models are fit through weighted least squares estimation. A typical empirical variogram with a parametric variogram fit is given in Figure 2.1.

Kriging is the best linear unbiased predictor (BLUP) when error normality and second order stationarity are assumed. In addition, the stochastic nature of this predictor enables kriging to quantify uncertainty and construct prediction intervals under error normality.

For all kriging methods, including ordinary kriging and kriging with external drift, a variogram model needs to be explicitly specified, and a single variogram model may not be appropriate for all hours of a rainfall event. Further, when the sample size is small, the variogram modeling procedures tend to be unstable and lead to poor uncertainty and interval estimates as well as spatial structure misspecification.

The ordinary kriging (OK) predictor is a stochastic linear interpolator,

$$\hat{y}_{i'} = \sum_{i=1}^n \lambda_i y_i,$$

where the λ_i are chosen to minimize the mean squared prediction error, $E(y_{i'} - \hat{y}_{i'})^2$, subject to the restriction $\sum_{i=1}^n \lambda_i = 1$.

Kriging with external drift (KED) is a variant of kriging that models a non-spatial trend as well as a spatial trend through

$$\hat{y}_{i'} = \mathbf{x}_{i'}^T \boldsymbol{\beta} + \sum_{i=1}^n \lambda_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}),$$

where \mathbf{x} and $\boldsymbol{\beta}$ are $p \times 1$ vectors of covariates and regression coefficients, respectively. λ_i and $\boldsymbol{\beta}$ are chosen to minimize $E(y_{i'} - \hat{y}_{i'})^2$ among all λ_i such that $\sum_{i=1}^n \lambda_i x_{ij} = x_{i'j}$ for $j = 0, \dots, p$.

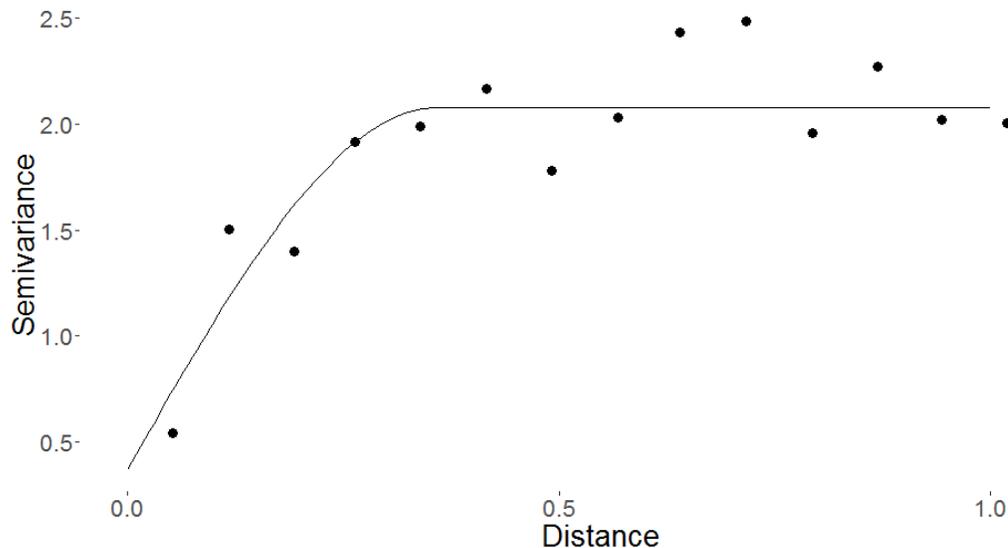


Figure 2.1: A visual representation of an empirical semivariogram (denoted by points), and parametric variogram fit (denoted by the solid line).

In the radar rainfall prediction analysis, the radar reflectivity rate is used as the covariate. This estimator is also the BLUP for $y_{i'}$, and stochastic interval estimates can be obtained under the assumptions of error normality and second order stationarity (Cressie, 1993). Similar to OK, the variogram estimator tends to be unstable for small n , which causes issues with uncertainty and interval estimation.

2.1.3 Regression-based Inverse Distance Weighting (RIDW)

RIDW models the global trend through a mean function $\mu(\mathbf{x}_i; \boldsymbol{\beta})$ in a regression setting

$$\begin{aligned} y_i &= \mu(\mathbf{x}_i; \boldsymbol{\beta}) + e_i \\ &= \mathbf{x}_i^T \boldsymbol{\beta} + e_i, \end{aligned}$$

where \mathbf{x}_i is the covariate vector, y_i is the response measurement, and e_i is the residual error. The RIDW predictor for a new location i' is given by

$$\begin{aligned} \hat{y}_{i'}(\mathbf{x}_{i'}) &= \mu(\mathbf{x}_{i'}; \hat{\boldsymbol{\beta}}) + \mathbf{v}_{i'}^T \mathbf{e} \\ &= \mathbf{x}_{i'}^T \hat{\boldsymbol{\beta}} + \mathbf{v}_{i'}^T \mathbf{e}, \end{aligned} \tag{2.4}$$

where $\mathbf{e} = (e_1, \dots, e_n)$, $e_i = y_i - \mu(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) = \mathbf{y} - \mathbf{X}^T \hat{\boldsymbol{\beta}}$, $\mu(\mathbf{x}_{i'}; \hat{\boldsymbol{\beta}}) = \mathbf{x}_{i'}^T \hat{\boldsymbol{\beta}}$ is the mean function at location i' with least squares estimator $\hat{\boldsymbol{\beta}}$, and $\mathbf{v}_{i'}$ is the vector of weights whose i th element is a function of the inverse distance between station i' and station i . This formulation accounts for global trends through $\mu(\mathbf{x}_{i'}; \hat{\boldsymbol{\beta}})$ and local spatial trends through the IDW-based residual adjustment, $\mathbf{v}_{i'}^T \mathbf{e}$.

2.1.3.1 Uncertainty estimate. Joseph and Kang (2012) proposed a heuristic method to assess uncertainty in the RIDW predictions. The local variance estimator of the RIDW prediction at the i' th location is

$$s_{i'}^2 = \sum_{i=1}^n v_i (e_i - \mathbf{v}_{i'}^T \mathbf{e})^2, \quad (2.5)$$

where v_i is the i th element of $\mathbf{v}_{i'}$. As noted in the paper, this estimator fails to take the uncertainty in the estimation of $\boldsymbol{\beta}$ into account. Let \mathbf{X} be the $n \times b$ design matrix, $\mathbf{c}_{i'}(\mathbf{x}) = \mathbf{x}_{i'} - \mathbf{X}^T \mathbf{v}_{i'}$, and $\mathbf{d}_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{X}^T \mathbf{v}_{i'}$. To incorporate the uncertainty in the regression estimates, the following estimator was proposed

$$\begin{aligned} s_{i'}^2 &= \sum_{i=1}^n v_i (e_i - \mathbf{v}_{i'}^T \mathbf{e})^2 \\ &+ \sigma^2 \sum_{i=1}^n v_i \mathbf{d}_i(\mathbf{x})' (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{d}_i(\mathbf{x}) \\ &+ \sigma^2 \mathbf{c}_{i'}(\mathbf{x})' (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}_{i'}(\mathbf{x}). \end{aligned} \quad (2.6)$$

When applied, the estimator $\hat{\sigma}^2 = \sum_{i=1}^n e_i^2 / (n - b - 1)$ is used in place of σ^2 .

2.1.3.2 Interval estimate. As distributional assumptions have not been made, stochastic prediction intervals cannot be created. Instead Joseph and Kang (2012) proposed the prediction interval estimate

$$\hat{y}_{i'}(\mathbf{x}_{i'}) \pm \kappa_{\alpha} s_{i'}. \quad (2.7)$$

Define

$$\begin{aligned}\hat{y}_{(-i)} &= \mu(\mathbf{x}_{(-i)}; \hat{\boldsymbol{\beta}}) + \mathbf{v}_i^T \mathbf{e}_{(-i)}, \\ CV_i &= y_i - \hat{y}_{(-i)},\end{aligned}\tag{2.8}$$

where $\mu(\mathbf{x}_{(-i)}; \hat{\boldsymbol{\beta}})$ and $\mathbf{e}_{(-i)}$ are found by cross validation of the least squares regression estimates. Then, κ_α is set as the upper α sample quantile of

$$\frac{|CV_i|}{s_{-i}}, \quad i = 1, \dots, n,$$

where s_{-i} is found through cross validation of (2.6) (see (A.1) for a full formulation).

CHAPTER THREE

RIDW for Spatial Interpolation with Spatio-temporal Data

3.1 Introduction

As previously stated, QPE algorithms solely based on radar measurements are known to be error prone. Radar measurement prediction accuracy can be bolstered through adjustment based on integration of surface level rain gauge validation data. Because rainfall data are geospatial in nature, an adjustment algorithm with a spatial interpolation component is necessary. A popular approach to problems concerning spatial interpolation in tandem with regression is kriging (Cressie, 1993).

As rainfall events are dynamic, and each event requires tuning for accurate radar based rainfall estimation, an automatic approach is desired. Further, with the nature of errors that exist in radar observations, prediction intervals for QPE are desired to aid in understanding the uncertainty associated with predictions.

When data are collected across time and space, incorporation of previous time information may bolster spatial structure specification, leading to more accurate predictions. Using past information is especially beneficial when current information is insufficient for spatial structure specification and temporal correlation is significant. Spatio-temporal kriging methods are computationally costly and require more extensive tuning than spatial kriging (Pebesma, 2004). Spatio-temporal data analysis commonly aims at prediction across time and space, but this paper focuses on only current time spatial prediction while borrowing previous time information, which is more appropriate to the motivating problem concerning radar calibration for QPE. However, it would be straightforward to include temporal prediction by adding a temporal component to the mean function in the proposed method.

The objective of the study is to propose a flexible, automatic, and efficient spatial predictor, called spatio-temporal regression-based inverse distance weighting (stRIDW). This extension of RIDW (Joseph and Kang, 2012) incorporates previous time error interpolation into current time spatial structure estimation. This method gives several advantages over other predictors, in that it requires no stochastic assumptions, tunes the necessary weight parameter automatically, provides efficient computation, and allows us to borrow information from previous time steps.

The remainder of this chapter is organized as follows: Section 3.2 introduces stRIDW for incorporating current and previous time information in spatial interpolation. Section 3.3 gives a discussion of the data analyzed, and follows with an application and comparison of stRIDW, RIDW, and KED. Conclusions and a discussion are given in Section 3.4.

3.2 *stRIDW*

In cases where current time information is sparse, it may be beneficial to borrow information from previous time steps when making predictions at the current time step to bolster spatial structure specification. Because distance in time and space are not on the same scale, the inverse distance weights chosen must appropriately reflect the relationship between observations (or residuals) over time and space. The inverse distance weight between a new location i' in time t and observed location i in time k is given by

$$v_{i,k} = \frac{w_{i,k}}{\sum_{k \in W} \sum_{j=1}^{n_k} w_{j,k}} \quad (3.1)$$

where W denotes the window length of time steps in consideration, t is the current time, $w_{i,k}$ is determined by some function of the inverse distance, and n_k is the number of observations in time step k .

The stRIDW predictor proposed in this paper is similar to (2.4), but uses current and past time residuals, giving

$$\hat{y}_{i'} = \mu(\mathbf{x}_{i',t}; \boldsymbol{\beta}_t) + \sum_{k \in W} \sum_{i=1}^{n_k} e_{i,k} v_{i,k}, \quad (3.2)$$

where $e_{i,k} = y_{i,k} - \mu(\mathbf{x}_{i,k}; \boldsymbol{\beta}_k)$ and $v_{i,k}$ is the weight in (3.1).

3.2.1 Distance Function

As addressed previously, it is essential to find a reasonable way to scale distance between the time and space domain. In this paper, we propose a new distance function between (i', t) and (i, k)

$$d_{i,k} = \sum_{k \in W} \sum_{i=1}^{n_k} \sqrt{\theta(\mathbf{l}_{i'} - \mathbf{l}_i) + (1 - \theta)(t - k)},$$

where $(\mathbf{l}_{i'} - \mathbf{l}_i)$ is the Euclidean distance between i' and i , $t - k$ is the temporal lag, and $\theta \in [0, 1]$ calibrates the relative importance of spatial and temporal distance for the residual adjustment in (3.2). θ is determined by minimizing the mean squared cross validation error (MSCV)

$$\text{MSCV} = \sum_{i=1}^{n_t} (e_{i,t} - \hat{e}_{i,t})^2,$$

where

$$\hat{e}_{i,t} = \sum_{j \neq i}^{n_t} e_{j,t}^{(-i)} v_{j,t} + \sum_{k \in W, k \neq t} \sum_{i=1}^{n_k} e_{i,k} v_{i,k},$$

and $e_{j,t}^{(-i)}$ is the residual at location j for time t obtained through leave one out cross validation for the i th observation. This distance can be used to determine the weight in (3.1), which depends on the nature of data in question. We consider $w_{i,t} = 1/d_{i,k}^2$ in the application later.

3.2.2 Uncertainty and Interval Estimate

The variance estimator for the residual error adjustment for stRIDW is given by

$$s_{i'}^2 = \sum_{k \in W} \sum_{i=1}^{n_k} v_{i,k} (e_{i,k} - \sum_{k \in W} \sum_{i=1}^{n_k} v_{i,k} e_{i,k})^2,$$

but, just as with (2.5), this estimator fails to account for uncertainty in mean modeling. The following variance estimate for stRIDW is proposed, which, in the case of linear regression for mean modeling, accounts for the uncertainty in $\mu(\mathbf{x}_{i',t}; \boldsymbol{\beta}_t)$ through

$$\begin{aligned} s_{i'}^2 &= \sum_{k \in W} \sum_{i=1}^{n_k} v_{i,k} (e_{i,k} - \sum_{k \in W} \sum_{i=1}^{n_k} v_{i,k} e_{i,k})^2 \\ &+ \sum_{k \in W} \sum_{i=1}^{n_k} \sigma_k^2 v_{i,k} \mathbf{d}_i(\mathbf{x}_{i,k}, k)' (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{d}_i(\mathbf{x}_{i,k}, k) \\ &+ \sum_{k \in W} \sigma_k^2 \mathbf{c}_{i'}(\mathbf{x})' (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{c}_{i'}(\mathbf{x}), \end{aligned} \quad (3.3)$$

where $\mathbf{x}_{i,k}$ is the covariate vector at i at time k , $\mathbf{c}_{i'}(\mathbf{x}) = \mathbf{x}_{i'} - \sum_{k \in W} \sum_{i=1}^{n_k} \mathbf{x}_{i,k} v_{i,k}$, and $\mathbf{d}_i(\mathbf{x}_{i,k}, k) = \mathbf{x}_{i,k} - \sum_{k \in W} \sum_{i=1}^{n_k} \mathbf{x}_{i,k} v_{i,k}$. When applied, σ_k^2 is replaced with its estimate $\hat{\sigma}_k^2 = \sum_{k \in W} \sum_{i=1}^{n_k} e_{i,k}^2 / (n - b - 1)$.

Similar to RIDW, distributional assumptions have not been made, so stochastic interval estimates cannot be created. A variant of (2.7) is used, but the cross validation formulation is adjusted to include information borrowed from previous time steps by the spatio-temporal residual adjustment. Recall, however, that the stRIDW predictor is only concerned with spatial prediction in current time, so cross validation through the previous time steps is not necessary. Let

$$\hat{y}_{(-i)} = \mu(\mathbf{x}_{i,t}; \boldsymbol{\beta}_t) + \sum_{j \neq i}^{n_t} e_{j,t} v_{j,t} + \sum_{k \in W, k \neq t} \sum_{j=1}^{n_k} e_{j,k} v_{j,k},$$

$CV_i = y_i - \hat{y}_{(-i)}$, and s_{-i}^2 be the cross validation realization of (3.3) (see (A.2) for a

full formulation). Then, we set κ_α as the upper α quantile of

$$\frac{|CV_i|}{s_{-i}}, \quad i = 1, 2, \dots, n.$$

The stRIDW prediction interval estimate for new location i' is then created through

$$\hat{y}_{i'} \pm \kappa_\alpha s_{i'}.$$

3.3 Application to Radar Rainfall Estimation

Recall the radar rainfall intensity rate, taken from the relationship in (1.1), and rainfall intensity gauge data described in the motivating example. Figure 3.1 shows rain gauge and radar observation pairs pooled across all time steps. It is clear that there is a strong linear relationship between the gauge and radar measurements, which indicates that a simple linear regression model may be appropriate to account for the radar information in rainfall estimation. Boxplots for hourly rain gauge and radar data are given in Figure 3.2, which indicate similar empirical distributional attributes across all hours for these two variables. Given the clear linear relationship between radar and rain gauge measurements of rainfall, it is more appropriate to examine the properties of the residuals resulting from simple linear regression between gauge and radar measurements when analyzing error distribution properties.

Figure 3.3 displays the histogram of pooled hourly simple linear regression residuals, which are clearly non-normal and heavily positively skewed. The histogram appears to be heavy tailed, and there are some highly extreme residual values, suggesting that the errors are leptokurtic. The residuals resulting from hourly simple linear regression between gauge observations and radar estimates data are summarized in Table 3.1. The summary statistics in Table 3.1 show that the hourly residuals tend to be right skewed, indicating that there may be some extreme positive residuals, as well as leptokurtic, with all but three hours showing a sample kurtosis larger than three. The Shapiro-Wilk test (Shapiro and Wilk, 1965) is used to test normality in the residuals, and the small p-values in Table 3.1 indicate that

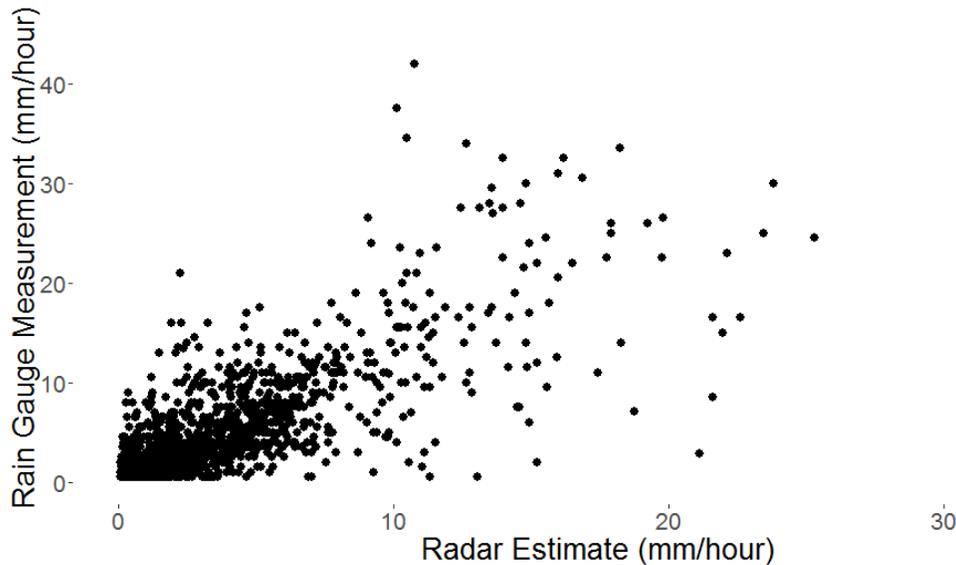


Figure 3.1: A scatter plot of rain gauge and radar rainfall measurements, pooled across all hours.

they may be non-normal except in a few time steps. All of these properties tend to occur regardless of sample size, which varies from $n_3 = 4$ to $n_{12} = 168$.

All empirical variograms and parametric variogram fits, which are of the exponential (hours 18 and 21), spherical (hours 3, 5, 6, 8, 10, 11, 14, 15, 19, 20, and 23), Gaussian (hours 4, 7, 13, 17, and 24), or Matern (hours 1, 2, 9, and 12) family, for the KED predictor are plotted in Figure 3.4. It is apparent that a single parametric variogram model fit across all hours is inappropriate. In particular, the heterogeneity of variogram fit is prominent when sample size is small, such as in hours 2, 3, and 4. The variogram typically increases as a function of the distance under positive spatial autocorrelation, but some cases, including hours 2, 7, 8, and 22, show problematically large empirical semivariances for smaller distance pairs indicating that the variance may not be increasing as a function of distance. In each case, the parametric variogram fit was chosen from the aforementioned families such that the sum of weighted squared errors was minimized.

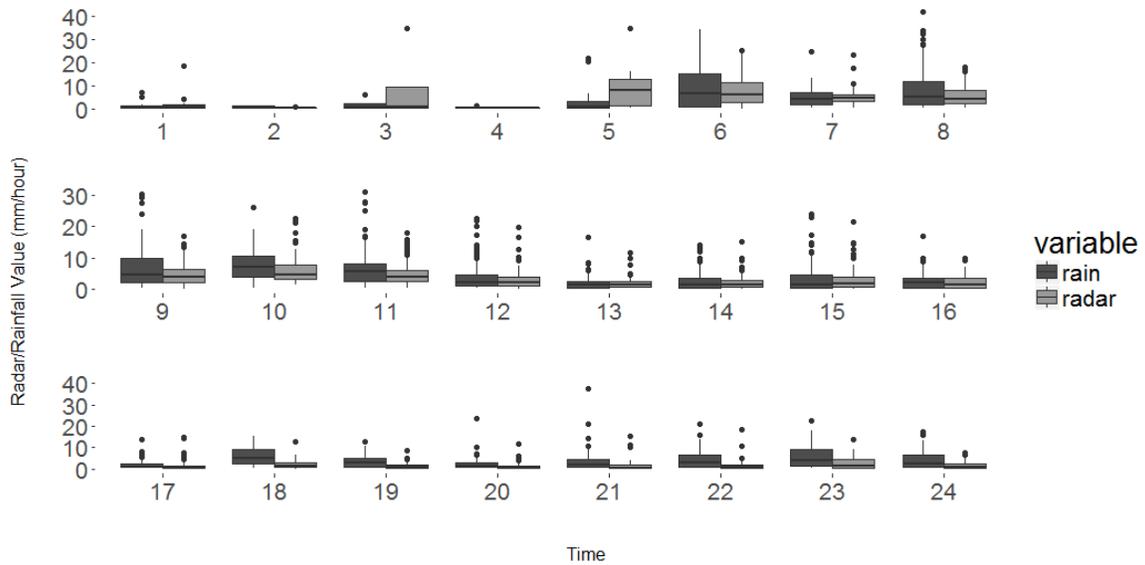


Figure 3.2: Boxplots of hourly rain gauge and radar data, one rain gauge observation (84mm) omitted in hour 23.

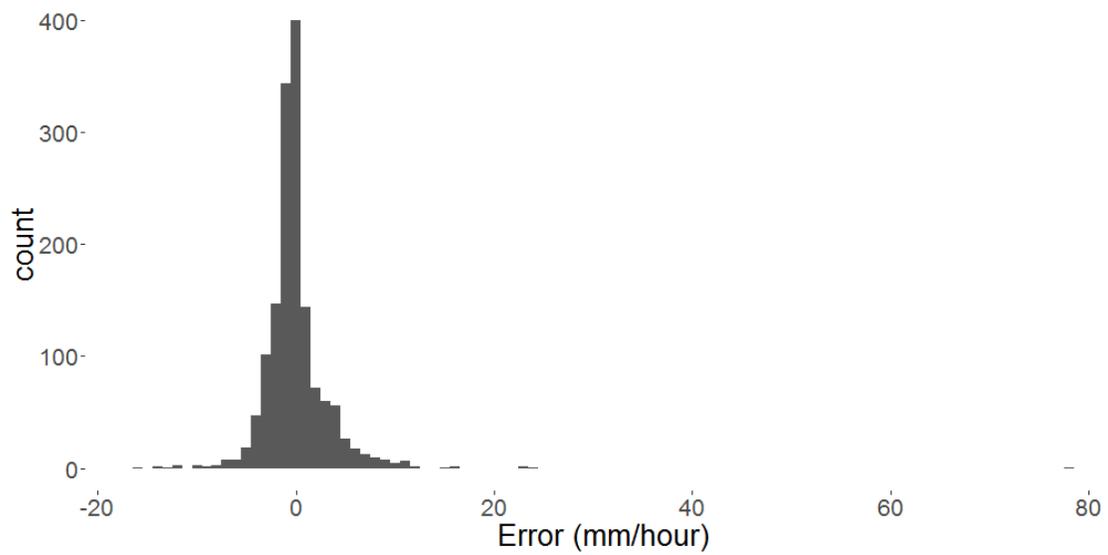


Figure 3.3: Histogram of pooled residuals resulting from a simple linear regression.

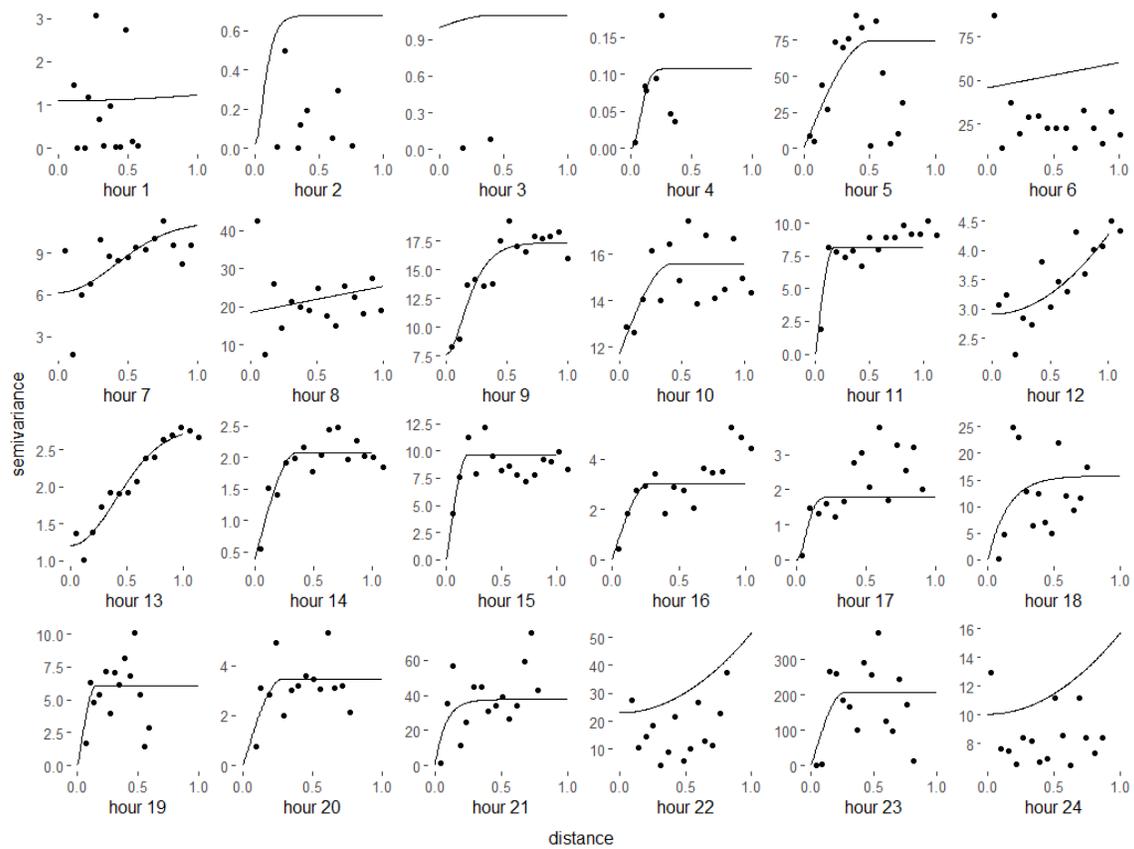


Figure 3.4: Empirical variogram (points) and parametric variogram fit (lines) by hour.

Table 3.1: Descriptive statistics of the simple linear regression residuals by hour.

Hour	n	Skewness	Kurtosis	SW p-value
1	16	2.7398	10.0402	<0.0001
2	8	0.6714	1.9216	0.0986
3	4	0.8479	2.0356	0.1785
4	9	0.8971	3.2753	0.4755
5	14	1.4440	4.0968	0.0025
6	47	2.0699	11.2603	<0.0001
7	71	0.9298	4.0546	0.0028
8	88	1.2559	10.3020	<0.0001
9	96	0.6853	4.7698	0.0001
10	106	-0.0116	4.4936	0.0185
11	150	0.5474	6.7343	<0.0001
12	168	0.8012	8.6911	<0.0001
13	151	1.9740	9.8583	<0.0001
14	107	0.8139	4.1735	<0.0001
15	88	0.1447	8.0373	<0.0001
16	76	1.7668	7.0435	<0.0001
17	51	1.2135	4.6640	<0.0001
18	20	0.9071	2.9105	0.0300
19	30	1.8066	6.6041	0.0002
20	36	-0.8612	6.2664	0.0005
21	37	1.7565	9.7791	<0.0001
22	32	1.4249	4.9746	0.0001
23	47	5.9402	38.9284	<0.0001
24	62	0.3760	4.9648	0.0001

3.3.1 Results

Given the strong linear relationship apparent between radar and gauge observations in Figure 3.1, simple linear regression is used to model the mean function in both RIDW and stRIDW. The performance of these predictors are compared with KED predictions through leave-one-out cross validation (CV).

Figure 3.5 gives boxplots of the residual errors for each of the three predictors (one extreme residual from hour 23 is omitted for each estimator). As we can see, there are a large number of outliers for all of the predictors. The errors for all three predictors appear to be centered about zero, indicating no obvious bias. The mean

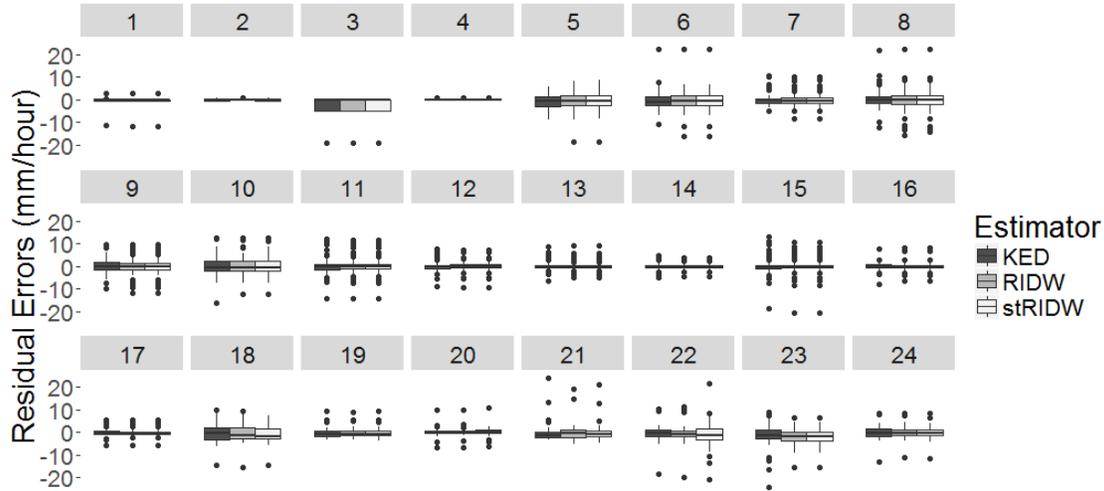


Figure 3.5: Residual errors for KED, RIDW, and stRIDW by hour.

square prediction error achieved by the KED estimator is marginally smaller than that of the RIDW/stRIDW estimators.

The efficacy of the prediction interval estimates obtained for RIDW, stRIDW, and KED is compared on two criteria: CV estimated coverage probability and prediction interval length. For all cases, $\alpha = 0.05$ was used to create 95% prediction intervals. Figure 3.6 shows the estimated coverage probability for the three methods. It is found that the probabilities for RIDW and stRIDW were equal in every hour. The medians of estimated coverage probabilities for RIDW and stRIDW were closer to the targeted 95% value, and the variation of the probabilities for the methods is much smaller than that of KED. RIDW and stRIDW also avoided overcoverage, which can result in interval lengths that are too large. Note that KED has an estimated coverage probability of 100% in three time steps out of 24 hours, and seven time steps have estimated coverage probabilities larger than 95%. In three time steps, RIDW and stRIDW exhibited lower than typically acceptable estimated coverage probabilities, however, sample sizes were very small, with $n = 8, 4$ and 9. Appendix B gives a series of prediction maps for the full radar grid for RIDW (Figures B.1 - B.4) and stRIDW (Figures B.5 - B.8).

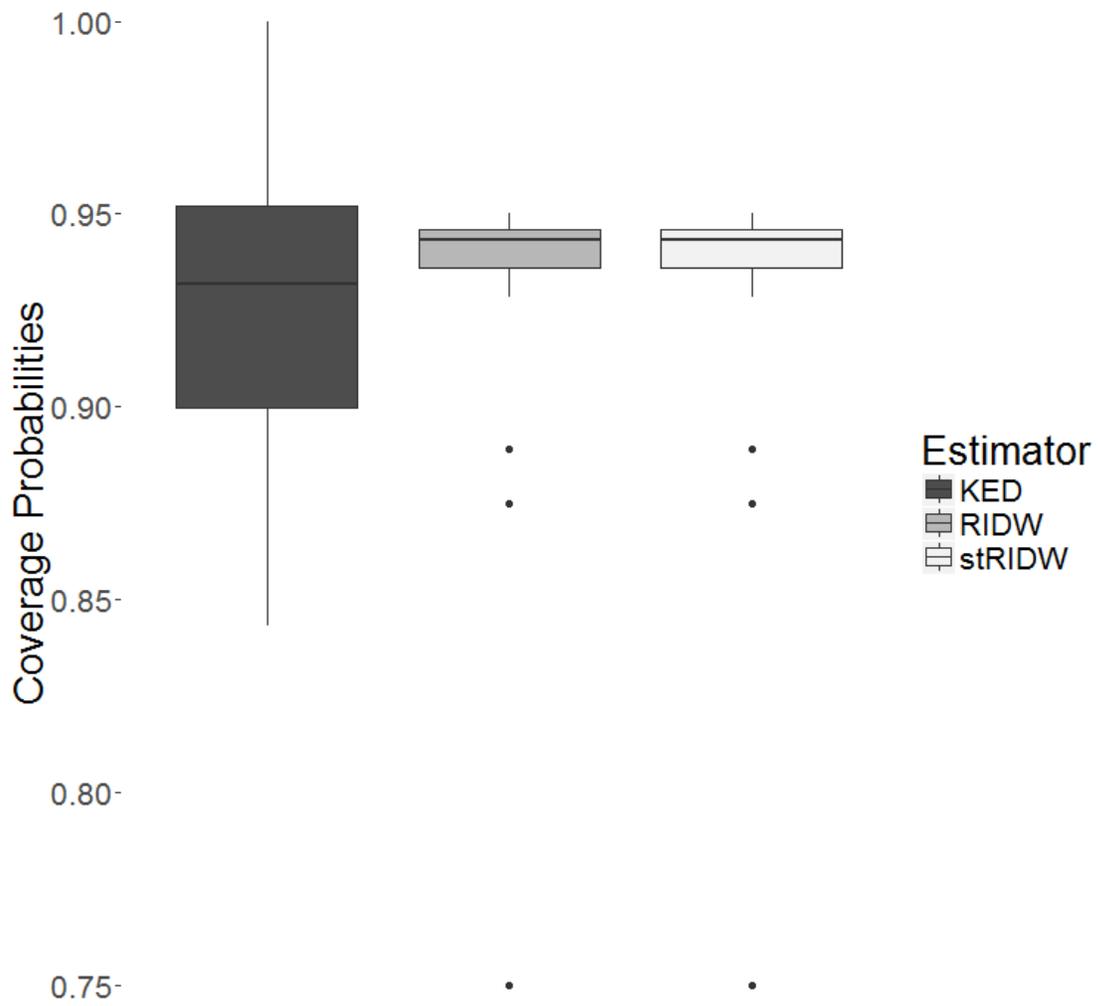


Figure 3.6: Boxplots of CV estimated coverage probabilities by estimator.

Although the estimated coverage probabilities for stRIDW and RIDW were equal for each hour, the lengths of the prediction intervals were different. Both of the median and mean of all prediction interval lengths for stRIDW were smaller than those for RIDW and KED, as seen in Table 3.2. The 25th and 75th percentiles for the interval lengths are given as well. This shows that, not only are the coverage probability properties for stRIDW and RIDW more optimal, the prediction interval estimators are more efficient at allocating uncertainty.

3.4 Discussion

In this paper, we have proposed a spatial prediction method that a) combines global mean modeling with local residual spatial adjustment, b) tunes the necessary weight parameter automatically, c) avoids problematic assumptions often required in common stochastic interpolations, and d) incorporates previous spatial information into current spatial interpolation. The method was applied to the motivating problem, radar based rainfall prediction, and outperformed KED in uncertainty quantification as expected because KED requires error normality for uncertainty quantification, which is not typically valid in rainfall data.

We found that KED often fails to estimate the spatial structure and perform prediction when sample size is too small (< 50), whereas RIDW and stRIDW succeed in estimating the predicted value and its uncertainty. It is of utmost important to perform radar rainfall estimation continuously for the purpose of operational use, which is easily actionable with the automatic RIDW and stRIDW estimators.

Although RIDW and stRIDW outperform KED in uncertainty quantification in terms of coverage probability and prediction interval length, one important issue still remains: the intervals created by KED are stochastic in nature, while those created by RIDW and stRIDW, which are free from problematic distributional assumptions, are heuristic. Heuristic estimates, by definition, rely solely on the observed

Table 3.2: Summary of CV estimated prediction interval lengths for KED, RIDW, and stRIDW.

Estimator	Mean	25%	Median	75%
stRIDW	11.13201	6.507291	9.756801	14.193878
RIDW	11.44977	6.463644	10.123043	14.782826
KED	12.07792	5.959374	10.309028	14.958806

data, and are valid when the observed data gives an accurate representation of reality. This may not be the case when extreme outliers exist, which is typical in rainfall estimation. Further, when small sample sizes are present, the cross-validation based prediction variance is often very large, resulting in overly conservative estimates (Efron, 1983). Hence, it would be promising to study a sophisticated method for quantifying uncertainty for the proposed spatial predictor.

A linear model for mean modeling is only considered in this paper, but it may be too restrictive. We envision flexible mean modeling to enhance global trend fitting using a more general class of models including non-linear models. It is expected to increase the model flexibility and improve prediction.

CHAPTER FOUR

RIDW for Anisotropic Spatial Interpolation

4.1 Introduction

In spatial prediction, an isotropy assumption simplifies the analysis by allowing the spatial dependence to be modeled uniformly in all directions. When necessary, anisotropy, the opposite of isotropy, can be modeled as zonal, geometric, or both. Modeling zonal anisotropy concerns adjusting the range of the spatial dependence fit based on direction, while modeling geometric anisotropy allows the degree of spatial dependence fit to differ depending on direction.

While an isotropy assumption is convenient, modeling anisotropy may be beneficial, and is common in atmospheric science applications. Analysis of measurements affected by wind may be benefited by dropping the isotropy assumption, as wind blown elements may vary less and show greater continuity when more information is borrowed in the direction of the prevailing wind (Houlding, 2000). Measurements of wind speed can also be better modeled with an anisotropic process (Tomczak, 1998). Flood estimation studies that incorporate both radar and rain gauge based rainfall data have been found to be better modeled with anisotropic spatial interpolation processes (Sun et al., 2000).

Kriging predictors are often employed for spatial prediction, but may not be flexible enough to accommodate rainfall data, as observed in Chapter Three. We propose directional regression-based inverse distance weighting (dRIDW), a flexible spatial predictor that incorporates mean modeling and residual interpolation using an anisotropic IDW method.

This chapter is organized as follows: in Section 4.2, we introduce the anisotropic IDW-based predictor, dRIDW. Section 4.3 gives a comparison of dRIDW, RIDW, and an anisotropic kriging predictor. Section 4.4 follows with a conclusion and discussion.

4.2 Methodology

Anisotropic kriging is often used in spatial analyses where an isotropy assumption is unreasonable and linear modeling is used to model the mean, as in the regression case. Anisotropic IDW predictors (Tomczak, 1998) are more flexible than anisotropic kriging predictors in spatial prediction, but do not incorporate mean modeling, which is also desired. In this section, we present the dRIDW predictor to incorporate mean modeling and anisotropic residual spatial interpolation.

4.2.1 Anisotropy in Kriging Estimators

When an isotropy assumption is unreasonable, fitting a single theoretical variogram to all directions of the data may lead to prediction variance misspecification in kriging estimators (Maity and Sherman, 2012). Let θ denote the relative direction between locations i and i' in radians. For some tolerance w , let θ' be the observed relative direction between i and i' such that several directional empirical variograms are created similar to (2.3) through

$$2\hat{\gamma}(h_k, \theta') = \frac{1}{N_{k,\theta'}} \sum_{j=1}^{N_{k,\theta'}} I(\theta') (\mu(y_i) - \mu(y_{i+h_k}))^2,$$

$$I(\theta') = \begin{cases} 1 & \text{if } \theta' \in \theta \pm w, \\ 0 & \text{otherwise,} \end{cases}$$

where $N_{k,\theta'}$ is the number of point pairs in the k th binned distance interval where $\theta' \in \theta \pm w$. Just as with isotropic variogram modeling, when $N_{k,\theta'}$ is small, variogram estimates in bin k for direction θ' may be a poor representation of the spatial process.

For parametric anisotropic variogram fitting, a parameter is added that controls the relative weight between the major axis, or direction of least variability, and minor axis, which is perpendicular to the major axis. Note that, all anisotropic kriging estimators enforce symmetry, such that, for variogram fit $f(\cdot)$ and two locations s_i and s_j , $f(\|s_i - s_j\|) \equiv f(\|s_j - s_i\|)$. The angle $\theta \in (0, \pi)$ describes the angle of the major axis. Parametric anisotropic variograms are fit through the same methods as their isotropic counterparts.

4.2.2 Anisotropic Weight Function for IDW

Although the data may indicate that spatial dependence tends to be stronger in some principal direction, it is not beneficial to ignore spatial dependence in other directions. Let w_i be the weight between location i and new prediction location i' ,

$$w_i = f(m_i, m_{i'}, q_i, q_{i'}, \theta, \gamma)^{-p}, \quad (4.1)$$

where θ is the direction of the dependence chosen, γ represents the magnitude of anisotropy, m is the location along a horizontal axis, q is the location along a vertical axis, and $p > 0$. In the application presented, $p = 2$ is considered. Tomczak (1998) specified

$$f(m_i, m_{i'}, q_i, q_{i'}, \theta, \gamma) = \sqrt{A_{mm} (\Delta m)^2 + A_{mq} \Delta m \Delta q + A_{qq} (\Delta q)^2}, \text{ where}$$

$$A_{mm} = \left(\frac{\cos(\theta)}{\gamma} \right)^2 + (-\sin(\theta))^2,$$

$$A_{mq} = 2 \left(\frac{\cos(\theta) \sin(\theta)}{\gamma^2} - \sin(\theta) \cos(\theta) \right),$$

$$A_{qq} = \left(\frac{\sin(\theta)}{\gamma} \right)^2 + (\cos(\theta))^2,$$

$$\Delta m = m_{i'} - m_i,$$

$$\Delta q = q_{i'} - q_i,$$

for $\gamma \geq 1$ and $\theta \in (0, 2\pi)$ to create (4.1) for an anisotropic IDW predictor. Intuitively speaking, increasing γ “shrinks” the distance in the direction θ , such that values in the direction θ are more heavily weighted than their equidistant counterparts in other directions. At $\gamma = 1$, these weights are equal to those created for isotropic IDW. The vector of directionally adjusted inverse distance weights $\mathbf{v}_{i'}(\theta, \gamma)$ is calculated through (2.2). The anisotropic IDW predictor at a new location i' has the form

$$\hat{y}_{i'} = \mathbf{v}_{i'}(\theta, \gamma)^T \mathbf{y}.$$

4.2.3 Point Estimate

To create the dRIDW point estimate, the same mean function and residual adjustment scheme in (2.4) is used, but the weights for the IDW vector in (2.2) are created through (4.1). Let $\mathbf{e} = \mathbf{y} - \mu(\mathbf{x}; \boldsymbol{\beta})$. The dRIDW point estimate is

$$\hat{y}_{i'}(\mathbf{x}_{i'}) = \mu(\mathbf{x}_{i'}; \boldsymbol{\beta}) + \mathbf{v}_{i'}(\theta, \gamma)^T \mathbf{e}. \quad (4.2)$$

For the purposes of this chapter, $\mu(\mathbf{x}; \boldsymbol{\beta})$ is estimated with simple linear regression.

4.2.4 Uncertainty and Interval Estimate

The local variance estimator of the dRIDW prediction at the i' th location is

$$s_{i'}^2 = \sum_{i=1}^n v_i (e_i - \mathbf{v}_{i'}(\theta, \gamma)^T \mathbf{e})^2, \quad (4.3)$$

where $v_i(\theta, \gamma)$ is the i th element of $\mathbf{v}_{i'}(\theta, \gamma)$.

The variance estimator in (4.3) fails to take the uncertainty in the estimation of $\boldsymbol{\beta}$ into account. Let \mathbf{X} be the $n \times b$ design matrix, where $\mathbf{c}_{i'}(\mathbf{x}) = \mathbf{x}_{i'} - \mathbf{X}^T \mathbf{v}_{i'}(\theta, \gamma)$, and $\mathbf{d}_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{X}^T \mathbf{v}_{i'}(\theta, \gamma)$. To incorporate the uncertainty in the regression

estimates,

$$\begin{aligned}
s_{i'}^2 &= \sum_{i=1}^n v_i(\theta, \gamma)(e_i - \mathbf{v}_{i'}(\theta, \gamma)^T \mathbf{e})^2 \\
&+ \sigma^2 \sum_{i=1}^n v_i(\theta, \gamma) \mathbf{d}_i(\mathbf{x})' (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{d}_i(\mathbf{x}) \\
&+ \sigma^2 \mathbf{c}_{i'}(\mathbf{x})' (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}_{i'}(\mathbf{x}).
\end{aligned} \tag{4.4}$$

When applied, the estimator $\hat{\sigma}^2 = \sum_{i=1}^n e_i^2 / (n - b - 1)$ is used in place of σ^2 .

As distributional assumptions have not been made, we cannot create interval estimates based on known error structures. Instead, we use a cross validation based approach. Let

$$\hat{y}_{(-i)} = \mu(\mathbf{x}_{(-i)}; \hat{\boldsymbol{\beta}}) + \mathbf{v}_i(\theta, \gamma)^T \mathbf{e}_{(-i)}, \tag{4.5}$$

$CV_i = y_i - \hat{y}_{(-i)}$, and s_{-i}^2 be the cross validation realization of (4.4) (see (A.1) for a full formulation). Then, we set κ_α as the upper α quantile of

$$\frac{|CV_i|}{s_{-i}}, \quad i = 1, 2, \dots, n.$$

The dRIDW prediction interval estimate for new location i' is then created through

$$\hat{y}_{i'}(\mathbf{x}_{i'}) \pm \kappa_\alpha s_{i'}.$$

4.2.5 Anisotropy Parameter Selection

In some cases, the angle of dependence is a “known” value, such as the case in estimation of storm cell properties when wind direction is observed. In other cases, these values are unknown, and must also be estimated. In both cases, it is recommended that the weight function is tuned to the strength of the anisotropic dependence. The parameters in the weight function (4.1) are estimated by minimizing

$$\text{MSCV}(\theta, \gamma) = \sum_{i=1}^n (e_i - \hat{e}_i)^2, \tag{4.6}$$

where

$$\hat{e}_i = \sum_{j \neq i} e_j^{(-i)} v_j(\theta, \gamma),$$

$e_j^{(-i)}$ is the residual error resulting from reestimating the regression coefficients without information from location i , and e_i is the residual error resulting from regression at location i . In cases where θ is known, we only minimize (4.6) with respect to γ .

4.3 Application to Rainfall Data

This chapter will focus on the case of hour 20 2012-07-14. The radar rainfall intensity is calculated through (1.1) over the same study area described in Chapter One. The ground level measurements of rainfall intensity are observed at 168 tipping bucket rain gauges in the study area. Figure 4.1 presents a map of locations for the rain gauges active in this data set.

As seen in Figure 4.2, the relationship between radar and rain gauge measurements appears to be linear. This suggests that, for rainfall estimation, the radar data could be used to model the mean rainfall intensity through simple linear regression.

As previously mentioned, uncertainty quantification for kriging predictors requires an error normality assumption. Figure 4.3 gives a histogram of the residual errors arising from mean modeling through simple linear regression. From this histogram, it is apparent that a normality assumption may not be reasonable for this analysis. A Shapiro-Wilk test (Shapiro and Wilk, 1965) for normality yields a p-value < 0.0001 , suggesting that an error normality assumption is inappropriate.

The RIDW predictor assumes isotropy when adjusting for the spatial effect. Figure 4.4 gives directional and omnidirectional variograms, which shows that the semivariance tends to differ depending on direction. Assuming isotropy may yield estimates that use information inefficiently.

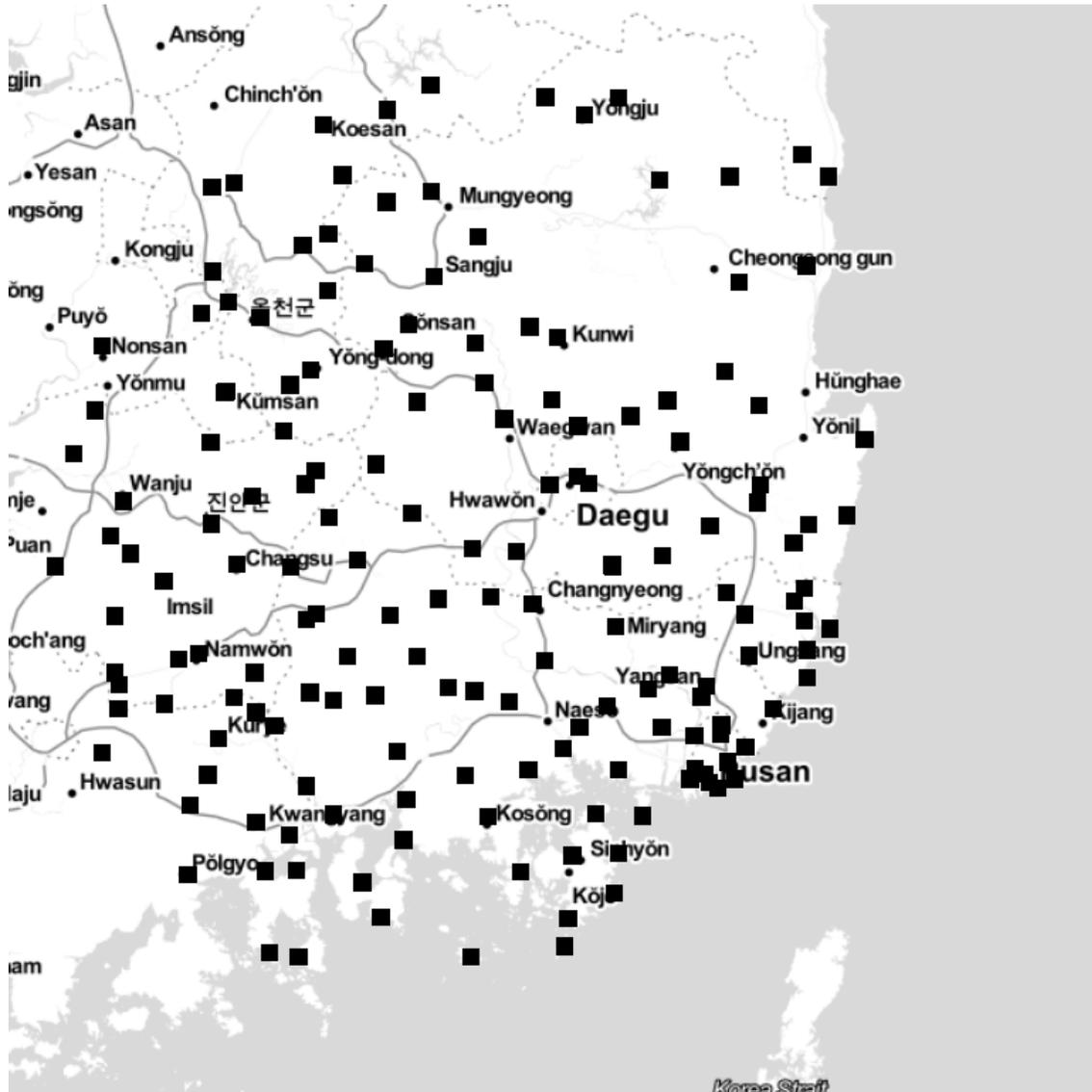


Figure 4.1: Rain gauge locations reporting, hour 20 2012-07-14, created with ggmap (Kahle and Wickham, 2013).

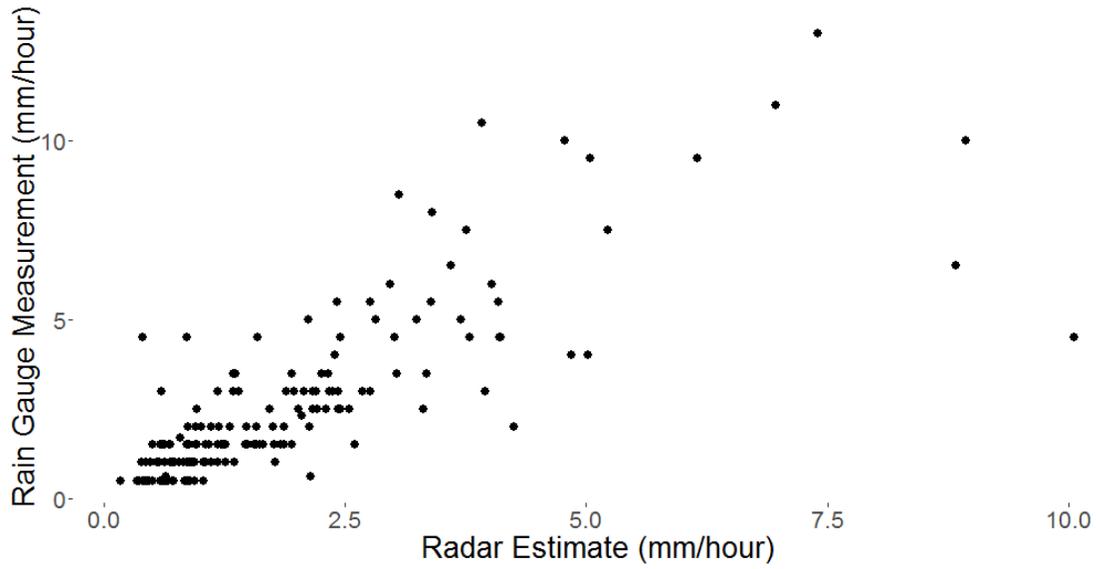


Figure 4.2: Rain gauge measurements by radar estimates.

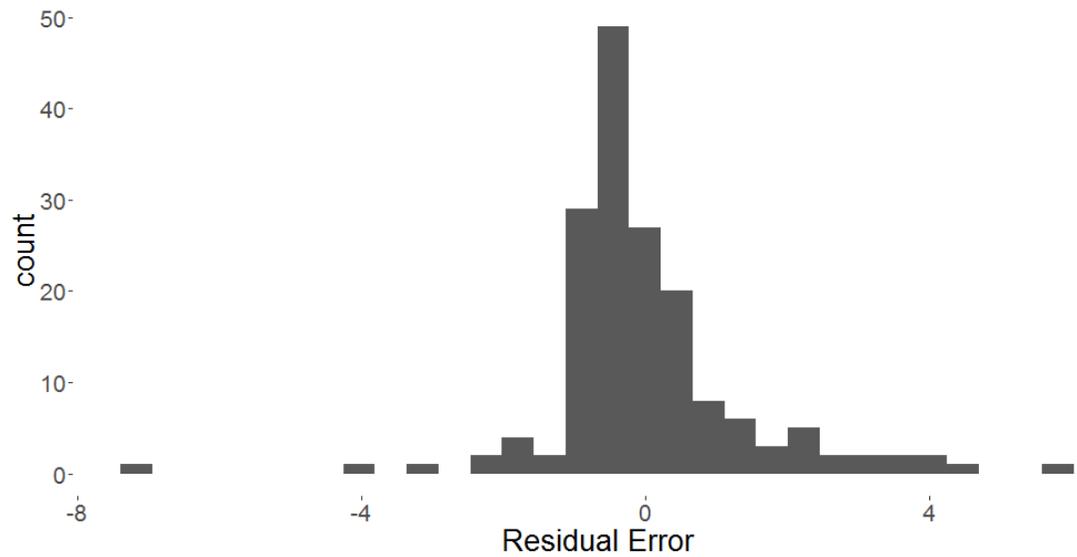


Figure 4.3: Residuals resulting from simple linear regression between the rainfall measurements and radar estimates.

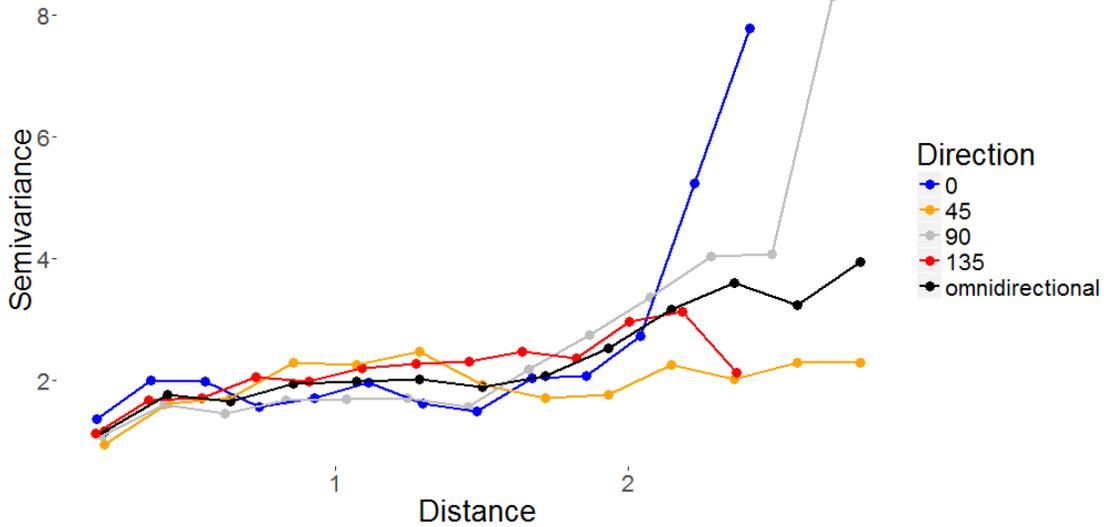


Figure 4.4: Directional empirical variogram.

4.3.1 Results

We evaluate the performance of RIDW, dRIDW, and anisotropic KED using three cross validation based criteria: cross validation estimated squared error (CVSE), 95% prediction interval length (PIL), and estimated coverage probability (ECP). CVSE is calculated by

$$CVSE_i = (y_i - \hat{y}_{(-i)})^2,$$

where $\hat{y}_{(-i)}$ is calculated through (2.8) and (4.5) for the RIDW and dRIDW predictions respectively. The PIL is defined as the difference in the upper and lower bounds of the cross validation 95% prediction intervals. ECP describes the proportion of instances the 95% prediction intervals capture the observed ground level rainfall intensity measurements. The PIL is a more meaningful metric than prediction variance for our heuristic estimator, since values of κ_α can differ by estimator. Because rainfall is a non-negative process, we create point mass confidence intervals with a mass at zero; that is, if the lower or upper prediction interval bound for one of our estimators falls below zero, we set the value for the quantity at zero.

Summary information for $CVSE_i$, including boxplots, the mean, and quartiles are given in Figure 4.5 and Table 4.1. When compared to RIDW and KED, dRIDW produces a similar median $CVSE_i$, but also the largest mean $CVSE_i$ value, which is 25% and 21% larger than the KED and RIDW mean $CVSE_i$ respectively. Figure 4.6 gives a map of $\sqrt{CVSE_i}$ for all three predictors. Edge effects aside, there does not appear to be a distinct spatial pattern in these values, indicating that the three predictors effectively accounted for the spatial effect.

Because the heuristic interval estimate is created to capture $\approx 95\%$ of all values, estimators using this cross validation approach with different weight functions often share an ECP when implemented on the same data. However, they do not necessarily capture values at the same locations, as seen in Figure 4.7. In this case, both RIDW and dRIDW captured 159 of 168 sites for an estimated 94.6% coverage, while KED only captured 154 sites for an estimated 91.7% coverage.

The 95% PILs are summarized in Table 4.2. As we can see, the Q1 and median summary values for the PILs favor dRIDW. When these values are compared at each location, 100% and 65.5% of the PILs created by dRIDW are smaller than their RIDW and KED counterparts respectively. A map comparing the size of the PILs is given in Figure 4.8. We can see a similar pattern for the three predictors, but again, dRIDW is able to borrow information more effectively while retaining the flexibility of IDW-based spatial interpolation.

Table 4.1: Comparison of $CVSE_i$ error mean and quantiles for RIDW, dRIDW, and KED.

Estimator	Mean	25%	Median	75%
RIDW	2.0252	0.0406	0.2703	1.0805
dRIDW	2.4422	0.0694	0.2516	0.8952
KED	1.9551	0.0671	0.2389	0.9355

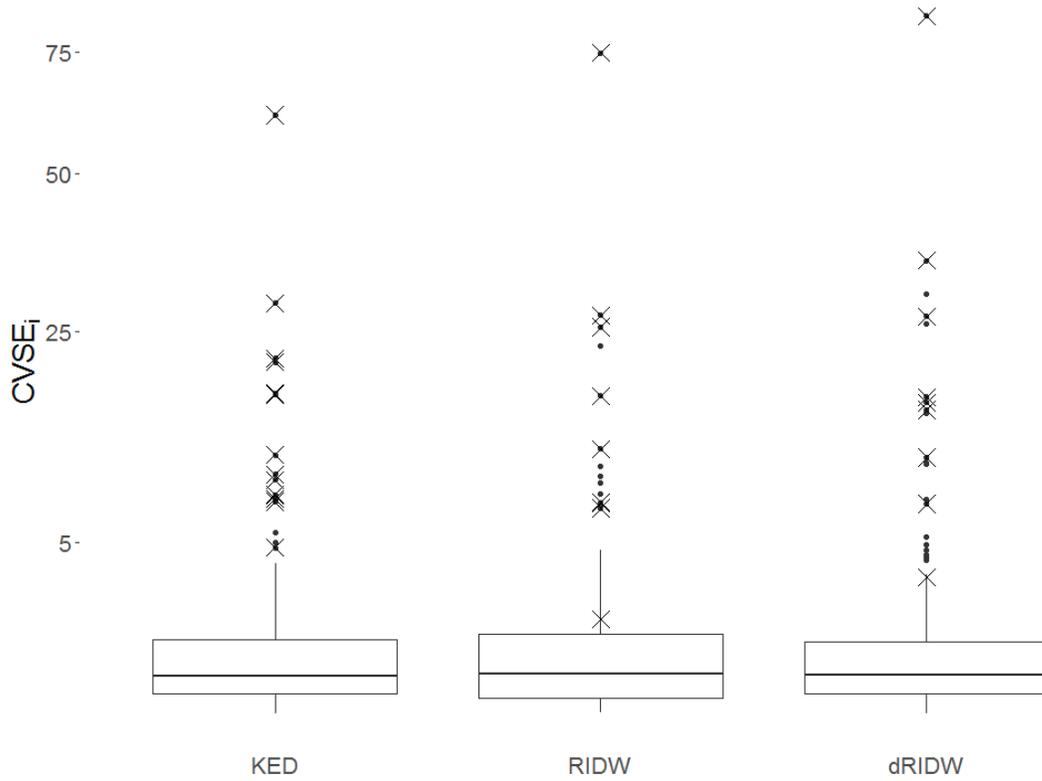


Figure 4.5: Box plots of $CVSE_i$ (on square root scale) plotted by estimator. An “X” indicates that this prediction was not captured in the prediction interval.

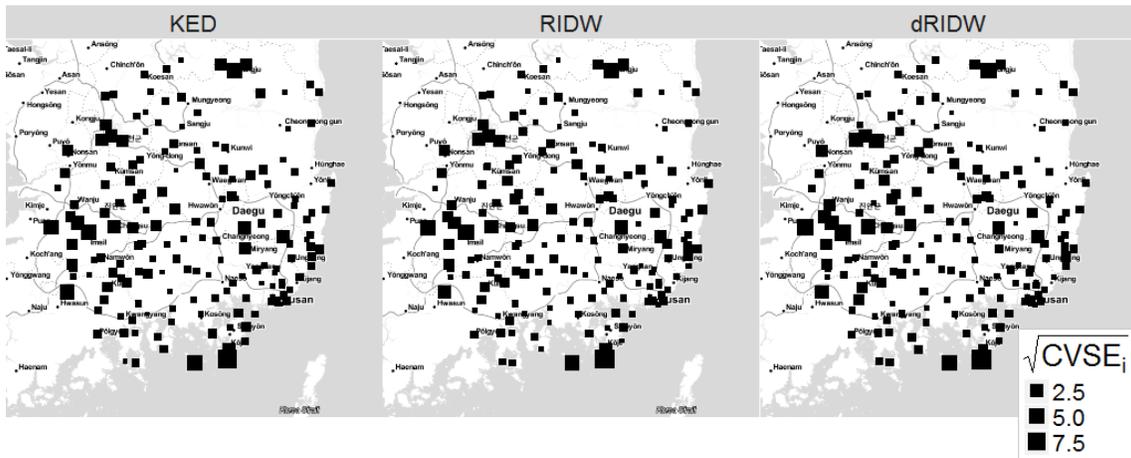


Figure 4.6: $\sqrt{CVSE_i}$ at each location for RIDW, dRIDW, and KED.

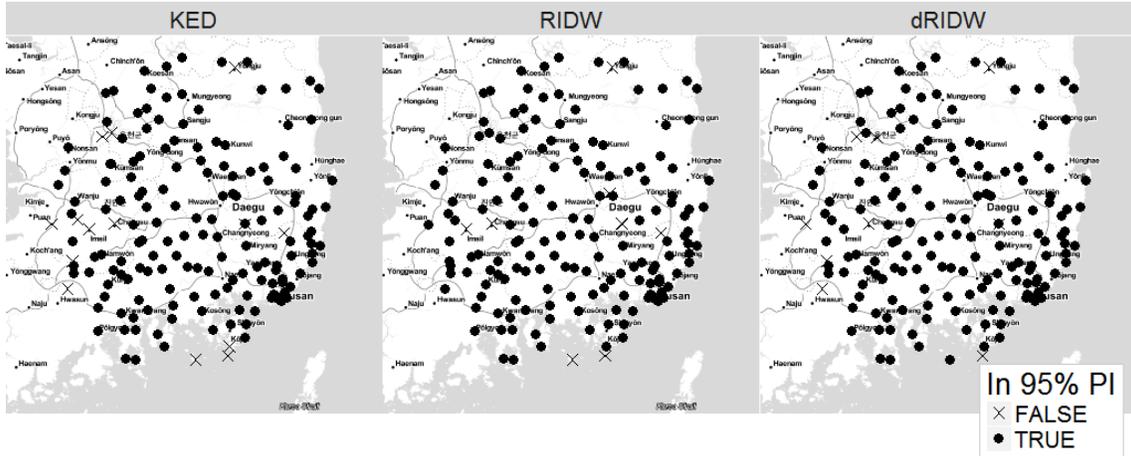


Figure 4.7: Rain gauge measurements captured in prediction intervals, by site.

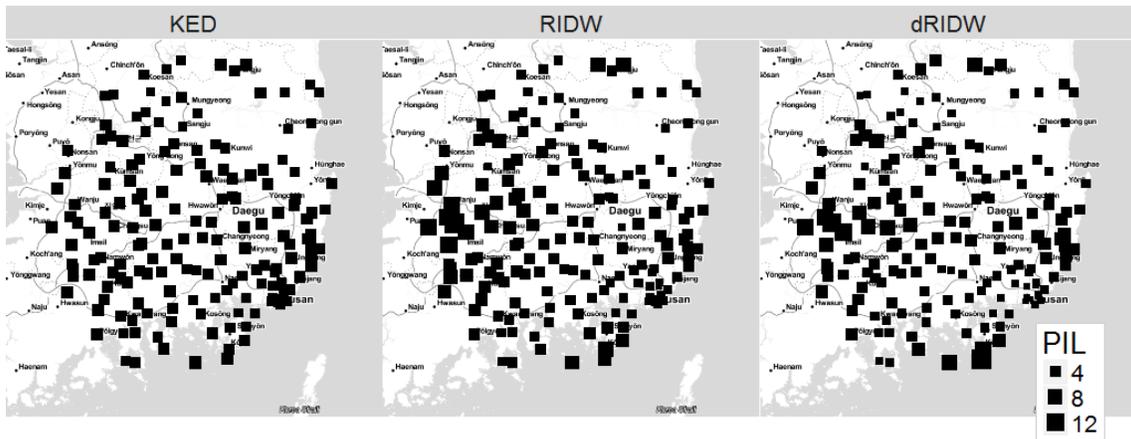


Figure 4.8: 95% PIL for RIDW and dRIDW.

Table 4.2: Comparison of PIL mean and quantiles for RIDW, dRIDW, and KED.

Estimator	Mean	25%	Median	75%
RIDW	4.302	3.075	3.815	4.941
dRIDW	4.013	2.660	3.545	4.702
KED	3.992	3.507	4.143	4.437

4.4 Conclusion

In this chapter, we have proposed a flexible spatial prediction scheme that combines mean modeling through regression and spatial interpolation of the residuals through anisotropic IDW. Using three cross validation based criterion, we compared the predictions and interval estimates resulting from dRIDW to those obtained by RIDW and KED using radar rainfall data. We found that the flexible dRIDW predictor outperforms KED in attaining the targeted coverage probability, while creating tighter prediction interval estimates than RIDW by using an anisotropic IDW scheme.

Similar to the stRIDW predictor proposed in Chapter Three, dRIDW is deterministic rather than stochastic in nature. In part, the dRIDW predictor’s ability to effectively achieve the targeted coverage probability is a function of the cross validation approach to uncertainty quantification. It may be of interest to develop a robust stochastic uncertainty quantification scheme that retains the necessary flexibility.

A simple linear model was considered for mean modeling in this proposal. A more sophisticated model would likely be of great benefit to the problem at hand, although prudence would be warranted to avoid overfitting.

CHAPTER FIVE

Stochastic Uncertainty Attachment for Gridded Deterministic Estimates

5.1 Introduction

Deterministic computer models, such as the radar rainfall prediction algorithm WSR-88D Precipitation Processing Subsystem, give a network wide approach to QPE (Hunter, 1996). Other closely related weather applications for deterministic models are wind forecasting (Tomczak, 1998), isolated storm cell tracking, and tornadic event identification (Johnson et al., 1998). Although deterministic predictions are popular for these applications due to their speed and ability to incorporate a vast amount of information, these algorithms often give no notion of uncertainty about the predictions made. Popular deterministic estimators in statistics commonly employ interval estimation procedures based on cross validation, which can result in overly conservative and larger than necessary interval estimates (Efron, 1983). Rather than rely on inefficient cross validation procedures to estimate uncertainty, we propose a robust Bayesian data fusion model that incorporates deterministic prediction model output and validation site data to produce gridded spatial estimates with stochastic uncertainty.

A general Bayesian approach, named Bayesian synthesis, to map stochastic uncertainty to deterministic model outputs was given by Raferty et al. (1995), and applied to baleen whale population dynamics. Bayesian melding, which alleviated instabilities in the Bayesian synthesis approach, was applied to the same problem later (Poole and Raferty, 2000). These approaches specified and attached error structures to aspects of the model inputs and outputs through systematic creation of likelihood and prior structures, yielding stochastic uncertainty in an otherwise deterministic model. Ignoring spatial correlation can result in poor variance estimates (Cressie,

1993). Ghosh et al. (2012) proposed a stochastic uncertainty attachment structure that accounts for spatial correlation, and allows for the incorporation of external validation data in a nearly fully conjugate model that facilitates efficient Bayesian sampling. Further, this model accounts for measurement error in the validation data. Paci et al. (2015) modified this model specification, but argue that, in situations where the desired predictions are in a lattice or areal spatial structure, a conditional autoregressive model (Besag, 1974), shortened to CAR, is more appropriate. While these specifications account for spatial correlation amongst the predictions, they employ a normal error structure, which may be insufficient. Modeling uncertainty involving rainfall extremes is a prevalent topic in the QPE literature (Ghosh et al., 2011, Wang et al., 2015).

We must also address the change of support problem (Gelfand et al., 2001), shortened to COSP. Our validation site data are geospatial, while our deterministic model output and desired predictions have a lattice structure. To facilitate this, we implement the same downscaling approach used by Paci et al. (2015).

In QPE, the random variable of interest is nonnegative. As most applications of the CAR model involve strictly positive or full real line continuous processes, this is somewhat problematic as zeroes are possible and should not be disallowed in modeling. The Tobit model (Tobin, 1958) can be applied to this task at the cost of a normality assumption on the errors. The parameters in this model are commonly chosen via Newton-Raphson, which struggles when the underlying errors are non-normal (Min and Agresti, 2002). Jørgensen (1987) proposed use of a compound Poisson construction, but parameter estimation is a computationally costly process. In this method, we instead propose to model semi-continuous processes through point mass methods, similar to the approach proposed in Ghosh et al. (2012).

In fully Bayesian linear modeling with the student's t distribution, prior specification for the degrees of freedom parameter ν is nontrivial. In many analyses, ν is

treated as a fixed and known quantity, which is typically unrealistic. As the support of ν exists in the interval $(0, \infty)$, ν may be modeled as a continuous parameter, but could also be restricted to the set of integers (or any other discrete enumeration scheme) in this interval. In discrete and continuous modeling, several authors advocate for setting some upper bound, forcing $\nu < c$ in sampling (Jacquier et al., 2004, Villa and Walker, 2014, Zellner, 1976). These priors preclude the possibility that ν approaches infinity, as in the normal case. For normal hierarchical models that give a marginal student's t distribution, there is a lower bound m , such that for $\nu < m$, the variance term drives to zero, which drives the likelihood to infinity (Berger et al., 2005, Fernandez and Steel, 1999). In specific model structures, m is tractable. Both Anscombe (1967) and Gelman and Hill (2007) seek to avoid issues with the lower bound m by imposing a uniform prior over the interval $(0, 0.5)$ on the inverse of ν . This parameterization is also popular as it is valid in WinBUGS (Speigelhalter et al., 1999), which restricts $\nu > 2$ for the student's t distribution. Although this parameterization is convenient in terms of computation, it is somewhat informative, and tends to give more posterior mass to small values of ν , which can lead to overfitting (Simpson et al., 2014). In an attempt to create a less subjective prior distribution that still fostered posterior propriety, Geweke (1993) specified an exponential prior on ν with a user specified hyperparameter g . However, it can be shown that a poor choice for g has disastrous modeling consequences (Fonseca et al., 2008), and even in large sample situations, the choice of g has an immense influence on the resulting posterior distribution of ν (Simpson et al., 2014). Martins and Rue (2013) and Simpson et al. (2014) advocate for a weakly informative "penalized complexity" prior, which gives more prior mass to areas that represent a simpler model, which, in this case, is a model with a normal error structure. This prior is focused on prediction with the student's t distribution, and will be implemented in this analysis.

This chapter is organized as follows: in Section 5.2, we give our strategy for attaching uncertainty to both the deterministic model and our validation data. In Section 5.3, we explore the data fusion model inputs. In Section 5.4, we compare the results between our robust specification and the model specification from Paci et al. (2015). Section 5.5 follows with a conclusion and discussion.

5.2 An Approach to Uncertainty Quantification for Deterministic Outputs

Let $R(A_i)$ be the deterministic prediction in grid cell A_i , $i = 1, \dots, n_i$. This quantity is interpreted as the average deterministic prediction throughout A_i . Let $V(s_j)$ denote the observed value of the validation data at locations s_j , $j = 1, \dots, n_j$. For A_i where $s_j \in A_i$, or the grid cells that contain a validation site, we can compute a set of *observed* residuals through $R(A_i) - V(s_j)$. Recall that, due to the COSP, $V(s_j)$ is not the “true” value of interest in the grid cell A_i . Additionally, due to the tip-bucket rain gauge measurement error issues presented by González et al. (2015), $V(s_j)$ is not the “true” value of the rainfall process exactly at location s_j . Further, these observed residuals do not exist for grid cells that do not contain a validation site, while we need to quantify uncertainty in all grid cells of interest.

Now, let the “true” average value in grid cell A_i be $\tilde{R}(A_i)$. To attach uncertainty in all grid cells of interest, we focus on the *realized* residuals (Zellner, 1975), $\varepsilon_R(A_i) = R(A_i) - \tilde{R}(A_i)$. If $[\cdot]$ denotes a probability distribution, inference is focused on modeling the posterior distribution $[\varepsilon_R(A_i)|\text{Data}]$ within the Bayesian framework.

5.2.1 Model Specification

Using the formulation from (Paci et al., 2015), the “true” average value $\tilde{R}(A_i)$ in grid cell A_i is

$$R(A_i) = \tilde{R}(A_i) + \varepsilon_R(A_i), \text{ where} \tag{5.1}$$

$$\varepsilon_R(A_i) \sim \text{Normal}(0, \sigma_R^2(A_i)).$$

This is an error-in-variables measurement error model, which has been chosen for its simplicity in this application. A measurement error model that incorporates bias, such as $R(A_i) = \beta_{R0} + \beta_{R1}\tilde{R}(A_i) + \varepsilon_R(A_i)$ could be employed. In cases that involve modeling of extra-normal errors, like those often observed in weather events, the normality assumption on the errors in (5.1) may prove problematic. Instead, we propose modeling these errors with a t distribution,

$$\varepsilon_R(A_i) \sim T(\nu_R, 0, \sigma_R^2(A_i)). \quad (5.2)$$

We can also account for error heteroskedasticity based on the magnitude of $\tilde{R}(A_i)$ in (5.1) and (5.2). From Ghosh et al. (2012), we specify

$$\sigma_R^2(A_i) \sim \text{Lognormal}(\alpha_0 + \alpha_1\tilde{R}(A_i), 1) \quad (5.3)$$

for the variance term in both (5.1) and (5.2). In order to enable a Gibbs/Metropolis sampling algorithm, the t distribution in (5.2) is

$$\begin{aligned} \varepsilon_R(A_i) &\sim \text{Normal}(0, \omega_R^2(A_i)), \\ \frac{1}{\omega_R^2(A_i)} &\sim \text{Gamma}\left(\frac{\nu_R}{2}, \frac{\nu_R \sigma_R^2(A_i)}{2}\right), \\ \nu_R &\sim \pi(\nu_R), \end{aligned}$$

where $\pi(\nu_R)$ is a prior density for the degrees of freedom parameter ν_R .

Given the areal nature of the prediction grid, a CAR model is appropriate to model the relationship between the $\tilde{R}(A_i)$. Let $i' \sim i$ designate that i' and i are neighbors in a rook neighborhood structure, and let the number of neighbors to grid cell A_i be w_i . Similar to Paci et al. (2015), we implement this as an intrinsic conditional autoregressive model (Besag et al., 1991), shortened to ICAR, where,

$$\begin{aligned} \tilde{R}(A_i) | \tilde{R}(A_{i'}) &\sim \text{Normal}\left(\sum_{i' \sim i} \frac{\tilde{R}(A_{i'})}{w_i}, \frac{\tau^2}{w_i}\right), \\ \tau^2 &\sim \text{Inverse Gamma}(a, b). \end{aligned} \quad (5.4)$$

In an investigation into several scaled hyperpriors for variance terms in intrinsic Gaussian Markov random fields, Sørbye and Rue (2014) found no difference between using scaled and unscaled diffuse gamma distributions as a hyperprior on the precision term for an ICAR modeled spatial effect, so we set $a = 2$ and $b = 1$.

Recall that $V(s_j)$ denotes the observed value of the validation data at locations $s_j, j = 1, \dots, n_j$. The measurement error for the validation data is modeled through

$$\begin{aligned} V(s_j) &= \tilde{V}(s_j) + \varepsilon_v(s_j), \text{ where} & (5.5) \\ \varepsilon_v &\sim \text{Normal}(\mathbf{0}, \sigma_v^2 H(\phi_v)), \\ \sigma_v^2 &\sim \text{Inverse Gamma}(\alpha_\sigma, \beta_\sigma), \\ H(\phi_v)_{jj'} &= \exp(-\phi_v \|s_j - s_{j'}\|), \end{aligned}$$

where $\|s_j - s_{j'}\|$ is the Euclidean distance between validation sites j and j' . We set

$$\frac{\beta_\sigma}{\alpha_\sigma - 1} = \frac{MSE}{2} \text{ and } \frac{\beta_\sigma^2}{(\alpha_\sigma - 1)^2(\alpha_\sigma - 2)} = 100$$

to center the diffuse prior for σ_v^2 at half of the MSE arising from simple linear regression of the $V(s_j)$ on the corresponding $R(A_i)$. As before, a normal distribution may be insufficient to properly model the errors. Instead, we propose

$$\begin{aligned} \varepsilon_v &\sim \text{T}(\nu_v, \mathbf{0}, \sigma_v^2 H(\phi_v)), & (5.6) \\ \sigma_v^2 &\propto \frac{1}{\sigma_v^2}, 0 < \sigma_v^2 < \infty, \\ H(\phi_v)_{jj'} &= \exp(-\phi_v \|s_j - s_{j'}\|) \end{aligned}$$

Similar to Paci et al. (2015) and Ghosh et al. (2011), ϕ_v is set at 60% of $\max \|s_j - s_{j'}\|$ for (5.5) and (5.6). Both of (5.5) and (5.6) force a spatial relationship on the measurement error for the validation data. Because the covariance matrix in (5.6) is the result of multiplying a scalar parameter by a matrix, we impose a multivariate

t distribution by specifying

$$\begin{aligned}\varepsilon_v &\sim \text{Normal}(\mathbf{0}, \omega_v^2 H(\phi_v)), \\ \frac{1}{\omega_v^2} &\sim \text{Gamma}\left(\frac{\nu_v}{2}, \frac{\nu_v \sigma_v^2}{2}\right), \\ \nu_v &\sim \pi(\nu_v),\end{aligned}$$

where $\pi(\nu_v)$ is a prior density for the degrees of freedom parameter ν_v

Because the validation data are geospatial and the $\tilde{R}(A_i)$ are areal, we need to specify error to account for the COSP between point measurement and grid cell estimate. For $s_j \in A_i$, we set

$$\begin{aligned}\tilde{V}(s_j) &= \tilde{R}(A_i) + \varepsilon_{\tilde{v}}(s_j), \text{ where} & (5.7) \\ \varepsilon_{\tilde{v}}(s_j) &\stackrel{iid}{\sim} \text{Normal}(0, \sigma_{\tilde{v}}^2), \\ \sigma_{\tilde{v}}^2 &\sim \text{Inverse Gamma}(\alpha_\sigma, \beta_\sigma).\end{aligned}$$

The specifications above can be thought of as components in a larger model, where we use either a normal (5.1) or robust (5.2) deterministic error proposal, and either a normal (5.5) or robust (5.6) validation site error proposal. These separate error specifications can be combined to create four distinct modeling schemes: NN, where the errors on the validation set and deterministic predictions are both normal, NT, where the validation set errors are normal and the errors on the deterministic predictions have a t distribution, TN, where the validation set errors have a t distribution and the errors on the deterministic prediction are normal, and TT, where both the validation set errors and errors on the deterministic predictions have a t distribution.

5.2.2 Modeling Semi-continuous Processes

Let the $\tilde{R}(A_i)$ be a representation of a non-negative semi-continuous process, with a point mass at 0. After all MCMC draws are complete, for any $\tilde{R}(A_i) < 0$,

we set $\tilde{R}(A_i) = 0$. As a result, all inferences performed on the uncertainty terms of interest will be achieved through analysis of the posterior samples drawn.

5.2.3 Prior Parameterization for the Degrees of Freedom Parameter

Martins and Rue (2013) and Simpson et al. (2014) implement a “penalized complexity” prior for the degrees of freedom parameter ν to discourage overfitting by giving more prior mass to values closer to the “base model,” which is the normal error structure. Using the Kullback-Liebler distance (Kullback and Liebler, 1951), shortened to KLD, to represent the degree of similarity between a t and normal distribution, more prior mass is allocated to areas where the KLD is smaller. The resulting prior is

$$\xi = \frac{1}{\nu},$$

$$\pi(\xi) = \lambda \exp[-\lambda d(\xi)] \left| \frac{\delta d(\xi)}{\delta \xi} \right|, \quad (5.8)$$

where $d(\xi) = d(1/\nu) = \sqrt{(2KLD(p(1/\nu), q))}$ for a standard unit student t density with ν degrees of freedom $p(\nu)$ and standard normal density q , $\left| \frac{\delta d(\xi)}{\delta \xi} \right|$ is the Jacobian of the transformation, and $\lambda = -\log(\alpha)/d(1/U)$, where the user specifies some prior belief regarding $\text{Prob}(\nu < U) = \alpha$. Although the specification of λ involves subjectivity, Simpson et al. (2014) show that inference is robust to poor choices for U and α , even for moderate sample sizes. For the purposes of this analysis, we set $U = 10$ and $\alpha = 0.5$.

The KLD, as given by van Zyl (2015), between $p(\nu)$ and q is

$$KLD(p(\nu), q) = \log \left(\frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu + 1)\sqrt{\nu}} \right) - \left(\frac{\nu + 1}{2} \right) \left(\psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) \right) + \frac{1}{2} \left(\log \left(\frac{\nu}{\nu - 2} \right) + 1 \right)$$

for $\nu > 2$, where $\psi(\cdot)$ is the digamma function. (5.8) is a proper prior after normalization by a constant and has support for $2 < \nu < \infty$.

5.3 Data Analysis

The data fit are arranged in grid cells $12 \text{ km} \times 12 \text{ km}$ in size across the Mount Bisl region centered near Daegu in South Korea. In the data set of interest, we observe $n_j = 168$ rain gauge measurements, and deterministic predictions in $n_i = 726$ distinct grid cells. The $R(A_i)$ are created through the deterministic RIDW predictor given in (2.4). A map of the validation sites and deterministic RIDW predictions is given in Figure 5.1.

The heteroscedastic error specification in (5.3) accounts for increasing error relative to magnitude of $\tilde{R}(A_i)$. Figure 5.2 is a plot of the residual error between the validation data and deterministic predictions for all hours of 2012-07-06. The magnitude of the error resulting from estimating the validation data with the deterministic predictions clearly increases as the deterministic prediction increases. Although there appears to be a trend indicating that error slightly decreases overall as $R(A_i)$ increases, this pattern does not appear to be prevalent in the hour 12 data.

A histogram of the residual error arising from predicting the $V(s_j)$ with the deterministic $R(A_i)$ is given in Figure 5.3. The errors appear to be symmetric about zero, but are clearly leptokurtic and non-normal.

5.3.1 Results

All four models detailed above were fit with Gibbs/Metropolis sampling using the C++ programming language with the GSL (Gallasi et al., 2015) and Eigen (Guennebaud et al., 2010) libraries. An overview of the full conditional distributions is given in Appendix C. A burnin of 10000 was followed by 100000 draws from the posterior, which was thinned to every tenth iteration such that a total of 10000 posterior samples remained. Posterior parameter summaries are given in Table 5.1. The TN and TT models yield smaller values of σ_v^2 and ν_v , suggesting that these models attribute much of the variability due to validation measurement error, given

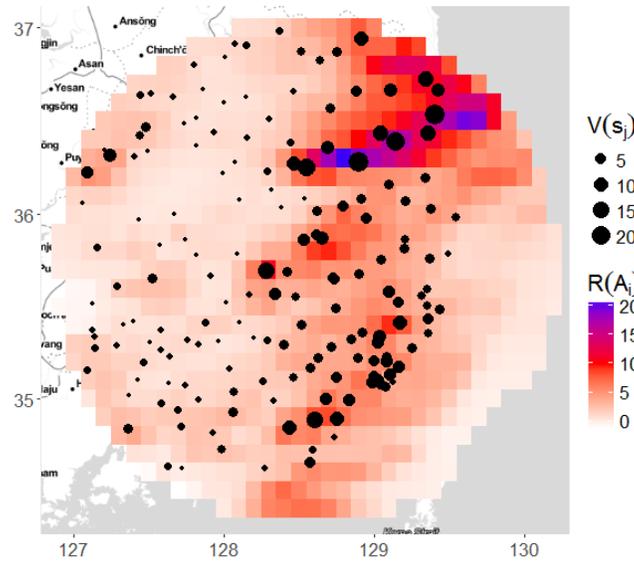


Figure 5.1: Validation data (black points) and RIDW grid cell rainfall predictions across study area, hour 12 2012-07-06.

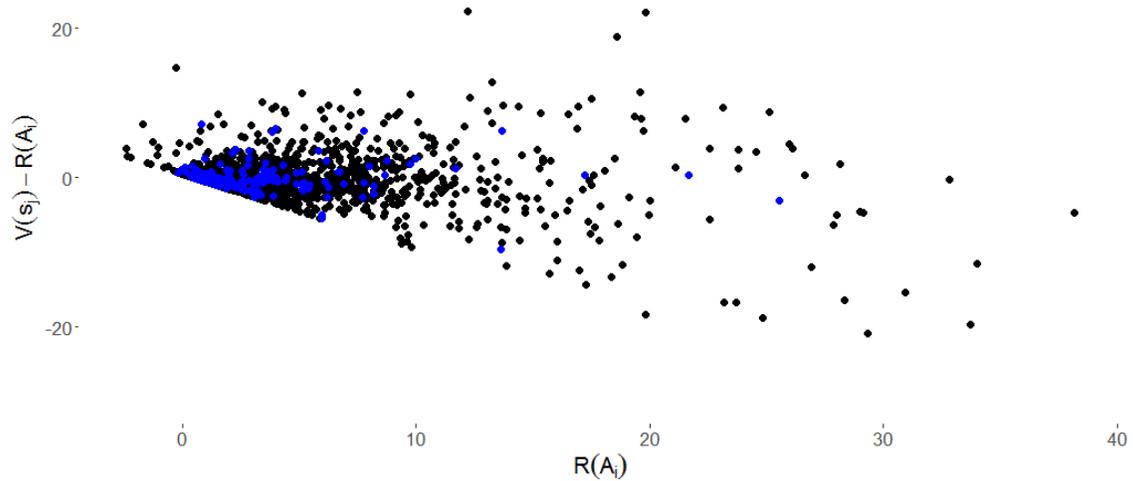


Figure 5.2: Residual error resulting from deterministic prediction of rainfall in grid cells with validation sites, 2012-07-06. Data from hour 12 is highlighted in blue.

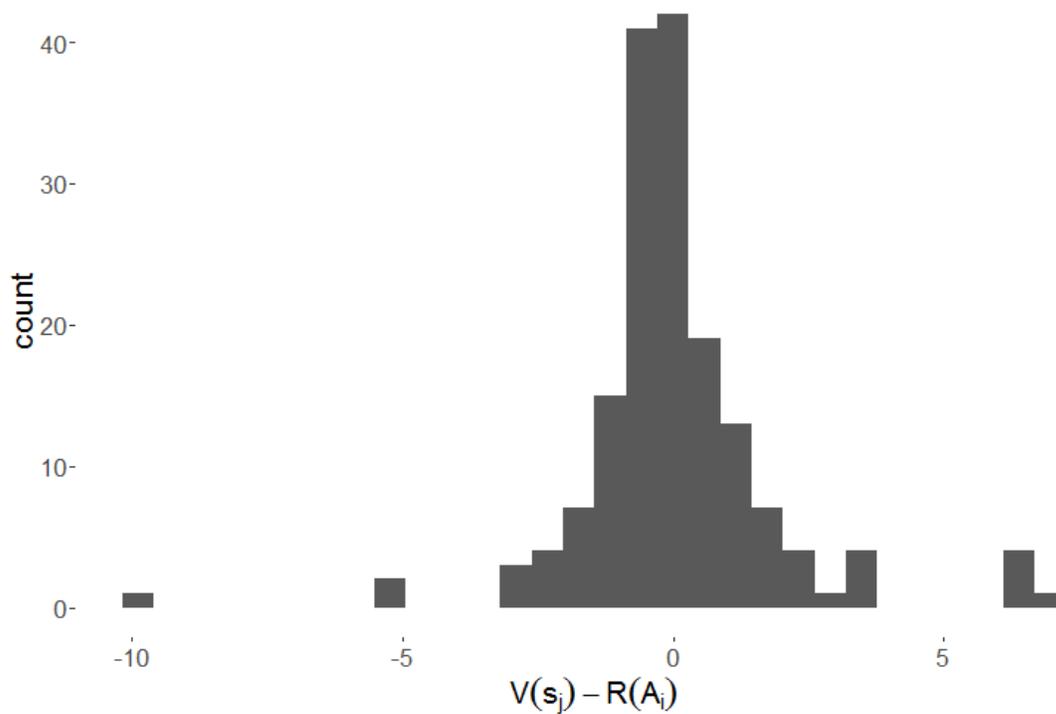


Figure 5.3: Histogram of errors resulting from deterministic prediction of rainfall in grid cells with validation sites, hour 12 2012-07-06.

by $V_{s_j} - \tilde{V}_{s_j}$, to the tails of the t distribution. Using the posterior means to represent the \tilde{V}_{s_j} , Figure 5.4 presents boxplots by model for $V_{s_j} - \tilde{V}_{s_j}$.

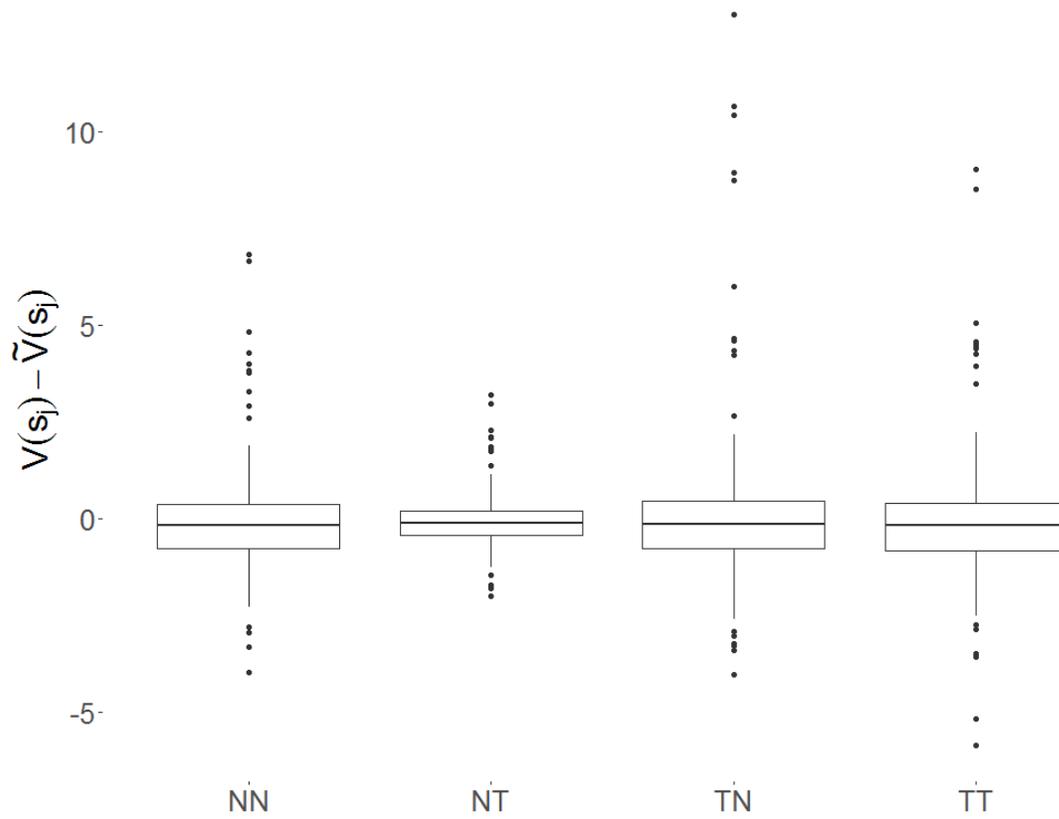


Figure 5.4: Boxplots of the validation site error $V(s_j) - \tilde{V}(s_j)$.

Table 5.1: Posterior parameter mean and 95% credible intervals, by model.

Parameter	NN	NT	TN	TT
α_0	0.0654 (-0.0652, 0.1938)	2.9992 (2.2297, 3.9609)	-2.9341 (-3.5378, -2.3494)	3.0107 (2.2225, 4.0537)
α_1	-0.0079 (-0.0364, 0.0203)	0.0779 (-0.0799, 0.2768)	0.4568 (0.3554, 0.5635)	0.1098 (-0.0673, 0.3769)
σ_v^2	16.7570 (10.5411, 24.2628)	5.0820 (0.0982, 11.1087)	0.0001 (<0.0001, 0.0004)	0.0001 (<0.0001, 0.0003)
σ_v^2	0.9993 (0.5825, 1.5586)	1.6390 (1.1037, 2.2908)	3.0407 (1.5048, 5.74685)	7.7984 (4.0493, 14.4046)
τ^2	15.1432 (12.9884, 17.5351)	12.7161 (11.3377, 14.2655)	6.5013 (5.4412, 7.7591)	12.6673 (11.3113, 14.2091)
ν_R	-	5.0556 (2.1014, 9.5899)	-	5.2310 (2.1222, 9.5651)
ν_v	-	-	3.6074 (2.0373, 7.5573)	5.3567 (2.1461, 9.6140)

The fits are assessed based on the balanced loss criterion (Gelfand and Ghosh, 1998), given by

$$\frac{1}{n_i} \sum_{i=1}^{n_i} \text{var} [\varepsilon_r(A_i)] + \frac{c}{n_j} \sum_{j=1}^{n_j} E \left[\left(\tilde{R}(A_i) - V(s_j) \right)^2 \right], \quad (5.9)$$

where c weights the relative regret of the two losses. Because the choice of c is subjective, we use $c = 1$ to describe a total for (5.9), but also report the components as

$$(5.9 - P) = \frac{1}{n_i} \sum_{i=1}^{n_i} \text{var} [\varepsilon_r(A_i)], \text{ and}$$

$$(5.9 - G) = \frac{1}{n_j} \sum_{j=1}^{n_j} E \left[\left(\tilde{R}(A_i) - V(s_j) \right)^2 \right].$$

Additionally, in the point mass interval estimation scheme used, using the mean of $\tilde{R}(A_i)$ may be inappropriate to compute (5.9-G). As such, the median is used for the typical value for $\tilde{R}(A_i)$. The mean 95% prediction interval length, as measured by the average distance between the 2.5th and 97.5th quantiles of the posterior $\tilde{R}(A_i)$ draws, is included as well as it lends a meaningful notion of uncertainty in the point mass interval application at hand. The values for (5.9-P), (5.9-G), and the average 95% prediction interval lengths are given in Table 5.2.

Table 5.2: Comparison of (5.9)

Model	(5.9-P)	+	(5.9-G)	=	(5.9)	Mean 95% PI Length
NN	0.7903	+	2.3414	=	3.1317	3.4245
NT	0.1066	+	2.1568	=	2.2634	1.3214
TN	0.3585	+	4.7520	=	5.1106	2.0926
TT	0.1004	+	2.1985	=	2.2988	1.2780

As we can see, the NT and TT models produce similar values for (5.9-G), with the NN model producing a slightly larger value. The NT and TT models also have similar (5.9-P) terms, while the TT model produced marginally smaller prediction

interval lengths on average. Strictly speaking, there is no value of the relative regret constant c in (5.9) that would indicate the NN or TN models outperformed the NT or TT models. Appendix D gives a series of maps (created with `ggmap` (Kahle and Wickham, 2013)) that show model predictions (Figure D.1), prediction interval lengths (Figure D.2), $\left| R(A_i) - \tilde{R}(A_i) \right|$ (Figure D.3), and proportion of $\tilde{R}(A_i) = 0$ in the point mass interval scheme (Figure D.4).

5.4 Conclusion

In this chapter, we have proposed a robust hierarchical data fusion model, and implemented it with radar rainfall data and the deterministic RIDW predictor. Using composition sampling for posterior inference, we showed that two of the robust models, the NT and TT model, outperform the strictly normal model in both the standard balanced loss criterion (5.9) and produced smaller average 95% prediction interval lengths for the radar-rainfall data in question.

Given that most weather related data are spatio-temporal in nature, extending the models given above to incorporate a time element could be of great interest. In many radar-rainfall systems, radar and gauge data are updated at different time intervals, which complicates the model. Previous time information could be used to bolster error prediction in (5.2) and (5.6). If we assume separability of the temporal and spatial effects, (5.4) could be rewritten as a latent model, and extended to include a temporal effect.

In this analysis, RIDW prediction was used as the deterministic model output. Although the literature is saturated with suggestions to calibrate radar with rainfall amounts for greater accuracy (Song et al., 2015), (Sun et al., 2000), (Sempere-Torres et al., 1999), it may be of interest to use the radar rainfall intensity rate from (1.1) as our deterministic model output, which could allow for faster uncertainty quantification, especially in areas where the radar information is already well calibrated.

CHAPTER SIX

Conclusion

In this dissertation, we considered several spatially adjusted modeling schemes and applied them to radar based rainfall prediction. Chapters Three and Four focused on flexible deterministic prediction combining mean modeling and spatial interpolation, while Chapter Five gave a fully Bayesian data fusion model that could be used in tandem with the deterministic predictions given in Chapters Three and Four.

In Chapter Three, we introduced stRIDW, which incorporates mean modeling and spatial residual adjustment to create deterministic spatial predictions. Specifically, stRIDW incorporates current and previous time information into the spatial adjustment. The stRIDW estimator automatically tunes the parameter necessary for controlling the influence of the previous time aspect of the spatial adjustment, and creates ad hoc interval estimates based on cross validation. We compared stRIDW to KED, and showed that stRIDW gave more desirable estimated coverage probabilities in interval estimation than KED over 24 hours of radar rainfall data. This is due to the lack of distributional assumptions necessary for stRIDW, which results in more flexible interval estimates for extreme data. We compared prediction interval length for stRIDW, RIDW, and KED, and showed that, by incorporating previous time information, stRIDW creates smaller prediction interval estimates than both RIDW and KED.

In Chapter Four, we introduced dRIDW, which also provides mean modeling and spatial residual adjustment, but allows for anisotropic spatial interpolation. The dRIDW predictor weights the spatial adjustment aspect more heavily in the direction θ by some magnitude parameter γ . This predictor also tunes θ and γ

automatically through simulated annealing, or, in the case where θ is known, γ can be tuned through most bounded optimization methods. We compared dRIDW to KED, and showed that dRIDW was able to attain the desired level of estimated coverage probability, again due to the lack of distributional assumptions necessary for interval estimation. When dRIDW and RIDW were compared, the mean and median prediction interval lengths for dRIDW were smaller than its RIDW counterpart.

In Chapter Five, we gave a fully Bayesian robust data fusion scheme for deterministic spatial predictions and validation data. This model incorporates a grid of deterministic predictions and a geospatial set of validation measurements in a data fusion approach that accounts for errors in both data sources. In particular, we introduce three modeling schemes that allow for departures from normality in the deterministic error, the validation measurement error, or both. This work is compared to the model presented by Paci et al. (2015), which assumes error normality from both the deterministic predictions and validation measurements. We show that two of our robust models outperform the model given in Paci et al. (2015) for any choice of the constant c in (5.9).

Given that storm systems travel in a specific direction over time, it may be interesting to combine the attributes in the estimators in Chapters Three and Four to create a spatial adjustment scheme that borrows information more heavily from the same part of the storm cell over the course of a few hours. This may create a more efficient spatial adjustment based prediction. The models presented in Chapter Five could be extended to the spatio-temporal case, which could use a state space representation for many of the sources of error deemed of interest.

APPENDICES

APPENDIX A

Cross Validation through Variance Estimators for RIDW, stRIDW, and dRIDW

For RIDW and dRIDW, recomputing (2.6) in the case of leave-one-out cross validation for location i gives

$$\begin{aligned}
s_{-i}^2 &= \sum_{j \neq i} v_j (e_j - \mathbf{v}_i^T \mathbf{e}_{(-i)})^2 \\
&+ \hat{\sigma}^2 \sum_{j \neq i} v_j \mathbf{d}_j(\mathbf{x})' (\mathbf{X}_{(-i)}^T \mathbf{X}_{(-i)})^{-1} \mathbf{d}_j(\mathbf{x}) \\
&+ \hat{\sigma}^2 \mathbf{c}_i(\mathbf{x})' (\mathbf{X}_{(-i)}^T \mathbf{X}_{(-i)})^{-1} \mathbf{c}_i(\mathbf{x}), \tag{A.1}
\end{aligned}$$

where $\mathbf{X}_{(-i)}$ is the covariate matrix with the i th row removed, $\mathbf{c}_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{X}_{(-i)}^T \mathbf{v}_i$, $\mathbf{d}_j(\mathbf{x}) = \mathbf{x}_j - \mathbf{X}_{(-i)}^T \mathbf{v}_i$, and $\hat{\sigma}^2 = \mathbf{e}_{(-i)}^T \mathbf{e}_{(-i)} / ((n-1) - 2)$.

For stRIDW, recomputing (3.3) in the case of leave-one-out cross validation at location i for time t is

$$\begin{aligned}
s_{(-i)}^2 &= \sum_{j \neq i}^{n_t} v_{j,t} (e_{j,t} - \sum_{j \neq i}^{n_t} v_{j,t} e_{j,t})^2 \\
&+ \hat{\sigma}_{t,(-i)}^2 \sum_{j \neq i}^{n_t} v_{j,t} \mathbf{d}_j(\mathbf{x}_{(-i),t}, t)' (\mathbf{X}_{t,(-i)}^T \mathbf{X}_{t,(-i)})^{-1} \mathbf{d}_j(\mathbf{x}_{(-i),t}, t) \\
&+ \hat{\sigma}_{t,(-i)}^2 \mathbf{c}_i(\mathbf{x}_{(-i)})' (\mathbf{X}_{t,(-i)}^T \mathbf{X}_{t,(-i)})^{-1} \mathbf{c}_i(\mathbf{x}_{(-i)}) \\
&+ \sum_{k \in W, k \neq t} \sum_{i=1}^{n_k} \sigma_k^2 v_{i,k} \mathbf{d}_i(\mathbf{x}_{i,k}, k)' (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{d}_i(\mathbf{x}_{i,k}, k) \\
&+ \sum_{k \in W, k \neq t} \sigma_k^2 \mathbf{c}_i(\mathbf{x})' (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{c}_i(\mathbf{x}), \tag{A.2}
\end{aligned}$$

where $\mathbf{X}_{k,(-i)}$ is the covariate matrix at time k with the i th row removed.

APPENDIX B

Hourly Prediction Maps for RIDW and stRIDW

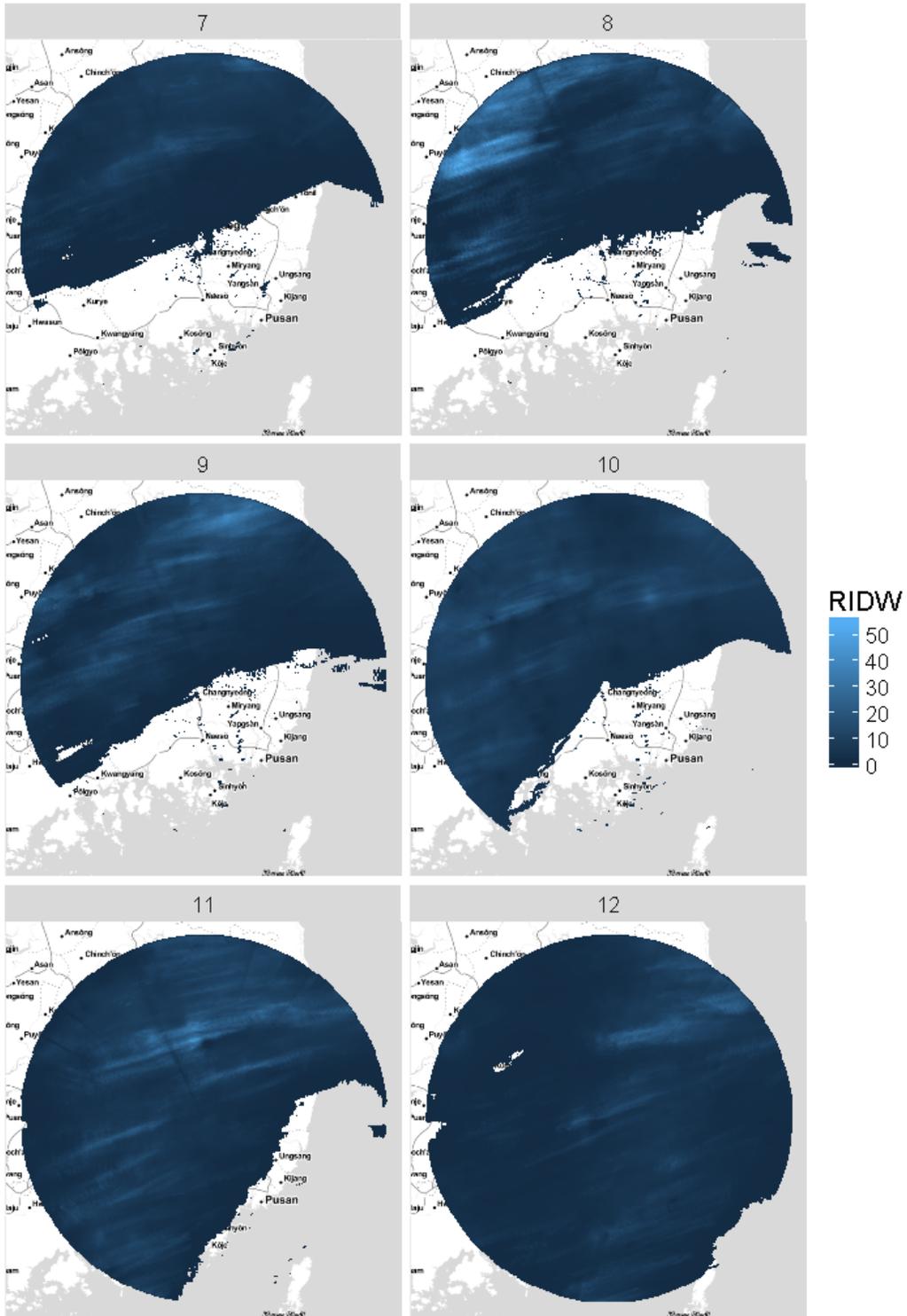


Figure B.2: RIDW predictions, hours 7-12.

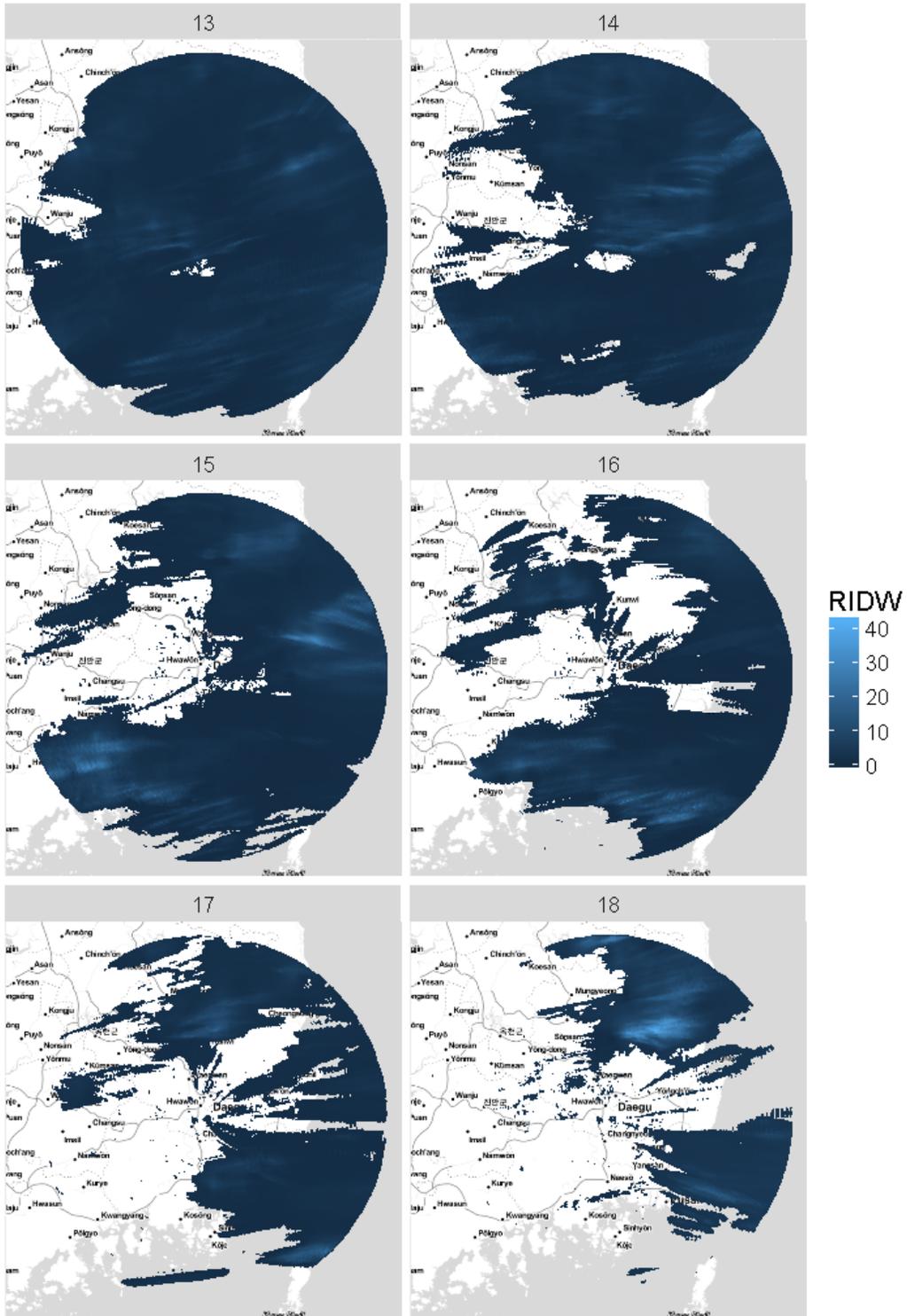


Figure B.3: RIDW predictions, hours 13-18.

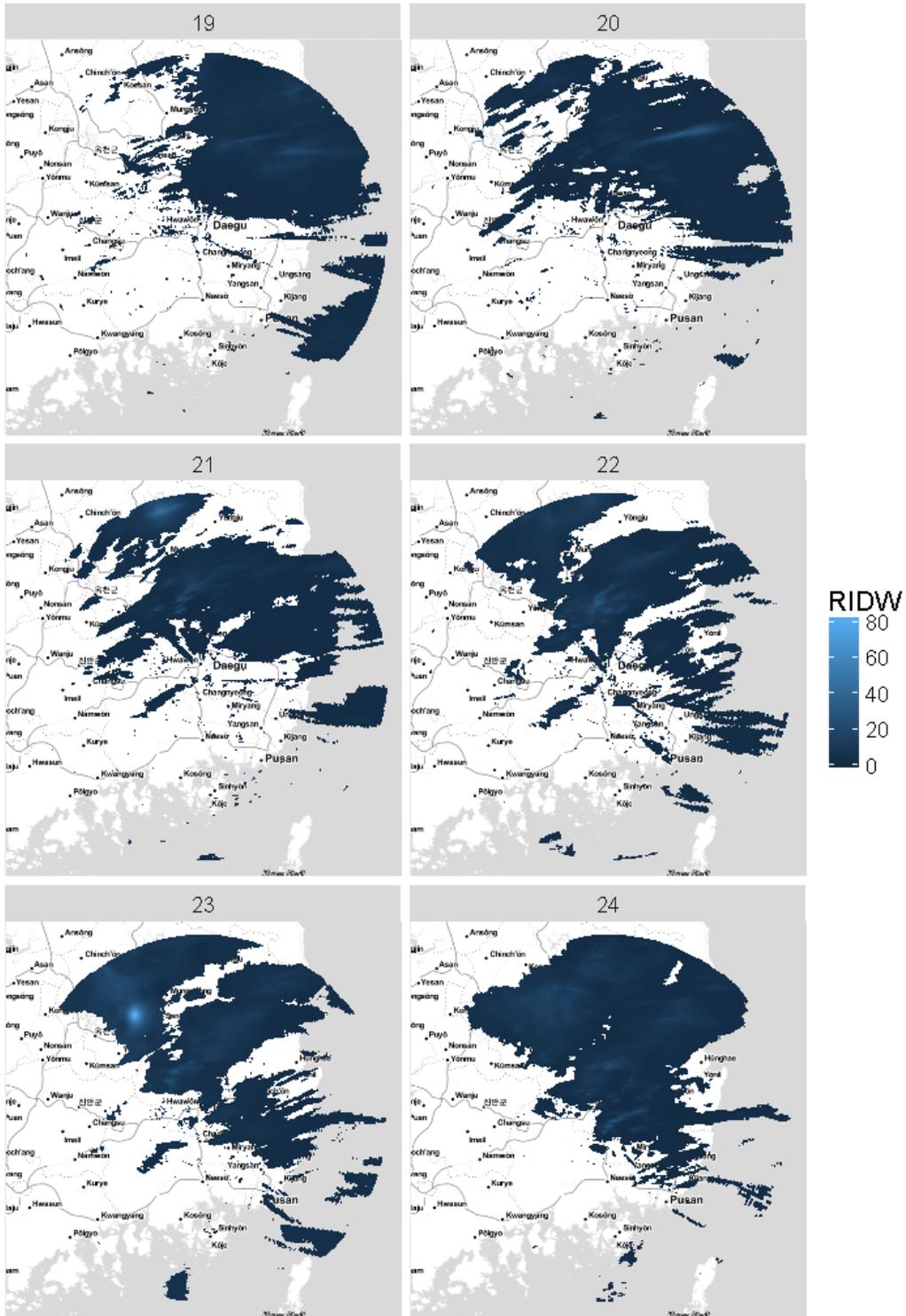


Figure B.4: RIDW predictions, hours 19-24.

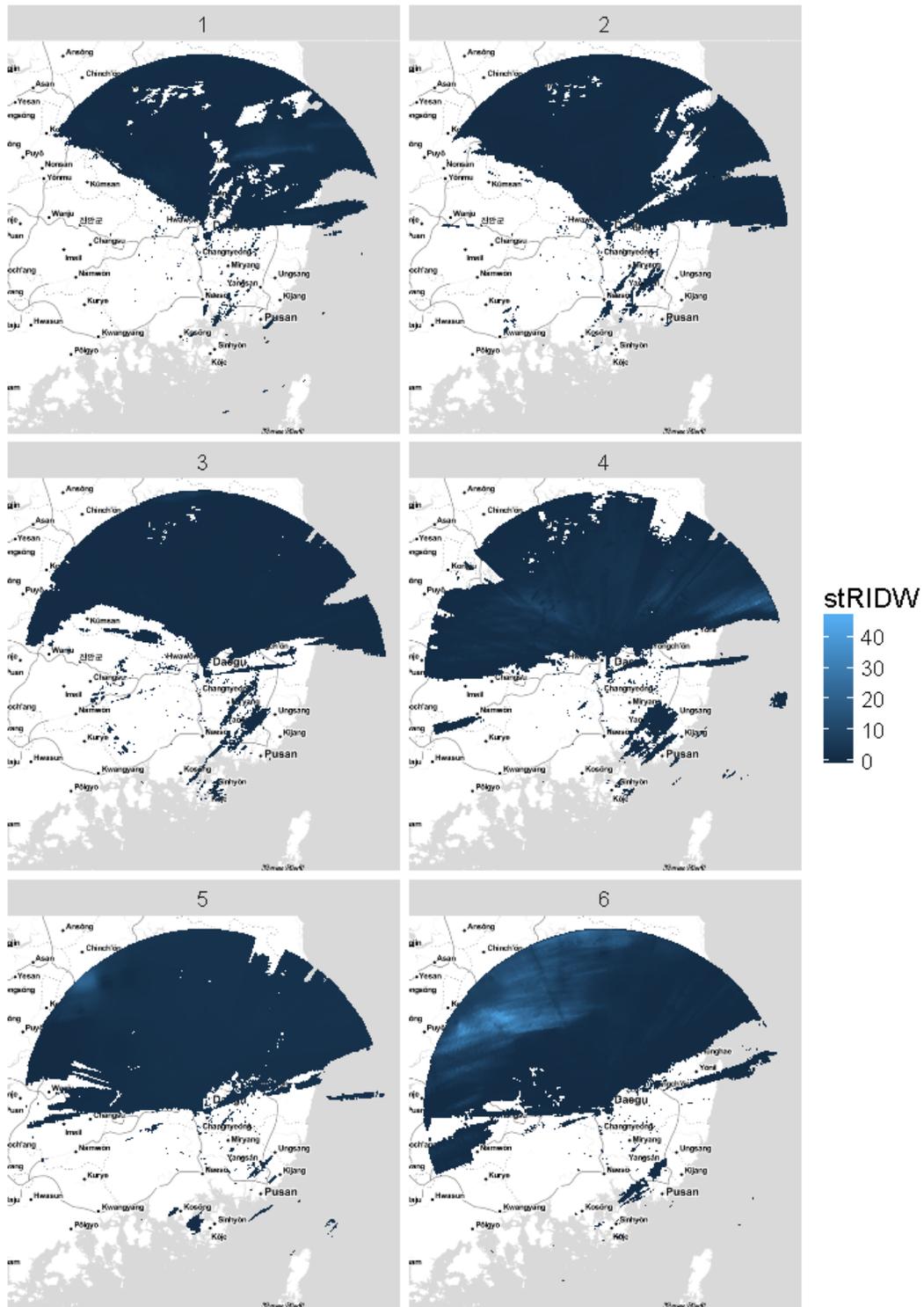


Figure B.5: stRIDW predictions, hours 1-6.

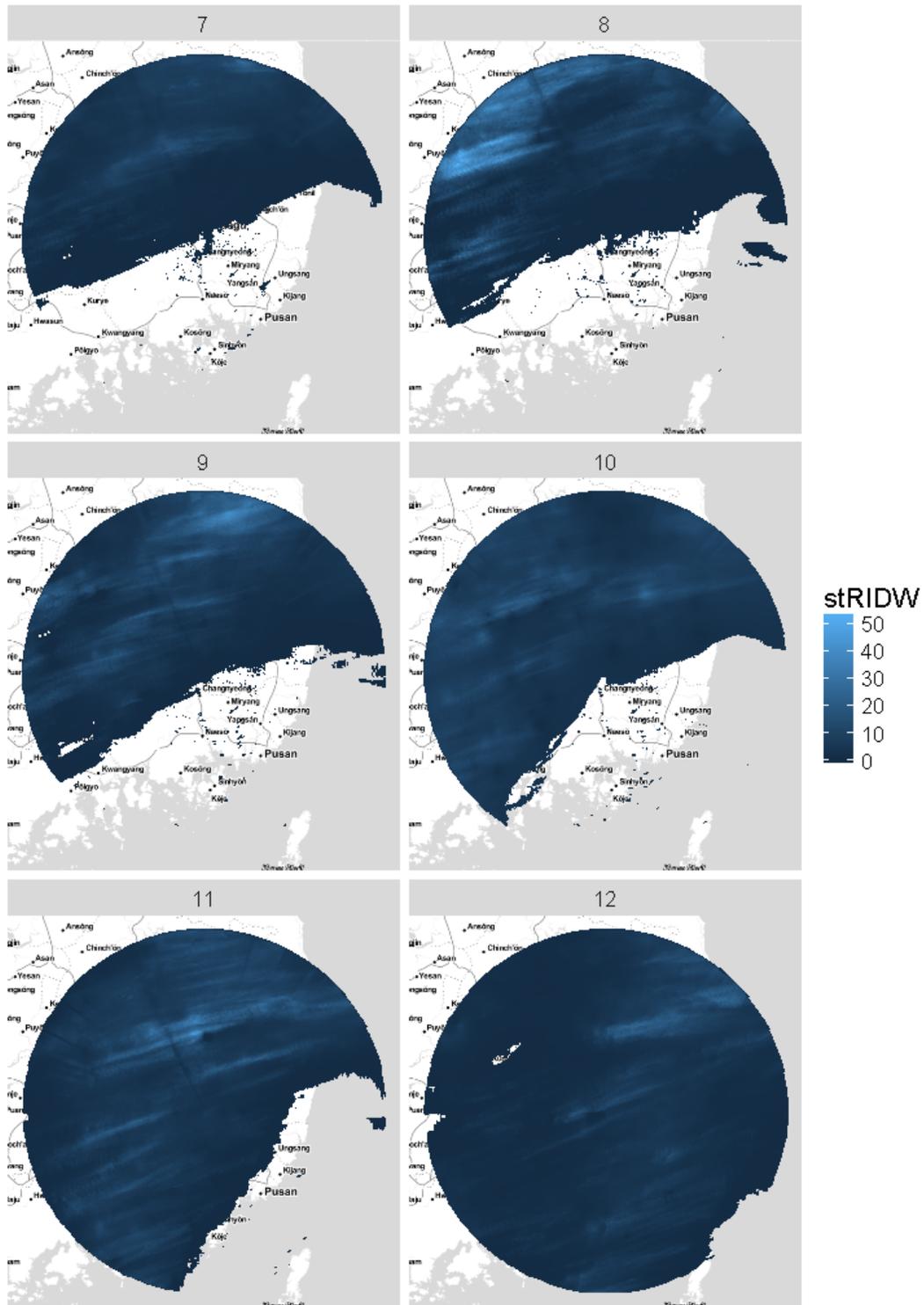


Figure B.6: stRIDW predictions, hours 7-12.

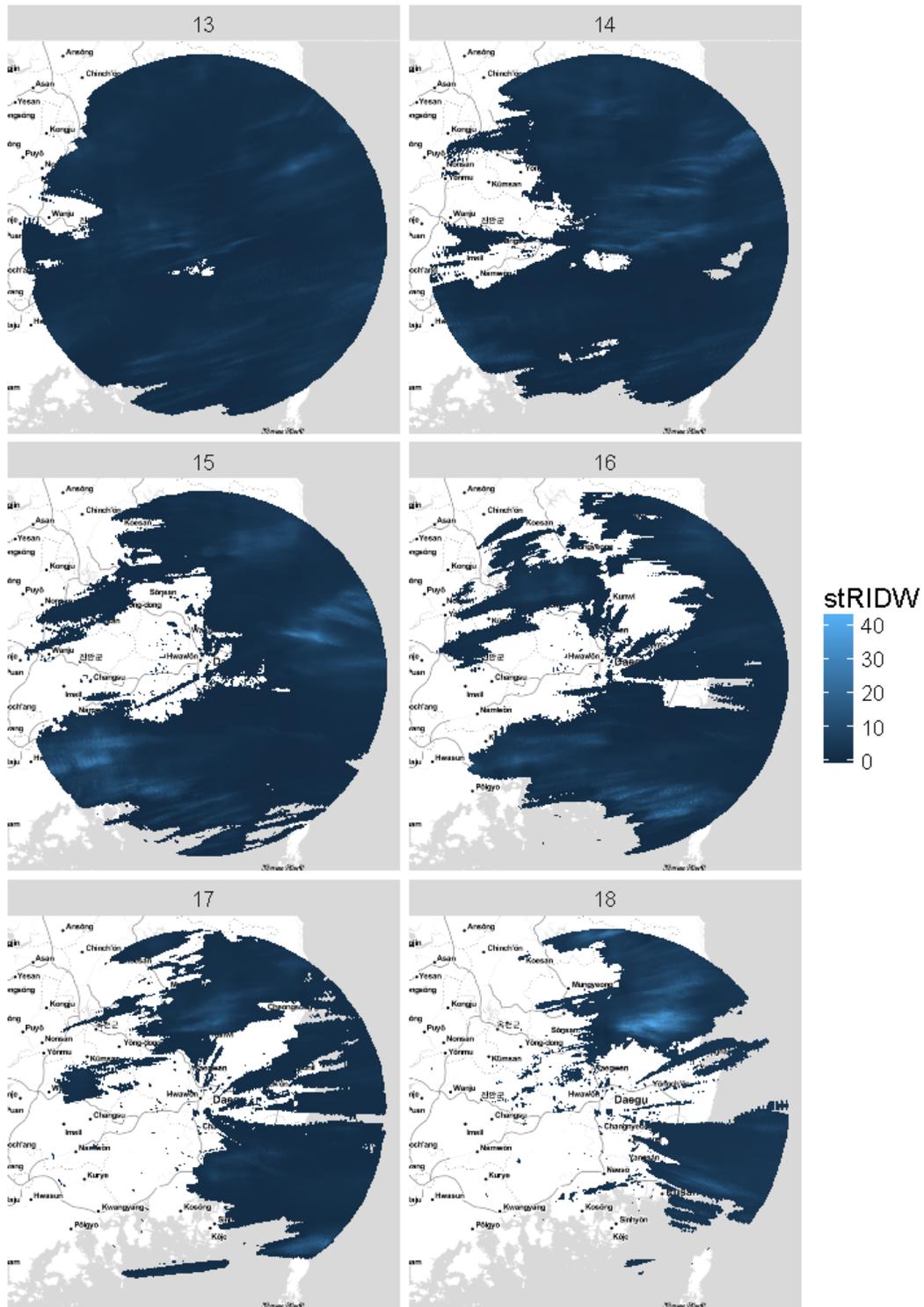


Figure B.7: stRIDW predictions, hours 13-18.

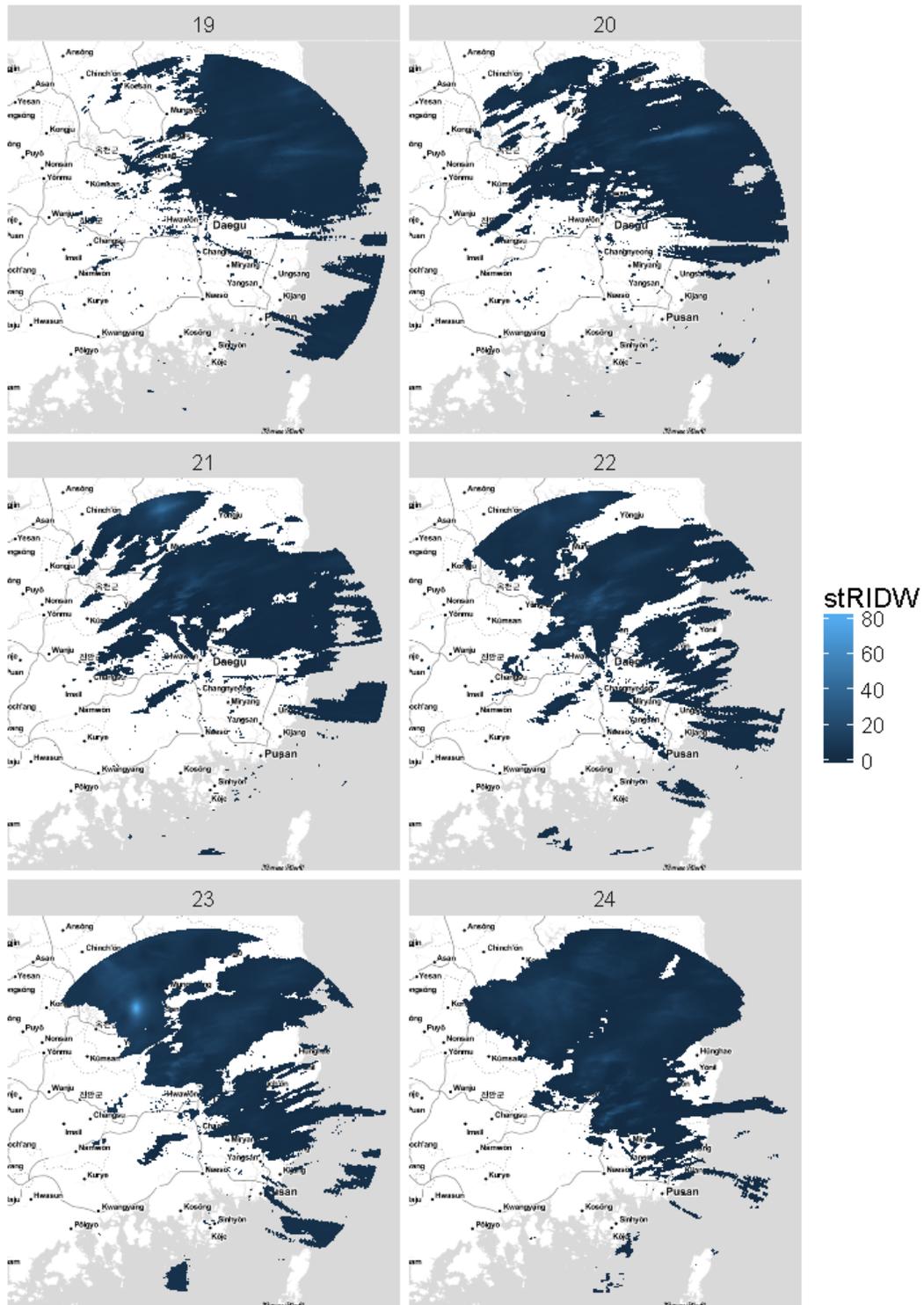


Figure B.8: stRIDW predictions, hours 19-24.

APPENDIX C
Full Conditional Distributions

For all models, let $\tilde{\mathbf{R}}^{(1)}$ denote the vector of predictions made for grid cells where $V_j \in A_i$, and let $\tilde{\mathbf{R}}^{(2)}$ denote the vector of predictions made for grid cells where $V_j \notin A_i$.

C.1 Full Conditionals

The posterior conditional distribution for σ_v^2 and τ^2 are

$$\begin{aligned} \frac{1}{\sigma_v^2} | \dots &\sim \text{Gamma} \left(\alpha_\sigma + \frac{n}{2}, \beta_\sigma + \frac{1}{2} \left(\mathbf{V} - \tilde{\mathbf{R}}^{(1)} \right)^T \left(\mathbf{V} - \tilde{\mathbf{R}}^{(1)} \right) \right), \\ \frac{1}{\tau^2} | \dots &\sim \text{Gamma} \left(2 + \frac{I}{2}, 1 + \frac{1}{2} \tilde{\mathbf{R}}^T (\mathbf{D}_w - \mathbf{W}) \tilde{\mathbf{R}} \right), \end{aligned}$$

where \mathbf{W} is the neighborhood matrix with rook structure and \mathbf{D}_w is the diagonal matrix whose i th entry describes the number of grid cell neighbors to grid cell A_i .

The posterior conditional distribution for $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ is multivariate normal with mean vector $\lambda_\alpha \chi_\alpha$ and covariance χ_α , where

$$\begin{aligned} \chi_\alpha^{-1} &= \mathbf{G}_\alpha^T \mathbf{G}_\alpha + \mathbf{I}_2, \\ \lambda_\alpha &= \mathbf{G}_\alpha^T \log(\boldsymbol{\sigma}_R^2), \\ \mathbf{G}_\alpha &= (\mathbf{1}, \tilde{\mathbf{R}}). \end{aligned}$$

In all proposed models, the $\sigma_R^2(A_i)$ must be generated via a Metropolis-Hastings algorithm from the lognormal specification in (5.3).

3.1.1 Degrees of Freedom Parameter ν .

Both degrees of freedom parameters ν_R and ν_v must be generated through a Metropolis-Hastings scheme. Let ν' be the value resulting from the previously accepted draw of ν , let k be a realization from a $N(0, \sqrt{3})$ draw, and let c be a realization from a truncated $N_{(2, \infty)}(0, \sqrt{3})$ draw. In order to facilitate proper

exploration of the parameter space, we suggest the proposal scheme for ν .

$$\nu = \begin{cases} \nu' + k & \text{if } \nu' + k > 2, \\ c & \text{otherwise.} \end{cases} \quad (\text{C.1})$$

3.1.2 NN Model

For the NN model, whose errors take on the forms given in (5.1) and (5.5), the remainder of the parameters can be drawn from their full conditional distributions as shown below. The posterior conditional distribution for σ_v^2 is

$$\frac{1}{\sigma_v^2} | \dots \sim \text{Gamma} \left(\alpha_\sigma + \frac{n}{2}, \beta_\sigma + \frac{1}{2} (\mathbf{V} - \tilde{\mathbf{V}})^T \mathbf{H} (-\phi_v)^{-1} (\mathbf{V} - \tilde{\mathbf{V}}) \right). \quad (\text{C.2})$$

The posterior conditional distribution for $\tilde{\mathbf{V}}$ is a multivariate normal with mean vector $\mathbf{D}_{\tilde{v}} \mathbf{d}_{\tilde{v}}$ and covariance $\mathbf{D}_{\tilde{v}}$, where

$$\begin{aligned} \mathbf{D}_{\tilde{v}}^{-1} &= \frac{1}{\sigma_v^2} \mathbf{H} (-\phi_v)^{-1} + \mathbf{I}_n \frac{1}{\sigma_{\tilde{v}}^2}, \\ \mathbf{d}_{\tilde{v}} &= \frac{1}{\sigma_v^2} \mathbf{H} (-\phi_v)^{-1} \mathbf{V} + \frac{1}{\sigma_{\tilde{v}}^2} \tilde{\mathbf{R}}^{(1)}. \end{aligned} \quad (\text{C.3})$$

We sample the elements of $\tilde{\mathbf{R}}$ with the following univariate scheme: if $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(1)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R1} d_{R1}, D_{R1})$, where

$$\begin{aligned} D_{R1}^{-1} &= \frac{1}{\sigma_R^2(A_i)} + \frac{1}{\sigma_{\tilde{v}}^2} + \frac{w_i}{\tau^2}, \\ d_{R1} &= \frac{R(A_i)}{\sigma_R^2(A_i)} + \frac{\tilde{V}(A_i)}{\sigma_{\tilde{v}}^2} + \frac{1}{\tau^2} \sum_{i' \sim i} \tilde{R}(A_{i'}). \end{aligned} \quad (\text{C.4})$$

For $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(2)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R2} d_{R1}, D_{R2})$, where

$$\begin{aligned} D_{R2}^{-1} &= \frac{1}{\sigma_R^2(A_i)} + \frac{w_i}{\tau^2}, \\ d_{R2} &= \frac{R(A_i)}{\sigma_R^2(A_i)} + \frac{1}{\tau^2} \sum_{i' \sim i} \tilde{R}(A_{i'}). \end{aligned} \quad (\text{C.5})$$

3.1.3 NT Model

We can again generate all remaining parameters from their full conditional distributions for the NT model, whose error structures are given in (5.2) and (5.5).

The conditional posterior distribution for σ_v^2 is given in (C.2), and the conditional posterior distribution for $\omega_R^2(A_i)$ is

$$\frac{1}{\omega_R^2(A_i)} | \dots \sim \text{Gamma} \left(\frac{\nu_R}{2}, \frac{\nu_R \sigma_R^2(A_i)}{2} \right). \quad (\text{C.6})$$

The posterior conditional distribution for $\tilde{\mathbf{V}}$ is a multivariate normal with mean vector $\mathbf{D}_{\tilde{v}} \mathbf{d}_{\tilde{v}}$ and covariance $\mathbf{D}_{\tilde{v}}$ using the elements from (C.3).

We sample the elements of $\tilde{\mathbf{R}}$ with the following univariate scheme: if $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(1)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R1} d_{R1}, D_{R1})$, where

$$\begin{aligned} D_{R1}^{-1} &= \frac{1}{\omega_R^2(A_i)} + \frac{1}{\sigma_v^2} + \frac{w_i}{\tau^2}, \\ d_{R1} &= \frac{R(A_i)}{\omega_R^2(A_i)} + \frac{\tilde{V}(A_i)}{\sigma_v^2} + \frac{1}{\tau^2} \sum_{i' \sim i} \tilde{R}(A_{i'}). \end{aligned} \quad (\text{C.7})$$

For $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(2)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R2} d_{R2}, D_{R2})$, where

$$\begin{aligned} D_{R2}^{-1} &= \frac{1}{\omega_R^2(A_i)} + \frac{w_i}{\tau^2}, \\ d_{R2} &= \frac{R(A_i)}{\omega_R^2(A_i)} + \frac{1}{\tau^2} \sum_{i' \sim i} \tilde{R}(A_{i'}). \end{aligned} \quad (\text{C.8})$$

3.1.4 TN Model

In the TN model case, whose error structures are given in (5.1) and (5.6), we can no longer gain conjugacy by implementing the $\sigma_v^2 \sim IG(\cdot)$ prior parameterization suggested by Paci et al. (2015). Instead, we use the improper uniform prior parameterization given in (5.6), and borrow from (Zellner, 1976), which gives the

posterior conditional distribution for σ_v^2 is

$$\begin{aligned} \frac{\sigma_v^2}{s_v^2} &\sim F(\nu_V, n), \\ s_v^2 &= (\mathbf{V} - \tilde{\mathbf{V}})^T \mathbf{H}(-\phi_v)^{-1} (\mathbf{V} - \tilde{\mathbf{V}}). \end{aligned} \quad (\text{C.9})$$

The conditional posterior distribution of ω_v^2 is then

$$\frac{1}{\omega_v^2} | \dots \sim \text{Gamma} \left(\frac{\nu_V}{2}, \frac{\nu_V \sigma_v^2}{2} \right). \quad (\text{C.10})$$

The posterior conditional distribution for $\tilde{\mathbf{V}}$ is a multivariate normal with mean vector $\mathbf{D}_{\tilde{v}} \mathbf{d}_{\tilde{v}}$ and covariance $\mathbf{D}_{\tilde{v}}$, where

$$\begin{aligned} \mathbf{D}_{\tilde{v}}^{-1} &= \frac{1}{\omega_v^2} \mathbf{H}(-\phi_v)^{-1} + \mathbf{I}_n \frac{1}{\sigma_{\tilde{v}}^2}, \\ \mathbf{d}_{\tilde{v}} &= \frac{1}{\omega_v^2} \mathbf{H}(-\phi_v)^{-1} \mathbf{V} + \frac{1}{\sigma_{\tilde{v}}^2} \tilde{\mathbf{R}}^{(1)}. \end{aligned} \quad (\text{C.11})$$

If $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(1)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R1} d_{R1}, D_{R1})$, and if $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(2)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R2} d_{R1}, D_{R2})$, where the elements are taken from (C.4) and (C.5) respectively.

3.1.5 TT Model

The TT model, whose error structures are given in (5.2) and (5.6), combines both of the robust proposals given. ν_V and ν_R are given the proposal distributions described in (C.1). Just as with the TN model, the posterior conditional distribution for σ_v^2 is given by (C.9). The conditional posterior distributions for $\omega_R^2(A_i)$ and ω_v^2 are given in (C.6) and (C.10) respectively.

The posterior conditional distribution for $\tilde{\mathbf{V}}$ is a multivariate normal with mean vector $\mathbf{D}_{\tilde{v}} \mathbf{d}_{\tilde{v}}$ and covariance $\mathbf{D}_{\tilde{v}}$, $\mathbf{d}_{\tilde{v}}$ are taken from (C.11). If $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(1)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R1} d_{R1}, D_{R1})$, and if $\tilde{R}(A_i) \in \tilde{\mathbf{R}}^{(2)}$, the full conditional distribution for $\tilde{R}(A_i)$ is $\text{Normal}(D_{R2} d_{R1}, D_{R2})$, where the elements are taken from (C.7) and (C.8) respectively.

APPENDIX D

Uncertainty Quantification Maps

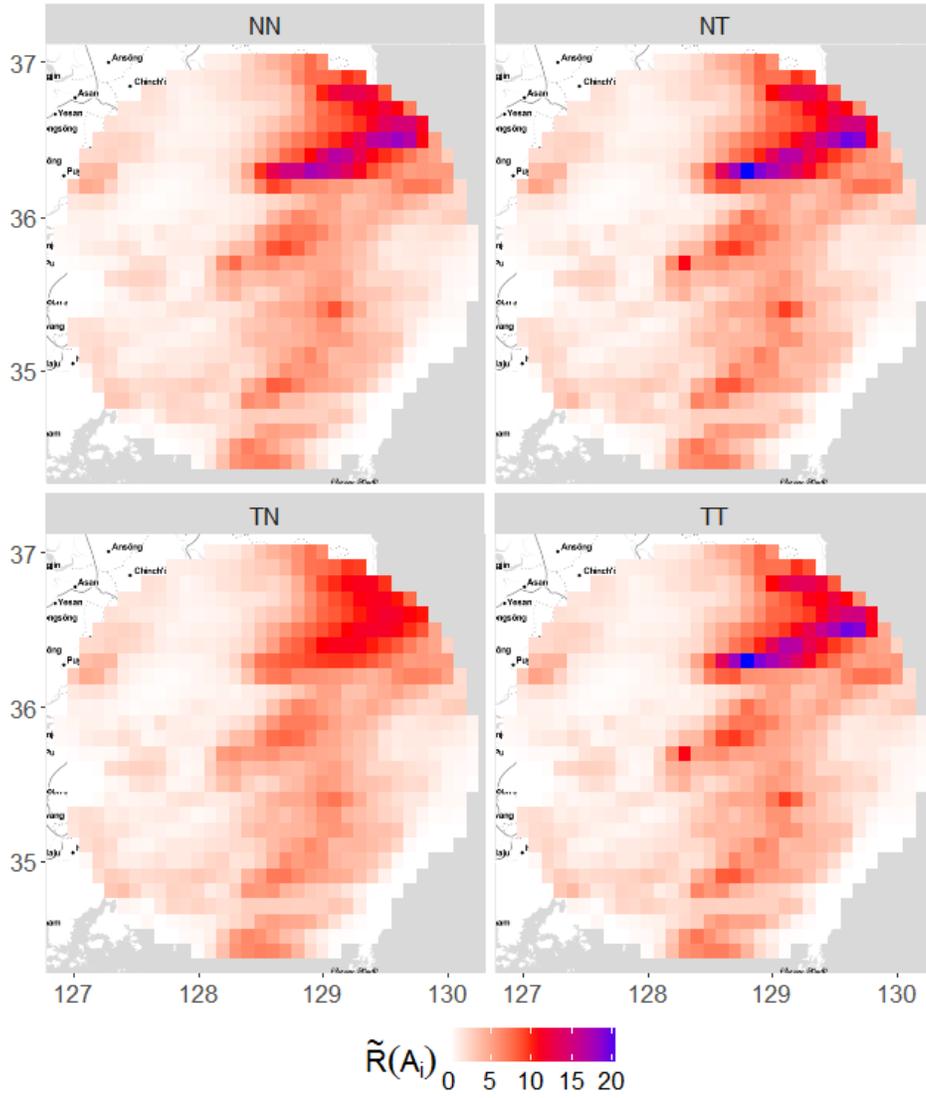


Figure D.1: Predictions in grid cells by model.

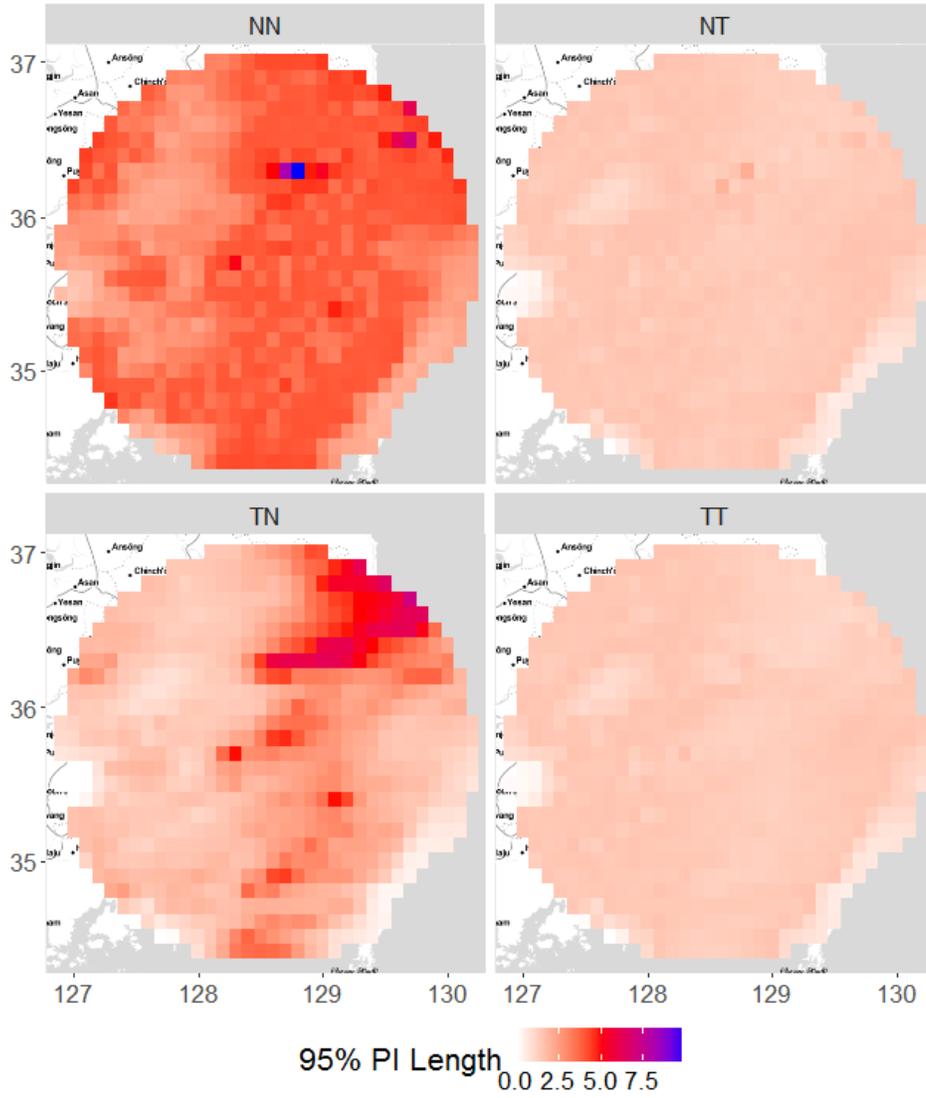


Figure D.2: Predictions interval width in grid cells, by model.

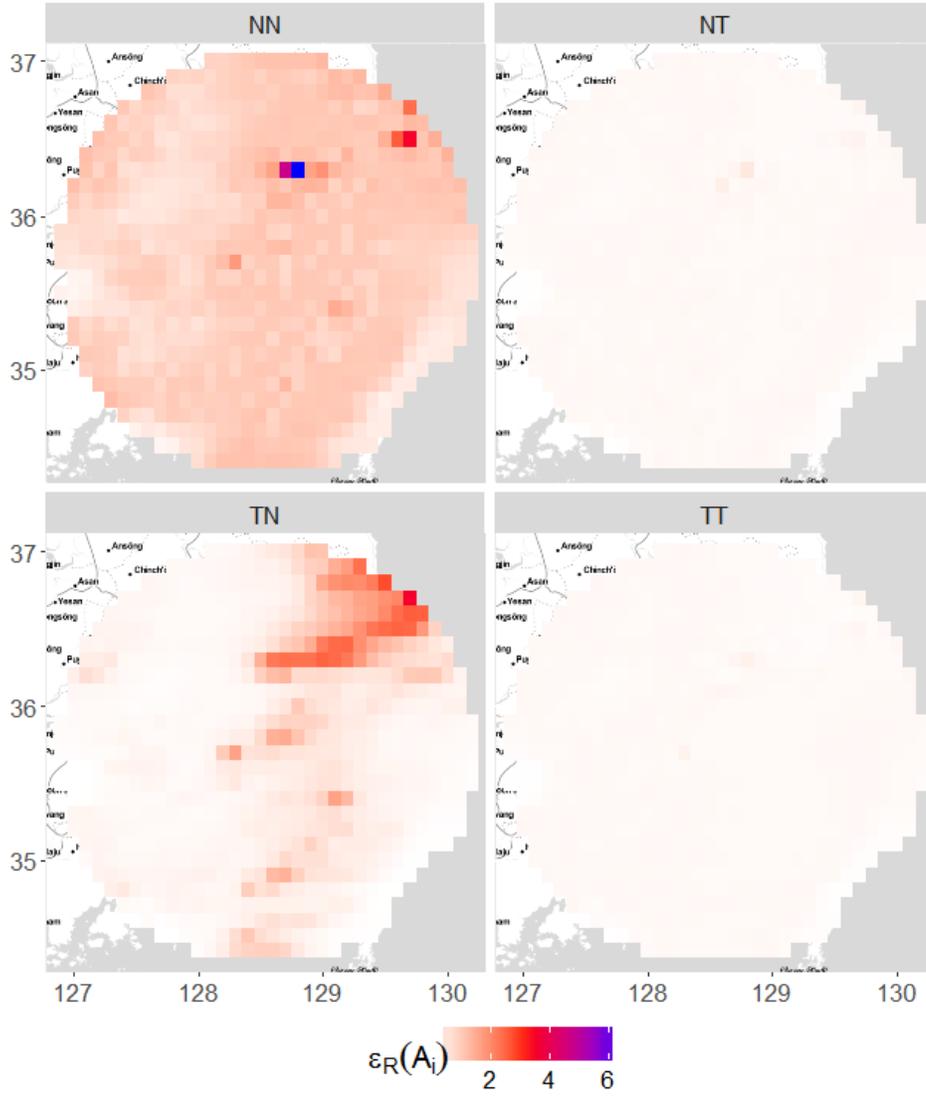


Figure D.3: $\left| R(A_i) - \tilde{R}(A_i) \right|$ in grid cells, by model.

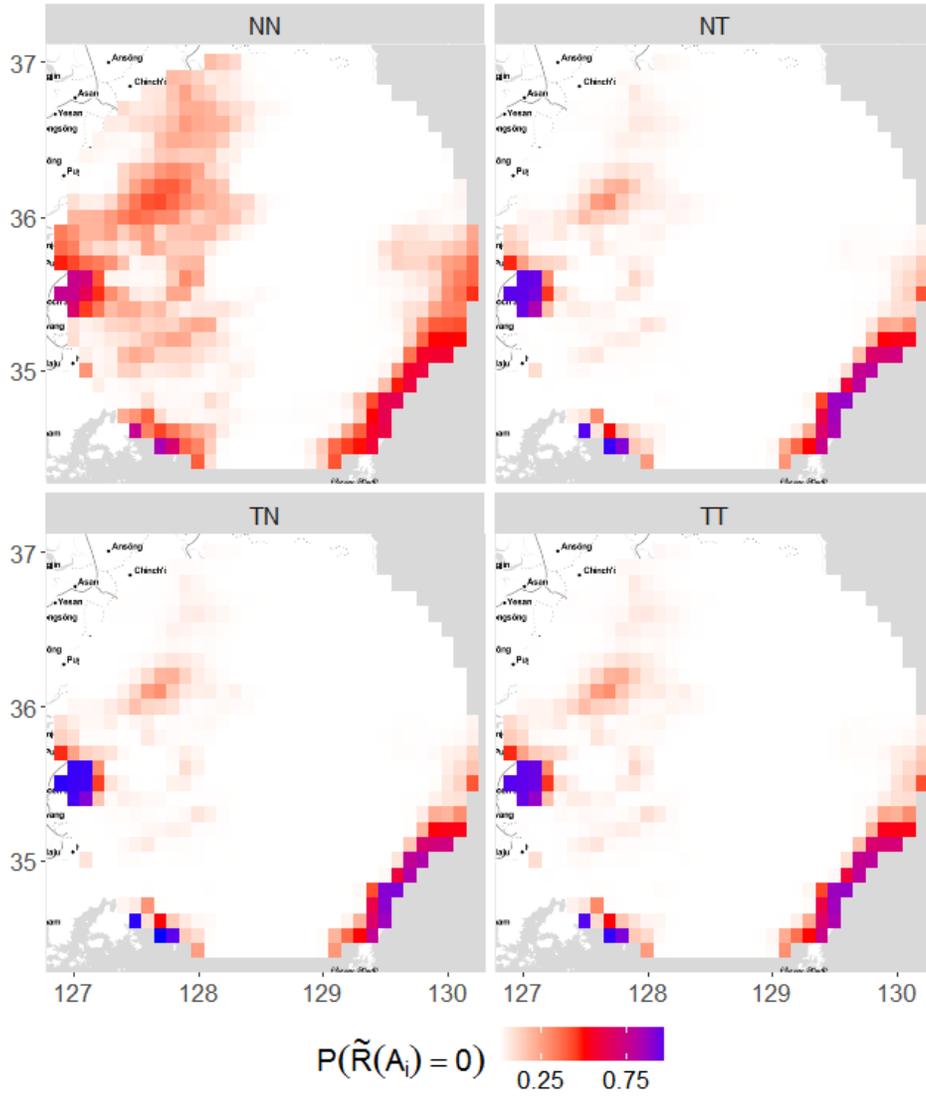


Figure D.4: Proportion of zeros resulting from point mass scheme in grid cells, by model.

BIBLIOGRAPHY

- Anscombe, F.J. (1967). "Topics in the Investigation of Linear Relations Fitted by Method of Least Squares." *Journal of the Royal Statistical Society, Series B.* 29, 1-52.
- Berger, J.O., Strawderman, W., and Tang, D. (2005). "Posterior Propriety and Admissibility of Hyperpriors in Normal Hierarchical Models." *The Annals of Statistics.* 33, 606-46.
- Besag, J. (1974). "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society, Series B.* 36, 192-236.
- Besag, J., York, J., and Mollie, A. (1991). "Bayesian Image Restoration with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics.* 43, 1-20.
- Cho, Y., Lee, G.W., Kim, K., and Zawadzki, I. (2006), "Identification and Removal of Ground Echoes and Anomalous Propagation Using the Characteristics of Radar Echoes," *Journal of Atmospheric and Oceanic Technology*, 23, 1206-22.
- Cressie, N. (1993), *Statistics for Spatial Data*. Wiley-Interscience.
- Efron, B. (1983). "Estimating the Error Rate of a Prediction Rule: Improvement on Cross Validation." *Journal of the American Statistical Association.* 78, 316-31.
- Fernandez, C. and Steel, M. (1999). "Multivariate Student-t Regression Models: Pitfalls and Inference." *Biometrika.* 86, 153-167.
- Fonseca, T., Ferreria, M., and Migon, H. (2008). "Objective Bayesian Analysis for the Student-T Regression Model." *Biometrika.* 2, 325-333.
- Fulton, R.A., Breidenbach, J.P., Seo, D.J., Miller, D.A., and O'Bannon, T. (1998), "The WSR-88D Rainfall Algorithm." *Weather and Forecasting.* 13, 377-95.
- Galassi, M., et al. (2015). *GNU Scientific Library Reference Manual, 2nd Edition*. <http://www.gnu.org/software/gsl/>

- Gelfand, A.E. and Ghosh, S.K. (1998). "Model Choice: A Minimum Posterior Predictive Loss Approach." *Biometrika*. 75, 651-9.
- Gelfand, A.E., Zhu, L., and Carlin, B.P. (2001). "On the Change of Support Problem for Spatio-temporal Data." *Biostatistics*. 2, 31-45.
- Gelman, A. and Hill, J. (2007). "Data Analysis Using Regression and Multilevel/Hierarchical Models." *Cambridge University Press*.
- Geweke, J. (1993). "Bayesian Treatment of the Independent Student-t Linear Model." *Journal of Applied Econometrics*. 8, 19-40.
- Ghosh, S., Das, D., Kao, S., Gaguly, A.R. (2011). "Lack of Uniform Trends but Increasing Spatial Variability in Observed Indian Rainfall Extremes." *Nature Climate Change*. 2, 86-91.
- Ghosh, S., Gelfand, A.E., and Mølhave, T. (2012). "Attaching Uncertainty to Deterministic Spatial Interpolations." *Statistical Methodology*. 9, 251-64.
- González, J., Lajara, J., and Díaz, S. (2015). "Errors Analysis in Extreme Rainfall Recording Due to Tipping Bucket Rain Gauge Performance." *E-proceedings of the 36th IAHR World Congress*. Presented 28 June - 3 July, 2015, The Hague, Netherlands.
- Goudenhooff, E. and Delobbe, L. (2009). "Evaluation of Radar-Gauge Merging Methods for Quantitative Precipitation Estimates." *Hydrology and Earth System Sciences*. 13, 195-203.
- Guennebaud, G., Jacob, B., et al. (2010). "Eigen v3."
<http://eigen.tuxfamily.org>
- Houlding, S. (2000). *Practical Geostatistics: Modeling and Spatial Analysis*. Springer.
- Hunter, S.M. (1996). "WSR-88D Radar Rainfall Estimation: Capabilities, Limitations, and Potential Improvements." *National Weather Digest*. 20, 26-38.
- Isaaks, E.H. and Srivastava, R.M. (1989). *Applied Geostatistics*. Oxford University Press.
- Jacquier, E., Polson, N., and Rossi, P. (2004). "Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors." *Journal of Econometrics*. 122, 185-212.

- Jørgensen, B. (1987). “Exponential Dispersion Models. *Journal of the Royal Statistical Society, Series B, Methodological*. 49, 127-45.
- Johnson, J.T., MacKeen, P.L., Witt, A., Mitchell, D.E., Stumpf, G.J., Eilts, M.D., and Thomas, K.W. (1998). “The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm.” *Weather and Forecasting*. 13, 263-76.
- Joseph, V.R., and Kang, L. (2012). “Regression-based Inverse Distance Weighting with Applications to Computer Experiments.” *Technometrics*. 53, 254-65.
- Journal, A.G. and Huijbregts, C.J. (1978). *Mining Geostatistics*. The Blackburn Press.
- Juarez, M. and Steel, M. (2010). “Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions.” *Journal of Business and Economic Statistics*. 28, 52-66.
- Kahle, D. and Wickham, H. (2013). “ggmap: Spatial Visualization with ggplot2.” *The R Journal*. 5, 144-61.
- Kwon, S., Lee, G., and Kim, G. (2015). “Rainfall Estimation from an Operational S-Band Dual-Polarization Radar: Effect of Radar Calibration.” *Journal of the Meteorological Society of Japan*. 93, 65-79.
- Kullback, S. and Leibler, R.A. (1951). “On Information and Sufficiency.” *Annals of Mathematical Statistics*. 22, 79-86.
- Lee, G.W. (2006). “Sources of Errors in Rainfall Measurements by Polarimetric Radar: Variability of Drop Size Distributions, Observational Noise, and Variation of Relationships between R and Polarimetric Parameters.” *Journal of Atmospheric and Oceanic Technology*. 23, 1005-28.
- Lloyd, C. (2006). *Local Models for Spatial Analysis, 1st Edition*. CRC Press.
- Maddox, R.A., Zhang, J., Gourley, J.J., and Howard, K.W. (2002). “Weather Radar Coverage over the Contiguous United States.” *Weather Forecasting*. 17, 927-34.
- Maity, A., and Sherman, M. (2012). “Testing for Spatial Isotropy Under General Designs.” *Journal of Statistical Planning and Inference*. 142, 1081-91.

- Martins, T.G. and Rue, H.. (2013). “Prior for Flexibility Parameters: the Student’s t Case.” Technical Report No. S8-2013, Department of Mathematical Sciences, NTNU, Norway.
- Min, Y. and Agresti, A. (2002). “Modeling Nonnegative Data with Clumping at Zero: A Survey.” *Journal of the Iranian Statistical Society.* 1, 7-33.
- Paci, L., Gelfand, A.E., Cocchi, D. (2015). “Quantifying Uncertainty for Temperature Maps Derived from Computer Models.” *Spatial Statistics.* 12, 96-108.
- Pebesma, E.J. (2000). “The Role of External Variables and GIS Databases in Geostatistical Analysis.” *Transactions in GIS.* 10, 615-32.
- Pebesma, E.J. (2004). “Multivariable Statistics in S: the `gstat` package.” *Computers and Geosciences.* 30, 683-91.
- Poole, D. and Raftery, A.E. (2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach.” *Journal of the American Statistical Association.* 95, 1244-55.
- Raftrey, A.E., Givens, G.H., and Zeh, J.E. (1995). “Inference from a Deterministic Population Dynamics Model for Bowhead Whales.” *Journal of the American Statistical Association.* 90, 402-16.
- Ryzhkov, A.V., Giangrande, S.E., and Schmur, T.J. (2005). “Rainfall Estimation with a Polarimetric Prototype of WSR-88D.” *Journal of Applied Meteorology.* 44, 502-15.
- Sempere-Torres, D., Corral, C. Raso, J., and Malgrat, P. (1999). “Use of Weather Radar for Combined Sewer Overflows Monitoring and Control.” *Journal of Environmental Engineering.* 125, 372-9.
- Seo, D.J. and Smith, J.A. (1991). “Rainfall Estimation Using Raingages and Radar - A Bayesian Approach: 1. Derivation of Estimators.” *Stochastic Hydrology and Hydraulics.* 5, 17-29.
- Seo, D.J. and Smith, J.A. (1991). “Rainfall Estimation Using Raingages and Radar - A Bayesian Approach: 2. An Application.” *Stochastic Hydrology and Hydraulics.* 5, 31-44.
- Shapiro, S.S., and Wilk, M.B. (1965). “An Analysis of Variance Test for Normality.” *Biometrika.* 52, 591-611.

- Simpson, D.P., Rue, H., Martins, T.G., Riebler, A., and Sørbye, S.H. (2014). “Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors.” *ARXIV:stat.ME*.
<https://arxiv.org/pdf/1403.4630v4.pdf>
- Song, J.J., Kwon, S., and Lee, G.W. (2015). “Incorporation of Parameter Uncertainty into Spatial Interpolation Using Bayesian Trans-Gaussian Kriging.” *Advances in Atmospheric Science*. 32, 413-23.
- Sørbye, S.H., and Rue, H. (2014). “Scaling Intrinsic Gaussian Random Field Priors in Spatial Modelling.” *Spatial Statistics*.8, 39-51.
- Spiegelhalter, D.J., Thomas, A., and Best, N.G. (1999). “WinBUGS Version 1.2 User Manual.” *MRC Biostatistics Unit*.
- Sun, X., Mein, R.G., Keenan, T.D., and Elliott, J.F. (2000). “Flood Estimation Using Radar and Raingauge Data.” *Journal of Hydrology*. 239, 4-18.
- Tobin, J. (1958). “Estimation of Relationships for Limited Dependent Variables.” *The Econometric Society*. 26, 24-36.
- Tomczak, M. (1998). “Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach.” *Journal of Geographic Information and Decision Analysis*. 2, 18-30.
- van Zyl, J.M. (2015). “The Efficiency of the Likelihood Ratio to Choose between a T-Distribution and a Normal Distribution.” *ARXIV:Stat.Co*.
<https://arxiv.org/ftp/papers/1505/1505.01979.pdf>
- Villa, C. and Walker, S. (2014). “Objective Prior for the Number of Degrees of Freedom of a t-Distribution.” *Bayesian Analysis*. 9, 197-220.
- Vivekanandan, J., Zhang, G., Ellis, S.M., Rajopadhyaya, D., and Avery, S.K. (2003). “Radar Reflectivity Calibration Using Differential Propagation Phase Measurement.” *Radio Science*. 38, 8049-63.
- Wang, L.P., Ochoa-Rodriguez, S., Onof, C., and Willems, P. (2015). “Singularity-sensitive Gauge-based Radar Rainfall Adjustment Methods for Urban Hydrological Applications.” *Hydrology and Earth System Sciences*. 19, 4001-21.
- Xia, R. and Chandrasekar, V. (1997). “Development of a Neural Network Based Algorithm for Rainfall Estimation from Radar Observations.” *IEEE Transactions on Geoscience and Remote Sensing*. 35, 160-71.

Zellner, A. (1975). "Bayesian Analysis of Regression Error Terms." *Journal of the American Statistical Association*. 70, 138-44.

Zellner, A. (1976). "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms." *Journal of the American Statistical Association*. 71, 400-5.