8-2019

Interval-Valued Kriging Models with Applications in Design Ground Snow Load Prediction

Brennan L. Bean
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Mathematics Commons

INTERVAL-VALUED KRIGING MODELS WITH APPLICATIONS IN DESIGN

GROUND SNOW LOAD PREDICTION

by

Brennan L. Bean

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Mathematical Sciences

Approved:

_____   _____
Yan Sun, Ph.D.   Adele Cutler, Ph.D.
Major Professor   Committee Member

_____   _____
Richard Cutler, Ph.D.   Marc Maguire, Ph.D.
Committee Member   Committee Member

_____   _____
Shih-Yu (Simon) Wang, Ph.D.   Richard S. Inouye, Ph.D.
Committee Member   Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

ABSTRACT

Interval-Valued Kriging Models with Applications in Design Ground Snow Load Prediction

by

Brennan L. Bean, Doctor of Philosophy

Utah State University, 2019

Major Professor: Yan Sun, Ph.D.
Department: Mathematics and Statistics

Design snow loads in the western United States are largely undefined due to complex geography and climates, leaving the individual states to publish detailed studies for their regions. These state-level studies vary widely in methodology, yet little has been written to compare the quality of their results. This dissertation begins such a comparison through a cross validation analysis of several common geostatistical mapping techniques as applied to design ground snow load prediction in Utah and Idaho. This analysis shows that regression-kriging models and our adaptation of PRISM have lower errors than Idaho and Utah's current methods across three independently developed datasets. However, the accuracy results in this analysis are based upon design snow loads that are estimates subject to uncertainty. These estimates are better characterized as intervals rather than single values. Despite the utility of interval-valued data in this and other contexts, most geostatistical mapping techniques are not equipped to handle interval-valued inputs. In response to this need, this dissertation proposes and develops interval-valued kriging models based on the theory of random sets and a generalized L2 distance. Previous developments of interval-valued kriging employed an intractable interval-valued covariance, which led to unnecessary complexities and limitations for the models. This work extends the mathematical advancements made in interval-valued regression to a spatial framework by providing a well-defined, real valued notion of spatial

covariance between intervals. This notion of covariance allows for increased flexibility in interval-valued inputs by allowing for negative weights and predictions in certain contexts. Numerical implementation of our interval-valued kriging is provided using a penalty-based constrained optimization algorithm as part of the `intkrige` package. This package formalizes the interval spatial data workflow in conjunction with these interval-valued kriging models. The methodology is used to predict interval-valued design snow loads in Utah and the results are compared to predictions made using traditional (point-valued) kriging. This application demonstrates the advantages of our interval-valued kriging in climate research, and motivates further developments of interval-valued kriging and other spatial methods.

(146 pages)

PUBLIC ABSTRACT

Interval-Valued Kriging Models with Applications in Design Ground Snow Load Prediction

Brennan L. Bean

One critical consideration in the design of buildings constructed in the western United States is the weight of settled snow on the roof of the structure. Engineers are tasked with selecting a design snow load that ensures that the building is safe and reliable, without making the construction overly expensive. Western states use historical snow records at weather stations scattered throughout the region to estimate appropriate design snow loads. Various mapping techniques are then used to predict design snow loads between the weather stations. Each state uses different mapping techniques to create their snow load requirements, yet these different techniques have never been compared. In addition, none of the current mapping techniques can account for the uncertainty in the design snow load estimates. We address both issues by formally comparing the existing mapping techniques, as well as creating a new mapping technique that allows the estimated design snow loads to be represented as an interval of values, rather than a single value. In the process, we have improved upon existing methods for creating design snow load requirements and have produced a new tool capable of handling uncertain climate data.

ACKNOWLEDGMENTS

Adequately expressing thanks to all family, friends, and colleagues who played a role in the completion of this research would require its own dissertation. I hope those not included in this partial list of acknowledgments may feel of my gratitude for your role in my life.

Thank you, Dr. Yan Sun, for your incredible support and patience as my advisor. You provided a perfect balance of guidance when I needed help, while still giving me the time and space to figure things out on my own. Thanks also to each of my committee members, Drs. Adele Cutler, Richard Cutler, Marc Maguire, and Simon Wang. Each of you provided opportunities and mentoring above and beyond your assigned committee member role. I look forward to working more with each of you in the future.

Thanks to the Structural Engineers Association of Utah, Utah State's Department of Mathematics and Statistics, and the Utah Climate Center for financially supporting various parts of my research. Thank you also to the faculty, staff, and students that I interacted with during my time at Utah State. I consider many of you to be lifelong friends.

Thank you, Tanya Johnson, for instilling in me a love of math and a belief that I was capable of quantitative studies as a high school student. Thank you, Jill Lundell, for your pivotal role in my decision to study statistics as an undergraduate. Thank you, Mom and Dad, for raising me in an environment of faith, curiosity, empathy, and creativity that has helped me to succeed as a teacher and researcher. Finally, thank you, Jessica Bean, for your love and support through long days and even longer nights during my graduate student tenure. My name may be printed on the cover, but your name gave life to the text.

Brennan L. Bean

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ASCE	American Society of Civil Engineers
CEC	Commission for Environmental Cooperation
COOP	Cooperative Observer Network
CRAN	Comprehensive R Archive Network
EPA	Environmental Protection Agency
GEV	Generalized Extreme Value
HUC	Hydrologic Unit Code
IDW	Inverse Distance Weighting (Idaho's implementation)
LRFD	Load Resistance Factor Design
MAE	Mean Absolute Error
ME	Mean Error
medAE	Median Absolute Error
ML	Maximum Likelihood
NACSE	Northwest Alliance for Computational Science and Engineering
NCEI	National Centers for Environmental Information
NGSL	Normalized Ground Snow Loads
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service
NWS	National Weather Service
OK	Ordinary Kriging
PANR	Penalized Approximate Newton-Raphson
PCIC	Pacific Climate Impacts Consortium
PRISM	Parameter-elevation Regressions on Independent Slopes Model
RMCD	Rocky Mountain Conversion Density
RMSE	Root Mean Square Error
SC	Snow Course
SEAU	Structural Engineers Association of Utah

SK	Simple Kriging
SKLM	Simple Kriging with Varying Local Means
SNLW	The 1992 Utah Ground Snow Load Equations
SNOTEL	Snowpack Telemetry
SNWD	Snow Depth
SUMT	Sequential Unconstrained Minimization Technique
TRI	Linear Triangulation Interpolation
UK	Universal Kriging (i.e. Kriging with an External Drift)
USGS	United States Geological Survey
WESD	Water Equivalent of Snow on the Ground
WRCC	Western Regional Climate Center

CHAPTER 1

INTRODUCTION

[1] Nearly all buildings in the United States (U.S.) are designed to strike a crucial balance between safety and economy: a building must reasonably withstand the anthropogenic and environmental forces induced upon it throughout its lifetime, yet use materials and designs that are realistically affordable to the future occupants. The need for safe, yet economical structures has given rise to state, local, and national building codes, which specify design load requirements for the various forces borne by a structure during its lifetime. These load requirements are then used in a Load Resistance Factor Design (LRFD), which treats the resistance (R) of a structure and various loads induced on a structure $(Q_i, i = 1, \cdots n)$ as random variables. The implication of the LRFD design is that a structure whose loads exceed its resistance results in building collapse. Consequently, a structure is then designed to ensure that the resistance of the structure will exceed the load i.e.

$$ G = R - \left. \sum_i Q_i \right) \geq 0 $$

for an acceptable probability level (Nowak and Collins, 2012).

The term "acceptable" may seem puzzling in this context, as the probability of structure collapse would ideally be zero. However, as mentioned previously, practical economic constraints make the creation of perfectly safe structures impossible. As articulated by Nowak and Collins (2012): "Conceptually, we can design [a] structure to reduce the probability of failure, but increasing the safety... beyond a certain optimum level is not always economical." The use of overly conservative load requirements unnecessarily burdens construction costs,

---

[1] Portions of Chapters 1-3 are adapted from papers published in the Journal of Structural Engineering (Bean et al., 2017) and the Journal of Cold Regions Engineering (Bean et al., 2019). These articles were both co-authored by Dr. Marc Maguire and Dr. Yan Sun and available at the following links:

- https://ascelibrary.org/doi/10.1061/%28ASCE%29ST.1943-541X.0001870
- https://ascelibrary.org/doi/full/10.1061/(ASCE)CR.1943-5495.0000190

which is particularly problematic in states like Utah with persistent housing shortages and declining housing affordability (Wood, 2019). Surely a structure built using reasonable load requirements will do a better job protecting its inhabitants than the *designs* of a "perfectly" safe structure that is too expensive to build.

In an attempt to achieve the necessary balance between safety and economy, the Structural Engineering Institute of the American Society of Civil Engineers (ASCE) oversees the publication of the consensus design load standard "Minimum Design Loads for Buildings and Other Structures' (ASCE 7). This standard was first released in 1988 and is updated on a six-year cycle through a series of committees and sub-committees (Goupil, 2013). State and local building officials adopt the ASCE 7 standards into building codes that largely govern the design and construction of nearly all inhabitable buildings in the U.S.

One load of particular interest to mountainous and northern states is the force induced by settled snow on the roof of a structure (q_s). This force is commonly referred to as the snow load with a corresponding design snow load specified in the building code. A design snow load (q_s^*) is typically derived from a 50 year ground snow load event, which is the weight of settled snow on the ground expected to occur once every 50 years at a given location. Supposing that these ground snow loads at a particular location are characterized by a probability distribution, a 50 year event corresponds to the 98th percentile. This approach to defining design ground snow loads is referred to as a uniform hazard approach. An alternative approach to defining these load requirements is through a uniform risk approach as is used in the most recent Colorado snow load study (DeBock et al., 2017; Liel et al., 2017). This dissertation focuses on methodologies for snow loads defined using the uniform hazard approach.

The U.S. has a long and unfortunate history of structure failure and damage due to snow. A Travelers Insurance report identified the weight of snow and ice as the second most common cause of insurance claims in the northeast U.S. (Business Wire, 2016). The states of Connecticut, Massachusetts, New York, and Rhode Island reported 382 full or partial building collapses due to snow in the 2010-2011 winter (O'Rourke, 2013). More recently,

heavy snowstorms in the winter of 2017 filled local newspapers across the western U.S. with reports of snow related building collapses and fatalities (Associated Press, 2017; Fisicaro, 2017; Glover, 2017; Kato and Florio, 2017; Lafferty, 2017; Mieure, 2017). These snow-related failures can be catastrophic to local economies, like the recent \$100 million in losses incurred by Idaho/Oregon's onion industry (Ellis, 2017). In a separate study, Strobel and Liel (2013) reported an average cost of \$166 per square meter and 122 days of business interruption for repairs in 40 snow-induced building failures across the U.S. Snow-related damages can extend beyond building repairs, as Geis et al. (2011) reported more than 300 fatalities in 1,100 domestic and international snow-induced building failures. Few details are made public about the true causes of the above damages, as they could be agricultural buildings not designed to code or even suffer from construction error, but these reports and articles provide a sample of the serious consequences associated with design snow load prediction.

However, recall that subtler costs are also associated with overly conservative design snow loads. The following two examples demonstrate this point by exploring the relationship between design snow loads and roof construction costs. Roof costs are selected for these illustrations as they are likely the aspect of a structure most sensitive to snow load design.

The first example is found in the 2017 Craftsman National Building Cost Manual, which includes a table of estimated roof costs for manufactured homes rated for different snow loads. In this manual, a doubling of the roof snow load requirement from 1.44 to 2.88 kilopascals (kPa) results in an approximate threefold increase in the estimated cost per unit meter of roof (\$11 to \$36) (Moselle, 2016). The second example comes from roof joist costs provided to the authors by Vulcraft Utah (Brothersen and Fisher, 2018). These roof-only designs assume varying snow loads with the constant depths, typical joist spacings and a L/240 deflection limit, as indicated in Figure 1.1. These costs do not include the effects of the snow and larger roof components on the remainder of the gravity or seismic systems' cost. For this system, doubling the roof snow load requirement from 1.44 to 2.88 kPa leads to a 40-90% increase in the cost of the joists.

These two examples may represent highly sensitive situations with respect to cost
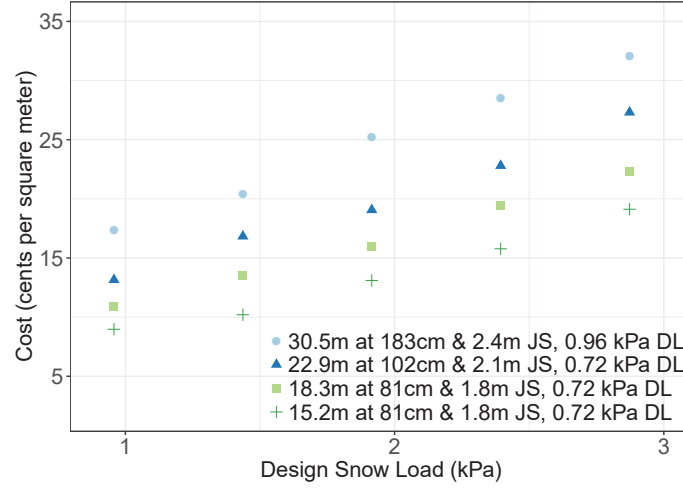
Fig. 1.1: Cost-to-snow load comparison for five different roof joist types. Data provided to the authors by Vulcraft Utah (Brigham City, Utah) in January 2018.

and snow load. Other systems and components would likely not experience such dramatic cost increases. Regardless, the potential economic burdens created by overly conservative requirements likely explain recently amended ground snow load requirements in Rich County, Utah, where a new requirement of 2.73 kPa for major communities in the county (Utah Legislature, 2016) are less than half the 6.3-7.2 kPa dictated previously by the Structural Engineers Association of Utah (SEAU, 1992) .

Defining snow load requirements for the western U.S. is particularly challenging given the complex geography. As a result, ASCE 7 requirements have historically remained unspecified for this region (ASCE, 2017). Western states have responded by defining design ground snow load requirements for their jurisdictions in state-specific studies (Sack, 2015), usually under the direction of the state Structural Engineers Associations. Many of these reports (or portions of them) are freely available to the public (Al Hatailah et al., 2015; NACSE, 2012; SEAU, 1992; Theisen et al., 2004; Torrents et al., 2016) and provide a wealth of information on dataset development, model predictions, and implications for building design. Each state employs a unique approach for defining snow load requirements, though all follow the same general workflow visualized in Figure 1.2. Note that the focus of this workflow is on determining design ground snow loads. Appropriate conversions from ground
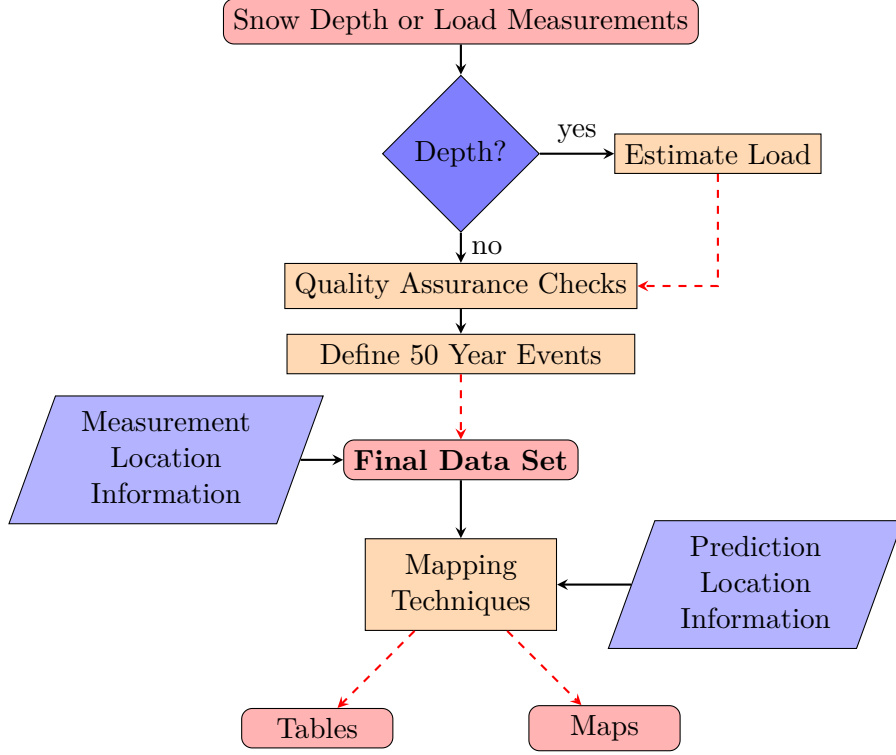
Fig. 1.2: Basic workflow for creating region-level design ground snow load requirements. Dashed, red arrows indicate steps where estimation uncertainty is introduced into the workflow.

snow loads to roof snow loads are provided in ASCE 7. For this reason, the terms "ground snow load" and "snow load" will be used interchangeably throughout this dissertation.

We demonstrate the details of this workflow by creating a new design snow load dataset for Utah in Chapter 2. For convenience, the main steps of this process are also summarized here:

- **Collect measurements of snow depth (SNWD) and water equivalent of snow on the ground (WESD):** Whenever possible, use direct measurements of WESD to determine snow load, as measured in kilopascals (kPa). When WESD measurements are not available, estimate the snow load from snow depth.

- **Estimate design (50 year) loads from annual maximum load measurements:**

This is typically accomplished by fitting the annual maximum snow loads to a right-skewed probability distribution and determining the 98th percentile.

- **Predict/Interpolate design snow loads between measurement locations:** Various mapping techniques are used to create continuous maps of design snow loads by predicting loads between measurement locations.

The western states may follow the same general workflow for defining the snow loads, but the methodologies at each particular step vary widely. Sack (2015) and Sack et al. (2016) discuss differences between state methodologies and acknowledge discrepancies these methodologies create along state boundaries. Despite these acknowledged differences, no formal comparison of design snow load prediction methods is found in the literature. A lack of accuracy comparisons makes it difficult to reconcile these inter-state differences in snow load requirements. Further, a lack of an established and reproducible method for estimating design snow loads makes it difficult to regularly revise and update snow load requirements as new data becomes available. The serious societal costs incurred by inadequate snow load design, coupled with increased uncertainty in snow patterns across the county due to climate change, motivates the need for reproducible, scalable, statistically-based approaches to defining design snow load requirements in the U.S.

There is also the need for new methods better equipped to handle the inherent imprecision in the design snow load estimates. This measurement imprecision should be a important consideration in the ensuing reliability analysis as load adjustments in areas with imprecise design snow load estimates should be larger than load adjustments in areas with precise estimates. Current mapping techniques treat all design snow load estimates as precise "observations" when predicting loads between regions. This approach robs the design snow load estimates of critical context in the subsequent reliability adjustments, which motivates the need for mapping techniques designed to handle imprecise inputs.

This dissertation address the methods and shortcomings of each of the major steps outlined in the Figure 1.2 workflow, with a focus on comparing mapping techniques for design snow loads. The result of our analyses is a rigorous set of comparisons between different

interpolation techniques, the definition of a new set of design snow load requirements for the state of Utah, and the creation of an interval-valued kriging which accommodates the imprecision in design snow load estimates when creating snow load maps. The remainder of this dissertation is organized as follows: Chapter 2 summarizes the independently created datasets used in cross validation, as well as the various spatial mapping techniques that we adapted to design snow load prediction. Chapter 3 conducts a rigorous comparison of these methods across multiple datasets and discusses considerations and implications for each approach. Chapter 4 introduces our interval-valued kriging models, which are designed to address the imprecision inherent in the design snow load estimation problem. Finally, Chapter 5 formalizes the process for analyzing interval-valued spatial data through a demonstration of the `intkrige` package. We discuss the conclusions and implications of our findings in Chapter 6 and provide links to R packages associated with this research. The appendix provides an excerpt from the unpublished Washington snow load study that illustrates our continual improvements to the design snow load estimation process.

CHAPTER 2

DATA AND METHODS

A formal comparison of design snow load prediction methodologies necessarily requires the adaptation of multiple methods across multiple datasets. This chapter is devoted to summarizing the various datasets and methods used in the cross validation comparisons provided in Chapter 3. Some of these datasets and methods were created in conjunction with the 2018 Utah Snow Load Report (Bean et al., 2018), while others were obtained from other snow load studies in Utah (SEAU, 1992) and Idaho (Al Hatailah et al., 2015).

The use of cross validation is limited to replicable methods that are separable from the input observations. For example, snow load predictions in Colorado involve a contour map of input parameter values that includes allowed discontinuities along mountain ridges (Liel et al., 2017; Torrents et al., 2016). These contours and discontinuities are inextricably connected to the measurement location observations and thus eliminate the option to use cross validation. In addition, the Montana and Oregon snow load reports do not include enough details to replicate their methods on new datasets (NACSE, 2012; Theisen et al., 2004). For this reason, our comparisons in Chapter 3 focus on the current methodologies being used in Utah and Idaho, as well as a suite of commonly used spatial mapping techniques that can be readily applied to design snow load prediction. The following subsections describe the datasets and methods we use in our comparisons.

2.1 Data

The three datasets used in the cross validation comparisons are the new Utah dataset (UT-2017), the 1992 Utah snow load report dataset (UT-1992) and the 2015 Idaho snow load report dataset (ID-2015). Table 2.1 provides an overview of each dataset. The variable of interest in each dataset is the design snow load. These readily available datasets were selected to compare the effectiveness of various spatial methods in predicting design snow

loads for different climates, terrain, and measurement location coverage. In addition, the main mapping techniques considered in this chapter are all associated with one of these datasets, including the current Utah snow load equations (UT-1992), Idaho's normalized design snow loads based on inverse distance weighting (ID-2015), kriging (UT-2017) and PRISM (UT-2017). The consideration of these three independently developed data sources ensures that the cross validation comparisons provided in Chapter 3 are not limited to one isolated dataset.

Each of these datasets use observations from Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) and Snow Course (SC) weather stations, as well as data from the National Weather Service (NWS) cooperative observer network (COOP) weather stations. Daily data from these sources can be conveniently obtained from the National Oceanic and Atmospheric Administration's (NOAA) global historical climatological network (GHCN) as maintained by the National Centers for Environmental Information (NCEI) (Menne et al., 2018). Many SNOTEL stations were installed to replace discontinued SC stations, thus creating situations where two separate stations have the same geographical location. Identical decimal degree locations for two distinct stations creates singularity issues in many spatial interpolation methods. We resolve this issue in each dataset by adding an arbitrarily small number r, $(|r| < .001)$ to the decimal degree locations to create well defined but negligible spatial separation between such stations. In the appendix, stations with nearly identical geographical locations and elevations are combined to create a single record. For this reason, we often refer to weather stations more generally as "measurement locations" throughout the remainder of this dissertation.

Figure 2.1 reveals the distinct log-linear relationship between measurement location design snow load estimates and elevation for each dataset. These scatterplots include lines representing ordinary and generalized least squares regression estimates of this log-linear relationship (using elevation as the predictor). The development of these regression lines will be discussed further in Section 2.2. In addition, marginal densities along the bottom of each plot visualize the different elevation profiles of the three datasets. For example, the

Table 2.1: Summary of the three design snow load datasets used in method comparisons.

Dataset	Measurement Locations	Depth-to-Load Conversions	Distribution
UT-2017	415	Sturm's Equation	log-normal
UT-1992	413	RMCD	log-Pearson Type III
ID-2015	651	RMCD	log-Pearson Type III
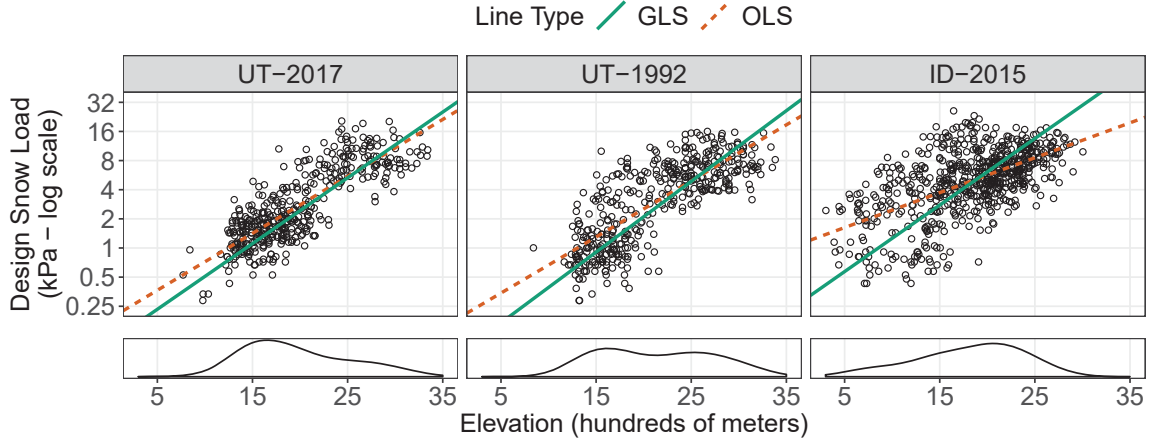
Fig. 2.1: Measurement location elevation plotted against design snow loads on a log-scale for each dataset. The lines represent the ordinary (OLS) and generalized (GLS) least squares model estimates using elevation as the sole predictor. Marginal measurement location densities across elevation are provided at the bottom of each plot.

density of measurement locations decreases sharply for elevations above 2500 meters (m) and slowly for elevations below 1400 m in the Idaho dataset, but such density trends are exactly opposite in both Utah datasets. The cross validation results of Chapter 3 must be interpreted in the context of measurement location elevation, as higher elevations relative to the dataset of interest tend to have higher snow loads and consequently more variability in predictive accuracy. Further, the areas of most importance in any snow load study are the populated locations typically located at elevations below 2500 m.

2.1.1 The new Utah dataset (UT-2017)

The new Utah dataset was created in conjunction with the 2018 Utah Snow Load Report (Bean et al., 2018) and a copy of this dataset can be found at this reference. This dataset was created with the intention of creating a reproducible data processing framework

that could be easily updated on a regular basis. Details regarding the development of this dataset are given in the following subsections.

Collection

Data were obtained primarily from the GHCN and supplemented by a handful of additional NRCS SNOTEL stations. We originally processed and visualized these raw measurements in R 3.3.0 (R Core Team, 2016), with with the help of several ancillary packages to R (Bivand et al., 2018; Douglas Nychka et al., 2015; Hijmans, 2016; Neuwirth, 2014; Wickham, 2011). This dataset contains 279 (192 COOP, 87 SNOTEL) Utah measurement locations with an additional 136 measurement (103 COOP, 33 SNOTEL), all located within 100 kilometers (km) of the Utah border. Figure 2.2 shows that most of the SNOTEL stations are concentrated in the Wasatch mountains, while most COOP stations occur in populated locations along the Wasatch front. We considered snow measurements for years 1970 to 2017. This range focuses on years where SNOTEL station measurements are available, as the earliest available measurements from active SNOTEL stations in Utah is 1978 (NRCS, 2017). SNOTEL stations are known for providing reliable real time WESD measurements in remote, high elevation areas where human access is difficult. COOP stations, while commonly used in design snow load estimates (Sack, 2015), do not provide the unprecedented precision and coverage of SNOTEL stations, often leaving users to estimate WESD from snow depth readings. Despite the inferior measurements, such stations are necessary for any appropriate estimations at low lying elevations in the state as observed in Figure 2.2. As such, lower elevation COOP data is required for a legitimate analysis, even if such measurements are of lower quality than SNOTEL readings.

Estimating Loads

Many COOP stations only provide snow depth readings, requiring the estimation of snow load from snow depth. Snow load is a function of snow depth and density, which are both highly variable and time dependent. The average depth-to-water ratio of freshly fallen snow is around 0.13 (Baxter et al., 2005) with a ratio as high as 0.5 for end of season, settled
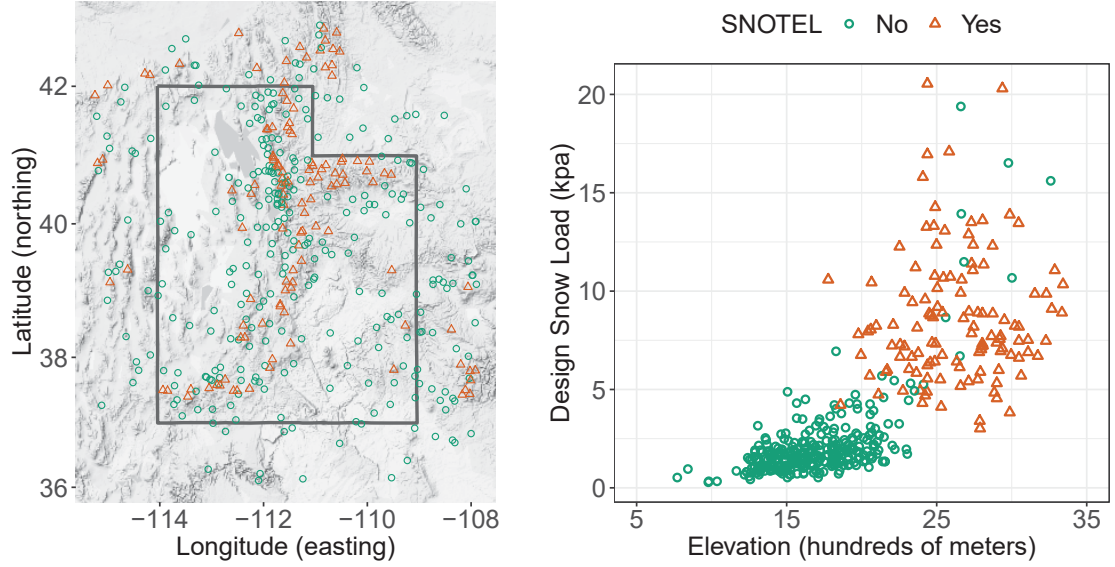
Fig. 2.2: Map of measurement location types in UT-2017. SNOTEL stations provide direct measurements of WESD.

snowpack. The western states have a history of using several different methods to convert snow depths to snow loads. Colorado converts snow depth to snow load using two non-linear curves, one created by Tobiasson and Greatorex (1996) and another developed by DeBock et al. (2017). These curves are defined as

$$q_s = g_1(h, A(\boldsymbol{x})) = p(A(\boldsymbol{x})) * f^{(1)}(h) + (1 - p(A(\boldsymbol{x}))) * f^{(2)}(h)$$

with

$$f^{(1)} = (0.0479)(0.279) \left(\frac{h}{2.54} \right)^{1.36} \qquad f^{(2)} = (0.0479)(0.584) \left(\frac{h}{2.54} \right)^{1.15}$$

where h is the snow depth measured in centimeters (cm). The load parameter $p \in [0, 1]$ reaches its lower and upper limits for elevations (A) of around 1800 m and 2600 m respectively.

Idaho uses the Rocky Mountain Conversion Density (RMCD) redefined for metric units as

$$q_s(h) = g_2(h) = \begin{cases} 0.017h & h < 55.88 \text{cm} \\ 0.0445h - 1.5274 & h \geq 55.88 \text{cm} \end{cases}$$

Table 2.2: Climate specific parameters for Sturm's equation.

Class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	0.2578	0.0010	0.0038
Prairie	0.5940	0.2332	0.016	0.0031
Tundra	0.3630	0.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000

where h represents snow depth (cm) (Sack and Sheikh-Taheri, 1986).

For UT-2017, we elect to use a method developed by Sturm et al. (2010). This method models snow load with the equation

$$q_s = g_3(h, d) = 0.0981h\left[(\rho_{max} - \rho_0)\left[1 - \exp\left(-k_1 h - k_2 d\right)\right] + \rho_0\right]$$

where d represents day of the snow season starting on October 1st (-92) and ending June 30th (181) with no zero value. Additionally, $\rho_o, \rho_{max}, k_1,$ and k_2 are parameters specific to a particular climate class defined in Table 4 of Sturm et al. (2010) and provided for convenience in Table 2.2 of this chapter. For convenience, this method is referred to hereafter as "Sturm's equation."

Sturm et al. (1995) classifies nearly all of Utah as a "Prairie" climate type. However, the coarse resolution of their classification (50 km by 50 km) makes it reasonable to believe that high elevation locations in Utah would likely be considered "alpine" if the grid was finer. Thus, we performed depth-to-load conversions for "prairie" and "alpine" terrains using the equation

$$q_s = \begin{cases} 0.0981h\left[.3608 * (1 - \exp\left(-.0016h - .0031d\right)) + .2332\right] & A < 2113.6\,\mathrm{m} \\ 0.0981h\left[.3738 * (1 - \exp\left(-.0012h - .0038d\right)) + .2237\right] & A >= 2113.6\,\mathrm{m}. \end{cases}$$

Such a method assumes a continuous increase in snow density throughout the snow season, perhaps overestimating the snow density of late season storms in low lying areas not subject to continuous snow accumulation. As such, Sturm's equation generally leads to larger design snow load estimates than similar estimates made with the RMCD. On average the resulting

design snow load estimates using Sturm's equation are nearly 50% larger than those made with the RMCD for UT-2017. The discrepancies between these methods are largely irrelevant whenever the predicted load falls below Utah's mandated minimum design snow load of approximately 1 kPa. For loads above this threshold, the more conservative load estimates offered by Sturm's equation are acceptable and perhaps desirable in the context of building design.

Quality Control

UT-2017 assumes all yearly maximums at a particular station come from the same log-normal distribution. Any false maximums due to faulty daily measurements or missing readings tend to artificially inflate the standard deviation estimate for right-skewed probability distributions, resulting in an severe overestimate of the 98th percentile. This is especially true of measurement locations with small sample sizes.

Coverage Filters

Coverage filters seek to remove artificially low annual snow load maximums from the data without throwing out excessive amounts of information. The spread of measurements across the period of the snow season in which a maximum is most likely to occur is more important the actual number of measurements in a given snow season. The low springtime temperatures at high elevations allow for consistent snow pack accumulation into April or May, while warmer, lower elevation locations will see a peak in snow accumulation much earlier in the season, as observed in Figure 2.3.

The observed difference in peak snow pack occurrence between low and high elevations in Figure2.3 prompts a two level coverage filter. All measurement locations above 2115 m (which roughly corresponds to 3rd quantile of the unfiltered set of measurement location elevations) must have at least one observation in every month from March to May in order to be considered. Measurement locations below 2115 m must have at least one observation in every month from December to March of the water year to be considered. We retain all annual maximums strictly greater than the median max at a given measurement location regardless of the yearly coverage. This exception ensures that true annual maximums are
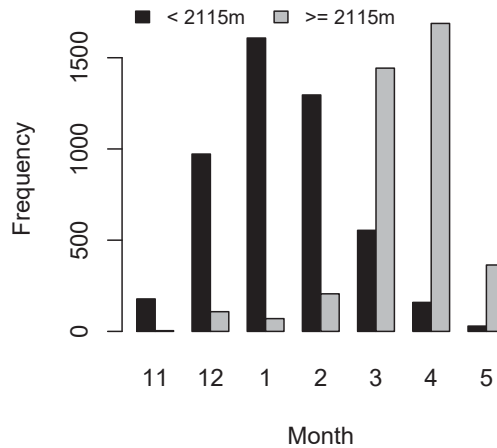
Fig. 2.3: Yearly maximum design snow loads from measurement locations with coverage in every month of the snow season as separated by elevation.

not inadvertently thrown out by the coverage filter.

In addition to this coverage filter, we remove the lowest 10% of maximums at each measurement location prior to distribution fitting. This removal is a practical solution to a systematic problem and the selection of 10% is somewhat arbitrary. The 10% threshold helps us remove low outliers that passed the coverage filters without throwing out excessive amounts of data. The effect of this coverage filter approach is illustrated in Figure 2.4. These quantile-quantile plots show the improved log-normal distribution fit that occurs at Levan, Utah after all screening measures are applied. The effect of this improved fit is a reduction in the estimated design snow load. This example illustrates the importance of screening for artificially low annual snow load maximums prior to estimating design snow loads.

Misreported Values

Retaining only the annual maximums makes the distribution fitting process particularly susceptible to high outliers. Fortunately, the NCEI provides a suite of data measurement and quality control flags to detect such outliers (Durre et al., 2010; NOAA, 2016). We removed all measurements that failed *any* of the NCEI quality assurance flags prior to analysis. We
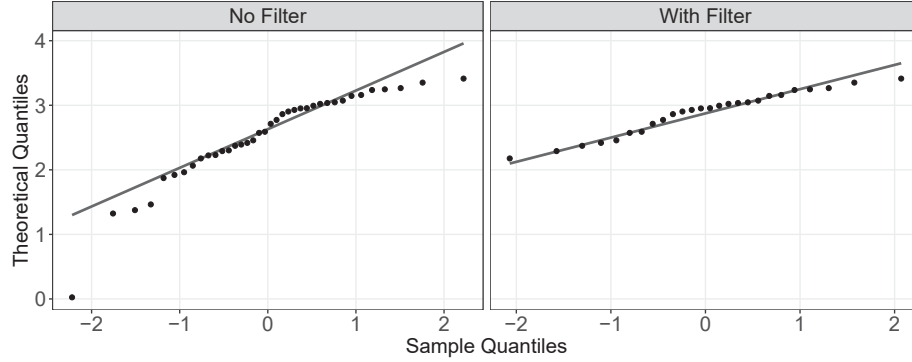
Fig. 2.4: Theoretical log-normal distribution quantiles vs empirical (observed) quantiles for annual maximum snow loads at Levan, Utah with and without coverage filters applied. The coverage filters result in the reduction of the estimated standard deviation from 0.69 to 0.33.

also conducted a manual search for additional outliers by flagging all sets of three consecutive measurements where the estimated design snow load varied by more than 1.44 kPa in a 10 day period. These arbitrary thresholds allowed us to determine candidate outlier points, from which we removed 16 observations at nine measurement locations. Given the millions of observations in the raw data, we felt the number of observations removed was modest.

Metadata Anomalies

Many measurement locations experience small changes in their latitude, longitude, and elevation over time. To handle such cases, we use a measurement location's median latitude, longitude, and elevation values as long as the maximum difference among recorded elevations for the same measurement location is less than 100 meters and the maximum geographic distance between coordinates is less than 10 km. Three stations in Kanosh, Utah, Bright Angel Ranger Station, Arizona, and Colorado National Monument, Colorado, all had measurement locations with elevations varying more than 100 meters during their periods of record. In each case, measurements at the anomalous elevation were treated as a separate measurement location.

Zero Values

Even after applying quality assurance filters, there were still measurement locations in Southern Utah, Northern Arizona and Southern Nevada with zero-valued annual maximum loads. For data to follow a log-normal distribution, all values observations must be *strictly*

positive. We initially tried to assign arbitrary small values to non-snow years to satisfy log-normal distribution requirements. However, these attempts caused measurement location maximums to no longer follow a log-normal distribution, which produced poor estimates of the distribution's upper tail. This in mind, we required all measurement locations to have at least five years of non-zero maximums and only used non-zero maximums when fitting the log-normal distribution. Section A.6 describes an alternative method of fitting a log-normal distribution that still considers zero-valued loads.

Estimating 50 Year Recurrence Intervals

The `fitdistrplus` package (Delignette-Muller and Dutang, 2015) is used to fit a log-normal distribution to the yearly maximum values at each measurement location via maximum likelihood (ML) estimation. The estimated distribution parameters are then used to determine the 98th percentile of the distribution, which is defined as the design (i.e. 50 year) snow load for that measurement location. The use of the log-normal distribution to describe the probabilities of maximum snow events is well established and used in the snow load reports of Colorado, Montana, Oregon, and New Mexico (Sack, 2015). Other right-skewed probability distributions have also been used, such as the log-Pearson III in SEAU (1992) and Al Hatailah et al. (2015). The use of formal goodness of fit tests to select distributions are not helpful and perhaps detrimental in this context, as a "good" fit of the bulk of the data in no way guarantees an appropriate fit for the extreme right tail of the distribution.

In an effort to guard against invalid design load estimates, we only include measurement locations with at least 12 years of record prior to removing the lowest 10% of maximums. This minimum sample size is similar to the thresholds used in Idaho (Al Hatailah et al., 2015) and Colorado (DeBock et al., 2017). These relatively small thresholds for the distribution fitting process reflect practical efforts on the part of researchers to produce reasonable 50 year estimates without excluding measurement locations with moderate periods of record. However, these thresholds also mean that we are often trying to predict snow accumulation events larger than has ever been recorded at the measurement location. This extrapolation,

coupled with known sensitivity issues in the distribution fitting process, reminds us that the 50 year estimates are at best, "... one[s] of good faith and not statistical in nature" (Scholz, 1995). See Section 3.4.3 for an expanded discussion of the limitations of 50 year estimates.

2.1.2 The 1992 Utah Dataset (UT-1992)

These data consist of 413 measurement locations (210 SC, 203 COOP), all located in Utah. The method used to calculate the log-Pearson type III parameters is not specified. Depth-to-load conversions using the RMCD were occasionally adjusted when the resulting snow water equivalents exceeded the measurement location's winter cumulative precipitation.

SEAU (1992) provides a copy of these data but does not provide precise measurement location information. Since 1979, many of the SC stations used in this report have been discontinued and precise location information is unavailable. Measurement location information was determined for all but seven locations through a combination of station number matching in NRCS and NOAA station databases, as well as personal contact with Randall Julander at the Utah Snow Survey Office in Salt Lake City. Locations for the seven remaining measurement locations were approximated using Google Earth to determine approximate coordinates given information from the Utah Snow Survey Office and county information given in SEAU (1992).

2.1.3 The 2015 Idaho Dataset (ID-2015)

These data consist of 394 (246 SC/SNOTEL, 148 COOP) Idaho measurement locations with an additional 257 (222 SC/SNOTEL, 35 COOP) measurement locations near the Idaho border. Log-Pearson type III distribution parameter estimates were determined using the method of moments. Al Hatailah (2015) provides these data as well as further details regarding their approach to estimating design snow loads.

2.2 Methods

Each of the following methods predict design snow loads at a state level using design snow loads at surrounding measurement locations as input. These methods were selected due

to their ability to be easily applied to datasets of varying size and location, an important prerequisite for calculating the cross validated errors discussed in Chapter 3. For comparative convenience, the primary methods of consideration are defined using a common set of notation. Let $q_s(\boldsymbol{x})$ denote the design snow load at a location $\boldsymbol{x}$ (with q_s^* representing the estimated design snow load) and let $A(\boldsymbol{x})$ denote location elevation. Further, let $\boldsymbol{x}_\alpha$ represent a measurement location $\alpha\,(\alpha = 1, \cdots, N)$ and let $D(\boldsymbol{x}_i, \boldsymbol{x}_j)$ represent the geographic distance between locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

The defining feature of each method is in the way that elevation is accounted for in the design snow load predictions. With the exception of the design snow load equations in SEAU (1992), each of the considered methods use normalized design snow loads (NGSL) or some variant of linear regression. NGSL are calculated as design snow load divided by elevation $\left(\frac{q_s^*(\boldsymbol{x}_\alpha)}{A(\boldsymbol{x}_\alpha)}\right)$. They "appear to mask out the effects of the environment on the snow-making mechanism" and "reduce the entire area to a common base elevation" (Sack et al., 2016). NGSL have a long history of use in western state snow load studies, including the current snow load reports of Idaho, Montana and Washington (Sack et al., 2016).

On the other hand, regression based estimators seek to characterize the log-linear relationship between design snow loads and elevation observed in Figure 2.1. This relationship can be characterized using simple linear regression (LR) defined as

$$\log(q_s^*(\boldsymbol{x})) = \beta_0 + \beta_1 A(\boldsymbol{x}) \tag{2.1}$$

where β_0 and β_1 are calculated using ordinary least squares regression. The cross validated results in the following section show that differences in method accuracy can be largely attributed to differences in the characterization of the elevation/snow load relationship.

2.2.1   Current Utah Law (SNLW)

Table 2.3: Explanation of coefficients used in SNLW.

Coefficient	Description	Value
P_0	Base design snow load at a given elevation	1.4-4.1 kPa
S	Rate at which design snow load changes with elevation	9.896 kPa/km
A	Elevation above sea level at location	km
A_0	Base design snow load elevation	1.25-2.13 km

The 1992 Utah snow load report defined design snow load requirements using the equation referred to hereafter as the "Utah snow law" (SNLW):

$$
q_s^*(\boldsymbol{x}) = \begin{cases} \left(P_0^2 + S^2 \left(A(\boldsymbol{x}) - A_0 \right)^2 \right)^{\frac{1}{2}} & A(\boldsymbol{x}) > A_0 \\ P_0 & A(\boldsymbol{x}) \leq A_0 \end{cases}
$$

where the coefficients A, A_0, and P_0 are county specific parameters (SEAU, 1992). Explanations and ranges (as converted from the original English units) for the county specific parameters are given in Table 2.2.1.

Load observations by building officials and others since the formation of these equations prompted the Utah legislature to release updated snow load requirements for select cities in the state (Utah Legislature, 2016), generally resulting in a reduction of design snow load requirements at these locations. Further discussion regarded these amended requirements is provided in Section 3.1.

The county specific coefficients attempt to address the highly diverse climate of Utah. However, these varying coefficients can, at times, lead to significant discrepancies in snow load predictions along county borders, particularly at high elevations. A great example of this can be found along the border of Cache and Box Elder counties at the point US highway 89 crosses between counties (1800 m in elevation). A structure built in Cache County at this point would have a snow load requirement of $\left(2.394^2 + 9.896^2 \left(1.8 - 1.372 \right)^2 \right)^{0.5} = 4.87$ kPa, while a structure built in Box Elder County would have a snow load requirement of $\left(2.0594^2 + 9.896^2 \left(1.8 - 1.585 \right)^2 \right)^{0.5} = 2.96$ kPa, approximately 40 percent lower than the Cache county requirement. Discrepancies like this are one of the reasons the state of Utah

chose to update their snow load requirements in 2018 (Bean et al., 2018).

2.2.2 Parameter-elevation Regressions on Independent Slopes Model (PRISM)

PRISM was originally developed by Chris Daly in the early 1990's to produce computerized climate maps as a satisfactory replacement to maps hand made by climate experts. Unlike traditional interpolation methods which only account for measurements of the primary (and perhaps one secondary) variable, PRISM is designed to think like a climatologist expert: taking into account a variety of climatic factors known to influence the variable of interest (Daly and Bryant, 2013).

Framework

In order to create a continuous map of design snow load requirements, we must use surrounding measurement location data to predict design snow loads between measurement locations. Such data would confirm that a rise in elevation tends to be associated with a rise in design snow loads. However, a model that only uses elevation to predict design snow loads fails to account for other important climate factors. This in mind, PRISM fits a unique linear model to each area of interest, giving measurement locations most relevant to the area of interest higher influence during model fitting through a series of user-defined weights.

The use of PRISM to predict precipitation and temperature is well established (see prism.oregonstate.edu) and is used as part of Oregon's most recent snow load report, made in partnership with the Northwest Alliance for Computational Science and Engineering (NACSE, 2012). The Oregon report generates 30 year mean snowfall predictions with PRISM and uses these predictions to estimate design snow loads. We alternatively use PRISM to directly predict design snow loads. In addition, recall from Figure 2.1 that design snow loads share a log-linear relationship with elevation. For this reason we make PRISM predictions on a log-scale and exponentiate predictions for final load estimates.

Weights

Weights provide a way to account for additional factors influencing snow loads beyond elevation in a particular region. Daly et al. (2008) defines the PRISM weighting scheme as

$$w = w_c \left[F_d w_d^2 + F_z w_z^2 \right]^{\frac{1}{2}} w_p w_f w_l w_t w_e, \tag{2.2}$$

where w_c, w_d, and w_z represent the cluster, distance, and elevation weights. F_d and F_z are scalars defining the importance of the distance and elevation weights and must sum to one. The additional weights w_p, w_f, w_l, w_t, and w_e represent the coastal proximity, topographic facet, vertical layer, topographic position, and effective terrain weights.

We use an adaptation of (2.2) defined as

$$w = w_c \left[F_d w_d^2 + F_z w_z^2 \right]^{\frac{1}{2}} w_b,$$

where w_b represents a basin weighting factor. Each weighting vector must individually and collectively sum to one. We provide summaries of each weight as follows:

Distance Weighting

The closer a measurement location lies to the area of interest, the more weight that measurement location receives. Measurement locations within a user minimum radius of influence (r_m) receive full weight, while measurement locations outside the radius of influence receive a weight inversely proportional to their geographic distance (d) to the area of interest. This is represented as

$$w_d = \begin{cases} 1 & d - r_m \leq 0 \\ \frac{1}{(d-r_m)^a} & d - r_m > 0 \end{cases}$$

where (a) allows the user to adjust the shape of the distance function (Daly et al., 2008).

Elevation Weighting

Measurement locations are given more weight if their elevation is similar to the elevation of the area of interest. A measurement location's absolute elevation difference (Δz) is compared to user specified minimum (Δz_m) and maximum (Δz_x) elevation thresholds. This

is represented as

$$w_z = \begin{cases} \frac{1}{\Delta z_m^b} & \Delta z \leq \Delta z_m \\ \frac{1}{\Delta z^b} & \Delta z_m < \Delta z < \Delta z_x \\ 0 & \Delta z \geq \Delta z_x \end{cases}$$

where (b) is an weighting factor that allows the user to adjust the shape of the elevation function (Daly et al., 2002).

Cluster Weighting

"Cluster weighting seeks to limit the influence of stations that are clustered with other nearby stations, which can lead to over-representation in the regression function" (Daly et al., 2008). This weight is defined as

$$w_c = \frac{1}{1 + s_c}$$

with

$$s_c = \sum_{j=1}^{n} h_{ij} v_{ij}$$

where (h_{ij}) and (v_{ij}) represent the horizontal and vertical cluster factors between measurement locations i and j respectively.

Any pair of measurement locations i and j that have a geographic distance (d_{ij}) between them that is within 20% of the minimum radius of influence (r_m) will have a non-zero horizontal cluster factor, i.e.

$$h_{ij} = \begin{cases} 0 & d_{ij} > .2r_m \\ \frac{.2r_m - d_{ij}}{.2r_m} & 0 \leq d_{ij} \leq .2r_m \end{cases} .$$

Any pair of measurement locations i and j that have an absolute elevation difference (e_{ij}) between them that is within a user-defined elevation precision (p) will have a non-zero
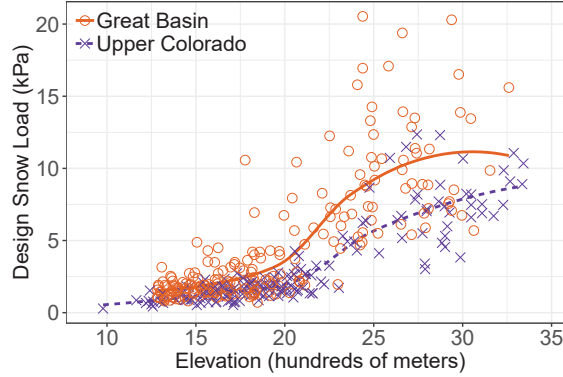
Fig. 2.5: Comparison of design snow loads vs elevation in the two major water basins of Utah, illustrating how the elevation/snow load relationship changes across water basins.

vertical cluster factor, i.e.

$$v_{ij} = \begin{cases} 0 & (e_{ij} - p) > p \\ \frac{2p - e_{ij}}{p} & (e_{ij} - p) \leq p \end{cases}$$

Basin Weighting

Most PRISM implementations have a set of weights devoted to handling the influence of mountainous terrain on climate patterns. Mountain ranges ultimately govern the flow of water and a map of watershed boundaries is a quick way to identify the major mountains and valleys of a region. Watershed data is readily available through the U.S. Geological Survey (USGS, 2019b) and defined by a series of Hydrologic Unit Codes (HUC). These HUCs define a hierarchy of water basins using up to 12 digits, with each pair of two digits identifying a sub-basin within the previous two-digits (reading left to right). Justification for these basin weights is demonstrated by the observable difference in design snow load/elevation profiles of the Great Basin Watershed (West of the Wasatch Front) and Upper Colorado Basin Watershed (East of the Wasatch Front) as observed in Figure 2.5. This figure shows (as modeled by a linear loess smoothing curve) that measurement locations on the west side of the Wasatch Front have a more drastic increase in design snow loads as elevation increases.

This observation highlights the need to weight measurement locations according to the similarity of their watersheds to the area of interest. We therefore create a water basin
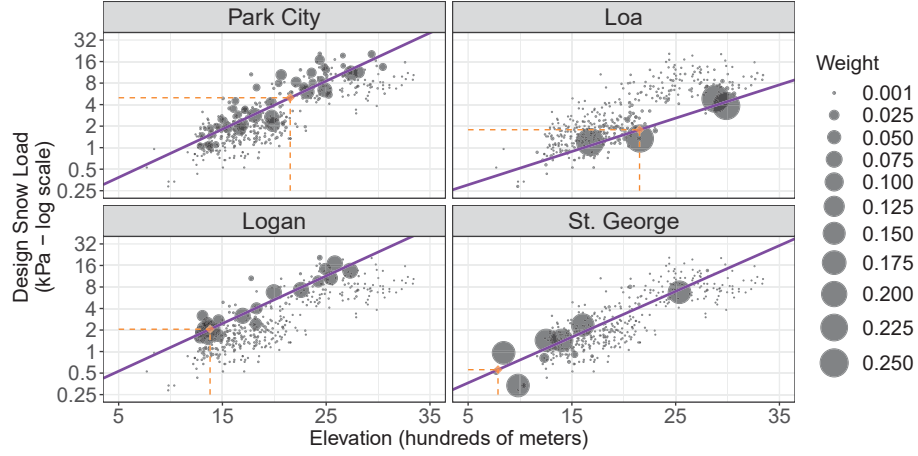
Fig. 2.6: Illustration of PRISM predictions at select locations in Utah.

weight with equation

$$w_{b_i} = \left(\frac{s_i + 1}{5}\right)^c,$$

where s represents the number of common watersheds (four levels ranging from HUC 2 through 8) shared by measurement location i and the target grid cell and c is a user-defined weighting factor that changes the shape of the weighting function.

Illustration

Figure 2.6 illustrates PRISM predictions at four locations in Utah. Notice that that slope of the linear model changes at each location to best fit highest weighted points. This reaffirms the important idea that each PRISM prediction uses a different linear model. These models can be heavily influenced by the combination of parameters selected for each of the weighting functions. Approaches and implications for PRISM weighting parameter selection are discussed further in Chapter 3.

2.2.3   Idaho's NGSL Based on Inverse Distance Weighting (IDW)

In IDW, the predicted design snow load at a particular location is a weighted average of the NGSL of surrounding measurement locations, multiplied by the location's elevation. Adapting the original notation given by Shepard (1968), this method is defined for snow

load prediction as

$$q_s^*(\boldsymbol{x}) = \frac{A(\boldsymbol{x})}{\sum_{\alpha=1}^{N} D\left(\boldsymbol{x}_\alpha, \boldsymbol{x}\right)^{-c}} \sum_{\alpha=1}^{n} \left[D\left(\boldsymbol{x}_\alpha, \boldsymbol{x}\right)^{-c} \frac{q_s^*(\boldsymbol{x}_\alpha)}{A(\boldsymbol{x}_\alpha)} \right].$$

The variable $c > 0$ allows for adjustments to the weighting factor, with larger values of c accelerating the weight decay as distance increases. This method is an exact interpolator, meaning if a prediction location exactly matches a measurement location (i.e. $D\left(\boldsymbol{x}_\alpha, \boldsymbol{x}\right) = 0$), the IDW method will simply predict using the NGSL of the matching measurement location.

Idaho's implementation of IDW separates the state and measurement locations into two layers above and below 1,219 m. Predictions in the lower layer use $c_1 = 2$ and predictions in the upper layer use $c_2 = 6$ (Al Hatailah et al., 2015). One difference in our implementation is the use of geographic distances rather than euclidean distances from the Idaho Transverse Mercator Projection (Al Hatailah, 2015). The use of geographic distances eliminates the spatial distortion that may occur when applying a euclidean based map projection to a larger geographical area.

2.2.4 Linear Triangulation Interpolation (TRI)

The TRI method partitions the area of interest into a set of non-intersecting triangles with vertices at each measurement location. Predictions use a weighted average of the NGSL at the three measurement locations forming the triangle overlaying the point of interest (Akima, 1978). The R implementation of this strategy creates a grid of predicted values within the convex hull of the given data points (Akima and Gebhardt, 2015)). This leads to missing value predictions at outer locations that do not fall within the convex hull. These missing values are ignored when computing cross validated errors in Chapter 3.

2.2.5 Kriging (SKLM and UK)

The `gstat` R package (Gräler et al., 2016) provides a numerical implementation of many kriging variations. Details regarding these family of estimators are given in Goovaerts (1997). One kriging extension of (2.1) is called simple kriging with varying local means

(SKLM) (Goovaerts, 2000) defined symbolically as

$$\log(q_s^*(\boldsymbol{x})) = \beta_0 + \beta_1 A(\boldsymbol{x}) + \sum_{\alpha=1}^{N} \lambda_\alpha(\boldsymbol{x}) r(\boldsymbol{x}_\alpha).$$

This method proceeds in three steps. First, a linear model is calculated identical to Equation 2.1. Then, simple kriging uses the residuals of the linear model to predict a residual value at the location of interest. Finally, this residual value is used to update the original linear model prediction. The simple kriging coefficients $(\lambda_\alpha(\boldsymbol{x}))$ are calculated by solving the kriging system

$$\sum_{\alpha=1}^{N} \lambda_\beta(\boldsymbol{x}) C_R \left(D(\boldsymbol{x}_\alpha, \boldsymbol{x}_\beta) \right) = C_R \left(D(\boldsymbol{x}_\alpha, \boldsymbol{x}) \right) \quad \beta = 1, \cdots, n.$$

where C_R represents the covariance between any two observations and is assumed to be a function of distance. More often, this system is solved using semi-variances $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$, assuming the covariance exists. Semi-variances are typically preferred as semi-variances can be well defined even in cases where the covariance is not. These semi-variances are modeled using variograms. A theoretical variogram is often used to approximate the empirical variogram defined as

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N_{\boldsymbol{h}_1}} \sum_{\alpha_{\boldsymbol{h}}=1}^{N_{\boldsymbol{h}}} \left[r(\boldsymbol{x}_{\alpha_{\boldsymbol{h}_1}}) - r(\boldsymbol{x}_{\alpha_{\boldsymbol{h}_2}}) \right]^2 \tag{2.3}$$

where $\left[r(\boldsymbol{x}_{\alpha_{\boldsymbol{h}_1}}), r(\boldsymbol{x}_{\alpha_{\boldsymbol{h}_2}}) \right]$ represents each pair of regression model residuals located $||\boldsymbol{h}||$ distance away from each other (Goovaerts, 1997). Figure 2.7 provides an example of the empirical and associated theoretical variograms for each dataset.

An alternative method for accounting for elevation in kriging predictions is through universal kriging (UK), or kriging with an external drift, which calculates the trend implicitly within the kriging system, rather than separately as in SKLM (Goovaerts, 1997). When
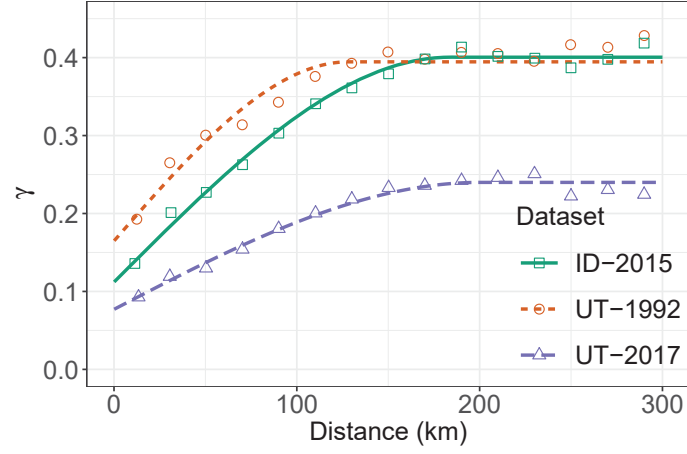
Fig. 2.7: Empirical (points) and theoretical (lines) semivariograms for each of the three datasets.
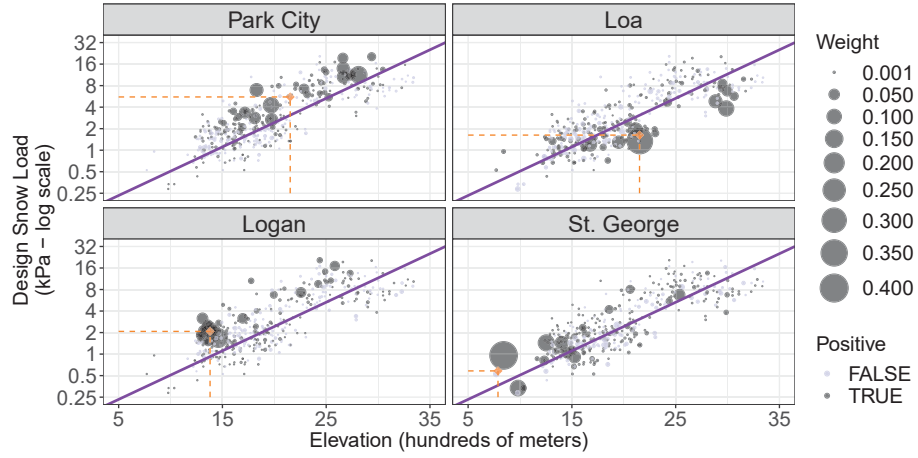
Fig. 2.8: Illustration of UK predictions at select locations in Utah.

elevation is the only trend coefficient, the universal kriging estimates are equivalent to

$$\log(q_s^*(\boldsymbol{x})) = \beta_0^* + \beta_1^* A(\boldsymbol{x}) + \sum_{\alpha=1}^{N} \lambda_\alpha(\boldsymbol{x}) r(\boldsymbol{x}_\alpha)$$

where β_0^* and β_1^* are calculated using generalized least squares regression based on the assumed spatial covariances. Figure 2.1 showed the difference in the trend lines resulting from SKLM and UK, while Figure 2.8 shows an example of the UK predictions at the same four locations used in the PRISM example. Notice that the linear model is identical for each location, yet the final predictions do not reside on the line as they did in PRISM.

Kriging predictions provide theoretical estimates of the prediction error uncertainty (often called kriging variance) (Moral, 2010). This kriging variance is based upon the distribution of data in space and is higher for prediction locations with few surrounding measurement locations. This kriging variance can be used to identify areas lacking measurement location density. This is different from the uncertainty inherent in the design snow load estimates at the measurement locations. The contrast between kriging variance and input precision is discussed further in Chapter 4.

2.3 Reflections

This chapter has outlined a standardized method for creating a design snow load dataset. This process highlighted the need for data filtering beyond those automatically provided by the NCEI in order to ensure practical design snow load estimates at measurement locations. We have continued to refine this data processing workflow as illustrated in the (unpublished) Washington snow load study. Portions of this study describing our improved data processing techniques are provided in the Appendix.

It was shown that design snow loads experienced a log-linear relationship with elevation. It is likely that this observed log-linear relationship does not hold outside of the range of observed data. This is particularly true at mountain peaks exceeding 3500 m, where exposure to the wind and sun could result in these peaks having less settled snow than the measurement locations tucked away at slightly lower elevations. Regardless, it is clear that elevation is a defining characteristic of design snow loads. The following chapter discusses how the proper treatment of elevation in design snow load mapping affects the accuracy of the mapping predictions.

CHAPTER 3

COMPARISONS AND CONSIDERATIONS

This chapter is devoted to comparisons and considerations of the data and methods described in Chapter 2. We first compare maps of PRISM and SNLW to visualize the recent updates to Utah's design snow load requirements. We also compare these recently updated requirements to the current design snow load requirements in Idaho and Colorado. Next, we compare the accuracy of all the methods introduced in Chapter 2 through a cross validation analysis on UT-2017, UT-1992, and ID-2015. Finally, we discuss important cautions and considerations for estimating design snow loads.

3.1 Comparing PRISM to SNLW

Figure 3.1 compares PRISM to SNLW. This comparison is particularly important as these PRISM predictions replace the SNLW predictions in the Utah building code. The PRISM predictions are bounded below by the state minimum design snow load requirement ($\approx$1 kPa) and above by the highest design snow load in UT-2017 ($\approx$ 21 kPa). This figure demonstrates that the PRISM predictions are lower than SNLW predictions for the majority of the state, particularly along the Wasatch front. There are a few areas where PRISM predictions are higher than SNLW predictions, including higher elevation areas in the Uintah Basin and in the mountains immediately north of St. George, Utah. Figure 3.2 provides a comprehensive comparison of PRISM, UK, and IDW predictions at cities in Utah with recently amended design snow load requirements (Utah Legislature, 2016). While the motivation for these amendments is not entirely clear, it is assumed that the amendments resulted from site-specific studies intended to improve upon the SNLW requirements. The sensitivity of the considered mapping techniques is illustrated by making separate predictions using UT-2017 and UT-1992. In nearly every case using both datasets, the amended requirements are more consistent with the PRISM and UK predictions than the SNLW predictions. This
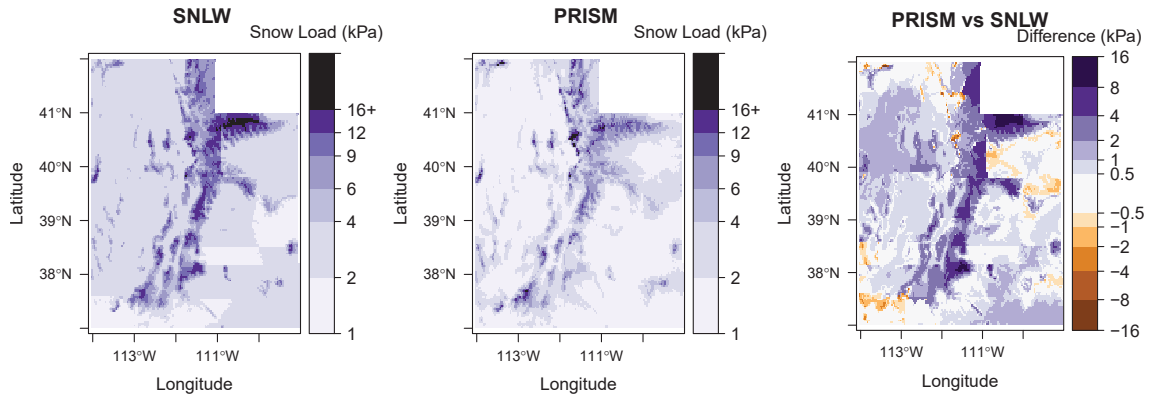
Fig. 3.1: Maps comparing design snow load predictions using PRISM and SNLW.

general agreement of PRISM and UK with the amended requirements provides evidence that these load requirements are an improvement from those outlined by SNLW.

We are most interested in instances in Figure 3.2 where the amended snow load requirements differ greatly from PRISM estimates. For example, PRISM estimates in Coalville (elevation 1,700 m) and Kamas (1,977 m) are nearly half those required in the amendments (2.11 vs 4.12 kPa and 2.68 vs 5.46 kPa). However, there is a consensus among the various mapping techniques that the design loads for these locations should be reduced. This consensus suggests that the previous design snow load requirements in Summit County are intentional over-predictions as explained further in Section 3.3.2.

Monticello, Utah is another interesting location as it is the only city for which each method predicts higher design snow loads than both the original and amended requirements. Monticello sits at the base of the Abajo Mountains at an elevation of 2150 m, nearly 300 m higher than its closest neighbor Blanding, Utah (1860 m) 30 km south. Both measurement locations have nearby measurement locations and the Monticello measurement location has a much higher design load (4.4 vs 2 kPa). It is difficult to determine if this drastic difference in design loads is legitimate given the sparsity of data in the region. This in mind, the available information in light of this data scarcity justify increases in the predicted design
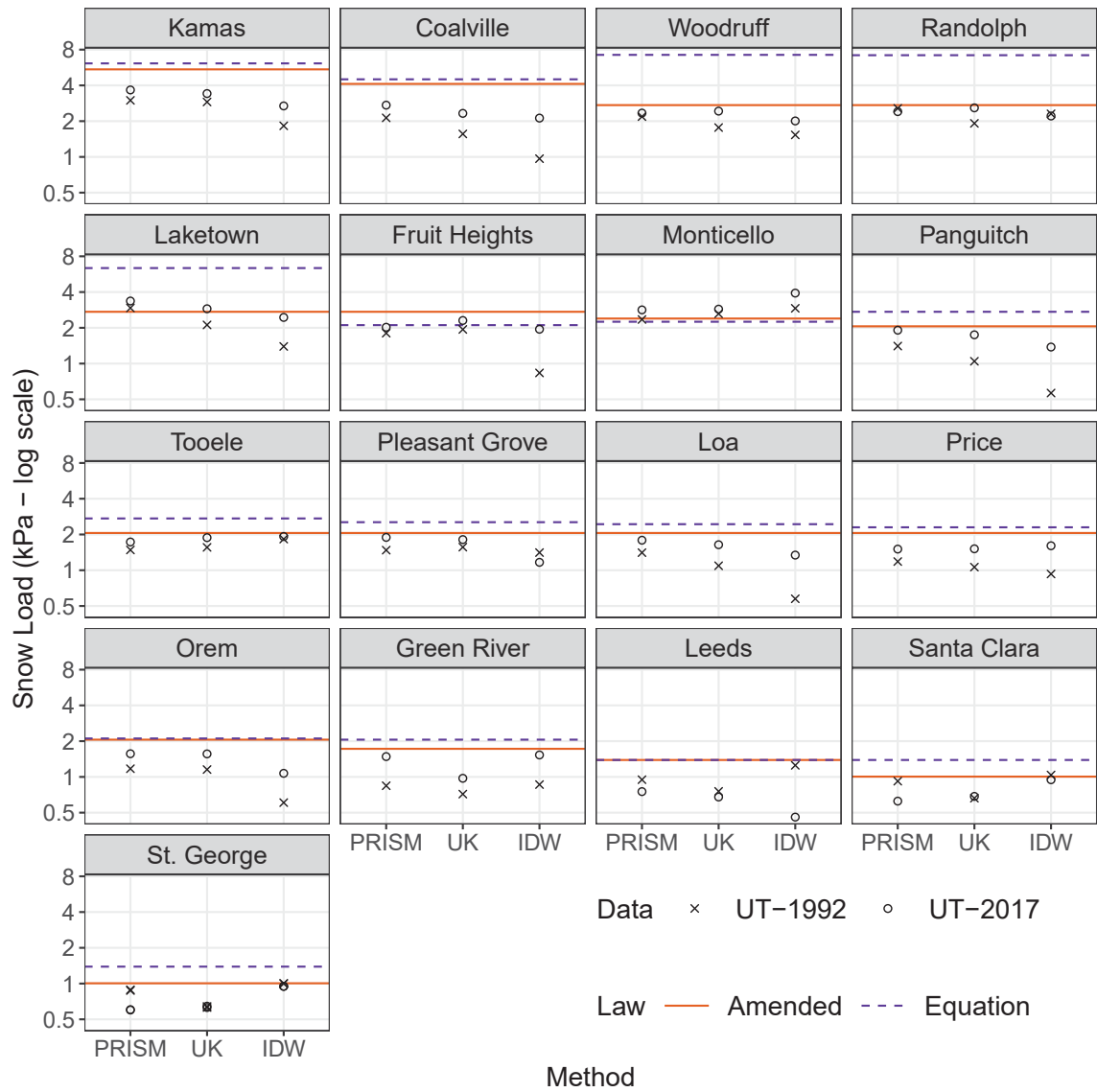
Fig. 3.2: A comparison of design snow load predictions using PRISM, UK, and IDW to the 2016 amendments and SNLW. The use of the log-scale emphasizes differences in predictions at locations with lower design snow loads.
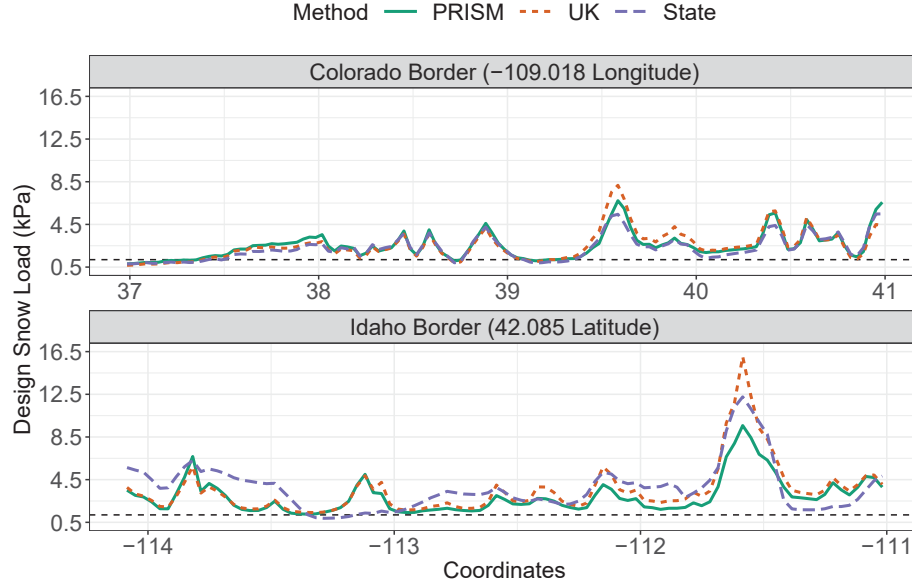
Fig. 3.3: Comparisons of Utah design snow load requirements along the Colorado and Idaho borders.

loads using PRISM or UK (approximately 3.2 kPa instead of 2.39 kPa).

3.2 Border Comparisons

We also compare our design snow load estimates to the estimates made by Colorado and Utah. Colorado snow load values were obtained using the contour map and snow load equation included in the Colorado snow load report (Torrents et al., 2016). Idaho predictions were obtained using our implementation of IDW with ID-2015 data. PRISM and UK estimates using UT-2017 had mean absolute differences of only 0.29 and 0.42 kPa respectively when compared to Colorado predictions. Predictions along the Idaho border did not share the same level of agreement, with mean absolute differences of 1.42 kPa and 1.22 kPa respectively. Figure 3.3 confirms that the Idaho predictions share less agreement with PRISM and UK than the Colorado predictions.

3.3 Cross Validation

Recall that each state uses their own mapping technique to predict design snow loads

between measurement locations. Because these techniques were independently developed, there has been no formal comparison of accuracy between methods. Cross validation is a common tool used for model selection and refinement in many disciplines (Arlot and Celisse, 2010), including structural engineering (Chang et al., 2017). We use 10-fold cross validation to compare the methods described in Chapter 2 using the datasets described in this same chapter. Cross validated errors are defined as

$$\text{Err}(\boldsymbol{x}_\alpha) = q_s(\boldsymbol{x}_\alpha) - \hat{q}_s(\boldsymbol{x}_\alpha)$$

where $q_s(\boldsymbol{x}_\alpha)$ and $\hat{q}_s(\boldsymbol{x}_\alpha)$ are the actual and predicted design snow loads at measurement location $\boldsymbol{x}_\alpha$ respectively. Defined in this way, a positive error indicates under-predictions and a negative error indicates over-predictions.

Cross validated comparisons of a single model are often performed using the scale of the modeled response variable. Thus, if a log-transformation has been applied to the response, the cross validated errors would be reported on the log-scale. In this setting, we are comparing multiple models with different transformations of the design snow load (i.e. the response variable). For this reason, all cross validated errors in this section are reported in the original scale. Given the exponential nature of snow, these cross validated errors are heteroscedastic and occasionally very large at higher elevations. Figure 3.4 shows an example of these heteroscedastic errors on UT-2017.

We use the mean absolute error (MAE) and mean error (ME) as the primary means of summarizing errors. These are defined (similar to Maguire et al. (2014)) as

$$\text{MAE} = \frac{1}{N} \sum_{\alpha=1}^{N} |\text{Err}(\boldsymbol{x}_\alpha)|$$
$$\text{ME} = \frac{1}{N} \sum_{\alpha=1}^{N} \text{Err}(\boldsymbol{x}_\alpha)$$

where N represents the total number of measurement locations with design snow load measurements and $\hat{q}_s(\boldsymbol{x}_\alpha)$ represents model predictions for each measurement location $\boldsymbol{x}_\alpha$. Given the inherent skewness in the errors. We also consider the median absolute error
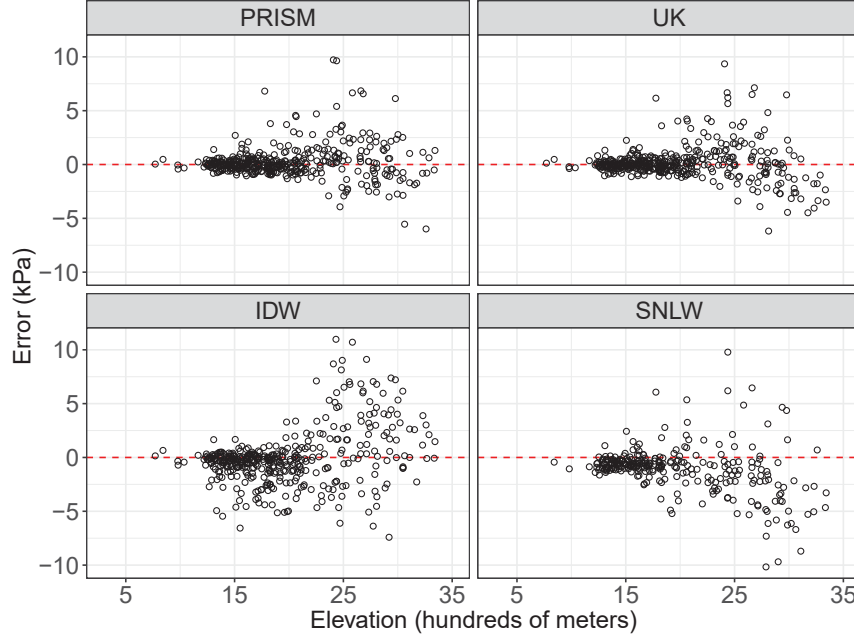
Fig. 3.4: Scatterplot of cross validated errors for select methods on UT-2017, highlighting the heteroscedasticity of the errors on the original scale.

(medAE) and the root mean squared error (RMSE) defined as

$$\text{medAE} = \{|\text{Err}(\boldsymbol{x}_k)| : P\left(|\text{Err}(\boldsymbol{x}_\alpha)| \leq |\text{Err}(\boldsymbol{x}_k)|\right) \leq 0.5\}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{\alpha=1}^{N} \text{Err}(\boldsymbol{x}_\alpha)^2}.$$

The medAE is less sensitive than the MAE to large errors, while the RMSE more sensitive. Reporting each measure allows us to determine how the heteroscedasticity influences in the final model comparisons. We also perform cross validation 100 times in order to measure the sensitivity of each measure to the random separation into the 10 groups.

3.3.1 Parameter Selection

Many of the model parameters described in Chapter 3 must be specified prior to prediction. Some of these parameters are physically based and could be selected using prior experience or expert opinion. This heuristic approach seems to guide the selection of parameters in the Idaho Snow Load Study (Al Hatailah, 2015). For this reason, we elect

Table 3.1: Weight parameters used for final predictions in Utah.

Parameter	Description	Final	Recommended
F_d	Distance weighting importance	.8	.8
r_m	Minimum radius of influence	50km	30-100km
a	Distance weighting exponent	2	2
b	Elevation weighting exponent	1	1
z_m	Minimum elevation threshold	100m	100-300m
z_x	Maximum elevation threshold	1500m	500-2500m
p	Elevation precision	100m*	n/a
d	Basin weighting exponent	2 (2)	n/a

* assumed $p = z_m$

to use identical parameters for IDW on each considered dataset. One alternative approach would be to select parameters via cross validation. For example, PRISM parameters in Bean et al. (2017) were selected using an eight-dimensional grid search, looking for the combination or parameters that minimized or nearly minimized the MAE. This eight-dimensional grid search was eventually reduced to four dimensions in Bean et al. (2018) by fixing parameters for which Daly et al. (2002) suggested only one value. The final parameters selected for the Utah Snow Load study are provided alongside the recommended values from Daly et al. (2002) in Table 3.3.1

Note that the tuning of PRISM in Bean et al. (2017) was performed prior to reporting the cross validation results. When the number of parameter combinations in the grid search is large, the cross validation tuning approach has the potential to over-fit the dataset of interest and produce smaller cross validation errors than would be seen on new data. To avoid over-fitting, Bean et al. (2019) used the same set of tuned parameters obtained in Bean et al. (2017) when making comparisons on UT-1992 and UT-2015. We alternatively tune PRISM as part of the cross validation process, selecting a new set of parameters when predicting for each fold of the data. The numerical implementation of this auto-tuning approach minimizes the MAE on the available data using the scale of the response variable. Thus, unlike the original parameter tuning, the auto-tuning minimizes the MAE for the log-transformation of design snow load predictions. Table 3.2 shows that the MAE using PRISM is fairly insensitive to our selection of parameters, although the range and values

Table 3.2: Comparison of the minimum (min), median (med) and maximum (max) MAE obtained from 100 iterations of cross validation with PRISM using the auto-tuning method, vs the original single-tune method.

| | UT-2017 | | | UT-1992 | | | ID-2015 | | |
	min	med	max	min	med	max	min	med	max
Single Tune	0.886	0.927	1.014	1.159	1.212	1.268	1.631	1.679	1.743
Auto Tune	0.891	0.942	1.015	1.207	1.273	1.505	1.539	1.625	1.939

of the MAE tend to be slightly larger with the auto-tuning approach. This insensitivity is partially due to the fact that the original parameters were partially determined using recommendations from Daly et al. (2002), rather than using a strict model tuning. The slight increase in MAE observed in Table 3.2 can be attributed to two sources. First, the auto-tuning approach selects the best combination of parameters for the MAE of the log-scale predictions. The log-transformation reduces the influence that large errors at high elevations have on the final results. Thus the "best" combination of parameters on the log-scale might not be the best parameters on the raw scale. Second, the auto-tuning approach only uses the available portion of the dataset at any step in cross validation, while the original single-tune approach used all observations. We expect a tuning method using all observations to be more robust than a tuning method using only a subset of observations. Tuning within cross validation provides a more realistic estimate of the model error as the chance for over-fitting the available data is reduced.

A similar approach could be applied to the variogram fitting required in UK and SKLM. The most recent implementation of these two models allows for an automatic fitting of the variogram at each stage of cross validation. However, Bean et al. (2019) demonstrated that the MAE is likewise fairly insensitive to modest changes in the variogram parameters. For example, the average MAE (over 100 iterations of cross validation) for ID-2015 and UT-1992 using the dataset-specific variograms in Figure 2.7 are within 0.01 kPa of the MAE using the UT-2017 variogram.

3.3.2 Error and Elevation

The locally weighted regression (loess) (Cleveland and Devlin, 1988) curves in Figure

3.5 reveal the elevation dependent structure of the error scatter-plots previously shown in Figure 3.4. These curves compute local weighted averages of raw and absolute measurement location errors across elevation and map these local averages as smooth polynomial curves. The gray tick marks drawn between each set of plots represent the elevations of the individual measurement locations. These tick marks help to visualize measurement location density across elevation. This characterization of density gives context to plotted curves, as the loess estimates will be more reliable at elevations with a higher density of measurement locations.

Figure 3.5 shows that PRISM, SKLM, and UK are fairly unbiased at low elevations (2000 meters or less) and tend to under-predict at higher elevations (2000 - 3000 meters). The errors of all methods are very unstable in ID-2015 at high elevations. The sinusoidal shape of the ME curves for IDW reveal the tendency of this method to over-predict design snow loads at low elevations and under-predict at high elevations. This behavior is a result of the correlation between NGSL and elevation discussed further in Section 3.4.2. Finally, Figure 3.5 shows the strong tendency of SNLW to over-predict design snow loads. In terms of relative errors, the Utah equations on average predict design snow loads 34% higher than measurement location design snow load estimates from UT-2017 and 57% higher than estimates from UT-1992 (with median relative errors of 25% and 41% respectively). Recall that Equation 2.2.1 was intentionally designed to over-predict design snow loads and it is no surprise that this method would have higher cross validated errors when compared to models designed to minimize error. However, these accuracy comparisons are still useful as they quantify the magnitude of the over-prediction of design snow loads using SNLW. Such over-predictions are understandable when considering the consequences of under-predictions discussed in Chapter 1. However, we agree with Nowak and Collins (2012) that load estimates should be as accurate and reliable as possible, with conservative adjustments being made to load predictions through the selection of load factors from a proper reliability analysis.

3.3.3   Accuracy Comparisons

Figure 3.6 compares the ME, medAE, MAE, and RMSE for each method over each dataset. Error measures for SNLW on ID-2015 are excluded as SNLW predictions are Utah
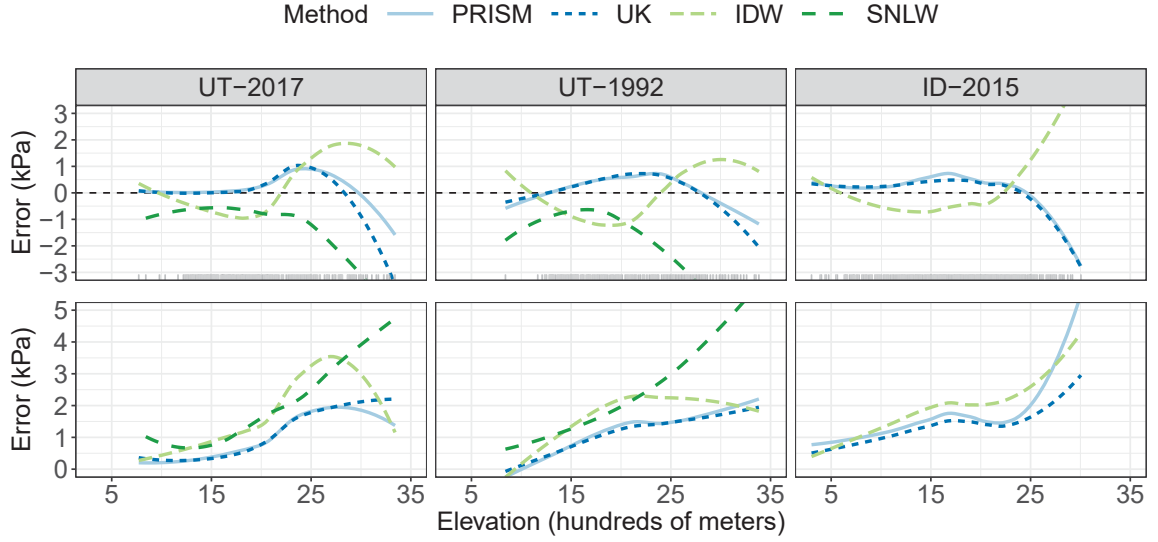
Fig. 3.5: Smoothed errors and absolute errors for each considered dataset. The gray tick marks plotted along the x-axis of the three upper figures denote the individual measurement location elevations.

specific. The points represent the median measure from the 100 iterations of cross validation, while the whiskers represent the range. In terms of MAE, Figure 3.6 shows that PRISM, SKLM and UK notably outperform all other methods on both Utah datasets, with an MAE approximately 40-45% lower than SNLW and IDW on UT-2017. These improvements are not as pronounced for ID-2015, likely due to the less pronounced log-linear relationship between design snow loads and elevation. Likewise, differences in RMSE across methods is also not as drastic as the MAE results, illustrating the difficulty that all methods have at appropriately predicting snow loads at high elevations. The very large range of RMSE measures for PRISM on UT-1992 and ID-2015 reveal the sensitivity of this method to very poor predictions at a small subset of locations. These sensitivities are further explained in Section 3.4.1. Further, the relatively poor performance of IDW on the Utah datasets is largely due to the inability of NGSL to account for the effect of elevation on Utah snow loads, as explained in Section 3.4.2.

UK has a noticeably lower medAE, MAE, and RMSE than all other methods on ID-2015. The competitive errors obtained by UK across all considered datasets are even more
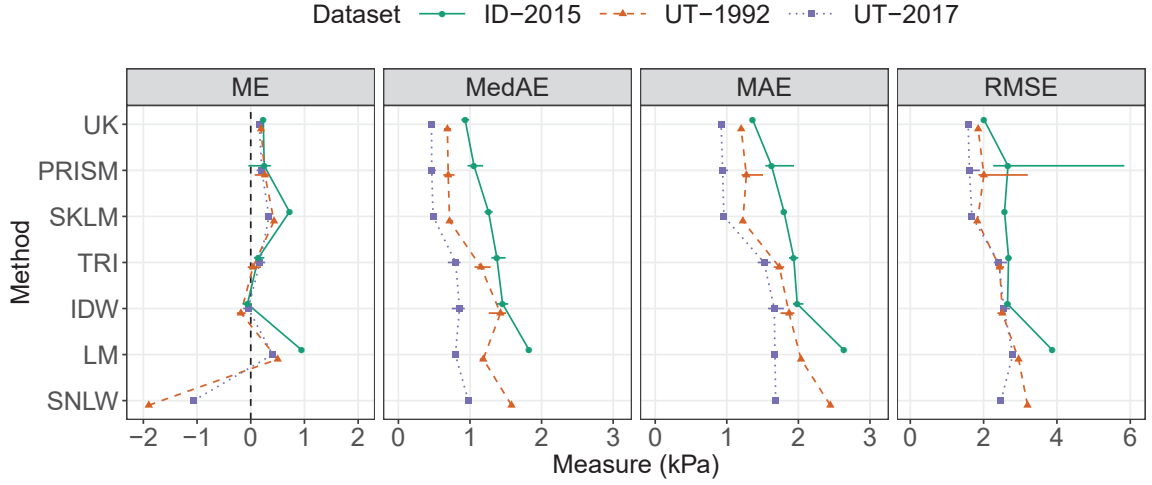
Fig. 3.6: Barchart of mean errors (ME), median absolute errors (medAE), mean absolute errors (MAE) and root mean square errors (RMSE) of spatial prediction methods on each dataset.

compelling given the very little tuning required to implement UK on new data. This in mind, we recommend UK as the best considered mapping technique in terms of its ability to fit the input data.

3.4 Practical Limitations

It is critical that these design snow load predictions and accuracy comparisons be placed in the context of observational limitations. Each mapping technique relies on accurate estimates of design snow loads within each dataset, which are subject to various sources of uncertainty. These datasets also contain a disproportionate number at elevations high above most populated locations in each state. Future research will likely involve adaptations to traditional cross validation measures that emphasize accuracy in the areas of greatest concern. Further, it is not clear if the best methods for mountainous terrain will be the best for design snow load predictions in other terrains. The following subsections discuss some notable limitations associated with predicting design snow loads.

3.4.1 Limitations of Regression-Based Estimators

There are extrapolation issues for the regression based estimators (PRISM, SKLM, UK, and LR) when attempting to predict snow loads at locations with elevations far exceeding all nearby measurement location elevations. In Utah, these situations most often occur at mountain peaks lacking measurement locations. In such cases, these estimators begin to predict unreasonably high snow load values, exceeding all observed snow load values in the dataset. This issue is resolved by restricting the regression-based predictions to extend no higher than the largest design snow load in the input dataset. In addition, the prediction of the global trend (as used in SKLM, UK, and LR) is not allowed to extend beyond the predicted trend for the highest elevation measurement location in the dataset. Such constraints are only imposed when creating design snow load maps and are not imposed for the cross validation results presented in this paper.

Further, the alarming RMSE range for PRISM on ID-2015 in Figure 3.6 reveal that certain combinations of PRISM weights create instances of local extrapolation. This illustrated by way of example at Robinson Lake, Nevada in Figure 3.7, which is one of six locations in ID-2015 where severe over-predictions were observed during cross validation. If this measurement location is removed from the data and the auto-tune method were to select the maximum possible basin weight parameter of five, PRISM assigns virtually all weight to only two measurement locations in ID-2015. These two measurement locations are less than 10 km apart with similar elevations, yet have very different design snow loads. This drastic difference is likely a result of the distribution fitting process, or a measurement bias at one of the locations, rather than a true difference in loads. Robinson Lake has an elevation higher than these two measurement locations and the local extrapolation results in a severe over-prediction. This issue can be avoided by reducing the basin weighting exponent or shifting all snow load values by 1 kPa prior to taking the log-transformation. This arbitrary shift has the effect of moderating the calculated slopes and thus reducing the severity of the extrapolation, yet is difficult to justify. Ekwaru and Veugelers (2018) explore ways in which an optimal shift could be selected empirically, but such an approach adds yet another parameter for the user to estimate. Future work will involve exploring
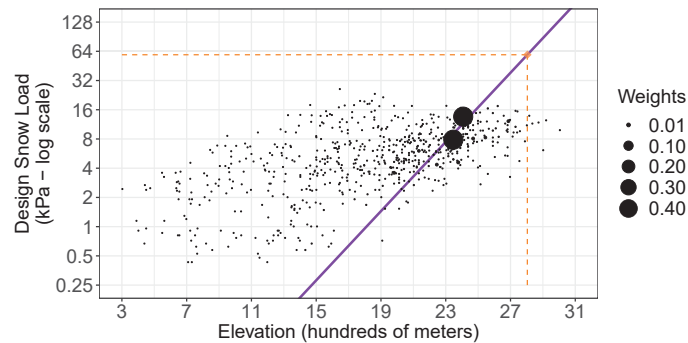
Fig. 3.7: Example of a severe over-prediction using PRISM at Robinson Lake, Nevada.

optimal transformations for regression predictions, as well as non-parametric alternatives to regression based estimators. Most importantly, the described scenario illustrates a potential shortcoming of PRISM for inappropriate combinations of weight parameters.

3.4.2 Limitations of NGSL-Based Estimators

Figure 3.8 reveals an unintended consequence of using NGSL in IDW predictions at several locations in Utah. The large difference between the PRISM and IDW predictions in each case is explored in detail at Farmington, Utah (elevation 1316 meters). As observed in Table 3.3, three of the four measurement locations nearest to Farmington are all located at elevations above 2000 meters with NGSL much higher than the NGSL of the lone, low elevation measurement location. This results in a likely over-prediction of the design snow load at Farmington. This shortcoming is due to the strong positive correlation between elevation and NGSL at measurement locations above the separating elevation of 1219 m in Utah as observed on a log-scale in Figure 3.9. This correlation explains the sinusoidal error patterns for IDW observed previously in Figure 3.5. NGSL should be independent of elevation with a non-significant correlation coefficient. However, the Spearman correlation coefficient for NGSL and elevation at measurement locations above 1219 m on UT-2017 is 0.63, which is highly statistically significant ($p < 0.0001$). The overall Spearman correlation between elevation and NGSL on ID-2015 for elevations above 1219 m is only 0.14, yet still highly significant ($p = .0009$). While the separating elevation
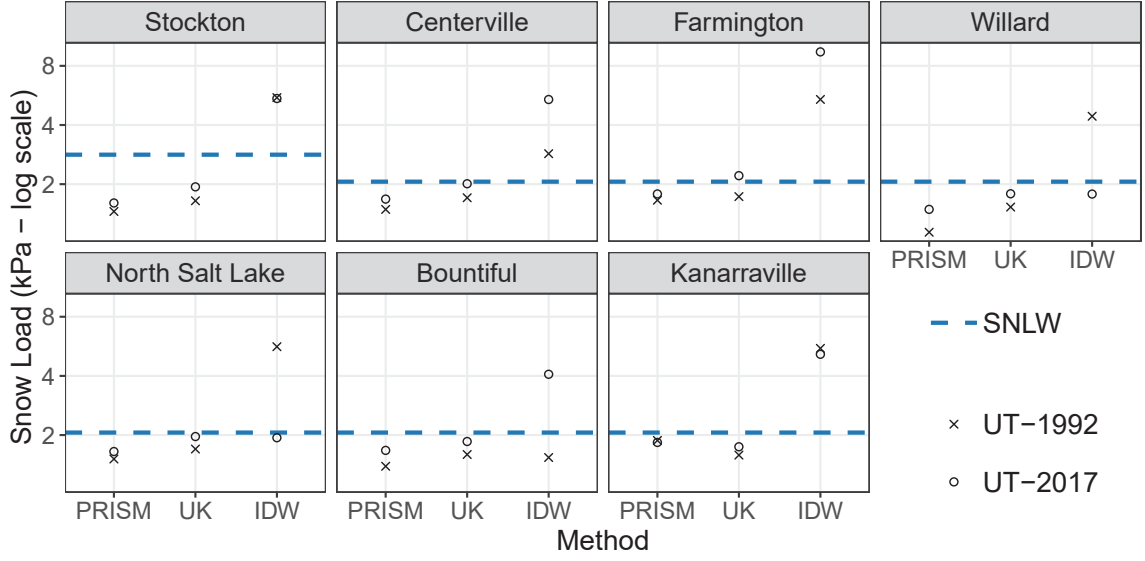
Fig. 3.8: Comparisons of spatial prediction methods using UT-2017 and UT-1992 that illustrate instances where NGSL predicts exceptionally high design snow loads.

seems to produce non-significant correlations between elevation and NGSL in the lower elevation layer, the significant correlations in the upper elevation layer cause inappropriate extrapolations of NGSL when the prediction location elevation is very different than the surrounding measurement location elevations. Recalling the cost implications shown in Figure 1.1, differences in design snow load prediction similar in magnitude to those observed at Farmington could easily double or triple the cost of the roof of a structure at these locations if this issue is not recognized and addressed.

One potential solution to this issue is the creation of an adaptive separating elevation that minimizes the correlations between NGSL and elevation in the lower and upper elevation groups. Letting ρ_l and ρ_u represent the correlation coefficient of choice (in our case, Spearman) in the lower and upper layers, our implementation of Idaho's method allows the separating elevation to be selected to minimize.

$$|\rho_c| = \frac{N_l|\rho_l| + N_u|\rho_u|}{N_l + N_u} \tag{3.1}$$

where N_l and N_u represent the number of measurement locations in the lower and upper
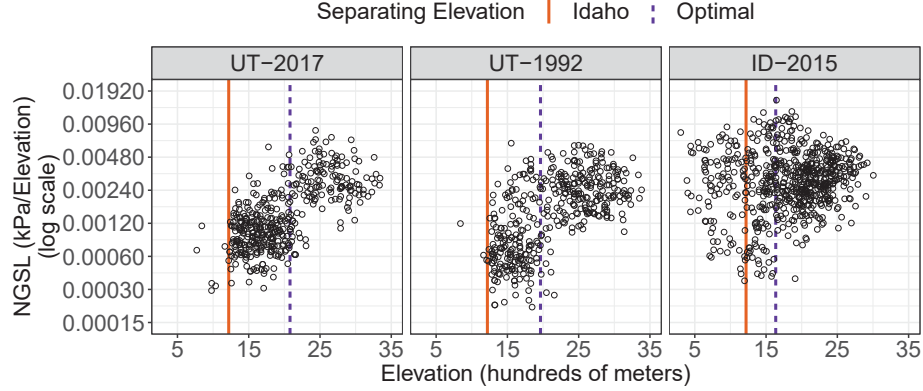
Fig. 3.9: Plots of elevation and NGSL (log-scale) for each dataset, showing a clear and unaccounted for relationship between NGSL and elevation. Idaho's original separating elevation is also compared to the separating elevation that minimizes the Spearman correlation between NGSL and elevation.

layers respectively. The function considers elevations within the interquartile range of the data (at 10 percentile increments) and selects the separating elevation that minimizes $|\rho_c|$. All measurement locations are placed in the lower layer if $|\rho_c|$ is not smaller than the absolute value of the overall correlation. The dotted lines in Figure 3.9 compare the "optimal" separating elevation to the original separating elevation used by Idaho.

This approach has the potential to make IDW predictions competitive with other methods. For example, the automatic selection of the separating elevation reduces the median MAE (with 100 iterations of cross validation) for IDW on UT-2017 by 25% (from 1.66 kPa to 1.25 kPa). Further improvements to IDW can be made using normalized, log-transformed design snow loads defined as log-NGSL $= \frac{\log(q_s(\boldsymbol{x}+1))}{A(\boldsymbol{x})}$. The +1 shift, though arbitrary as discussed in the previous section, is necessary to avoid negative log-NGSL values. This transformation results in a MAE on UT-2017 of 1.03, a 38% reduction in the MAE when applying the Idaho method directly. Similar gains can be obtained across all measures and all three datasets as observed in Table 3.4. It is important to note that the problems and potential solutions for NGSL were only exposed through a rigorous error analysis of IDW. This highlights the need for error analysis and model comparison in *any* statistical mapping problem.

Table 3.3: NGSL at the four measurement locations nearest to Farmington, Utah (111.884 W, 40.981 N).

ID	Elevation (m)	Distance to Location (km)	NGSL (kPa/m)
USC00422726	1335	5.4	0.0013
USS0011J11S	2438	5.5	0.0070
USS0011J12S	2066	6.4	0.0050
USS0011J68S	2359	8.4	0.0047

Table 3.4: Comparison of median measures of 100 iterations of cross validation for original (Org) and adjusted (Adj) IDW predictions (using the $\log(x+1)$ transformation and automatically selected separating elevation) to UK and PRISM.

	UT-2017			UT-1992			ID-2015		
	medAE	MAE	RMSE	medAE	MAE	RMSE	medAE	MAE	RMSE
Org	0.85	1.66	2.55	1.43	1.87	2.51	1.46	1.98	2.65
Adj	0.5	1.03	1.79	0.83	1.33	1.92	1.16	1.67	2.55
PRISM	0.47	0.94	1.61	0.7	1.27	2.01	1.06	1.63	2.67
UK	0.46	0.92	1.58	0.68	1.2	1.85	0.93	1.36	2.01

3.4.3 Limitations of Design Snow Load Estimates

Even with an "adequate" sample size, the inherently messy nature of real data (outliers, missing values, inaccurate measurements, and poor estimates of snow load from snow depth) adds uncertainty to 50 year (i.e. design) snow load estimates resulting from the distribution fitting process. In addition, potential violations of two assumptions inherent to the distribution fitting process add additional uncertainty to design snow load estimates. The first assumption is that the yearly maximums at each measurement location all come from the same distribution, implying that the conditions at each measurement location remain constant over time. However, Julander and Bricco (2006) document changes in measurement tools, sampling site conditions, and human influence at measurement locations that bring this assumption into doubt. The second assumption is that the yearly maximums are statistically independent, implying that snow measurements at each measurement location are uncorrelated across time. However, there is a wealth of evidence that suggests that time cannot be ignored when measuring climatic events. Gillies et al. (2012) claims that the proportion of precipitation falling as snow in Utah has declined by nine percent over the

last half century, accompanied by long term decreases in overall snow cover. This agrees with multiple sources indicating that yearly snow packs are declining across the Pacific Northwest (Mote, 2006; Scott and Kaiser, 2004). These sources indicate that the assumption of independence between yearly maximums is likely violated. Consequences resulting from these violations of assumptions will inevitably become more prevalent when trying to predict recurrence intervals beyond 50 years, such as those explored in DeBock et al. (2017). Future efforts to predict extreme snow load events should account for long-term, time-dependent trends in the snowpack. Such attempts could leverage advancements in snowpack forecasting made by Rhoades et al. (2016, 2018), which project significant declines in the snowpack of western state mountains through the end of the century.

One way to illustrate the effect of these uncertainties is through a comparison of estimated design snow loads for COOP station USC00109638 in Weiser, Idaho (NOAA, 2017). This station was selected due to the series of snow related collapses occurring in Weiser during the winter of 2017, where design snow loads were estimated to be as high as 1.89 kPa (Arcement, 2017). The reader should be cautioned that the reported collapses could be due to any number of factors (design, construction, etc.), not just design snow load prediction. We can not comment on the safety of those structures, only to illustrate the uncertainty in design snow loads based on the selected distribution and depth-to-load prediction. Station records at this location extend as far back as 1912. Data from this station were processed using the same procedures and filters used in the creation of UT-2017, resulting in a sample size of 73 yearly maximum snow loads. The normal, log-normal, Gumbel and generalized extreme value distributions each predict the design snow load estimate at this location, the latter two distributions being fit using the `extRemes` package (Gilleland and Katz, 2016). Efforts to fit a log-Pearson type III distribution via ML estimation were non-convergent and thus were excluded from the comparison. Each distribution was fit twice: once using Sturm's equation to convert snow depths to loads and again using the RMCD. Table 3.5 compares each of the resulting estimates to the 0.81 kPa design snow load estimate from the Idaho snow load report (Al Hatailah et al., 2015).

Table 3.5: Design snow load estimates for Weiser, Idaho using a variety of distributions.

	Design Load (kPa)	
Method	**Sturm**	**RMCD**
Log-Normal	1.64	1.04
Normal	1.54	1.07
Gumbel	1.55	0.99
GEV	2.34	1.25
Idaho Report		0.81

Table 3.5 shows that different distributions can provide notably different estimates of 50 year events. The differences in distribution estimates shown in Table 3.5 are relatively larger than distribution comparisons at the Denver-Stapleton, Colorado snow site provided in DeBock et al. (2017). This is likely due to Colorado's approach of fitting only the upper third of all maximums, which seems to reduce differences in the distribution tail approximations. Perhaps more important, however, is the difference in design snow load predictions resulting from changes to the depth-to-load conversion method. Table 3.5 shows that, using the same distribution, design snow loads using Sturm's equation results are more than 44% higher than design snow loads using the RMCD. Differences of this magnitude are not unique to this particular station, but are most pronounced at low elevation locations such as Weiser. Table 3.6 shows the median absolute relative difference of 50 year estimates for 261 measurement locations on UT-2017 relative to the original log-normal distribution estimates. Of the 415 measurement locations, 120 were excluded as they did not require any depth-to-load conversions and 21 were excluded for not having stable generalized extreme value 50 year estimates. These results suggest that differences in depth-to-load conversion methods are generally more influential on design snow loads than differences in distribution selection. These large differences reinforce the need for increased scrutiny in the process used to estimate design snow loads.

3.5 Reflections

Cross validation showed that UK was the most accurate across all three datasets and is our current top recommendation for mapping design snow loads. In addition, the relative

Table 3.6: Median absolute relative difference in design snow load estimates as compared to the original log-normal distribution estimates.

	Absolute Relative difference (%)	
Method	**Sturm**	**RMCD**
log-Normal		35%
Normal	13%	42%
Gumbel	8%	40%
GEV	21%	29%

ease of implementing UK and PRISM on new data demonstrate the feasibility of making predictions for multi-state regions. In addition, these prediction methods readily lend themselves to other water resource mapping problems. For example, we have used PRISM to visualize changes in the water content of Utah's April 1$^{\text{st}}$ snowpack from 1930-2015.

This chapter also discussed the limitations underlying the current distribution based methods for estimating design snow loads (or similar variants) at measurement locations. Comparisons of various distributions and snow load conversion methods in Tables 3.5 and 3.6 revealed that estimated design snow loads are very sensitive to changes in the depth-to-load conversion method.

This in mind, the following conclusions can be made:

- Changes in design snow loads up to a factor of nearly 290% in the Weiser, Idaho case study and more than 40% on average in UT-2017 occur based on changes in the selected probability distribution and depth-to-load conversion method.

- The best methods (in terms of MAE) account for log-linear relationship between design snow loads and elevation. The improvements in cross validated accuracy using these methods was as much as 45% on UT-2017 when compared to the current prediction methods used in Idaho and Utah.

- NGSL do not always adequately account for the elevation effect in design snow load prediction. The consequence is a tendency for IDW to over-predict snow loads at low elevations and under-predict at high elevations. The problems with NGSL can be

fixed by log-transforming snow loads prior to the NGSL calculation and adjusting the separating elevation.

- UK was similar in accuracy to PRISM on UT-2017 (MAE $\approx$ 0.9kPa) and UT-1992 (MAE $\approx$ 1.2kPa) and more accurate on ID-2015 (MAE $\approx$ 1.4kPa vs MAE $\approx$ 1.7kPa). Given its relative simplicity, well defined prediction variance, and robustness to differences in input data, we recommend universal kriging as the optimal method for predicting design snow loads in Utah and Idaho.

It is imperative to remember that these cross validation results are based on design snow loads which are estimates themselves, subject to uncertainty. Each mapping technique ignores this uncertainty by treating the design snow loads at measurement locations as exact observations. These approaches rob the design snow loads of crucial context, treating each measurement location as equally reliable in the spatial mapping technique. It may be more appropriate in these cases to characterize design snow loads as intervals expressing a range of possible design snow loads. This introduction of imprecision to the model inputs is discussed further in Chapter 4.

CHAPTER 4

INTERVAL-VALUED KRIGING

4.1   Introduction

[1] Recall that design snow loads are estimates subject to uncertainty. This uncertainty is effectively ignored by states when mapping design snow loads between measurement locations. The consequence of ignoring the imprecision in the model inputs is a set of outputs that lack context, as all model inputs (i.e. snow loads at measurement locations) are treated as equally certain. A reliability analysis considers both the magnitude and variability of anticipated loads for a structure. Treating uncertain loads as precise values robs the reliability analysis of crucial information about load variability. There is therefore a great need to develop interpolation methods that better characterize the input uncertainty to improve the ensuing reliability analysis.

For addressing the geographical mapping problems such as the aforementioned design snow loads, geostatistics and the associated kriging models are at the center of the current state-of-art science. It continues to be widely applied in many geosciences and related fields (Jin et al., 2018; Mao et al., 2018; Shtiliyanova et al., 2017). Entire textbooks have been devoted to explaining the many variations of kriging with virtually countless applications (Goovaerts, 1997). In particular, Bean et al. (2019) demonstrated that regression-kriging is significantly superior than a series of existing methods in predicting design snow loads for the state of Utah. Despite this variety and prevalence, increases in the availability and complexity of spatial data prompt discussions as to how kriging can better accommodate new data sources. New measurement technologies are more pervasive, yet many lack the precision of traditional measurements subject to human oversight. In addition, summaries of data across time or space create another layer of uncertainty, regardless of the precision

---

[1] This chapter is an adaptation of a manuscript submitted to the Annals of Applied Statistics. This manuscript was co-authored by Dr. Yan Sun and Dr. Marc Maguire.

of the individual measurements. There are several attempts in the literature to address uncertainties in the inputs (see Goovaerts (1997) for an outline), which typically create variants of kriging that handle the uncertainty indirectly. For example, one such method involves the use of indicator functions to characterize a set of response variable thresholds (Hohn, 1998). The kriging output in this case is a series of probability predictions, rather than a direct prediction on the response variable.

In a different framework, uncertainty is directly characterized by the data itself. This is done by expressing the data as intervals, as opposed to single numbers, which means that the true value is somewhere in the interval but not precisely known. Imprecise geostatistics models that aim at handling such uncertain inputs were originally studied in the late 1980s. (See Loquin and Dubois (2010) for a relatively comprehensive review.) The earliest attempt was probably the interval-valued kriging by Diamond (1988), which takes interval input and produces interval output. This model is a direct extension of the point-valued kriging (Matheron, 1963, 1971) with an interval-valued random function and a proposed covariance structure for intervals. The methodology was extended to a fuzzy kriging in Diamond (1989), intended for fuzzy interval input to allow membership degrees for each interval. At about the same time, Bardossy et al. (1990a,b) separately developed another type of fuzzy kriging by considering uncertainties in the variograms whose parameters are fuzzy intervals. Despite the well-justified theoretical foundations, these models did not gain popularity, largely due to a lack of consensus regarding the notion of covariance between intervals and the computational limitations of the time. Recently, Loquin and Dubois (2012) proposed an algorithmic extension to the Bardossy's fuzzy kriging, which effectively solved the computational issue in the original model and potentially improved its applicability. Additionally, Bandemer and Gebhardt (2000) proposed a Bayesian extension of Diamond's fuzzy kriging with an associated numerical implementation.

This chapter focuses on uncertain kriging with real (non-fuzzy) interval inputs. As reviewed above, Diamond's interval-valued kriging (Diamond, 1988) provided an initial solution to this problem. However, the definitions of covariance and stationarity underlying

this model may have been questionable, leading to an over complicated formulation and computation. In this chapter, we propose a modification to Diamond's interval-valued kriging based on the recent developments of the random set theory and a generalized L2 distance. These models overcome many of the mathematical and computational difficulties of previous interval-valued kriging attempts through the use of a well-established real-valued covariance between intervals. The numerical implementations of these models in R 3.5.1 (R Core Team, 2018) leverage existing geospatial workflows in the sp and gstat packages (Bivand et al., 2013; Gräler et al., 2016; Pebesma, 2004; Pebesma and Bivand, 2005) described further in Chapter 5. Using these implementations, we apply our interval-valued kriging to the design snow load prediction problem. The analysis includes the creation of an interval-valued design snow load dataset that characterizes traditionally ignored sources of uncertainty in the design snow load estimation process.

The rest of the chapter proceeds as follows. Section 4.2 reviews the random sets framework underlying the proposed interval-valued kriging. Section 4.3 introduces our interval-valued kriging models and discusses the associated properties. Section 4.4 reviews the details of the numerical implementation of the algorithm while Section 4.5 demonstrates the empirical convergence of the numerical implementation through a series of simulations. Section 4.6 presents the application of our interval-valued kriging to predicting design snow loads in Utah. We give concluding remarks in Section 4.7. Technical proofs are deferred to Section 4.8.

4.2   Random sets preliminaries

Denote by $\mathcal{K}\left(\mathbb{R}^d\right)$ or $\mathcal{K}$ the collection of all non-empty compact subsets of $\mathbb{R}^d$. The Hausdorff metric ρ_H

$$\rho_H\left(A, B\right) = \max\left(\sup_{a \in A} \rho\left(a, B\right), \sup_{b \in B} \rho\left(b, A\right)\right), \ \forall A, B \in \mathcal{K},$$

where ρ denotes the Euclidean metric, defines a natural metric in $\mathcal{K}$. As a metric space, $(\mathcal{K}, \rho_H)$ is complete and separable (Debreu, 1967). In the space $\mathcal{K}$, a linear structure can be

defined by Minkowski addition and scalar multiplication as

$$A + B = \{a + b : a \in A, b \in B\}, \quad \lambda A = \{\lambda a : a \in A\}, \ \ \forall A, B \in \mathcal{K}, \ \ \lambda \in \mathbb{R}.$$

Note however that $\mathcal{K}$ is not a linear space (or vector space) as there is no inverse element of addition. Let $(\Omega, \mathcal{L}, P)$ be a probability space. A random compact set is a Borel measurable function $A : \Omega \to \mathcal{K}$, $\mathcal{K}$ being equipped with the Borel σ-algebra induced by the Hausdorff metric. If $A(\omega)$ is convex almost surely, then A is called a random compact convex set (Molchanov, 2005). The collection of all compact convex subsets of $\mathbb{R}^d$ is denoted by $\mathcal{K}_{\mathcal{C}}\left(\mathbb{R}^d\right)$ or $\mathcal{K}_{\mathcal{C}}$.

Particularly, $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ contains all the non-empty bounded closed intervals in $\mathbb{R}$ and a measurable function that maps Ω to $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ is called a random interval. From now on, an element in $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ will be denoted by $[x]$, whose lower/upper bounds and center/radius are denoted by x^L/x^U and x^C/x^R, respectively. Bold letters denote vectors and random versions are denoted by capital letters. For example, $[\boldsymbol{x}] = [[x_1], \cdots, [x_p]]^T$ denotes a p-dimensional hyper interval and its random version is denoted by $[\boldsymbol{X}]$. The expectation of a random compact convex random set A is defined by the Aumann integral of set-valued function (Artstein and Vitale, 1975; Aumann, 1965) as $E(A) = \{E\xi : \xi \in A \text{ almost surely}\}$, which for a random interval $[X]$ is $E([X]) = [E\left(X^L\right), E\left(X^U\right)]$.

For interval-valued data analysis, the measure of distance is a critical issue. According to the embedding theorems (Hörmander, 1954; Rådström, 1952), $\mathcal{K}_{\mathcal{C}}$ can be embedded isometrically into the Banach space $C(S)$ of continuous functions on the unit sphere S^{d-1}, which are realized by the support function of $X \in \mathcal{K}_{\mathcal{C}}$. Therefore, a compact convex set can be represented by its support function s_X and $\rho_2(X, Y) := \|s_X - s_Y\|_2$, $\forall X, Y \in \mathcal{K}_{\mathcal{C}}$, defines an L_2 metric on $\mathcal{K}_{\mathcal{C}}$. It is known that ρ_H and ρ_2 are equivalent metrics, but ρ_2 is more preferred for statistical inference, due to many of its established properties (Körner, 1995, 1997). The ρ_2-metric for an interval $[x]$ has the particularly simple form $\|[x]\|_2^2 = \frac{1}{2}\left(x^L\right)^2 + \frac{1}{2}\left(x^U\right)^2 = \left(x^C\right)^2 + \left(x^R\right)^2$ and the ρ_2-distance between two intervals

is $\rho_2\left([x],[y]\right) = \left[\frac{1}{2}\left(x^L - y^L\right)^2 + \frac{1}{2}\left(x^U - y^U\right)^2\right]^{\frac{1}{2}} = \left[\left(x^C - y^C\right)^2 + \left(x^R - y^R\right)^2\right]^{\frac{1}{2}}$. A more general metric for $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ was proposed Gil et al. (2001) which essentially takes form

$$\rho_W^2\left([x],[y]\right) = \left(x^C - y^C\right)^2 + \left(x^R - y^R\right)^2 \int_{[0,1]} (2\lambda - 1)^2\, dW(\lambda),$$

where W is any non-degenerate symmetric measure on $[0,1]$. This allows for weighting between the center and radius. Separately, for a more general space, Körner and Näther (2001) proposed another L_2 metric, which when restricted to $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ is

$$\rho_K^2([x],[y]) = \sum_{(u,v)\in S^0 \times S^0} \left(s_{[x]}(u) - s_{[y]}(u)\right)\left(s_{[x]}(v) - s_{[y]}(v)\right) K(u,v),$$

where K is a symmetric positive definite kernel. It can be represented by the upper/lower bounds as

$$\begin{aligned}\rho_K^2([x],[y]) = {}& \left(x^U - y^U\right)^2 K(1,1) + \left(x^L - y^L\right)^2 K(-1,-1) \\ & - \left(x^U - y^U\right)\left(x^L - y^L\right)\left[K(1,-1) + K(-1,1)\right],\end{aligned}$$

or equivalently in the center-radius form as

$$\rho_K^2([x],[y]) = A_{11}(x^C - y^C)^2 + A_{22}(x^R - y^R)^2 + 2A_{12}(x^C - y^C)(x^R - y^R),$$

where

$$A_{11} = K(1,1) + K(-1,-1) - \left[K(1,-1) + K(-1,1)\right],$$
$$A_{22} = K(1,1) + K(-1,-1) + \left[K(1,-1) + K(-1,1)\right],$$
$$A_{12} = A_{21} = K(1,1) - K(-1,-1).$$

Apparently, when K is symmetric positive definite, so is A. Thus the essence of ρ_K lies in its further generalization of ρ_W that takes into account the interaction between the center

and the radius.

4.3 The interval-valued kriging

The key to developing interval-valued kriging is to define a proper second-order structure for the interval-valued random function $[Z(\boldsymbol{x})] = [Z^L(\boldsymbol{x}), Z^U(\boldsymbol{x})]$. Diamond's approach used an interval-valued covariance as

$$[C(\boldsymbol{x}, \boldsymbol{h})] = E\left[Z(\boldsymbol{x})Z(\boldsymbol{x} + \boldsymbol{h})\right] - E\left[Z(\boldsymbol{x})\right]E\left[Z(\boldsymbol{x} + \boldsymbol{h})\right]. \tag{4.1}$$

This definition is only conceptual, because there is no inverse element in the space $\mathcal{K}_\mathcal{C}(\mathbb{R})$ and thus the subtraction of intervals is not defined. In addition, it requires the multiplication of intervals, which is complicated in general. For simplification purposes, Diamond restricted considerations to positive intervals, i.e., intervals that contain only positive numbers. Under this restriction, the multiplication is seen to be

$$[Z(\boldsymbol{x} + \boldsymbol{h})]\,[Z(\boldsymbol{x})] = \left[Z(\boldsymbol{x} + \boldsymbol{h})^L Z(\boldsymbol{x})^L, Z(\boldsymbol{x} + \boldsymbol{h})^U Z(\boldsymbol{x})^U\right].$$

Diamond's notion of second-order stationarity was subsequently defined only for positive intervals with conditions:

1. $E\left[Z(\boldsymbol{x})\right] = [m]$ exists and is independent of $\boldsymbol{x}$;

2. $E\left[Z(\boldsymbol{x})Z(\boldsymbol{x} + \boldsymbol{h})\right] = [C(\boldsymbol{x}, \boldsymbol{h})] + [m]^2$ exists and is independent of $\boldsymbol{x}$, assuming $[C(\boldsymbol{x}, \boldsymbol{h})]$ is a positive interval.

There are several mathematical difficulties in this framework. First, as mentioned above, there is no well-defined subtraction operation for intervals and therefore the interval-valued covariance $[C(\boldsymbol{x}, \boldsymbol{h})]$ cannot be determined by (4.1). Notice how the covariance stationarity (condition 2) is stated indirectly by $E\left[Z(\boldsymbol{x})Z(\boldsymbol{x} + \boldsymbol{h})\right]$, instead of by the covariance $[C(\boldsymbol{x}, \boldsymbol{h})]$ itself. While one can theoretically impose a covariance structure by assumption, in practice it is not obvious how the covariance can be estimated from the data, which limits its applicability. The second difficulty is with regard to the mathematical coherence of the

variance and covariance, that is, the covariance of two identical quantities should be the same as the variance. However, it can be seen that $\text{Var}\,[Z(\boldsymbol{x})] \neq C(\boldsymbol{x}, \boldsymbol{0})$. In fact, the former is real-valued and the latter is an interval. Lastly, to ensure the non-negativity of the prediction variance, all of the interpolation weights are assumed to be non-negative, which makes the model even more restrictive.

According to the recent development of set-valued statistics (e.g., Körner (1997); Körner and Näther (1998)), the covariance of random intervals, and in general random sets, should be defined as real-valued. This could be the potential solution to the aforementioned problems in Diamond's formulation. Motivated by this, we propose to re-construct the second-order structure based on the random sets theory to modify Diamond's interval-valued kriging into a more rigorous and more computationally feasible method. To this, the notion of variance of a random set (Körner, 1995, 1997; Lyashenko, 1982; Näther, 1997) plays the key role. Given a metric ρ in the space $\mathcal{K}$, the variance of a random compact set A is defined as $\text{Var}_\rho(A) = E\rho^2[A, E(A)]$. Now if we restrict to $\mathcal{K}_\mathcal{C}$, according to the embedding, a random compact convex set X can be represented by its support function s_X and the space $\mathcal{K}_\mathcal{C}$ is equipped with an L_2 metric ρ_2. Considering $< \cdot, \cdot >$ as the inner product in the Hilbert space $L_2(S^{d-1})$, the variance is defined as

$$\text{Var}(X) = E\left\|s_X - s_{E(X)}\right\|_2^2 = E\int_{S^{d-1}}[s_X - s_{E(X)}]^2\mu d(u) = E < s_X - s_{E(X)}, s_X - s_{E(X)} > .$$

This leads to the natural extension to the covariance function for $X, Y \in \mathcal{K}_\mathcal{C}(\mathbb{R}^d)$ as

$$\text{Cov}(X, Y) = E < s_X - s_{E(X)}, s_Y - s_{E(Y)} >= E\int_{S^{d-1}}[s_X - s_{E(X)}][s_Y - s_{E(Y)}]\mu d(u).$$

Such a definition of covariance has been shown to be very favorable for statistical analysis (Körner, 1995, 1997). Consider random intervals $[X], [Y] \in \mathcal{K}_\mathcal{C}(\mathbb{R})$ and the general metric ρ_K. The variance is seen to be

$$\text{Var}([X]) \quad = \quad E\left\{\rho_K^2([X], E([X]))\right\}$$

$$
\begin{aligned}
&= E\big[A_{11}\left(X^C - E(X^C)\right)^2 + A_{22}\left(X^R - E(X^R)\right)^2 \\
&\quad + 2A_{12}\left(X^C - E(X^C)\right)\left(X^R - E(X^R)\right)\big] \\
&= A_{11}\mathrm{Var}(X^C) + A_{22}\mathrm{Var}(X^R) + 2A_{12}\mathrm{Cov}(X^C, X^R), \qquad (4.2)
\end{aligned}
$$

The covariance is a little more complex. We notice that the inner product associated with the ρ_K metric is

$$
\begin{aligned}
< s_X - s_{E(X)}, s_Y - s_{E(Y)} > &= \left(X^U - E(X^U)\right)\left(Y^U - E(Y^U)\right)K(1,1) \\
&\quad + \left(X^L - E(X^L)\right)\left(Y^L - E(Y^L)\right)K(-1,-1) \\
&\quad - \left(X^U - E(X^U)\right)\left(Y^L - E(Y^L)\right)K(1,-1) \\
&\quad - \left(X^L - E(X^L)\right)\left(Y^U - E(Y^U)\right)K(-1,1),
\end{aligned}
$$

which can be rewritten in terms of the center and radius as

$$
\begin{aligned}
< s_X - s_{E(X)}, s_Y - s_{E(Y)} > &= B_{11}\left(X^C - E(X^C)\right)\left(Y^C - E(Y^C)\right) \\
&\quad + B_{22}\left(X^R - E(X^R)\right)\left(Y^R - E(Y^R)\right) \\
&\quad + B_{12}\left(X^C - E(X^C)\right)\left(Y^R - E(Y^R)\right) \\
&\quad + B_{21}\left(X^R - E(X^R)\right)\left(Y^C - E(Y^C)\right),
\end{aligned}
$$

where B is a symmetric positive definite matrix uniquely determined by K. The covariance is consequently defined as

$$
\begin{aligned}
\mathrm{Cov}([X],[Y]) &= E\left\{< s_X - s_{E(X)}, s_Y - s_{E(Y)} >\right\} \\
&= B_{11}\mathrm{Cov}(X^C, Y^C) + B_{22}\mathrm{Cov}(X^R, Y^R) \\
&\quad + B_{12}\mathrm{Cov}(X^C, X^R) + B_{21}\mathrm{Cov}(X^R, X^C).
\end{aligned}
$$

We are now ready to introduce our interval-valued kriging and the definition of stationarity. Recall that $[Z(\boldsymbol{x})] = [Z^L(\boldsymbol{x}), Z^U(\boldsymbol{x})]$ denotes the interval-valued random function,

which can be alternatively represented by the center function $Z^C(\boldsymbol{x})$ and the radius function $Z^R(\boldsymbol{x})$. As in Diamond (1988), our interval-valued kriging interpolator is defined as

$$\widehat{[Z]}(\boldsymbol{x}^*) = \sum_{i=1}^{n} \lambda_i [Z](\boldsymbol{x}_i),$$

according to the Minkowski addition and scalar multiplication. It can be expressed equivalently in the center-radius form as

$$\hat{Z}^C(\boldsymbol{x}^*) = \sum_{i=1}^{n} \lambda_i Z^C(\boldsymbol{x}_i), \quad \hat{Z}^R(\boldsymbol{x}^*) = \sum_{i=1}^{n} |\lambda_i| Z^C(\boldsymbol{x}_i).$$

Given the preceding discussion of the second-order structure of random intervals, the stationarity of $[Z(\boldsymbol{x})]$ is derived from a natural extension of the stationarity for point-valued random function. We formally state it in the following.

Definition 1. *The interval-valued random function $[Z(\cdot)]$ is second-order stationary if it satisfies*

1. *(Mean Stationarity) $E([Z(\boldsymbol{x})]) = [m]$, for some fixed interval $[m]$ independent of $\boldsymbol{x}$, i.e. $E(Z^C(\boldsymbol{x})) = m^C$ and $E(Z^R(\boldsymbol{x})) = m^R \geq 0$ independent of $\boldsymbol{x}$;*

2. *(Covariance Stationarity) $Cov([Z(\boldsymbol{x} + \boldsymbol{h})], [Z(\boldsymbol{x})]), \boldsymbol{h} \in \mathbb{R}^n$ is a function of $\boldsymbol{h}$ only, i.e., the four covariance functions $Cov\left(Z^I(\boldsymbol{x} + \boldsymbol{h}), Z^J(\boldsymbol{x})\right) = C^{I,J}(\boldsymbol{h}), I, J \in \{C, R\}$, are all independent of $\boldsymbol{x}$.*

As a remark on the covariance stationary, $C^{R,C}$ is completely determined by $C^{C,R}$ in that $C^{R,C}(\boldsymbol{h}) = C^{C,R}(-\boldsymbol{h})$, so only three covariance functions, i.e., $C^{C,C}(\boldsymbol{h})$, $C^{R,R}(\boldsymbol{h})$, and $C^{C,R}(\boldsymbol{h})$ are needed to define stationarity. Under the assumption of second-order stationarity, the prediction variance of the kriging estimator is calculated in the following Theorem 1.

Theorem 1. *Up to an additive constant, the prediction variance of the interval-valued kriging interpolator is equal to*

$$E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right] \quad = \quad A_{11}\left[\sum_i \sum_j \lambda_i \lambda_j C^{C,C}(x_i - x_j) - 2\sum_i \lambda_i C^{C,C}(x_i - x^*)\right]$$

$$+A_{22}\left[\sum_i\sum_j|\lambda_i\lambda_j|\,C^{R,R}(x_i-x_j)-2\sum_i|\lambda_i|\,C^{R,R}(x_i-x^*)\right]$$

$$+2A_{12}\left[\sum_i\sum_j\lambda_i|\lambda_i|C^{C,R}(x_i-x_j)\right]$$

$$-2A_{12}\left[\sum_i|\lambda_i|C^{C,R}(x^*-x_i)+\sum_i\lambda_iC^{C,R}(x_i-x^*)\right]. \qquad (4.3)$$

4.3.1 Simple Kriging (SK)

Assume $E\left([Z(\boldsymbol{x})]\right)=[m]$ for a known fixed interval $[m]$. We can replace $[Z]$ by $[Z]-m^C$ (and add m^C back after the model is fitted) so that the center function has a constant mean of zero. Then, $\hat{Z}^C(\boldsymbol{x}^*)$ is automatically unbiased and the unbiasedness of $\hat{Z}^R(\boldsymbol{x}^*)$ implies $\sum_{i=1}^n|\lambda_i|=1$.

Hence, the interval-valued SK estimator is defined as the minimizer of the prediction variance under the unbiasedness constraint, i.e.,

$$[\hat{Z}^{SK}(x^*)]=\arg\min E\left[\rho_K^2\left([\hat{Z}(x^*)],[Z(x^*)]\right)\right],\quad\text{subject to}\quad\sum_{i=1}^n|\lambda_i|=1. \qquad (4.4)$$

4.3.2 Ordinary Kriging (OK)

OK still assumes that $E\left([Z(\boldsymbol{x})]\right)=[m]$, but the interval-valued mean $[m]$ is unknown. Thus, we can no longer demean the center and instead have to impose the additional condition $\sum_{i=1}^n\lambda_i=1$ to ensure that the center prediction is unbiased. This together with the condition $\sum_{i=1}^n|\lambda_i|=1$ implies that the weights need to be all non-negative. Therefore, the interval-valued OK estimator is defined as

$$\begin{aligned}[\hat{Z}^{OK}(x^*)]\quad&=\quad\arg\min E\left[\rho_K^2\left([\hat{Z}(x^*)],[Z(x^*)]\right)\right],\\&\qquad\text{subject to}\quad\sum_{i=1}^n\lambda_i=1,\quad\lambda_i\ge0,\quad i=1,\cdots,n.\qquad(4.5)\end{aligned}$$

4.3.3 The Variogram

In a slightly different situation, instead of assuming $Z(\boldsymbol{x})$ is stationary, it is assumed that the increment $Z(\boldsymbol{x}+\boldsymbol{h})-Z(\boldsymbol{x})$ is stationary. A point-valued random function that is

second-order increment stationary is defined such that the first two moments of the increment are both independent of $\boldsymbol{x}$, i.e.,

$$E\left[Z(\boldsymbol{x}+\boldsymbol{h})-Z(\boldsymbol{x})\right]=m\left(\boldsymbol{h}\right);$$

$$\text{Var}\left[Z(\boldsymbol{x}+\boldsymbol{h})-Z(\boldsymbol{x})\right]=2\gamma\left(\boldsymbol{h}\right),$$

where $\gamma(\boldsymbol{h})$ is called the semi-variogram. Usually the function $m(\boldsymbol{h})$ is assumed to be constantly zero, namely, $Z(\boldsymbol{x})$ has a constant mean. Increment stationarity is a slightly weaker condition than stationarity, mainly because it allows the variance of $Z(\cdot)$ to be infinite. For interval-valued random function $[Z(\boldsymbol{x})]$, it is difficult to define "increment" as in the point-valued case, because there is no inverse element of addition in the space $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$. Nevertheless, the assumptions for increment stationarity can be equivalently specified for interval-valued process through an interval-valued drift function and a real-valued semi-variogram as follows:

Definition 2. *The interval-valued random function $[Z(\cdot)]$ is increment stationary if it satisfies*

1. *(Mean Stationarity) $E\left([Z(\boldsymbol{x})]\right)=[m]$, for some fixed interval $[m]$ independent of $\boldsymbol{x}$;*

2. *(Variogram Stationarity) $E\rho_K^2\left([Z(\boldsymbol{x}+\boldsymbol{h})],[Z(\boldsymbol{x})]\right)=2\gamma(\boldsymbol{h})$, $h\in\mathbb{R}^n$, independent of $\boldsymbol{x}$.*

According to the definition of variance (4.2), the semi-variogram $\gamma(\boldsymbol{h})$ breaks down into the center, radius, and center-radius semi-variograms as

$$\gamma(\boldsymbol{h})=A_{11}\gamma^C(\boldsymbol{h})+A_{22}\gamma^R(\boldsymbol{h})+2A_{12}\gamma^{C,R}(\boldsymbol{h}),$$

where

$$\gamma^C(\boldsymbol{h})=\frac{1}{2}\text{Var}\left(Z^C(\boldsymbol{x}+\boldsymbol{h})-Z^C(\boldsymbol{x})\right),$$
$$\gamma^R(\boldsymbol{h})=\frac{1}{2}\text{Var}\left(Z^R(\boldsymbol{x}+\boldsymbol{h})-Z^R(\boldsymbol{x})\right),$$
$$\gamma^{C,R}(\boldsymbol{h})=\frac{1}{2}\text{Cov}\left(Z^C(\boldsymbol{x}+\boldsymbol{h})-Z^C(\boldsymbol{x}),Z^R(\boldsymbol{x}+\boldsymbol{h})-Z^R(\boldsymbol{x})\right).$$

Thus, variogram stationarity means that all of the semi-variograms γ^C, γ^R, and $\gamma^{C,R}$, are independent of $\boldsymbol{x}$. In practice, their forms can be chosen by the corresponding sample estimates. If the covariance functions exist, they are related to the semi-variograms by the following equations:

$$
\begin{aligned}
\gamma^C(\boldsymbol{h}) &= C^{C,C}(\boldsymbol{0}) - C^{C,C}(\boldsymbol{h}), \\
\gamma^R(\boldsymbol{h}) &= C^{R,R}(\boldsymbol{0}) - C^{R,R}(\boldsymbol{h}), \\
\gamma^{C,R}(\boldsymbol{h}) &= C^{C,R}(\boldsymbol{0}) - \frac{1}{2}\left[C^{C,R}(\boldsymbol{h}) + C^{R,C}(\boldsymbol{h})\right], \\
&= C^{C,R}(\boldsymbol{0}) - \frac{1}{2}\left[C^{C,R}(\boldsymbol{h}) + C^{C,R}(-\boldsymbol{h})\right].
\end{aligned}
$$

Theorem 2. *Under the unbiasedness constraints $\sum_{i=1}^{n} \lambda_i = 1$ and $\sum_{i=1}^{n} |\lambda_i| = 1$, $i = 1, \cdots, n$, the prediction variance is equal to*

$$
\begin{aligned}
&E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right] \\
=\ & A_{11}\left[-\sum_i \sum_j \lambda_i \lambda_j \gamma^C(x_i - x_j) + 2\sum_i \lambda_i \gamma^C(x_i - x^*)\right] \\
+\ & A_{22}\left[-\sum_i \sum_j \lambda_i \lambda_j \gamma^R(x_i - x_j) + 2\sum_i \lambda_i \gamma^C(x_i - x^*)\right] \\
+\ & 2A_{12}\left[-\sum_{i \neq j} \lambda_i \lambda_j \gamma^{C,R}(x_i - x_j) + 2\sum_i \lambda_i \gamma^{C,R}(x_i - x^*)\right].
\end{aligned}
$$

4.4 Numerical implementation

Implementation of the proposed SK and OK models amounts to minimizing the prediction variance subject to certain constraints. If the covariance functions exist, which implies that the variograms also exist, the prediction variance can be expressed either by (4.3) or (4.6). Otherwise, under a weaker assumption when the covariance functions do not exist but the variograms exist, the prediction variance is given by (4.6). In either case, direct differentiation is impossible due to the involvement of $|\boldsymbol{\lambda}|$. In addition, the inequality constraints associated with the OK in (4.5) are a form of the Karush-Kuhn-Tucker conditions, for which an analytical solution usually does not exist. Considering all these, to implement

the interval-valued kriging models, we propose a penalized approximate Newton-Raphson (PANR) algorithm that finds a numerical solution to the constrained optimization problem.

4.4.1 The penalty method for constraints

Denote the prediction variance by

$$V(\boldsymbol{\lambda}) = E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right].$$

Recall from (4.4) and (4.5) that SK minimizes $V(\boldsymbol{\lambda})$ subject to $\sum_{i=1}^{n} |\lambda_i| = 1$ and OK subject to $\sum_{i=1}^{n} \lambda_i = 1$, $\lambda_i \geq 0$, $i = 1, \cdots, n$, respectively. We employ a penalty method to account for these optimization constraints. (See, e.g., Jensen and Bard (2003) for a review of algorithms for constrained optimization.) The idea is to approximate the constrained optimization problem by an unconstrained problem formulated as

$$\arg\min\left\{V(\boldsymbol{\lambda}) + P(\boldsymbol{\lambda}, c)\right\},$$

where $P(\boldsymbol{\lambda}, c)$ is a continuous penalty function that equals to zero if and only if the constraints are satisfied and c is a positive constant controlling the magnitude of the penalty function. With an appropriately chosen penalty function, the solution of (4.4.1) is approximately the same as the constrained minimizer of the original objective function $V(\boldsymbol{\lambda})$. Corresponding to the equality constraint $\sum_{i=1}^{n} |\lambda_i| = 1$ for SK, the most natural penalty is the quadratic loss penalty

$$P^{SK}(\boldsymbol{\lambda}, c) = c\left(1 - \sum_{i=1}^{n} |\lambda_i|\right)^2.$$

When c is large enough, any violation of the constraint will result in a heavy cost from the penalty and thus minimizing the penalized objective function will yield a feasible solution. For OK, there are an equality constraint $\sum_{i=1}^{n} \lambda_i = 1$ and inequality constraints $\lambda_i \geq 0$, $i = 1, \cdots, n$. To tackle such a problem, the most common strategy is to employ

the logarithmic-quadratic loss function

$$P^{OK}(\boldsymbol{\lambda}, c) = -c \sum_{i=1}^{n} \ln (\lambda_i) + \frac{1}{c} \left(1 - \sum_{i=1}^{n} |\lambda_i| \right)^2,$$

where the logarithmic terms take care of the inequality constraints. Similar to the pure quadratic loss penalty, small values of c will lead to a solution within the feasible region. A simple straightforward strategy to implement the penalty method is known as the sequential unconstrained minimization technique (SUMT) (Jensen and Bard, 2003). It starts with an initial value of the penalty parameter c_0 and iteratively updates it until the convergence criterion is satisfied. It was shown in Fiacco and McCormick (1968) that for a sequence of monotonically increasing (or decreasing depending on the nature of the problem) $\{c_k\}$, the SUMT converges to the penalized objective within the feasible region. A small issue with this algorithm is that values of c that are too large (or too small) will create ill-behaved surfaces for which gradient and Hessian calculations will be unstable. Therefore, slowly changing the values of c balances the influence of the penalty term with the rest of the objective function and is the key to the success.

For SK, the penalty parameter c needs to strictly increase to a large enough value. So we implement the SUMT algorithm as follows:

1. (*Initialization*) Set the initial parameter $c_0 = 0$ and determine an initial value $\boldsymbol{\lambda}^{(0)}$. Let $k = 0$.

2. (*Minimization*) Minimize $V(\boldsymbol{\lambda}) + P^{SK}(\boldsymbol{\lambda}, c_k)$ to obtain $\boldsymbol{\lambda}^{(k+1)}$. Let $k = k + 1$.

3. (*Check Constraint*) Compute $p_k = 1 - \sum_{i=1}^{n} |\lambda_i^{(1)}|$. If $|p_k| < tolp$, where *tolp* is a user defined tolerance for the constraint, end the iteration. Otherwise, go to Step 4.

4. (*Update Parameter*) Set $c_{k+1} = \eta c_k$, with $\eta > 1$ being a user defined growth parameter ($\eta > 1$), and repeat Step 2 and 3.

For OK, the nonnegative constraints are ensured through the use of a barrier (logarithmic) penalty that slowly allows entries in $\boldsymbol{\lambda}$ to approach the boundaries of the feasible

region by shrinking an initially large penalty parameter. We therefore implement the SUMT algorithm as follows:

1. (*Initialization*) Set a relatively large value for c_0 (we use $c_0 = 100$) and determine an initial value $\boldsymbol{\lambda}^{(0)}$.

2. (*Minimization*) Minimize $V(\boldsymbol{\lambda}) + P^{OK}(\boldsymbol{\lambda}, c_k)$ to obtain $\boldsymbol{\lambda}^{(k+1)}$. Let $k = k + 1$.

3. (*Check Constraints*) Check that $\min(\boldsymbol{\lambda}^{(k)}) \geq 0$. Also, compute $p_k = 1 - \sum_{i=1}^{n} |\lambda_i^{(k)}|$ and check if $|p_k| < tolp$. If both criteria are satisfied, end the iteration. Otherwise, go to Step 4.

4. (*Update Parameter*) Set $c_{k+1} = \eta c_k$, with $\eta < 1$ being a user defined shrinkage parameter, and repeat Step 2 and 3. The shrinkage parameter η must be close enough to 1 to ensure that $\min(\boldsymbol{\lambda}^{(z_t)}) \geq 0$ until $|p_k| < tolp$ is satisfied.

For both SK and OK, the algorithm is terminated if the minimization (Step 2) fails for the final value of c_k after a user-specified maximum number of iterations ($maxq$), or the constraints are not satisfied within the maximum number of penalty iterations ($maxp$).

4.4.2 Approximation of $|\boldsymbol{\lambda}|$

The key step of the aforementioned SUMT algorithm is the minimization of the penalized prediction variance

$$Q(\boldsymbol{\lambda}) = V(\boldsymbol{\lambda}) + P(\boldsymbol{\lambda}, c). \tag{4.6}$$

Numerically, this can be carried out by the PANR algorithm. In order to guarantee convergence, the objective function must be second-order continuously differentiable. However, the absolute value $|\lambda|$ in the prediction variance is not differentiable at $\lambda = 0$. To address this issue, we propose to approximate the absolute value by a local quadratic function. Consider the Taylor expansion of $|\lambda|$ at $\lambda_0 \neq 0$:

$$|\lambda| = |\lambda_0| + |\lambda|'(\lambda_0)(\lambda - \lambda_0) + o\left(|\lambda - \lambda_0|^2\right), \quad \lambda \approx \lambda_0.$$

Replacing the derivative $|\lambda|'(\lambda_0) = sgn(\lambda_0)$ by $\frac{\lambda}{|\lambda_0|}$, $|\lambda|$ is approximated by a quadratic function as

$$|\lambda| \approx |\lambda_0| + (\lambda - \lambda_0)\frac{\lambda}{|\lambda_0|}, \quad \lambda_0 \neq 0. \tag{4.7}$$

Under this approximation, the gradient $\boldsymbol{G}$ and Hessian $\boldsymbol{H}$ of $Q(\boldsymbol{\lambda})$ are defined as

$$\boldsymbol{G} = \nabla Q(\boldsymbol{\lambda}),$$
$$\boldsymbol{H} = \nabla^2 Q(\boldsymbol{\lambda}).$$

and the iteration of the PANR algorithm is given by

$$\boldsymbol{\lambda}^{(m+1)} = \boldsymbol{\lambda}^{(m)} - \boldsymbol{H}^{-1}\left(\boldsymbol{\lambda}^{(m)}\right) * \boldsymbol{G}\left(\boldsymbol{\lambda}^{(m)}\right).$$

where $\boldsymbol{\lambda}^{(m)}$ is the value of $\boldsymbol{\lambda}$ from the mth iteration. In the Theorem 3 below, the gradient and Hessiann of $Q(\boldsymbol{\lambda})$ are explicitly calculated for both SK and OK.

Theorem 3. *Consider minimizing the penalized prediction variance $Q(\boldsymbol{\lambda})$ defined in (4.4.1) and (4.6) using Newton-Raphson algorithm. Let $|\boldsymbol{\lambda}|$ be approximated by (4.7), where $\boldsymbol{\lambda}_0$ is an approximated value of $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}_0 \neq \boldsymbol{0}$. Define*

$$f^{R,R}(\boldsymbol{\lambda}, \lambda_k) = \left[\sum_i |\lambda_i|\left[C^{R,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_i)\right] - 2C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right].$$

Then the gradient $\boldsymbol{G}$ and Hessian $\boldsymbol{H}$ of $Q(\boldsymbol{\lambda})$ for SK are

$$
\begin{aligned}
\boldsymbol{G}_k = {} & A_{11}\left[\sum_i \lambda_i\left[C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_i)\right] - 2C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right] \\
& + A_{22}\left[\frac{\lambda_k}{|\lambda_{k0}|}f^{R,R}(\boldsymbol{\lambda}, \lambda_k)\right] \\
& + 2A_{12}\left[\sum_i\left[|\lambda_i|C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) + \frac{\lambda_k}{|\lambda_{k0}|}\lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k)\right]\right] \\
& - 2A_{12}\left[\frac{\lambda_k}{|\lambda_{k0}|}C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) + C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right] \\
& - \frac{2c\lambda_k}{|\lambda_{k0}|}\left(1 - \sum_i |\lambda_i|\right) \quad k = 1, \cdots, n;
\end{aligned}
$$

$$\begin{aligned}
\boldsymbol{H}_{k,l} =\ & A_{11}\left[C^{C,C}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right]\\
& + A_{22}\left[\frac{\lambda_k \lambda_l}{|\lambda_{k0}\lambda_{l0}|}\left[C^{R,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right] + \frac{I_{\{k=l\}}(\boldsymbol{\lambda})}{|\lambda_{k0}|}f^{R,R}\left(\boldsymbol{\lambda}, \lambda_k\right)\right]\\
& + 2A_{12}\left[\frac{\lambda_l}{|\lambda_{l0}|}C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_l) + \frac{\lambda_k}{|\lambda_{k0}|}C^{C,R}(\boldsymbol{x}_l - \boldsymbol{x}_k)\right]\\
& + \frac{2A_{12}I_{\{k=l\}}(\boldsymbol{\lambda})}{|\lambda_{k0}|}\left[\sum_i \lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) - C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right]\\
& + \frac{2c\lambda_k\lambda_l}{|\lambda_{k0}||\lambda_{l0}|} - I_{\{k=l\}}(\boldsymbol{\lambda})\left[\frac{2c}{|\lambda_{k0}|}\left(1 - \sum_i |\lambda_i|\right)\right] \qquad k,l = 1, \cdots, n.
\end{aligned}$$

For OK,

$$\begin{aligned}
\boldsymbol{G}_k =\ & A_{11}\left[\sum_i \lambda_i\left[C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_i)\right] - 2C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right]\\
& + A_{22}\left[\sum_i \lambda_i\left[C^{R,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_i)\right] - 2C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right]\\
& + 2A_{12}\left[\sum_i \lambda_i\left[C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) + C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k)\right]\right]\\
& - 2A_{12}\left[C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) + C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right]\\
& - \frac{2}{c}\left(1 - \sum_i \lambda_i\right) - \frac{c}{\lambda_k} \qquad k = 1, \cdots, n;
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{H}_{k,l} =\ & A_{11}\left[C^{C,C}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right]\\
& + A_{22}\left[C^{R,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right]\\
& + A_{12}\left[C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_l) + C^{C,R}(\boldsymbol{x}_l - \boldsymbol{x}_k)\right]\\
& + \frac{2}{c} + \frac{c}{\lambda_k^2}I_{\{k=l\}}(\boldsymbol{\lambda}) \qquad k,l = 1, \cdots, n.
\end{aligned}$$

4.4.3 Adjustments for effective zero weights

The assumption for approximation (4.7) to work is that $\lambda_{i0} \neq 0$, $(i = 1, \cdots, n)$. Thus, zero estimates are not allowed. Also, the calculation of $\ln(\lambda_i)$ gets very unstable when λ_i is close to zero. To guard against zero estimates, a natural strategy is to set small values to zero and exclude them from the next iteration. Such a strategy reduces the dimension of the gradient and Hessian and in some instances greatly speeds the computation. However, it

suffers from a big drawback that once a weight parameter is set to zero, it remains zero in the ensuing iterations. This is potentially problematic, for example, in the event that some parameters are close to zero initially but would tend away from zero as the penalty iterates. To avoid this problem, instead of sub-setting the parameter vector at each iteration, we propose adjustments to the penalties, which still allow the λ_i's to approach zero using the tolerance criteria described previously, yet initially prevents movement to zero.

For SK, we simply add a barrier function that prohibits the λ_i's from approaching zero. This leads to the following new penalty:

$$\tilde{P}^{SK}(\boldsymbol{\lambda}, c) = -\frac{c}{n^2} \sum_i \ln\left(\lambda_i^2\right) + \frac{(1 - \sum_i |\lambda_i|)^2}{c}. \tag{4.8}$$

Dividing the barrier penalty parameter by n^2 ensures a balance between the penalty terms so that one does not dominate the other. Now that the estimates are guarded against zero, the approximation of the absolute value function is no longer needed. Thus, the gradient and Hessian associated with the new penalty have the following simplified forms.

Theorem 4. *The gradient $\boldsymbol{G}$ and Hessian $\boldsymbol{H}$ of $Q(\boldsymbol{\lambda})$ for SK with the new penalty $\tilde{P}^{SK}$ defined in (4.8) are given as*

$$
\begin{aligned}
\tilde{\boldsymbol{G}}^{SK} &= A_{11}\left[\sum_i \lambda_i \left[C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_i)\right] - 2C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right] \\
&\quad + A_{22}\left[\operatorname{sgn}(\lambda_k) f^{R,R}(\boldsymbol{\lambda}, \lambda_k)\right] \\
&\quad + 2A_{12}\left[\sum_i \left[|\lambda_i| C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) + \operatorname{sgn}(\lambda_k)\lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k)\right]\right] \\
&\quad - 2A_{12}\left[C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) + C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}^*)\right] \\
&\quad - \frac{2c}{n^2 \lambda_k} - \frac{2\operatorname{sgn}\lambda_k}{c}\left(1 - \sum_i |\lambda_i|\right) \qquad k = 1, \cdots, n;
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\boldsymbol{H}}^{SK} &= A_{11}\left[C^{C,C}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right] \\
&\quad + A_{22}\left[\operatorname{sgn}(\lambda_k \lambda_l)\left[C^{R,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_l)\right]\right]
\end{aligned}
$$

$$+2A_{12}\left[\operatorname{sgn}(\lambda_l)C^{C,R}(\boldsymbol{x}_k-\boldsymbol{x}_l)+\operatorname{sgn}(\lambda_k)C^{C,R}(\boldsymbol{x}_l-\boldsymbol{x}_k)\right]$$
$$+\frac{2\operatorname{sgn}(\lambda_k\lambda_l)}{c}+I_{\{k=l\}}(\boldsymbol{\lambda})\frac{2c}{(n\lambda_k)^2}\qquad k,l=1,\cdots,n.$$

For OK, in the current version of the penalty, the cost of reaching the zero-valued boundary is too high relative to the equality constraint. As a result, the algorithm tends to produce a non-zero weight for each lambda and predictions very close to the global average. Therefore, we add a small tolerance to the logarithmic penalty, which shifts the boundary of the penalty slightly below zero to allow λ_i to reach zero without leaving the feasible region. In the mean time, any value that effectively dips below zero will be set equal to zero before the next iteration. In addition, we divide the weight of the quadratic penalty by n to ensure that growth of the equality constraint (as c decreases) does not outpace the corresponding decrease in the n inequality constraints. With these adjustments, the new penalty is given as

$$\tilde{P}^{OK}=-c\sum_i\ln\left(\lambda_i+tolz\right)+\frac{(1-\sum_i\lambda_i)^2}{c*n}.$$

4.5 Simulation

A simulation study is carried out to demonstrate the finite sample performances of the numerical implementation of the proposed interval-valued kriging models. This simulation and the ensuing application to design snow load prediction make use of several ancillary packages to R (Bivand et al., 2018, 2013; Douglas Nychka et al., 2015; Hijmans, 2016; Kahle and Wickham, 2013; Neuwirth, 2014; Pebesma and Bivand, 2005; Wickham, 2017). The experimental design of choice is a Gaussian random field with an exponential covariance structure for both center and radius spanning a $\mathcal{R}=[0,2]\times[0,2]$ square. In the first scenario, we simulate a random field with no center-radius interaction. Let $X(\boldsymbol{x})$ and $Y(\boldsymbol{x})$ be independent standard normal random functions of location $\boldsymbol{x}\in\mathcal{R}$. The center and radius functions are generated by

$$Z^C(\boldsymbol{x})\quad=\quad X(\boldsymbol{x}),\tag{4.9}$$

$$Z^R(\boldsymbol{x})\quad=\quad \theta_r Y(\boldsymbol{x})+z_0,\tag{4.10}$$

where θ_r and z_0 are the scale and shift parameters respectively. The shift parameter is used to ensure that all of the simulated values of the radius remain positive. The covariance functions for this field are defined as

$$
\begin{aligned}
C^{C,C}(\boldsymbol{h}) &= \exp\left\{-\frac{||\boldsymbol{h}||_2}{4}\right\}, & (4.11)\\
C^{R,R}(\boldsymbol{h}) &= \theta_r^2 \exp\left\{-\frac{||\boldsymbol{h}||_2}{5}\right\}, & (4.12)\\
C^{C,R}(\boldsymbol{h}) &= C^{R,C}(\boldsymbol{h}) = \boldsymbol{0}. & (4.13)
\end{aligned}
$$

The second scenario assumes a strong interaction between the center and radius. We simulate this with

$$
Z^R(\boldsymbol{x}) = \theta_r Z^C(\boldsymbol{x}) + z_0 + \epsilon
$$

where $\epsilon \sim N(0, \theta_\epsilon^2)$ and $Z^C(\boldsymbol{x})$ is defined identical to (4.9). The covariance for the centers are defined identical to (4.11), while the covariance for the radii and center/radius interaction are seen to be

$$
\begin{aligned}
C^{R,R}(\boldsymbol{h}) &= \theta_r^2 \exp\left\{-\frac{||\boldsymbol{h}||_2}{4}\right\} + \theta_\epsilon^2,\\
C^{C,R}(\boldsymbol{h}) &= \theta_r \exp\left\{-\frac{||\boldsymbol{h}||_2}{4}\right\}.
\end{aligned}
$$

In both scenarios, the simulations of the radii are re-run whenever $\min\left(Z^R(\boldsymbol{x})\right) < 0$.

We generate $n = \{50, 100, 150, 200, 300, 400, 500, 600\}$ measurement locations in the random field from a complete spatial random process. Interval-valued SK and OK models are fit to these measurements to make predictions at $q = 100$ randomly selected locations on a regular 10×10 grid. The only difference between the SK and OK kriging models in this process is that SK assumes a known mean (which is set to zero), while OK makes no such assumption. Figure 4.1 shows an illustration of simulated data for both scenarios when $n = 100$, $\theta_r = \{\frac{1}{5}, \frac{1}{3}, 1\}$, and $\theta_\epsilon = \frac{1}{10}$. The data values are displayed by the degrees of darkness in the color maps, with darker colors implying larger values. Figure 4.2 plots the empirical variograms obtained from the simulated data for both scenarios visualized in

Figure 4.1 and compares them to the theoretical variograms. We can see that the empirical variograms overall match the theoretical ones fairly well, except for a few large deviations for some cases, which is expected due to randomness.

The matrix A of the ρ_K metric is selected as the sample covariance matrix between center and radius, i.e.

$$\begin{bmatrix} \frac{1}{n-1}\sum_{x^*}\left(Z^C(x^*)-\bar{Z}^C\right)^2 & \frac{1}{n-1}\sum_{x^*}\left(Z^C(x^*)-\bar{Z}^C\right)\left(Z^R(x^*)-\bar{Z}^R\right) \\ \frac{1}{n-1}\sum_{x^*}\left(Z^C(x^*)-\bar{Z}^C\right)\left(Z^R(x^*)-\bar{Z}^R\right) & \frac{1}{n-1}\sum_{x^*}\left(Z^R(x^*)-\bar{Z}^R\right)^2 \end{bmatrix}$$

where $\bar{Z}^C$ and $\bar{Z}^R$ are the sample mean for the sample interval centers and radii respectively. In theory, this matrix is known to be positive semidefinite and thus qualifies for use in ρ_K.

Simulations and estimations for each sample size are repeated 10 times independently and each time we record the RMSE for $q = 100$ prediction locations with respect to the ρ_K metric defined as:

$$\text{RMSE} = \sqrt{\frac{1}{q}\sum_{x^*}\rho_k^2\left([\hat{Z}(x^*)],[Z(x^*)]\right)}$$

where the values of A correspond to the theoretical covariance matrix between interval centers and radii. We also record the RMSE's for the centers (C) and radii (R) individually defined as

$$\begin{aligned} \text{RMSE(C)} &= \sqrt{\frac{1}{q}\sum_{x^*}\left(\hat{Z}(x^*)^C - Z(x^*)^C\right)^2} \\ \text{RMSE(R)} &= \sqrt{\frac{1}{q}\sum_{x^*}\left(\hat{Z}(x^*)^R - Z(x^*)^R\right)^2}. \end{aligned}$$

Tables 4.1 and 4.2 list the average values of the RMSE over 10 independent repetitions for each scenario and sample size combination when $\theta_r = \frac{1}{3}$. These values are likewise visualized for all considered values of θ_r in Figure 4.3. The SK results include errors using the original (SK-Fast) and adjusted (SK-Full) penalty functions described in Section 4.4.3. In all cases, the average prediction error for each interval-valued implementation decreases as sample size increases, demonstrating the empirical convergence of our method to th'e optimal kriging solution. Further, for sample sizes greater than 50, the errors for the full

and fast implementations of SK are nearly identical to OK. In fact, when variance of the interval centers and radii are roughly equal, the fast implementation outperforms the full method when the sample size is 500 or more. The out-performance of SK (Fast) in this instance demonstrates the difficulty of preventing zero-valued kriging weights when the sample size is large. Finally, all kriging models are compared to the sample mean as the benchmark predictor. In each instance, the kriging predictions result in a 50% improvement when compared to the global mean predictions for sample sizes of 300 or greater. These results demonstrate the empirical convergence of our interval-valued kriging models to the optimal solution using the ρ_k^2 metric.

Table 4.1: Average RMSE from 10 simulations of an interval-valued random field with no radius/center interaction and $\theta_r = \frac{1}{3}$.

Model		sample size (n)							
		50	100	150	200	300	400	500	600
	RMSE	0.81	0.74	0.64	0.6	0.55	0.52	0.5	0.46
SK-Full	RMSE(C)	0.77	0.7	0.6	0.56	0.51	0.49	0.46	0.43
	RMSE(R)	0.27	0.24	0.22	0.21	0.19	0.17	0.17	0.17
	RMSE	0.88	0.75	0.64	0.6	0.55	0.52	0.5	0.46
SK-Fast	RMSE(C)	0.82	0.72	0.6	0.56	0.51	0.49	0.46	0.43
	RMSE(R)	0.32	0.24	0.22	0.21	0.19	0.17	0.17	0.17
	RMSE	0.82	0.74	0.65	0.6	0.55	0.53	0.5	0.47
OK	RMSE(C)	0.77	0.7	0.61	0.57	0.52	0.5	0.47	0.44
	RMSE(R)	0.28	0.24	0.22	0.21	0.2	0.18	0.17	0.17
	RMSE	0.99	1.01	1.03	1.02	1.08	1.04	0.97	1.01
Mean	RMSE(C)	0.93	0.95	0.98	0.96	1.03	0.99	0.91	0.95
	RMSE(R)	0.34	0.33	0.32	0.32	0.33	0.32	0.32	0.33

4.6 Design snow load predictions for Utah

In this section, we demonstrate an application of our interval-valued kriging models to predict design snow loads in Utah. We create an interval-valued design snow load dataset with the same daily data used to create UT-2017. The final intervals account for various uncertainties in the design load estimation process outlined in Figure 1.2. These imprecise estimates of design snow load are then used as input to interval-valued kriging. Lastly, the
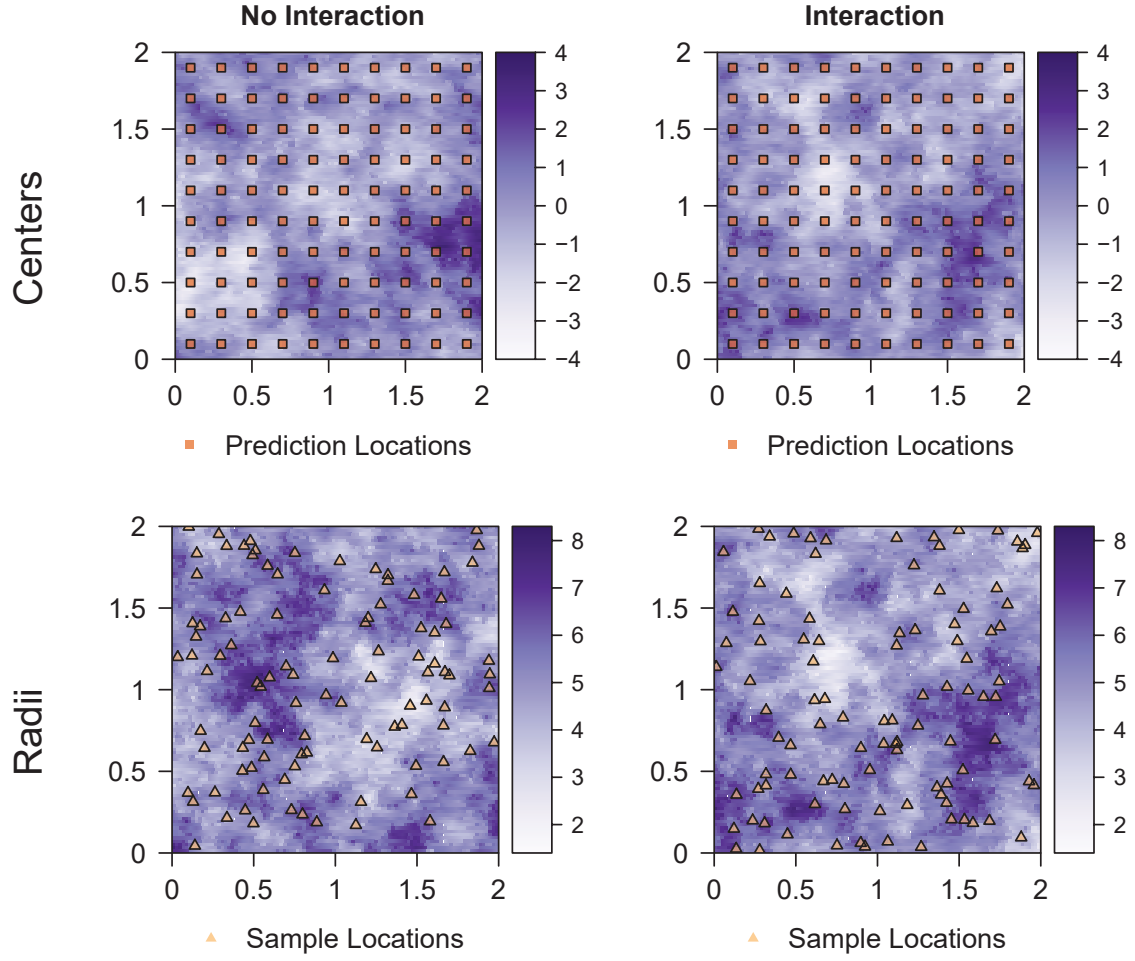
Fig. 4.1: Maps of simulated data for both scenarios with 100 sample and prediction locations $(\theta_r = 1, z_0 = 5, \text{ and } \theta_\epsilon = \frac{1}{10})$.

Table 4.2: Average RMSE from 10 simulations of an interval-valued random field with a radius/center interaction, $\theta_r = \frac{1}{3}$, and $\theta_\epsilon = \frac{1}{10}$.

Model		sample size (n)							
		50	**100**	**150**	**200**	**300**	**400**	**500**	**600**
	RMSE	0.82	0.72	0.65	0.61	0.56	0.53	0.51	0.47
SK-Full	RMSE(C)	0.77	0.67	0.61	0.56	0.52	0.49	0.47	0.44
	RMSE(R)	0.28	0.25	0.23	0.23	0.2	0.2	0.19	0.18
	RMSE	0.86	0.72	0.65	0.61	0.56	0.53	0.51	0.47
SK-Fast	RMSE(C)	0.81	0.68	0.61	0.56	0.52	0.49	0.47	0.44
	RMSE(R)	0.29	0.25	0.24	0.23	0.2	0.2	0.19	0.18
	RMSE	0.81	0.72	0.66	0.61	0.57	0.54	0.51	0.48
OK	RMSE(C)	0.76	0.68	0.61	0.57	0.53	0.5	0.47	0.44
	RMSE(R)	0.28	0.25	0.23	0.23	0.21	0.2	0.19	0.18
	RMSE	1.05	1.05	1.03	1.04	1.04	1.02	1.04	0.99
Mean	RMSE(C)	0.99	1	0.97	0.98	0.98	0.96	0.98	0.93
	RMSE(R)	0.35	0.35	0.34	0.35	0.34	0.34	0.34	0.33
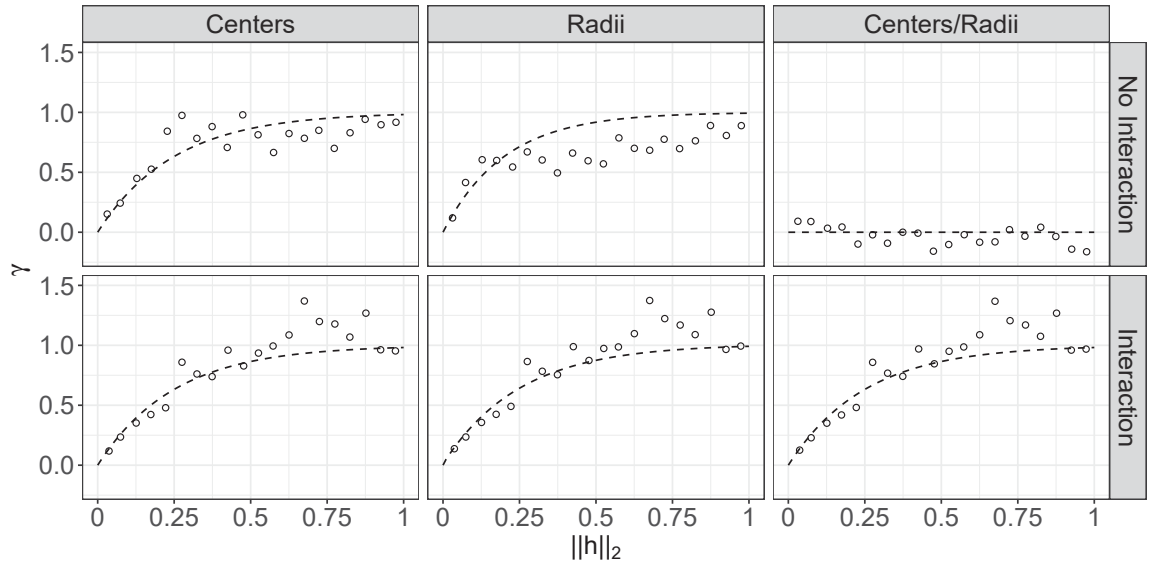
Fig. 4.2: Plots of the theoretical variograms (lines) and their empirical counterparts (points) computed from the simulated data shown in Figure 4.1. In some instances, the empirical variograms do not exactly match the theoretical ideal.
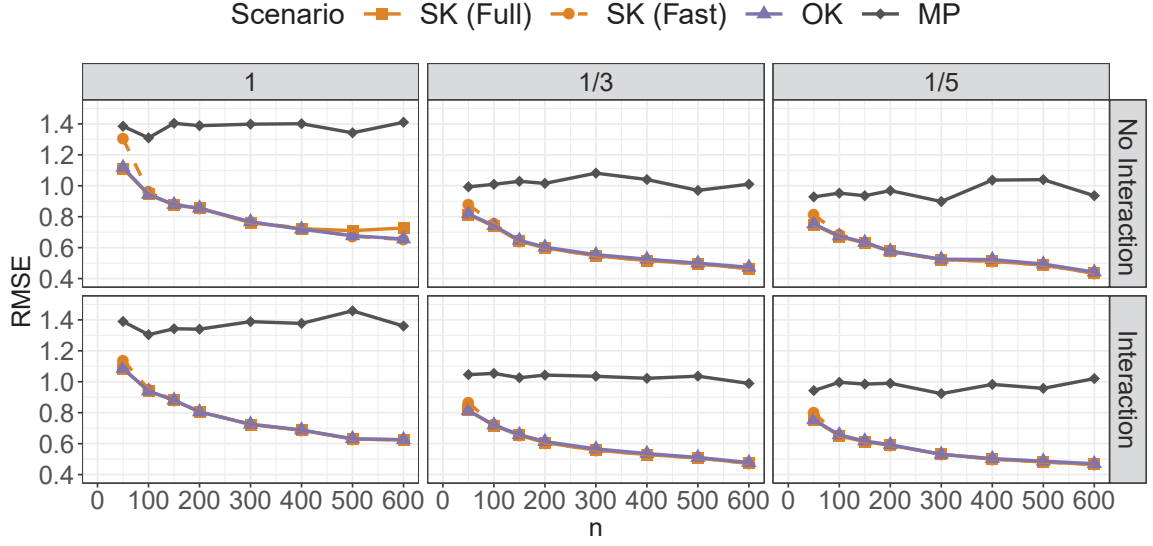
Fig. 4.3: Comparison of the overall root mean square error (RMSE), as well as separate comparisons of the RMSE for the center and radius, for several interval-valued kriging models. Results are calculated for each model using various values of θ_r at $q = 100$ prediction locations on a regular grid.

resulting interval-valued predictions are compared to estimates from SKLM, which is the most similar point-valued method.

4.6.1 Defining the interval-valued data

We define interval-valued design snow loads ($[q_s^*(\boldsymbol{x}_\alpha)]$) through the following process. This process combines multiple sources of uncertainty in the design snow load estimation process into a single interval. The resulting interval-valued data are then used as input into our interval-valued kriging in the subsequent analyses.

1. *Create daily snow load intervals using various depth-to-load conversion methods.*

   Recall that many measurement locations do not measure the load of settled snow directly, thus requiring an estimate from snow depth. Chapter 2 summarized the various methods used by western states estimate snow load from snow depth. Figure 4.4 compares depth-to-load prediction methods for various snow depths on different days of the year. This figure shows that load predictions can vary widely depending
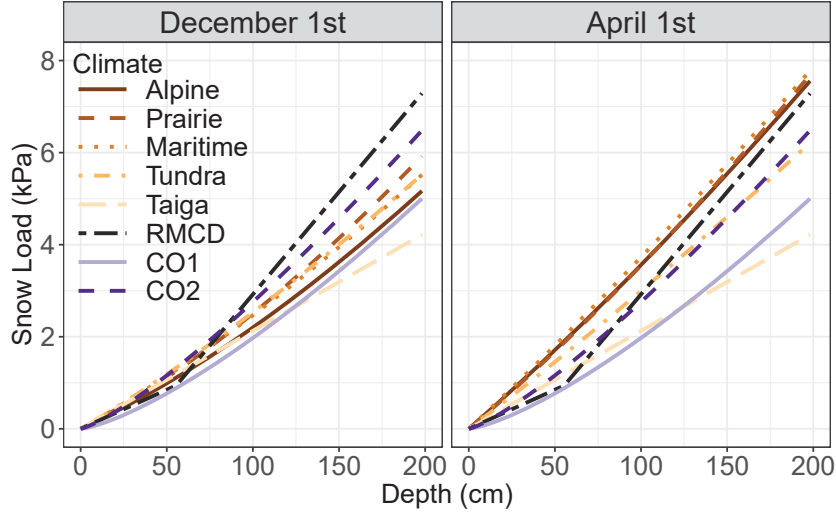
Fig. 4.4: Comparisons of the different depth-to-load conversion methods for various depths and days of the year.

on the selected conversion method. In order to capture the imprecision induced when a depth-to-load conversion is required, we characterize these predictions using the interval $[q_s(\boldsymbol{x}_\alpha)] = [\min(\boldsymbol{g}), \max(\boldsymbol{g})]$ for each recorded measurement, where $\boldsymbol{g}$ represents the vector of all considered depth-to-load predictions. This interval is graphically characterized as the envelope of depth-to-weight conversions observed in Figure 4.4. Under this definition, direct measurements of snow load are characterized as intervals of length zero.

2. *Create lower and upper annual maximums snow loads using daily interval-valued data.*

   We create two sets of annual maximum snow loads through separate considerations of the upper and lower endpoints of the daily snow load intervals created in Step 1. Using this strategy, the upper and lower maximum snow load for any given snow season need not occur on the same day.

3. *Account for uncertainty from distribution fitting.*

The American Society of Civil Engineers (ASCE) conventionally assumes a log-normal distribution for the annual maximum snow loads at a given station and the design snow load is defined as the associated 98th percentile (ASCE, 2017). However, very often in practice, this assumption is not appropriate and can lead to inaccurate estimates of the design snow load. We seek a more robust estimate of the design snow load by considering five distinct probability distributions: log-normal, gamma, and extreme value distributions types I, II, and III. Each distribution is characterized by its location ψ and scale θ parameters, whose values are estimated via ML.

Assume that the upper and lower annual maximum snow load from the previous step each follows one of the five distributions denoted by $F_k(\cdot)$, $k = 1, \cdots, 5$. The ML estimates of the parameters are $(\hat{\psi}_k, \hat{\theta}_k)$, which are asymptotically normal under the hypothesized i.i.d. assumption. To account for the uncertainty of fitting, we estimate the 98th percentile of each distribution using all values of $\hat{\psi}$ and $\hat{\theta}$ in their 95% asymptotic confidence interval and calculate the interval that contains all the estimates.

4. *Create final interval-valued design snow load estimates.*

   The above process generates two intervals for each of the five distributions fit to the upper/lower annual maximums resulting from Step 2. The final interval is created by considering the five upper endpoints of the intervals created with the upper annual maximums and the five lower endpoints of the intervals created with the lower annual maximums. It is very likely that at least one of the five probability distributions used in the previous step will provide a poor fit for at least one of these endpoints. To guard against poor fits, the final design snow load interval is created using only the median upper and lower endpoints of each interval. This means that the median of the five upper endpoints from the upper annual maximum intervals is selected as the final upper endpoint and the median of the five lower endpoints from the lower annual maximum intervals is selected as the final lower endpoint. This median-based strategy is a practical method for ensuring that the final intervals have a realistic range.

4.6.2 Analyses and results

Denote the final design snow load intervals by $[q_s^L(\boldsymbol{x}), q_s^U(\boldsymbol{x})]$. All ensuing analyses use these intervals on the log-scale, i.e.

$$[\log(q_s^L(\boldsymbol{x})), \log(q_s^U(\boldsymbol{x}))] = [l_s^C(\boldsymbol{x}) - l_s^R(\boldsymbol{x}), l_s^C(\boldsymbol{x}) + l_s^R(\boldsymbol{x})].$$

Figure 4.5 compares $l_s^C(\boldsymbol{x})$ to the original point-valued data $(\log(q_s^P(\boldsymbol{x}_\alpha)))$ across space and elevation. To achieve stationarity, we need to remove the elevation effect in both interval center and radius (shown in Figure 4.5) prior to input into the kriging model. The elevation trends for both cases are modeled as

$$
\begin{aligned}
l_s^C(\boldsymbol{x}) &= \beta_0 + \beta_1 A(\boldsymbol{x}) + R^C(\boldsymbol{x}), &\quad (4.14)\\
l_s^R(\boldsymbol{x}) &= R^R(\boldsymbol{x}) * \left(\frac{c}{A(\boldsymbol{x})}\right), &\quad (4.15)
\end{aligned}
$$

where c is a scaling factor, which in this case is set as the median elevation value in the snow load dataset. It is the residual intervals, $\left[R^C - R^R, R^C + R^R\right]$, that are used as input in interval-valued kriging.

Figure 4.6 shows the empirical variograms of the interval data alongside an interval-valued design snow load prediction map. The variograms reveal a spatial relationship between the centers, radii, and centers/radii interaction. In practice, these empirical variograms are used to define the theoretical variograms shown as dotted lines in this same figure (Bivand et al., 2013). We selected spherical models for both the center and radius variograms and a Gaussian model for the center/radius interaction. (See Goovaerts (1997) for a summary of other commonly used variograms.) Each model includes a nugget effect and parameters are selected using a weighted least squares fitting algorithm in the gstat package. Their
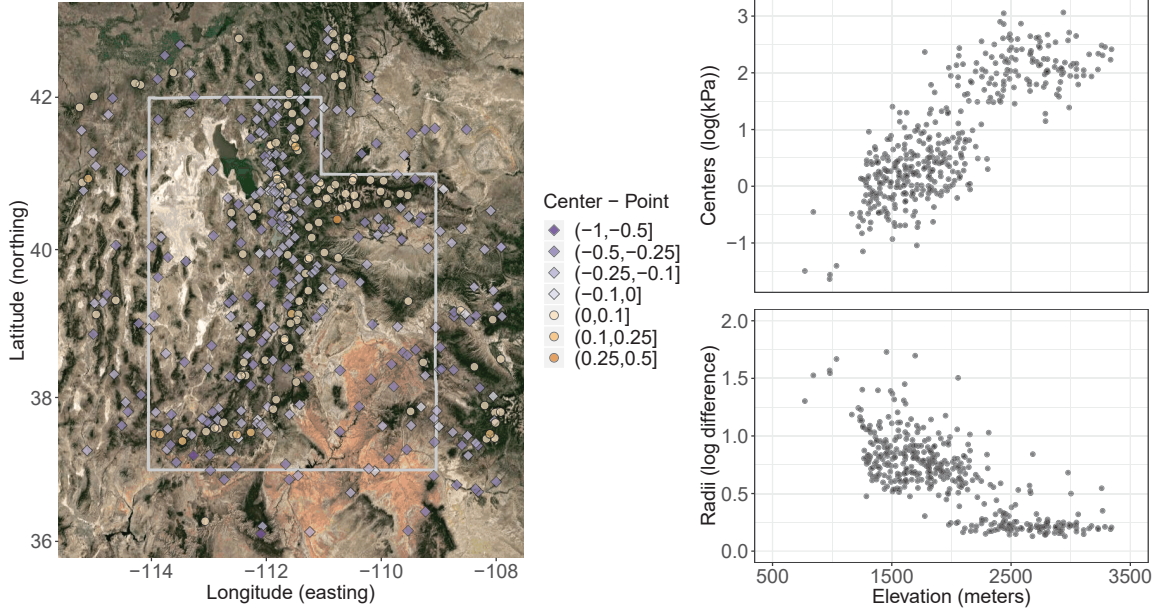
Fig. 4.5: (Clockwise from top left): (1) A map comparing interval-centers to the original point predictions on the log-scale. Purple diamonds indicate places where the centers of the log-intervals are less than the log-transformation of the point value predictions. (2) Interval centers (log-scale) across elevation. (3) Interval radii (log-scale) across elevation.

mathematical definitions are given as

$$
\gamma^C(||\boldsymbol{h}||) = \begin{cases} 0 & ||\boldsymbol{h}|| = 0, \\[2mm] 0.21\left[1.5\frac{||\boldsymbol{h}||}{193} - 0.5\left(\frac{||\boldsymbol{h}||}{193}\right)^3\right] + 0.09 & 0 < ||\boldsymbol{h}|| \le 193, \\[2mm] 0.3 & ||\boldsymbol{h}|| > 193; \end{cases}
$$

$$
\gamma^R(||\boldsymbol{h}||) = \begin{cases} 0 & ||\boldsymbol{h}|| = 0, \\[2mm] 0.03\left[1.5\frac{||\boldsymbol{h}||}{139} - 0.5\left(\frac{||\boldsymbol{h}||}{139}\right)^3\right] + 0.03 & 0 < ||\boldsymbol{h}|| \le 139, \\[2mm] 0.06 & ||\boldsymbol{h}|| > 139; \end{cases}
$$

$$
\gamma^{CR} = \begin{cases} 0 & ||\boldsymbol{h}|| = 0, \\[2mm] -0.05\left[1 - \exp\{\frac{-3||\boldsymbol{h}||^2}{187^2}\}\right] - 0.01 & ||\boldsymbol{h}|| > 0. \end{cases}
$$

We elected to use the SK model because $R^C(\cdot)$ has, both in theory and practice, a

known mean of zero. The resulting kriging predictions are input into (4.14) and (4.15) and exponentiated for final load predictions. An interval map of these final predictions shown in Figure 4.6 is intended for simultaneous visualizations of center and radius. The darkness of the grid indicates the interval center while the size of the circle within each grid represents the interval radius. The size of the circle is scaled so that the grid cell with the largest radius will have a circle exactly circumscribed within the cell. It is important to distinguish the predicted interval radius from the prediction variance. Recall that the prediction variance defined in (4.3) relies on the stationarity assumptions outlined in Definition 1. One of these assumptions is that the covariances that comprise the prediction variance are solely a function of the difference between locations. This essentially makes the kriging variance a measure of data quantity across space. In contrast, the interval radius measures the imprecision of the surrounding measurements used in predictions. The kriging variance and the interval radii measure different aspects of uncertainty in the prediction and need not be related. A comparison between the predicted radius and the kriging variance using interval-valued simple kriging is visualized in Figure 4.7. Kriging radii are taken directly from the interval predictions while kriging variance is calculated directly as in (4.3). The overall correlation between these two maps is 0.57, but there is certainly no one-to-one correspondence and the difference can be quite large at certain locations. For example, the two rectangles in this figure highlight two areas where the spatial patterns of the radii and kriging variances differ greatly. In the northern rectangle, the lowest variances occur in the eastern portion while the smallest interval radii occur in the western portion. This is because the highest concentration of measurements occur in the eastern portion of the rectangle, while the most precise design snow load estimates are located in the western portion of this rectangle. Similar conclusions can be drawn from spatial pattern differences in the southern rectangle. Thus, the interval-valued kriging allows for simultaneous measures of data quantity *and* quality, both of which are essential to evaluating the certainty of the model predictions.
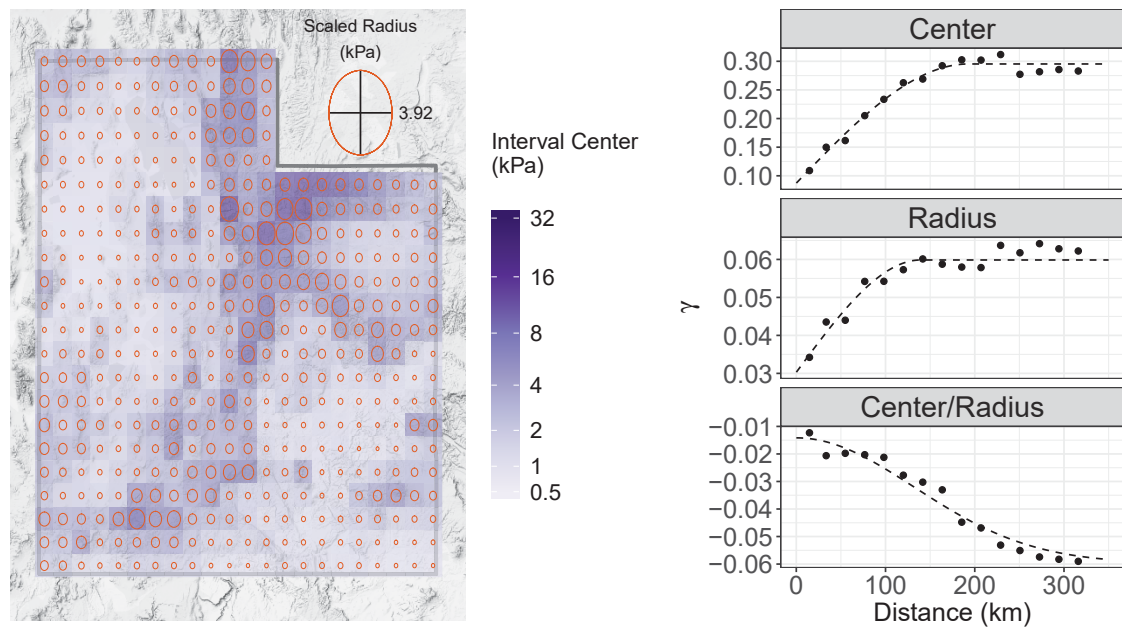
Fig. 4.6: (Left) An interval-valued design snow load prediction map for Utah. The darkness of the grid-cell indicates the interval center while the size of the circle within each cell represents the interval radius. Circles are scaled so that the maximum radius (indicated in the legend) is exactly circumscribed within the grid-cell. (Right) Plots of the empirical and theoretical variograms used in the numerical implementation of interval-valued simple kriging.
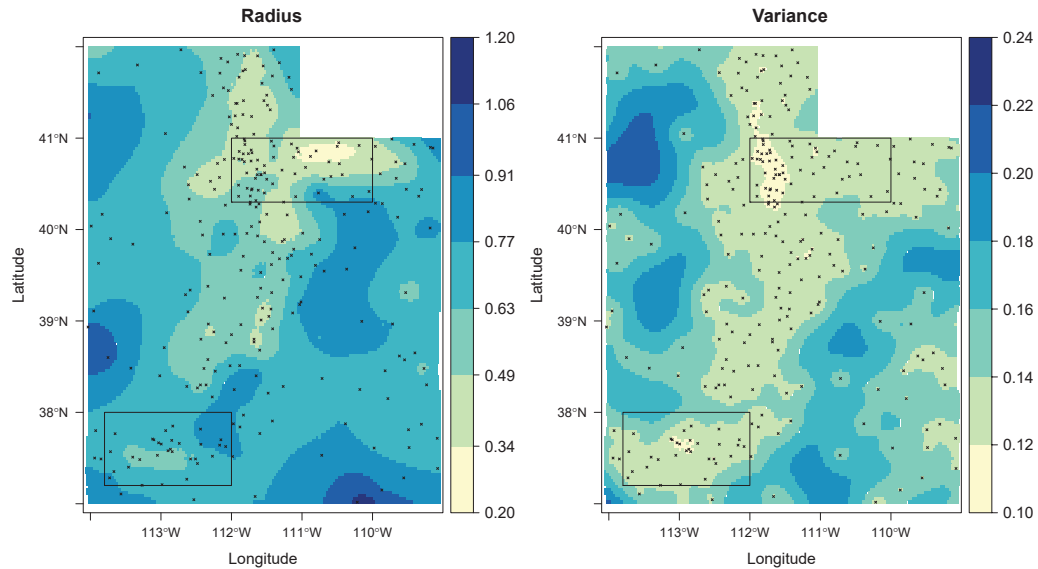
Fig. 4.7: Maps of the prediction variance and the predicted interval radius for the state of Utah. The black points denote measurement locations.

4.6.3 Discussion

Figure 4.8 compares the predicted intervals to the original point predictions. In the map, orange color represents areas where the point predictions are relatively higher than the interval centers and purple color represents the contrary. Generally, the interval centers are consistently lower along the Wasatch mountain range (running north to south through the middle of the state) and consistently higher in lower-elevation regions of the state. The strong effect of elevation on final design snow load predictions makes it difficult to compare the interval-valued and point-valued results at high elevation locations. At these locations, the effect of the kriging prediction is marginal compared to the regression prediction based on elevation. This results in small differences between the interval centers and original point valued predictions at high elevation locations. Rather, we focus on low-elevation areas where the original point-valued predictions are below the 40% quantile of the interval. These areas tend to have low population densities and nearly all receive less snow than is typical for the state. As such, the few measurement locations that do exist in these regions tend to have less

consistent measurements of snow. This creates region-wide data precision issues evidenced by the relatively large interval radii occurring at nearby measurement locations. Large radii on the log-scale, when exponentiated back to the original scale, disproportionately inflate the upper endpoints of the final intervals and leave the point predictions in the lower portion of the intervals.

In the right of plot Figure 4.8, we picked a few locations to show their point predictions directly on the interval predictions. These locations were selected to adequately represent the diverse geography and climates of populated locations in the state. Some of them demonstrate very interesting comparative results. For example, the original point predictions for St. George and Moab, Utah were almost identical, yet the interval length for St. George is roughly twice the length of Moab's. When looking at the measurement locations surrounding St. George and Moab, we found that the average data interval length for the top five measurement locations (in terms of kriging weights) for St. George is more than double the average for the top five measurement locations for Moab (3.4 kPa vs 1.6 kPa, as calculated on a log-scale after removing the effect of elevation). Snow accumulation is not typically of interest for the warm climate of St. George. For this reason, there is a general lack of recorded observations at measurement locations near St. George, which manifests itself in less precise (i.e. larger interval-radii) estimates of design snow load.

It is important that data imprecision is not confused with data coverage, which is measured by the kriging variance. For example, St. George has a less precise prediction than Wendover, Utah, as indicated by the interval lengths. Yet Wendover has a much larger kriging prediction variance than St. George (0.21 vs 0.14, as calculated on a log-scale after removing the effect of elevation). Thus the prediction at St. George is uncertain due to imprecise measurements at surrounding locations, while the prediction at Wendover is uncertain due to a lack of surrounding measurement locations.

Finally, the patterns observed in the left plot of Figure 4.8 show some interesting similarities with Colorado's reliability-based snow load study: locations with higher relative uncertainty in the 50 year estimates tend to have higher design snow load predictions. The

Colorado study used Monte-Carlo simulations of several load factors in light of existing resistance factors to determine the design snow load that would satisfy a desired reliability index. Load and resistance factors are partial safety factors which rationally increase the predicted loading and reduce the predicted resistance simultaneously. The goal of the Colorado study was to select snow loads that would result in a constant risk, in light of the observed differences in uncertainty across the state. Snow loads were simulated using a log-normal distribution estimated from the upper third of all annual maximums at each measurement location. This simulation-based strategy accounts for some of the year-to-year variability in observed design snow loads when determining appropriate reliability-based design snow loads. Using this strategy, mountainous locations had reliability-based snow loads about 90% of the typical 50 year load while the eastern Colorado plains region had reliability-based loads as much as double the 50 year loads (Liel et al., 2017).

The Colorado methodology represents an important step forward in the design snow load prediction problem by characterizing loads as distributions rather than single values. However, the strategy does not consider uncertainty in the depth-to-load conversions or the distribution fitting process. In addition, their ensuing spatial interpolation is still based on point-valued snow loads derived from fixed load factors. This again robs predictions between measurement locations of proper context, as it is unclear if the high predicted values are due to surrounding measurement locations with large means, or distributions with heavy tails. Preserving context through intervals opens the possibility for defining location specific load factors as a function of the relative variability of each interval. In addition to design snow load prediction, other models based on this framework could easily be constructed for similar predictions of other extreme climatic events where the data inputs are inherently imprecise.

4.7 Reflections

The primary contribution of this chapter is the development of a mathematically tractable, numerically feasible set of interval-valued kriging models. These models leverage the many advancements in set-valued statistics over the past 30 years to provide a well-
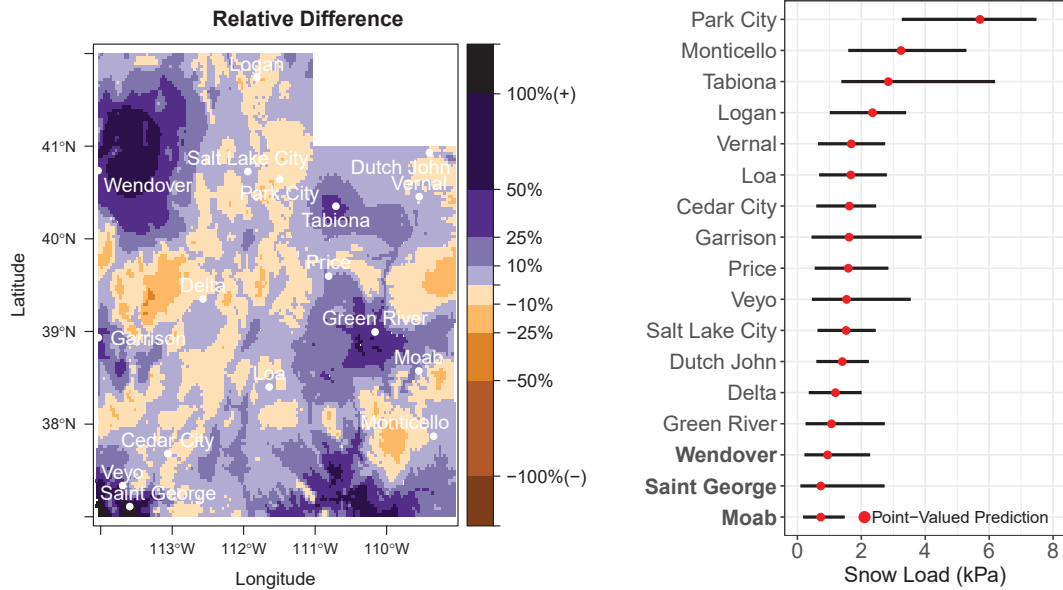
Fig. 4.8: (Left) Comparison of the relative difference between the predicted interval centers and the original point-valued predictions across space on the raw scale. Orange represents areas where the point predictions are higher than the interval centers and purple represents the contrary. (Right) Comparison of interval-valued simple kriging predictions to the original point-valued predictions at selected locations in Utah.

established notion of spatial covariance between intervals, which simplifies and clarifies the earlier development in Diamond (1988). The new models inherit much of their mathematical foundation from their point-valued counterparts and offer a natural interval-valued extension to anyone familiar with the kriging paradigm. We demonstrated the numerical feasibility of these models through a series of simulations on interval-valued stationary random fields, which showed the finite sample performance and empirical convergence of our numerical implementations.

Another important contribution of this chapter from the applied perspective is a novel approach suitable for estimating design loads as part of a reliability analysis where spatial variability is desired for the snow load, or other environmental load, prediction problems. This approach includes a new method that characterizes the uncertainties of the design load estimation process as interval-valued observations. This interval-valued dataset was used in a regression-kriging framework to make interval-valued design snow load prediction maps for the state of Utah. We also provided unique spatial visualizations of intervals that

allow for simultaneous comparisons of interval center and radius in a single figure. The comparison to the original point valued predictions revealed a systematic pattern: locations with imprecise estimates of design load tend to have interval centers larger than the point valued predictions. This phenomenon is consistent with the aims of any reliability analysis, where increased uncertainty in a process is ultimately balanced with higher load requirements, (i.e. load factors). The implications of these interval-predictions extend well beyond this initial comparison. In particular, the interval length (or the center/radius ratio) could be used to calculate location-specific reliability indices for design predictions rather simply relying on theoretical norms.

Although the center is a natural measure of level for an interval, it is not necessarily so in general. More rigorously, one needs to consider weighting within the interval and compute the level by means of an integral, e.g., $\int_{[0,1]} \left[\lambda s^L + (1-\lambda)s^U\right] \nu(d\lambda)$, where ν is a normalized measure on $[0, 1]$ characterizing the weighting. Future work on our interval-valued kriging models will include the development of such methods for computing the level, which is usually of great interest in practice. We also anticipate adjustments to the optimization algorithm that will allow for direct use of the variogram in the numerical implementation, rather than assuming the existence of $C(\mathbf{0})$. In addition, considerations of uncertainty in the theoretical variogram parameters, similar to Loquin and Dubois (2012), would be worth further investigation. Indeed, one of the crucial products of this chapter are the many future considerations made possible by a new mathematical foundation and computationally feasible implementation of interval-valued kriging. We conclude that our interval-valued kriging models provide a practical and important alternative for researchers looking to extend their kriging applications to accommodate interval-valued inputs. These accommodations of imprecise inputs are imperative to ensuring that the spatial models we create today can meet the data challenges of tomorrow. The following chapter outlines a generalized approach for handling imprecise spatial data.

4.8   Proofs

4.8.1 Proof of Theorem 1

Proof. The prediction variance is defined as

$$
\begin{aligned}
&E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right] \\
=\ & E\big[A_{11}\left(\hat{Z}^C(x^*) - Z^C(x^*)\right)^2 + A_{22}\left(\hat{Z}^R(x^*) - Z^R(x^*)\right)^2 + \\
& 2A_{12}\left(\hat{Z}^C(x^*) - Z^C(x^*)\right)\left(\hat{Z}^R(x^*) - Z^R(x^*)\right)\big] \\
:=\ & A_{11}\mathrm{I} + A_{22}\mathrm{II} + 2A_{12}\mathrm{III}.
\end{aligned}
\tag{4.16}
$$

First of all, by the unbiasedness of $\hat{Z}^C(\cdot)$,

$$
\begin{aligned}
\mathrm{I} &= E\left(\hat{Z}^C(x^*) - Z^C(x^*)\right)^2 \\
&= \mathrm{Var}\left(\hat{Z}^C(x^*) - Z^C(x^*)\right) \\
&= \mathrm{Var}\left(\sum \lambda_i Z^C(x_i)\right) + \mathrm{Var}\left(Z^C\right) - 2\mathrm{Cov}\left(\sum \lambda_i Z^C(x_i), Z^C(x^*)\right) \\
&= \sum_i\sum_j \lambda_i\lambda_j \mathrm{Cov}\left(Z^C(x_i), Z^C(x_j)\right) - 2\sum_i \lambda_i \mathrm{Cov}\left(Z^C(x_i), Z^C\right) + \mathrm{Var}\left(Z^C(x^*)\right) \\
&= \sum_i\sum_j \lambda_i\lambda_j C^{C,C}(x_i - x_j) - 2\sum_i \lambda_i C^{C,C}(x_i - x^*) + C^{C,C}(0).
\end{aligned}
\tag{4.17}
$$

Second, in the similar fashion,

$$
\begin{aligned}
\mathrm{II} &= E\left(\hat{Z}^R(x^*) - Z^R(x^*)\right)^2 \\
&= \sum_i\sum_j |\lambda_i\lambda_j|\, C^{R,R}(x_i - x_j) - 2\sum_i |\lambda_i|\, C^{R,R}(x_i - x^*) + C^{R,R}(0).
\end{aligned}
\tag{4.18}
$$

Finally,

$$
\begin{aligned}
\mathrm{III} &= E\left(\hat{Z}^C(x^*) - Z^C(x^*)\right)\left(\hat{Z}^R(x^*) - Z^R(x^*)\right) \\
&= \mathrm{Cov}\left(\hat{Z}^C(x^*) - Z^C(x^*), \hat{Z}^R(x^*) - Z^R(x^*)\right) \\
&= \mathrm{Cov}\left(\sum \lambda_i Z^C(x_i), \sum |\lambda_j| Z^R(x_j)\right) - \mathrm{Cov}\left(Z^C(x^*), \sum |\lambda_i| Z^R(x_i)\right) \\
&\quad - \mathrm{Cov}\left(\sum \lambda_i Z^C(x_i), Z^R(x^*)\right) + \mathrm{Cov}\left(Z^C(x^*), Z^R(x^*)\right)
\end{aligned}
$$

$$
\begin{aligned}
= \ &\sum_i \sum_j \lambda_i |\lambda_j| C^{C,R}(x_i - x_j) - \sum_i |\lambda_i| C^{R,C}(x_i - x^*) \\
&- \sum_i \lambda_i C^{C,R}(x_i - x^*) + C^{C,R}(0).
\end{aligned}
\tag{4.19}
$$

Plugging (4.17)-(4.19) into (4.16) completes the proof. $\qquad\square$

4.8.2 Proof of Theorem 2

Proof. Recall (4.16) in the proof of Theorem 1 that

$$
E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right] = A_{11}\mathrm{I} + A_{22}\mathrm{II} + 2A_{12}\mathrm{III}.
\tag{4.20}
$$

From (4.6)-(4.6), we have

$$
C^{C,C}(\boldsymbol{h}) = C^{C,C}(\boldsymbol{0}) - \gamma^C(\boldsymbol{h}),
\tag{4.21}
$$

$$
C^{R,R}(\boldsymbol{h}) = C^{R,R}(\boldsymbol{0}) - \gamma^R(\boldsymbol{h}),
\tag{4.22}
$$

$$
C^{C,R}(\boldsymbol{h}) + C^{C,R}(-\boldsymbol{h}) = 2\left[C^{C,R}(\boldsymbol{0}) - \gamma^{C,R}(\boldsymbol{h})\right].
\tag{4.23}
$$

Plugging (4.21) in (4.17) and by the unbiasedness constraints, we obtain

$$
\mathrm{I} = -\sum_i \sum_j \lambda_i \lambda_j \gamma^C(x_i - x_j) + 2\sum_i \lambda_i \gamma^C(x_i - x^*).
\tag{4.24}
$$

Similarly, by plugging (4.22) in (4.18), we get

$$
\mathrm{II} = -\sum_i \sum_j \lambda_i \lambda_j \gamma^R(x_i - x_j) + 2\sum_i \lambda_i \gamma^C(x_i - x^*).
\tag{4.25}
$$

Finally, the last term can be rewritten as

$$
\begin{aligned}
\mathrm{III} = \ &\left(\sum_i \lambda_i^2\right) C^{C,R}(\boldsymbol{0}) + \sum_{i<j} \lambda_i \lambda_j \left[C^{C,R}(x_i - x_j) + C^{C,R}\left(-(x_i - x_j)\right)\right] \\
&- \sum_i \lambda_i \left[C^{C,R}(x_i - x^*) + C^{C,R}\left(-(x_i - x^*)\right)\right] + C^{C,R}(\boldsymbol{0}).
\end{aligned}
\tag{4.26}
$$

Plugging (4.23) in (4.26), we get

$$
\begin{aligned}
III &= \left(\sum_i \lambda_i^2\right) C^{C,R}(\mathbf{0}) + 2\sum_{i<j} \lambda_i \lambda_j \left[C^{C,R}(\mathbf{0}) - \gamma^{C,R}(x_i - x_j)\right] \\
&\quad -2\sum_i \lambda_i \left[C^{C,R}(\mathbf{0}) - \gamma^{C,R}(x_i - x^*)\right] + C^{C,R}(\mathbf{0}) \\
&= C^{C,R}(\mathbf{0}) - 2\sum_{i<j} \lambda_i \lambda_j \gamma^{C,R}(x_i - x_j) - 2C^{C,R}(\mathbf{0}) + 2\sum_i \lambda_i \gamma^{C,R}(x_i - x^*) + C^{C,R}(\mathbf{0}) \\
&= -\sum_{i \neq j} \lambda_i \lambda_j \gamma^{C,R}(x_i - x_j) + 2\sum_i \lambda_i \gamma^{C,R}(x_i - x^*).
\end{aligned}
\tag{4.27}
$$

Plugging (4.24), (4.25), and (4.27) in (4.20) completes the proof. $\qquad\square$

4.8.3 Proof of Theorem 3

Proof. Define

$$
V^{C,C}(\boldsymbol{\lambda}) = \sum_i \sum_j \lambda_i \lambda_j C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_j) - 2\sum_i \lambda_i C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}^*),
\tag{4.28}
$$

$$
V^{R,R}(\boldsymbol{\lambda}) = \sum_i \sum_j |\lambda_i \lambda_j| \, C^{R,R}(\boldsymbol{x}_i - \boldsymbol{x}_j) - 2\sum_i |\lambda_i| \, C^{R,R}(\boldsymbol{x}_i - \boldsymbol{x}^*),
\tag{4.29}
$$

$$
\begin{aligned}
V^{C,R}(\boldsymbol{\lambda}) &= \sum_i \sum_j \lambda_i |\lambda_j| C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_j) - \sum_i |\lambda_i| C^{R,C}(\boldsymbol{x}_i - \boldsymbol{x}^*) \\
&\quad - \sum_i \lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}^*).
\end{aligned}
\tag{4.30}
$$

The prediction variance is then rewritten as

$$
E\left[\rho_K^2\left([\hat{Z}(\boldsymbol{x}^*)], [Z(\boldsymbol{x}^*)]\right)\right] = A_{11} V^{C,C}(\boldsymbol{\lambda}) + A_{22} V^{R,R}(\boldsymbol{\lambda}) + 2A_{12} V^{C,R}(\boldsymbol{\lambda}).
$$

It follows that

$$
Q(\lambda) = A_{11} V^{C,C}(\boldsymbol{\lambda}) + A_{22} V^{R,R}(\boldsymbol{\lambda}) + 2A_{12} V^{C,R}(\boldsymbol{\lambda}) + P(\boldsymbol{\lambda}, c),
\tag{4.31}
$$

and

$$
\begin{aligned}
\boldsymbol{G} = \nabla Q(\lambda) &= A_{11}\frac{\partial}{\partial\lambda_k}V^{C,C}(\boldsymbol{\lambda}) + A_{22}\frac{\partial}{\partial\lambda_k}V^{R,R}(\boldsymbol{\lambda}) + 2A_{12}\frac{\partial}{\partial\lambda_k}V^{C,R}(\boldsymbol{\lambda}) \\
&\quad + \frac{\partial}{\partial\lambda_k}P(\boldsymbol{\lambda},c) \quad (k=1,\ldots,n) \tag{4.32}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{H} = \nabla^2 Q(\lambda) &= A_{11}\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,C}(\boldsymbol{\lambda}) + A_{22}\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{R,R}(\boldsymbol{\lambda}) + 2A_{12}\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,R}(\boldsymbol{\lambda}) \\
&\quad + \frac{\partial^2}{\partial\lambda_k\lambda_l}P(\boldsymbol{\lambda},c) \quad (k,l=1,\ldots,n). \tag{4.33}
\end{aligned}
$$

For SK, we use the quadratic approximation in (4.7) to handle the non-differentiability of $|\lambda|$, which amounts to

$$
\frac{d}{d\lambda}|\lambda| = \frac{\lambda}{\lambda_0}, \qquad \frac{d^2}{d\lambda^2}|\lambda| = \frac{1}{\lambda_0}, \qquad \text{for } \lambda \approx \lambda_0 \neq 0.
$$

Then, the components of ∇Q are calculated as

$$
\begin{aligned}
\frac{\partial}{\partial\lambda_k}V^{C,C} &= \sum_i \lambda_i \left[C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_i) \right] - 2C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}^*), \\
\frac{\partial}{\partial\lambda_k}V^{R,R} &= \frac{\lambda_k}{|\lambda_{k0}|}\left[\sum_i |\lambda_i| \left[C^{R,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) \right] - 2C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}^*) \right] \\
&= \frac{\lambda_k}{|\lambda_{k0}|} f^{R,R}\left(\boldsymbol{\lambda}, \lambda_k\right), \\
\frac{\partial}{\partial\lambda_k}V^{C,R} &= \sum_i \left[|\lambda_i| C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) + \frac{\lambda_k}{|\lambda_{k0}|}\lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) \right] \\
&\quad - \frac{\lambda_k}{|\lambda_{k0}|}C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) - C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}^*), \\
\frac{\partial}{\partial\lambda_k}P(\boldsymbol{\lambda},c) &= -\frac{2c\lambda_k}{|\lambda_{k0}|}\left(1 - \sum_i |\lambda_i| \right).
\end{aligned}
$$

Similarly, the components of $\nabla^2 Q$ are

$$
\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,C} = C^{C,C}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_l),
$$

$$
\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{R,R} = \begin{cases} \frac{1}{|\lambda_{k0}|}f^{R,R}\left(\boldsymbol{\lambda}, \lambda_k\right) + 2C^{R,R}(\boldsymbol{0})\left(\frac{\lambda_k}{|\lambda_{k0}|}\right)^2 & k = l \\ \frac{\lambda_k\lambda_l}{|\lambda_{k0}\lambda_{l0}|}\left[C^{R,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_l) \right] & k \neq l \end{cases},
$$

$$\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,R} = \begin{cases} \frac{2\lambda_k}{|\lambda_{k0}|}C^{C,R}(\mathbf{0}) + \frac{1}{|\lambda_{k0}|}\left[\sum_i\lambda_iC^{C,R}(\boldsymbol{x}_i-\boldsymbol{x}_k) - C^{R,C}(\boldsymbol{x}_k-\boldsymbol{x}^*)\right] & k = l \\ \frac{\lambda_l}{|\lambda_{l0}|}C^{C,R}(\boldsymbol{x}_k-\boldsymbol{x}_l) + \frac{\lambda_k}{|\lambda_{k0}|}C^{C,R}(\boldsymbol{x}_l-\boldsymbol{x}_k) & k \neq l \end{cases},$$

$$\frac{\partial^2}{\partial\lambda_k\lambda_l}P(\boldsymbol{\lambda},c) = \begin{cases} 2c\left(\frac{\lambda_k}{|\lambda_{k0}|}\right)^2 - \frac{2c}{|\lambda_{k0}|}\left(1-\sum_i|\lambda_i|\right) & k = l \\ \frac{2c\lambda_k\lambda_l}{|\lambda_{k0}||\lambda_{l0}|} & k \neq l \end{cases}.$$

Plugging these derivations into (4.31 - 4.33) completes the SK case.

For OK, under the constraint that $\lambda_i \geq 0$, we have $|\lambda_i| = \lambda_i$ and $Q(\boldsymbol{\lambda})$ is a quadratic function, so no approximation is needed. The components of the gradient and Hessian are calculated straightforwardly as

$$\begin{aligned}
\frac{\partial}{\partial\lambda_k}V^{C,C} &= \sum_i\lambda_i\left[C^{C,C}(\boldsymbol{x}_i-\boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k-\boldsymbol{x}_i)\right] - 2C^{C,C}(\boldsymbol{x}_k-\boldsymbol{x}^*), \\
\frac{\partial}{\partial\lambda_k}V^{R,R} &= \sum_i\lambda_i\left[C^{R,R}(\boldsymbol{x}_i-\boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k-\boldsymbol{x}_i)\right] - 2C^{R,R}(\boldsymbol{x}_k-\boldsymbol{x}^*), \\
\frac{\partial}{\partial\lambda_k}V^{C,R} &= \sum_i\lambda_i\left[C^{C,R}(\boldsymbol{x}_k-\boldsymbol{x}_i) + C^{C,R}(\boldsymbol{x}_i-\boldsymbol{x}_k)\right] \\
&\quad -C^{R,C}(\boldsymbol{x}_k-\boldsymbol{x}^*) - C^{C,R}(\boldsymbol{x}_k-\boldsymbol{x}^*), \\
\frac{\partial}{\partial\lambda_k}P(\boldsymbol{\lambda},c) &= -\frac{2}{c}\left(1-\sum_i\lambda_i\right) - \frac{c}{\lambda_k}, \\
\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,C} &= C^{C,C}(\boldsymbol{x}_l-\boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k-\boldsymbol{x}_l), \\
\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{R,R} &= C^{R,R}(\boldsymbol{x}_l-\boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k-\boldsymbol{x}_l), \\
\frac{\partial^2}{\partial\lambda_k\lambda_l}V^{C,R} &= C^{C,R}(\boldsymbol{x}_k-\boldsymbol{x}_l) + C^{C,R}(\boldsymbol{x}_l-\boldsymbol{x}_k), \\
\frac{\partial^2}{\partial\lambda_k\lambda_l}P(\boldsymbol{\lambda},c) &= \frac{2}{c} + \frac{c}{\lambda_k^2}I_{(k=l)}.
\end{aligned}$$

Plugging these results into (4.31 - 4.33) completes the OK case. $\qquad\square$

4.8.4 Proof of Theorem 4

Proof. Recall the components of $V(\boldsymbol{\lambda})$ defined in (4.28 - 4.30). Because $\tilde{P}^{SK}(\boldsymbol{\lambda},c)$ (as defined

in (4.8)) now ensures that $\min|\boldsymbol{\lambda}| > 0$, we can now directly use derivatives of $|\lambda|$ expressed as

$$\frac{d}{d\lambda}|\lambda| = \text{sgn}(\lambda), \qquad \frac{d^2}{d\lambda^2}|\lambda| = 0.$$

This, in turn, simplifies the partial derivative calculations to become

$$
\begin{aligned}
\frac{\partial V^{C,C}(\boldsymbol{\lambda})}{\partial \lambda_k} &= \sum_i \lambda_i \left[C^{C,C}(\boldsymbol{x}_i - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_i) \right] - 2C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) \\
\frac{\partial V^{R,R}(\boldsymbol{\lambda})}{\partial \lambda_k} &= \text{sgn}(\lambda_k) f^{R,R}(\boldsymbol{\lambda}, \lambda_k)] \\
\frac{\partial V^{C,R}(\boldsymbol{\lambda})}{\partial \lambda_k} &= \sum_i \left[|\lambda_i| C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_i) + \text{sgn}(\lambda_k) \lambda_i C^{C,R}(\boldsymbol{x}_i - \boldsymbol{x}_k) \right] \\
&\quad - C^{R,C}(\boldsymbol{x}_k - \boldsymbol{x}^*) - C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}^*) \\
\frac{\partial \tilde{P}^{SK}(\boldsymbol{\lambda}, c)}{\partial \lambda_k} &= -\frac{2c}{n^2 \lambda_k} - \frac{2\,\text{sgn}\,\lambda_k}{c} \left(1 - \sum_i |\lambda_i| \right) \qquad k = 1, \cdots, n;
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 V^{C,C}(\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} &= C^{C,C}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{C,C}(\boldsymbol{x}_k - \boldsymbol{x}_l) \\
\frac{\partial^2 V^{R,R}(\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} &= \text{sgn}(\lambda_k \lambda_l) \left[C^{R,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) + C^{R,R}(\boldsymbol{x}_k - \boldsymbol{x}_l) \right] \\
\frac{\partial^2 V^{C,R}(\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} &= \text{sgn}(\lambda_l) C^{C,R}(\boldsymbol{x}_k - \boldsymbol{x}_l) + \text{sgn}(\lambda_k) C^{C,R}(\boldsymbol{x}_l - \boldsymbol{x}_k) \\
\frac{\partial^2 \tilde{P}^{SK}(\boldsymbol{\lambda}, c)}{\partial \lambda_k \partial \lambda_l} &= \frac{2\,\text{sgn}(\lambda_k \lambda_l)}{c} + I_{\{k=l\}}(\boldsymbol{\lambda}) \frac{2c}{(n\lambda_k)^2} \qquad k, l = 1, \cdots, n.
\end{aligned}
$$

Plugging these results into (4.31 - 4.33) completes this alternative SK case. $\qquad \square$

CHAPTER 5

ANALYZING INTERVAL-VALUED SPATIAL DATA

Now that our interval-valued kriging models have been formally defined and introduced, there remains the need to make interval-valued spatial data analysis accessible to the public. This is accomplished in the `intkrige` R software package through a series of functions designed to simplify the process for analyzing and exploring interval-valued spatial data. This process is demonstrated through interval-valued temperature predictions in the Ohio River basin.

5.1 Predicting interval-valued temperatures in the Ohio River Basin

Millions of Americans take interest in the forecasts of minimum and maximum temperature when deciding how to dress each day. DeGaetano and Belcher (2007) used various interpolation approaches to create minimum and maximum temperature maps of the northeastern United States. Their techniques treated minimum and maximum temperatures as separate observations. In a different framework, minimum and maximum temperatures could be considered as two elements of one and the same observation, which is daily temperature. In other words, these daily temperature measurements create a single *interval-valued* observation rather than two individual observations. This holistic view of temperature motivates the need for methods to analyze the interaction of temperature magnitude and range across space.

The Ohio river basin includes most of the states of Ohio, West Virginia, Kentucky, and Indiana, as well as parts of Pennsylvania, Illinois, and Tennessee. The region is approximately 204,000 square miles and home to nearly 25 million people (Ohio River Basin Consortium, 2019). The region was selected due its relatively homogeneous elevation profile when compared to the western United States. We wish to map the magnitude and range of temperatures across this region. This is done by creating intervals of the 30 year mean

maximum and minimum daily temperatures for available measurement locations.

Measurement locations are downloaded from the GHCN (Menne et al., 2018). The 161 remaining observations in the dataset all contained at least 30 years of record with at least 300 daily observation for calendar years 1988 to 2018. Intervals were created using a 5% trimmed 30 year mean for maximum and minimum temperatures respectively. These data are included in the package and can be accessed with the commands

```
library(intkrige)
#> Loading required package: sp
#> Loading required package: gstat
#> Registered S3 method overwritten by 'xts':
#>   method     from
#>   as.zoo.xts zoo
#> Loading required package: raster
data(ohtemp)
head(ohtemp)
#>           ID                 NAME STATE LONGITUDE LATITUDE ELEVATION
#> 1 USC00111302             CARMI 3    IL  -88.1805  38.0733     102.1
#> 2 USC00111436          CHARLESTON    IL  -88.1653  39.4761     213.4
#> 3 USC00112140            DANVILLE    IL  -87.6478  40.1392     169.2
#> 4 USC00112931 FAIRFIELD RADIO WFIW    IL  -88.3264  38.3806     131.1
#> 5 USC00114198           HOOPESTON    IL  -87.6850  40.4664     216.4
#> 6 USC00116446            OLNEY 2S    IL  -88.0817  38.7003     139.9
#>      minm     maxm
#> 1 76.11001 196.7273
#> 2 67.56999 180.6998
#> 3 57.90647 178.5897
#> 4 75.05457 197.1208
#> 5 57.97599 170.2541
```
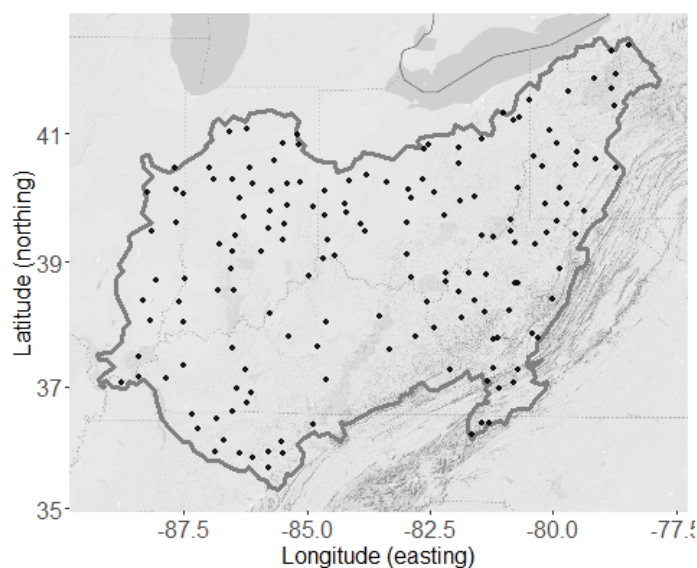
Fig. 5.1: Google map of the Ohio river basin with measurement locations overlaid.

```
#> 6 72.36740 191.8014
```

Figure 5.1 shows a map of this region with the 161 weather stations locations overlaid. The code for this map makes use of the `ggmap` package (Kahle and Wickham, 2013). The code to generate the map is not provided in this demonstration as the map rendering requires a registered Google API account.

5.2   Interval exploration

Our goal is to create an interval-valued map of the mean maximum and minimum temperatures across the region. Doing so allows us to simultaneously compare spatial patterns in the magnitude and range of average temperatures. The `intkrige` package provides a formal workflow for analyzing interval valued data by extending `SpatialPointsDataFrame` and `SpatialPixelsDataFrame` objects from the `sp` package (Bivand et al., 2013; Pebesma and Bivand, 2005). These objects are created using the `interval()` function as demonstrated below.

```r
# First, create a SpatialPointsDataFrame in the usual way

sp::coordinates(ohtemp) <- c("LONGITUDE", "LATITUDE")

sp::proj4string(ohtemp) <- CRS("+proj=longlat +ellps=WGS84")

interval(ohtemp) <- c("minm", "maxm")


head(ohtemp)
#>            coordinates                interval           ID
#> 1 (-88.1805, 38.0733) [76.11001, 196.7273] USC00111302
#> 2 (-88.1653, 39.4761) [67.56999, 180.6998] USC00111436
#> 3 (-87.6478, 40.1392) [57.90647, 178.5897] USC00112140
#> 4 (-88.3264, 38.3806) [75.05457, 197.1208] USC00112931
#> 5  (-87.685, 40.4664) [57.97599, 170.2541] USC00114198
#> 6 (-88.0817, 38.7003)  [72.3674, 191.8014] USC00116446
#>                   NAME STATE ELEVATION
#> 1               CARMI 3    IL     102.1
#> 2            CHARLESTON    IL     213.4
#> 3              DANVILLE    IL     169.2
#> 4 FAIRFIELD RADIO WFIW    IL     131.1
#> 5             HOOPESTON    IL     216.4
#> 6             OLNEY 2S    IL     139.9
```

This function creates either an `intsp` or `intgrd` object depending on the parent class. Both of these objects contain an interval slot that extend their parent classes. The interval slot is filled by specifying column names within the data frame, or specifying a two-column matrix with the same number of rows as the data object. These two forms of specification allow for convenient transformations of the interval slot values with the appropriate calls.

```r
interval(ohtemp) <- log(interval(ohtemp))

head(ohtemp)
```

```
#>          coordinates              interval          ID
#> 1 (-88.1805, 38.0733)  [4.33218, 5.281819] USC00111302
#> 2 (-88.1653, 39.4761) [4.213164, 5.196837] USC00111436
#> 3 (-87.6478, 40.1392) [4.058829, 5.185091] USC00112140
#> 4 (-88.3264, 38.3806) [4.318215, 5.283817] USC00112931
#> 5  (-87.685, 40.4664) [4.060029, 5.137292] USC00114198
#> 6 (-88.0817, 38.7003)  [4.281756, 5.25646] USC00116446
#>               NAME STATE ELEVATION
#> 1            CARMI 3    IL     102.1
#> 2         CHARLESTON    IL     213.4
#> 3           DANVILLE    IL     169.2
#> 4 FAIRFIELD RADIO WFIW    IL     131.1
#> 5          HOOPESTON    IL     216.4
#> 6            OLNEY 2S    IL     139.9
```

Note that intervals must be defined by their endpoints with the lower endpoint appearing in the first column. The `interval()` function will throw an error if it detects a lower endpoint greater than its corresponding upper endpoint. The function will also throw an error if a proposed transformation of the interval causes a lower endpoint to be greater than an upper endpoint. Recall from Chapter 4 that intervals can be alternatively characterized by their center and radius. It is easier to describe and characterize intervals by their endpoints, but it is easier to visualize and analyze *spatial* intervals using the center and radius. Thus interval objects must be initialized using their endpoints, but all interval-valued kriging and variogram functions automatically determine and use the center/radius form in analysis.

The interval-valued kriging functions require variograms for the center and radius, while a cross-variogram for the center/radius interaction is optional. The most common way of exploring these spatial relationships is through the use of an empirical variogram, defined previously in (2.3). The `intvariogram()` function provides wrappers to variogram functionality in the `gstat` package (Gräler et al., 2016; Pebesma, 2004). These wrapper
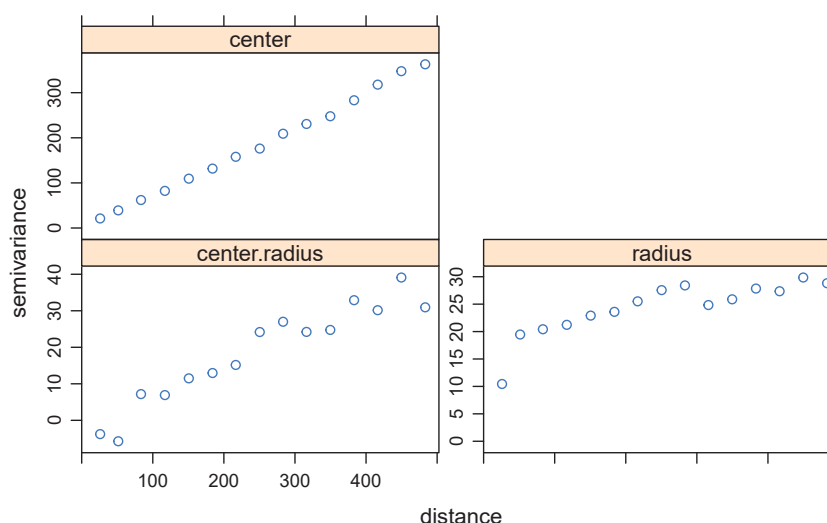
Fig. 5.2: Empricial variograms for the interval-valued data.

functions exploit the predictable structure of the interval slot in the `intsp` and `intgrd` objects to calculate the three empirical variograms simultaneously. The output of this function is an object of class `intvariogram`. This new object class distinguishes this output from other lists of empirical variograms.

```
# Revert back to the standard interval

interval(ohtemp) <- exp(interval(ohtemp))

varios <- intvariogram(ohtemp, cutoff = 500)



plot(varios)
```

It seems from Figure 5.2 that the radius and center/radius interaction could be appropriately modeled with spherical variograms while the center could be modeled with a linear variogram. The `fit.intvariogram` function provides a way to fit theoretical variograms to the three empirical variograms generated by the `intvariogram` function. The `fit.intvariogram` function uses the existing functionality of the `fit.variogram` function in the `gstat` package. The `fit.intvariogram` function only accepts objects of

class `intvariogram` and is designed to be used in tandem with the `intvariogram` function. The `intvcheck` function allows for a quick evaluation of the model output from `fit.intvariogram`.

```
varioFit <- fit.intvariogram(varios, models = gstat::vgm(c("Lin", "Sph", "Sph")))
#> Warning in gstat::fit.variogram(x_sub, model = models[[i]], ...): singular
#> model in variogram fit
varioFit
#> [[1]]
#>   model      psill      range
#> 1   Nug    1.003235    0.0000
#> 2   Lin 361.958967  497.7719
#>
#> [[2]]
#>   model    psill     range
#> 1   Nug   0.0000   0.00000
#> 2   Sph  23.1517  80.56012
#>
#> [[3]]
#>   model      psill    range
#> 1   Nug  -8.460132   0.000
#> 2   Sph  41.454114 470.397
intvCheck(varios, varioFit)
```

Note that for the observed data, the theoretical variogram models visualized in Figure 5.3 provide a decent fit to the data. However, the "singular model" warning reminds us that well-defined covariances cannot be recovered from linear variograms. The current interval-valued kriging models rely on the covariance equivalence of the variograms (which is calculated using `variogramLine(covariance = TRUE)`). A common practice in this case is to fit a spherical variogram with an arbitrary sill and range. These values are selected to
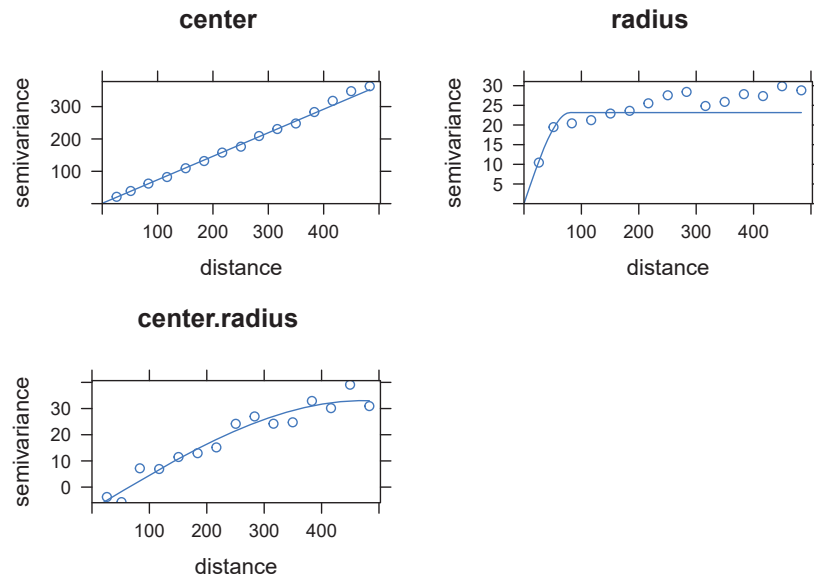
Fig. 5.3: Empirical variograms plotted against theoretical variograms for the interval-valued data.

create a spherical variogram that looks similar to the linear variogram within the viewing window as observed in Figure 5.4.

```r
# Replace non-convergent variogram fit with a surrogate that
# contains a reasonable range and sill.
varioFit[[1]] <- gstat::vgm(psill = 350, nugget = 4.608542,
                            range = 600, model = "Sph")


intvCheck(varios, varioFit)
```

We understand and acknowledge the shortcomings of this variogram fitting approach as articulated in Loquin and Dubois (2010). Despite its shortcomings, this common approach to variogram fitting works well for many spatial projects. Our substitution of a spherical variogram for a linear one certainly would demand more attention in a rigorous analysis of temperatures. However, for the purposes of demonstration, this model substitution is sufficient to demonstrate the interval-valued spatial data workflow.
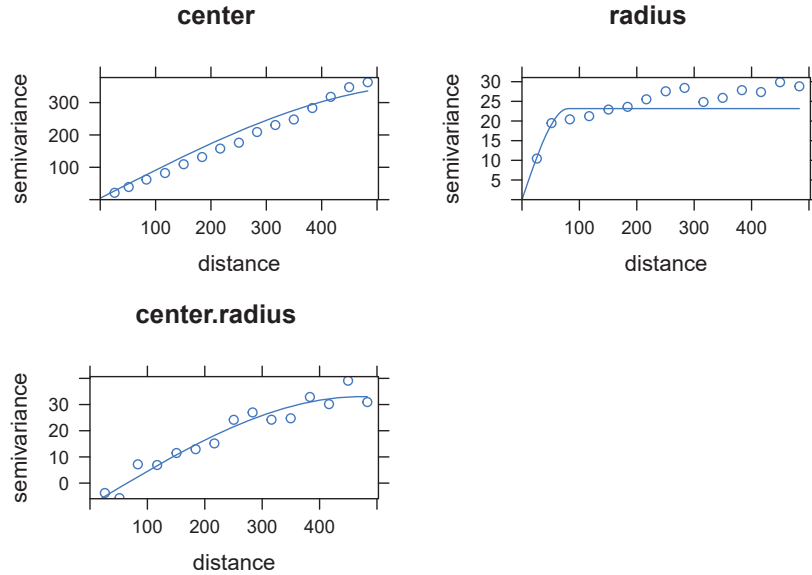
Fig. 5.4: Variogram plot where the linear variogram has been manually replaced with a spherical one.

5.3 Interval-valued kriging arguments

We are now ready to make temperature predictions using interval valued kriging. There are 17 different arguments for the `intkrige` function, yet all but three of these arguments have default options that allow for a quick implementation. These 17 arguments can be separated into three general categories: data, models, and optimization. A brief description of each group of variables is provided below:

- **Data:** the `locations` and `newdata` arguments define the measurement and prediction locations respectively. The location data object should also include a well-defined interval slot.

- **Models:** The `models` and `trend` arguments control the type and behavior of the interval valued kriging model. The `trend` argument corresponds to the known mean of the interval-valued centers. When specified, this argument leads the function to use simple instead of ordinary kriging. The `models` argument includes the list of variogram models used to calculate the kriging weights.

- **Optimization:** Recall that interval-valued kriging amounts the minimizing (4.3). The current numerical implementation assumes that the covariances exist and calculates them from the variograms using `variogramLine(covariance = TRUE)` from the `gstat` package. The $A_{11}, A_{12},$ and A_{22} parameters are specified with the `A` parameter in the `intkrige` function. The default option assumes no interaction between center and radius (i.e. `A[3] = 0`). `A[3]` is also effectively set to zero when a third variogram model is not specified.

  The `eta, thresh, tolq, maxq, tolp, maxp,` and `r` variables all directly relate to the PANR technique used minimize $E\left[\rho_K^2\left([\hat{Z}(x^*)], [Z(x^*)]\right)\right]$ as described in Chapter 4. The details of this optimization technique were provided in Chapter 4. These optimization parameters provide the user the necessary flexibility to adjust the optimization technique as needed to produce convergent results.

  These interval-valued kriging models are designed to predict hundreds, if not thousands, of locations in a single function call. Providing a separate initial guess for each prediction location quickly becomes impractical in such a context. As such, the initial guess for the kriging weights at each locations are derived from the traditional simple kriging weights obtained using only the interval centers. It is expected that the final weights will be close to this initial guess.

  Once the user has selected proper optimization parameters, the `fast`,`useR`, and `cores` variables all allow for potential speedups in the computation time. More about these arguments are found in the function documentation.

5.4   Troubleshooting Optimization Concerns

Each new dataset brings with it a new set of considerations for the optimization algorithm. The R version of the models (invoked when `useR = TRUE`, the default) will print warnings to the screen when the optimization fails. Both the R and c++ versions will flag troublesome predictions by setting `warn = 1` in the list of return arguments.

Determining the right optimization parameters to ensure a successful optimization for *all* locations can be tricky, especially when faced with a large number of prediction locations. It is recommended that the model first be run for a small set of prediction locations using the function defaults. Then, adjust the optimization parameters as needed based on the recommended guidelines in the list at the end of the section. Once the optimization is working correctly for a subset of locations, the model can be tested on all locations. Movement to the c++ version (via `useR = FALSE`) should only happen after the optimization parameters have been appropriately tuned, as the c++ version provides less descriptive messages to the user. Below is a list of common solutions for various optimization issues.

- In ordinary kriging (`trend = NULL`). The warning "left feasible region" is best handled by increasing the value of `r` and/or increasing the value of `eta` to be closer to 1.

- In simple kriging (`trend = !NULL`), the warning "convergent, feasible solution not obtained" is likely best handled by invoking the `fast = TRUE` option. The `fast = FALSE` is intended to prevent weights from converging to zero and is only relevant when the number of measurement locations is small (around 50 or less). In situations with more stations, the prevention of zero valued weights can cause a failure to converge. If this warning continues to be obtained. Consider increasing the value of `eta` closer to 1, or increasing the tolerance levels for the convergence criteria.

- The warning "feasible solution obtained from a non-convergent optimization step" is likely best handled by lowering the value of `eta`, which controls the rate of imposition of the penalty parameter.

5.5 Application

As previously mentioned, we can accept the default arguments for the majority of parameters in the `intkrige` model for our predictions of Ohio temperatures. Because no closed form solution exists for these interval-valued kriging models, the computational time for fine maps can be tremendous. This is one of the reasons for the parallel processing

option when the number of prediction locations is large. The parallel version of the model makes use of multiple packages included in the "Suggests" field of the package description file. Once these packages are installed, the argument `cores = -1` will use all available cores (minus 1) to make predictions at the specified locations. Alternatively, the user can explicitly define the number of cores to use in parallel with `n = <integer>`. We make predictions on a 10 by 10 grid of coordinates, defined using the bounding box of the Ohio river valley shapefile, using the following code.

```r
# Include the Ohio river basin shapefile
data(ohMap)


# New location data preparation
lon <- seq(-89.26637, -77.83937, length.out = 10)

lat <- seq(35.31332, 42.44983, length.out = 10)

newlocations <- expand.grid(lon, lat)

colnames(newlocations) <- c("lon", "lat")

sp::coordinates(newlocations) <- c("lon", "lat")

sp::proj4string(newlocations) <- sp::proj4string(ohtemp)

sp::gridded(newlocations) <- TRUE


# Adjust r and theta to ensure answers remain in feasible region.
preds <- intkrige(ohtemp, newlocations, varioFit,

                  A = c(1, 1, 0.5), r = 200, eta = 0.9, maxp = 225)


plot(preds, beside = FALSE, circleCol = "gray") +

  latticeExtra::layer(sp.lines(ohMap, col = "white"))
```
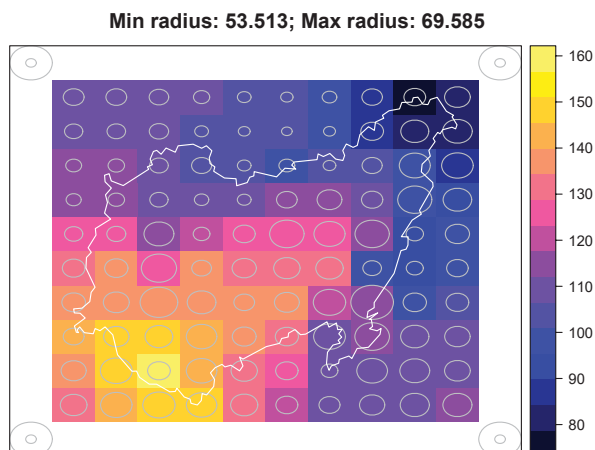
Fig. 5.5: Single plot method for interval-valued spatial grids. The reference circles in the corners of the image indicate the circle size for the maximum and minimum observed radius in the map. The range of radii are also given along the top of the plot.

```r
plot(preds, beside = TRUE)
```

5.6 Analysis

Figure 5.5 visualizes the center and radius of the interval-predictions in a single image. The reference circles in the corners of the image provide the circle sizes corresponding to the maximum and minimum radius. The numeric values of the maximum and minimum radius (rounded to three decimal places) are also given along the top of the plot. Figure 5.6 visualizes the components of the interval in a more traditional side-by-side plotting approach. These side-by-side plots make use of `gridArrange()` in the `gridExtra` package (Auguie, 2017). Both plotting approaches use the `lattice` package approach to plotting. This means that the user cannot make adjustments to the plotting window using `par()` from the `graphics` package. Both plotting styles are intended to give users a quick way to visualize the spatial patterns in predictions.

5.7 Reflections

These default plots help us to draw some quick conclusions from the data. First, notice
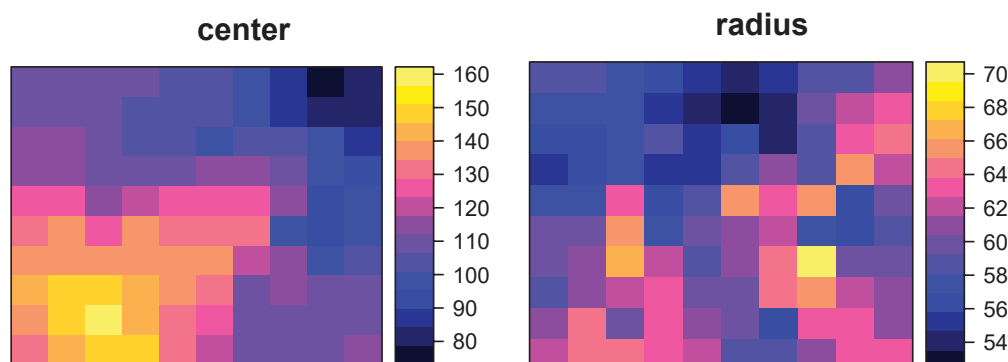
Fig. 5.6: Side-by-side plotting method for interval-valued spatial grids.

that the highest temperature magnitudes occur in the southern portion of the basin while the largest temperature *ranges* occur in the eastern portion of the basin. The eastern border of the basin runs through the Appalachian mountains and mountainous regions are notorious for experiencing larger daily swings in temperature. Conversely, warmer temperatures are more likely to persist throughout the night at lower elevations farther south. This likely explains why the southern corner of the basin has higher temperature magnitudes and smaller temperature ranges. It is also noteworthy that the smallest predicted interval radii occur along the northern border of the basin, possibly due to the moderating effect of lake Erie on extreme daily temperatures.

More important than the specific application provided in this chapter is the establishment of a pattern for exploring interval-valued spatial data. This demonstration has illustrated the use of a series of functions designed simplify and enhance interval-valued spatial analysis. We anticipate that future iterations of the `intkrige` package will result in a continual increase of functionality and applications.

CHAPTER 6

CONCLUSION

The ever-increasing availability and complexity of spatial data necessitate future adaptations, enhancements and replacements of traditional spatial methods. These methodological advancements require new paradigms to address old problems using a diverse set of approaches. This is especially true of design ground snow loads, which will become increasingly difficult to define as age old patterns in snow accumulation are disrupted by climate change. We have reviewed the various data processing and mapping techniques used by western states to predict ground snow and formally compared many of them in a cross validation analysis. This analysis revealed that regression-based mapping techniques performed noticeably better than the current mapping techniques used in Utah and Idaho. These results also demonstrated the shortcomings of using NGSL to predict snow loads between measurement locations. Our findings have culminated in a new set of design snow load requirements for Utah which have been adopted into Utah law (Article VI, Section 2 of the Utah Constitution). The new design snow load requirements can be obtained for any address in Utah at https://utahsnowload.usu.edu/.

Our review of the state-of-the-art methods for defining design snow load requirements revealed several important considerations discussed in Chapter 3. One of these important considerations is that the comparison of mapping techniques relied on the assumption that the design snow loads at each measurement location were known, precise values. This assumption ignores several sources of uncertainty inherent in the design snow load estimation process, which limits the opportunity to incorporate local measurement imprecision in a reliability analysis. The need for mapping techniques that can accommodate imprecise input gave rise to our interval-valued kriging models, which were developed and applied to the Utah design snow load prediction problem. The application highlighted apparent patterns in the relative precision of surrounding measurements: high elevation locations with direct

measurements of WESD had better relative precision than the sparse measurements of SNWD at low elevation locations. The process of creating an interval-valued design snow load dataset, as well as the ensuing interval-valued kriging predictions, are a crucial step forward in defining reliability-based design snow load maps.

These interval-valued kriging models gave rise to an interval-valued spatial data analysis workflow, which is formalized in the `intkrige` R software package. This workflow was demonstrated by making interval-valued predictions of temperature in the Ohio river basin, as well as interval-valued design snow load predictions in Utah. The `intkrige` package is available to the public via the Comprehensive R Archive Network (CRAN) or at https://github.com/beanb2/intkrige.

Perhaps equally important to these results are the momentum they have created for future projects. We are currently creating tools to fully automate the design snow load estimation workflow. A preliminary version of a design snow load dataset for the entire nation is available as part of the `snowload` package available at https://github.com/beanb2/snowload. This package also includes R implementations of the mapping methods described in Chapter 2. Other future projects include scaling the methods reviewed in this paper for nation-wide predictions of design snow load.

Our framework for data creation, model validation, and interval-valued spatial analysis provide a pattern for creating design snow load requirements that can be extended to a national scale. In particular, we provided a new approach for estimating design snow loads that accounts for imprecision at the measurement locations. This framework owes its existence to the many researchers actively engaged in the design snow load estimation problem. Inspiration was drawn from many of the state-level snow load reports and datasets made available to the public and cited throughout this dissertation. Our advancements and extensions will allow individual states, and perhaps the entire nation, to create data-driven design snow load requirements that can be quickly updated and evaluated with little marginal cost. Such updates will aid engineers across the U.S. as they design safe and economical buildings for future generations.

REFERENCES

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, 50(271):901–908.

Akima, H. (1978). A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Transactions on Mathematical Software (TOMS)*, 4(2):148–159.

Akima, H. and Gebhardt, A. (2015). *akima: Interpolation of Irregularly and Regularly Spaced Data*. R package version 0.5-12.

Al Hatailah, H., Godfrey, B. R., Nielsen, R. J., and Sack, R. L. (2015). Ground snow loads for Idaho–2015 edition. Technical report, University of Idaho, Department of Civil Engineering, Moscow, ID 83843.

Al Hatailah, H. A. (2015). Ground snow loads for the state of Idaho. Master's thesis, University of Idaho, Department of Civil Engineering, Moscow, Idaho.

Arcement, K. (2017). 'A lot of scared people': Relentless snow collapses hundreds of Idaho roofs, devastates rural county. Retrieved from washingtonpost.com on 05-15-2018.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Artstein, Z. and Vitale, R. (1975). A strong law of large numbers for random compact sets. *Annals of Probability*, 5:879–882.

ASCE (2017). *Minimum Design Loads and Associated Criteria for Buildings and Other Structures*. American Society of Civil Engineers, ASCE/SEI 7-16 edition.

Associated Press (2017). Roof on Provo building collapses under snow weight. Retrieved from ksl.com on 05-15-2018.

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

Aumann, R. (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.*, 12:1–12.

Bandemer, H. and Gebhardt, A. (2000). Bayesian fuzzy kriging. *Fuzzy sets and systems*, 112(3):405–418.

Bardossy, A., Bogardi, I., and Kelly, W. (1990a). Kriging with imprecise (fuzzy) variograms. i: application. *Mathematical Geology*, 22:81–94.

Bardossy, A., Bogardi, I., and Kelly, W. (1990b). Kriging with imprecise (fuzzy) variograms. i: theory. *Mathematical Geology*, 22:63–79.

Baxter, M. A., Graves, C. E., and Moore, J. T. (2005). A climatology of snow-to-liquid ratio for the contiguous United States. *Weather and Forecasting*, 20(5):729–744.

Bean, B., Maguire, M., and Sun, Y. (2017). Predicting Utah ground snow loads with PRISM. *Journal of Structural Engineering*, 143(9):04017126.

Bean, B., Maguire, M., and Sun, Y. (2018). The Utah snow load study. Technical report, Utah State University, Department of Civil and Environmental Engineering.

Bean, B., Maguire, M., and Sun, Y. (2019). Comparing design ground snow load prediction in Utah and Idaho. *Journal of Cold Regions Engineering*, 33(3):04019010.

Bivand, R., Keitt, T., and Rowlingson, B. (2018). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library.* R package version 1.3-4.

Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., and Pebesma, E. J. (2013). *Applied Spatial Data Analysis with R*, volume 747248717. Springer, 2 edition.

British Columbia (2018). Automated snow weather station data. Retrieved from gov.bc.ca on 10-11-2018.

Brothersen, B. and Fisher, J. (2018). Local contemporary costs. Personal correspondence with Vulcraft, Utah.

Business Wire (2016). Travelers identifies most common and costliest homeowners claims. Retrieved from businesswire.com on 05-15-2018.

CEC (1997). Ecological regions of North America: toward a common perspective. Technical report, Commission for Environmental Cooperation. Retrieved from cec.org on 10-9-2018.

Chang, M., Maguire, M., and Sun, Y. (2017). Framework for mitigating human bias in selection of explanatory variables for bridge deterioration modeling. *Journal of Infrastructure Systems*, 23(3):04017002.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610.

Daly, C. and Bryant, K. (2013). The PRISM climate and weather system – an introduction. Retrieved from prism.oregonstate.edu on 2018-05-15.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate research*, 22(2):99–113.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15):2031–2064.

DeBock, D. J., Liel, A. B., Harris, J. R., Ellingwood, B. R., and Torrents, J. M. (2017). Reliability-based design snow loads. i: Site-specific probability models for ground snow loads. *Journal of Structural Engineering*, page 04017046.

Debreu, G. (1967). Integration of correspondences. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability II*, pages 351–372.

DeGaetano, A. T. and Belcher, B. N. (2007). Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *Journal of Applied Meteorology and Climatology*, 46(11):1981–1992.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34.

Diamond, P. (1988). Interval-valued random functions and the kriging of intervals. *Mathematical Geology*, 20(3):145–165.

Diamond, P. (1989). Fuzzy kriging. *Fuzzy Sets and Systems*, 33:315–332.

Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain (2015). fields: Tools for spatial data. R package version 9.0.

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., and Vose, R. S. (2010). Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, 49(8):1615–1633.

Ekwaru, J. P. and Veugelers, P. J. (2018). The overlooked importance of constants added in log transformation of independent variables with zero values: A proposed approach for determining an optimal constant. *Statistics in Biopharmaceutical Research*, 10(1):26–29.

Ellis, S. (2017). Snow damage to Idaho-Oregon onion industry nears $100 million. Retrieved from capitalpress.com on 05-15-2018.

EPA (2018). Ecoregions of North America. Retrieved from epa.gov on 9-24-2018.

Fiacco, A. and McCormick, G. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, New York.

Fisicaro, K. (2017). Snow-load removal began day before gym collapse. Retrieved from lagrandeobserver.com on 05-15-2018.

Geis, J., Strobel, K., and Liel, A. (2011). Snow-induced building failures. *Journal of Performance of Constructed Facilities*, 26(4):377–388.

Gil, M., Lopez, M., Lubiano, M., and Montenegro, M. (2001). Regression and correlation analyses of a linear relation between random intervals. *Test*, 10(1):183–201.

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39.

Gillies, R. R., Wang, S.-Y., and Booth, M. R. (2012). Observational and synoptic analyses of the winter precipitation regime change over Utah. *Journal of Climate*, 25(13):4679–4698.

Glover, J. (2017). Wet, heavy snow in Spokane sparks concern that weak roofs could collapse. Retrieved from spokesman.com on 05-15-2018.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* Oxford University Press.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1):113–129.

Goupil, J. (2013). ASCE 7 and the standards development process. *Structure Magazine.* Retrieved from structuremag.org on 2019-06-07.

Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal*, 8:204–218.

Hijmans, R. J. (2016). *raster: Geographic Data Analysis and Modeling.* R package version 2.5-8.

Hohn, M. (1998). *Geostatistics and Petroleum Geology.* Springer Science & Business Media.

Hörmander, H. (1954). Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. *Arkiv för Mat*, 3:181–186.

Jensen, P. A. and Bard, J. F. (2003). *Operations Research Models and Methods [Supplementary Appendix]*, volume 1. John Wiley & Sons Incorporated.

Jin, Y., Ge, Y., Wang, J., Chen, Y., Heuvelink, G., and Atkinson, P. (2018). Downscaling amsr-2 soil moisture data with geographically weighted area-to-area regression kriging. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4):2362–2376.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate discrete distributions.* John Wiley & Sons, 2 edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994a). volume 1. John Wiley & Sons, 2 edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994b). volume 2, pages 207–227. John Wiley & Sons, 2 edition.

Julander, R. P. and Bricco, M. (2006). An examination of external influences imbedded in the historical snow data of Utah. In *74th Annual Western Snow Conference.* Accessed: 2018-05-15.

Kahle, D. and Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161.

Kato, D. and Florio, G. (2017). Montana theater's roof collapses under weight of snow. Retrieved from billingsgazette.com on 05-15-2018.

Körner, R. (1995). A variance of compact convex random sets. Technical report, Institut für Stochastik, Bernhard-von-Cotta-Str. 2 09599 Freiberg.

Körner, R. (1997). On the variance of fuzzy random variables. *Fuzzy Sets and Systems*, 92:83–93.

Körner, R. and Näther, W. (1998). Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Information Sciences*, 109:95–118.

Körner, R. and Näther, W. (2001). On the variance of random fuzzy variables. In Bertoluzza, C., Gil, M., and Ralescu, D., editors, *Statistical Model, Analysis and Management of Fuzzy Data*, pages 22–39. Physica, Heidelberg.

Lafferty, K. (2017). Snow buildup on porch roof causes collapse fatally injuring Deary woman. Retrieved from klewtv.com on 05-15-2018.

Lattin, J., Carroll, J. D., and Green, P. E. (2003). *Analyzing Multivariate Data.* Brooks/Cole, Cengage Learning.

Liel, A. B., DeBock, D. J., Harris, J. R., Ellingwood, B. R., and Torrents, J. M. (2017). Reliability-based design snow loads. ii: Reliability assessment and mapping procedures. *Journal of Structural Engineering*, 143(7):04017047.

Loquin, K. and Dubois, D. (2010). Kriging and epistemic uncertainty: a critical discussion. In *Methods for Handling Imperfect Spatial Information*, pages 269–305. Springer.

Loquin, K. and Dubois, D. (2012). A fuzzy interval analysis approach to kriging with ill-known variogram and data. *Soft Computing*, 16(5):769–784.

Lyashenko, N. (1982). Limit theorem for sums of independent compact random subsets of euclidean space. *Journal of Soviet Mathematics*, 20:2187–2196.

Maguire, M., Moen, C. D., Roberts-Wollmann, C., and Cousins, T. (2014). Field verification of simplified analysis procedures for segmental concrete bridges. *Journal of Structural Engineering*, 141(1):D4014007.

Mao, D., Qian, Z., Wei, S., and Hong, X. (2018). Constructing accurate radio environment maps with kriging interpolation in cognitive radio networks. *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, page 1.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

Matheron, G. (1971). *The Theory of Regionalized Variables and Its Applications.* École Nationale Supérieure des Mines.

Menne, M., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R., B.E.Gleason, and Houston, T. (2018). Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.24. Accessed: 9-24-2018.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910.

Mieure, E. (2017). Snow causes partial roof collapse at Sears, Axis, Hole Bowl. Retrieved from jhnewsandguide.com on 05-15-2018.

Molchanov, I. (2005). *Theory of Random Sets.* Springer-Verlag, London.

Moral, F. J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, 30(4):620–631.

Moselle, B. (2016). *National Building Cost Manual 41st Edition.* Craftsman Book Company, Carlsbad, CA.

Mote, P. W. (2006). Climate-driven variability and trends in mountain snowpack in western North America. *Journal of Climate*, 19(23):6209–6220.

NACSE (2012). An updated snow load map and internet map server for Oregon. Technical report, College of Engineering, Oregon State University, Corvallis, Oregon, Corvallis, OR.

Näther, W. (1997). Linear statistical inference for random fuzzy data. *Statistics*, 29(3):221–240.

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes.* R package version 1.1-2.

NOAA (2016). GHCN (global historical climatology network) – daily documentation. Retrieved from ncdc.noaa.gov on 2016-08-10.

NOAA (2017). Cooperative observer network (COOP). Retrieved from ncdc.noaa.gov 2017-09-07.

Nowak, A. S. and Collins, K. R. (2012). *Reliability of Structures.* CRC Press.

NRCS (2017). Active SNOTEL stations. Retrieved from https://www.wcc.nrcs.usda.gov 2018-08-20.

Ohio River Basin Consortium (2019). Retrieved from ohio.edu/orbcre on 5-28-2019.

O'Rourke, M. (2013). Snow related roof collapse and implications for building codes. *Structure Magazine.* Retrieved from structuremag.org on 7-13-2019.

PCIC (2018). BC station data - PCDS. Retrieved from data.pacificclimate.org 10-11-2018.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30:683–691.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rådström, H. (1952). An embedding theorem for spaces of convex sets. *Proc. Amer. Math. Soc.*, 3:165–169.

Rhoades, A. M., Huang, X., Ullrich, P. A., and Zarzycki, C. M. (2016). Characterizing Sierra Nevada snowpack using variable-resolution CESM. *Journal of Applied Meteorology and Climatology*, 55(1):173–196.

Rhoades, A. M., Ullrich, P. A., and Zarzycki, C. M. (2018). Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM. *Climate Dynamics*, 50(1):261–288.

Sack, R. L. (2015). Ground snow loads for the western United States: State of the art. *Journal of Structural Engineering*, 142(1):04015082.

Sack, R. L., Nielsen, R. J., and Godfrey, B. R. (2016). Evolving studies of ground snow loads for several western US states. *Journal of Structural Engineering*, page 04016187.

Sack, R. L. and Sheikh-Taheri, A. (1986). *Ground and Roof Snow Loads for Idaho*. University of Idaho, Department of Civil Engineering.

Scholz, F. (1995). Nonparametric tail extrapolation. Retrieved from faculty.washington.edu/fscholz on 2018-05-15.

Scott, D. and Kaiser, D. (2004). 5.2 variability and trends in United States snowfall over the last half century. In *15th Symposium on Global Change and Climate Variations*. American Meteorological Society.

SEAU (1992). Utah snow load study. Technical report, Structural Engineers Association of Utah, Salt Lake City, Utah. Provided in online format by Calder-Kankainen Consulting Engineers Inc. Salt Lake City, UT.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*, pages 517–524. ACM.

Shtiliyanova, A., David, B., Gianni, B., Pascal, C., Raphael, M., and Ulrich, E. (2017). Kriging-based approach to predict missing air temperature data. *Computers and Electronics in Agriculture*, 142:440 – 449.

Strobel, K. and Liel, A. (2013). Snow load damage to buildings: physical and economic impacts. *Proceedings of the Institution of Civil Engineers-Forensic Engineering*, 166(3):116–133.

Sturm, M., Holmgren, J., and Liston, G. E. (1995). A seasonal snow cover classification system for local to global applications. *Journal of Climate*, 8(5):1261–1283.

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., and Lea, J. (2010). Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, 11(6):1380–1394.

Theisen, G. P., Keller, M. J., Stephens, J. E., Videon, F. F., and Schilke, J. P. (2004). Snow loads for structural design in Montana. Technical report, Department of Civil Engineering, Montana State University, Bozeman, MT.

Tobiasson, W. and Greatorex, A. (1996). Database and methodology for conducting site specific snow load case studies for the United States. In *Proceedings of the Third International Conference on Snow Engineering*, pages 249–256.

Torrents, J. M., DeBock, D. J., Harris, J. R., Liel, A. B., and Patillo, R. M. (2016). Colorado design snow loads. Technical report, Structural Engineers Association of Colorado, Aurora, CO.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Company.

USGS (2019a). The national map. Retrieved from usgs.gov on 2019-06-07.

USGS (2019b). Watershed boundary dataset. Retrieved from usgs.gov on 2019-06-09.

Utah Legislature (2016). 15A-3-107 amendments to chapter 16 of IBC. Retrieved from le.utah.gov on 2018-05-15.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.

Wood, J. (2019). The year in charts: Utah's housing market 2018. Technical report, Kem C. Gardner Policy Institute, The University of Utah. Retrieved from gardner.utah.edu on 2019-06-07.

WRCC (2018). Climate of Washington. Retrieved from wrcc.dri.edu on 9-24-2018.

APPENDIX

AUTOMATING THE DESIGN GROUND SNOW LOAD ESTIMATION PROCESS

The following is an excerpt from the unpublished 2018 Washington Snow Load Report. The excerpt illustrates improvements and automations to the process for creating a design ground snow load dataset. This process has been subsequently used to create a national design snow load dataset as provided in the `snowload` package.

A.1 Washington Climate

The Western Regional Climate Center (WRCC) describes Washington as having a western coastal climate region and an eastern semi-arid climate region as separated by the Cascade Mountains (WRCC, 2018). Such a description is consistent with the Environmental Protection Agency's (EPA) designated ecoregions for the state. These ecoregions are "areas where ecosystems (and the type, quality, and quantity of environmental resources) are generally similar" (EPA, 2018). Further details are provided by the Commission for Environmental Cooperation (CEC, 1997). There are five different hierarchical levels of ecoregions defined for the United States. Figure A.1 visualizes the Level III ecoregions and similarly colored regions share the same Level I ecoregions. This base layer of this figure is obtained from Google's API using the ggmap package (Kahle and Wickham, 2013). These different ecoregions will differ in elevation, precipitation amount, and temperature and the type and amount of snow will inevitably differ greatly between regions. Details of how these ecoregions are used in the creation of the Washington design ground snow load dataset and corresponding PRISM predictions are provided in the ensuing sections.

A.2 Data Development

Data are collected from the Global Historical Climatology Network (GHCN) (Menne et al., 2018) as provided by the National Centers for Environmental Information (NCEI), a division of the National Oceanic and Atmospheric Administration (NOAA) (Menne et al.,
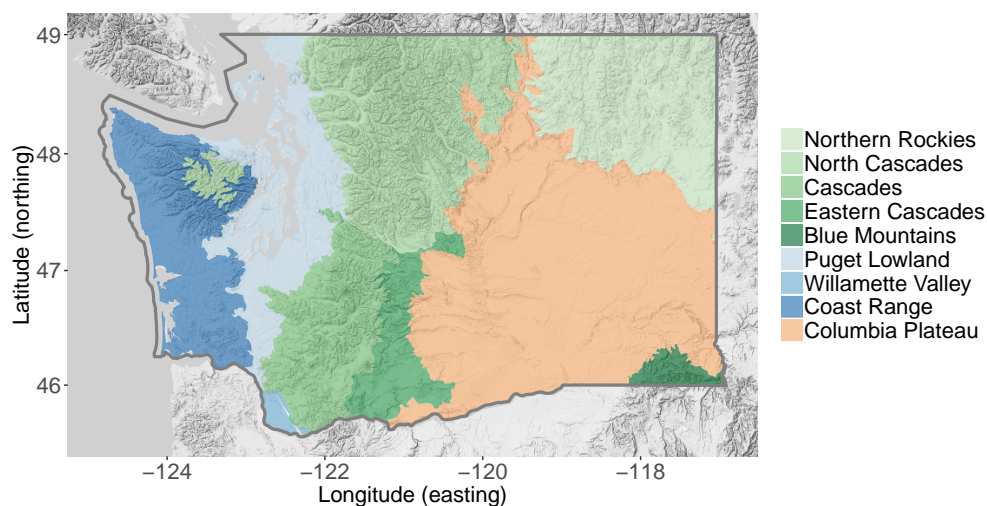
Fig. A.1: EPA level III ecoregions for the state of Washington.

2012). This network includes a variety of different stations, including the National Weather Service (NWS) first-order stations, the Natural Resources Conversation Service (NRCS) snowpack telemetry (SNOTEL) and snow course (SC) stations, and the NWS cooperative observer network (COOP) stations. The raw data include millions of snow depth (SNWD) and water equivalent of snow depth (WESD) observations at thousands of station locations in or within 60 miles of the Washington state border. These data are conveniently stored in ".dly" format by the NCEI as found at the following URL:

https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/.

These NOAA data are supplemented with observations taken from two British Columbia data sources, which aid predictions made along Washington's northern border. The Pacific Climate Impacts Consortium (PCIC) provides daily snow depth observations from several station networks in the province (PCIC, 2018). Additional SNOTEL data below the 50th latitude are provided by British Columbia's Snow Survey Network as operated by NOAA (British Columbia, 2018). Note that these data are not subject to the same set internal quality assurance measures provided with the NCEI ".dly" files. However, they are subject to
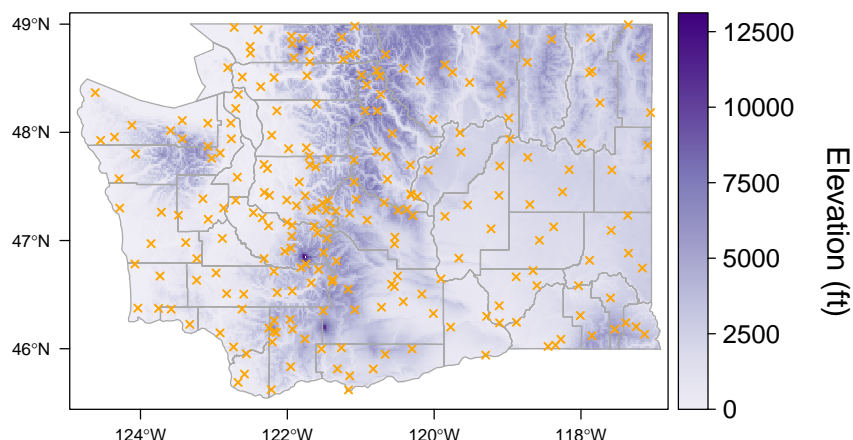
Fig. A.2: Snow measurement locations overlaid on an elevation map of Washington.

the authors' manual outlier screening described later in this section. All data manipulations and many of the ensuing visualizations make use of R's tidyverse (Wickham, 2017) as well as the reshape2 package (Wickham, 2007).

The final dataset contains 50 year ground snow load estimates at 234 measurement locations in Washington and 300 locations in surrounding states and provinces withing 60 miles of the Washington border. These data are provided in Tables 4 and 5 at the end of this report. Figure A.2 overlays the Washington measurement locations on an elevation map of the state. Elevations are obtained from digital elevation models (DEMs) obtained from the United States Geological Survey's (USGS) National Map (USGS, 2019a). Further details regarding these DEMs are provided in the Final Predictions section. Figure A.2 reveals that the highest concentration of measurement locations appear to be along the western slopes of the Cascade range.

The design snow loads in this dataset are calculated by fitting the annual maximum snow load measurements to five different probability distributions and retaining the median 98th percentile. Maximums are separated by water year rather than calendar year, so that each water year includes all observations from October 1 of the year previous to May 31 of the designated year. Data are only considered for water years beginning in 1967 in order to emphasize observations from recent history at newer stations with more reliable measurement

technologies. The ensuing subsections describe

- the clustering of stations with nearly identical locations/elevations into a single measurement location,

- the process of estimating snow load from SNWD when necessary,

- the quality assurance measures taken on daily observations and annual maximums,

- the distribution fitting process and the resulting 50 year loads.

A.3   Snow Site Clustering

There are occasional instances where two or more weather stations share a nearly identical location and elevation. This may occur when a new station is installed to replace an older one, or when measurements at a particular location are of interest to multiple independently-operated station networks. In many instances, combining station information extends the period of record for a particular location and prevents large differences in design snow load estimates over very short distances. The advantages of station combination must be balanced against the requirement that the annual maximums at a particular location follow the same theoretical probability distribution. Combining station measurements with distinct snow patterns could invalidate the distribution fitting assumptions and obscure key regional differences in snow patterns.

The Colorado Snow load report provides an example for clustering similar snow stations into consolidated measurement locations. This report combines stations that are within 2-12 miles and 300-500 feet elevation of each other, with the stricter clustering criteria being imposed on stations in mountainous locations (Torrents et al., 2016). This report uses the complete linkage agglomerative clustering technique (Lattin et al., 2003, pp. 264-288) as implemented in R 3.5.1 (R Core Team, 2018). The clustering uses a custom distance matrix formed from a combination of scaled geographic distances and scaled elevation differences. Each 2.5 miles of geographic separation contributes a distance equal to one, while each 165 feet of elevation separation likewise contributes a distance equal to one. The complete linkage cluster technique requires that the farthest neighbors in an individual cluster have
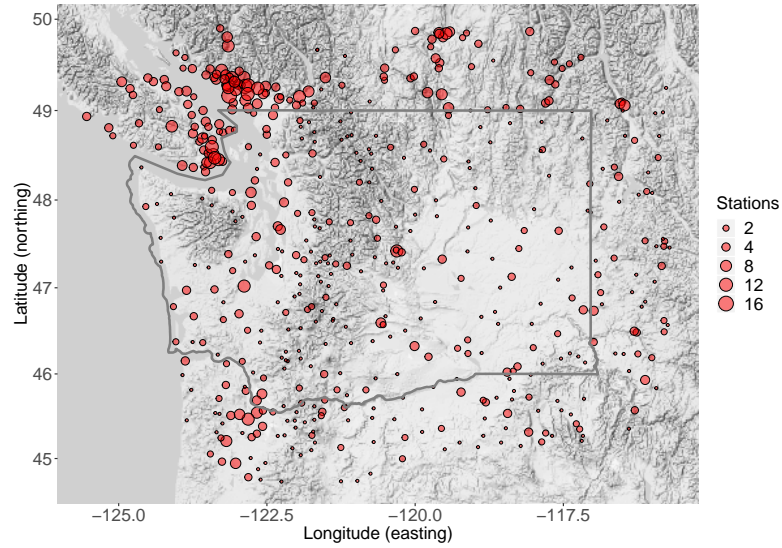
Fig. A.3: The number of stations comprising each measurement location for the Washington snow load dataset.

a distance of two or less. This means that the maximum possible geographic separation between stations in a cluster is five miles and the maximum possible elevation separation is 330 feet. The larger the geographic separation, the less elevation difference is tolerated and vice versa. Also, stations located in different level III ecoregions are prevented from clustering together regardless of their geographic or elevation separation. Figure A.3 visualizes these measurement locations and their respective cluster sizes. Notice the high number of stations included in or near Vancouver and Victoria, British Columbia. This concentration suggests a larger number of snow reporting stations in the British Columbia networks with sufficient coverage of the snow season than is found in the NOAA network. Measurement locations are not weighted by cluster size and the sizes of the points in Figure A.3 only represent the number of stations comprising the cluster, not the weight of the observation. Indeed, this clustering technique prevents the large number of coastal Canadian stations from being over-represented in the final dataset.

As in the Colorado report, when stations in a single cluster have overlapping periods of record, preference is given to direct measurements of WESD. If the overlapping measurements are of the same type, then only the maximum measurement is retained.

Table A.1: Climate specific parameters for Sturm's equation.

Class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	0.2578	0.0010	0.0038
Prairie	0.5940	0.2332	0.016	0.0031
Tundra	0.3630	0.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000

A.4 Estimating Snow Load From Snow Depth

The depth-to-load conversions in the most recent Washington Snow load report make use of the Rocky Mountain Conversion density (RMCD). This conversion equation models snow load as a function of snow depth and was designed in conjunction with the Idaho snow load report (Sack and Sheikh-Taheri, 1986). However, the density of snow will inevitably vary greatly between the coastal and inland regions and the RMCD provides no mechanism to account for these climatic differences. An alternative to the RMCD is a conversion equation developed by Sturm et al. (2010), referred to hereafter as "Sturm's equation", which estimates snow load as a function of depth, time, and climate class. This equation adapted for English units and defined as

$$p_g(h, d) = ((\rho_{max} - \rho_0) \left[1 - \exp\left(-k_1 * h * 2.54 - k_2 * d\right) \right] + \rho_0) * 2.048176 \qquad (A.1)$$

where h represents snow depth (in inches) and d represents day of the snow season starting on October 1st (-92) and ending June 30th (181) with no zero value. Additionally, $\rho_o, \rho_{max}, k_1,$ and k_2 are parameters specific to a particular climate class defined in Table 4 of Sturm et al. (2010) and provided for convenience in Table A.1 of this report. Locations residing in an Ephemeral snow region use the average parameter values for the Alpine, Maritime, and Prairie regions.

The climate classes are originally defined in Sturm et al. (1995) and this reference includes a map of the climate classes of North America at a 50 km resolution. This climate class map shows many similarities with the EPA's level III ecoregions. Consequently, the authors mapped each ecoregion to a particular climate class through a visual comparison
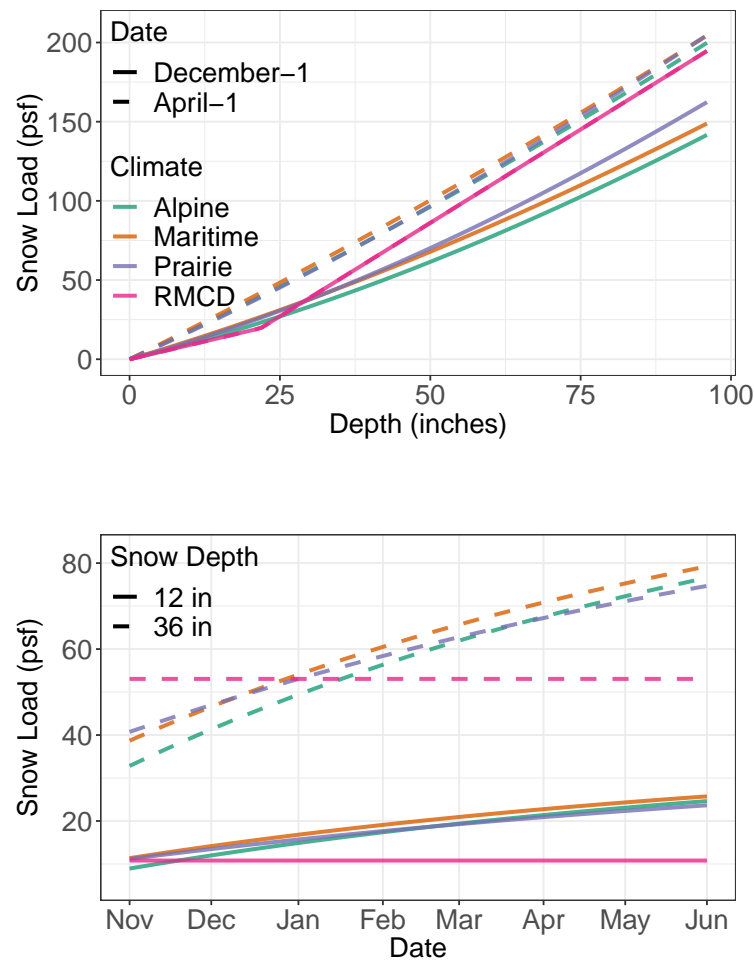
Fig. A.4: Comparison of the various depth-to-load conversion methods used in the Washington snow load report to the Rocky Mountain Conversion Density (RMCD).

of the climate class and ecoregion maps. In cases where the visual assignment was not immediately obvious, ecoregions were assigned to the climate class with characteristics most similar to the particular ecoregion according to the authors' best judgment.

Figure A.4 compares the depth-to-load conversions using the three major climate classes indicative of locations in Washington as compared to the RMCD predictions. This figure shows a tendency for Sturm's equation to estimate higher snow loads than the RMCD as the snow season progresses.

A.5   Quality Assurance

The proper detection of impossibly high snow load measurements is critical given that only the highest load value from each water year is retained. These outlier values have the effect of extending the tail of the selected probability distributions, producing artificially high design snow loads. The NCEI provides a suite of quality assurance checks for observations in the GHCN (Durre et al., 2010). All observations failing any of these checks were removed at the time of download. An additional check for outlier observations is made after collecting the annual maximum snow load measurements for each measurement location. Because the annual maximum snow loads at a particular location are typically right skewed, the log transformation of annual maximum snow loads should be roughly symmetric in distribution and suitable for traditional outlier checks. One common method of detecting outliers uses the interquartile range (IQR) to determine an outlier value threshold. The IQR is defined as the difference between the 25th (q_{25}) and 75th (q_{75}) percentiles of the log-transformed annual maximums for a given location (i.e, IQR $= q_{75} - q_{25}$). One very common outlier threshold is defined as $q_{75} + 1.5 * (\text{IQR})$ or $q_{75} + 3 * (\text{IQR})$ for "far out" observations (Tukey, 1977, pp. 43-44). Since the cost of removing a true value is much higher than failing to remove a misreported value, this report uses the $q_{75} + 3 * (\text{IQR})$ rule to screen outliers. This threshold is calculated for each individual measurement location and all daily observations exceeding the threshold are removed from the data prior to final analysis. One exception to this rule is when there are five or more daily observations exceeding the threshold at a measurement for a given water year. This exception guards against situations where the snow load legitimately exceeds the outlier threshold as corroborated by multiple measurements. Using this threshold, 19 observations were removed from relevant measurement locations in Washington, 17 in British Columbia, 15 in Oregon, and 1 in Idaho.

Figure A.5 visualizes the outlier detection screening at four locations in Washington. All values above the red line are removed prior to the final analysis. These plots show that there are other probable outliers that are not caught by this screening method. However, this screening method seems to strike an appropriate balance between catching the most
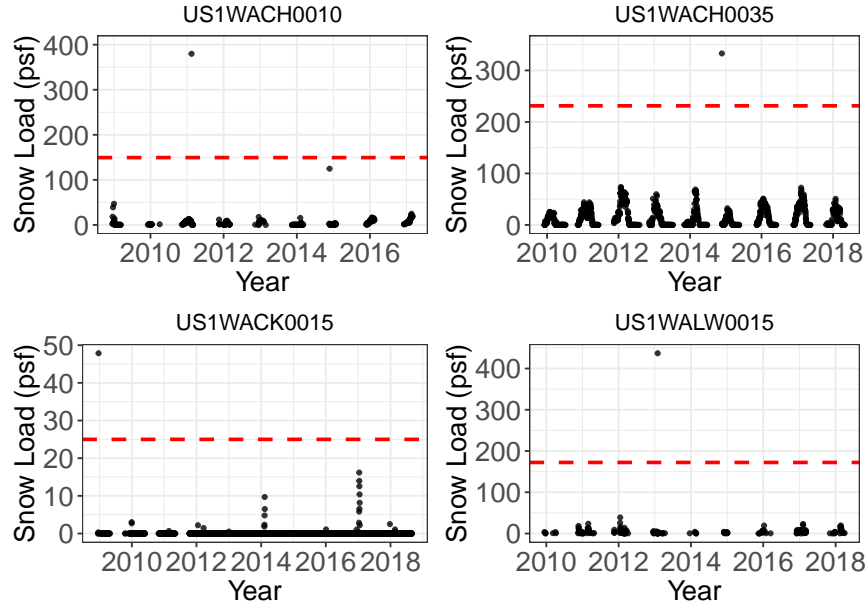
Fig. A.5: Sample plot of four different outlier screens for measurement locations included in the Washington design snow load data set.

extreme outliers, without compromising the integrity of the input data.

Artificially low snow load maximums are likewise problematic when estimating design snow loads. These values have the effect of inflating the estimated standard deviation of the fitted distribution and consequently produce artificially high design snow loads. These abnormally low snow loads are usually the product of a lack of daily measurements at a station for a given water year. To guard against the inclusion of such values, a yearly maximum is retained only if the maximum came from a water year with at least 30 observations spanning over at least 5 of the 8 months of the relevant snow season (October - May). This coverage filter is ignored if the resulting maximum is in the upper half of all recorded maximums for a given station. To further protect against artificially low maximums, all maximums at or below the 10th percentile are removed prior to analysis. This removal only occurs if more than 75 percent of the annual maximum snow loads are non-zero.

A.6 Distribution Estimates

ASCE 7-16 defines 50 year ground snow loads as the design ground snow load (ASCE,

2017). A 50 year ground snow load event is the ground snow load for which there is a 2% chance of exceeding on any given year. One common practice for defining these design loads is to fit a probability distribution to the annual maximum snow loads and extract the 98th percentile. ASCE 7-16 uses the log-normal distribution to define 50 year loads (ASCE, 2017). Colorado also uses the log-normal distribution but only fits the distribution to the upper third of the annual maximums (DeBock et al., 2017). Alternatively, Idaho uses the log-Pearson type III distribution (Al Hatailah et al., 2015).

This process assumes that the annual maximum snow loads at any given location are time-independent observations from the same probability distribution. The approach is problematic if the selected distribution cannot appropriately characterize the observations, or if there is a notable time-dependent trend in the annual maximums. For example, some locations have a much higher proportion of low-snow years than would be expected under the log-normal distribution. Fitting a log-normal distribution at such locations results in artificially high estimates of the standard deviation which leads to an unreasonably high estimate of the design snow load. In some cases, the design snow load estimates from the log-normal distribution at select sites in Washington are two to three times higher than the largest ever observed snow load at the location, which is indicative of a distribution that is no longer representative of the data. In addition, many of these probability distributions are not equipped to deal with measurement locations that have no-snow years. No snow years were simply ignored in the Utah report (Bean et al., 2018) but cannot be ignored given the high number of coastal measurement locations in Washington that record no snow during a given year.

In order to protect against unreasonably high estimates of design snow loads resulting from unstable distribution fits, this report fits five different distributions to the annual maximums at each measurement location using maximum likelihood (ML) estimation. These distributions include the log-normal (LN), gamma (GM), and extreme value distribution types I, II, and III. Comprehensive summaries of these distributions can be found in Johnson et al.'s two-volume Continuous Univariate Distributions (Johnson et al., 1994a, pp. 207-258;

Table A.2: Number of times each distribution was used to determine the design load for the 522 relevant locations in the Washington snow load dataset.

Distribution	n
log-normal	100
gamma	332
Type I	89
Type II	1
Type II	0

337-414; Johnson et al., 1994b, pp. 1-112). In theory, the maximum values of a random variable should converge to one of the three extreme value distributions (Johnson et al., 1994b). The extreme value type II distribution typically forms an upper bound of the design load and the predicted value is often too high to be feasible for use in design. Conversely, the extreme value type III distribution assumes observations are bounded above and typically forms a lower bound for prediction. The design load is selected as the median value from the five selected distributions. Figure A.6 shows examples of the distribution fitting process at four locations in the state of Washington. The gamma distribution provides the median value for the three cities while the log-normal distribution provides the median value for the Stevens Pass ski area. Table A.2 shows the number of times each distribution was used for the design ground snow load for 522 relevant measurement locations in the Washington snow load dataset.

Note that this strategy does not use any type of goodness-of-fit measurement to determine the selected distribution. The notion of "fit" is often vaguely defined and measures of fit do not always guarantee that the appropriate distribution will be selected, as discussed in the Montana snow load report (Theisen et al., 2004). Rather, this strategy guards against situations where the ML fit for any one of the distributions are unstable.

Measurement locations with multiple years of no recorded snow are accommodated by use of a "zero-contaminated" distribution (Johnson et al., 2005), pp. 312-313) originally introduced in Aitchison (1955). Let θ represent the proportion of yearly maximums X equal to 0. Assume that $X_s = \{x \in X : x > 0\} \sim F_{X_s}(x)$ where $F_{X_s}(x)$ represents the cumulative distribution function fit via ML from one of the five candidate distributions. If $\theta = 0$, the
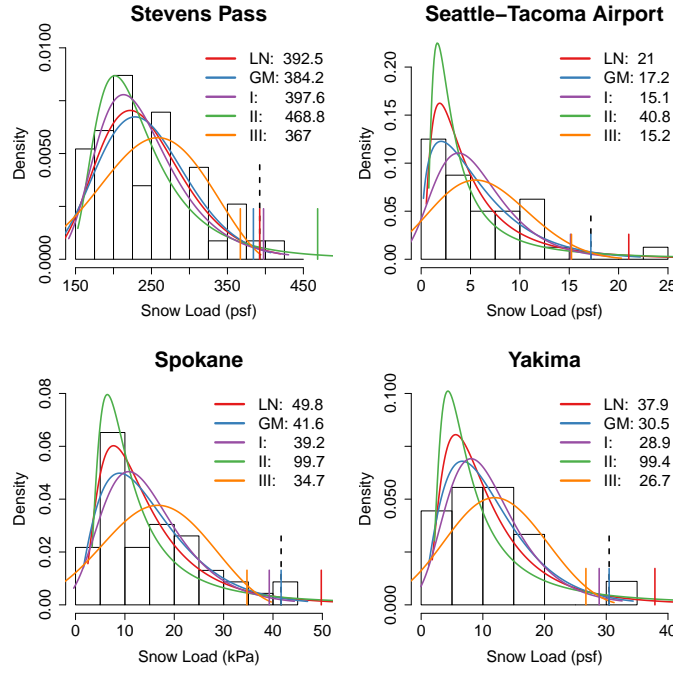
Fig. A.6: The five selected probability distributions overlaid on histograms of annual yearly maximum snow load measurements.

design snow load is defined as the value of x for which $F_{X_s}(x) = 0.98$. However, in situations where $\theta > 0$, the design snow load is defined as the value of x for which

$$F_{X_s}(x) = \frac{0.98 - \theta}{1 - \theta}.$$

For example, if a measurement location receives no snow for half of its period of record ($\theta = 0.5$) then the design snow load will be the value of x for which $F_{X_s}(x) = \frac{0.98-0.5}{1-0.5} = 0.96$.

Measurement locations are required to have at least 12 viable yearly maximums (prior to filtering the bottom 10 percent of observations) in order to be considered for the distribution fitting process. The 12 locations with less than five non-zero yearly maximums simply use the maximum max for the design snow load as there would not be enough observations to justify the distribution fitting process. This distribution strategy is far from perfect as there will always be instances where observed data do not meet the theoretical assumptions required for distribution fitting via ML. Despite its imperfections, this process provides reasonable estimates of design snow loads as informed by daily data.

CURRICULUM VITAE

Brennan Bean
brennan.bean.20@gmail.com

Education

Doctor of Philosophy in Mathematical Sciences *July 2019*
Utah State University Logan, Utah

- GPA: 4.0

Bachelor of Science in Applied Mathematics *December 2014*
Brigham Young University - Idaho Rexburg, Idaho

- Graduated *summa cum laude*, invited speaker at convocation.

Publications

Journal Articles

- **Bean, B.**, Maguire, M., & Sun, Y. (2019). "Comparing ground snow load prediction methods in Utah and Idaho," *Journal of Cold Regions Engineering*, *33*(3), 04019010. https://doi.org/10.1061/(ASCE)CR.1943-5495.0000190

- **Bean, B.**, Maguire, M., & Sun, Y. (2017). "Predicting Utah ground snow loads with PRISM," *Journal of Structural Engineering*, *143*(9), 04017126. https://doi.org/10.1061/(ASCE)ST.1943-541X.0001870

Proceedings and Reports

- Lundell, J., **Bean, B.**, & Symanzik, J. (2018). "Let's talk about the weather," In *JSM Proceedings*, Statistical Computing Section. Alexandra, VA: American Statistical Association. 1944-1958.

- Symanzik, J., McKinney, E., Studenka, B., **Bean, B.**, Athens, M., Hansen, M. (2018). "Eye-Tracking in Practice: Results from a Study on Human Postures," In *JSM Proceedings*, Statistical Computing Section. Alexandra, VA: American Statistical Association. 2696-2706.

- **Bean, B.**, Maguire, M., & Sun, Y. (2018). "The Utah snow load study," *Civil and Environmental Engineering Faculty Publications.* Paper 3589. https://digitalcommons.usu.edu/cee_facpub/3589

- **Bean, B.**, Maguire, M., & Sun, Y. (2017). "Comparing ground snow load prediction methods in Utah," In *the proceedings of Congress on Technical Advancement*, Duluth, MN, September 10-13, 2017, pp. 207-217. https://doi.org/10.1061/9780784481011.020

Other Articles

- **Bean, B.**, Maguire, M., & Sun, Y. (2019) "Reflections on the Utah snow load study," *Utah Engineers Council Journal*

- **Bean, B.** (2019) "USU Data Analytics Club reviews code submissions with ASA president-elect," *Amstat News* Alexandra, VA: American Statistical Association.

Working Papers

- **Bean, B.**, Sun, Y., & Maguire, M., "Interval-valued kriging models for geostatistical mapping with uncertain inputs." *Annals of Applied Statistics* (under review)

- Lundell, J., **Bean, B.**, & Symanzik, J., "Let's talk about the weather: A cluster-based approach to weather forecast accuracy" *Computational Statistics* (under review)

Presentations

- (November 27, 2018). "Lessons learned from snow: eight rules to live by for every data analytics project," Data Analytics Club Student Seminar Series, Utah State University Department of Mathematics and Statistics, Logan, Utah.

- (August 22, 2018). "Embracing vulnerability in teaching," Together We Teach Conference, Utah State University Department of Mathematics and Statistics, Logan, Utah.

- (July 30, 2018). "Interval-valued kriging and application in climate related predictions," Joint Statistics Meetings, Section on Statistics and the Environment, Vancouver, British Columbia.

- (April 4, 2018). "Review of the Utah snow load study (condensed)," Monthly Meeting, Utah Association of Building Officials, Farmington, Utah.

- (March 27, 2018). "Visualizing spatio-temporal Variations in the Utah snowpack," Spring Runoff Conference, Utah State University, Logan, Utah.

- (February 21, 2018). "Review of the Utah snow load study," Annual Education Conference, Structural Engineers Association of Utah, Layton, Utah.

- (September 11, 2017). "Comparing ground snow load prediction methods in Utah," Congress on Technical Advancement, American Society of Civil Engineers, Duluth, Minnesota.

Funded Research Projects

The Washington Snow Load Study *August 2018-December 2018*
Structural Engineers Association of Washington

- Generated county-scale design snow load maps for the state of Washington.

The Utah Snow Load Study   *May 2016-August 2017*
Structural Engineers Association of Utah &
National Council of Structural Engineering Associations

- Helped to develop the freely available ground snow load website: utahsnowload.usu.edu.

  - Provides estimates of the design ground snow load for any user-defined location in Utah.

  - Website estimates have been adopted into Utah state law (Article VI, Section 2 of Utah Constitution).

Professional Experience

Statistics Fellow   *May 2017-August 2017*
SAS Institute Inc.   Cary, North Carolina

- Independently researched the Pearson and Johnson Systems of distributions, as well as the theory of Pharmacokinetics, in order to create test programs for new features in PROC NLMIXED, PROC MCMC, and the DATA step.

- Developed test tables to automatically run and benchmark test program results on a Linux server.

DHS-STEM Summer Intern   *May 2015-July 2015*
Engineer Research and Development Center -
U.S. Army Corps of Engineers   Vicksburg, Mississippi

- Generated continental scale river networks with ARCGIS for use in global streamflow modeling and flood forecasting.

- Informed a variety of research projects through data visualizations conducted in R and MATLAB.

Awards and Honors

Graduate PhD Researcher of the Year   *April 12, 2018*
USU - Department of Mathematics and Statistics   Logan, Utah
Second Place - 2018 Data Expo   *August 30, 2018*
JSM - Statistical Computing and Graphics Sections   Vancouver, British Columbia
Data Analytics Leadership   *April 12, 2018; March 29, 2019*
USU - Department of Mathematics and Statistics   Logan, Utah
First Place - PhD Student Poster   *March 27, 2018*
USU - Spring Runoff Conference   Logan, Utah
Excellence in Teaching   *April 13, 2017*
USU - Department of Mathematics and Statistics   Logan, Utah
Academic Excellence   *April 15, 2016*
USU - Department of Mathematics and Statistics   Logan, Utah

Teaching

Instructor	STAT 1040 - Intro to Statistics	*Spring 2019*
Instructor	MATH 1210 - Calculus I	*Fall 2016*
Teaching Assistant	STAT 2300 - Business Statistics	*Spring 2016; Spring 2017*
Teaching Assistant	MATH 1210 - Calculus I	*Fall 2015*

Leadership and Service

President	*January 2018 - May 2019*
Utah State University Data Analytics Club	Logan, UT

- Coordinated a series of student-led presentations to demonstrate various analytical softwares and methods.

Professional Memberships

American Statistical Association	since June 2017
USU Data Analytics Club	since October 2017
Institute of Mathematical Statistics	since December 2018

Programming Experience

Proficient	**Working Knowledge**	**Exposure**
R	Python, C++, SAS, MATLAB	SQL, HTML