The 2020 National Snow Load Study

Brennan Bean,[1*] Marc Maguire (A.M.ASCE),[2] Yan Sun,[1] Jadon Wagstaff,[1] Salam Al-Rubaye,[2] Jesse Wheeler,[1] Scout Jarman,[1] and Miranda Rogers[1]

[1]*Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, Utah, 84322*
[2]*Durham School of Architectural Engineering and Construction, University of Nebraska - Lincoln, 1110 S. 67th Street, Omaha, Nebraska 68182*

*Corresponding Author: brennan.bean@usu.edu

Contents

Foreword

While great efforts have been made to ensure that the reliability-targeted design ground snow load predictions resulting from this research are as accurate as possible, the authors cannot accept responsibility for prediction errors or any consequences resulting therefrom. Responsibility for the final design snow loads rests with the builder or designer in charge of the project.

Acknowledgments

This research is made possible through funding from the Structural Engineering Institute of the American Society of Civil Engineers in collaboration with several private engineering firms. The groups that provided significant monetary support to this effort were (in alphabetical order): Factory Mutual, Metal Building Manufacturer's Association, National Council of Structural Engineering Associations, Nucor, Simpson Gumpertz and Heger, the State of Montana, the Steel Deck Institute, the Steel Joist Institute, Structural Engineers Association of Montana, Wiss Janney and Elstner Associates.

Oversight of the work was provided by a steering committee headed by Abbie Liel (University of Colorado - Boulder) and Scott Russell (Nucor Steel) with members

- Mike O'Rourke (RPI)
- Jim Harris (JR Harris)
- Jim Buska (CRREL)
- Jerry Stephens (U of MT)
- R. Nielson (U of ID)
- D. Jared DeBock (Chico State)
- Johnn Judd (U of WY)
- David Thompson (KTA)
- Hossein Mostafaei, (FM)
- John Corless (SEAOC)
- John-Paul Cardin (AISI)
- Sean Homem (SGH)
- Gary Ehlrich (NAHB)
- Sterling Strait (SEAAK)
- Vince Sagan (MBMA)
- Thomas DiBlasi (SEA)
- John Duntemann (WJE)

Additionally, interim results were reviewed by a steering committee led by Bruce Ellingwood (Colorado State University) with additional members Jeanette Torrents (Structural Engineers Association of Colorado) and Therese McCallister (National Institute of Standards and Technology). Each committee member and reviewer generously gave of their time to improve the quality of the results presented in this report. The authors would like thank all those who have generously provided their time as part of this effort.

Acronyms

AISC	American Institute of Steel Construction
CDF	Cumulative Distribution Function
COV	Coefficient of Variation
CP	Cost-Complexity Parameter
CRREL	Cold Regions Research and Engineering Laboratory
D2C	Distance-to-Coast
EPA	Environmental Protection Agency
FOS	First Order Stations
GAM	Generalized Additive Models
GEV	Generalized Extreme Value
GHCND	Global Historical Climatology Network - Daily
GR	Ground-to-Roof Conversion Factor

MLE	Maximum Likelihood Estimation
MME	Method of Moments Estimation
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service
RF	Random Forest
RGAM	Regional Generalized Additive Model
RMCD	Rocky Mountain Conversion Density
RSS	Residual Sum of Squares
RT(L)	Reliability Targeted (Load)
SNODAS	Snow Data Assimilation System
SNOTEL	Snowpack Telemetry
SNOW	Snow Course
USGS	United States Geological Survey
WESD	Water Equivalent of Snow Depth

Chapter 1

Introduction

The United States has a rich history of snow load studies at the state and national level. The current ASCE 7 snow loads are based on studies performed at the Cold Regions Research and Engineering Laboratory (CRREL) ca. 1980 and updated ca. 1993. The map includes large regions where a site-specific case study is required to establish the load. Many state reports attempt to address the "case-study regions" designated in the current ASCE 7 design snow load requirements. The independently developed state-specific requirements vary in approach, which can lead to discrepancies in requirements at state boundaries. In addition, there has been great interest to develop site-specific reliability-targeted loads that replace the current load and importance factors applied to 50-year snow load events as defined in ASCE 7-16. This interest stems from the fact that the relative variability in extreme snow load events is not constant across the country, leading to a non-constant probability of failure for a given design scenario.

This report describes efforts to achieve three objectives:

1. Identify a representative reliability-target design snow load scenario that incorporates advancements made in the reliability and construction of structural members as well as changes made to snow-related provisions in ASCE 7.

2. Obtain site-specific probability distributions of annual snow loads at locations with sufficiently long histories of snow measurements. Use these distributions to estimate the nominal ground snow load required to achieve a desired level of reliability based on the scenario developed in Objective 1.
3. Estimate reliability-targeted loads between measurement locations to provide high resolution maps of reliability-targeted loads for the conterminous United States that varies smoothly across the landscape and eliminates inconsistencies at state boundaries.

Emphasis was placed on finding reproducible and data-driven solutions for each objective. This allows updates to this national effort to be made reasonably quickly for relatively little marginal cost as improved and updated information becomes available. This chapter summarizes the steps taken to obtain reliability-targeted loads on a national scale.

Chapter Highlights:

- The illustration of the need for a uniform risk approach to defining ground snow load as opposed to a uniform hazard approach.
- The summary of a reproducible workflow for obtaining reliability-targeted ground snow loads.
- A brief summary of the remaining chapters in the report.
- A high-level comparison of the changes in design snow load requirements from the current provisions with the move to reliability-targeted loads.

1.1. Project Aims

The final product of this project is a modern, universal, and reproducible approach for generating design ground snow loads for the conterminous United States. A natural consequence of this approach is the significant reduction of areas currently designated as case-study regions. The estimated loads resulting from this effort target a uniform risk for the entire country. This is in contrast to the current ASCE 7 approach for snow loads which target a uniform hazard (i.e. 50-year event) subject to a constant load factor and a discrete set of importance factors. Design loads targeting a uniform risk will be referred to hereafter as reliability-targeted design snow loads, or reliability-targeted loads (RTL).

The need for RTLs, as opposed to uniform hazard loads, stems from the fact that there are regional and local differences in the nature of the hazard itself. This is represented analytically by the shape of the probability distributions describing annual maximum ground snow load events. The term "annual maximum" describes the maximum snow load event occurring in the snow season beginning in October of the previous year and ending in June of the listed year. The shape of each distribution can be roughly classified as light-tailed, exponential-tailed, or heavy-tailed with examples provided in Figure 1.1. Note that the area under the curve for any specified range of values on the x-axis denotes the probability of observing an event in that range. It is the area under the curve in the upper tail of the distribution that is of greatest interest for structural safety. For example, a 50-year event is a value for which the area under the curve above the specified value is equal to 0.02 or 2%. Table 1.1 compares the 20-year (0.05), 50-year (0.02), and 100-year (0.01) standardized (unit-less) events resulting from each distribution. Included also

in Table 1.1 is the relative increase between 20, 50, and 100-year events. Note that the magnitude of the extreme events, *and the rate of increase between the extreme events*, are significantly larger for heavy-tailed distributions than for light-tailed distributions.

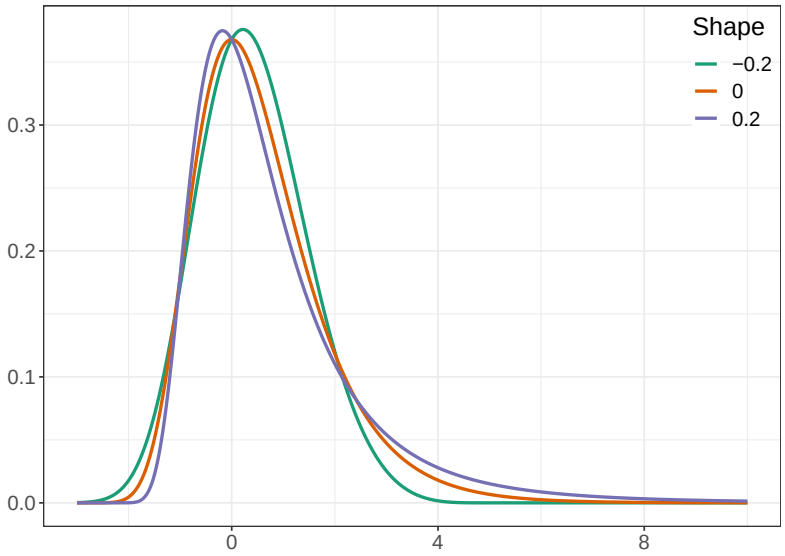

Figure 1.1: Example of a light (shape = -0.2), exponential (shape = 0), and heavy (shape = 0.2) tailed probability distribution.

Table 1.1: Comparing the relative increase in estimated extreme events for light (shape = -0.2), exponential (shape = 0), and heavy-tailed (shape = 0.2) distributions.

Event	Extreme Event			Relative Increase (%) (from 20-year event)		
	Light	Exponential	Heavy	Light	Exponential	Heavy
20 Year	2.2	3	4.1			
50 Year	2.7	3.9	5.9	23	30	44
100 Year	3	4.6	7.5	36	53	83

The crucial implication of these differing tail behaviors is that a 50-year

load multiplied by a constant load factor does not achieve a uniform design reliability. For locations whose annual maximum snow events are described by a heavy-tailed distribution, the constant load factor approach tends to under-estimate the load required to achieve the desired reliability target. For locations with light-tailed distributions, this same approach tends to over-estimate the RTL. This argument is demonstrated in DeBock et al. [2017] and Liel et al. [2017] which show that the constant load factor approach was conservative in the mountains of Colorado but unsafe in the eastern plains of Colorado. For this reason, this report pursues the identification of site-specific RTLs, rather than 50-year ground snow loads.

The implications of this paradigm shift are best illustrated by way of example. Figure 1.2 shows three histograms of annual snow load maximums in Baltimore, MD; Rochester, NY; and Duluth, MN. Note that Baltimore has a heavy-tailed distribution, Rochester has an exponential tail, and Duluth has a light tail. In places like Duluth, the light upper tail leads to an RTL slightly less than current requirements while the heavy tail in Baltimore is much greater than current requirements.

1.1.1. "The Next Storm"

Recall that both current and new snow load requirements shown in Figure 1.2 need to be multiplied by 1.6 in order to obtain the design ground snow load. This multiplication almost always results in the design ground snow load exceeding any of the observed snow loads in a 50-100 year period. In some cases, particularly at stations with short periods of record, the design ground snow load may greatly exceed any observed snow loads. The target probability of failure for a Risk Category II structure in a 50 year period is a mere 0.13%,

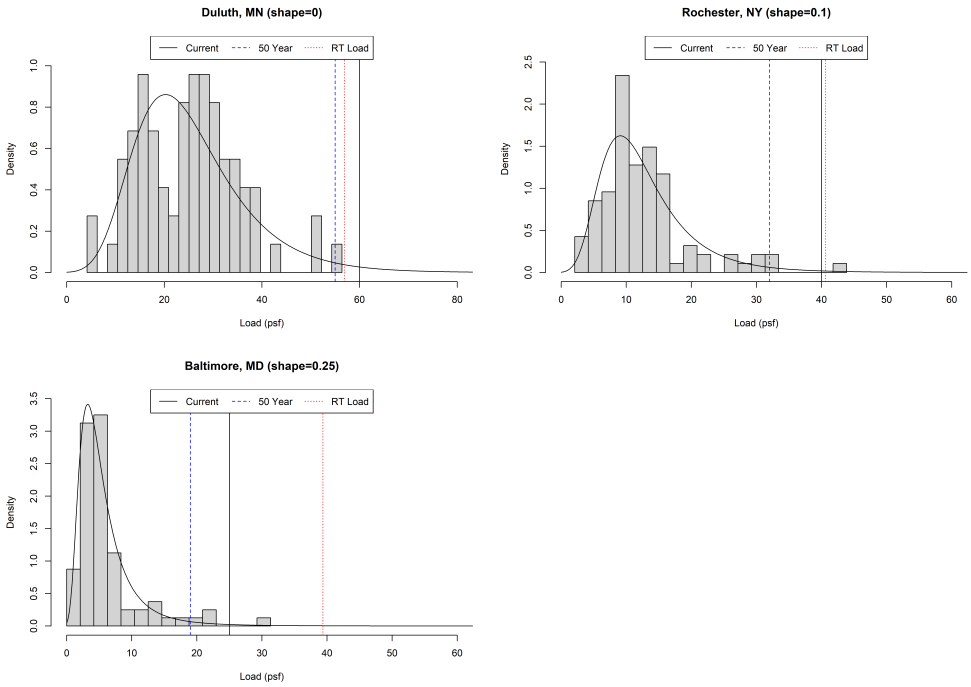

Figure 1.2: Histograms of annual maximum snow loads with fitted probability distributions overlaid. Included also is a comparison of the new 50-year and RTLs (divided by 1.6) to the current ASCE 7 requirements.

or one failure every 37,000 years. This exceedingly low probability might be thought of as the probability of a building being required to withstand the peak snow load in a year with two consecutive "superstorms," the kind of storm observed only once every 50-100 years, let alone *twice*. For places like Baltimore, MD, an additional "Snowmageddon" NESDIS [2020] storm would result in a proportionally larger increase in the annual peak snow load than if that same storm hit Duluth, MN. This is because Baltimore's peak snow loads tend to be the product of a few large storms, while Duluth's peak snow loads tend to result from an accumulation of many storms through the year. Such an explanation is consistent with the observation, made both in this report

as well as in Liel et al. [2017], that the difference between the RTL and the 50-year load are smaller in locations that consistently accumulate snow each year. Remembering that design loads are intended to be larger than observed snow loads aids in the evaluation of the results presented in this report.

1.2. Project Workflow

Figure 1.3 visualizes the workflow for estimating RTLs. Red boxes indicate data/information, tan boxes indicate actions, and blue boxes indicate decision points. The reliability analysis conducted in this report follows the pattern for reliability analysis set forth in Ellingwood et al. [1980]. The primary difference is that the reliability analysis is conducted using site-specific probability distributions, rather than using an aggregation of several site-specific probability distributions to derive a constant load factor. DeBock et al. [2017] and Liel et al. [2017] provide the template for the site-specific reliability analysis approach pursued in this report. This template was supplemented by lessons learned from many state-specific snow load studies [Tobiasson et al., 2002, Theisen et al., 2004, SEAO, 2013, Al Hatailah et al., 2015, Bean et al., 2018, Meehleis et al., 2020] as well as national snow load studies [Tobiasson and Greatorex, 1997, Buska et al., 2020].

The process starts with raw measurements of snow depth (SNWD) or snow load (water equivalent of snow on the ground, denoted WESD) and ends with maps of RTLs that can be used by practicing engineers. Several intermediate steps are required to derive design snow loads from these raw measurements. Some of those steps require assumptions/estimates that introduce uncertainty into the workflow and are denoted by the red arrows. It is not practical to fully

account for every possible source of uncertainty in the estimation process. Fortunately, DeBock et al. [2016] demonstrated that some sources of uncertainty, such as the uncertainty resulting from the estimation of snow load from snow depth, are not consequential in the estimation of RTLs as long as the estimates of snow load from snow depth are unbiased. This study accounts for sources of uncertainty known to be of greatest consequence in the RTL estimations, namely:

- The uncertainty in the extreme ground snow load events.
- The uncertainty in the conversion from ground loads to roof loads.
- The uncertainty in the resistance members of the target-reliability scenario.

Decisions regarding how to characterize uncertainty in the workflow were made using expert judgement on the part of the authors in collaboration with the project steering committee. The following subsections provide brief summaries of each step in this workflow.

1.2.1. (Chapter 2) Define Reliability-Target Scenario

The reliability-target scenario is a steel beam supporting a heated flat roof in normal exposure conditions. This chapter describes the selection of probability distribution parameters that properly characterize this target scenario. These distributions reflect changes that have been made in the production and understanding of structural steel, as well as changes that have been made to ASCE 7 provisions since the development of the 1.6 load factor for snow loads in Ellingwood et al. [1980]. This chapter discusses changes made to ASCE 7 since the original load factor calibrations as well as the implications of those changes on the resulting RTL calculations.

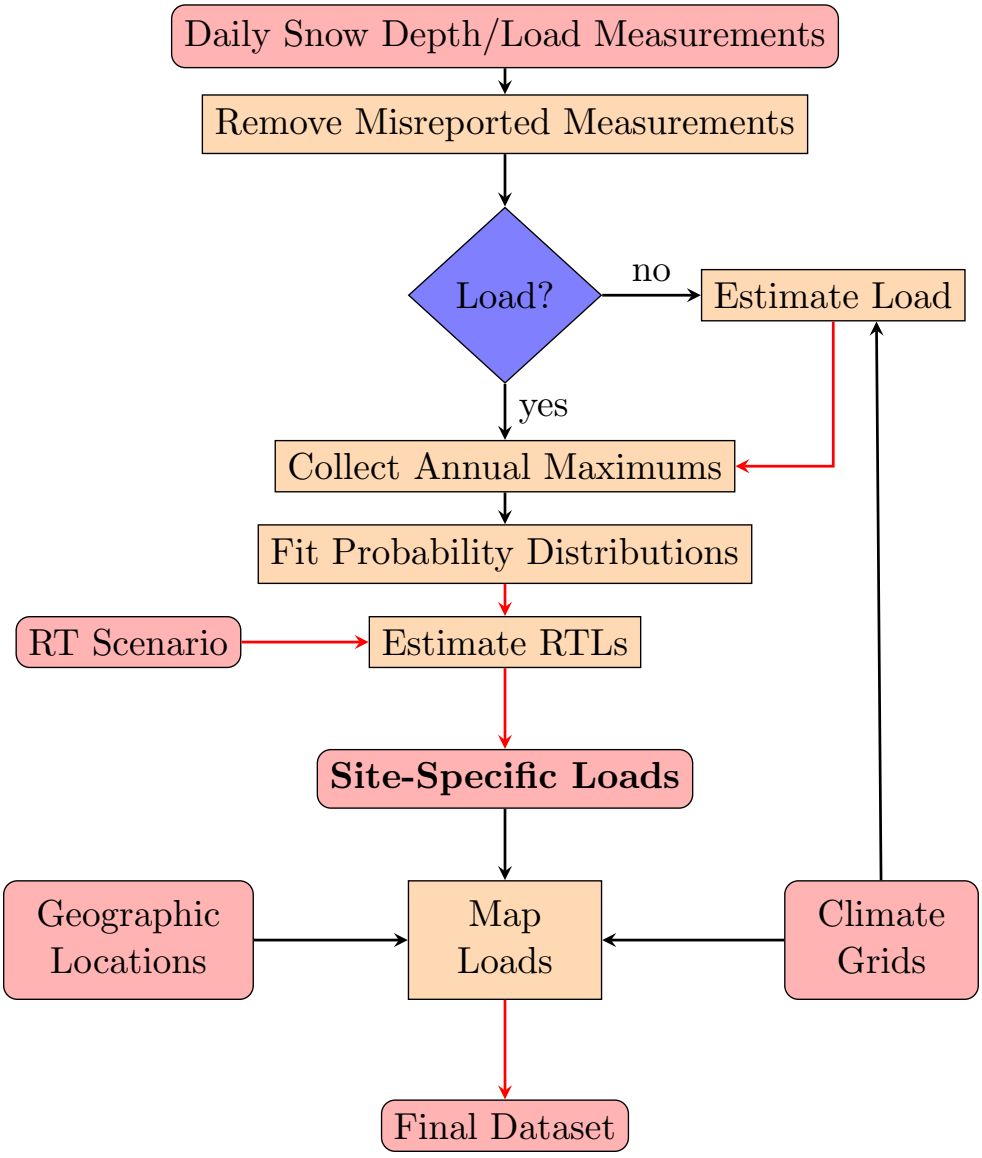

Figure 1.3: Workflow for obtaining reliability-targeted (RT) maps from daily measurements of snow.

1.2.2. (Chapter 3) Create Ground to Roof Conversion Models

One crucial element of the reliability analysis described in Chapter 2 is the assumed probability distribution characterizing the ratio between the maximum ground and roof snow loads, referred to as GR. This chapter reviews existing methods and available datasets for estimating GR and proposes a new ground snow load dependent GR model using the best available data.

1.2.3. (Chapter 4) Clean and Process Data

Site-specific RTLs are very sensitive to the probability distribution used to describe annual maximum snow load events. This chapter describes efforts made to download, clean, and process daily snow measurements from the National Oceanic and Atmospheric Administration's Global Historical Climatological Network [Menne et al., 2012]. The raw dataset contained more than 236 million observations at more than 65 thousand locations across North America. Observations considered extended from the late 1800s through June of 2020. Only stations with sufficiently long histories of high quality measurements were retained, resulting in RTL estimates at nearly 8,000 measurement locations in the conterminous United States and southern Canada.

1.2.4. (Chapter 5) Estimate Load from Depth

There are relatively few snow measurement locations that make direct measurements of snow load. This makes it necessary in many situations to estimate the snow load from the snow depth. There is an extensive history of models aimed at relating a 50-year/annual snow depth to a 50-year/annual snow load. Most of these models have used high altitude snow depth/load measurement pairs, though others have used the National Weather Service's first-order stations.

Snow density is fundamentally different at high altitude/high load locations as compared to locations that receive intermittent snow. This makes it impossible to use any single existing depth to load conversion model to characterize snow density for all locations across the country. This chapter develops an approach for estimating snow load from snow depth that can accurately predict both mountainous and non-mountainous snow density with a single random forest model. Included also in the chapter are site-specific comparisons of snow densities using a variety of depth-to-load conversion models, as well as overall comparisons of accuracy between existing and proposed methods.

1.2.5. (Chapter 6) Fit Ground Snow Load Probability Distributions

The most significant piece of the reliability analysis described in Chapter 2 is the distribution of annual maximum snow loads. The reliability analysis requires the estimation of loads whose magnitudes far exceed any observed snow loads. This extrapolation can cause two distributions that produce similar 50-year loads to produce divergent RTLs. This chapter describes a series of steps intended to ensure robust and reasonable site-specific RTLs. Annual maximums are modeled with a generalized extreme value (GEV) distribution, which includes a third parameter that provides more flexibility in the distribution fitting process. The shape parameter is smoothed at a regional level to ensure that nearby and otherwise similar measurement locations have consistent RTLs. This chapter demonstrates that the distribution fitting process is more robust to outlier values than other distribution fitting strategies.

1.2.6. (Chapter 7) Map Reliability-Targeted Loads

Chapters 2-5 result in a table of RTLs at nearly 8,000 measurement locations. This chapter describes the method for estimating RTLs between these measurement locations. The method of choice for this task is called regional generalized additive models (RGAM), which fit trends between snow loads, elevation, winter precipitation, and temperature at a regional level. A smoothing scheme is used between predictions in adjoining regions which eliminates sharp changes in estimated loads along region boundaries. The accuracy of the RGAM approach is evaluated by means of cross validation.

1.3. Project Implications

In order to make comparisons to existing 50-year loads, the new RTLs in Figures 1.4 and 1.5 are divided by 1.6. This division by 1.6 makes the Risk Category II loads from the current study directly comparable to 50-year loads provided in current design requirements. The move to RTLs necessitates a change in the load factor from 1.6 to 1.0. This makes the new design snow load requirements substantially larger than the current snow load requirements defined by 50-year loads. Figure 1.4 shows a map of the newly proposed design snow load requirements for the country. Figure 1.5 shows the ratio between new and existing requirements at all locations where new and existing requirements are both between 10 and 100 psf. Current requirements were obtained from the ASCE 7 Hazard Tool using requirements available in ASCE 7-16.

In general, new requirements tend to be smaller than existing requirements in places where the maximum load is a product of consistent snow accumulation throughout the snow season. RTLs tend to be higher than current re-

quirements in locations where the maximum snow load is a product of only a handful of major storms. In general, more consistent snow accumulation patterns throughout the season are associated with lower RTLs relative to the 50-year loads.

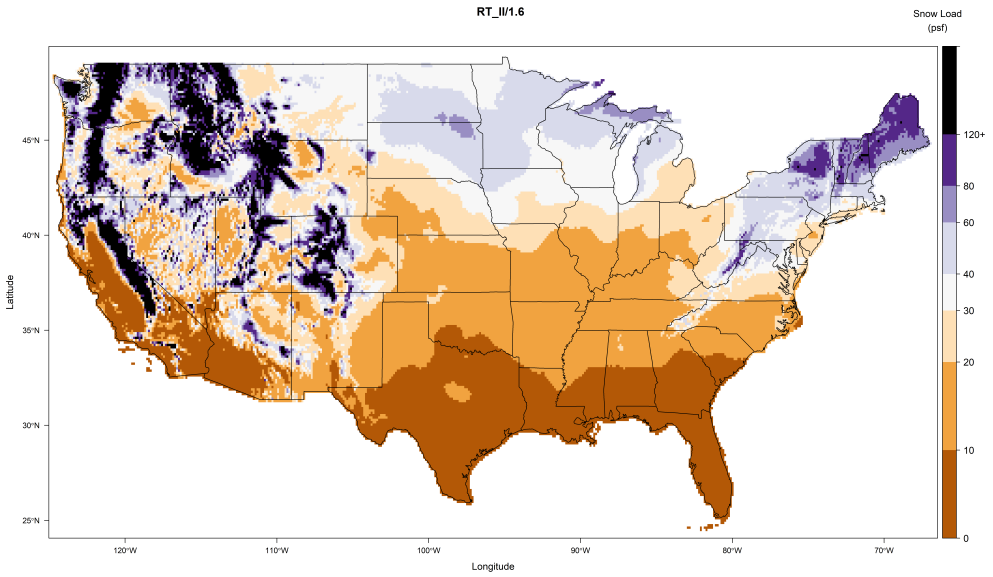

Figure 1.4: Map of Risk Category II ground snow loads (divided by 1.6) resulting from the 2020 National Snow Load Study.

The new requirements make a continuously varying set of design ground snow load predictions on a 0.5 mile (800 meter) resolution grid. This is in contrast to current requirements available in the ASCE 7 Hazard Tool, which define a single load for an entire geographic/elevation zone. The discreteness of the current requirements partially explains the large relative differences in design loads in western states as observed in Figure 1.5.

Mapping techniques described in Chapter 7 drastically reduce the number

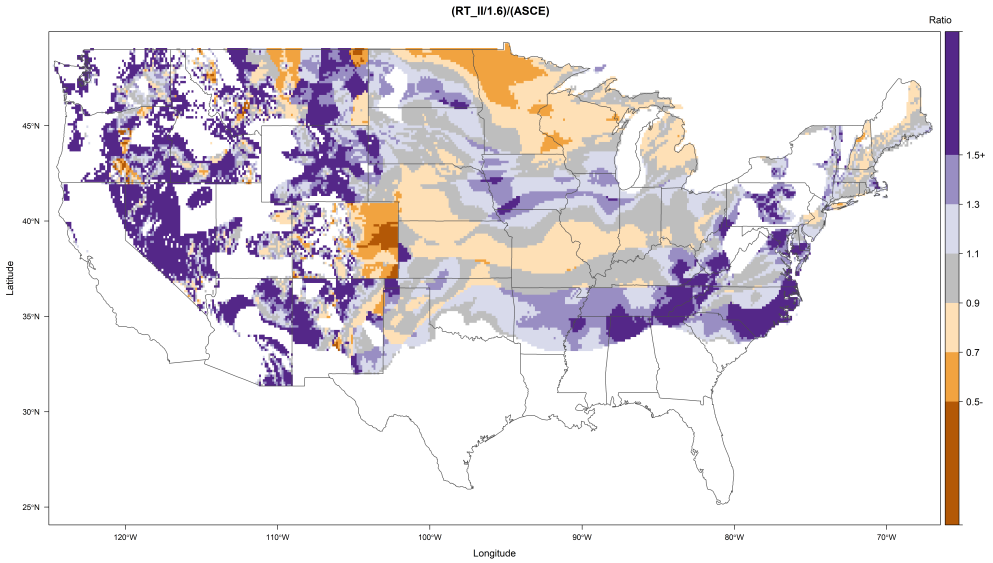

Figure 1.5: Map of the ratio between newly proposed and existing design ground snow load requirements. Ratios are only calculated in areas where both the new and existing snow load requirements are between 10 and 100 psf. Note that limitations in the resolution of mapped values for existing requirements in the ASCE 7 Hazard Tool make comparisons difficult in most western states.

of previously defined "case study" regions. This approach reduces case study regions by 91% from what they are in ASCE 7-16 and 96% of what they were in ASCE 7-2010. The difference in the reduction is due to the addition of state-level studies to the ASCE 7 standard between 2010 and 2016 which eliminate case study regions in some states. Case study regions are now confined to locations with elevations far exceeding the elevations of surrounding measurement locations, typical of high mountain peaks in the intermountain west.

Another important consequence of this work is the elimination of load and importance factors. Rather, the RTL is directly provided to the user for each risk category. Figure 1.6 shows a map of the ratio between the Risk Category

II and Risk Category IV RTLs. Under current ASCE 7 provisions, the ratio between these two quantities is a constant value of 1.2. However, this map illustrates that this ratio is highly dependent upon the shape of the annual snow load probability distributions in the region. The move to direct estimates of RTLs ensures that the same structure will have the same probability of failure due to snow, regardless of its location for all Risk Categories in the United States. The ensuing chapters illustrate the creation of site-specific RTLs in a new era of design snow load requirements.

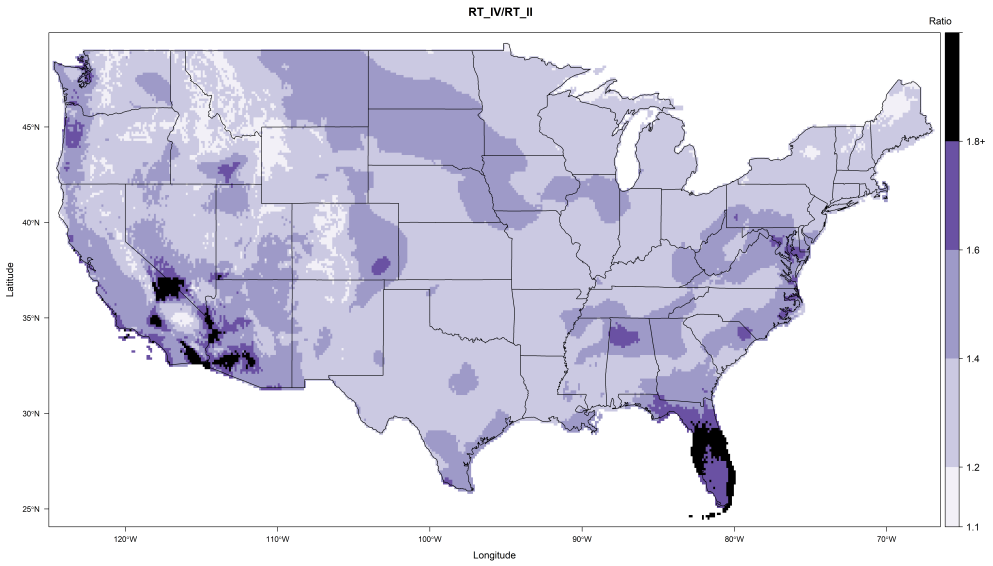

Figure 1.6: Comparison of the ratio between Risk Category II and IV loads resulting from the 2020 National Snow Load Study.

Bibliography

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. Accessed: 12-1-2020.

Bean, B., Maguire, M., and Sun, Y. (2018). The Utah snow load study. Technical Report 4591, Utah State University, Department of Civil and Environmental Engineering.

Buska, J. S., Greatorex, A., and Tobiasson, W. (2020). Site specific case studies for determining ground snow loads in the United States. Technical report, Engineer Research and Development Center, Hanover, NH. Accessed: 11-30-2020.

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

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.

Ellingwood, B., Galambos, T. V., MacGregor, J. G., and Cornell, C. A. (1980). Development of a probability based load criterion for American National Standard A58: Building code requirements for minimum design loads in buildings and other structures, volume 13. US Department of Commerce, National Bureau of Standards.

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.

Meehleis, K., Folan, T., Hamel, S., Lang, R., and Gienko, G. (2020). Snow load calculations for alaska using ghcn data (1950–2017). Journal of Cold Regions Engineering, 34(3):04020011.

Menne, M., Durre, I., Korzeniewski, B., Vose, R., Gleason, B., and Houston, T. (2012). Global historical climatology network - daily (ghcn-daily), version 3.26. Accessed: 4-6-2020.

NESDIS (2020). 10 years later, Snowmageddon records still stand. www.nesdis.noaa.gov. Accessed: 11-30-2020.

SEAO (2013). Snow load analysis for Oregon. Structural Engineers Association of Oregon, Portland, OR, fourth edition.

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., Buska, J., Greatorex, A., Tirey, J., and Fisher, J. (2002). Ground snow loads for New Hampshire. Technical report, Cold Regions Research and Engineering Laboratory.

Tobiasson, W. and Greatorex, A. (1997). Database and methodology for conducting site specific snow load case studies for the United States. In Proc., 3rd Int. Conf. on Snow Engineering, Izumi, I., Nakamura, T., and Sack, RL, eds., AA Balkema, Rotterdam, Netherlands, pages 249–256.

Chapter 2
Selected Conditions for Reliability-Targeted Loads

Reliability analysis requires modeling the relationship between the hazard (snow), the resistance members (a steel beam), and the design code. While the distribution of the ground snow load is constant regardless of the structure, interaction between the hazard, resistance, and design provisions changes based on the target structure. This chapter defines the target design scenario as well as the parameters used to describe the structural resistance and conversion from ground load to roof loads.

Since the seminal study of Ellingwood et al. [1980], the probability-based method for calculating load factors and load combinations have been widely used and proven satisfactory. It provides a foundational framework for calibrating load factors based on the available information. As a brief review on the historical development, this chapter will also discuss how the Ellingwood et al. [1980] framework has been adapted by various researchers to re-calibrate load factors to account for changes to the design load provisions in ASCE 7, as well as an improved understanding of the distribution of resistance members and ground to roof conversion factors (GR) over the decades. In particular, this framework enables the direct calculation of site-specific reliability-targeted design ground snow loads (RTLs), as opposed to a single load factor. The re-

mainder of this chapter is devoted to documenting the conditions for resistance and GR used in the RTL calculations.

Chapter Highlights:

- A summary of the original load factor calibration by Ellingwood et al. [1980].
- A summary of changes to design snow load provisions that prompt a re-calibration of the snow load factor.
- A description of the design reliability-target scenario along with associated probability distribution parameters.
- A summary of the simulation strategy used to estimate RTLs.

2.1. Previous Snow Load Calibration and Required Context

A summary of pertinent research related to the snow load combination and reliability analysis is provided for context. The goal of this section is to provide a basis for comparison to the selected reliability-target scenario described later in this chapter. When necessary, original notation has been adjusted to ensure consist notation among the referenced literature.

2.1.1. Ellingwood et al. (1980)

The original calibration for the ANSI A58 and the later ASCE 7-88 load factors related to the dead plus snow load case has remained unchanged until the current version of ASCE 7-16. The seminal load and resistance factor cal-

ibration for ASNI A58 load combinations was performed by Ellingwood et al. [1980] with additional information provided in Ellingwood et al. [1982] and Galambos et al. [1982]. The nominal load combination recommended in 1980 as well as ASCE 7-16 is presented in (2.1) as

$$1.2D_n + 1.6S_n = \phi R_n \tag{2.1}$$

where

- D_n is the nominal dead load,
- S_n is the nominal snow load,
- ϕR_n is the nominal factored resistance.

The arrival at these load and resistance factors is described in some detail and is based on a weighted approach intended to arrive at an optimal selection of partial safety factors applied to the nominal resistance, dead load, and snow load. The load and resistance factors are highly interdependent across hazards, often constraining researchers to propose updated values under the limitation that other relevant load and resistance factors be held constant. Such was the case for Ellingwood et al. [1980], who was constrained to use a dead load factor of 1.2 when defining the snow load factor.

Table 2.1 presents the Ellingwood et al. [1980] optimal load and resistance factors for a steel beam as estimated using the information available at the time. There is a strong dependency between the load (γ) and resistance factors (ϕ), namely that larger values of ϕ require larger load factors to achieve the desired reliability index. A 1.6 load factor for snow, in tandem with a resistance factor of 0.79 and a load factor of 1.2 for dead load, were shown to achieve the desired reliability target index of 3.0. The presentation of optimum factors, as

well as optimum resistance factors for a 1.6 snow load factor and a 1.2 dead load factor, illustrates the need for Ellingwood et al. [1980] to accommodate constraints outside the scope of the referenced study.

Table 2.1: Steel Beam Optimal Load and Resistance Factors for Gravity Loads (excerpt from Table 5.3 Ellingwood et al. [1980]).

Material	Combination	Optimum Values		Optimum ϕ for
		ϕ	Y_L, Y_S	$Y_D = 1.2, Y_L = 1.6$
Steel Beam	$D + L$	0.96	2.10	0.78
($\beta_0 = 3$)	$D + S$	1.05	2.32	0.79

The roof snow load model used in the previous calibration is critical to discuss with relation to the current study. The roof snow load model used in the ANSI A58 calibration was:

$$S = G_r G_l$$

where

- S is the random variable associated with roof snow loading,
- G_r is the random variable representing the ratio between the max ground load and the max roof load,
- G_l is the random variable for ground snow load.

The distribution of G_l and associated parameters for reliability analysis were developed from eight sites (shown in Table 2.2) that were part of a larger statistical analysis of 180 first order weather stations and other sites between the winter of 1952-1978 as documented in Tobiasson and Redfield [1980]. These sites made up the basis for the ASCE nominal (i.e. 50-year) ground snow loads (P_g) and/or maps for ANSI A58.

Table 2.2: Water-Equivalent Ground Snow Load Data (excerpt from Ellingwood et al. [1980]).

Site	Annual Extreme Ground Load			A58.1-1972	50-yr Maximum Roof Load	
	Years of Record	λ	ζ	q_n (i.e. P_g)	u	a
Green Bay, WI	26	2.01	0.70	28	0.87	5.07
Rochester, NY	26	2.49	0.56	34	0.83	6.16
Boston, MA	25	2.28	0.51	30	0.70	6.63
Detroit, MI	20	1.63	0.58	18	0.69	5.97
Omaha, NB	25	1.60	0.69	25	0.62	5.20
Cleveland, OH	26	1.50	0.58	19	0.60	6.30
Columbia, MO	25	1.21	0.84	20	0.69	4.05
Great Falls, MT	26	1.77	0.49	15	0.80	7.16

Log-normal distributions were fit using annual extreme ground snow loads. These log-normal distributions were combined with G_r to develop distributions for 50-year roof loads that were assumed to follow a Type II distribution. The Type II distributions in the final column of Table 2.2 were averaged to obtain $\mu = 0.72$ and $\alpha = 5.82$. These correspond to a bias of 0.82 and coefficient of variation (COV) of 0.26 for the roof snow load distribution.

The nominal ground-to-roof conversion factor (C_n) is nominally 0.8 in this version of the ANSI A58 standard, but is currently 0.7. The random variable G_r was assumed to follow a normal distribution with a mean of 0.5 and a COV of 0.23.

With respect to the current study, flexural yielding of a simply supported beam is the most critical resistance parameter. The Ellingwood et al. [1980] resistance statistics were mean-to-nominal ratio (bias) of 1.07, COV of 0.13, and followed a log-normal distribution. The reliability analysis primarily em-

ployed the Rackwitz-Fiessler procedure, which is a quickly converging iterative procedure that can accurately accept any distribution type.

2.1.2. Bennett (1988)

Shortly after the development of the ANSI A58 and ASCE 7-88 standard which imposed the load and resistance factors described above, Bennett [1988] performed a reliability analysis that investigated changes to both the code (such as changing the ground-to-roof conversion factor from 0.8 to 0.7) and the statistical model for the ground-to-roof conversion factor. These changes were based on a CRREL sponsored study by O'Rourke et al. [1983], which measured ground to roof conversion factors across the United States. Further details regarding this study are provided in Ellingwood and O'Rourke [1985] and O'Rourke and Stiefel [1983]. This model is described as having a mean of 0.47 and COV of 0.42.

It is expected that a lower nominal ground to roof conversion factor and a more variable GR model would require larger loads to achieve the same reliability-targets. Using the ground snow load model provided in Ellingwood et al. [1980], Bennett [1988] confirmed that these changes resulted in reliability indices less than 3.0 in all cases. Bennett [1988] ultimately recommended increasing the snow load factor from 1.6 to 2.0 to obtain a target reliability of only 2.0 and indicated a load factor of up to 4.6 may be needed to obtain a reliability index of 3.0.

Bennett [1988] also opined that it is difficult to develop models for snow suitable for reliability analysis due to the nature of the data. This is because reliability analysis requires the modeling of N-year recurrence intervals which could be very large and in excess of what may be possible for a theoretical

distribution. This was also discussed in Ellingwood and Redfield [1983] where 1000-year events may be needed for the reliability index. Chapter 6 discusses strategies to ensure consistent tail extrapolations in the face of limited periods of record.

2.1.3. Bartlett et al. (2003)

Bartlett et al. [2003] sought to update statistical parameters for steel members to reflect those of current A992, Grade 50, materials rather than the A36 parameters from the 1960s and 1970s from Galambos and Ravindra [1978] and ultimately update the resistance parameters for reliability calibration. Table 2.3 is a reproduction of the original and proposed resistance calibration parameters presented by Bartlett et al. [2003]. These numbers without discretization, which was not considered in the original calibration, have lower bias and COV. Bartlett et al. [2003] performed a reliability analysis considering the dead plus live load case, but did not investigate snow.

Table 2.3: Reproduction of Table 9 from Bartlett et al. [2003].

Factor	Original Calibration		Current Calibration			
			No Discretization		With Discretization	
	Bias	CoV	Bias	CoV	Bias	CoV
Geometric	1.00	0.05	1.00	0.034	1.00	0.034
Material	1.05	0.10	1.028	0.058	1.028	0.058
Professional	1.02	0.06	1.02	0.06	1.02	0.06
Discretization	1.00	0.00	1.00	0.00	1.05	0.043
Total	1.07	0.127	1.049	0.090	1.101	0.100

2.1.4. Lee and Rosowsky (2005)

Lee and Rosowsky [2005] proposed a new snow roof load for three different regions in the US suitable for reliability analysis, which intended to improve upon the original calibration ground and roof snow statistical models. Ground snow parameters were calculated for several sites in the United States and it was found that log-normal distributions fit best among most stations, though it seems that only the Type I Extreme Distribution was alternatively considered. In this case, as has been previously done up to this point, the entirety of the data was used to fit the distribution, lending little weight to the tail of the ground snow load distribution. This will be described in more detail later.

The ground to roof conversion factor was selected from the Ellingwood and O'Rourke [1985] and O'Rourke and Stiefel [1983] models and was combined with the distribution that fit the ground snow load. This was then simulated to obtain a 50-year roof load where the upper 10% of the tail was fit. By fitting the entire ground snow dataset and fitting only the upper 10% of the transformed roof snow load simulated data, it is unclear if the ground snow load tail dynamics are preserved in the final presented roof snow load distribution. For roof snow the resulting regional log-normal distributions had a bias of 0.61 and COV of 0.53 for Northeast, a bias of 0.84 and COV of 0.60 for Midwest/Mid-Atlantic, and a bias of 0.8 and COV of 0.58 for Northern Midwest/Mountain West. These bias and COV are scaled based on the nominal values and are suitable for comparison with the original calibration 0.82 and 0.26 for bias and COV, respectively. While biases are largely similar for two regions as compared to the original calibration, the COV are approximately double for each region. These are also more severe than those investigated by Bennett [1988] which resulted in reliability indices below 2.0.

2.1.5. Galambos (2006)

Galambos [2006] investigated the reliability of the 2005 American Institute of Steel Construction (AISC) Specification in light of the information contained in Bartlett et al. [2003] and corrected for dynamic yield stress similar to that performed in Galambos and Ravindra [1978] and Jaquess and Frank [1999]. The material factor was used with a mean of 1.06 and COV of 0.06 (compare to Table 2.3). The fabrication factor was obtained from Galambos et al. [1982] with a mean of 1.0 and a COV of 0.05. The professional factor mean of 0.99 and COV of 0.06 were based on extensive tests found in White and Barker [2008], White and Duk Kim [2008], and White and Jung [2008]. Combining material, fabrication, and professional factors for comparison resulted in a mean of 1.05 and COV of 0.1 with the resistance parameter following a log-normal distribution (nearly identical to that in Table 2.3).

Galambos [2006] also investigated the effects of snow plus dead load reliability using the Ellingwood et al. [1980] roof snow distribution. The reliability analysis method was the log transform of the first order second moment reliability index introduced by Hasofer and Lind [1974] which assumes both load and resistance are log-normal random variables:

$$\beta = \frac{\ln\left(\frac{\bar{R}}{\bar{Q}}\right)}{\sqrt{V_R^2 + V_Q^2}} \tag{2.2}$$

where

- $\bar{R}$ and $\bar{Q}$ are the mean values of the resistance and the load, respectively
- V_R and V_Q are the corresponding COVs

This method, a first order second moment method, is known to produce

issues when COVs are large [Turkstra and Putcha, 1985], but the log-transform should help account for this with log-normally distributed inputs. The random variable C_s was assumed to follow a normal distribution with a mean of 0.5 and a COV of 0.23. This process consistently produced reliability indices above 3.0 as shown in Figure 2.1.

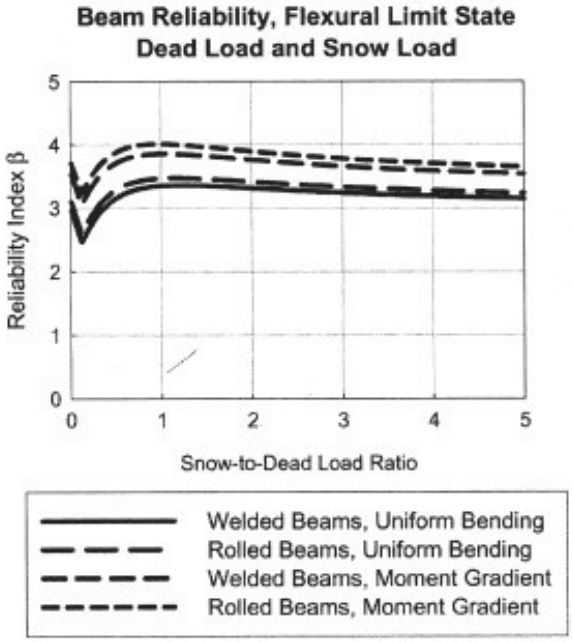

Figure 2.1: Reproduction of Dead Plus Snow Reliability Indices from Galambos [2006].

2.1.6. The Colorado Study: Reliability Targeted Loads

The 2016 Colorado Study [DeBock et al., 2016] addressed the aforementioned lack of snow load requirements for mountainous states discussed in Chapter 1. What sets this study apart from other state-specific studies is the pursuit of site-specific reliability-targeted design ground snow loads. These site specific

RTLs addressed a pressing concern in the state that design ground snow loads were too low on the eastern plains of Colorado [DeBock et al., 2017, Liel et al., 2017]. Coefficients of variation at stations of lower elevations were much larger than those found in the mountains. This resulted in the site-specific RTL concept. To accomplish this task, DeBock et al. [2016] performed analyses to identify the load at each Colorado station that would result in the target reliability index of 3.0. In many locations this resulted in a dramatic increase above the ASCE 7 stipulated 50-year ground snow load, but in many higher elevations resulted in a slight reduction.

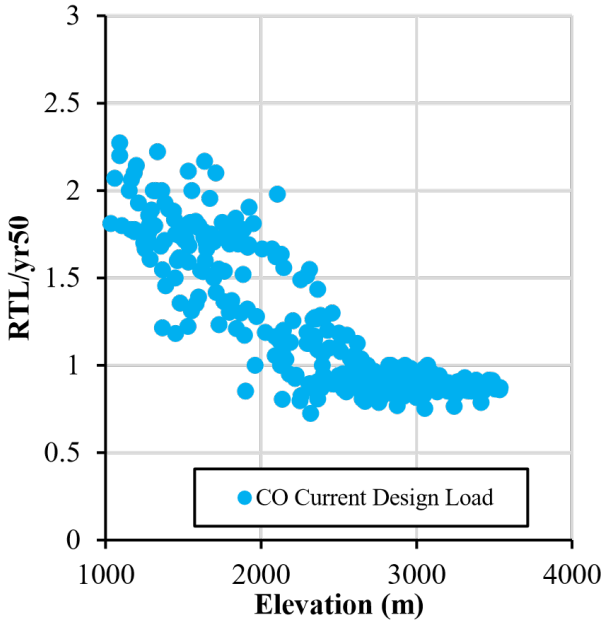

Figure 2.2: RTL/50-year ground snow load versus elevation (adapted from cover of DeBock et al. [2016]).

The Colorado study targeted steel flexural yielding (i.e., $R = Z_x F_y$) as the resistance limit state and obtained steel yield strength (F_y) log-normal random variable parameters with a bias of 1.10 and COV of 0.09 from Ellingwood

et al. [1980]. The plastic section modulus (Z_x) was modeled as a normally distributed random variable with a bias of 1.05 and COV of 0.05 [Galambos and Ravindra, 1978, Lind, 1977] to account for discretization (similar to Table 2.3). For comparison with the previous studies, these were combined by the authors to produce a normally distributed random variable with bias of 1.155 and COV of 0.103.

Like the other studies, DeBock et al. [2016] used the original calibration dead load random variable parameters (normally distributed, 1.05 and 0.1 bias and COV respectively), but rather than varying dead-to-live-load ratio, they targeted a constant dead load of 15 psf, reflecting the fact that larger loads may not result in significant increases in dead load for many light roof systems.

Snow station parameters were clustered using an expert-based superstation approach to arrive at fairly controlled tails of log-normal distributions. A tail fitting approach was used for the upper 10% of observations in the superstations (used to establish the ground snow load COV) and the upper 33% at the original measurement locations (used to establish the ground snow load magnitudes).

The GR model was based on observations in Norway presented by Thiis and O'Rourke [2015]. This is one of the largest databases of its kind, but is held largely in strict confidence by the Norwegians. Little other information about these data are known, beyond what is in the Thiis and O'Rourke [2015] publication. The data seem to imply that there is a ground snow load trend that makes some physical sense in that larger loads will persist longer, thus reducing

the maximum potential GR. The equations for the distribution parameters are

$$\mu_{ln} = \ln(0.50 \exp(-0.034 + g_l 0.4)$$

$$\sigma_{ln} = \min(0.007 g_l + 0.1, 0.33)$$

where g_l represents simulated values from the ground snow load distribution G_l.

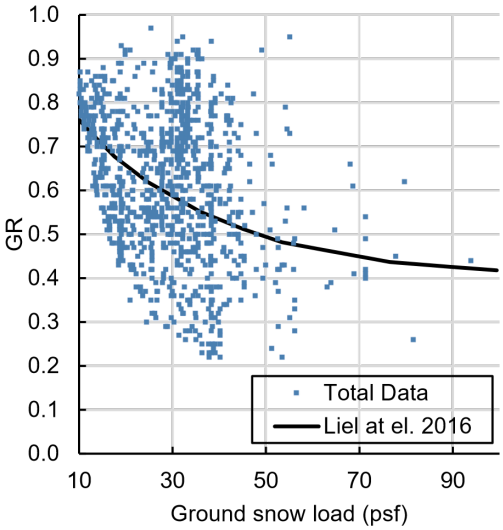

Figure 2.3: Ground to roof conversion factor (GR) versus Ground Snow Load (psf) from Liel et al. [2017].

Once station parameters and GR were known, they were simulated and combined with the dead load and compared with the simulated resistance to result in an annual probability of failure. This was then converted to a 50-year probability of failure. The process was repeated using a different nominal load until each location achieved the target 50-year reliability. Following this, loads were estimated between measurement locations using an interpolation approach.

2.1.7. Synthesis of the literature

The purpose of this literature investigation is not to be critical of past studies, but to illustrate the need for re-calibrating the snow load factor with the benefit of more information and updated design provisions. Each study illustrates that reliability-targeted design ground snow loads are very sensitive to assumptions regarding resistance members, ground to roof conversions, and design provisions. It is clear that the studies in this area have used a wide variety of reliability analyses and random variables. Table 4 attempts to summarize the main distribution parameters selected for the most important variables.

Table 2.4: Compilation of statistical parameters from previous work. The abbreviations N, LN, and Type II stand for Normal, Log-normal, and Type II Extreme Value distributions, respectively.

Study	Snow Roof Load			Resistance			Ground to Roof Conversion			
	Bias	COV	Shape	Bias	COV	Shape	Nominal	Mean	COV	Shape
Ellingwood et al. 1980	0.82	0.26	Type II	1.07	0.13	LN	0.8	0.5	0.23	N
Bennett 1988	1.17	0.47	Type II	1.07	0.13	LN	0.7	0.47	0.42	N
Lee and Rosowsky 2005	0.61-0.84	0.53-0.6	LN	-	-	-	0.7	0.47	0.42	LN
Galambos 2006	0.82	0.26	Type II	1.05	0.1	LN	0.8	0.5	0.23	N
Liel et al. 2017	Site specific	Site specific	LN	1.155	0.103	N	0.7	Eq	Eq	LN

Because of the differences in both analysis methods (first order second moment, Rackwitz-Fiessler, and Monte Carlo simulations), the data from each study was reproduced using each method and various input parameters. Figure 2.4a illustrates reproduction of the recommended load and resistance factors from Ellingwood et al. [1980] compared to the authors attempted reproduction. Figure 2.4b illustrates the effect of changes to the code made by industry as produced by Bennett [1988] and reproduced by the authors for model valida-

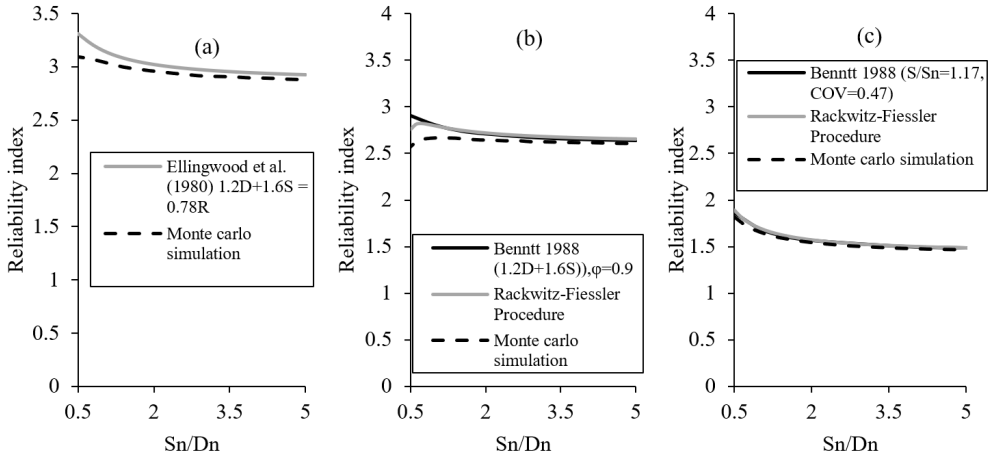

Figure 2.4: Reliability index vs nominal snow to dead load ratio using (a) Ellingwood et al. [1980] recommended load and resistance factors for steel flexural members (b) A58.1- 1982 code provisions and Ellingwood et al. [1980] loads as determined by Bennett [1988] (c) using Bennett [1988] worst case roof snow load.

tion purposes. Figure 2.4c shows the Bennett [1988] worst case scenario. The Ellingwood et al. [1980] case illustrates the expected target reliability scenario given the information available at the time. In both Bennett [1988] cases, the reliability is lower than the target 3.0 due to code changes that occurred following the Bennett [1988] calibration. Bennett [1988] showed that using the same assumptions in Table 2.4 in conjunction with the Ellingwood et al. [1980] roof snow model, the estimated reliabilities are near 2.5. In further analysis, Bennett [1988] developed additional snow load models, with the worst case presented in Table 2.4, that show reliability indices well below 2.0.

The Galambos [2006] study used an updated resistance parameter, but duplicated the original calibration with the exception of the reliability method. Figure 2.5 (left) presents digitized Galambos [2006] data along with reproductions of this analysis using different reliability methods. Using the non-

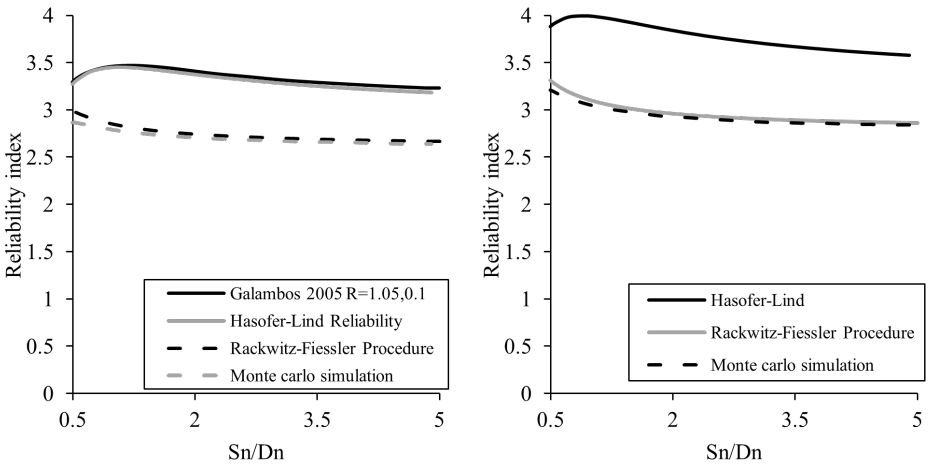

Figure 2.5: (left) Reliability index vs snow to dead load ratio using various reliability processes and parameters associated with Galambos [2006] and the same analysis using R based on Liel et al. [2017] (right).

simplified methods, the reliability indices are slightly below the target reliability, but significantly below the first order second moment method based on Table 2.2 [Lind, 1977]. To illustrate the effect the resistance parameter selection has on the reliability index, in Figure 2.5 (right), the same analysis was performed, but changing only the resistance parameter to that of Liel et al. [2017]. The Hasofer-Lind index is most greatly affected and shows an increase of approximately 0.5 and the other procedures result in an increase of approximately 0.25.

Figure 2.6 illustrates the reliability indices versus nominal snow to dead load ratio when using the Lee and Rosowsky [2005] regional roof snow load distribution parameters combined with resistance parameters from Ellingwood et al. [1980]. In all cases, the reliability indices are below the target values.

After digitizing all stations investigated by Lee and Rosowsky [2005], the reliability of each station was calculated to illustrate the effects of the updated

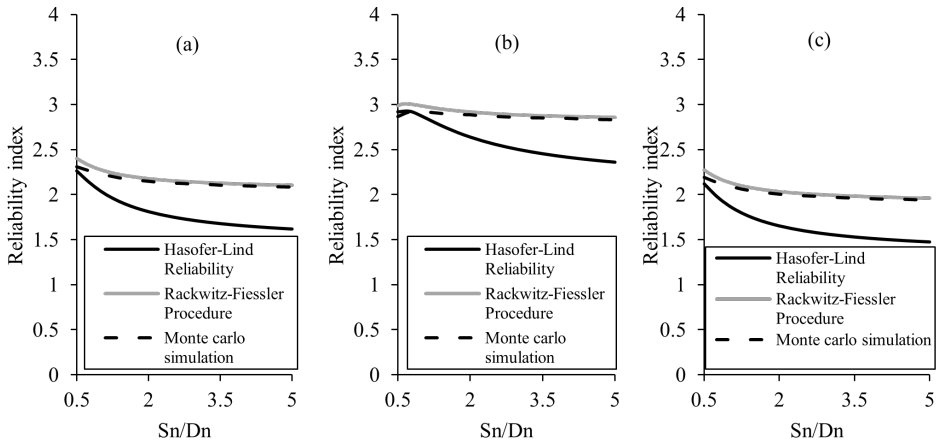

Figure 2.6: Reliability index vs snow to dead load ratio and Lee and Rosowsky [2005] snow roof load (a) bias of 0.8, and COV of 0.58 (b) bias of 0.61, and COV of 0.53 (c) bias of 0.84, and COV of 0.60.

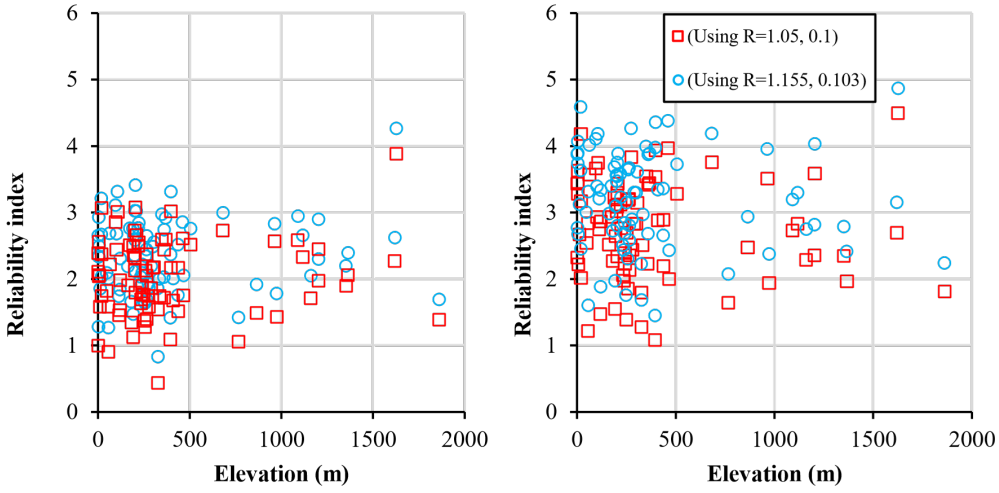

Figure 2.7: Reliability index versus elevation for (left) Monte Carlo Analysis and (right) First Order Second Moment using Lee and Rosowsky [2005] Roof Snow Load Station Parameters using Liel et al. [2017] resistance statistics (blue circles) and Bartlett et al. [2003] resistance statistics (red squares). Plot assumes constant dead load of 15psf as described by Liel et al. [2017].

roof snow load parameters and to justify future increases in snow loads as presented in Figure 2.7. Only a handful of actual stations achieved reliability indices above 3.0 when using the Bartlett et al. [2003] resistance statistics indicating that local reliabilities are likely lower than when using large regional composite statistics. Again, there is an approximate drop of 0.25 in the reliability indices when using Bartlett et al. [2003] resistance statistics when compared to those used in the Colorado study. Interestingly, there does not seem to be elevation dependence on the reliability index in the stations selected by Lee and Rosowsky [2005]. The same conclusions can be drawn from the first order second moment calculations in Figure 2.7. These results seem to be inflated, but are more reproducible and can be checked by hand using the tabulated values in Lee and Rosowsky [2005] and Table 2.2.

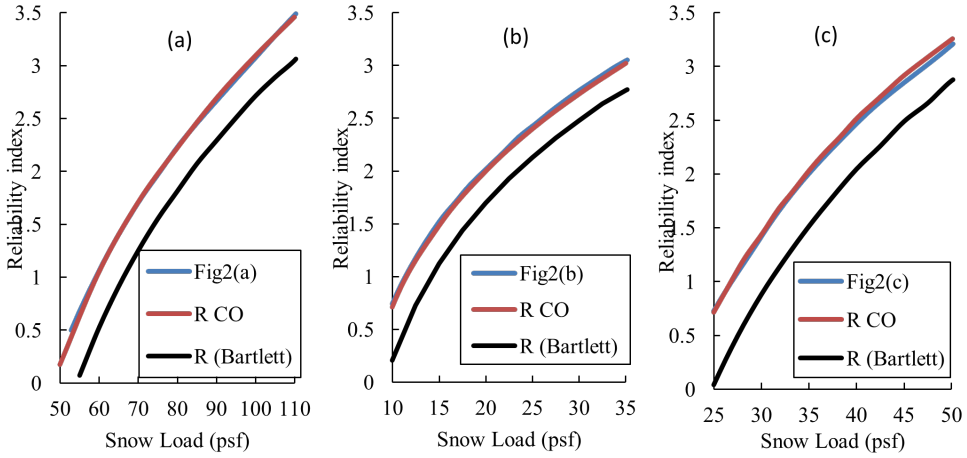

Figure 2.8: Digitized reproduction of DeBock et al. [2016] data (blue) Fig. 2a,b,c plotted alongside reliability reproduction by the authors using station specific parameters with Liel et al. [2017] parameters from Table 2.4 (red) and changing only the resistance to the recommended resistance parameters (black).

While the reliability targeted load (RTL) procedure presented in this paper

is yet to be described, the Colorado approach was reproduced and is demonstrated on three stations (the only stations presented with actual distribution parameters) in Figure 2.8. Using only the updated resistance distribution parameters from Bartlett et al. [2003], assuming no discretization, there is another significant drop in reliability indicating similar results to those in Figure 2.5b.

Using the framework developed by Colorado (and reproduction validated by the authors in Figure 2.8) the station specific RTLs were calculated for the Lee and Rosowsky [2005] station roof snow load models to estimate how much the loads would need to be increased from the current 50-year loads (calculated from the same distribution). Figure 2.9a presents the RTLs minus 50-year loads versus elevation, showing that loads will need to be raised substantially, on the order of 20-25psf on average and by as much as 75 psf for the worst-case station. There does seem to be some elevation dependence for RTL 50-year loads as higher elevations exhibit lower increases and also seem to exhibit smaller changes when using the updated resistance model. In Figure 2.9b, the ratio of the RTL and 50-year load also shows some elevation dependence where ratios generally decrease with increasing elevation, but this may be due to the lack of stations at the higher elevations. Based on Figure 2.9b, some low elevation stations would see increases over 3.5 times what the 50-year load would estimate. Figure 2.9 also indicates some stations would decrease in loads, though not as dramatically. The use of the updated resistance parameters when compared to the Colorado resistance model seems to increase loads on the order of 2.5 to 10psf and 10% to 25% greater.

In the preceding sections, the authors have illustrated how state-of-the-art snow load design has evolved since 1980. The purpose of the above exercises was to show that the authors could accurately produce and validate various

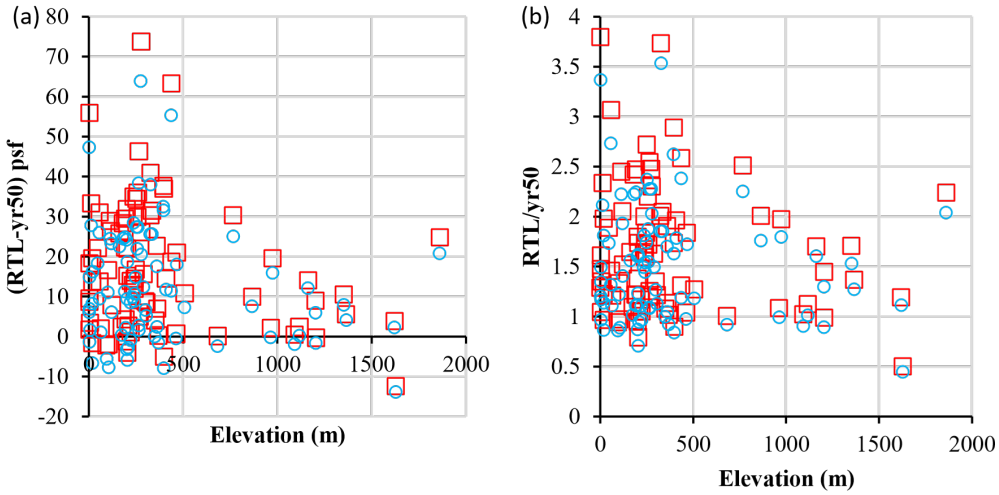

Figure 2.9: (a) reliability targeted load (RTL) minus the 50-year load and (b) RTL to 50-year load ratio versus elevation for Lee and Rosowsky [2005] Roof Snow Load Station Parameters using Liel et al. [2017] resistance statistics (blue circles) and Bartlett et al. [2003] resistance statistics (red squares). Plot assumes constant dead load of 15psf as described by Liel et al. [2017]

reliability analysis frameworks through reproduction of historical analyses from the literature. Furthermore, it is important to demonstrate the differences in change of parameters and methods on independently developed datasets and input parameters. From these analyses it is clear that, due to changes in design provisions and the distribution of resistance members, reliability indices no longer meet the criteria outlined in ASCE 7 for target reliabilities for snow loads on a national basis.

2.2. The Selected Target Scenario

Estimating RTLs requires the selection of a target situation, likely to be considered the most common case and then modify ground snow load values to

meet this target reliability based on the random variables associated with the structural situation. The selection of a reliability target does not guarantee a uniform reliability in all design scenarios, but a uniform reliability for all geographical locations given the target design scenario. In collaboration with the project steering committee, the selected target situation is a heated flat roof supported by a steel beam in normal exposure conditions.

The analysis is limited to the nominal snow load controlling load case in Table 2.1. Lambda (λ) will be defined as the ratio of the mean to the nominal value or bias:

$$\lambda_X = \frac{\mu_X}{X_n}$$

where

- μ_X is the mean of the random variable X
- X_n is the nominal parameter of interest for random variable X.

The COV is defined as the ratio of the standard deviation to the mean of the parameter of interest

$$V_X = \frac{\sigma_X}{\mu_X}$$

where σ_X is the standard deviation of random variable X.

2.2.1. Resistance Parameters

Plastic yielding of steel flexural member is selected as the target resistance limit state characterized as

$$\phi R_n = 0.9 Z_{x,n} F_{y,n} \tag{2.3}$$

where

- $Z_{x,n}$ is the nominal plastic section modulus
- $F_{y,n}$ is the nominal yield stress of the steel
- 0.9 is the resistance factor ϕ.

After lengthy discussion with the steering committee it was decided that the updated resistance statistics presented by Bartlett et al. [2003], assuming **A992 steel** are to be used in the analysis. The combined material, fabrication and professional random variable bias and COV are:

$$\lambda_R = 1.049 \qquad V_R = 0.09$$

assuming no discretization.

2.2.2. Load Parameters

Dead Load (D) is assumed to follow a normal distribution with statistical parameters taken from Ellingwood et al. [1980]:

$$\lambda_D = 1.05 \qquad V_D = 0.1.$$

The use of the normal distribution rather than log-normal distribution improves the computational efficiency of the Monte-Carlo simulations and does not affect the resulting reliability indices or RTLs.

Snow Load Statistical Parameters assume the ASCE 7-16 Nominal Load Model for flat roofs defined as

$$S_n = P_f = 0.7 C_e C_t I_s P_g. \tag{2.4}$$

where:

- P_f is the nominal flat roof snow load
- 0.7 is the nominal ground to roof snow load conversion ratio
- C_e is the exposure coefficient
- C_t is the thermal coefficient
- I_s is the snow importance factor
- P_g is the nominal ground snow load (currently 50-year per ASCE 7)

The load factor for nominal roof snow load was changed from 1.6 to 1.0. This study also elected to assume $C_s = C_t = C_e = 1.0$ and I_s will be removed as this study directly provides estimates for each risk category. As an aside, 20-year loads are also provided for use as service loads.

The proposed roof snow load model attempts to incorporate the uncertainties associated with the roof snow loading process. The ground to roof conversion factor (GR) is used to convert from ground to roof snow load. The coefficients in (2.4) ($0.7C_eC_t$) represent the nominal flat-roof GR model in ASCE 7. The target scenario results in $C_e = C_s = 1$ and the roof snow load model becomes

$$S = G_r * G_l$$

where

- G_r is the ground to roof conversion factor statistical model
- G_l is the statistical model for the ground snow load at a specific site which has site specific distribution parameters.

The ground snow load is assumed to come from a Generalized Extreme Value (GEV) Distribution. This distribution has three parameters called the location (μ_{G_l}), scale (σ_{G_l}) and shape (ξ_{G_l}). Chapter 6 provides the details regarding the GEV distribution fits in this National Study. Further, the G_r

model is assumed to follow a square-root-normal distribution [Stidd, 1970], with a ground snow load dependent mean

$$E\left(\sqrt{G_r}|G_l = g_l\right) = 0.9865 - 0.1192 * \log(g_l)$$

and standard deviation

$$\sigma_{G_r} = 0.18645$$

Details regarding the derivation of the G_l model are provided in Chapter 3.

In general, the greater the relative uncertainty in G_r and G_l, the larger the reliability-targeted load (all else equal). Other notable sources of uncertainty include the uncertainty in depth-to-load conversions, as well as the spatial, mapping, and the ground snow load distribution parameters. Bean [2019] illustrates the potential explosion of load magnitudes (and corresponding increase in loads) that occurs when accounting for compounding uncertainties in distribution parameter estimates, rather than simply accounting for the variability defined by the distribution itself. One issue with accounting for distribution parameter uncertainty is that it is primarily a function of data availability, rather than a function of snow dynamics. For example, all else equal, a station measuring snow depth would have a larger RTL than a station directly measuring snow load due to the increased uncertainty in parameter estimates resulting from the depth-to-load conversion. This makes it difficult to distinguish if the hazard, or the lack of information related to the hazard, is driving the estimated RTLs. For these reasons, the reliability-analysis will only characterize the variability in the ground snow, the roof conversion, and the resistance members, similar to related studies.

2.2.3. Reliability Analysis

Monte Carlo Analysis is used to combine distributions and determine the number of failures based on the selected load case and limit state above. The limit state equation to simulate is

$$G(R, Q) = R - Q$$

where

- R is the random variable describing the structural resistance
- Q is the random variable describing the load combination.

Q is defined as the 50-year roof snow load plus dead load (D_l). The targeted probability of failure is calculated as

$$Pr(R < Q) = \Phi(-\beta) \tag{2.5}$$

where:

- $Pr(R < Q)$ is the probability of failure of the member or system in a 50-year period
- Φ is the CDF of the standard normal distribution
- β is the reliability index.

Figure 2.10 illustrates the workflow for the Monte-Carlo simulations. While the GEV distribution models annual ground snow loads, direct simulations of 50-year ground snow loads are obtained using the relation from Lee and

Rosowsky [2005]:

$$F_{G_L}^{(50)}(x) = (F_{G_l}(x))^{(50)}$$
$$\left(F_{G_l}^{(50)}(x)\right)^{1/50} = F_{G_l}(x) \tag{2.6}$$

where F_{G_l} and $F_{G_l}^{(50)}$ represent the cumulative distribution of annual and 50-year ground snow loads respectively. The direct simulation of 50-year ground snow loads ($G_l^{(50)}$) differs from DeBock et al. [2016], but proved necessary to ease the computational burdens of carrying out the simulations on a national scale without affecting the RTL estimates. Similarly, the assumption that both R and D_l are normally distributed allows for the simulation of a single "adjusted resistance $R^* = R - D_l$ (which is also, by definition, normally distributed) which eases computation times.

Simulated events $\left(R^* - \left(G_l^{(50)} G_r\right)\right) < 0$ are considered failures. The number of tolerated failures corresponds with the target probability of failure defined in (2.5). Table 2.5 shows the tolerated number of failures in 1 million simulations for each risk category. One million simulations was shown to achieve stability in the RTL estimates while still being computationally feasible.

Table 2.5: Target number of failures from 1 million Monte-Carlo simulations for each Risk Category.

Category	β	Failures
I	2.5	6,209
II	3.0	1,349
III	3.25	577
IV	3.5	232

2.2.4. Monte-Carlo Simulation Steps

1. Define nominal P_g.
2. Count number of simulated failures using P_g:
   - Simulate $G_l^{(50)}$
     - Simulate a random number u between 0 and 1.
     - Calculate $u^* = u^{1/50}$ (see Equation (2.6)).
     - Calculate $g_l = F_{G_l}^{-1}(u*; \mu_{G_l}, \sigma_{G_l}, \xi_{G_l})$.
   - Simulate g_r and $r - d_l$.
   - Count number of times that $r - d - g_l * g_r < 0$.
3. If simulated failures exceed the target number of failures, increase P_g and repeat Step 2.

2.3. Related Chapters

This chapter describes a probability-based computationally feasible framework for estimating site-specific reliability-targeted loads. The result is a change in the snow load factor from 1.6 to 1.0 and the elimination of the importance factor I_s. Further details regarding the simulation process are provided in Chapter 3, which describes the derivation of the new G_r model, and Chapter 6, which describes the process of estimating annual ground snow load probability distributions.

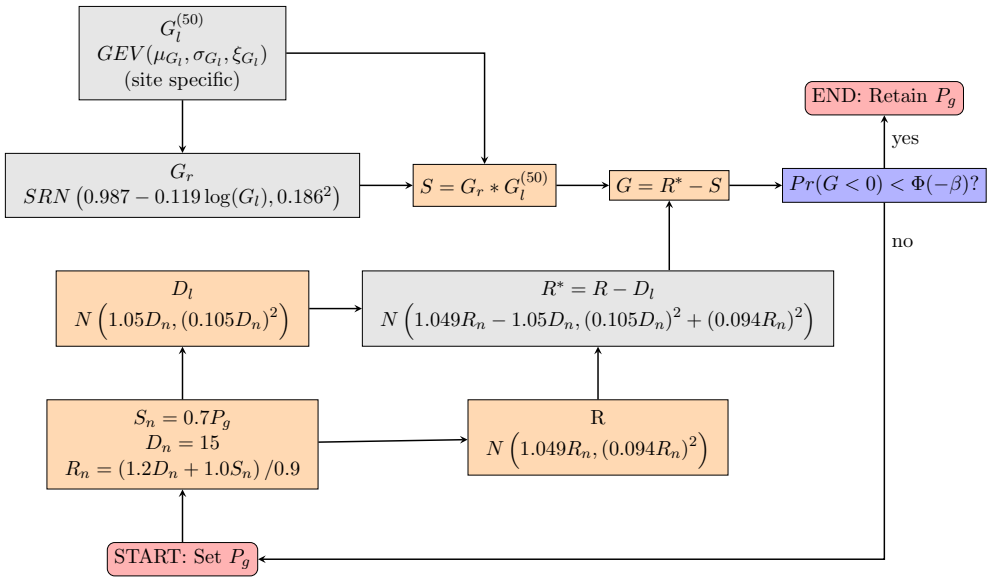

Figure 2.10: Flowchart summarizing the RTL estimation process. Grey squares indicate the distributions that are directly simulated from as part of the Monte-Carlo analysis. Orange squares indicate calculations. Distributions include generalized extreme value (GEV), Normal (N) and Square-Root Normal (SRN).

Bibliography

Bartlett, F. M., Dexter, R. J., Graeser, M. D., Jelinek, J. J., Schmidt, B. J., and Galambos, T. V. (2003). Updating standard shape material properties database for design and reliability. Engineering Journal-American Institute of Steel Construction Inc, 40(1):2–14.

Bean, B. (2019). Interval-Valued Kriging Models with Applications in Design Ground Snow Load Prediction. PhD thesis, Utah State University, Logan, UT. https://doi.org/10.26076/c805-1951.

Bennett, R. M. (1988). Snow load factors for lrfd. Journal of Structural Engineering, 114(10):2371–2383.

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

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.

Ellingwood, B., Galambos, T. V., MacGregor, J. G., and Cornell, C. A. (1980). Development of a probability based load criterion for American National Standard A58: Building code requirements for minimum design loads in buildings and other structures, volume 13. US Department of Commerce, National Bureau of Standards.

Ellingwood, B., MacGregor, J. G., Galambos, T. V., and Cornell, C. A. (1982). Probability based load criteria: load factors and load combinations. Journal of the Structural Division, 108(5):978–997.

Ellingwood, B. and O'Rourke, M. (1985). Probabilistic models of snow loads on structures. Structural safety, 2(4):291–299.

Ellingwood, B. and Redfield, R. (1983). Ground snow loads for structural design. Journal of Structural Engineering, 109(4):950–964.

Galambos, T. V. (2006). Reliability of the member stability criteria in the 2005 aisc specification. Engineering journal, 43(4):257.

Galambos, T. V., Ellingwood, B., MacGregor, J. G., and Cornell, C. A. (1982). Probability based load criteria: Assessment of current design practice. Journal of the Structural Division, 108(5):959–977.

Galambos, T. V. and Ravindra, M. K. (1978). Properties of steel for use in lrfd. Journal of the Structural Division, 104(9):1459–1468.

Hasofer, A. M. and Lind, N. C. (1974). An exact and invariant first order reliability format. Journal of Engineering Mechanics, 100(1):111–121.

Jaquess, T. K. and Frank, K. H. (1999). Characterization of the material properties of rolled sections. SAC Joint Venture.

Lee, K. H. and Rosowsky, D. V. (2005). Site-specific snow load models and hazard curves for probabilistic design. Natural Hazards Review, 6(3):109–120.

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.

Lind, N. C. (1977). Rationalizations of sections properties tables. Journal of the Structural Division, 103(3):649–662.

O'Rourke, M., Koch, P., and Redfield, R. (1983). Analysis of roof snow load case studies. Technical report, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire 03755. CRREL Report 83-1.

O'Rourke, M. J. and Stiefel, U. (1983). Roof snow loads for structural design. Journal of Structural Engineering, 109(7):1527–1537.

Stidd, C. K. (1970). The nth root normal distribution of precipitation. Water Resources Research, 6(4):1095–1103.

Thiis, T. K. and O'Rourke, M. (2015). Model for snow loading on gable roofs. Journal of Structural Engineering, 141(12):04015051.

Tobiasson, W. and Redfield, R. (1980). Snow loads for the United States, parts i and ii. Technical report, Cold Regions Research and Engineering Laboratory.

Turkstra, C. and Putcha, C. (1985). Safety index analysis for problems with large variances. Structural Safety and Reliability.

White, D. W. and Barker, M. G. (2008). Shear resistance of transversely stiffened steel i-girders. Journal of Structural Engineering, 134(9):1425–1436.

White, D. W. and Duk Kim, Y. (2008). Unified flexural resistance equations for stability design of steel i-section members: Moment gradient tests. Journal of Structural Engineering, 134(9):1471–1486.

White, D. W. and Jung, S.-K. (2008). Unified flexural resistance equations for stability design of steel i-section members: Uniform bending tests. Journal of Structural Engineering, 134(9):1450–1470.

Chapter 3
Converting Ground Loads to Roof Loads

The core element of the reliability-analysis described in Chapter 2 is the simulated 50-year roof snow loads. Roof loads are almost always inferred from the ground snow load due to the general lack of direct roof load measurements. The ratio between the annual maximum ground snow load and roof snow load, referred to as GR, has the potential to dominate the proposed reliability analysis given the high variability of GR due to roof geometry, heat loss, and exposure conditions. This chapter describes efforts to create GR models compatible with the reliability-target scenario, using ground and roof load measurements from a decade of snow surveys on a variety of structures across Canada [Allen, 1956, 1958, Allen and Peter, 1963, Faucher, 1967, Hebert and Peter, 1963, Ho and Lutes, 1968, Kennedy and Lutes, 1968, Pernica and Peter, 1966, Scott and Peter, 1961, Watt and Thorburn, 1960]. The models include a ground snow load dependency that assumes that GR tends to decrease as the ground snow load increases. The model behavior accounts for the expected loss of snow on roofs due to wind, sublimation, heat loss, etc., which tends to be greater in high snow load regions where persistent snow is subject to longer periods of exposure.

This chapter reviews previously existing GR models [Ellingwood et al.,

1980, O'Rourke et al., 1983, DeBock et al., 2016]. Included also is a comparison of the reliability-targeted loads (RTLs) that result at locations across the United States using each GR model. The recommended model includes the ground snow load dependency observed in Thiis and O'Rourke [2015] and DeBock et al. [2016], while leveraging the detailed metadata in the Canadian snow surveys to create a subset of data most relevant to the target scenario.

Chapter Highlights:

- A summary of datasets that have been used to create GR models.
- A summary of previous GR models that have been used in reliability-analyses.
- A description of a newly proposed ground snow load dependent GR model based on flat roof GR measurements taken from a decade of Canadian snow surveys.
- A comparison of the differences in RTLs that result from the use of different GR models.

3.1. Available Datasets

The difficulty of obtaining simultaneous measurements of ground and roof loads likely explains the relative lack of available GR data, especially recent GR data. The authors contacted some of the authors involved in the ongoing update of the Eurocode [Croce et al., 2019], who were willing to share some recent GR data but not in a compatible format. The general lack of available measurements is exacerbated by varied roof geometries, exposure, slope, and

thermal conditions among the measurements that further reduce the number of observations relevant to the target scenario, which is a heated flat roof under normal exposure conditions. After a thorough review of the GR literature, there appears to be three major datasets that are available for GR model development: one Norwegian, one American, and one Canadian (citations provided in following subsections). Table 3.1 shows the number of observations along with varying subsets based on roof slope (θ).

Table 3.1: Number of observations in each available GR dataset at varying slopes θ.

Dataset	Sample Size		
	All	$\theta \leq 30$	$\theta \leq 15$
Norwegian	991	430	n/a
American	230	203	140
Canadian	477	434	337

The model proposed at the end of this chapter only considered roofs with slopes less than 15 degrees to reduce the chance of underestimating GR due to snow sliding. Similar logic could be used to justify such sub-setting based on the thermal (C_t) and exposure (C_e) properties of the roof. However, no further subsets of the data were considered for a variety of reasons:

- The definitions of thermal and exposure classes were not constant across datasets. For example, the Canadian data only includes two classes for exposure instead of three, as is the case with the American data. Establishing equivalency among thermal categories in the American and Canadian datasets was similarly unclear. Additionally, exposure information was not available in the Norwegian dataset.
- Any differences in GR measurements due to thermal properties were domi-

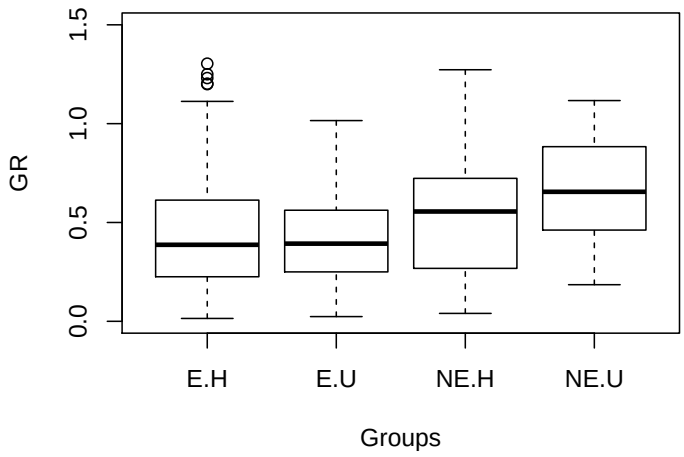

Figure 3.1: Comparisons of four combinations of scenarios considering Heated (H) vs. Unheated (U), as well as Exposed (E) vs. Not Exposed (NE) roofs in the American and Canadian datasets.

nated by other site-specific measurement factors. Figure 3.1 shows that the differences in GR measurements taken from heated and unheated buildings were small relative to the variability due to other factors among flat roof observations in the combined American and Canadian datasets.

- More importantly, there was no significant decrease in the variability of GR measurements when focusing on a single Thermal/Exposure scenario.
- Most importantly, subsets based on thermal or exposure resulted in too small of sample sizes to reliably estimate GR models. A mere 51 American observations and 23 Canadian observations exactly matched the reliability-target scenario of a heated flat roof with normal exposure.

Figure 3.2 shows a scatterplot of GR vs. ground snow load for all observations taken on roofs with $\theta \leq 30$. The points show a slight decreasing trend

in GR as ground snow load increases, though substantial variability remains both within and between datasets. Figure 3.3 shows smoothed histograms of GR measurements in each dataset, which shows that the Norwegian measurements have higher values with less variability than the American and Canadian measurements. The following subsections provide details for each of the candidate datasets.

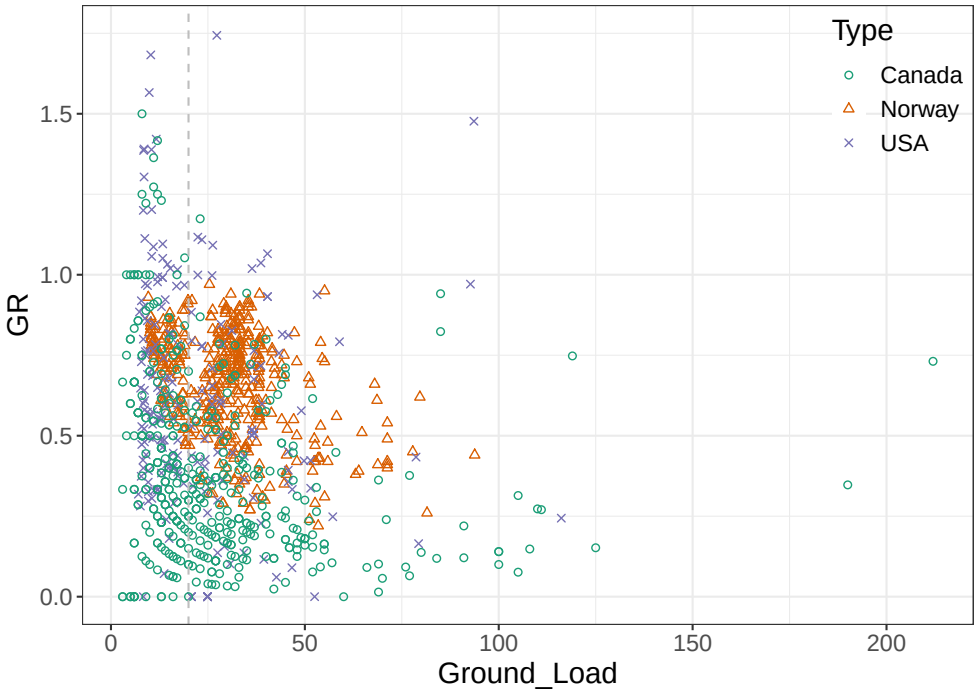

Figure 3.2: Scatterplot of GR vs. ground snow load for the Norwegian, American, and Canadian datasets.

3.1.1. Norwegian Dataset

Collection of the Norwegian dataset is described in Høibø [1988] and Høibø [1989] and subsequently analyzed in Thiis and O'Rourke [2015]. These data

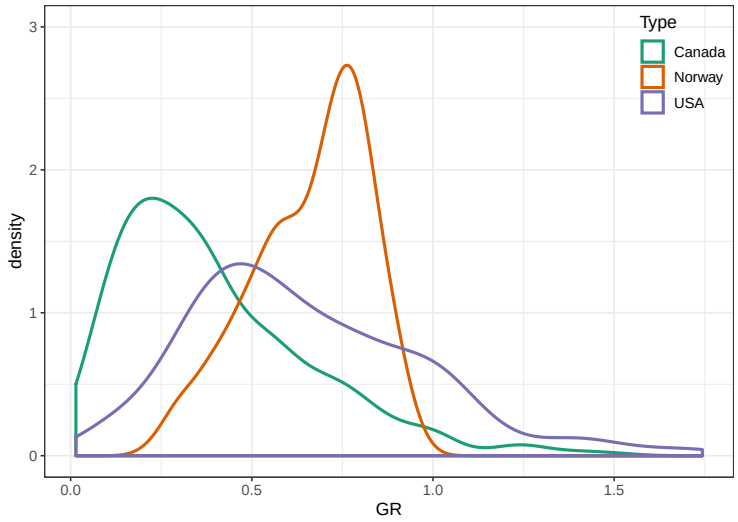

Figure 3.3: Smoothed histograms of GR measurements in the Norwegian, American, and Canadian datasets.

are not publicly available, though author Thomas Thiis graciously provided measurements for use in this national snow load study. These data, which includes unheated gabled roofs with slopes from 0 to 45 degrees, were subsequently used to develop the ground snow load dependent GR models described in Liel et al. [2017].

One disadvantage of these data is the lack of meta-data regarding roof type, exposure, or geolocation. This makes it difficult to determine the similarity in climate conditions between Norway and the United States. Additionally, the Norwegian measurements were "intended to give the 'μ-factor,' given as the ratio of the roof load to ground when the snow load on the roof was at its highest during the winter" [Høibø, 1988]. This is different from "GR" as measured in O'Rourke et al. [1983] and subsequently used in ASCE 7, which is the ratio between the max ground load and the max roof load, which may not occur at the same time during the snow season. This difference likely explains

why the Norwegian measurements tend to be higher than the North American GR measurements.

3.1.2. American Dataset

The American dataset collected by O'Rourke et al. [1983] and further analyzed in O'Rourke and Stiefel [1983] forms the foundation of many roof related provisions in ASCE 7. Measurements were taken 2-4 times during the snow season in an attempt to capture the maximum ground and roof snow loads. Proposed GR models resulting from these data only considered GR measurements for which the associated ground snow load was greater than 20 psf. Measurements were taken at structures in Idaho, Colorado, South Dakota, Oregon, and New York in an attempt to obtain a representative dataset for the country. Figure 3.4 shows that increases in the number of visits during the snow season is associated with a decrease in median GR measurement across locations. This reinforces the point that maximum ground and roof snow loads often occur at different times of the year.

3.1.3. Canadian Dataset

The Canadian dataset is a compilation of a decade of snow surveys used in the development of Canadian design snow load provisions [Allen, 1956, 1958, Allen and Peter, 1963, Faucher, 1967, Hebert and Peter, 1963, Ho and Lutes, 1968, Kennedy and Lutes, 1968, Pernica and Peter, 1966, Scott and Peter, 1961, Watt and Thorburn, 1960]. The surveys included four Tiers of measurement stations including:

- "A" Buildings: Detailed descriptions are provided of roof geometry and site-specific snow conditions. Measurements of ground and roof snow loads are

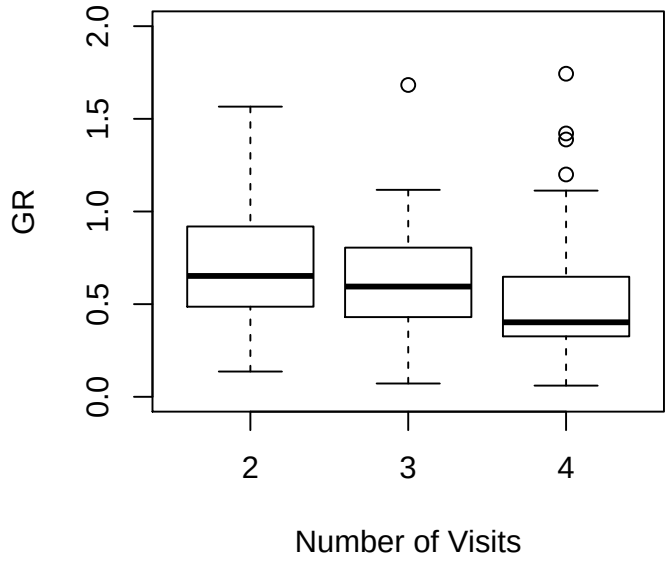

Figure 3.4: Boxplots of GR measurements based on the number of measurements made during the snow season. Typical (i.e. median) measurements tend to decrease as the number of measurements increase.

taken weekly throughout the course of the snow season, usually over the course of several years. Measurements are taken at several locations on the roof and averaged to obtain a roof snow load. These buildings are regarded as the best available data in the snow survey.

- "C" Buildings: These are similar to A stations in terms of measurement procedures and quality. These buildings are all large, flat roofs located on military bases throughout the country.
- "B" Buildings: These are measurements taken by volunteers due to anomalous circumstances such as a building failure. These data are not subject to

the same quality standards as A and C stations.

- "D" Buildings: These measurements are from surveys of snow loads on roofs in residential neighborhoods after particularly large snow storms. Measurements seem to be taken only once during the snow season for these locations, rather than weekly.

The authors decided to only use measurements at "A" and "C" structures due to the frequency and consistency of measurements throughout each snow season. Additionally, one unusual roof geometry observed amongst the "C" structures was removed where measurements were being taken on a flat roof that was adjacent to an arch hangar. Sliding snow from the arch hangar would consistently result in GR measurements on the flat roof portion well above one. This unusual situation, perhaps only seen on military bases, did not seem representative of the target scenario of interest. Additionally, one observation was removed for having a GR measurement greater than 2, which was deemed unrealistic for the target scenario.

The climate conditions of the Canadian GR measurement locations were compared to locations from the American dataset to ensure that the Canadian data were representative. This determination was made considering approximate climate metrics obtained from gridded climate maps from the climateNA project [Wang et al., 2016] by geolocating measurements based on city names from both the American and Canadian datsets. Figure 3.5 shows the mean annual temperatures of the coldest month as plotted against the average winter (December-February) precipitation. The Canadian data is fairly representative across both metrics, but the three annotated locations were removed for having significantly colder temperatures than those observed at the American locations. The sample sizes provided in Table 3.1 reflect only the measurements

retained for analysis.

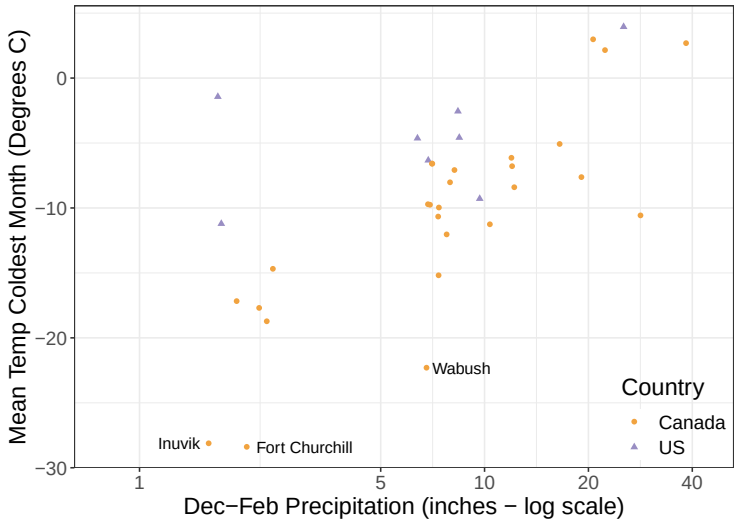

Figure 3.5: Plots of mean annual temperature of the coldest month vs. winter precipitation from measurement locations in the American and Canadian datasets.

3.2. Previous Methods

Ellingwood's original partial safety factor calibrations were conducted before O'Rourke et al. [1983] and without access to the Canadian surveys. This in mind, through collaborations with Wayne Tobiasson of CRREL, GR was assumed to follow a normal distribution with a mean (μ_{gr}) of 0.5 and a standard deviation (σ_{gr}) of 0.115. Subsequent data analysis by O'Rourke et al. [1983], from which the American GR measurements for this manuscript were obtained, proposed the new GR model

$$\mu_{gr} = 0.47 * E * T$$

where E and T represent exposure and thermal factors respectively. Under the target reliability scenario, $E = T = 1$. The residuals were assumed to follow a lognormal distribution with $\mu_r^* = 0$ and $\sigma_r^* = 0.42$.

Colorado's recent pursuit of RTLs made use of all the Norwegian GR data described in Thiis and O'Rourke [2015], which includes gable roofs with slopes from 0 to 45 degrees. This model assumed that GR followed a lognormal distribution, with log-scale parameters μ_{gr}^* and σ_{gr}^*, dependent on ground snow (p_g) and calculated as

$$\mu_{gr}^* = \log\left(0.5 \times \exp\left(-0.034 * p_g\right) + 0.4\right)$$
$$\sigma_{gr}^* = \min\left(.007 * p_g + 0.1, 0.33\right).$$

For low snow loads, the resulting probability distribution can lead to simulated GR values much larger than 1. To control for this, simulated GR values were capped to never exceed 1.2 in the reliability analysis [Liel et al., 2017].

Figure 3.6 compares the shape of the resulting GR distributions for each method, including the Colorado GR distribution at ground snow loads of 10, 30, and 60 psf. Note the significant shift in the mean of the GR distribution as ground snow load increases for the Colorado method. Also note that the O'Rourke et al. [1983] model has more variability than the other considered methods.

3.3. Proposed Model

Originally, efforts were made to combine the three datasets to create a new GR model. However, the different GR distributions in each dataset caused

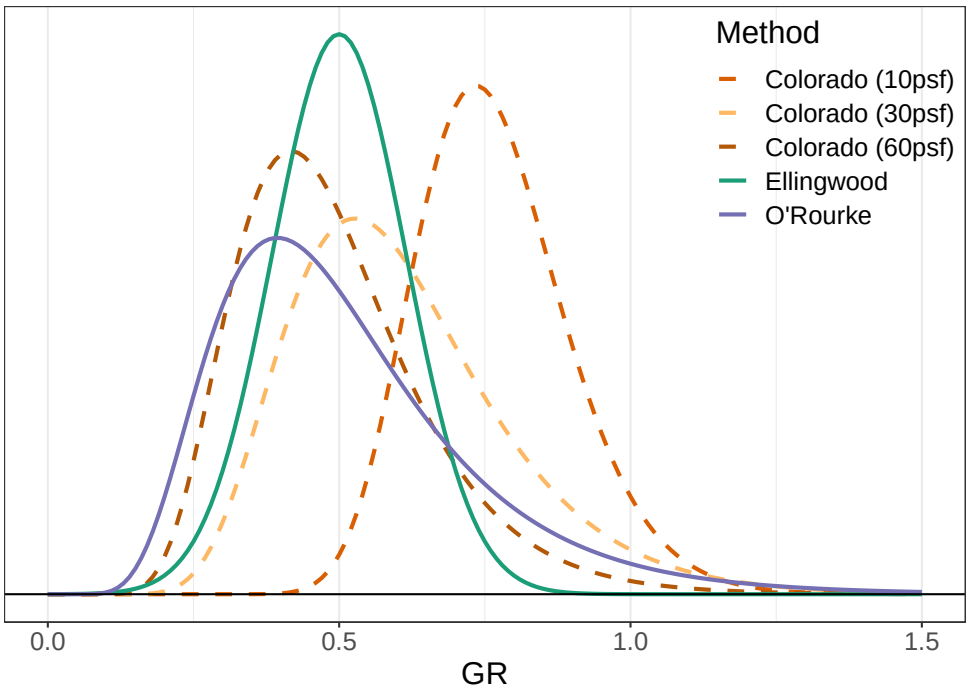

Figure 3.6: Comparison of the shape of the assumed GR distribution from Ellingwood et al. [1980], O'Rourke et al. [1983], and Liel et al. [2017]. Loads in parenthesis indicate the ground snow load associated with the GR distribution.

the consolidated models to inherit the high average values of the Norwegian data, as well as the high variability of the Canadian and American data. The resulting models exhibited higher RTLs than would be obtained using existing models and did not seem reasonable for use. The newly proposed model instead uses only Canadian observations at "A" and "C" buildings with roof slopes less than 15 degrees. This model incorporates the ground snow load dependency of the Colorado GR curve, while avoiding the use of the Norwegian μ-factor measurements which are known to overestimate GR.

Figure 3.2 illustrates that GR measurements tend to decrease as ground snow load decreases. This relationship can be modelled linearly with appro-

priate variable transformations. The ground snow load dependency in GR is assumed to be of the form:

$$E\left[\sqrt{G_r}\right] = \beta_0 + \beta_1 \log(p_g).$$

This dependency is estimated via least squared regression to obtain:

$$E\left[\sqrt{G_r}\right] = 0.99 - 0.12 \log(p_g). \tag{3.1}$$

Because there are few GR measurements with associated ground snow load values above 50 psf, $E\left[\sqrt{G_r}\right]$ were capped below at $0.99 - 0.12 * 50 = 0.27$ to avoid inappropriate extrapolations of the ground snow load dependent trend. Figure 3.7 visualizes the trend line (with the 50 psf ground snow load cap) as compared to a local polynomial regression model on the transformed scale. The agreement between the proposed model and the local regression model verifies that the $\sqrt{G_r}$ and $\log(p_g)$ are linearly related, with $\log(p_g)$ explaining about 17% of the variability of $\sqrt{G_r}$. This reduction in the variance of $\sqrt{G_r}$ serves to reduce RTL estimates as compared to those obtained using a GR model developed with the same data that assumes no ground snow load dependency.

Figure 3.8 shows that the residuals of this regression model follow a normal distribution (except perhaps at the extreme endpoints) and are centered around zero. The variance of the residuals is estimated to be $\hat{\sigma_{G_r}} = 0.19$. Values of the square root of GR are simulated from a normal distribution, and then squared to return to the original scale of GR.

Figure 3.9 shows the back-transformed estimates of (3.1). The dashed lines represent the thresholds for which 95% of the simulated GR values (after back-transforming) are expected to fall. These simulated values are occasionally

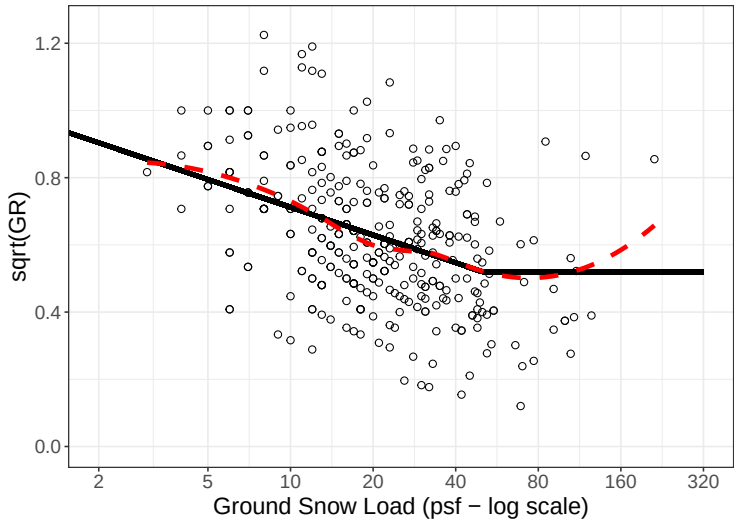

Figure 3.7: Comparison of the proposed regression model with the trend flat-lined after 50 psf (black) to a local-polynomial regression (red) model.

above 1.0 when the ground snow load is small. DeBock et al. [2016] had a similar issue with their simulated values which they resolved by capping simulated values at 1.2. It was decided in consultation with the steering committee associated with this national study that simulated values should be capped at 1.0 since the maximum roof load is not expected to exceed the maximum ground snow load on a heated roof under the uniform loading scenario. Simulated GR values are also necessarily capped below at 0. Table 3.2 shows the expected percentage of simulated values that require a 0.0 or 1.0 GR cap for various ground snow loads.

3.4. Implications

Figure 3.10 compares the assumed GR distributions under these new models to those obtained in DeBock et al. [2016]. As expected, the new models have

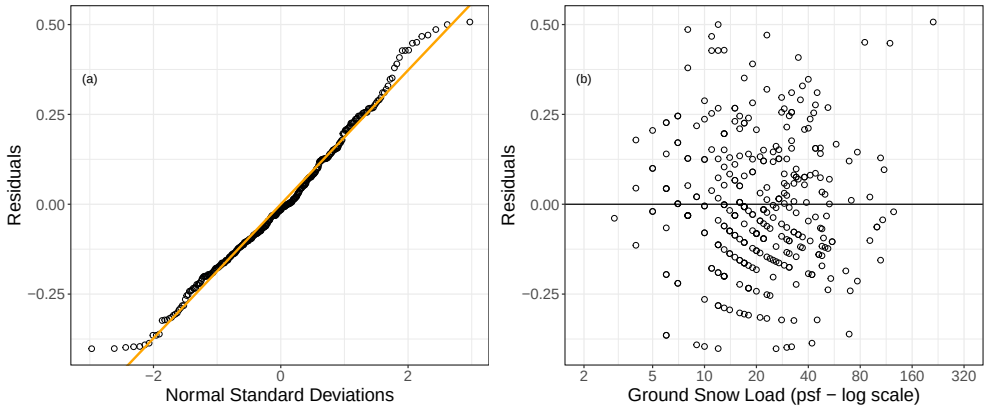

Figure 3.8: (a) Shows that the residuals of model (3.1) are normally distributed. (b) Scatterplot of residuals vs. ground snow load to illustrate that the residuals are unbiased with constant variance until at least 50 psf ground snow load.

Table 3.2: Percentage of simulated G_r values that are capped at the 1.0 or 0.0 threshold for various ground snow loads. Recall that the G_r distributions are identical after 50 psf.

Ground Load (psf)	% Capped at 1.0	% Capped at 0.0
10	6.1	< 0.1
20	2.3	< 0.1
30	1.2	0.1
40	0.8	0.2
50	0.5	0.3

lower averages than the Colorado model, but also have more variability. The larger variance of the newly proposed model reduces the expected reduction in loads due to the smaller average measurements.

Figure 3.11 compares the estimated RTLs using different GR models at the 81 non-Alaska locations considered in Lee and Rosowsky [2005]. Note that larger values of the distribution shape in the left plot indicate a heavier distribution tail for the ground snow load distribution. The results show that the new model usually estimates lower RTLs than would be obtained using previ-

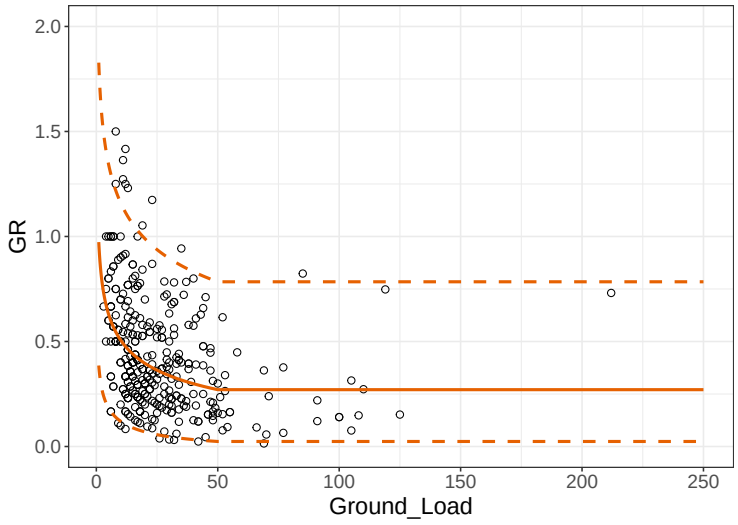

Figure 3.9: Back-transformed estimates of the average of $\sqrt{G_r}$. Dashed lines represent the range for which 95% of simulated GR values fall for a given ground snow load.

ous GR models, and substantially lower RTLs than the O'Rourke and Stiefel [1983] GR model. The "tempering" effect that the ground snow load dependency in the new GR model has on heavy-tailed ground snow load distributions is highlighted by the increasing ratios with increases in the distribution shape. The reduction in RTLs as compared to using the Colorado GR model illustrates the effect of the high GR bias in the Norwegian measurements. The final RTL values are comparable to what would have been obtained using the Ellingwood et al. [1980] GR model, though the ratio is dependent upon distribution shape.

The reduction in RTLs that occur with the use of the new ground snow load dependent GR model is likely due to the following:

1. The frequency of measurements in the Canadian dataset make it the most likely dataset to capture the true GR value each snow season.
2. Accounting for the ground snow load dependency observed in each dataset

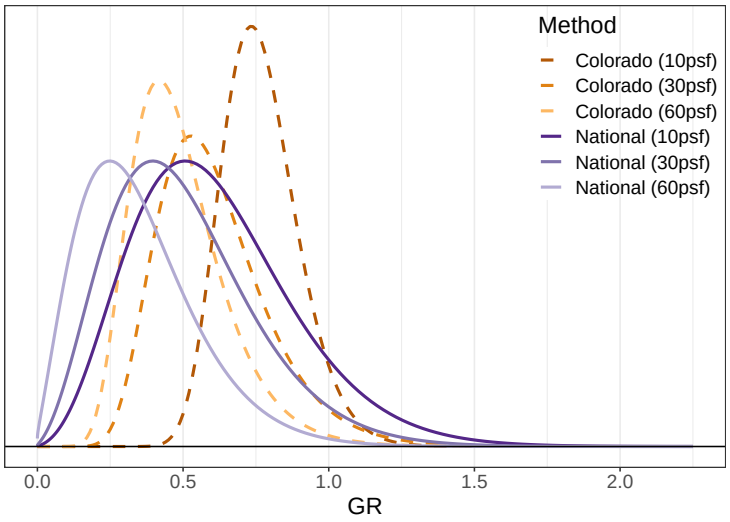

Figure 3.10: Comparison of distributions for the Colorado and National Study GR models for various ground snow loads.

reduces the variability in GR as relative to equivalent "flat-line" models.

3. Capping simulated values at one (which reduces loads) reflects expected conditions for the target design scenario.

Chapter 6 explains the development of the ground snow load distribution models used in these GR model comparisons.

Bibliography

Allen, C. and Peter, B. (1963). Snow loads on roofs 1962-63: Seventh progress report. Technical report, National Research Council, Division of Building Research.

Allen, D. E. (1956). Snow loads on roofs. the present requirements and a proposal for a survey of snow loads on roofs. Technical report, National

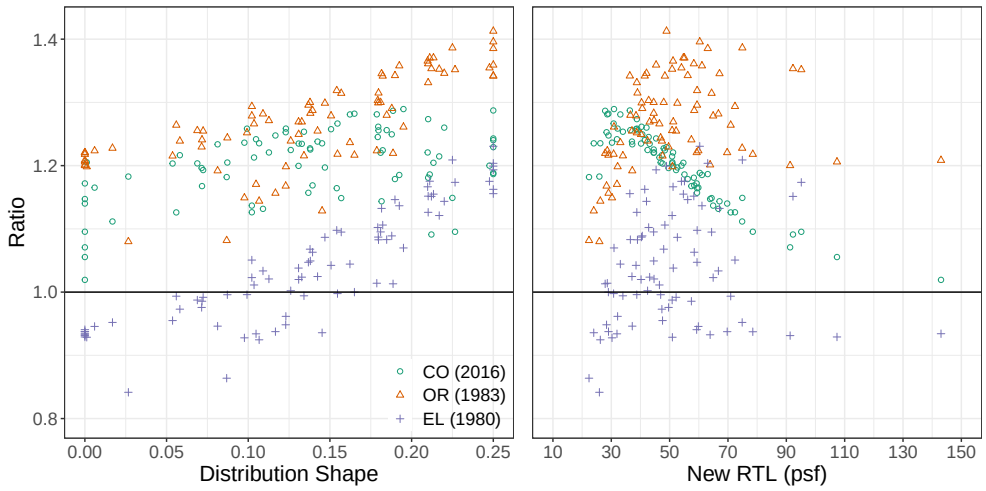

Figure 3.11: Comparisons of the ratio between estimated RTLs using the Ellingwood et al. [1980] (EL), O'Rourke and Stiefel [1983] (OR), and Liel et al. [2017] (CO) GR models and the newly proposed ground snow load dependent GR models based on Canadian data. Values above one indicate instances where previous methods would predict higher RTLs than the newly developed GR model.

Research Council, Division of Building Research.

Allen, D. E. (1958). Snow loads on roofs 1956-57: a progress report. Technical report, National Research Council, Division of Building Research.

Croce, P., Formichi, P., Landi, F., and Marsili, F. (2019). Harmonized european ground snow load map: Analysis and comparison of national provisions. Cold Regions Science and Technology, 168:102875.

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

Ellingwood, B., Galambos, T. V., MacGregor, J. G., and Cornell, C. A. (1980). Development of a probability based load criterion for American National

Standard A58: Building code requirements for minimum design loads in buildings and other structures, volume 13. US Department of Commerce, National Bureau of Standards.

Faucher, Y. (1967). Snow Loads on Roofs 1964-65: Ninth Progress Report. National Research Council of Canada, Division of Building Research.

Hebert, P. and Peter, B. (1963). Snow loads on roofs 1961-62: Sixth progress report with an appendex on roof to ground load ratios. Technical report, National Research Council, Division of Building Research.

Ho, M. and Lutes, D. A. (1968). Snow Loads on Roofs 1965-66: Tenth Progress Report. National Research Council of Canada, Division of Building Research.

Høibø, H. (1988). Snow load on gable roofs-results from snow load measurements on farm buildings in norway. In Proceedings of the First International Conference on Snow Engineering, pages 89–6.

Høibø, H. (1989). Form factors for snow load on gable roofs: Extending use of snow load data from inland districts to wind exposed areas. In Proceedings of the 11th International Congress on Agricultural Engineering, Dublin, Ireland, pages 4–8.

Kennedy, I. and Lutes, D. (1968). Snow Loads on Roofs 1966-67: Eleventh Progress Report.

Lee, K. H. and Rosowsky, D. V. (2005). Site-specific snow load models and hazard curves for probabilistic design. Natural Hazards Review, 6(3):109–120.

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.

O'Rourke, M., Koch, P., and Redfield, R. (1983). Analysis of roof snow load case studies. Technical report, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire 03755. CRREL Report 83-1.

O'Rourke, M. J. and Stiefel, U. (1983). Roof snow loads for structural design. Journal of Structural Engineering, 109(7):1527–1537.

Pernica, G. and Peter, B. (1966). Snow Loads on Roofs 1963-64: Eighth Progress Report. National Research Council Canada, Division of Building Research.

Scott, J. and Peter, B. (1961). Snow loads on roofs 1960-61: Fifth progress report. Technical report, National Research Council, Division of Building Research.

Thiis, T. K. and O'Rourke, M. (2015). Model for snow loading on gable roofs. Journal of Structural Engineering, 141(12):04015051.

Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for north america. PLOS ONE, 11(6):1–17.

Watt, W. and Thorburn, H. J. (1960). Snow loads on roofs 1959-60: Fourth progress report. Technical report, National Research Council, Division of Building Research.

Chapter 4
Data Processing

The time intensive nature of data collection and cleaning makes it difficult to quickly update design snow load estimates as new information becomes available. These difficulties are partially overcome by improvements in the quality and accessibility of snow measurements at the national level. Despite these improvements, significant challenges in data quality remain. Proper methods and strategies for handling misreported values are particularly important given this project's focus on extreme events, which are particularly sensitive to misreported outlier values. Estimates of extreme events are likewise sensitive to pseudo maximums which are caused by inconsistent coverage of the snow season. This chapter describes a systematic procedure to screen daily observations of SNWD and snow load for misreported values and detect unreasonably low pseudo maximums due to lack of coverage. The iterative outlier detection schemes described in this chapter err on the side of caution by retaining observations when an observation is only suspected to be misreported. The distribution fitting approach described in Chapter 6 is designed to tolerate the inevitable outlier observations that remain in the record. The key advantage to the data cleaning approaches described in this chapter is that they can be quickly implemented and easily updated, with the exception of the manual outlier verification. This allows the final project results to be updated in future

years as improved information becomes available with little marginal cost.

Chapter Highlights:

- A description of the data sources used to define the reliability-targeted loads (RTLs).
- Descriptions and examples of recurring outlier issues that were identified and removed from the dataset.
- A summary of a series of data screens that were used to identify candidate stations with sufficient information to estimate site-specific RTLs.
- An explanation of the process used to merge snow records at geographically close stations.
- An outline of the observation preference hierarchy when multiple measures of a snow load are provided for the same season.
- Maps of the locations of the qualifying measurement locations as well as the definition of a three-tier system for describing the reliability of the station measurements.

4.1. Data Summary

The core dataset for this project was the Global Historical Climatological Network Daily Dataset (GHCND) Menne et al. [2012]. The GHCND includes observations from the following station networks:

- National Weather Service (NWS) first-order stations (FOS). These stations are typically located at airports and are regarded as the most reliable measurements in the GHCND.

- The Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) stations. SNOTEL stations are primarily located in the intermountain west and began replacing or supplementing the once-monthly snow course measurements in the late 1970s.
- NWS Cooperative Observer Network (COOP) stations. Measurements at these stations are taken by volunteers in collaboration with the NWS. These measurements are subject to less quality control measures than FOSs and often measure only snow depth.
- Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) measurements. Like COOP stations, these measurements are also taken by volunteers. They are subject to less quality control measures than FOSs but do occasionally contain direct measurements of snow load.

GHCND measurements are freely available for mass download (https://www.ncdc.noaa.gov/ghcnd-data-access). The variables of interest are snow depth (SNWD) and water equivalent of snow on the ground (WESD). Measurements from additional station networks were used to develop the depth-to-load conversion models. Additional details about those supplemental networks are provided in Chapter 5.

The original data download included more than 237 million observations at more than 65,000 weather stations in the United States and Canada. Observations extend as far back as 1857, though the vast majority of measurements are taken post 1948.

4.2. Outlier Detection

The distribution fitting process described in Chapter 6 relies on the annual snow maximums, with snow seasons extending from October of the previous year to June of the listed year. The focus on seasonal maximums makes the distribution fitting process particularly sensitive to abnormally high and misreported measurements in the period of record. This chapter describes efforts to remove the most grievous misreported values.

Every observation in the GHCND data set has a quality flag (QFLAG) which indicates whether the observation has failed any of a series of automatic and manual outlier checks [Durre et al., 2010]. All observations flagged by the GHCND for quality control were removed prior to analysis. While the removal of these flagged observations greatly improved the quality of the dataset, many misreported observations remained.

Additional automatic checks of observations implemented by the authors proved insufficient for removing the persistent misreported observations. However, manual checks of all 65,000 candidate stations was also not feasible given time and funding constraints. In light of these constraints, a hybrid approach was adopted where stations would be flagged for potential issues using a series of automatic checks, then manually checked for outliers by the authors. The iterative process is as follows:

1. Fit distributions at candidate measurement locations with no outlier points removed.
2. Identify stations for manual inspection at locations with anomalous distribution parameter estimates.
3. Visually inspect SNWD and WESD measurements at the flagged stations for outlier values.

4. Remove observations only if the measurements are "obviously" misreported,
5. Refit distributions with the anomalous points removed.

If there was any doubt as to whether the value in question was legitimate, the observation was left in the dataset.

Stations were flagged for manual inspection if they met any of the following problematic conditions:

- Stations where an observation exceeded verified state-level snow depth records, or the difference in sequential observations exceeded county-level snowfall records [SCEC, 2020].
- Stations with unusually heavy distribution tails (as fit in Chapter 6).
- Stations with distribution shapes significantly different than neighboring observations.
- Stations where the majority of seasonal maximums were zero, yet the station had an observed snow load above 20 psf.

Occasionally, all stations within a region would be checked for misreported observations if the estimated design loads were higher than expected. This was the case in coastal Washington and Oregon as well as the eastern slopes of the Rocky Mountains in Colorado. In many cases, few to none of the observations were removed. In other cases, entire snow years were found to be incorrect and were removed. Some of the recurring outlier issues that were discovered during the manual checks included:

- SNWD measurements that were incompatible with corresponding WESD measurements (Figure 4.1).
- Incorrectly reported units of measurements (usually a factor of 10) for a

portion of the period of record (Figure 4.2).

- Consecutive WESD measurements at an impossibly high value during months when little to no snow is expected (Figure 4.3).
- Single, anomalous observations in an otherwise well-behaved set of measurements (Figure 4.4).
- Consecutive years of zero-valued maximums in locations where some snow is expected every year (Figure 4.5).

The manual checks also confirmed that many of the state and county snow records are out of date. Figure 4.6 shows a Montana SNOTEL station that consistently exceeds the state-verified snow depth record, yet there is no sign of anomalous values. Values that were flagged but showed no visual evidence of being an outlier value were ultimately retained in the dataset.

The original data download considered 237 million observations at more than 65,000 stations in the United States and Southern Canada, though roughly 100 million of those observations were during summer months, missing, or misreported. Table 4.1 shows the number of observations (in millions) before and after each data cleaning step. Table 4.2 shows the percentage breakdown of outliers that were manually removed by the authors. Note that the vast majority of these roughly 650,000 manually removed observations were misreported zero-values or systematic issues in the measurement units (usually off by a factor of 10). Only about 0.2% of the values removed were isolated incidents of anomalously high values that were inconsistent with the surrounding observations. Most of the "other" outliers were also units issues, though the issues were not as pervasive as they were at the 29 weather stations with systematic measurement unit issues.

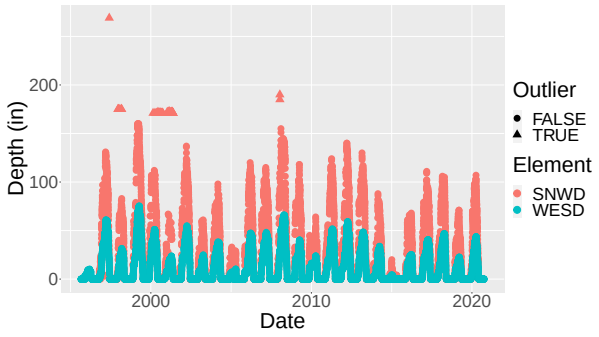

Figure 4.1: Station USS0021A32S in Washington, illustrating incompatible SNWD and WESD measurements. Triangle points indicate removed observations.

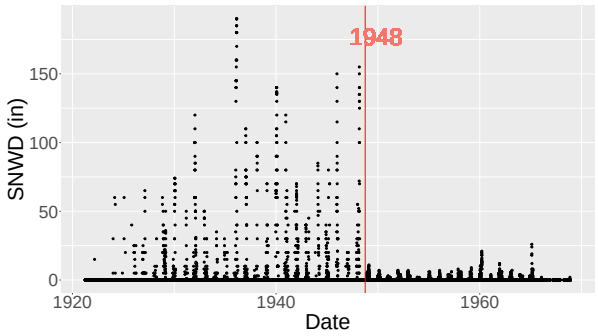

Figure 4.2: Station USC00254790 in Nebraska which illustrates a systematic measurement units issue.

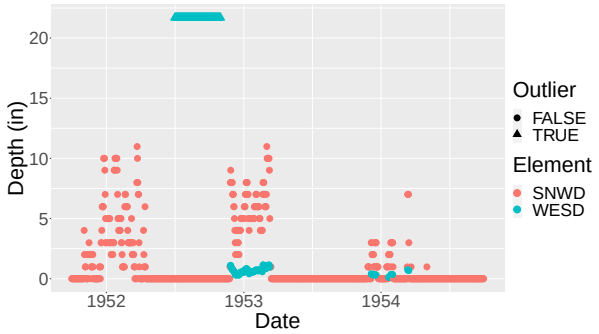

Figure 4.3: Station USW00014925 in Minnesota which demonstrates consecutive misreported measurements of WESD.

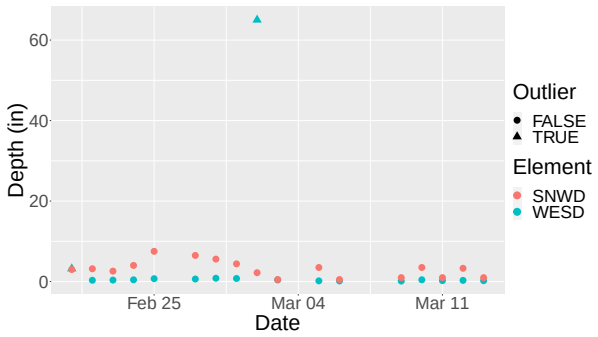

Figure 4.4: Station US1COBO0290 in Lafayette, Colorado, illustrating isolated measurements incompatible with the long term accumulation patterns.

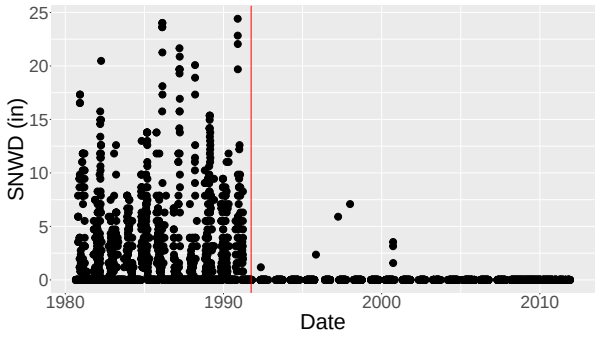

Figure 4.5: Station CA003031400 in Alberta, Canada, illustrating an impossibly long series of zero-valued snow years in a location where snow is expected every year.

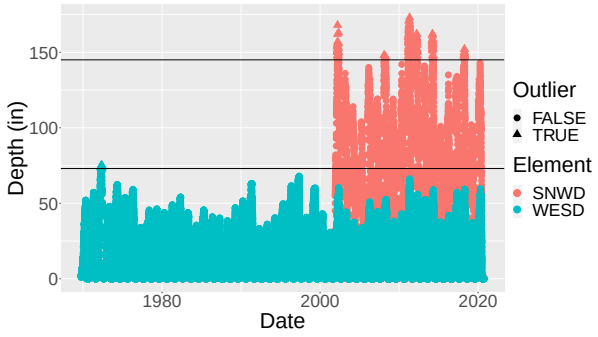

Figure 4.6: Station USS0013A19S in Montana. The top line represents the SNWD Montana maximum, and the bottom line represents the WESD Montana maximum. The flagged points, which are any point above its respective line, were not removed since they are clearly not outliers.

Table 4.1: Summary of the remaining observations after each data cleaning step.

Cleaning Steps	Observations (in millions)
Original Dataset	237.09
Remove July-September and "missing presumed zero"	138.74
Remove quality control issues identified by GHCN	138.49
Remove manually identified misreported observations	137.84

Table 4.2: Composition of the 653,000 manually removed outliers.

Type	Percentage
Misreported zero-valued observations	93.7%
Systematic unit issues ($>$ 250 points removed per station)	5.3%
Isolated outlier values ($<$ 10 points removed per station)	0.2%
Other	0.8%

4.3. Coverage Filters

The manual outlier checks described in Section 4.2 revealed some obvious issues of systematic misreported observations. However, those checks were not effective at identifying pseudo maximum values caused by a partial lack of coverage of the snow season. These pseudo maximums can be screened through coverage filters that ensure sufficient coverage of each snow season. These screens need to be strict enough to remove pseudo maximums, yet lenient enough to avoid throwing out entire years of record unnecessarily.

4.3.1. Coverage Filter Algorithm #1

The balance between strictness and leniency is achieved in this effort through the following algorithm applied after outliers were removed:

1. Where necessary, estimate snow load from snow depth (see Chapter 5 for

details).

2. Collect all seasonal snow load maximums from every available year applying no coverage filter.
3. Calculate the median seasonal maximum snow load at each station location.
4. Separate stations into "high" and "low" accumulation groups based on their median seasonal maximum. High accumulation stations are those whose median seasonal maximum is greater than 50 psf.
    - The 50 psf cutoff separates most SNOTEL and "SNOTEL-like" stations from the rest, since SNOTEL stations are known to have different accumulation patterns than typical COOP stations. Approximately 80% of considered SNOTEL stations have a median annual maximum above this threshold.
5. A seasonal maximum "passes" the coverage filter if it has at least one observation in each of the four months where a seasonal maximum is most likely to occur. This four month window depends on accumulation group as observed in Figure 4.7.
    - High accumulation stations: January-April.
    - Low accumulation stations: December-March.
6. Discard stations with less than five seasonal maximums passing the coverage filter.

January-April was selected for high accumulation stations instead of February-May to accommodate COOP stations that are less likely to have consistent snow records in May. Most May maximums occur at SNOTEL stations that are very likely to also have consistent records in January. This coverage fil-

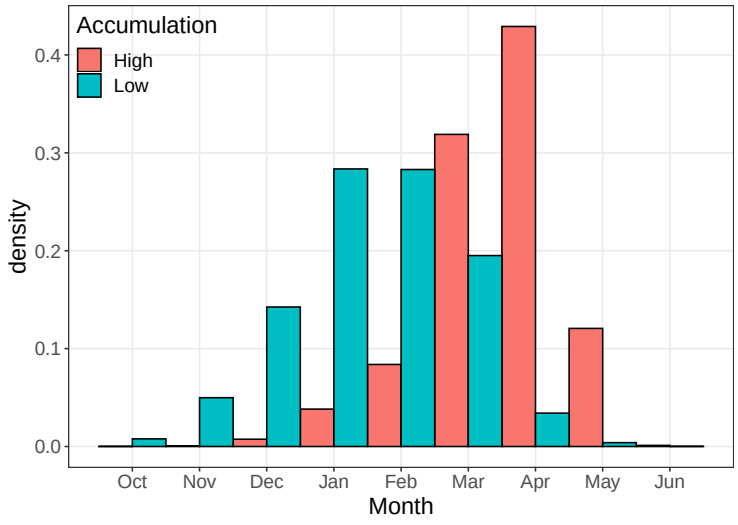

Figure 4.7: Counts (normalized by accumulation group) of the number of non-zero snow load maximums occurring in each month at stations with coverage in every month of the snow season.

ter reduces the number of candidate stations from roughly 65,000 to 20,000. All seasonal maximums, regardless of coverage filter status, are retained for these 20,000 stations. A second coverage filter is applied to the seasonal maximums after grouping geographically close stations as described in the following section.

4.4. Station Clustering

It is often the case that the geographic location of a station will change slightly during its lifetime. Occasionally, this change in location results in a new station identifier being assigned to the ensuing measurements. This creates situations where what should be a single, extended period of record is incorrectly regarded as two shorter periods of record. The distribution fitting process described

in Chapter 6 is most reliable when applied to long periods of record. The need for long periods of record encourages the combination of observations at geographically close stations. This is accomplished through a hierarchical clustering algorithm using a custom distance metric that assigns a "distance" unit d of one for every:

- 0.6 miles of geographical separation between groups
- 50 feet of elevation difference between groups

The clustering algorithm creates groups of stations for which $d \leq 4$ between the farthest neighbors in the cluster. This means that stations in a cluster are separated by no more than 2.4 miles and 200 feet in elevation. For stations separated by 100 feet in elevation, the geographical separation can be no more than 1.2 miles for stations to be combined. This clustering approach is an adaptation of the approach described in DeBock et al. [2017], yet creates smaller clusters.

The $d = 4$ cluster threshold creates roughly 18,000 "measurement locations" from the 20,000 qualifying weather stations. The clustering scheme serves to extend the period of record for a measurement location, especially when one weather station was intended to replace another. The clustering scheme also eliminates the model instability issues that occur when co-located (or nearly co-located) stations are used as input into the spatial mapping models described in Chapter 7. These advantages come at the risk of combining observations at stations whose annual maximum snow load follow different probability distributions due to differences in measurement conditions. The $d = 4$ threshold is intended to balance the advantages of clustering with the risk of losing the small-scale variability in snow loads. This balancing act results in occasional sets of "sister stations" that should be combined but are

ultimately treated as distinct locations. Any discrepancies in the fitted distributions that occur at sister stations are reconciled with the shape parameter smoothing approach described in Chapter 6.

4.5. Collecting Seasonal Maximums

The combination of weather stations into consolidated measurement locations inevitably creates situations where there are overlapping measurements of the same snow season. Even measurement locations comprised of a single weather station often have overlapping direct (WESD) and indirect (SNWD) measurements of snow load. This means there are usually multiple candidate seasonal maximums for each snow season obtained from different weather stations and measurement types. A single maximum is obtained for each snow season though the following preference hierarchy:

1. Prefer seasonal maximums obtained from measurements that pass coverage filter #1.
2. Prefer non-zero seasonal maximums.
3. Prefer seasonal maximums obtained from direct measurements of snow load (WESD) to those obtained from indirect measurements of load (SNWD).
4. All else equal, prefer the largest available seasonal maximum.

The preference hierarchy only proceeds to the next preference option if multiple candidate seasonal maximums satisfy the current preference option. Decisions 1 and 2 protect against artificial zero maximums, while Decision 3 gives preference to direct measurements of snow load.

4.5.1. Coverage Filter Algorithm #2

With preferred seasonal ground snow load maximums in hand, a second coverage filter is applied to further reduce the prevalence of pseudo maximums. For each measurement location:

1. Determine the median seasonal maximum.
2. Retain maximums that meet at least one of the following two conditions:
   - The maximum passes coverage filter check #1.
   - The maximum is above the median seasonal maximum.

The coverage filter exception for maximums above the median ensures that the largest seasonal maximums are never excluded due to lack of coverage of the snow season. At the same time, the coverage filter protects against the pseudo maximums that can wreak havoc on estimated distribution fitting parameters in high load locations.

4.6. Final Stations

The previous sections of this chapter ensure the quality of retained seasonal maximums at the measurement locations. Probability distributions are fit to these annual maximums as described in Chapter 6. Reliable estimates of probability distribution parameters rely on sufficiently large sample sizes of seasonal maximums. This is especially true for site specific RTLs, which are more sensitive to slight changes in probability distribution parameters as compared to 50-year snow loads. Minimum sample sizes for distribution fitting have historically included seven [SEAU, 1992] or ten [Theisen et al., 2004, Al Hatailah et al., 2015, Meehleis et al., 2020, Buska et al., 2020]. DeBock et al. [2016] uses

a minimum sample size of 30 in the only comparable site-specific reliability analysis available.

In light of the sample size limitations for certain portions of the country, a three tier station designation was adopted:

1. Tier 1 stations have at least 30 years of record and 15 years of non-zero seasonal maximums.
2. Tier 2 stations have at least 15 years of record and 7 years of non-zero seasonal maximums.
3. Tier 3 stations have at least 30 years of record with 20% or less of the seasonal maximums being non-zero.

Tier 2 stations are only considered in the analysis if there is not a Tier 1 station close by. Similarly, Tier 3 stations are only considered if there are no Tier 1 or 2 stations close by. "Closeness" is defined using the same clustering scheme proposed in Section 4.4 but uses a threshold of $d = 20$ instead of $d = 4$. The hierarchical nature of the clustering ensures that $d = 4$ clusters will be fully contained within $d = 20$ clusters. Note that Tier 3 stations retained in the analysis treat the $d = 20$ clusters as a single measurement location. This prevents the Tier 3 stations from being over-represented in the analysis. Figure 4.8 shows a map of the final set of stations with color denoting the tiers. Only Tier 1 Canadian stations within 60 miles of the U.S. border are retained in the analysis. Table 4.3 shows a breakdown of the final set of stations by Tier. Chapter 6 describes the distribution fitting process at these measurement locations while Chapter 7 describes how RTLs are estimated between measurement locations.

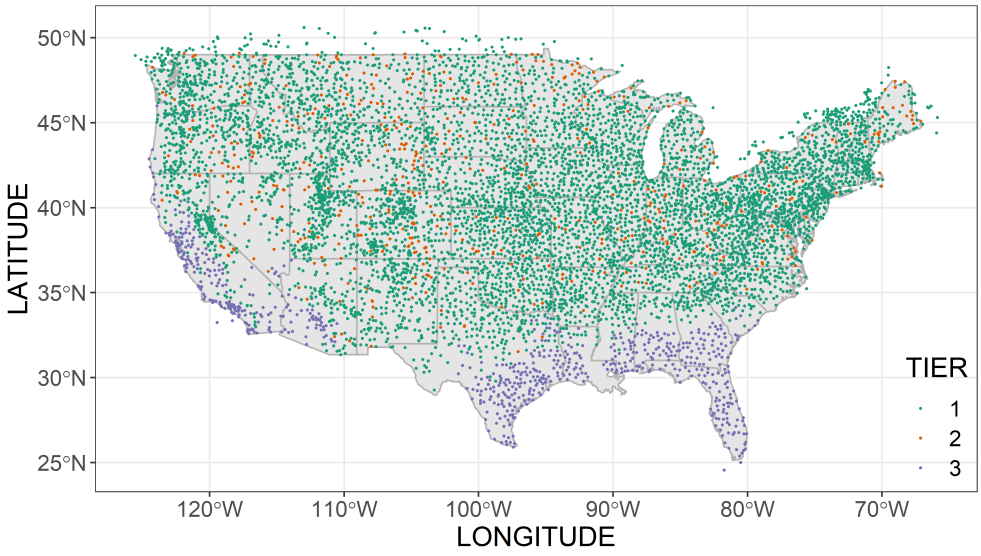

Figure 4.8: Map of Tier 1, 2, and 3 stations retained for analysis.

Table 4.3: Counts of station Tiers used for distribution fitting.

Tier	Count
1	6775
2	509
3	680

Bibliography

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. Accessed: 12-1-2020.

Buska, J. S., Greatorex, A., and Tobiasson, W. (2020). Site specific case studies for determining ground snow loads in the United States. Technical report, Engineer Research and Development Center, Hanover, NH. Accessed: 11-30-2020.

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

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.

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.

Meehleis, K., Folan, T., Hamel, S., Lang, R., and Gienko, G. (2020). Snow load calculations for alaska using ghcn data (1950–2017). Journal of Cold Regions Engineering, 34(3):04020011.

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.

SCEC (2020). State climate extremes. https://www.ncdc.noaa.gov/extremes/scec/records.

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.

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.

Chapter 5
Depth-to-Load Conversions

With the exclusion of SNOTEL stations, relatively few weather stations provide direct measurements of snow load. This requires the snow load to be estimated from snow depth. There are multiple national [Tobiasson and Greatorex, 1997], regional [Sack and Sheikh-Taheri, 1986, Sturm et al., 2010], and state-specific [Theisen et al., 2004, SEAO, 2007, DeBock et al., 2016, Meehleis et al., 2020] depth-to-load conversion models that are currently used to obtain design ground snow loads. These models characterize the relationship between the maximum (or 50-year) snow load with the maximum (or 50-year) snow depth. Each of these models effectively characterize expected snow densities for a particular region or station type, but none are equipped to characterize snow loads at a continental scale in both high and low accumulation regions.

This chapter describes efforts to develop a universal depth-to-load conversion model that accounts for differences in local climate and resolves the non-linear density relationship that occurs between low and high (usually mountainous) accumulation regions. The analysis draws inspiration from Hill et al. [2019], but the model is specifically designed to predict annual maximum snow loads, rather than daily snow loads. Additionally, this model is the first to resolve the non-linear gap between the depth/density relationships observed among high altitude Snowpack Telemetry (SNOTEL) stations, and low alti-

tude first-order stations (FOS).

The new models are shown to be competitively accurate in estimating snow loads on a variety of station networks. This is in contrast to existing methods which show strong accuracy on the specific station type or region for which it was developed. Such a model allows for the use of a single depth-to-load conversion method for all locations in the conterminous United States, eliminating the need for different depth-to-load conversion models in different regions or circumstances.

Chapter Highlights:

- A brief summary of the datasets that were used to develop a universal depth-to-load conversion model.
- A review of current depth-to-load conversion methods, including hydrologic models that predict daily, rather than annual maximum, snow loads.
- The introduction of a universal depth-to-load conversion model using the random forests method.
- A comparison of the accuracy of new and existing models on various station networks.

5.1. Data Consolidation

The core dataset for the depth-to-load conversion models was the global historical climatological network - daily (GHCND) [Menne et al., 2012] described in Chapter 4. Relevant observations from the GHCND (excluding Canadian locations) were taken from SNOTEL stations, located at high altitudes in western

states, and FOSs generally located at airports scattered across the country. These observations were supplemented with Snow Course (SC) observations from the Natural Resources Conservation Service (NRCS). Additional supplemental data came from region specific datasets in Maine (ME) [Maine Geological Survery, 2020], New York (NY) [NRCC, 2020], and California (CA) [CDWR, 2020]. These supplemental data were necessary to overcome the lack of direct load measurements in eastern states. However, only GHCND data was used for reliability-targeted load (RTL) calculations as these data were most consistent and dependable in terms of accessibility and quality control.

Whenever a station was simultaneously reporting in two separate networks, measurements were retained only from the station network that was easier to access. The data from each of these sources include measurements of snow depth (SNWD) and the water equivalent of snow on the ground (WESD), which is equivalent to snow load. Available station location information in each network includes elevation (E), latitude (LAT) and longitude (LON). Table 5.1 shows the measurement frequency, sample size (yearly maximum ratio), and indication of quality control checks prior to data publication. Measurements

Table 5.1: Comparison of measurement frequency, data availability, and provided quality control (QC) checks for the considered station networks.

Network	Frequency	Stations	N	QC
SNOTEL	Daily	825	13,465	Yes
FOS	Daily	177	4,265	Yes
SC	Monthly	742	13,640	No
ME	weekly	218	3,046	Yes
NY	Bi-Monthly	456	10,862	No
CA	Daily	55	601	No

from these data sources were combined into a single dataset of SNWD/WESD

pairs and grouped by snow season, which covers October of previous year to June of listed year.

The variable of interest is the ratio $\rho_d(i,j) = \max(\text{WESD}_{i,j}) / \max(\text{SNWD}_{i,j})$, where i and j represent stations and years respectively. Because the annual maximum measurements of WESD and SNWD need not occur on the same day, final values of the ρ_d are not observed ratios, but representations of the maximum snow density for each station/water year pair. The ratio ρ_d will be referred to as "specific gravity" throughout the remainder of this chapter.

5.1.1. Climate Normals

Station meta data were supplemented with 30-year climate normals (i.e. averages) obtained from 800 meter resolution PRISM maps [Daly et al., 2008]. The inclusion of these climate normals makes it possible to account for the effect of climate on snow densities, motivated by the recent success of Hill et al. [2019] in a similar approach. Table 5.2 lists the PRISM climate normals considered in model development. Site-specific values of each climate variable were extracted from the PRISM maps using bilinear interpolation. Other variables considered in model development but not obtained via the climate grids are provided in Table 5.3.

Table 5.2: Description of 30-year normals used as explanatory variables in the regression tree models.

Name	Description	Units	Variable
MCMT	Mean Coldest Month Temperature	°C	T_c
MWMT	Mean Warmest Month Temperature	°C	T_w
TD	MWMT - MCMT	°C	T_d
PPTWT	Winter Precipitation (Dec - Feb)	mm	P_t

Table 5.3: Description of 30-year normals used as explanatory variables in the regression tree models.

Name	Description	Units	Variable
SNWD	Snow Depth	mm	h
D2C	Distance to Coast	km	D_c
Elevation	Elevation	m	E
SMONTH	Month of Max Depth (Oct - 1, Jun - 9)		M_s

5.2. Data Processing

Quality control checks were performed both on the data and the meta-data. For station meta-data, misreported geographical coordinates created mismatches between the mapped climate normals and the actual climate of the measurements. Potentially misreported locations were flagged by comparing the official station elevation and the PRISM elevation map. Stations were removed from consideration if the officially listed elevation was less than 0.8 times the lowest PRISM elevation, or greater than 1.2 times the highest PRISM elevation, observed in a 3 mile radius. This resulted in the removal of 24 candidate stations: one from the CA network, five from the NY network, and 18 from the SC network.

GHCND data were subject to the same automatic and manual quality control measures described in Chapter 4. For supplemental networks, any observations flagged by the data administrators were also removed prior to analysis, but no additional manual checks of individual observations were performed. All station networks were subject to coverage filters described in Chapter 4, though the exception allowing observations above the median to be retained regardless of coverage was not allowed. The removal of the median exception was in part due to smaller periods of record where both SNWD and WESD

are recorded which makes estimates of the median less robust.

The smaller sample size issue is exacerbated by the need for both measures to pass coverage filters each year. This is in contrast to the condition required for distribution fitting which was that *at least one* measure passed the coverage filter. At the same time, it is crucial that the maximum SNWD and snow load for each year are correctly represented in order to ensure the validity of the ρ_d measurements. In order to prevent excessive loss of observations at under-represented locations, stations in the FOSs, NY, and ME were only required to have observations in three of the four months in which a maximum snow load was most likely to occur.

Observations of ρ_d above 0.8 were removed from consideration, which is a density typical of "firn" (i.e. pre-glacial) snow [Copland, 2020]. Similarly, observations of ρ_d below 0.05 were also removed per recommendations from members of the project steering committee. The sample sizes provided in Table 5.1 represent the observations that remain after data filters are applied.

5.3. Current Methodologies

Numerous region-specific models have been developed to estimate snow loads from snow depth. Some methods focus only on estimating annual maximum or 50-year snow loads, while other methods attempt to estimate snow loads on a monthly or daily scale. Additionally, some methods predict load (p_g) directly, while others predict specific gravity which is easily converted to snow load. This section considers a variety of density methods that can be readily used to predict annual maximum snow loads. For convenience, all equations are converted from their original forms to show the estimated values of p_g.

5.3.1. Rocky Mountain Conversion Density

The Rocky Mountain Conversion Density (RMCD) models snow loads in western states solely as a function function of SNWD, denoted as h in the equation and measured in inches [Sack and Sheikh-Taheri, 1986]. This method is a two part linear regression represented as

$$p_g(h) = \begin{cases} 0.90 * (h)\,, & h \leq 22 \\ 2.36 * (h) - 31.9, & h > 22 \end{cases}.$$

Coefficients were determined using high elevation snow course data, making the model most suitable to to predict snow loads in western states where snow is expected to accumulate throughout the season.

5.3.2. Colorado Models

The state of Colorado developed a similar depth-to-load model using high elevation SNOTEL and snow course data specific to their state [DeBock et al., 2016]. Their study acknowledged that the resulting power curve was most appropriately applied to "compacted" snow sites subject to consistent snow accumulation. This model is given as

$$p_g(h) = \mathrm{COLH}(h) = 0.584 * (h)^{1.25}\,.$$

However, this curve overestimates snow loads at "settled" snow sites, which are locations where the snow does not always persist throughout the season. For such locations, the Colorado study made use of the depth-to-load model (TOB) developed by Tobiasson and Greatorex [1997] using data from FOSs. These stations tend to be more representative of populated locations not subject to

consistent snow accumulations. This model is defined as

$$p_g(h) = \text{TOB}(h) = 0.279 * (h)^{1.36}\,.$$

The combined Colorado model (COL) takes a weighted average of the predictions from both curve at locations with elevations falling between those typical of SNOTEL stations and Colorado's FOSs.

It is worth noting that the TOB model was developed by relating 50-year snow depths to 50-year snow loads. An annual version of this same curve was obtained via personal communication with the TOB model authors. This equation is defined as

$$p_g(h) = 0.342 * (h)^{1.32}\,.$$

It has been confirmed that the loads resulting from this annual alternative are not appreciably different than those obtained from the original TOB model.

5.3.3. Sturm's Equations

Alternative depth-to-load conversion models come from research in hydrology and attempt to model daily snow densities for various climate classes. One notable method is described by Sturm et al. [2010], who created a bulk density equation with varying coefficients based on climate class. This model can be summarized by the following equation:

$$p_g(h, d, C_c) = ((\rho_{max} - \rho_0)\left[1 - e^{-k_1 * (h * 2.54) - k_2 * d}\right] + \rho_0) * 0.2048 * h$$

Here, C_c is the distinct climate class indicating where the measurement of h

was taken and ρ_{max}, ρ_0, k_1 and k_2 are parameters specific to the particular C_c. These values are summarized in Table 5.4. This model was used in the most recent Utah snow load study [Bean et al., 2018] and is referred to as STURM for the remainder of this chapter. It has been noted that this model likely

Table 5.4: Parameters for Sturm's equation for each distinct climate class.

CC	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.598	0.224	0.001	0.004
Maritime	0.598	0.258	0.001	0.004
Prairie	0.594	0.233	0.02	0.003
Tundra	0.363	0.243	0.003	0.005
Taiga	0.217	0.217	0.0000	0.0000

over-estimates ground snow loads at most low elevation locations in Utah. This conservatism was a desirable feature in the context of Utah snow load study, but perhaps not appropriate on a national scale.

5.3.4. Hill's Climate Map Approach

Like Sturm et al. [2010], Hill et al. [2019] developed a regression model for estimating WESD that can account for environmental variability in a continuous fashion rather than using discrete climate classes. This was done by using 30-year gridded climate normals obtained from the ClimateNA project [Wang et al., 2016]. This model has separate equations for the snow accumulation and ablation phases of each water year. The model was fit using SNOTEL station

data and is expressed in final form as

$$
\begin{aligned}
&p_g(h, P_t(\boldsymbol{u}), T_d(\boldsymbol{u}), D_Y) \\
&= \begin{cases} 0.2048 * 0.053 h^{0.948} P_t^{0.170} T_d^{-0.131} D_Y^{0.292} & D_y < 180 \\ 0.2048 * 0.0481 h^{1.0395} P_t^{0.1699} T_d^{-0.0461} D_y^{0.1804} & D_y \geq 180. \end{cases}
\end{aligned}
$$

where D_y represents the day of he snow season and P_t and T_d are defined in Table 5.2. Hill et al. [2019] demonstrates that the consideration of climate variables improves upon Sturm et al. [2010] in terms of accuracy. While not specifically designed for annual maximum depths, the model can be readily used for this purpose.

5.3.5. Bulk Density Equations

There exists a large body of research that aims at directly modeling ρ_d. These methods are often referred to as "bulk density equations" and tend to be simple and easy to scale nationally. The bulk density equations considered in this chapter are compared in Avanzi et al. [2015] on a limited number of SNOTEL stations. This chapter expands the original comparison by Avanzi et al. [2015] to a national scale.

5.3.6. Other Methods

Other depth-to-load conversion methods do exist, most notably the Montana-specific depth-to-load equations applied in Theisen et al. [2004]. Other, more complicated time series models also exist [Meløysund et al., 2007, McCreight and Small, 2014], but require measurements on a time scale not feasible at most weather stations. As such, model comparisons in this chapter are limited

to methods that can readily extended to a national scale.

5.4. Modern Regression Approach

One major limitation of all of the above described approaches is that each model was developed using a particular weather station type, which limits its efficacy in different regions or climates. Most are developed using only high elevation SNOTEL data with snow accumulation patterns very different from most populated locations. The Tobiasson and Greatorex [1997] model is an important exception as it was developed with FOS data that is more relevant to most populated locations, but perhaps not relevant to populated locations that receive more snow than is typically observed at FOS locations. Using existing depth-to-load conversion models would require different models to be selected for use in different parts of the country, requiring extensive knowledge of the varied climate of the country that is beyond the expertise of the authors.

Rather, the authors use modern regression approaches to characterize differences in snow density properties across the country. These models are able to characterize high-ordered interactions and non-linear effects across time, depth, and climate, to provide accurate estimates of snow densities at both FOS and SNOTEL locations. The key advantage of the modern regression approach is the elimination of the need for different models in different climates and at different elevations.

The model of choice is named random forests (RF), which is an extension of regression trees (rtree). Both models make use of gridded climate data similar to Hill et al. [2019]. The following subsections describe the structure of these models as well as a brief descriptions of their implementation.

5.4.1. Regression Trees

Regression trees [Breiman et al., 1984] are a machine learning technique that are popular due to their relatively straightforward representations. Regression trees are comprised of a number of binary splits on the predictor variables which results in a set of disjoint prediction "branches." The tree is fit using a greedy algorithm that at each step makes a split on the predictor variable that results in the greatest possible reduction in the Residual Sum of Squares (RSS):

$$\text{RSS} = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

where $\hat{y}_{R_j}$ represents the predicted values of the response variable for all observations falling into the R_jth terminal node (i.e. bin with no more splits), and J represents the total number of terminal nodes that result from the proposed split in the regression tree. Predicted values from the tree in this case are simply the average value of ρ_d for all observations that fall into the same terminal node.

Fully grown trees can fit the input data perfectly, which usually leads to poor accuracy when predicting new observations. Instead, trees are "pruned" (by means of a cost-complexity parameter) so that the tree is large enough to be accurate, but small enough to generalize to new observations. To prune the tree in this analysis, it was required that each terminal node have no less than 1% of the total number of observations and that each split resulted in at least a 0.1% increase in the total variance of ρ_d explained by the model, similar to Hill et al. [2019].

A representation of the final regression tree for predicting ρ_d is observed in Figure 5.1. Observations fall to the left if the listed condition at each split is

met, and falls to the right otherwise. For example, at the first split, observations fall to the left if the given snow depth is less than $e^7/25.4 \approx 43$ inches. For the low depth measurements, observations fall again to the left if the mean temperature of the warmest month is greater than 21 degrees Celsius. This decision making process continues until the observation falls into a terminal node and is assigned the average value of the node. Notice that the second level splits occur on different variables for the low and high depth observations. These differences highlight the ability of the regression tree to characterize interactions among the variables, as certain variables are only important in characterizing certain subsets of the data.

The tree predicts ρ_d rather than p_g. These predictions are readily converted to snow loads as

$$p_g = 0.2048\rho_d h$$

While more complicated than linear regression, the regression tree is a relatively simple alternative among possible machine learning approaches. Despite its relative simplicity, the model is surprisingly effective at estimating snow loads as discussed in Section 5.5. However, the discrete "jumps" in densities as observations transition between nodes creates significant issues in the distribution fitting approaches described in Chapter 6. This problem is resolved by smoothing the transitions between terminal nodes with a RF model.

5.4.2. Random Forests

Random forest models are simply collections of regression trees where each tree is fit using a bootstrap sample of the original data, and each split in the tree is made using a random subset of the available variables. The random

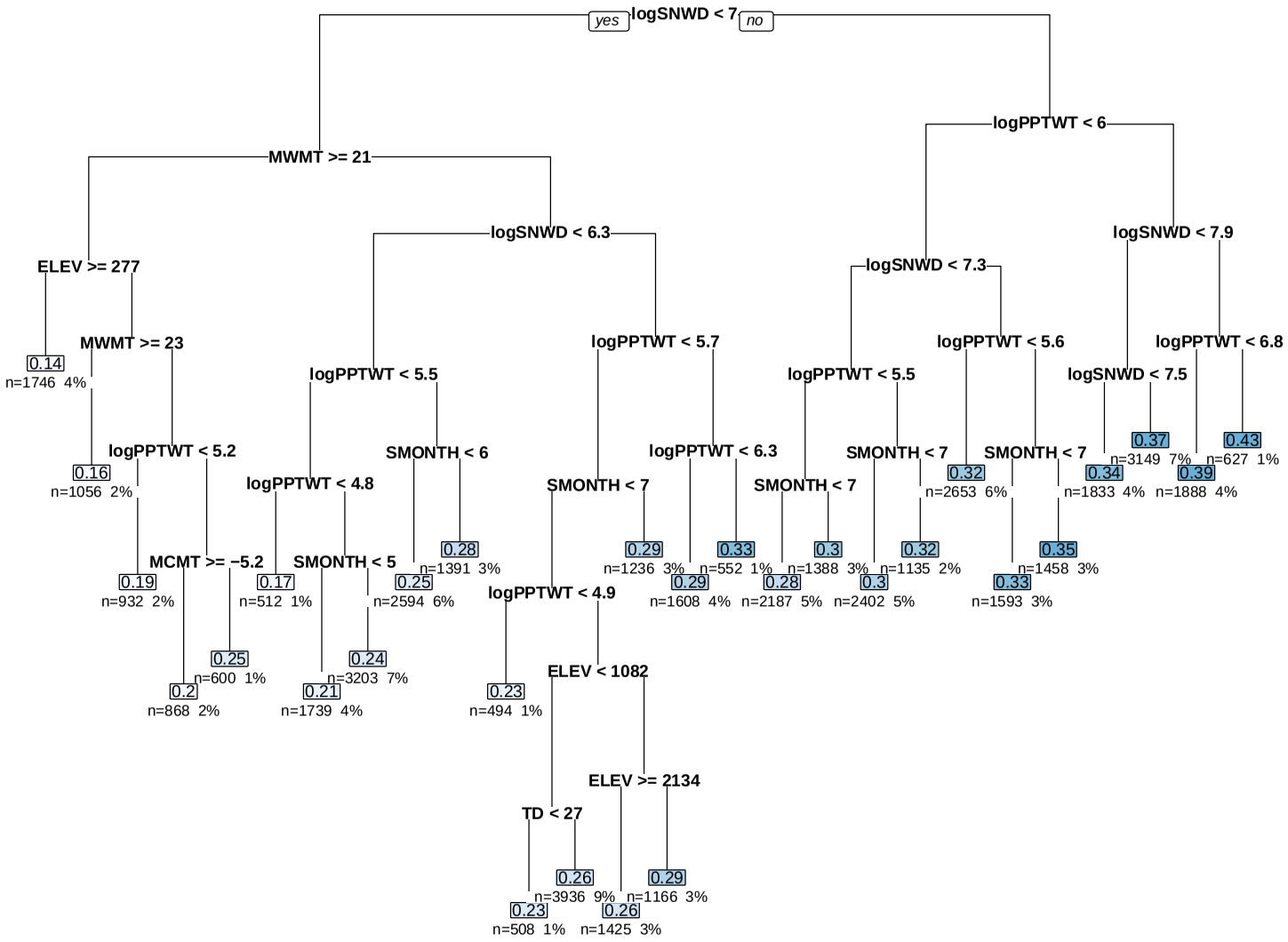

Figure 5.1: Regression tree fit on all available data. The text under each terminal node represents the number and percentage of observations that fall into each region.

bootstraps and subsets encourage diversity among the trees, so that each tree predicts a slightly different value for each input observation. Final predictions for an observation are simply an average prediction made from all trees in the forest. This average has the effect of smoothing the regression tree predictions and tends to improve the accuracy compared to using a single regression tree. For this analysis, the RF model is fit using 201 regression trees, and each tree is allowed to grow until each terminal node contains no less than 0.5% of the total number of training observations.

The improvement in predictive accuracy, however, comes at the cost of model interpretability, as the RF model consists of many trees that are impossible to effectively visualize. However, it could be argued that a large collection of linear regression models bound together by a model selection algorithm has its own interpretation difficulties. RF models are best visualized by their outputs, as is explored in Section 5.6. Random forests also provide a unique and robust measure of variable importance not possible in linear models. This metric is obtained by determining the decrease in accuracy that occurs when the information for one of the explanatory variables is randomly permuted. Variables that, when permuted, result in greater losses of predictive accuracy are deemed more important. Figure 5.2 shows that the most important variable in predicting ρ_d (not p_g) is the month in which the maximum snow depth occurred, followed by winter precipitation and snow depth.

The importance of the month variable reflects the tendency for snow density to increase throughout the season. It is therefore possible that the RF model will predict a higher snow load for depth slightly lower than the annual maximum that occurs later in the season. While the models were estimated using only annual maximum depths, estimates of snow load are made using monthly

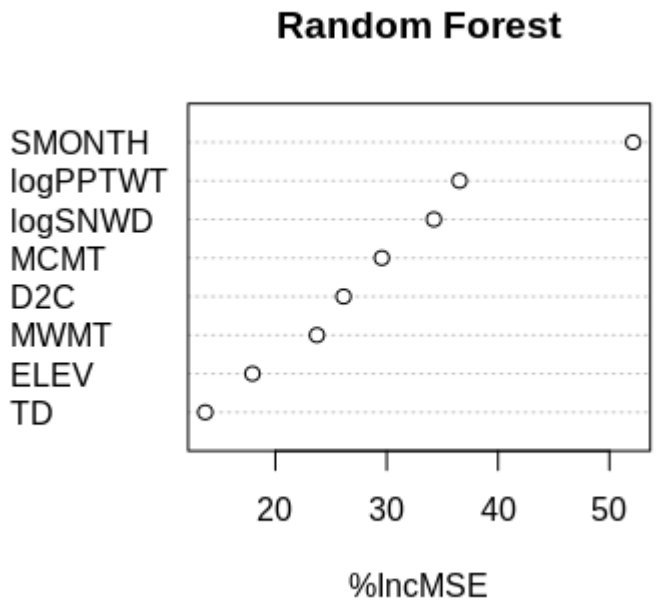

Figure 5.2: Variable importance plot for the RF model. Variables associated with a higher percentage increase in mean square error (MSE) are more important in prediction.

maximum depths, retaining only the maximum estimated load for each season. This approach will give the same or slightly higher estimates of annual snow load than compared to predictions made using only annual maximum depths.

5.5. Accuracy Comparisons

All available observations were used when creating the final RF model. However, it is also important to determine the effectiveness of both new and existing models in predicting snow depths on new information. To do this, a secondary version of the random forest and regression tree models were split using only a

subset of the available data (i.e. a training set) and evaluated on the remaining data not used during model fitting (i.e. a test set). The creation of these two subsets was performed at the station level, not the observation level, such that the locations in the training and test sets are fully distinct. A summary of the total number of stations and observations for each station network is given in Table 5.5. Figure 5.3 shows the geographical locations of the training and test set stations.

Table 5.5: Summary of the number of stations (ST) and observations (N) in the training and test sets.

Network	Train		Test	
	N	ST	N	ST
SNOTEL	7,244	434	6,221	391
FOS	2,413	103	1,852	74
SC	6,481	355	7,159	387
ME	1,688	113	1,358	105
NY	5,431	207	5,431	249
CA	266	335	25	30

Splitting the data in this manner demonstrates the ability of the regression tree and RF methods to generalize to locations that were not used in their training. For this analysis, four different accuracy metrics were considered, all comparing the difference between the actual and estimated snow loads (psf). These metrics include:

- Mean Absolute Error (MAE): shows the average error (skewed high by large loads).
- Mean Error (ME): gives a sense of any systematic bias in predictions.
- Median Absolute Error (MedAE): shows the "typical" error and is less sensitive to occasionally large errors.

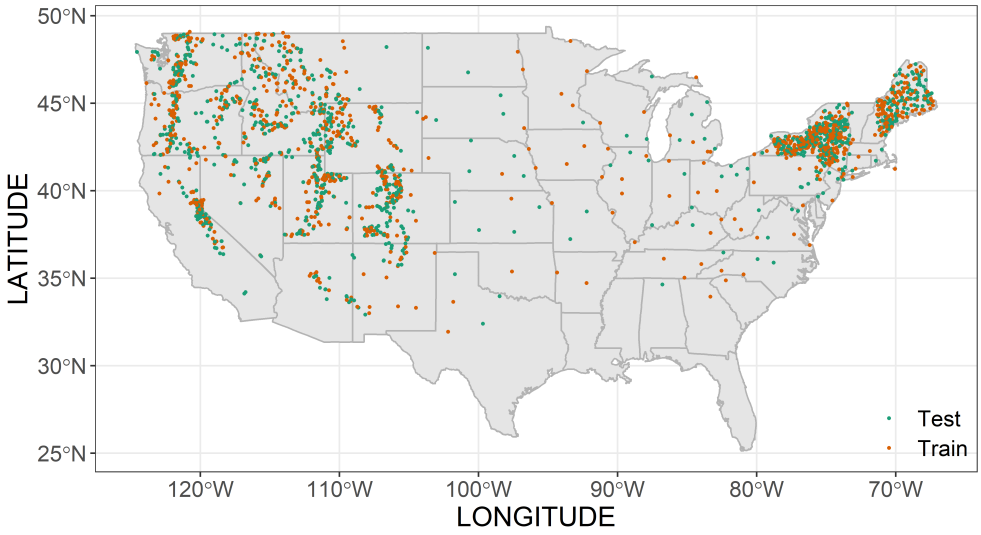

Figure 5.3: Geographic locations of training and testing stations.

- Root Mean Square Error (RMSE): most sensitive to large errors, when compared to the MedAE it provides a sense of the error skewness (greater difference between RMSE and MedAE means greater skew).

The results of the final comparison of methods on the testing dataset can be seen in Figure 5.4. This shows that the RF model outperforms all other methods on the combined dataset and is unbiased in prediction, with the rtree method not far behind. This is not unexpected as this is the only model developed using a combined dataset. Note that the Colorado (COL) method employs the weighted average approach based on elevation as developed in the Colorado report. This was designed to be Colorado specific and not expected to scale nationally. The same could be said for the RMCD. In spite of this, the COL and RMCD methods outperform most of the considered bulk density equations on the p_g dataset, as well as the hydrologic approaches of Hill and

Sturm.

The TOB method performs poorly on the combined dataset but was never intended to be used at SNOTEL station locations, which make up the majority of the combined dataset. A true test of the new RF approach is its ability to maintain performance on specific subsets of the data, such as FOS, which are more relevant stations for most populated locations. This accuracy comparison is visualized in Figure 5.5. For this subset, the accuracy of the TOB model is best (with the COL model being nearly identical to the TOB model in this situation), but the RF model is a close second. This suggests that the Tobiasson and Greatorex [1997] model is effective when used in its intended dataset, and that the RF model is similarly effective. The important implication is that the RF model is competitive (in terms of accuracy) *across* station networks, demonstrating its ability to learn differences in snow densities that occur between different climates and station types. Similar results were observed on the other station networks and using other accuracy comparison approaches, such as spatial cross validation Meyer et al. [2019]. This validates the use of the RF model as a universal approach for estimating snow densities on a national scale.

5.6. Site-Specific Implications

The ability of the RF model to model the interaction between ρ_d, time, and climate is demonstrated in Figures 5.6, 5.7. The low and high elevation depth-to-load models from DeBock et al. [2016], serve as reference lines for the RF predictions. Figure 5.6 shows that the RF predictions follow the TOB curve almost exactly in Salt Lake (a valley location) but follow the Colorado mountain

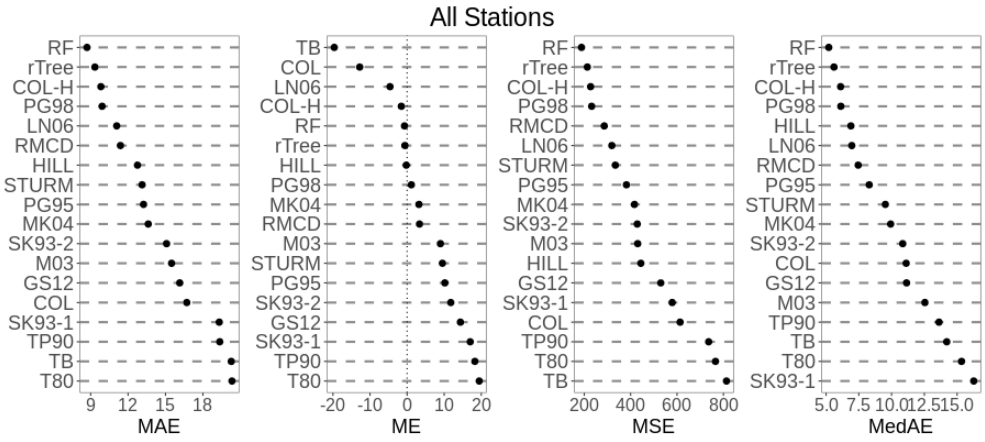

Figure 5.4: Comparison of snow load estimation methods on all stations in the testing dataset. The x-axis is measured in psf.

snow curve almost exactly in Brighton (a popular ski area). Similar effects can be seen at two eastern locations in Figure 5.7, though the RF model seems to slightly over-predict average loads in the Concord, NH case. While not perfect, these figures demonstrate the ability of the RF model to appropriately adjust load predictions based on climate, strengthening the argument for its use as a universal depth-to-load conversion approach.

5.7. Future Work

This chapter has demonstrated that the RF model provides accurate estimates of annual maximum snow loads from snow depths across a variety of station networks. All considered models estimated annual maximum snow loads, which allows for combinations of direct and indirect measurements of loads in the distribution fitting step described in Chapter 6. Depth-to-load conversion methods not considered in this chapter, including those intended for use with 50-year

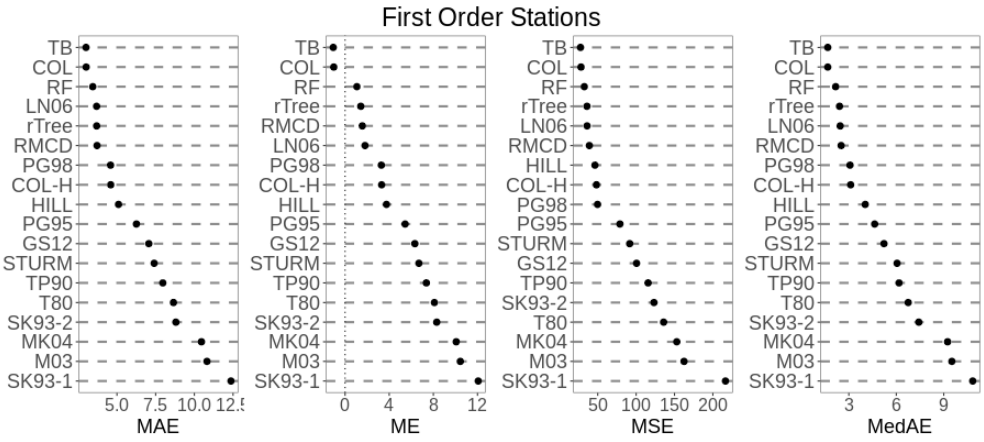

Figure 5.5: Comparison of snow load estimation methods on FOS in the testing dataset. The x-axis is measured in psf.

snow depths rather than annual snow depths, deserve further investigation in future comparisons.

Another area where further investigation is warranted is the impact of the depth-to-load conversion model on the distribution fitting of the annual ground snow load. All of the models considered in this chapter, including the proposed regression tree and random forests, are smoothing methods which have the intrinsic effect of a reduced variability in the predicted loads. As such, they all have the potential to condense the distribution fitted to the loads, leading to an underestimation of any extreme event like the 50-year load. DeBock et al. [2016] demonstrated that the Colorado depth-to-load conversion models resulted in unbiased estimates of annual ground snow load probability distribution parameters, but the available data to make this determination was overwhelmingly from SC and SNOTEL stations. Such results are consistent with the authors' observations of unbiased estimations of 50-year events using indirect measurements of snow loads at SNOTEL stations, but similar

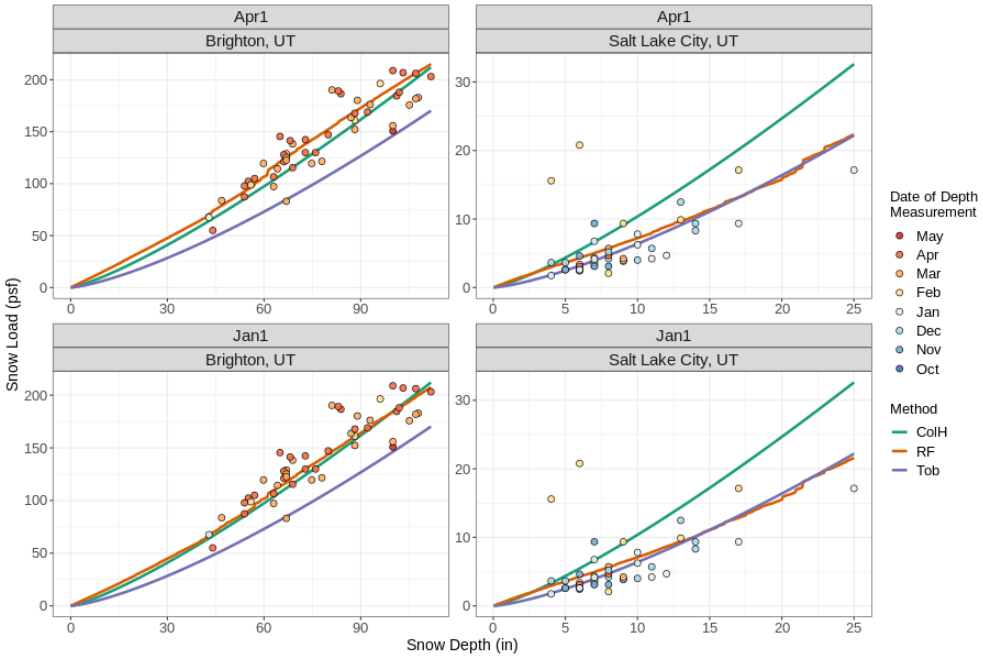

Figure 5.6: Comparison of the newly proposed depth-to-load conversion predictions against the high (COLH) and low (TOB) elevation models described in DeBock et al. [2016] at locations in the state of Utah. Scatterplots show measured depth/load pairs at each location.

comparisons of 50-year loads using direct and indirect measurements at FOSs proved much more variable and biased (high or low) across all approaches (new and existing). Limited information was available for these preliminary distribution fitting comparisons at FOSs and more study is needed to investigate the potential bias on distribution fits at these locations.

While there is no substitute for direct measurements of load, the RF model presented in this chapter has proven effective in estimating snow loads at an annual scale. These estimates are crucial to supplementing the lack of snow load measurements at most weather stations across the country. The key advantage of the RF approach is the elimination of the need for different model

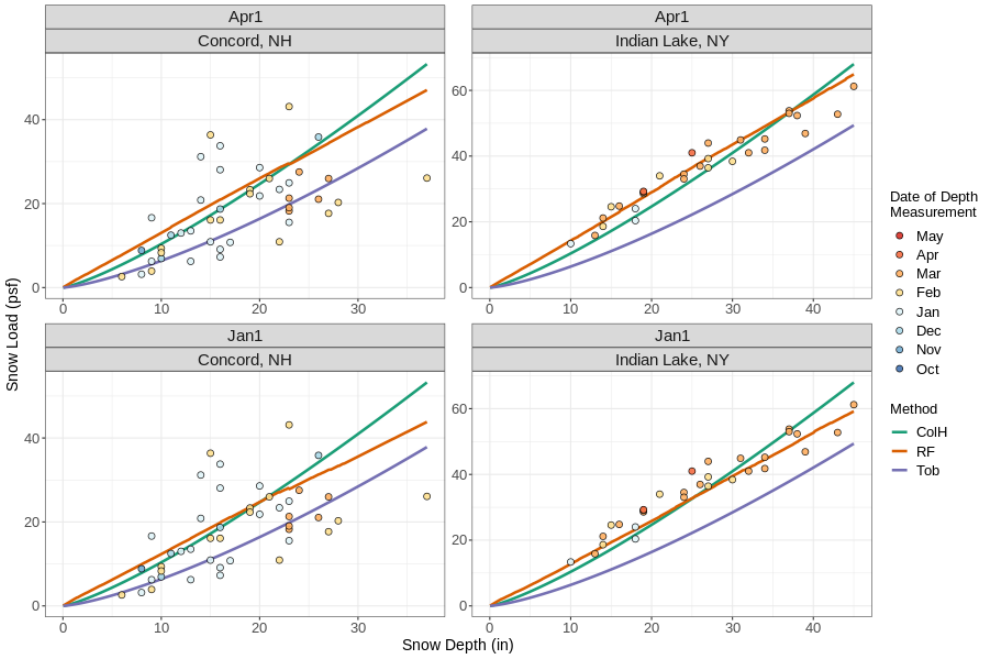

Figure 5.7: Comparison of the newly proposed depth-to-load conversion predictions against the high (COLH) and low (TOB) elevation models described in DeBock et al. [2016] at locations in eastern states. Scatterplots show measured depth/load pairs at each location.

equations for different regions/elevations, which allows for the newly proposed methodology to be easily deployed on a national scale.

Bibliography

Avanzi, F., de michele, C., and Ghezzi, A. (2015). On the performances of empirical regressions for the estimation of bulk snow density. Geografia Fisica e Dinamicca Quaternaria.

Bean, B., Maguire, M., and Sun, Y. (2018). The Utah snow load study. Tech-

nical Report 4591, Utah State University, Department of Civil and Environmental Engineering.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). Classification and regression trees. CRC press.

CDWR (2020). California snow survey data. water.ca.gov/Contact. Data obtained through personal correspondence with organization.

Copland, L. (2020). Properties of glacial ice and glacier classification. In Reference Module in Earth Systems and Environmental Sciences. Elsevier.

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

Hill, D. F., Burakowski, E. A., Crumley, R. L., Keon, J., Hu, J. M., Arendt, A. A., Wikstrom Jones, K., and Wolken, G. J. (2019). Converting snow depth to snow water equivalent using climatological variables. The Cryosphere, 13(7):1767–1784.

Maine Geological Survery (2020). Maine snow survey data. Accessed: 2020-04-03.

McCreight, J. L. and Small, E. E. (2014). Modeling bulk density and snow water equivalent using daily snow depth observations. The Cryosphere, 8(2):521–536.

Meehleis, K., Folan, T., Hamel, S., Lang, R., and Gienko, G. (2020). Snow load calculations for alaska using ghcn data (1950–2017). Journal of Cold Regions Engineering, 34(3):04020011.

Meløysund, V., Leira, B., Høiseth, K. V., and Lisø, K. R. (2007). Predicting snow density using meteorological data. Meteorological Applications, 14(4):413–423.

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.

Meyer, H., Reudenbach, C., Wöllauer, S., and Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. Ecological Modelling, 411:108815.

NRCC (2020). New york snow survey data. http://www.nrcc.cornell.edu/. Data obtained through personal correspondence with Northeast Regional Climate Center.

Sack, R. L. and Sheikh-Taheri, A. (1986). Ground and roof snow loads for Idaho. University of Idaho, Department of Civil Engineering.

SEAO (2007). Snow load analysis for oregon.

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. (1997). Database and methodology for conducting site specific snow load case studies for the United States. In Proc., 3rd Int. Conf. on Snow Engineering, Izumi, I., Nakamura, T., and Sack, RL, eds., AA Balkema, Rotterdam, Netherlands, pages 249–256.

Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for north america. PLOS ONE, 11(6):1–17.

Chapter 6
Site-Specific Distribution Fitting

6.1. Introduction

The central-element of the reliability-targeted design ground snow load (RTL) estimation problem is the assumed distribution of the annual ground snow loads. The RTLs require accurate estimations of the extreme right tail of the ground snow load probability-distribution, which makes RTLs sensitive to even small changes in the estimated distribution parameters. Robust estimates of the ground snow load distribution parameters are difficult to obtain given the short periods of record relative to the targeted probabilities of failure. The problem is exacerbated by the occasional misreported maximums that go undetected in the quality control step described in Chapter 4.

This chapter describes a regional generalized extreme value (GEV) distribution fitting approach, where estimates of the distribution tail shape are informed by geographically close stations with similar patterns of snow accumulation. The third parameter of the GEV distribution, called the shape parameter, provides greater flexibility in modeling the shape of the upper tail of extreme ground snow loads than can be obtained with traditional two-parameter distributions. A lower shape parameter results in a "lighter" upper-tail where

extreme snow events are less likely than the same distribution with a higher shape parameter or "heavier" upper-tail where extreme snow events are more likely. The analysis reveals that high altitude and far north locations have lighter upper-tailed distributions than would be expected with the log-normal distribution, while certain mid-latitude locations known for their occasional "superstorms" have heavier distribution tails than would be expected with the log-normal distribution.

The chapter also describes an alternative distribution fitting approach employed in places that consistently have annual maximum snow loads equal to zero. The result is a set of geographically consistent RTLs that accurately reflect regional differences in snow accumulation patterns across the country.

Chapter Highlights:

- A review of alternative distribution fitting approaches.
- A description of the regional smoothing of the GEV shape parameter to ensure robust estimations of the upper tail of the ground snow load distributions.
- A summary of an alternative distribution fitting approach for locations with mostly zero-valued annual snow load maximums.
- A discussion of practical constraints used to ensure consistent RTL estimates.

6.2. Previous Approaches

Extreme value analysis has a relatively long history with a wide variety of applications [Gumbel, 2004]. Probabilistic characterizations of environmental hazards have been an integral piece of structural reliability analysis since the inception of load resistance factor design in the late 1960s [Ellingwood, 2000]. The seminal work of Ellingwood et al. [1980] defined a load factor that related a 50-year ground snow load to the RTL by fitting log-normal distributions to annual maximum ground snow loads at eight locations and deriving a Extreme Value Type I roof load distribution at each location. The mean and coefficient of variation (COV) at each location were then averaged to create a single probability model for roof loads that resulted in the current 1.6 snow load factor defined in ASCE 7. The averaging of the site-specific coefficients made this final probability model less sensitive to changes in the input data than would have been the case if RTLs were calculated for each individual site. A region-specific set of RTLs were proposed by Lee and Rosowsky [2005], though this approach also relied upon an averaging of individual probability distributions within each region using an expanded set of ground snow load measurement locations.

Since that time, the primary focus of extreme snow load analysis has been on accurate characterizations of 50-year events, both at the national [Tobiasson and Greatorex, 1997] and state levels [Tobiasson et al., 2002, Sack, 2015, Sack et al., 2016, Meehleis et al., 2020]. Each of these national and state-specific reports have used a variety of two-parameter probability distributions to model annual maximum ground snow loads, though the log-normal distribution appears to be most common. The distribution fitting approaches have varied widely in each study, with some studies fitting distributions to all annual max-

imums and others fitting distributions to the upper tails of the distributions. In every case, the focus on 50-year events reduces the need for extreme tail extrapolation, which is why estimated 50-year events tend to be less sensitive to changes in the estimated distribution parameters than direct estimates of RTLs.

The challenge of robust estimates of site-specific RTLs is demonstrated in DeBock et al. [2017], which acknowledged the difficulty of obtaining consistent estimates of RTLs from short periods of record. Their remedy for this issue involved clustering measurement locations into six (consolidated to four for this study) climate regions in an adaptation of a region of influence approach [BURN, 1990]. Annual maximum snow loads from individual sites were then scaled to have a common 95th percentile to create "super-stations" with more observations in the distribution tails than could be obtained at any individual site. The estimated parameters resulting from the combined distributions were then adjusted to better reflect site specific conditions [DeBock et al., 2017]. This process highlights the perhaps unavoidable need for a site-specific reliability analysis to be partially informed by available information at neighboring locations with similar snow accumulation patterns.

The region of influence approach requires expert opinion and local knowledge to cluster the stations, both of which are difficult to scale nationally. Further, it is unclear how clustering might be employed in locations where climate regions are not highly correlated with elevation. Regardless, the region of influence approach described in DeBock et al. [2017] provides a template for leveraging information from surrounding stations in the calculation of site-specific RTLs.

6.3. The Generalized Extreme Value Distribution

One critical observation in DeBock et al. [2017] are the distinctly different probability distribution tail behaviors observed in high vs. low elevation locations. High elevation stations subject to consistent snow accumulation had lighter distribution tails than low elevation locations subject to intermittent snow accumulation. These different tail behaviors were expressed via differences in the COV, though all measurement locations in DeBock et al. [2016] were assumed to follow a log-normal distribution. The GEV distribution is a collection of three two-parameter extreme value distributions tied together by a third parameter called the shape parameter. The flexibility in modeling the distribution shape offered by the third parameter allows for better characterizations of the differing tail behaviors observed in the Colorado study and is a popular distribution for estimating extreme hydrologic events [Martins and Stedinger, 2000, Feng et al., 2007, Panagoulia et al., 2014].

The GEV distribution has the nice theoretical property that any set of extreme measurements (such as annual maximum loads) are guaranteed to converge to one of the three GEV distribution types given a sufficiently large sample size. A shape parameter of zero results in the Type I or Gumbel distribution, which has a lighter tail than the log-normal distribution. A shape parameter greater than zero results in a Frechet or Type II distribution and may have a heavier tail than the log-normal distribution based on the magnitude of the shape parameter. Finally, a shape parameter less than zero follows a reversed Weibull distribution with a finite upper bound. The probability density function of the GEV distribution, $f(x)$ in Equation 6.1, and cumulative

distribution function (CDF), $F(x)$ in Equation 6.2, are defined below in terms of location μ, scale σ, and shape ξ. The possible range of values for Equations 6.1 and 6.2 are given in Equation 6.3. Figure 1.1 in Chapter 1 provides examples of distribution shapes resulting from each distribution type.

$$f(x) = \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1}\exp\left[-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right] & \xi \neq 0 \\ \frac{1}{\sigma}\exp\left[-\left(\frac{x-\mu}{\sigma}+\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right)\right] & \xi = 0 \end{cases} \tag{6.1}$$

$$F(x) = \begin{cases} \exp\left[-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right] & \xi \neq 0 \\ \exp\left[-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right] & \xi = 0 \end{cases} \tag{6.2}$$

$$x \in \begin{cases} (-\infty, \mu - \frac{\sigma}{\xi}] & \xi < 0 \\ (-\infty, \infty) & \xi = 0 \\ [\mu - \frac{\sigma}{\xi}, \infty) & \xi > 0 \end{cases} \tag{6.3}$$

6.4. Distribution Fitting

One common approach for estimating extreme events is distribution tail fitting, where distribution parameters are derived by focusing on only a portion of the observations in the upper tail [Nowak and Collins, 2012]. The result has the advantage of providing more accurate estimates of the upper tail of the distribution (which is usually the primary interest in reliability analysis), but with the disadvantage of increased sensitivity in parameter estimates due to the effective reduction in sample size. Figure 6.1 show example distributions fit to annual maximum snow loads in Denver, Colorado. It should be noted that

there is record of at least one 30+ psf snow load in the city though this value was not recorded in the available data from the Global Historical Climatological Network - Daily (GHCND) Dataset. Regardless, Figure 6.1 still effectively illustrates differences in the estimated quantiles from each distribution fitting approach. The log-normal distribution fit to all of the observations fails to properly characterize the upper tail of the distribution as well as the tail fit log-normal distribution. However, notice that the GEV distribution characterizes both the upper and lower tail of the distribution of the data. The flexibility offered by the shape parameter is reflected in the curve of the GEV distribution on the probability plot. This provides evidence that the GEV distribution can properly characterize tail behavior like a tail-fitting approach, while maintaining the relative stability in parameter estimates that comes when fitting all observations. In this study, GEV distributions are fit to annual maximum snow loads using L-moments, a variant of probability weighted moments [Hosking et al., 1985] known to produce parameter estimates that are robust to outlier values and small sample sizes [Hosking, 1990]. Success using L-moments to estimate 50-year ground snow loads was recently demonstrated by Cho and Jacobs [2020].

6.4.1. Low Outlier Screens

The shape parameter has a substantial influence on the relative magnitude of the RTL estimates. The shape parameter estimates are sensitive to anomalously low maximums usually due to poor reporting during a particular snow year. While Chapter 4 describes extensive efforts to remove such observations, the undetected anomalous low values that persist disrupt accurate estimations of the distribution shape. These low values often manifest themselves in the

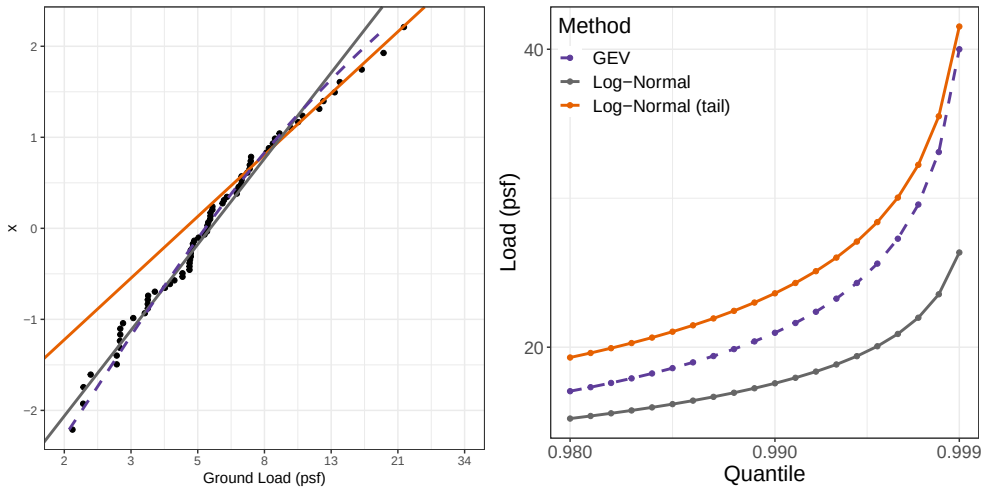

Figure 6.1: Example of various distributions fit to annual maximum loads observed in Denver, CO. The left shows probability plots with distributions overlaid. The right shows estimated quantiles (0.98 - 50 year to 0.99 - 1,000 year) for each distribution.

form of a negative estimate of the shape parameter. While the GEV distribution is intended to be fit to all observations, it is reasonable to assume that the lowest valued maximums should not have undue influence on the estimated distribution shape. This in mind, an automatic screening strategy is employed that:

1. Fits three separate GEV distributions using L-moments at each location:
   (a) using all data,
   (b) using all data except the lowest recorded maximum,
   (c) using all data except the lowest two recorded maximums.
2. If the shape parameter in (b) or (c) is 0.1 units larger than the shape parameter in (a) and the shape parameter in (a) is negative, then discard the (a) distribution fit. If both conditions are not met, then use the (a) distribution fit and skip step 3.

3. If the shape parameter in (c) is 0.1 units larger than the shape parameter in (a) and the shape parameter in (b) is negative then use the (c) distribution fit. Otherwise, use the (b) distribution fit.

This strategy results in the removal of 765 low non-zero maximum values from the more than 0.5 million original maximum values.

6.4.2. Distribution Screens

Despite best efforts to remove misreported values from the dataset, the realities of imperfect data make distribution fits untenable at some locations. Poor fits are flagged by detecting anomalous values of the shape parameter. Hosking [1990] notes that $-0.5 < \xi < 0.5$ in practice and that estimated parameters are no longer asymptotically efficient outside of this range. For this reason, all measurement locations with initial shape parameter estimates below -0.5 or above 0.5 were removed from consideration. This resulted in the removal of 83 of the 9715 candidate Tier 1 and 2 stations.

6.4.3. Shape Parameter Smoothing

The key advantage of the GEV distribution is greater flexibility in modeling the upper tail of the distribution with the shape parameter. However, this flexibility comes with the need to estimate an additional parameter, which is difficult to accomplish with short periods of record. Even small changes in parameters, especially the shape parameter, can cause substantial changes in the estimated RTLs. Similar sensitivity is also observed fitting log-normal distributions. Left unrestrained, this sensitivity can result in large disparities in estimated RTLs within the same municipality. Consider for example the disparities in RTLs

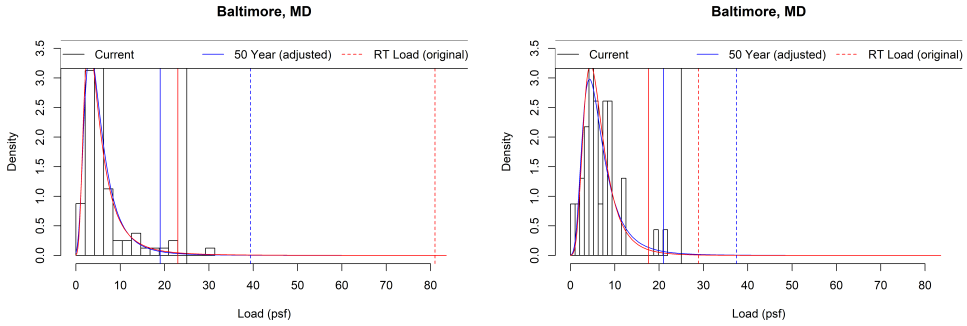

Figure 6.2: Sample of raw (red) and adjusted (blue) 50 year (solid) and RTL (dashed) at two measurement locations in Baltimore, MD.

observed at two separate measurement locations in Baltimore, MD, observed in Figure 6.2. One location recorded a 30 psf snow load event while the other, due to differences in recording periods, records no measurements much larger than 20 psf. Left unrestrained (i.e. red lines), the RTL (which are divided by 1.6 to be comparable to 50 year loads) in one location is nearly triple the RTL in the other location. These disparities reinforce the need to leverage surrounding information to inform parameter estimates. The blue lines shown in Figure 6.2 illustrate the results of measures described in this section to ensure consistency in geographically close and climatically similar locations.

Despite these occasionally large site-specific differences due to misreported maximums or small sample sizes, the *average* distribution shape parameters show strong and consistent local patterns. These patterns seem to be strongly related to local snow accumulation patterns: locations whose peak loads are the result of a few major storms tend to have large shape parameters while locations whose peak loads are the result of the accumulation of many storms throughout the snow season tend to have small (or even negative) shape parameters. Patterns in typical snow accumulation are represented by the median

annual maximum snow load from the available period of record. Examples of these patterns in four ecoregions are shown in Figure 6.4. The smoothed shape parameters include a manually applied lower bound at zero for reasons described in Section 6.4.4. Note that some regions, such as a plains of Colorado, show no relationship between median loads and the estimated shape parameters. In such cases, the shape parameter smoothing proceeds by simply modeling any geographical trends.

The shape parameter is smoothed using the RGAM approach described in detail in Chapter 7. The regional models adopt the following form using the median annual ground snow load $p_g^{(\text{med})}$:

$$E(\xi|p_g^{(\text{med})}) = \beta_0 + \beta_1 \log(p_g^{(\text{med})} + 1) + f_s\left(\text{LON}, \text{LAT}\right) \tag{6.4}$$

where $f_s\left(\text{LON}, \text{LAT}\right)$ is a spatial smoothing strategy described in Chapter 7. The key model assumptions is that the shape parameter varies as a function of snow accumulation (modeled with median load), but also exhibits spatial patterns not fully explained by median loads. Figure 6.3 shows a map of the smoothed shape parameter values across the country. There are certain regions (such as the coastal Washington/Oregon, the Mid-Atlantic, Eastern Colorado, and central North/South Dakota) that have particularly heavy-tailed annual maximum ground snow load distributions. On the other hand, the Rocky Mountains, Northern Minnesota, and the New England states have lighter-tailed distributions.

After obtaining smoothed estimates of the shape parameter, the location and scale parameters of the GEV distribution are fit using constrained maximum likelihood. This strategy allows the shape of the ground snow load probability distributions to be defined regionally, but only use site-specific data to

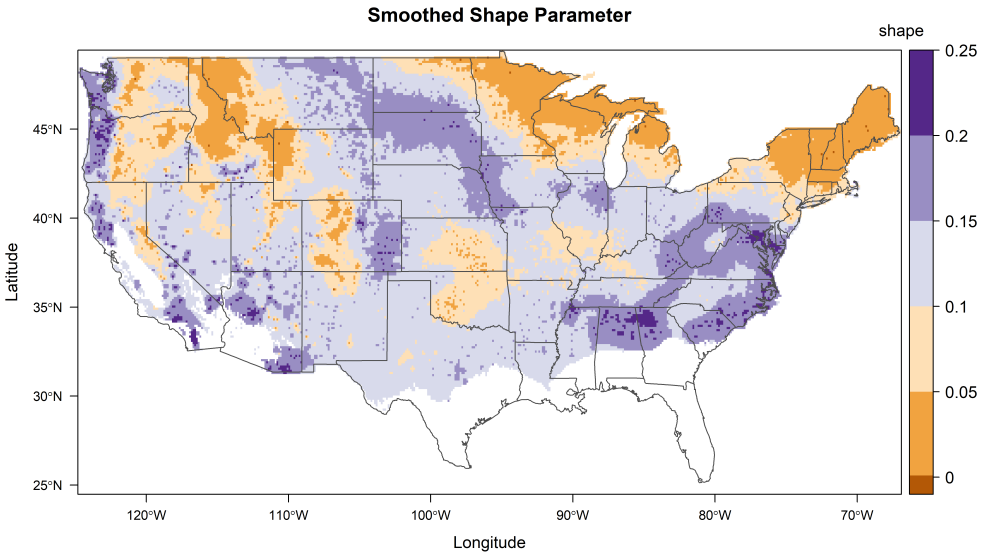

Figure 6.3: Map of smoothed shape parameter values. Larger values of the shape parameter indicate heavier-tailed probability distributions.

model the mean and variance of the snow loads.

Figure 6.5 shows that the shape parameter smoothing results in unbiased estimates of 50-year events as compared to the original distribution estimates, but produces slightly lower estimates of 50-year events than would have been obtained using tail-fit parameter estimates of a log-normal distribution. Most importantly, the shape parameter smoothing ensures consistent estimates of the distribution tails despite the size and quality limitations of the input data.

6.4.4. Practical Constraints

The smoothing strategy described in Section 6.4.3 proved effective in ensuring consistency in RTL estimates across the country. Two additional practical constraints on the shape parameters are included to ensure feasibility in design.

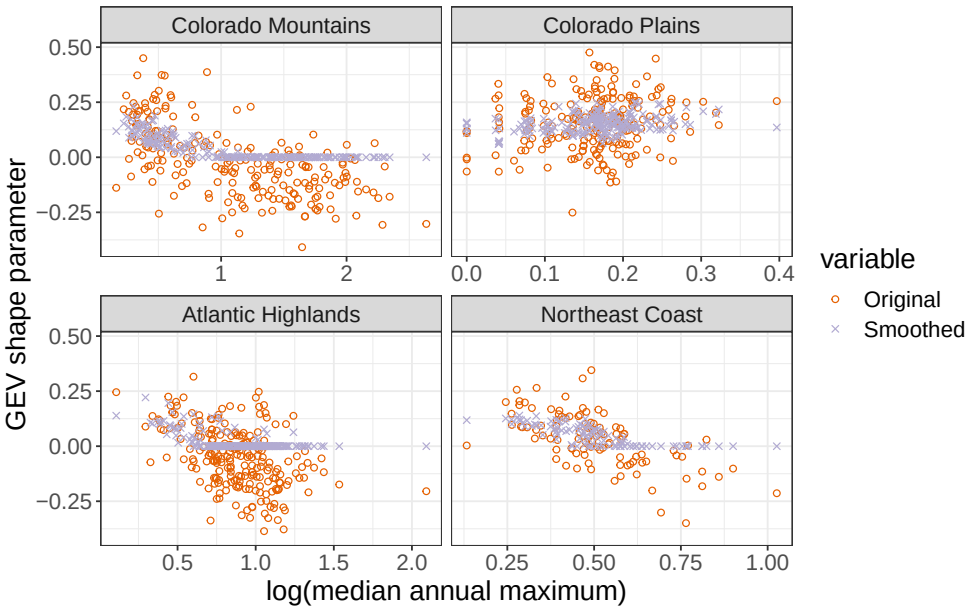

Figure 6.4: Examples of original and smoothed GEV shape parameters plotted against median annual maximum loads.

The first constraint is that smoothed shape parameters are bounded below by zero. Negative GEV shape parameters assume a finite upper bound of the simulated distribution values, which results in non-conservative estimates of reliability-targeted loads for Risk Category III and IV structures. Additionally, historical preference for the log-normal distribution means that distributions with shape parameters equal to 0 will likely already reduce loads from their currently defined requirements. Any further reductions of loads due to negative shape parameters seem unwarranted until more research is done to investigate the consequence of bounded distributions on RTL estimates.

The second constraint is that shape parameters are limited to be no larger than 0.25. This is in line with Hosking [1990] and Ragulina and Reitan [2017], who indicate that nearly all GEV distribution fits fall below 0.23 in hydrological

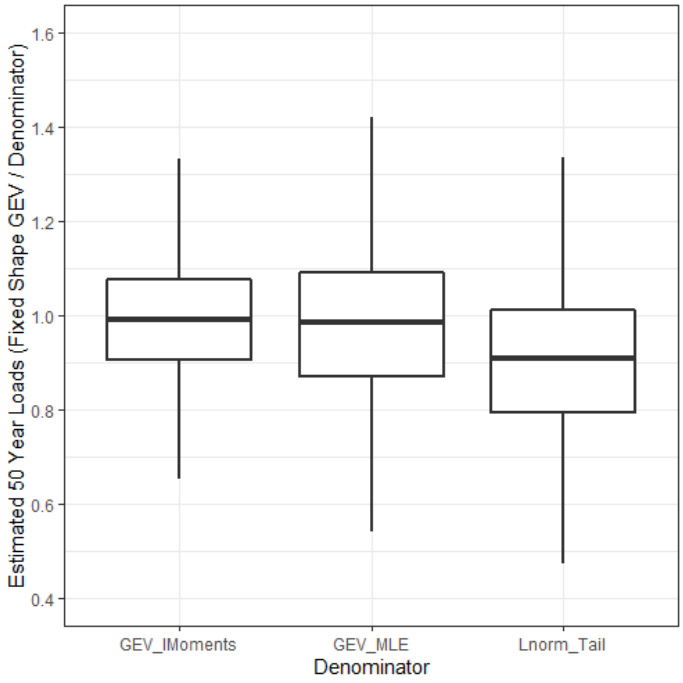

Figure 6.5: A comparison of 98th percentile estimates for different distributions fit to maximum load, data for all stations. Shape parameters are fixed using and location and scale parameters are estimated using maximum likelihood estimation.

applications. 256 of the 9715 candidate stations (2.6%) were subject to the 0.25 shape parameter cap. Empirical results suggest that shape parameters beyond this value result in untenable loads.

Finally, despite every effort to ensure high-quality distribution fits in spite of the data limitations, there are still 23 of the final 7987 stations whose RTLs were more than 3.5 times their estimated 50-year loads with a difference greater than 40 psf. These locations highlight the difficulty of site-specific RTLs calculations and are removed for practical reasons as the resulting RTLs are simply too high to be tolerated. The loss of these locations is countered by the high-quality mapping techniques described in Chapter 7 that make reasonable

inferences of design loads in the absence of the anomalous station.

6.5. Considerations for "No-Snow" Years

All distribution parameters are estimated using non-zero maximum values. To account for areas with zero-valued maximums, a point mass at zero is added to the CDF [Aitchison, 1955]. This point mass is proportional to the number of years with zero-valued maximums (n_0) passing coverage filters divided by the total number of years (i.e. $p_0 = \frac{n_0}{N}$). This is modeled by the CDF $F'(x) = p_0 + (1 - p_0)F(x), x > 0$, where $F(x)$ represents the CDF with parameters estimated using only non-zero snow years. This representation can be used to adjust the estimated quantiles during Monte-Carlo simulations. For a given $x_k > 0$ and $F'_(x_k) = p_k$, the effective quantile for the non-zero portion of the distribution is calculated as

$$F(x) = \frac{p_k - p_0}{1 - p_0}.$$

Thus, the 98th percentile of a site with 50% zero-valued maximums is estimated using the $\frac{0.98-0.5}{1-0.5} = 0.96$, or 96th percentile of the distribution fit only to the non-zero maximums. The consideration of zero-valued snow years, recommended by Buska et al. [2020], avoids bias at the nearly 50% of measurement locations in the final dataset recording at least one zero-valued snow year.

There are some locations with such high proportions of zero-valued snow years that there are simply not enough non-zero observations to fit a probability distribution. Simply defining the RTLs as being exactly equal to zero is not appropriate as virtually all locations in the United States have received some snow, including Florida [SCEC, 2020]. In order to ensure smooth transi-

tions between "low-snow" and "no-snow" locations, it is imperative to obtain reasonable (albeit small), non-zero estimates of RTLs at these locations. Such locations are the motivation for the creation of the Tier 3 measurement locations (see Chapter 6), which have:

- More than 30 years of observations.
- More than 80% of the recorded maximums are zero.
- Are not already a Tier 1 or 2 station.

For these locations, the clustering threshold described in Chapter 4 is increased from $d = 4$ to $d = 20$. The newly formed clusters are only retained if they do not include any measurement locations already being considered as Tier 1 or 2 stations.

Tier 3 locations are located in areas that receive hardly any snow over large geographical areas. The uniformity in snow conditions in such regions allows for more aggressive combinations of measurements to overcome the small sample size constraints. Such combinations are only appropriate in areas where the lack of snow is widespread, which is why Tier 3 station combinations are only performed in Level III ecoregions where at least 25% of the stations were Tier 3 stations. For the qualifying ecoregions:

1. Combine all Tier 3 measurements in qualifying ecoregions to create a single super-station.
2. Determine a single annual maximum snow load for each year by:
   - Retain all non-zero snow load maximums within the combined records.
   - When more than one non-zero maximum exists for a given year, take the median of the non-zero maximums as the representative measurement for the year.

3. Fit a gamma distribution to the non-zero maximums resulting from the previous step. Due to its shape, the gamma distribution more naturally characterizes values that are arbitrarily close to zero and only requires the estimation of two parameters instead of three.
4. Use the resulting gamma distribution parameters for all Tier 3 measurement locations within the ecoregion, but use the site-specifics estimate of the proportion of zero-valued snow years.

This strategy prevents spurious extrapolations of large snow loads due to the inevitable instability in site-specific distribution fits that would result from small sample sizes. The Tier 3 distribution fits are combined with the Tier 1 and 2 station fits to provide appropriate transitions in RTLs from "low snow" to "no snow" regions. These transitions are accomplished via the mapping scheme described in Chapter 7.

Bibliography

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.

BURN, D. H. (1990). An appraisal of the "region of influence" approach to flood frequency analysis. Hydrological Sciences Journal, 35(2):149–165.

Buska, J. S., Greatorex, A., and Tobiasson, W. (2020). Site specific case studies for determining ground snow loads in the United States. Technical report, Engineer Research and Development Center, Hanover, NH. Accessed: 11-30-2020.

Cho, E. and Jacobs, J. M. (2020). Extreme value snow water equivalent and snowmelt for infrastructure design over the contiguous united states. Earth and Space Science Open Archive, page 40.

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

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.

Ellingwood, B., Galambos, T. V., MacGregor, J. G., and Cornell, C. A. (1980). Development of a probability based load criterion for American National Standard A58: Building code requirements for minimum design loads in buildings and other structures, volume 13. US Department of Commerce, National Bureau of Standards.

Ellingwood, B. R. (2000). Lrfd: implementing structural reliability in professional practice. Engineering Structures, 22(2):106–115.

Feng, S., Nadarajah, S., and Hu, Q. (2007). Modeling annual extreme precipitation in china using the generalized extreme value distribution. Journal of the Meteorological Society of Japan. Ser. II, 85(5):599–613.

Gumbel, E. J. (2004). Statistics of extremes. Courier Corporation.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B (Methodological), 52(1):105–124.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27(3):251–261.

Lee, K. H. and Rosowsky, D. V. (2005). Site-specific snow load models and hazard curves for probabilistic design. Natural Hazards Review, 6(3):109–120.

Martins, E. S. and Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research, 36(3):737–744.

Meehleis, K., Folan, T., Hamel, S., Lang, R., and Gienko, G. (2020). Snow load calculations for alaska using ghcn data (1950–2017). Journal of Cold Regions Engineering, 34(3):04020011.

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

Panagoulia, D., Economou, P., and Caroni, C. (2014). Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. Environmetrics, 25(1):29–43.

Ragulina, G. and Reitan, T. (2017). Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the bayesian approach. Hydrological Sciences Journal, 62(6):863–879.

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.

SCEC (2020). State climate extremes. https://www.ncdc.noaa.gov/extremes/scec/records.

Tobiasson, W., Buska, J., Greatorex, A., Tirey, J., and Fisher, J. (2002). Ground snow loads for New Hampshire. Technical report, Cold Regions Research and Engineering Laboratory.

Tobiasson, W. and Greatorex, A. (1997). Database and methodology for conducting site specific snow load case studies for the United States. In Proc., 3rd Int. Conf. on Snow Engineering, Izumi, I., Nakamura, T., and Sack, RL, eds., AA Balkema, Rotterdam, Netherlands, pages 249–256.

Chapter 7
Mapping Reliability-Targeted Design Ground Snow Loads

7.1. Introduction

In addition to the move to reliability-targeted design ground snow loads (RTLs), this research aims to drastically reduce the number and size of case study regions in the United States. This requires high quality estimates of RTLs between the measurement locations to create continuous maps of requirements. Newly mapped values rely upon the 7,987 site-specific RTLs computed in Chapter 6 as input. This chapter describes efforts to create mapping techniques that are:

- **Accurate:** Mapped values should closely reflect the input data, without over-fitting the input data.
- **Adaptive:** The relationship between ground snow load and explanatory variables such as elevation changes regionally. Mapping approaches should account for these non-constant trends.
- **Smooth:** Small changes in location and/or elevation should result in proportionally small changes to the estimated load.
- **Scalable:** Predictions should be computationally feasible on standard com-

puters to facilitate reproducibility.

Of all mapping approaches considered, the best method for achieving the listed objectives was a regionalized adaptation of generalized additive models (GAMs). GAMs fit smooth trends between explanatory and response variables without having to specify a particular model form. GAMs also seamlessly model spatial trends in snow loads not accounted for by other variables such as elevation. Unique GAMs were fit to site-specific RTLs within each of the Environmental Protection Agency's (EPA) ecoregions and use a buffering approach to smooth the mapped values between regions. The resulting regionalized GAMs (i.e. RGAMs) create accurate, high resolution snow load maps that drastically reduce the number of case study regions and eliminate the discrepancies in load requirements that currently exist along the borders of many western states.

Chapter Highlights:

- A brief summary of previous mapping approaches.
- A description of the RGAM mapping approach, including region-specific examples.
- Comparisons of accuracy between the new and previous mapping approaches.

7.2. Previous Methods

The number of locations with sufficiently long histories of snow depth/load measurements is sparse relative to the number of locations requiring design snow load estimates. This issue is almost always addressed by estimating de-

sign loads between measurement locations using mapping techniques. Perhaps the most common mapping approach is inverse distance weighting, where predictions at any location on the map are a weighted average of the surrounding measurement locations with preference given to locations that are closer to the prediction location [Shepard, 1968]. This approach remains a popular approach [Lu and Wong, 2008] and is representative similar interpolation approaches that seek to fit the input data exactly. These interpolation approaches leverage the inuitive spatial assumption that observations located close to each other in space tend to be more similar than observations that are far away from each other.

Another popular set of mapping approaches are regression-based models such as PRISM [Daly et al., 2008]. These models also account for the similarities between observations due to location, but do not try to fit the input data exactly. Rather, these models try to model the changing relationships between the response and explanatory variables over space. There are advantages and disadvantages to both interpolation and regression approaches, but one key consideration in favor of regression approaches are that the site-specific RTLs are estimates, not observations. Fitting the input RTLs exactly can lead to unreasonably sharp changes in mapped values over short geographical distances, even within the boundaries of a single municipality. In contrast, regression approaches smooth over the uncertainties present in the RTL values, while still respecting the rapid changes in load that can occur due to changes in elevation or climate.

There is a rich history of interpolation and regression approaches for mapping design ground snow loads in the United States. The current ASCE 7 snow loads are based on studies performed at the Cold Regions Research and Engi-

neering Laboratory (CRREL) ca. 1980 and updated ca. 1993. These maps focus on defining loads for most populated locations, but label many topographically complex locations as "case-study regions." Guidance for conducting case study regions is provided by Tobiasson and Greatorex [1997] and more recently in Buska et al. [2020], though many western states have elected to define snow load requirements through state-level studies using a wide variety of mapping techniques (see Sack [2015] for a relatively comprehensive review).

The states of Idaho [Al Hatailah et al., 2015], Montana [Theisen et al., 2004], and Washington [Sack, 2015] use interpolation based approaches that use normalized ground snow loads (NGSL) Sack and Sheikh-Taheri [1986] to account for the effect of elevation. In contrast, the states of Colorado DeBock et al. [2016], Utah Bean et al. [2018], Oregon SEAO [2013], and New Hampshire Tobiasson et al. [2002] employ regression based approaches to account for the effect of elevations on design loads. Each approach acknowledge the strong spatial dependencies among observations that cannot be explained solely by elevation. For this reason, each of the referenced reports attempt to account for both elevation and spatial/climate effects in design load estimations.

While the NGSL approach has proven popular, Bean et al. [2019] illustrated the difficulties of using NGSLs to account for the effect of elevation in the state of Utah. The difficulty arises from the changing relationship between ground snow loads and elevations in different states. For example, the relationship between RTLs and elevation is log-linear in the state of Wyoming, linear in Maryland, and virtually non-existent in Ohio (see Figure 7.1). NGSL's work well for linear effects, but poorly for non-linear effects. In order to effectively map loads nationally (including current case study regions), it is crucial to employ a mapping technique that can adaptively model the relationship

between ground snow loads and elevation (or any other potential explanatory variables). Further, new approaches are required to appropriately employ these state-specific mapping approaches on a national scale.

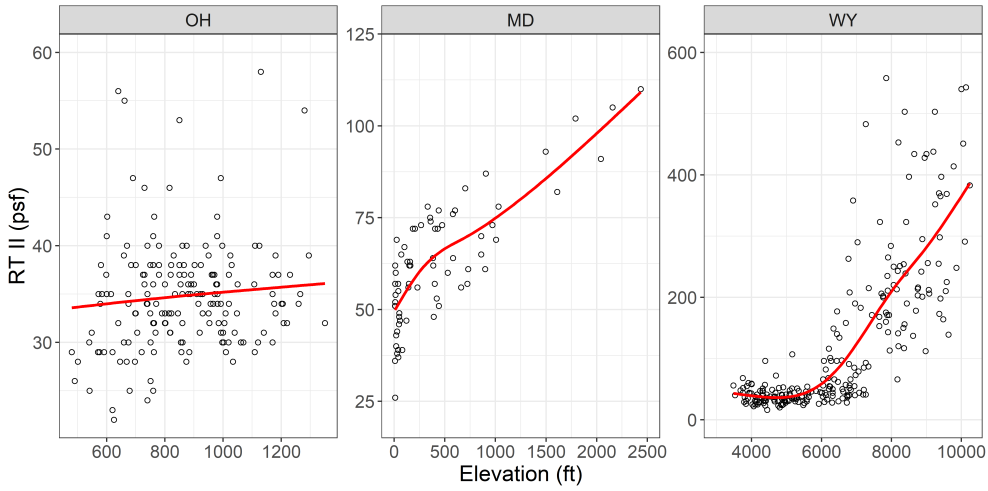

Figure 7.1: Comparison of the relationship between RTLs (Risk Category II) and elevation in Ohio (OH), Maryland (MD), and Wyoming (WY).

To address this need for a new national mapping approach, the authors have created an adaptive mapping technique called regional generalized additive models (RGAMs) that map RTLs between measurement locations. The remainder of the chapter is devoted to describing the data and methodology underlying the RGAM approach.

7.3. Incorporating Climate Data

The core data of the RGAM approach are RTLs defined at the nearly 8,000 measurement location. Available meta-data for each location includes its geographical coordinates and elevation. While elevation is a strong predictor of

snow loads in many western states, other variables such as temperature prove to be better predictors of design loads in many eastern states.

Recent advances in streaming data mechanisms have led to the rise of gridded (i.e. mapped) climate products produced by the PRISM climate group [Daly et al., 2002, 2008]. These maps provide daily, monthly, or 30-year averages of climate-related measurements, such as temperature and precipitation, and were used previously in Chapter 5. The state of Oregon took advantage of these gridded data by replacing elevation with custom PRISM output as the explanatory variable in their most recent design ground snow load predictions [SEAO, 2013]. This is accomplished by matching the measurement locations with their mapped climate variables using the measurement location coordinates. Other uses of gridded climate data include a recent attempt by Cho and Jacobs [2020] to define 50-year ground snow loads using output from the Snow Data Assimilation System (SNODAD) maintained by the National Operational Hydrologic Remote Sensing Center [NOHRSC, 2004]. This attempt defines 50-year loads for each grid and entirely circumvents the use of traditional measurement locations.

The mapping approach described in this chapter uses elevation as the primary explanatory variable for predicting RTLs. In addition to elevation, the maps make use of the PRISM climate variables:

- 1981-2010 Mean Temperature of the Coldest Month
- 1981-2010 Mean Annual Winter Precipitation (December - February).

The three variables elevation, temperature, and winter precipitation often explain large proportions of the variability in RTLs, though their influence changes drastically from region to region. There are also spatial patterns in RTLs that these explanatory variables do not fully explain. For this reason,

the RGAMs described in the following section include a regional subset strategy that accounts for the ever-changing influence of the predictor variables, as well as a spatial modeling step that accounts for local variability in RTLs left unexplained by other climate variables.

7.4. Generalized Additive Models

GAMs provide a framework for generalizing ordinary least squares (OLS) regression models to account for non-linear effects. The method can be represented as

$$E(p_g|\boldsymbol{x}) = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \ldots + f_p(x_{ip}) \tag{7.1}$$

where p_g represents the reliability-targeted design ground snow load and $\boldsymbol{x}$ represents the potential explanatory variables such as elevation and temperature. There are a variety of different approaches for fitting GAMs, but typically each smooth term $f_k()$ is estimated using penalized regression splines with smoothing parameters that are selected using some form of cross-validation. These smoothing parameters control the smoothness of each term in the model. The cross-validation approach automatically calibrates the smoothing terms to generalize to new data. This drastically reduces the number of parameters that need to be defined prior to modeling, which makes the approach more objective than comparable approaches.

To account for the spatial variability in RTLs left unexplained by other climate variables, a geographic smoothing spline is added to Equation 7.1. The final model used to estimate RTLs is given as 7.2.

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \ldots + f_p(x_{ip}) + f_s(\text{LON}, \text{LAT}) + \epsilon_i \qquad (7.2)$$

where f_s is modeled using a "splines on the sphere" approach [Wahba, 1981, Wood, 2003]. This additional term models spatial patterns not explained by other explanatory variables.

7.5. The Regional Smoothing Approach

The GAM modeling approach is effective at characterizing non-linear trends between RTLs and the variables elevation, temperature, and precipitation, but there still exists the need for a way to allow the estimated trends to vary regionally. The spatial smoothing term described in Equation 7.2 is not fully adequate in explaining continental-scale differences in RTLs. Rather there is the need for separate models to be defined for different regions of the country. The main issue with regional models is the inevitable discrepancies in predictions that occur along region boundaries. Such is the case currently along the boundaries of western states [Sack, 2015].

To address the boundary issues, the authors developed the following regional smoothing approach.

1. Partition the country into well-defined regions. This is accomplished using the Environmental Protection Agency's (EPA) Level III ecorergions [CEC, 1997], which are areas that are regarded as having similar climate and ecological characteristics. Figure 7.2 shows an example of the level III ecoregions in the state of Colorado, though these ecoregions pay no

respect to political boundaries such as state borders. Figures 7.3 and 7.4 provide examples of the different trends between RTLs and climate variables depending on the eco-region.

2. Fit a regional GAM for all observations in a level III ecoregion as well as all observations within 30 miles of the boundaries of the ecoregion (buffer zone #1). Figure 7.5 shows an example of included stations within and near an ecoregion boundary.
   - To ensure reliable trend estimates, at least 150 observations are required to fit a GAM model within an ecoregion. If this is not automatically satisfied, the buffer zone of 30 miles is increased until 150 observations are in range.
3. Make predictions on a 0.5 mile resolution grid for all locations in the ecoregion, as well as those within 15 miles of the ecoregion boundary (buffer zone #2).
4. Smooth predictions by taking a weighted average of ecoregion model predictions in grid cells with predictions from two or more ecoregions due to the second buffering. See Section 7.5.1 for details regarding the weighted average calculation.

The described algorithm has the precision that comes with local modeling, without the undesirable sharp boundary changes that normally come with regional models.

7.5.1. Weighted Averaging Approach

Given a location x, RTL predictions $y_1, y_2, ..., y_m$ from models corresponding to ecoregions 1 through m are obtained. Let $d_1, d_2, ..., d_m$ represent the shortest

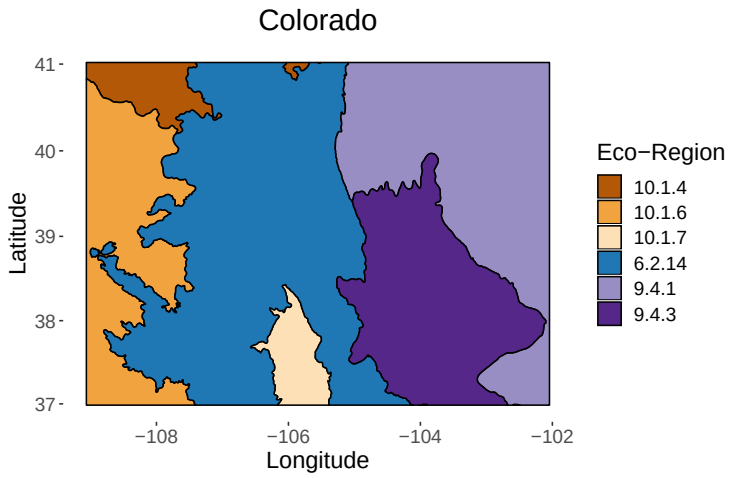

Figure 7.2: An illustration of level III ecoregions in the state of Colorado.

distances between location x and the boundaries of ecoregions 1 through m respectively. If x is located within the boundaries of an ecoregion, the distance between x and that ecoregion is zero. Given an arbitrary ecoregion j, the weight w_j given to y_j is non-zero when d_j is within some threshold S (Equation 7.3).

$$w_j = \begin{cases} \left(\frac{S-d_j}{S}\right)^2 & d_j \leq S \\ 0 & d_j > S \end{cases} \tag{7.3}$$

A final prediction y' for location x is calculated by Equation 7.4:

$$y' = \frac{w_1 y_1 + w_2 y_2 + ... + w_m y_m}{w_1 + w_2 + ... + w_2} \tag{7.4}$$

When a prediction is made in ecoregion j and the prediction location is further than S units from any ecoregion border, then $d_j = 0$ and Equation 7.4 reduces to $y' = y_j$. As predictions in ecoregion j approach the border of ecoregion k, then the weight of y_k increases gradually and $y' = \frac{y_j + ((S-d_k)/S)^2 y_k}{1+((S-d_k)/s)^2}$. At

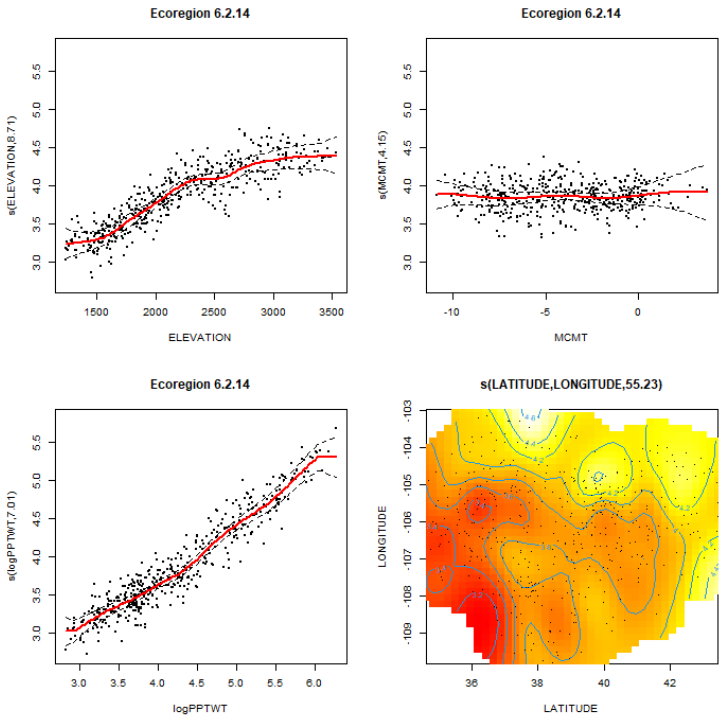

Figure 7.3: Log of RTL event vs. variables used in GAM for ecoregion 6.2.14. The points shown in each plot represent partial residuals, which are the residuals that would be obtained by dropping the term concerned from the model while leaving all other estimates fixed.

the border of ecoregions j and k, $y = \frac{y_j + y_k}{2}$. Finally, as predictions progress into ecoregion k, weights for y_j decrease gradually to zero and $y' = \frac{((S-d_j)/S)^2 y_j + y_k}{((S-d_j)/S)^2 + 1}$. Figure 7.6 is a simple example of smooth transition between three different regions given a constant predicted value for each region.

7.6. Cross Validated Results

The efficacy of using GAMs rather than alternative modeling techniques is evaluated by means of cross validation. Ten-fold cross-validation involves randomly

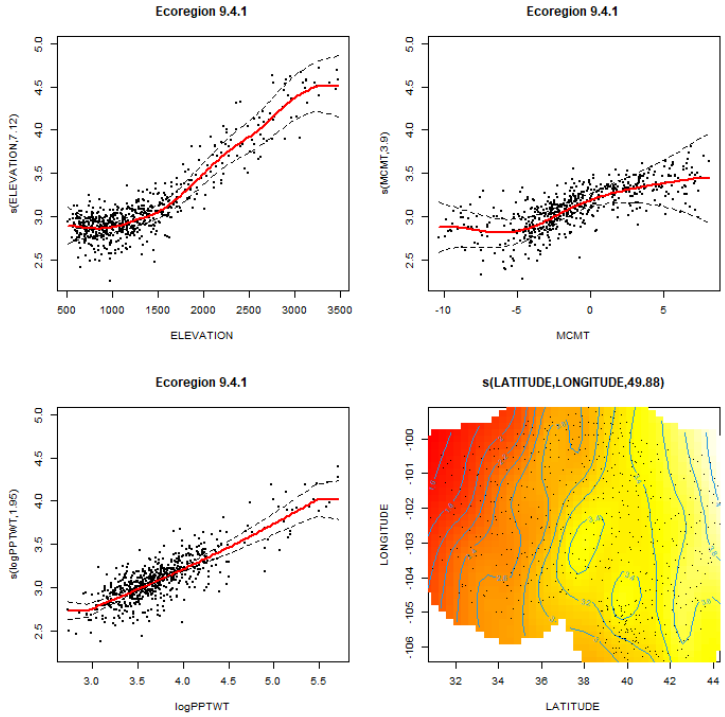

Figure 7.4: Log of RTL event vs. variables used in GAM for ecoregion 9.4.1. The points shown in each plot represent partial residuals, which are the residuals that would be obtained by dropping the term concerned from the model while leaving all other estimates fixed.

separating the data into ten groups then using nine of the ten groups to fit the model to then make predictions on the tenth group. This process is repeated ten times, each time withholding a different group of observations, refitting the model with the remaining observations, and evaluating the difference between the actual and predicted values. The process of removing observations helps to determine how well the model will generalize to new data and discourages models that fit the input data closely, but generalize poorly. A spatial variant of cross validation was also attempted [Meyer et al., 2019], which considers the geographic distribution of the locations when forming model groups, though

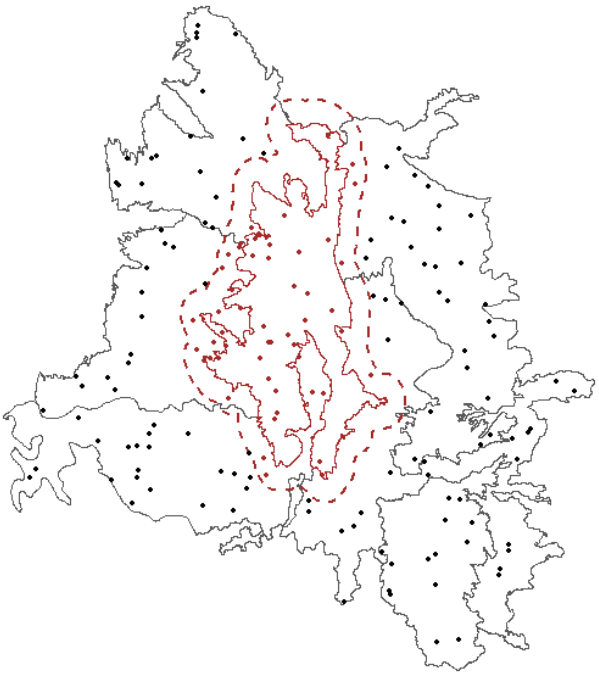

Figure 7.5: Example of a buffer zone being applied to an ecoregion to determine qualifying stations for a regional model fit.

the spatial variant of the method yielded nearly identical results.

Table 7.1 shows the results from several spatial modeling approaches included traditional regression, kriging with an external drift [Goovaerts, 1997, Bean et al., 2019], PRISM [Bean et al., 2017], and inverse distance weighting [Al Hatailah et al., 2015]. The regional smoothing approach described in the previous section improved the accuracy of all considered models, though the RGAM models stood out as the models having the lowest errors across every considered metric.

Because the RGAMs model performed the better than any of the other models considered, the cross-validated error rates of this model are considered

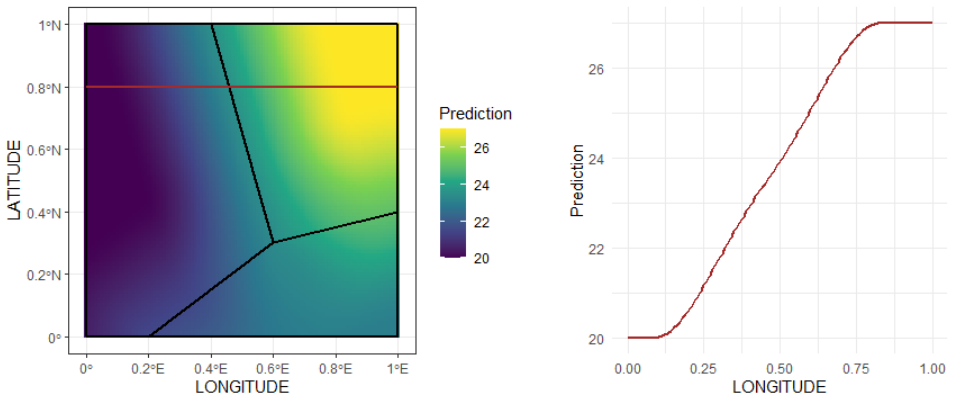

Figure 7.6: Border smoothing example. Each region gives a constant valued prediction, then the borders are smoothed as described in Equation 7.4. The red line is predicted values at 0.8°N.

in depth. In particular, summarized values of the spatial cross-validation of the globally smoothed GAMs model are displayed for each ecoregion in which data are available. Figure 7.7 shows that mapped RTLs (divided by 1.6 so as to be comparable to current ASCE 7 requirements) tend to be within 4 psf of the site-specific values for the vast majority of the country, with accuracy slightly worse in intermountain states. Further, Figure 7.8 shows that the mean relative errors are within 2%, on average, for virtually all of the country with the exception of some slight biases (blue represents under-predictions and red represents over-predictions) in the Cascade mountains, western deserts, and areas with exceptionally small RTL values (such as Southern Texas). This demonstrates the efficacy of the RGAM approach to maintain accuracy in mountainous regions, allowing for the near elimination of case study regions.

Table 7.1: Standard cross-validated results on RTL.

Model	Fitting Technique	MAE	MedAE	MSE
GAM	National Scale	8.51	3.2	504
GAM	**Locally Smoothed**	**6.43**	**2.35**	**235**
OLS	National Scale	16.8	6.53	1810
OLS	Locally Smoothed	8.44	3.35	361
Kriging	National Scale	15.3	5.74	1280
Kriging	Locally Smoothed	9.24	2.94	553
Prism	National Scale	8.33	3.12	518
Prism	Locally Smoothed	6.94	2.65	272
IDW	National Scale	28.9	15.6	2200
IDW	Locally Smoothed	18.3	6.3	1480

7.7. Implications and Future Work

Figure 7.9 show maps of the relative difference between the new maps and the current maps provided in ASCE 7-16, excluding western states with state-specific standards that have been adopted in ASCE 7-16. Many current ASCE 7-16 snow load zones have different prescribed loads for different layers of elevation. The mapped comparisons in Figure 7.9 only compare to the primary ASCE 7-16 load, which explain the large relative increases design loads in the Appalachian Mountains. Many of the difference between current and design-loads are a result of the move to RTLs and are not due to differences in the mapping approach.

Future work may consider the use of different climate variables besides temperature and winter precipitation for making predictions. Additional efforts may also be devoted to understand the rate of increase in design snow loads in areas with highly volatile elevations. For example, large changes in design loads were noted along the benches of the municipalities of Missoula, MT and Park City, UT. While increases in design loads are expected in mountain bench

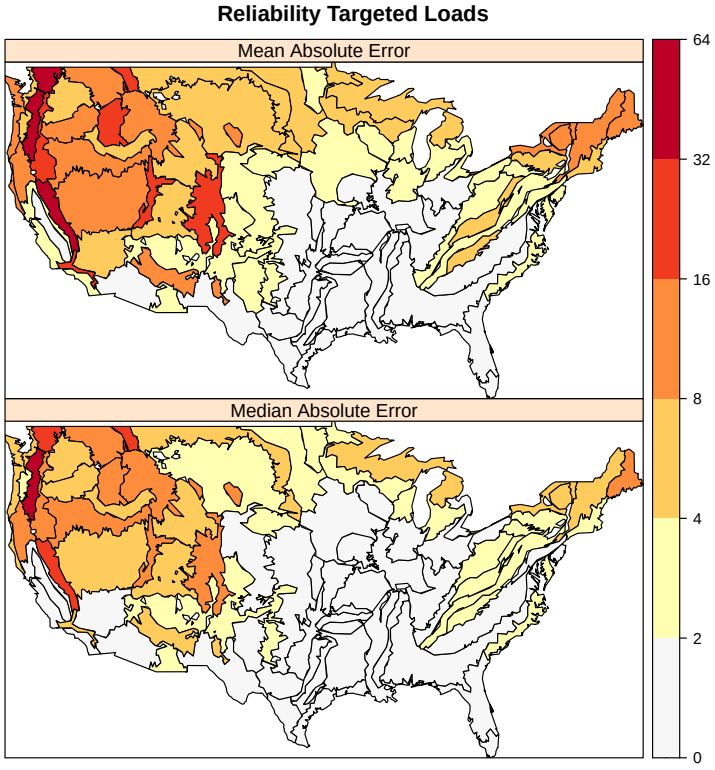

Figure 7.7: Mean and median absolute errors for each ecoregion (showing error magnitude).

neighborhoods, greater scrutiny could be devoted to ensuring that the rate of increase in these unique situations is consistent with expectations given local knowledge. Regardless, the newly proposed RGAM models play a key role in eliminating the case study regions that current exist in the ASCE 7-16 design ground snow load maps.

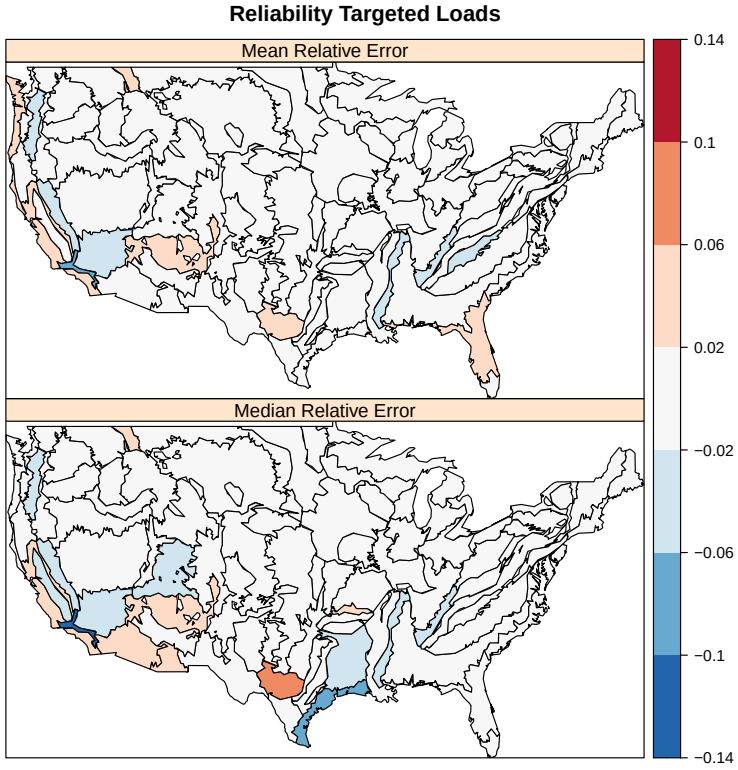

Figure 7.8: Mean and median relative error for each ecoregion. The relative error is calculated as (Predicted - Actual) / (Predicted + Actual).

Bibliography

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. Accessed: 12-1-2020.

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 4591, Utah State University, Department of Civil and Environ-

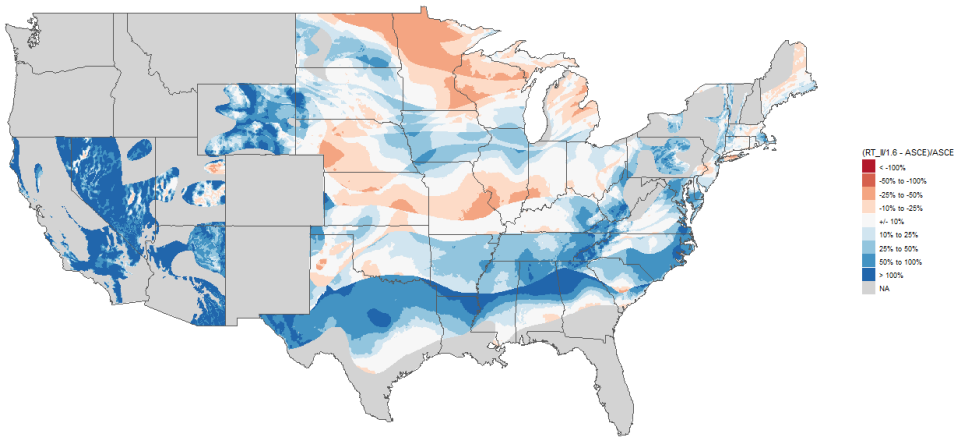

Figure 7.9: Comparison of the relative difference between mapped RTLs and current ASCE 7 requirements.

mental 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.

Buska, J. S., Greatorex, A., and Tobiasson, W. (2020). Site specific case studies for determining ground snow loads in the United States. Technical report, Engineer Research and Development Center, Hanover, NH. Accessed: 11-30-2020.

CEC (1997). Ecological regions of North America: toward a common perspective. Technical report, Commission for Environmental Cooperation.

Cho, E. and Jacobs, J. M. (2020). Extreme value snow water equivalent and

snowmelt for infrastructure design over the contiguous united states. Earth and Space Science Open Archive, page 40.

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

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

Lu, G. Y. and Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. Computers and Geosciences, 34(9):1044 – 1055.

Meyer, H., Reudenbach, C., Wöllauer, S., and Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. Ecological Modelling, 411:108815.

NOHRSC (2004). Snow data assimilation system (snodas) data products at nsidc version 1. https://doi.org/10.7265/N5TB14TC. Accessed: 8-1-2020.

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. and Sheikh-Taheri, A. (1986). Ground and roof snow loads for Idaho. University of Idaho, Department of Civil Engineering.

SEAO (2013). Snow load analysis for Oregon. Structural Engineers Association of Oregon, Portland, OR, fourth edition.

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

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., Buska, J., Greatorex, A., Tirey, J., and Fisher, J. (2002). Ground snow loads for New Hampshire. Technical report, Cold Regions Research and Engineering Laboratory.

Tobiasson, W. and Greatorex, A. (1997). Database and methodology for conducting site specific snow load case studies for the United States. In Proc., 3rd Int. Conf. on Snow Engineering, Izumi, I., Nakamura, T., and Sack, RL, eds., AA Balkema, Rotterdam, Netherlands, pages 249–256.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. SIAM Journal on Scientific and Statistical Computing, 2(1):5–16.

Wood, S. N. (2003). Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(1):95–114.

Chapter 8
Conclusions

This report has summarized the efforts of the 2020 National Snow Study to :

1. Significantly reduce the number of case study regions through a modern, universal, and reproducible approach for generating design ground snow loads for the conterminous United States.
2. Directly estimate reliability-targeted design ground snow loads (RTLs) for each Risk Category, resulting in both a reduction of the snow load factor from 1.6 to 1.0 and the elimination of importance factors.

This effort quantified the effect that changes to design provisions, as well as an evolving understanding of the distributions of resistance members and snow loads, have on the original load factor calibrations. Additionally, the move to direct predictions of RTLs identified the influence that snow accumulation patterns have on the difference between the 50-year snow load and the RTL. Locations whose peak snow loads are characterized by a few, large storms tend to have larger design snow loads than those currently defined in ASCE 7. Conversely, locations whose peak snow loads are characterized by the accumulation of many storms throughout the the snow season tend to have lower requirements than those currently defined.

The pursuit of a uniform method for estimating RTLs resulted in novel approaches for estimating snow load from snow depth (Chapter 5, leverag-

ing information at surrounding locations to improve distribution tail estimates (Chapter 6), and smoothing mapped values across a partition of regions (Chapter 7). These methods were designed to be reproducible, and the computer code underlying each step is available upon request. This framework allows for quick updates to estimated values as improved information becomes available with little marginal cost.

Tables 8.1 and 8.2 compare the new and current design snow load requirements for Risk Category II buildings with heated flat roofs in normal exposure conditions in cities across the country. These cities match those explored in Lee and Rosowsky [2005], though western state locations have been omitted since their ASCE 7-16 design snow loads are derived from state-specific studies. Note that the current requirements in these two tables are obtained by multiplying the 50-year snow load available in ASCE 7-16 by 1.6. Loads rose the most mid-latitude areas whose typical winters have little snow, but whose extreme winters have substantial snow. Loads fell slightly in areas that consistently experience high snow load winters every year. Figure 8.1 shows a boxplot of the ratio between the new and current requirements at these 65 locations. The average ratio is 1.12 with a standard deviation of 0.26, indicating a modest rise, on average, in design snow load requirements. The modest increase in design loads is consistent with expectations based on changes to design provisions since the original calibration.

The new snow load maps reduce the number and size of case study regions by 91% from what they were in ASCE 7-16 and 96% of what they were in ASCE 7-10. The remaining "case-study regions" have elevations exceeding all measurement locations and are virtually devoid of structures. This substantially reduces the burden, disproportionately carried by the topographically complex

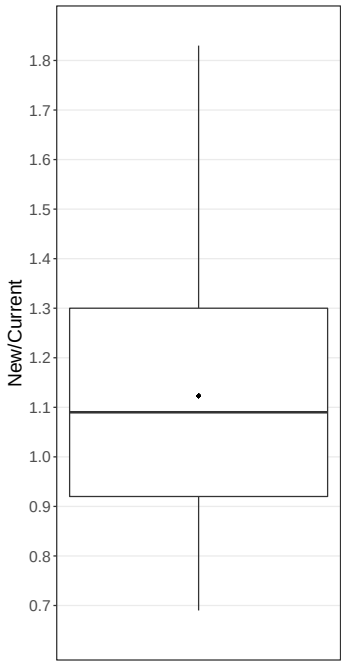

Figure 8.1: Boxplot of the ratio between new and current requirements at the 65 locations specified in Tables 8.1 and 8.2.

western states, of specifying design load requirements in the previously defined case study regions. A natural benefit of this effort is the elimination of the discrepancies in design load requirements that exist between the independently developed state-specific studies.

This research effort owes its success to the many state and national studies that preceded it. Many of the authors of those previous studies were directly involved in the steering committee that collaborated on this effort. Their collective knowledge and experience, coupled with the computational abilities of modern statistical software, result in a new, uniform, and reproducible set of design snow load requirements for the conterminous United States.

Bibliography

Lee, K. H. and Rosowsky, D. V. (2005). Site-specific snow load models and hazard curves for probabilistic design. Natural Hazards Review, 6(3):109–120.

Table 8.1: Comparison of new and current design ground snow load requirements for Risk Category II buildings in the United States.

Location	**New**	**Current**	**Ratio**
Bridgeport, CT	42	48	0.88
Hartford, CT	50	56	0.89
Washington, DC	61	40	1.52
Des Moines, IA	45	40	1.12
Dubuque, IA	53	48	1.1
Sioux City, IA	65	48	1.35
Waterloo, IA	49	48	1.02
Chicago, IL	53	40	1.32
Moline, IL	43	32	1.34
Peoria, IL	33	32	1.03
Rockford, IL	52	40	1.3
Springfield, IL	28	32	0.88
Evansville, IN	22	24	0.92
Fort Wayne, IN	33	32	1.03
Indianapolis, IN	29	32	0.91
Wichita, KS	23	24	0.96
Covington/Cincinnati, KY	29	32	0.91
Boston, MA	62	64	0.97
Worcester, MA	71	80	0.89
Baltimore, MD	62	40	1.55
Caribou, ME	139	160	0.87
Portland, ME	85	80	1.06
Alpena, MI	65	80	0.81
Detroit, MI	38	32	1.19
Grand Rapids, MI	58	56	1.04
Houghton Lake, MI	67	80	0.84
Lansing, MI	44	48	0.92
Sault Ste. Marie, MI	108	112	0.96
Duluth, MN	81	96	0.84
International Falls, MN	67	80	0.84
Minneapolis–St. Paul, MN	58	80	0.72
Rochester, MN	55	80	0.69

Table 8.2: Comparison of new and current design ground snow load requirements for Risk Category II buildings in the United States (continued).

Location	**New**	**Current**	**Ratio**
Bismarck, ND	72	56	1.29
Fargo, ND	62	80	0.78
Norfolk, NE	52	40	1.3
Omaha, NE	52	40	1.3
Scottsbluff, NE	33	24	1.38
Atlantic City, NJ	38	32	1.19
Newark, NJ	44	40	1.1
Reno, NV	42	24	1.75
Albany, NY	66	64	1.03
New York, NY	47	32	1.47
Rochester, NY	70	64	1.09
Akron, OH	32	32	1
Cleveland, OH	39	32	1.22
Columbus, OH	32	32	1
Mansfield, OH	37	32	1.16
Toledo, OH	35	32	1.09
Philadelphia, PA	35	32	1.09
Pittsburgh, PA	53	40	1.32
Providence, RI	49	48	1.02
Aberdeen, SD	95	80	1.19
Rapid City, SD	41	32	1.28
Sioux Falls, SD	80	64	1.25
Burlington, VT	83	64	1.3
Green Bay, WI	58	64	0.91
La Crosse, WI	46	64	0.72
Madison, WI	54	48	1.12
Milwaukee, WI	57	48	1.19
Beckley, WV	58	32	1.81
Charleston, WV	40	32	1.25
Huntington, WV	32	32	1
Casper, WY	44	24	1.83
Cheyenne, WY	46	32	1.44
Sheridan, WY	47	32	1.47

Appendix A
Relevant Software

This project was primarily completed in R 3.6 [R Core Team, 2019] with the help of the following ancillary packages.

- `gstat` [Pebesma, 2004, Gräler et al., 2016]: For kriging and inverse distance weighting.
- `maps` [code by Richard A. Becker et al., 2018]: For state and county shapefiles in visualizations.
- `mgcv` [Wood, 2003, 2004, 2011, 2017, Wood et al., 2016]: For generalized additive models.
- `randomforest` [Liaw and Wiener, 2002]: For random forest models.
- `rdgal` [Bivand et al., 2020]: For spatial projections.
- `rgeos` [Bivand and Rundel, 2020]: For coastal distance calculations.
- `sf` [Pebesma, 2018]: For spatial distance calculations.
- `sp` [Pebesma and Bivand, 2005, Bivand et al., 2013]: For reprojections of spatial data.

Bibliography

Bivand, R., Keitt, T., and Rowlingson, B. (2020). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.5-18.

Bivand, R. and Rundel, C. (2020). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-5.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). Applied spatial data analysis with R, Second edition. Springer, NY.

code by Richard A. Becker, O. S., version by Ray Brownrigg. Enhancements by Thomas P Minka, A. R. W. R., and Deckmyn., A. (2018). maps: Draw Geographical Maps. R package version 3.3.0.

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

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. R News, 2(3):18–22.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1):439–446.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. Computers and 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 (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Wood, S. (2017). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, 2 edition.

Wood, S., N., Pya, and S"afken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). Journal of the American Statistical Association, 111:1548–1575.

Wood, S. N. (2003). Thin-plate regression splines. Journal of the Royal Statistical Society (B), 65(1):95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association, 99(467):673–686.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B), 73(1):3–36.