Reliability-Based Design Snow Loads: I. Site-Specific Probability Models for Ground Snow Loads

D. Jared DeBock, PhD. Department of Civil Engineering, California State University, Chico. ddebock@csuchico.edu.

Abbie B. Liel, PhD, PE. Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder. Abbie.Liel@colorado.edu.

James R. Harris, PhD, PE, SE, NAE. J. R. Harris and Company, Denver, CO.

Bruce R. Ellingwood, PhD, PE, NAE. Department of Civil and Environmental Engineering, Colorado State University, Ft. Collins.

Jeannette M. Torrents, PE, SE, LEED AP. JVA, Incorporated, Boulder, CO. jtorrents@jvajva.com.

Abstract: This paper describes a new method for fitting probability distributions for modeling annual maximum ground snow loads for use in structural design. These probability models are intended for use in reliability assessments to determine reliability-targeted ground snow loads, as described in the companion paper. The proposed method emphasizes the upper tail of the distributions, because the upper tail is most critical for the reliability assessment and for determination of design loads. A combination of site-specific and region-of-influence approaches results in annual maximum snow load distributions whose magnitudes are consistent with historical data observed at the site of interest, but with upper tail shapes that are informed by historical snow records at a number of similar sites. Clusters of sites with similar snow accumulation patterns are utilized to improve the definition of the snow load distribution at sites in the cluster.

Introduction

The standard practice for defining ground snow loads for design of buildings is based on the concept of uniform hazard. In the *ASCE Standard 7-10* (ASCE 2010), for example, design snow loads are based on ground snow loads with a 2% annual probability of exceedance, *i.e.* ground snow loads with a 50-year mean recurrence interval (MRI) at all locations. Although different standards have used different mean recurrence intervals, this uniform hazard approach can be found in U.S., Canadian, Chinese and European codes (ASCE 2010, NRCC 2010, Ministry of Housing and Urban-Rural Development of the People's Republic of China 2012, BSI 2003), among others. To determine the uniform hazard loads for design, previous researchers have fitted probability distributions to historical snow weight or depth data obtained from weather stations (*e.g.* Ellingwood and Redfield 1983, Newark et al. 1989, Durmaz and Daloglu 2006, SEAC 2007, Blanchet and Lehning 2010, Blanchet and Davison 2011, Hong and Ye 2014, Fan and Hong 2015). Probability distributions that have been commonly used in the literature for this purpose include the Gumbel, Lognormal, and Generalized Extreme Value (GEV) distributions. At locations where data are not available, the snow hazard is predicted from nearby sites.

This is the first of two companion papers that develop, describe, and illustrate a new approach for determining design ground snow loads based on uniform risk or reliability, rather than on uniform hazard. The procedures are consistent with the philosophy of risk-targeted seismic design maps newly adopted by *ASCE Standard 7-10* (ASCE 2010; Luco et al. 2007). These procedures recognize that designing for uniform hazard (*i.e*., a specific return period value or MRI) may not achieve consistent reliability of structures at different sites due to variability in the shape of the probability distribution of the hazard at different sites. This paper describes the methods for determining the probability distributions of annual maximum ground snow load for different sites, requiring the definition of both a probability model (*e.g.*, Lognormal, GEV, etc.) and the parameters associated with that model. This probability distribution can be used to determine a load associated with a predefined hazard level (*i.e.* a uniform hazard approach), but it is also appropriate for use in reliability assessment due to its emphasis on robust determination of rare loads. The application of these distributions for reliability assessment and mapping design ground snow loads is the focus of the companion paper (Liel et al. 2016).

There are a number of challenges that affect the statistical analysis of ground snow loads to develop probability distributions, especially (1) the short historical snow load record, particularly in the western U.S., and (2) the fact that much of the available historical data are recorded in terms of snow depth rather than snow weight. It is relatively straightforward to find a probability model that fits the observed data well. However, due to the short historical record, it is difficult to define the upper tail of the distribution, which governs the prediction of loads beyond the historical record, with confidence. This part of the distribution is the most important because it controls the 50-year values that have been conventionally used for design. This tail becomes even more important when considering reliability-based approaches because reliability calculations involve loads with return periods of 100 to 1000 years. For problems like this with limited data (*e.g.* 30 years of data and interest in a 50-year or longer MRI load), the probability distribution model best fitting the extremes of the data may not be the same as the distribution that provides the best overall fit to the entire data set. This study addresses these challenges through a new approach for determining probability models for ground snow loads that emphasizes distribution fitting in the upper tail of the distribution. To increase confidence in the shape of the distribution in the upper tail, a region of influence approach is developed to combine data from nearby sites with similar climatic conditions. These methods are illustrated through a case study of Colorado, but the procedures are also appropriate in other places where there is significant climatic variability from site to site and a limited historical record.

Ground Snow Load Data

Data Collection

Historical ground snow records for Colorado are obtained from four sources: Snow Course stations, SNOTEL (SNOwpack TELementary) stations, National Weather Service (NWS) cooperative observer (CO-OP) stations, and first-order NWS stations.

In Colorado, Snow Course stations are primarily located in mountainous areas above altitudes of 8,500 ft (2,591 m), and data from 177 of these stations are collected from NRCS (2015) for this study. Snow Course stations consist of a series of ten locations along a course, typically about a half-mile (805 m) long, at which a trained observer measures snow depth (inches) and weight in inches of water (*i.e.* snow water equivalent, “SWE”) on a monthly basis (NRCS 2015). The measurements are obtained by plunging an aluminum pipe through the snow pack and measuring the depth of snow around the outside of the pipe and weight of the snow that is trapped inside. The average depth and weight for the ten locations are reported.

Automated SNOTEL stations measure the weight of snow (SWE) via a fluid-filled pillow that senses pressure from the snow on top of it (NRCS 2015). SNOTEL also measures snow depth using a sonic sensor. The data collected for both SWE and depth include daily maxima, minima and averages. SNOTEL stations are often located in remote mountainous areas where accessibility is limited; they have sometimes replaced existing Snow Course stations. Data from 117 SNOTEL stations are collected from NRCS (2015) for this study.

First-order NWS stations report depth of snow on a daily basis, and have historically reported daily snow weight (SWE) in many cases. There are six First-order stations in the state of Colorado (Alamosa San Luis Valley Regional Airport, Colorado Springs Municipal Airport, Denver-Stapleton, Fort Collins, Grand Junction Walker, and Pueblo Memorial Airport). NWS CO-OP stations record snow depth on a daily basis, and in isolated cases snow weight (NWS 2015). Since snow weight is not always recorded at NWS stations, weights must be obtained from snow depth if not available, as described later. Data from 303 NWS stations are collected from the Western Regional Climate Center (WRCC 2015).

Data Processing

Data collected from the 603 stations are consolidated to create 463 historical records. The consolidated records are referred to as "snow sites." This consolidation combines data from stations that have changed names (*i.e.* a new station replaced an older station at the same location) and from stations that are close enough that no substantial difference in snow accumulation patterns is expected.

In this study, the following criteria are used to identify snow stations that are in close enough proximity for their records to be combined: (1) stations on the plains (east of the Rockies and below altitudes of 5,000 ft. (1524 m)) within approximately twelve miles (19 km) and 500 ft. altitude (152 m) of each other; (2) mountain stations (above altitudes of 6,000 ft (1829 m).) within approximately two miles (3.2 km) and 300 ft. altitude (91 m) of each other; and (3) any other stations within approximately five miles (8.0 km) and 300 ft. altitude (91 m) of each other. These criteria are based on observations of concurrent snow recordings at adjacent stations, which confirmed that there were not significant differences in the magnitude of snow load at sites combined using this criteria. In combining the stations, when a snow site consists of multiple stations that have data for the same year, direct weight measurements are given priority over snow weights that are converted from depth measurements. If there are multiple stations with the same data type in a given year (*i.e.,* all having direct weight measurements or all having depth measurements that are converted to weights), then the maximum annual snow weight for that year is taken as the maximum from all of the contributing stations. It should be noted that similar combinations of stations are already present in the NWS data that is collected for this study, reflecting sites with the same name that have moved.

Snow sites with less than 30 years of historical data are eliminated from the basic analysis, leaving 327 snow sites with 30 or more years of data, with an average historical record of 59 years. However, 60 sites with 18 to 29 years of data are used in some areas to inform the mapping process in the companion paper (Liel et al. 2016).

Example Sites

Figure 1 provides a map showing the locations of all of the snow sites. For illustration in this paper, we present summary details at six sites, listed in Table 1, which are located in various climatic regions across the state and represent the variety of cases considered. For complete information about all snow sites in Colorado, refer to SEAC (2016).

Conversions between Snow Depth and Snow Weight

Overview and Site Classification

Direct measurements of snow weight or load, reported in the form of snow water equivalent (SWE), are desirable for determining the annual maximum snow weights to define the ground snow load probability distributions. However, many weather stations do not measure snow weights, so weight must be estimated from depth. This process is complicated by the fact that the density of snow is highly variable by season, climate, topography and location (Sturm *et al.* 2010). A number of models for relating snow density to snow depth have been developed (*e.g.,* Tobiasson and Greatorex 1997, Jonas *et al.* 2009, Sturm *et al.* 2010). Here, we produce a Colorado-specific density model so that the depth-to-weight conversions used will be consistent with observed snow densities in the region.

We begin by classifying sites in Colorado in three categories based on their season-long snow coverage. Snow sites are defined as “compacted”, *i.e.* continuous snow coverage over winter season, “settled”, *i.e.* sporadic snow coverage over winter season, or “intermediate”. Separate depth-to-weight relationships are employed for each category because of different patterns in snow density. Compacted snow generally occurs at mountain sites as the result of multi-storm accumulation (*i.e.* the snow does not melt off completely between snow storms) and tends to be denser. Sites above 8,500 ft (2491 m) altitude are assumed to have compacted snow, which is consistent with the observation that most Snow Course stations (*i.e.* stations intended to gauge winter snow pack for estimating spring run-off, and thus located where season-long accumulation is expected to occur) are located at altitudes of 8,500 (2491 m) and higher. Settled snow sites are at lower elevations and are characterized by ground snow that tends to melt off between storms, resulting in lower density snow due to less time for consolidation and settlement. Settled snow sites are assumed to be those located east of the Rocky Mountains with altitudes less than 6,500 ft (1981 m) and those located in or west of the Rocky Mountains with altitudes less than 5,500 ft (1676 m). The upper elevation limit for settled snow sites is higher on the eastern slope because the terrain is less mountainous and tends to have greater sun exposure to melt off the snow between storms. There is also a large portion of the state of Colorado that does not meet the criteria for compacted or settled snow; snow sites in these areas are classified as “intermediate” sites.

Compacted Snow Sites

To examine the relationship between snow depth and weight at compacted snow sites, annual maximum snow depth and annual maximum snow weight data from all Snow Course sites are plotted on Figure 2, including data only for years in which depth and weight data are both available (6,524 data points in total). Two possible depth-to-weight relationships are shown on Figure 2a: (1) a power curve developed by Tobiasson and Greatorex (1997), Equation 1, which was developed from U.S. NWS data and is commonly used in the U.S.,

$$w = 0.279 * d^{1.36} \qquad 1$$

and (2) a power curve fitted to the Colorado specific data, developed by the authors and provided in Equation 2,

$$w = 0.584 * d^{1.25} \qquad 2$$

In both equations, w is the annual maximum snow weight or load in psf and d is the annual maximum snow depth in inches. The power curve that is fitted specifically to the Colorado compacted snow data predicts snow weight better than the Tobiasson and Greatorex (1997) relationship, which underestimates density at compacted snow sites by approximately 20-30 percent, since it was not developed for high altitude sites where compacted snow conditions are expected. In this study, annual maximum depth data are converted to annual maximum weights by Equation 2 at compacted snow sites where only snow depth data are available.

Settled Snow Sites

In Colorado, few stations at settled snow sites report both depth and weight data, so the data set for fitting a Colorado-specific depth-to-weight relationship is sparse (only 125 data points). The available data are shown in Figure 2b, which reveals that a power curve relationship fitted to Colorado data is very similar to the Tobiasson and Greatorex (1997) relationship up to a snow depth of about 12 inches (305 mm). Beyond this depth, the Colorado dataset is limited. However, Tobiasson and Greatorex's equation was fitted to data from 204 first-order NWS stations across the U.S., many of which could be classified as settled snow sites. Therefore, the Tobiasson and Greatorex (1997) equation (Equation 1) is selected for use at settled snow sites.

Intermediate Snow Sites

Intermediate snow sites are those that are classified as neither compacted nor settled. These locations may be mountain valleys, for example, and may have some years with compacted snow from season-long accumulation and some years with settled snow characteristics. There are insufficient data from stations at altitudes between 5,500 ft. ft and 8,500 ft (1676 – 2591 m) in Colorado that report both depth and weight, which prohibits fitting a model to intermediate site data directly. Instead, the depth-to-weight conversion at intermediate snow sites is computed by linearly interpolating between the compacted and settled depth-to-weight conversions, based on site altitude.

Validation of Density Relationships

To assess the depth-to-weight conversions described in the previous sections, we consider the 133 sites for which both depth and weights are available, so it is possible to compare the measured weight with the weights calculated from depth data according to the selected predictive equations. For each site, we computed the ratio of statistics on the converted weight data, *e.g.* mean, coefficient of variation (COV), and 50-year load, to the same statistics from the measured weight data. As Figure 3 shows, the ratios of the mean, COV and 50-year load between the converted and measured statistics all have median values of about 1 (0.99-1.02), indicating that there is no significant bias added by using converted rather than measured weights. Note that the converted weights have additional epistemic uncertainty stemming from the uncertainty in the depth to weight conversions (as indicated by the variability in the COV). However, based on the observation from Figure 3 that unbiased ground snow load statistics result from snow weights that are converted from depth, this source of uncertainty is not considered in the development of probability models for ground snow load.

Ground Snow Load Probability Models and Parameter Fitting

A two-stage approach is proposed for fitting probability distributions to ground snow load data. In the first stage, we develop site-specific probability models based on the data gathered for each snow site. In the second stage, the tail shapes of the site-specific probability models are modified based on data from other snow sites where knowledge of climate and snow conditions indicates that their extreme upper tails should be similarly shaped. This stage essentially uses a region of influence approach to combine data from multiple sites to improve prediction of the upper tail of the probability distribution, while maintaining site-to-site variability in snow loads.

<u>Site-specific Distribution-Fitting</u>

In this study, site-specific probability distributions of annual maximum ground snow loads are obtained by fitting a lognormal distribution to the upper tail of the data for each site. This approach, hereafter referred to as "tail fitting," ensures that the site-specific distribution is fitted to the portion of the data set that is important for determining rare loads, *i.e.* the rarer snow load values observed in each data set. For the same reason, Ellingwood (1981) used a tail-fitting approach to develop probability models for ground snow loads in the northeastern quadrant of the U.S. Similar approaches were described for biological applications by Harding (1949) and have been used for snow loads by Lee and Rosowsky (2005), among others.

The tail fitting is performed at each site by first rank ordering the annual maximum ground snow load data points from that site. This rank ordering is conducted without removal of outliers, unless the outliers are found to indicate an error in recording or transcription. The rank-ordered data are then plotted on a probability plot, and a least-square linear regression line is fit to the top 33% of the data. A probability plot has nonlinear axes, such that the quantiles of a distribution plotted against their corresponding theoretical values will form a straight line (Montgomery and Runger 2007). For the normal and lognormal distributions, the y-axis on probability paper (*i.e.* “standardized plotting position”) is defined as theoretical number of standard deviations from the mean associated with each quantile. The top 33 percent of the data at each site are selected for tail fitting, reflecting a tradeoff between the desire to emphasize the upper tail (tending to inspire a choice of a few, most extreme data points) versus the need to ensure confidence in the regression (where more data points is advantageous). Since the minimum years of record considered is 30, we chose the top 33% to ensure that at least 10 points were used to make the fit. The regression parameters for the fitted line are used to determine the parameters of the probability model for that site. The estimated standard deviation is the reciprocal of the slope and the expected value (mean) is the location of the x-intercept (*i.e.* 50^{th} percentile location of the fitted line).

An example of the tail-fitting technique applied to the 121 years of data from the Denver-Stapleton snow site is shown in Figure 4. Values of snow loads greater than 6 psf (287 Pa) are used in the tail-fitting in this example. The tail-fit distribution captures the tail behavior, *i.e.* the largest observed snow recordings at Denver-Stapleton, which range from 22 to 32 psf (1.05 kPa to 1.53 kPa), well. If the lognormal distribution is instead fitted to the entire data set, the fit diverges from the largest recordings. In particular, the model fitted to the entire data set underpredicts the likelihood of occurrence of large loads for the Denver-Stapleton site (as indicated by the thin dashed line that falls to the left of the uppermost data points). For example, the tail-fit (solid) line in Figure 4 suggests that 33 psf (1.58 kPa) ground snow load is about 2.5 standard deviations from the median, while the dashed non-tail-fit line indicates that 33 psf (1.58 kPa) is about 3 standard deviations from the median. The tail-fit distribution in Figure 4 does not match the observed data for low snow loads particularly well, but these snow loads do not influence the reliability assessment for determination of design values.

The tail-fitting approach is further explored with the Denver-Stapleton data in Figure 5 and Table 2. Figure 5 compares the Denver-Stapleton recorded data to tail-fitted distributions using Normal, Lognormal, Gamma and Log-Gamma probability models. A Type II Extreme Value distribution is also shown in Figure 5. In contrast to the other four distributions, the Type II distribution has been fit to the entire data set, but incorporates a third parameter that allows it to adapt to the upper tail of the data. The Type II is the best-fit distribution of the *entire* data set based on the Anderson-Darling, Kolmogorov-Smirnov, and Cramer-von Mises goodness-of-fit tests. Table 2 shows the ground snow loads at Denver-Stapleton predicted by these five probability models for mean recurrence intervals ranging from 50 to 1000 years. The results show only a six percent range in predictions of the 50-year loads obtained from the five models (values range from 20.4 psf to 21.6 psf (977 to 1034 Pa)). Even the Normal distribution, which is clearly an inappropriate choice for the Denver-Stapleton data, provides a similar fit to the top 33% of the data to the other distributions. As Table 2 shows, emphasizing the upper tail allows for estimates of extreme loads that are not strongly dependent on the distribution type that is selected for predicting loads, provided the recurrence interval of interest is near or within the length of the historical record.

In this study, all site-specific distributions are tail-fit and use the Lognormal probability model. In this model, the tail-thickness and variability is quantified by a single parameter, the logarithmic standard deviation. The use of the Lognormal distribution is consistent with previous analyses of snow loads in the U.S., *e.g.* Ellingwood and Redfield (1983). In addition, the parameters of the Lognormal distribution can be adjusted to ensure that the range of tail thicknesses observed in Colorado are adequately matched over the range of MRIs for which data are available, as shown in Table 2. The Type II extreme value distribution was not considered further because its tail is heavier than appropriate at some Colorado sites, and because of the simplicity of the two-parameter Lognormal tail-fitting approach.

Site-specific Lognormal distributions fit to the example sites in Table 1 are reported in Table 3. In the eastern plains region of the state, annual maximum ground snow loads come from single storm events, which are highly variable from year to year. Ground snow load distributions at sites in this region have high COVs (0.6 to 1.2) and hence relatively heavy upper tails. In the Rocky Mountains, maximum annual ground snow loads are due to snow accumulation over the winter season and have much smaller year-to-year variation (COVs of 0.2 to 0.5). The variability of annual maximum ground snow loads at plains and at mountain sites can be adequately captured by the Lognormal distribution tail-fit models.

Clustered-Based Distribution Fitting Approach

Limitations of the Site-Specific Approach

The Lognormal tail-fitting technique described above can be used for predicting snow loads with recurrence intervals up to about the length of the historical record. However, as shown in Table 2, there is significant uncertainty when extrapolating the distribution to predict rare loads of interest in structural design whose recurrence intervals are much greater than the length of the data time-series. Therefore, the Lognormal tail-fit distributions may not accurately predict extremely rare loads (*e.g.* 500 or 1000 year loads).

Generally, the true distribution cannot be identified from an existing data set. First, sampling error can result in the selection of a distribution that is not the true underlying (parent) distribution. Second, data records at individual sites are not long enough to verify whether a best-fit distribution for the 30 to 100 years of recorded data is still applicable for loads as rare as 500-1000 years. The implications of misfitting, for example, Lognormal data with a Type II model is relatively inconsequential for predicting loads with return periods of 50 – 100 years, but can produce large disparities at larger return period loads, which are highly dependent on the distribution choice, as demonstrated in Table 2. Ellingwood and Redfield (1983) also identified the pitfalls of using a data set to select a probability model for use at large loads without consideration of common best-fit distributions for similar data sets (*i.e.* other snow sites).

Proposed Approach for Clustered Site Distribution Fitting

This paper proposes a method to increase confidence in the upper tail of the distribution by combining data from similar snow sites. This procedure determines the magnitude of the snow loads in the distribution tail with a site-specific analysis, but determines the shape of the tail by combining data from a number of stations whose distribution tails are expected to have similar shapes. The idea of using measurements from multiple stations to improve predictions of extreme values has a significant history in the field of hydrology, where flood frequency analysis employs recordings at multiple stream gauges to improve the extreme value analysis at the station of interest (Burn 1990). Similar approaches have also been explored for extreme value statistics in wind loading (*e.g.* Hong and Ye 2014). In these approaches, measurement locations are grouped according to physical (*e.g,* stream drainage area) or statistical (*e.g.* stream mean flow) attributes, and measurements from all locations in the group are weighted and used to compute extreme value statistics. The groups may be mutually exclusive, (typically termed the "regional frequency" approach), or overlapping (typically termed the "region of influence" approach). The region of influence approach, a variation of which is adopted here, groups each site with its most similar sites for combined analysis. Region of influence methods have been shown to reduce the root mean squared error in the estimate of the prediction of a particular return period event, in comparison to a solely site-specific analysis or the regional frequency approach (Burn 1990, Mo et al. 2015).

Clustering of Snow Sites

To carry out the proposed distribution fitting method, it is necessary to determine groups, or "clusters", of snow sites that share physical and statistical attributes such that their upper tails are expected to have similar shapes. In this study, the classification of snow sites into clusters is performed by first dividing snow sites into coarse regions for which different snow distribution characteristics are expected based on climatic and macro-topographical features, and then combining each site with a minimum of 20 additional sites that are in the same climatic region and have similar altitude.

For the Colorado study, snow stations are initially divided into six regions, as shown in Figure 1. The six regions were selected through a combination of k-means cluster analysis (MacQueen 1967) and judgment of the authors. In addition, DePaolo (2013) previously conducted k-means clustering analysis of Colorado, showing that clusters of Colorado sites selected by similarities in snow density patterns aligned with these climatic delineations.

To examine the annual maximum snow load distribution properties on a regional basis, the COV is taken as a proxy for the distribution shape representing both the variability and the skew of the data. Distributions with heavy upper tails tend to have larger COVs, as compared to distributions with light upper tails. Analysis of the data for annual maximum snow loads in the six climatic regions identified in Figure 1 shows that the COV is related to altitude and climatic region. For example, Figure 6 shows the trends in COV with altitude for sites within the western slope and Plains/Front Range climatic regions. In both cases, the COV decreases with increases in altitude due to season-long snow accumulation at higher altitudes, but the shape of the trend differs between the regions. There is little distinction in the trends of COV with altitude among the east central Rocky Mountains, southwest Rocky Mountains, and San Luis Valley sites, so these are combined into one climatic region for analysis. The remaining three climatic regions are considered separately.

Clusters of snow sites are formed by selecting all snow sites that are: (a) within the same climatic region as the site of interest, and (b) within 1,000 ft. of altitude (305 m) of the site of interest. If a cluster contains fewer than 20 snow sites, the altitude criterion is relaxed until a minimum of 20 snow sites are included. This approach implies that each site is grouped with the 20 (or more) most similar sites on the basis of region and altitude; note that the groups determined according to this method are not mutually exclusive, and a snow site could potentially be included in multiple groups. The climatic region and altitude are the selected physical attributes for the grouping because of their critical relationship with the COV and the shape of the distribution. Mo *et al.* (2015) use COV directly to spatially interpolate snow loads in a province in China, but the region of interest in that study has less variability in altitude than considered here.

The heterogeneity (H) of a group of sites in terms of COV can be quantified by the metric (Burn 1990):

$$H_{COV} = \frac{Max(COV) - Min(COV)}{Median\ (COV)} \tag{3}$$

Considering all sites in the region, the heterogeneity is 2.8. Considering a typical cluster, the heterogeneity is 1.5, quantifying the increase of within-cluster similarity accomplished by the clustering.

Quantification of Tail Shape for Clustered Sites

Once clusters of sites are established, a probability model is fit to the aggregated data for the cluster, considering data from all the constituent sites. This model is used to inform the shape of the site-specific probability models.

To illustrate, consider a simple case: a hypothetical cluster of snow sites that have the same underlying distribution, in terms of the mean, as well as the shape of the upper tail of the distribution. This assumption may be valid for the eastern plains of Colorado, for example, where the variability in climate from site to site is not significant. For this case, the procedure is as simple as combining the data from multiple stations to create one large data set. A collection of 20 snow sites having 50 years of annual maximum snow data per site, for example, is combined to make a single station with 1000 data points. A distribution can then be fit to the 1,000 year "record" of the clustered site, which permits prediction of rare loads up to a 1,000-year mean recurrence interval without extrapolation. If all of the contributing snow sites have the same distribution shape and magnitude, then the enlarged data set provides a better estimation of the ground snow load probability distribution, due to the lengthened record of data.

In most cases, however, the annual maximum snow weight distributions (*i.e.* the means) vary somewhat even between similar snow sites. The process of combining the data in this case is illustrated in Figure 7. After the sites belonging in a particular cluster are identified, the site-specific lognormal distribution models are used to determine the 20-year ground snow load at each site (Figure 7a). Then, the record from each site is scaled so that each site's altered record has a 20-year ground snow load equal to the average 20-year load for all of the sites in the cluster (i.e. the annual maximum snow loads at each site are multiplied by the ratio of cluster 20 year load to site 20 year load, as shown in Figure 7b). This scaling has the effect of removing site-to-site differences in the amplitude of the annual maximum snow loads so that the data can be analyzed as a group. The 20-year load is selected as the scaling point for three reasons: (1) the 20-year load is rare (large) enough to quantify the magnitude of the upper distribution tail; (2) site-specific tail-fit distributions are appropriate for predicting the 20-year load at a site, because all sites have more than 30 years of history; and (3) the quantile of the 20-year load (0.95) is located in the middle of the quantile range of the data that will be used for fitting a distribution to the clustered data (0.9-1), as described below.

Once this scaling has been completed, the combined data set is fitted with a lognormal distribution using the tail-fit approaches described previously (Figure 7c). In this case, the lognormal distribution is tail-fitted using the highest 10% of the data points in the enlarged (cluster) data set. A smaller (10%) fraction of the tail, as compared to the 33% used in the site-specific approach, is permissible because there are far more data points in the enlarged data set. The fitted cluster distribution is then rescaled so that it matches the 20-year load for the site of interest (Figure 7d). The distributions resulting from this procedure match the 20-year load at each site, but have upper tails whose shapes are informed by a number of sites whose upper tail behaviors are expected to be similar. Therefore, they maintain a site-specific scale, or magnitude, but exhibit a tail shape that is informed by a cluster of data.

The combination of data from different sites to create an aggregated historical record implicitly assumes that the snow data from the different sites are statistically independent. This independence is satisfied if sites are not affected by the same weather systems. However, clustered sites in a common region may be influenced by the same storms, and, as a result, the annual maxima at different sites may be correlated. Some of this correlation is eliminated by combining nearby snow sites into single stations before fitting the distributions (*i.e.* in the data processing stage). The remaining correlation in annual maxima between pairs of snow sites in a given cluster typically ranges from 0 to 0.6, but many of the correlations are small. Thus, the clustering could produce an artificial increase in correlation between snow sites because the same storm may be counted twice in the artificially lengthened cluster record that is used to determine tail shapes at multiple sites. This correlation has a statistically similar impact (with regard to fitting the distribution) to a decrease in the sample size. A sensitivity study conducted by the authors confirmed that these correlations decrease the confidence of the fitted distribution parameters, but do not introduce significant bias.

Bias Produced by Clustering Process

In the process of scaling the data from a group of snow sites to create a clustered data set, a systematic bias is introduced. This bias is a direct result of the scaling that was used here to combine data from similar sites and does not occur if this scaling is not needed as in other region of influence studies (e.g. Burn et al. 1980). The bias can be quantified through Monte Carlo simulation. To illustrate the source of this bias, consider two hypothetical sites. Each has the same parent distribution, *i.e.* both follow a lognormal distribution with the same median and logarithmic standard deviation (for illustration, 4.5 psf (215 Pa) and 0.6), but with unique sets of random data representing their historical records generated from that distribution. In its 50 years of record, the largest load experienced at Site 1 is 13 psf (622 Pa, a load with approximately a 25-year MRI), whereas Site 2 has experienced a 21 psf (1005 Pa, approximately 200-year MRI load), as illustrated in Figure 8. When distributions are fit to each site's historical record, 20-year loads of 11 psf (527 Pa)and 14 psf (670 Pa) are calculated, compared to the theoretical 20-year load of 12 psf (575 Pa) based on the parent distribution. When the data from both sites are clustered together, the data from each site are scaled so that its 20-year load is equal to the average of the two (12.7 psf), requiring scaling the data from Site 1 by 1.12 and Site 2 by 0.9.

We can then compare distributions fit to 1) the combined scaled data from Sites 1 and 2, and 2) the unscaled combined data from Sites 1 and 2 (recall that the two sites' data come from the same parent distribution). The unscaled clustered data set has a logarithmic standard deviation of 0.60, which is equal to that of the parent distribution used to generate the historical data at both sites. However, the scaled clustered data set has a logarithmic standard deviation of 0.52, which underestimates the true logarithmic standard deviation. This bias is a result of data sets with large rare loads, which tend to be scaled down in the clustering process because their calculated 20-year loads tend to be higher than those of sites that have not experienced large rare loads; this leads to a systematic reduction of large rare loads in the clustered data sets. For groups of sites in which the shape and scale of the distribution are expected to be the same, this problem could be circumvented by combining their data without performing any scaling, but this solution is impractical due to the variety of snow conditions in the state and in other regions where this approach may be used.

Since the bias produced is systematic, a correction factor is developed to ensure that the clustered data set has the correct standard deviation. To develop the correction factors, we consider first sites that have a Lognormal parent distribution by considering artificial data sets for which we know the correct standard deviation (like the example above). For a given Lognormal parent distribution, 1000 data records, representing 50 years each (*i.e.* 50 data points per site, times 1000 sites, for a total of 50,000 points) are randomly generated. Then, the cluster fitting approach is applied to the simulated data to compute a distribution shape. The shape of the fitted distribution is compared to the shape of the true underlying (parent) distribution through comparison of their logarithmic standard deviations. This process is repeated for Lognormal distributions with a range of logarithmic standard deviations from 0.2 to 1.05, which corresponds to COV ranging from 0.2 to 1.4, representative of the majority of the Colorado snow sites. Based on these studies, the logarithmic standard deviations, σ, that are obtained by the clustered station distribution fitting approach can be corrected by Equation 4:

$$\sigma_{Corrected} = 1.16 * \sigma_{Clustered\ Data} \qquad 4$$

where $\sigma_{Clustered\ Data}$ is the computed logarithmic standard deviation of the scaled site data that have been combined into an augmented cluster record, and $\sigma_{Corrected}$ is the standard deviation that would be computed if scaling had not been required in that process. The effect of the correction factor is illustrated in Figure 9. Equation 4 works equally well for all values of logarithmic standard deviations that were tested, as long as the parent distribution is Lognormal.

Although we use the Lognormal distribution to develop the site-specific distribution at all sites, not all sites have data that follow a Lognormal distribution over the entire range of values. For example, in the mountains of Colorado, the best-fit distribution (when the models are fitted to the entire data set) is typically the Normal or Gamma distribution. In the Plains/Front Range, where the tail is heavier due to infrequent but large storms that cause some year's maximum snow load to be much larger than typical, the best-fit distribution tends to be Lognormal, Log-Gamma or the Type II extreme value distribution.

Fortunately, the likely parent distribution for a given group of stations is strongly related to the shape of a Lognormal distribution that is tail-fitted to their combined (scaled) data. Based on comparisons of the best-fit distributions of historical snow records in Colorado to the

Lognormal tail-fit parameters of their corresponding cluster-fit distributions, sites belonging to clusters with uncorrected logarithmic standard deviations greater than 0.6 are primarily modeled appropriately with the Lognormal distribution, and Equation 4 applies. Cluster-fitted distributions with logarithmic standard deviations less than 0.3 are composed of historical records that are commonly best described by the Normal or Gamma distributions, so they are corrected with Equation 5,

$$\sigma_{Corrected} = 0.97 * \sigma_{Clustered\ Data} + 0.03 \qquad 5$$

Equation 5 was developed using a similar procedure to that by which Equation 4 was developed.

If the uncorrected logarithmic standard deviation is between 0.3 and 0.6, then the underlying parent distribution is thinner-tailed than Lognormal, but thicker-tailed than Gamma and Normal. In such cases, the correction is computed by a weighted average of Equations 4 and 5, where the weights are proportional to the uncorrected logarithmic standard deviation relative to the threshold values of 0.3 and 0.6.

Results of Clustering Process

Each site is grouped with the 20 or more sites that are most similar, according to the clustering criteria. Lognormal probability models are then tail-fit to the clusters of data, and the results are rescaled to the site of interest. The outcome of this process is illustrated for six sites in Table 3. The cluster-based distributions have the same 20-year ground snow loads as the site-specific distributions, but the shape of the upper tail is refined through the clustering of data from 20 or more similar sites, producing changes in the logarithmic standard deviation from the site-specific distribution. Note that the clustering may also produce moderate changes to the median of the site-specific distribution. However, the reliability assessment is relatively insensitive to the accuracy with which the median is modeled, because snow loads that cause failure tend to fall in the upper tail of the probability distribution, not its mid-range.

Results of the cluster-fitting also show that sites with heavy-tailed snow load distributions tend to require larger corrections to their fitted distribution shapes than sites with thin upper tails, as shown in Table 3. At sites with heavy tails (*i.e.* the plains), rare loads are on the order of five to ten times larger than typical loads so the shape of the site-specific distribution depends heavily on whether a rare load has occurred within that site's historical record or not. In contrast, sites with thin upper tails require less correction, because the presence (or absence) of one or two rare loads has less impact on the shape of the fitted distribution. The data for Colorado seems to suggest that sites with less than $\mathrm{COV} \approx 0.4$ are not significantly impacted by the region of influence approach. However, this observation depends on the length of the historical record in Colorado and a number of other factors. A more practical approach is to observe the degree to which distribution shapes vary from one site to the next where similar snow patterns are expected; if the distribution shapes vary significantly between "similar" sites, a region of influence approach will likely enhance the distribution fitting process.

Conclusions

This paper proposes an approach for fitting probability models for ground snow loads to historical weather data. The probability models are fit through a two-stage process in which a Lognormal probability model is fit to the tail (upper 33 percent) of the data for each site at which data are available. Then, sites are clustered together based on similarities in climate and altitude. The combined clustered site is used to improve the prediction of the shape of the upper tail of the distribution, which defines rare loads that are critical for definition of design loads and robust reliability assessment. In addition, the approach employs region-specific snow density relationships. The outcome of this process is probability models that improve the quantification of loads associated with longer return periods.

The proposed approach for fitting probability models is applied to snow sites in the western U.S. state of Colorado. The resulting probability distributions quantify the wide variability of snow conditions throughout the state. These differences significantly impact the reliability of buildings designed for the 50-year loads, as described in the companion paper (Liel et al. 2016). Although the analysis focuses on data from Colorado, the proposed procedure is adaptable to any place where there is need to predict snow loads with recurrence intervals that exceed a limited historical record. Analysts interested in applying similar methods to other regions may wish to consider a region-specific depth-density relation if dealing with depth-only snow records, adaptation of clustering criteria appropriate to the topography and climatic patterns in the region of interest, and modification of the number of data points used in fitting the distributions based on the lengths of the historical record.

Acknowledgments

This paper summarizes work done by the Structural Engineers Association of Colorado Snow Loads Committee. Other members of the committee have contributed extensively to the study, including Richard Cunningham and Robert Pattillo. The contributions of Michael DePaolo are also gratefully acknowledged. Support for this work also was provided by J. R. Harris & Co.; this support is gratefully acknowledged.

References

ASCE (American Society of Civil Engineers). (2010). *Minimum design loads for buildings and other structures* (ASCE/SEI 7–10). American Society of Civil Engineering, Reston.

Blanchet J, Lehning M (2010). Mapping snow depth return levels: smooth spatial modeling versus station interpolation. *Hydrology and Earth Systems Sciences* 14(12), 2527–2544.

Blanchet J, Davison AC (2011). Spatial modeling of extreme snow depth. *Annals of Applied Statistics*, 5(3), 1699–2264.

BSI (British Standards Institution BSI). (2003). *Eurocode 1: Actions on structures. I–3: General actions—Snow loads.* EN 1991-1-3, London.

Burn, DH. (1990). An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrological Sciences Journal*, *35*(2), 149-165.

DePaolo, M. (2013). *A Proposal for a Unified Process to Improve Probabilistic Ground Snow Loads in the United States using SNODAS Modeled Weather Station Data*. M.S. Thesis, University of Colorado Boulder.

Durmaz, M., & Daloğlu, A. T. (2006). Frequency analysis of ground snow data and production of the snow load map using geographic information system for the Eastern Black Sea region of Turkey. *Journal of Structural Engineering*, *132*(7), 1166-1177.

Ellingwood, B. (1981). Wind and snow load statistics for probabilistic design. *Journal of the Structural Division*, 107(7), 1345-1350.

Ellingwood, B., & Redfield, R. (1983). Ground snow loads for structural design. *Journal of Structural Engineering*, 109(4), 950-964.

Ministry of Housing and Urban-Rural Development of the People’s Republic of China (2012.) *Load code for the design of building structures* (GB 50009-2012). China Architecture & Building Press, Beijing (in Chinese).

Hong HP, Ye W (2014). Analysis of extreme of ground snow loads for Canada using snow depth records. *Natural Hazards*, 73(2), 355–371.

Jonas T, Marty C, Magnusson, J. (2009). Estimating the Snow Water Equivalent from Snow Depth Measurements in the Swiss Alps. *Journal of Hydrology* 378(1-2): 161-167.

Liel, Abbie, D. Jared DeBock, James R. Harris, Bruce R. Ellingwood, Jeannette M. Torrents. "Reliability-Based Design Snow Loads: II. Reliability Assessment and Mapping Procedures", *Journal of Structural Engineering.* (companion paper; under review).

Luco N, Ellingwood BR, Hamburger RO, Hooper JD, Kimball JK, Kircher C. (2007). Risk-targeted versus current seismic design maps for the conterminous United States. *SEAOC 2007 Convention Proc*., Structural Engineers Association of California, Sacramento, CA.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

Montgomery, D. C., & Runger, G. C. (2007). *Applied statistics and probability for engineers, 4th Edition*. John Wiley & Sons Inc.

Mo, H. M., Fan, F., & Hong, H. P. (2015). Snow hazard estimation and mapping for a province in northeast China. *Natural Hazards*, *77*(2), 543-558.

Newark MJ, Welsh LE, Morris RJ, Dnes WV (1989). Revised ground snow loads for the 1990 National Building Code of Canada. *Canadian Journal of Civil Engineering,* 16(3), 267–278

NRCC (National Research Council of Canada). (2010) *National Building Code of Canada.* Institute for Research in Construction, Ottawa, Ontario.

NRCC (Natural Resources Conservation Center). (2015). National Water and Climate Center. http://www.wcc.nrcs.usda.gov/snow/. (Last accessed March 17, 2015).

NWS (National Weather Service). (2015). http://www.weather.gov. (Last accessed March, 2015)

SEAC (Structural Engineers Association of Colorado). (2007). *Colorado Ground Snow Loads.* Prepared by the SEAC Snow Load Committee.

SEAC (2016). Colorado Design Snow Loads. Prepared by the SEAC Snow Load Committee.

Sturm M, Taras B, Liston G, Derksen C, Lea J (2010). Estimating Snow Water Equivalent Using Snow Depth Data and Climate Classes. *Journal of Hydrometeorology*, 11,1380-394.

Tobiasson W, Greatorex A (1997). Database and methodology for conducting site specific snow load case studies for the United States. *3rd International Conference on Snow Engineering* (pp. 249-256).

WRCC (Western Regional Climate Center). (2015). http://www.wrcc.dri.edu/. (Last Accessed March, 2015)

Figure Captions and Tables

Figure 1 . Map of snow sites in Colorado, showing climatic region assignment and location of numbered example sites in Table 1. (1 ft = 0.30 m)

Figure 2. Annual maximum snow weight vs. annual maximum snow depth for (a) compacted and (b) settled snow sites in the state of Colorado. (1 in = 25.4 mm; 1 psf = 47.88 Pa)

Figure 3. Boxplots showing the ratio of statistics (mean, COV, 50-year load) computed from converted weights to statistics computed from measured weights, for 133 sites where both depth and weight are available.

Figure 4. Probability plot showing Denver-Stapleton annual maximum snow load data set, a tail-fit lognormal distribution fitted to the data, and a lognormal distribution fitted to the entire data set. The thickened part of the solid line represents the range over which the tail-fitted distribution was fitted (*i.e.* upper 33% of the data points).

Figure 5. Histogram of the Denver-Stapleton annual maximum snow loads with four different tail-fit distributions overlaid, in addition to a Type II distribution fitted to the entire data set. (1 psf = 47.88 Pa)

Figure 6. Trends of COV with altitude for annual maximum snow load data in two climatic regions: eastern slope (Plains/Front Range) and western slope. (1000 ft = 300 m)

Figure 7. Illustration of cluster fitting distribution procedure, showing (a) top 33% of data for four hypothetical sites and their associated tail-fit site specific lognormal distributions, and their 20-year loads, (b) the same data, but after the data from all sites have been scaled to have a common 20-year load, (c) the scaled data combined into a single data set, with a lognormal distribution fit to the top 10% of the enlarged data set, and (d) the rescaling of the combined data set to represent a site of interest (Site 4). These site data are hypothetical, each generated from lognormal distributions having $\sigma = 0.6$, but different median values. Here we show the approach for 4 sites, but the actual procedure utilizes a minimum of 20 sites. (1 psf = 47.88 Pa)

Figure 8. Illustration of bias that is introduced by the cluster fitting procedure, using two example sites.

Figure 9. Logarithmic standard deviation (σ) of the clustered data set before and after the proposed correction, based on Equation 4.

Table 1. Example snow sites in Colorado.

Site No. (On Figure 1)	Site Name	Altitude (ft [m])	Type of Station	Years of Record	Snow Classification*	Climatic Region
1	Denver-Stapleton	5286 [1611]	First Order	121	Settled	Eastern Slope
2	Boulder	5484 [1672]	NWS	63	Settled	Eastern Slope
3	Lamar	3627 [1106]	NWS	118	Settled	Eastern Slope
4	Yampa	7857 [2395]	NWS	64	Intermediate	Western Slope
5	Copper Mountain	10550 [3216]	SNOTEL	38	Compacted	North Rockies
6	Geneva Park	9600 [2926]	Snow Course	64	Compacted	East-Central Rockies

*compacted: sites with continuous snow coverage over winter season; settled, sites with sporadic snow coverage over winter season

Table 2. Ground snow loads predicted by four tail-fitted distributions and the Type II distribution for Denver-Stapleton.

Mean Recurrence Interval	Ground Snow Loads				
	Tail-fitted Distributions				Type II
	Normal	Gamma	Lognormal	Log-Gamma	
50 yrs	21.5 psf	21.6 psf	20.5 psf	20.8 psf	20.4 psf
100 yrs	24.6 psf	26.0 psf	25.4 psf	26.8 psf	26.5 psf
500 yrs	30.9 psf	36.3 psf	39.4 psf	46.9 psf	47.6 psf
1000 yrs	33.3 psf	40.8 psf	46.6 psf	59.0 psf	60.8 psf

Table 3. Site-specific and cluster-based distributions fitted for example sites.

Site	Site-Specific Distributions		Cluster-Based Distributions	
	Median (psf)	**Logarithmic Std. Deviation*, σ_{ln}**	**Median (psf)**	**Logarithmic Std. Deviation*, σ_{ln}**
Denver-Stapleton	4.0	0.79	4.3	0.75
Boulder	7.8	0.51	5.1	0.77
Lamar	2.4	1.0	3.3	0.84
Yampa	25.1	0.30	25.2	0.30
Copper Mountain	75.7	0.18	73.3	0.2
Geneva Park	21.2	0.33	21.8	0.31

*For small values, COV is approximately equal to σ_{ln}. The values diverge for large quantities, but both parameters quantify the spread of the distribution.