PROBABILISTIC MODELS FOR GROUND SNOW ACCUMULATION

Bruce Ellingwood

Center for Building Technology
National Bureau of Standards
Gaithersburg, MD 20899

ABSTRACT

Snow loads specified in modern structural design standards, e.g., American National Standard A58 and the National Building Code of Canada, are calculated as the product of a ground snow load and a ground-to-roof conversion factor. The design-basis ground snow loads are sensitive to the choice of probability distribution, since they must be obtained by extrapolating into the upper tail of the distribution beyond the range covered by the historical data. This paper considers the selection of probability distributions for modeling annual extreme ground loads, sampling errors caused by limitations in the data, and the sensitivity of nominal design-basis snow loads to these factors. Extensive water-equivalent data from weather stations in the northern United States and Canada are analyzed. The analysis strongly suggests that the lognormal probability distribution is preferable for describing the annual extreme ground snow loads at sites in the north-central United States, while the Type I distribution appears preferable in the northeast United States and Canada.

INTRODUCTION

Snow loads furnish the governing load requirements for the structural design of roofs in many areas of the northern United States, Canada, and Europe. Current practice is to calculate the roof snow load, P_r, as the product of a design-basis ground snow load, P_g, and a ground-to-roof conversion factor, C, (ANS A58, 1982; NBCC, 1977):

$$P_r = C\,P_g \tag{1}$$

The factor C depends on the exposure of the roof and its geometric and thermal characteristics (Taylor, 1980; O'Rourke, et al, 1982). The design-basis ground snow load describes the regional variation in the load and is determined from meteorological data.

The design-basis ground snow load is obtained from a statistical analysis of data describing the annual extreme snow accumulation. A cumulative probability distribution function (cdf) is assumed as a model for the annual extremes, the parameters of the distribution are determined from the data at a given site, and a value of the distribution with a small probability, p, of being exceeded in any year is selected for design. These annual probabilities typically range from 0.01 to 0.04 (ANS A58, 1982; NBCC, 1977; ISO 4355, 1981), which correspond to mean recurrence intervals N = 1/p ranging from 25 to 100 years. Snow load contours that are consistent with the calculated values are drawn on a map. The map shows the regional variation in the ground snow load for code purposes.

The probability distribution for the annual extreme ground snow obviously is an important element of the snow load provisions. There is some question as to whether a single distribution is sufficient to model the annual extreme at all sites and what the

Proceedings, Eastern Snow Conference, V. 29, 41st Annual Meeting, Washington, D.C., June 7-8, 1984

distribution ought to be. This question is significant because it is necessary to extrapolate into the upper tail of the distribution beyond the range of observed data to determine P_g. In this range, the values of P_g are sensitive to the probability distribution chosen.

PROBABILITY MODELS

It is instructive to review briefly the history behind the current ground snow provisions in the United States and Canada (ANS A58, 1972 and 1982; NBCC, 1977).

Studies in the early 1960's of snow depths at some 200 sites in Canada led to a recommendation that the Type I distribution of largest values should be selected to model the annual extremes (Boyd, 1961). It was reasoned that since each annual extreme is the maximum of all snow depths measured during a winter, a Type I distribution should model the distribution of annual extremes. The ground snow loads in the National Building Code of Canada (NBCC, 1977) are based on this reasoning. Although the argument underlying the selection of the Type I distribution is intuitively appealing, no attempt to test this hypothesis statistically has been reported. Snow load provisions in Europe (ISO 4355, 1981) have also utilized the Type I distribution, apparently relying on the same intuitive reasoning.

In the United States, an analysis of 10 years (1952-1962) of water-equivalents of snow at 140 stations led to the conclusion that the water-equivalents should be described by a lognormal distribution (Thom, 1966). Thom's study was the basis of the ground snow maps in the 1972 edition of the A58 Standard on structural loads (ANS A58, 1972). As part of the work underlying the 1982 revision of the A58 Standard, Redfield analyzed snow depths at selected stations using a Chi-square test of fit. This analysis confirmed the choice of the lognormal distribution as a model of ground snow loads, and the ground snow maps in ANS A58.1-1982 are based on this probability model. A study (Steyaert, et al, 1980) of water-equivalents at first-order weather stations in the northeast United States found the lognormal distribution to be preferable at a majority of stations using the Chi-square test; however, the same study found the Type I to be preferable at a majority of the same stations using the Kolmogorov-Smirnov test. Finally, in a recent analysis of water-equivalents in the northern United States for the years 1952-1980 (Ellingwood and Redfield, 1983), the lognormal distribution was found to be preferable to the Type I distribution by a margin of over two to one. This analysis will be described subsequently.

The lognormal distribution for a variable X is given by,

$$F_{LN}(x) = \Phi\left(\frac{\ell n\, x - \lambda}{\zeta}\right); 0 < x < \infty \tag{2}$$

in which λ and ζ^2 are the mean and variance of ℓn X, and $\Phi(\ )$ is the standard normal probability integral (Johnson and Kotz, 1970). Estimates $\hat{\lambda}$ and $\hat{\zeta}$ of these parameters depend on the site. The N-year mean recurrence interval value of X, X_N, is estimated by,

$$\hat{X}_N = \exp\,[\hat{\lambda} + \hat{\zeta}\ \Phi^{-1}\ (1-1/N)] \tag{3}$$

in which $\Phi^{-1}(\ )$ is the inverse of $\Phi(\ )$.

The Type I distribution for X is given by,

$$F_I(x) = \exp\left(-\exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right]\right); -\infty < x < \infty \tag{4}$$

in which μ and σ are distribution parameters. Estimates of the N-year mean recurrence interval values for design purposes can be obtained by,

$$\hat{X}_N = \hat{\mu} - \hat{\sigma}\ \ell n\ [-\ell n\ (1-1/N)] \tag{5}$$

in which $\hat{\mu}$ and $\hat{\sigma}$ are estimates of μ and σ, obtained from the data sample at a particular site.

The lognormal distribution is a limiting form of the log-Pearson Type III distribution, which has three parameters. Similarly, the Type I distribution is a limiting form of a general three-parameter Type II distribution of largest values. These more complex probabilistic models of snow data were considered in a recent study (Ellingwood and Redfield, 1984). It was found, in most cases, that the third parameter, while affording an additional degree of freedom and flexibility, did not yield a significant improvement in the fit over the two-parameter lognormal and Type I distributions.

STATISTICAL ANALYSIS OF WATER-EQUIVALENTS

The investigations of probability models utilize sets of annual extreme water-equivalents at first-order stations in the northern United States. This is an area where snow loads tend to be important in structural design. There are many more stations that report snow depths. However, the relation between depth and density must be established to compute the load, and this relation is uncertain. Water-equivalents which can be converted to load using a multiplier of 5.2 psf/inch (1 g/cm) are more suitable for studying basic probabilistic models of ground snow load.

The water-equivalents are analyzed using the probability plot correlation coefficient (Filliben, 1975; Ellingwood and Redfield, 1983) as a measure of distribution fit. With this method, data being analyzed are rank-ordered and plotted (by computer) at their rank median plotting positions on probability paper for the distribution being tested. If the distribution being tested is a good model of the data, the plot of the data will be nearly linear and the probability plot correlation coefficient will be nearly unity. The distribution for which the probability plot correlation coefficient is closest to unity is selected for subsequent statistical modeling and testing. The procedure is computer-automated, enabling a wide variety of distributions to be tested.

Table 1 summarizes the results of the analysis of annual extreme water-equivalents at stations in the upper Midwest. These data are shown, for convenience, in psf units, obtained by multiplying the water-equivalent, in inches, by 5.2. Column 2 of Table 1 lists the number of years of record, column 3 the maximum value observed, and column 4 whether the lognormal (LN) or Type I (I) distribution provides the better fit to the data. Columns 5 and 6 show, respectively, estimates of the 50-year mean recurrence interval load and the standard deviation in the estimate of X_{50} that is obtained from the distribution of choice. The latter parameter, $SD(\hat{X}_{50})$, is a measure of the uncertainty in the estimate $\hat{X}_{50}$ based on a sample of the size listed in column 2. Under the assumption that the sampling error has a normal distribution, one could conclude that the true value of X_{50} is within $\pm SD(\hat{X}_{50})$ with a probability of 68 percent of being correct. The final three columns give the mapped (50-year mean recurrence interval) ground snow load in the 1972 and 1982 editions of the A58 Standard and in the map developed by Steyaert, et al (1980). Probability plots of the annual extremes at Milwaukee, WI and Duluth, MN, are shown in Figures 1 and 2.

Table 1 shows that the maximum observed value in a sample of 28 (or less) observations usually is less than the 50-year mean recurrence interval value, with a few exceptions (viz., Lansing, MI; International Falls, MN; Milwaukee, WI). The maximum in a sample of M is a reasonable estimate of the M-year mean recurrence interval value (Ellingwood and Redfield, 1983). Thus, the probability that the maximum in a sample of size M is exceeded in any given year is roughly equal to 1/M and is influenced strongly by the sample size, at least for values of M that are typical for climatological data samples. In contrast, the estimate of the M-year mean recurrence interval value is less sensitive to sample size than is the sample maximum and is more desirable as a basis for design. The sensitivity of the distribution to sample size can be seen in Figure 2, comparing the cdf fitting data from 1952-1962 to the cdf fitting data from 1952-1980 at Duluth, MN.

It should be noted that while there are some differences between the P_g recommended in A58.1-1982 and NUREG/CR-1389 (1980) (the data sets were slightly different), they both tend to be larger than the loads recommended in ANS A58.1-1972. The trend for the design

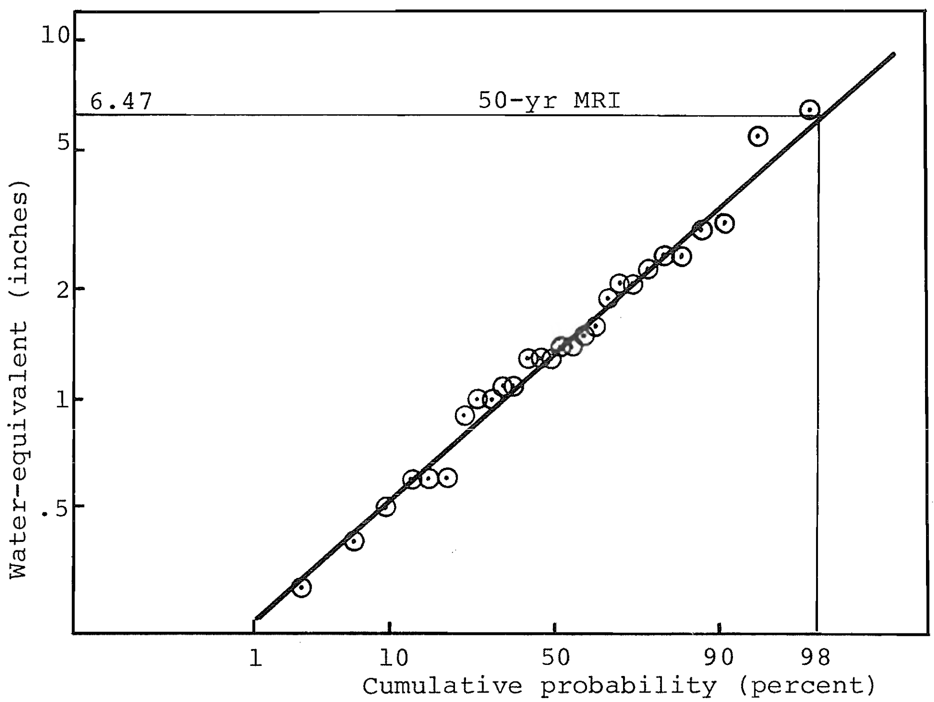

Figure 1. Lognormal distribution of annual extreme water-equivalents at Milwaukee, WI

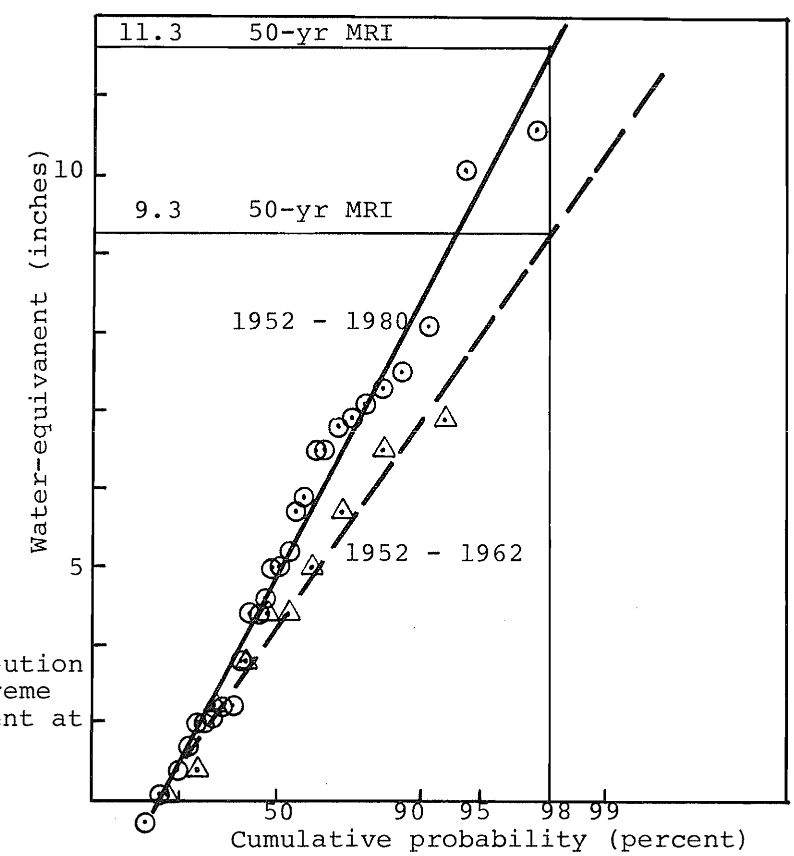

Figure 2. Type I distribution of annual extreme water-equivalent at Duluth, MN

Table 1. Ground Snow Loads from Annual Extreme Water-Equivalents at First-Order Weather Stations (psf)

Station	Sample statistics (psf)					Design snow load, P_g (psf)		
	Size, M	Maximum	cdf	$\hat{X}_{50}$	$SD(\hat{X}_{50})$	A58-72	A58-82	NUREG/CR-1389
Alpena, MI	19	34	I	44	8	42	50	47
Detroit Air, MI	22	14	I	15	3	18	20	20
Flint, MI	25	20	I	22	4	20	30	25
Grand Rapids, MI	28	32	LN	39	10	22	40	28
Houghton Lake, MI	16	33	I	42	7	30	50	50
Lansing, MI	23	34	I	34	6	20	35	25
Marquette, MI	16	44	LN	54	8	50	60	70
Muskegon, MI	28	40	LN	45	12	24	50	50
S. Ste. Marie, MI	28	68	LN	83	16	50	70	60
Duluth, MN	28	55	I	61	9	45	70	60
Intl. Falls, MN	28	43	I	41	5	35	50	50
Minneapolis, MN	28	34	LN	53	16	42	50	40
Rochester, MN	28	30	I	33	5	35	50	30
St. Cloud, MN	28	40	LN	56	18	40	60	40
Green Bay, WI	28	37	LN	39	11	28	40	35
LaCrosse, WI	16	23	LN	34	16	30	40	35
Madison, WI	28	32	LN	34	11	24	30	25
Milwaukee, WI	28	34	LN	34	9	24	35	30

* 1 psf = 47.88 Pa

snow loads to increase as the number of years of data is increased was also observed in the Canadian snow load update (Newark, 1984).

An analysis of water-equivalents at 76 first-order stations in the northern United States (Ellingwood and Redfield, 1983) revealed that the data are better fitted by a Type I distribution at 26 (or 34 percent) of the stations, while the data are better fitted by a lognormal distribution at 50 (or 66 percent) of the stations. If only those 38 stations with 28 years of data are considered, the Type I is the choice at 24 percent while the lognormal is the choice at 76 percent. There is a strong indication from these results that the lognormal distribution is a better probabilistic model of ground snow load than the Type I distribution at a majority of stations in the northern United States.

One possible explanation for the differences in distributions observed at different stations is sampling error due to the finite size of the data samples analyzed. As an illustration, a large number of data sets containing 28 values each was generated by Monte Carlo simulation from a parent lognormal population with parameters as defined for Rochester, NY. Each of these data sets was tested for distribution fit. The results of this experiment are summarized in Table 2. When sets of 28 are generated from a typical underlying lognormal population of annual extreme snow loads and are tested for distribution fit, approximately 25 percent actually will appear to be fitted better by a Type I distribution. These percentages are quite similar to the percentages observed. Conversely, when the process is repeated assuming that the parent distribution is Type I, about 29 percent will be fit better by a lognormal distribution.

Table 2. Percentage of Data Sets Fitted Better By Lognormal and Type I Distributions

Sets simulated from	Distribution of Choice LN	I
LN	75	25
I	29	71

The above explanation for the differences in probabilistic models would be plausible were it not for the fact that the stations characterized by the Type I distribution tend to occur in clusters rather than at random over the northeast United States. There are two distinct clusters of such stations: one in New England and a second in northeast Michigan. There appears to be a third grouping in northern Minnesota and the eastern Dakotas, but there are not enough stations in that area to infer any definite regional patterns. Many of the stations where the Type I distribution appears preferable experience long periods of continuous snow cover during the winter. The annual extremes at such stations may be the result of the progressive accumulation of several storms during the course of the winter rather than a single storm. In contrast, annual extremes at many stations in the Midwest and plains states, where the lognormal distribution is preferable, tend to be the result of a major winter storm. Moreover, these data sets are characterized by a few years of large values amidst a large number of years of moderate values.

Thus, there is a strong possibility that the differences in distributions may be the result of different climatological conditions. The dependence of probability models on local climatology has been observed in connection with a study of annual extreme wind speeds in hurricane-prone regions as opposed to extratropical regions (Simiu, 1980). In order to examine this possibility, water-equivalent data for 69 stations representing up to 25 years of data in northeast Canada were obtained*. These data were analyzed using, for consistency, the same statistical methods that were used to analyze the data in the northeast United States.

The analysis of ground snow loads in Canada has been presented recently (Newark, 1984), so only those results that pertain to the present discussion are presented herein. The characteristics of the Canadian data, taken as a whole, appear different from the characteristics of the data from stations in the United States. The Type I distribution is preferable at 43 (57 percent) of the Canadian stations, while it is the choice at only 26 (34 percent) of the U.S. stations. The sample coefficients of variation of the Canadian data sets typically are 0.25-0.45, while in the United States the sample coefficients of variation typically are 0.60-0.85. The sample means tend to be higher in Canada while the sample standard deviations are about the same (typically 1.5-4.0 in (37-102 mm)), leading to a smaller sample coefficient of variation.

The results of the tests of distribution fit for the northeast United States and Canada are summarized in Figure 3. A contour has been drawn (without a rigorous basis) around the region where the Type I distribution is predominant. The regions of Type I behavior noted previously in New England and Michigan appear to continue up into Canada. There are a few lognormal stations in this area, but their presence could be attributable to the distribution sampling errors described previously. With the exception of the stations in New York State, where the lognormal distribution clearly is preferable to the Type I distribution, it appears that ground snow loads in much of the northeast United States and Canada should be described by a Type I distribution. Conversely, ground snow loads at stations in the plains states and upper Midwest are described better by a lognormal distribution.

The fact that more than one distribution may be necessary to estimate X_N for design purposes may have a significant impact on current practice. When lognormal and Type I probability distributions are fitted to a sample of data for which the sample coefficient of variation is greater than approximately 0.35, the lognormal distribution has a longer upper tail than the Type I distribution. For this case, estimates of X_N for N = 25-100 years, will tend to be more conservative assuming a lognormal distribution. Conversely, if the sample coefficient of variation is less than approximately 0.35, the Type I distribution has a longer upper tail. Recall that the lognormal distribution has been used to develop snow load criteria in the United States, where the sample coefficients of variation are almost always higher than 0.50. Similarly, the Type I distribution has been used in Canada, where the sample coefficients of variation tend to be 0.35 or less. It appears fortuitous that the distributions selected in both Canada and the United States tend to result in conservative estimates of X_N for design.

While this conservatism may be reassuring from a public safety point of view, the snow load criteria that result may be unnecessarily conservative from the uniform risk standpoint that has been the basis of the A58 Standard (and other standards) for over 10 years. Table 3a lists several stations for which the distribution of choice is Type I and the corresponding estimate, $\hat{X}_{50}$. The estimate of X_{50} for these stations, under the assumption that the distribution is lognormal, is also given. Columns 4 and 5 of Table 3a show the probabilities that $\hat{X}_{50,LN}$ are exceeded annually and the corresponding MRI. These probabilities are smaller, by an order of magnitude at some stations, than the probability of 0.02 corresponding to the 50-year MRI value. However, the assumption of a Type I distribution at all stations would be unconservative in much of the United States. Table 3b lists several stations for which the distribution of choice is lognormal. Estimates of $X_{50,I}$, under the assumption that the underlying distribution is Type I, and the probability of exceeding $\hat{X}_{50,I}$ are presented. The mean recurrence intervals associated with design loads determined on this basis may be less than 25 years. This would be unacceptable in the United States, where a 50-year mean recurrence interval load has been recommended for most permanent structures in the 1972 and 1982 editions of the A58 Standard (1972, 1982).

*These data were made available to the writer courtesy of M. Newark of the Atmospheric Environment Service of Canada (AESC). However, the views expressed herein are those of the writer, and do not necessarily reflect the opinions of AESC.

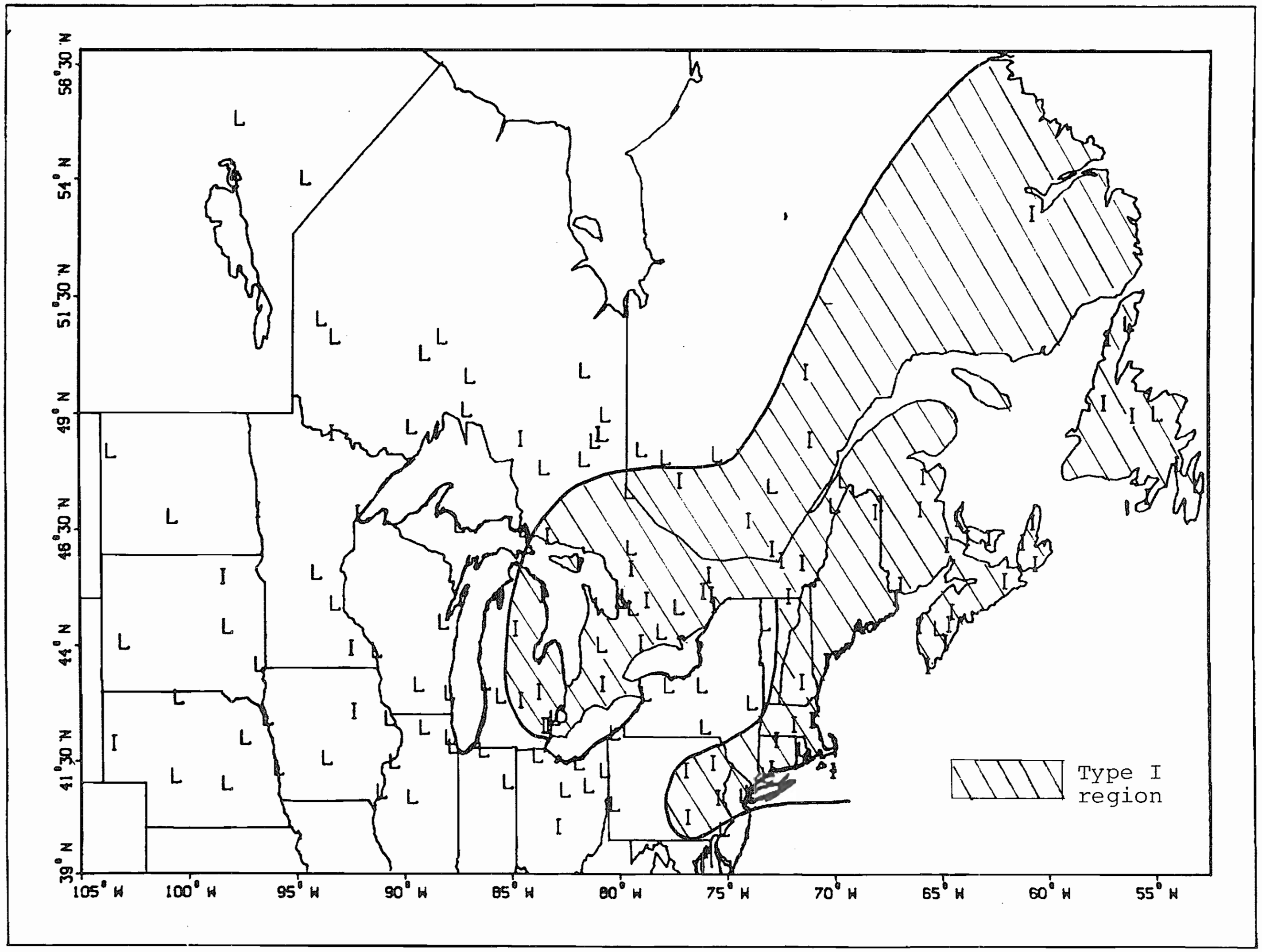

Figure 3. Regional variation in probability distribution of annual extreme water-equivalents

Table 3. Annual Probability of Exceeding X_{50} Determined from Lognormal Distribution

3a. Type I Stations

Station	$\hat{X}_{50}$ (psf)	$\hat{X}_{50,LN}$ (psf)	$P[X > \hat{X}_{50,LN}]$	N (yr)
Concord, NH	44	65	0.0014	719
Boston, MA	27	31	0.0080	124
Alpena, MI	44	54	0.0049	205
Lansing, MI	34	44	0.0046	218
Duluth, MN	61	67	0.0109	92

3b. Lognormal Stations

Station	$\hat{X}_{50}$ (psf)	$\hat{X}_{50,I}$ (psf)	$P[X > \hat{X}_{50,I}]$	N (yr)
Rochester, NY	40	36	0.0287	35
Grand Rapids, MI	39	31	0.0409	24
Minneapolis, MN	53	39	0.0446	22
Green Bay, WI	39	32	0.0344	29
Madison, WI	34	27	0.0359	28

1 psf = 47.88 Pa

SUMMARY AND CONCLUSIONS

Probability models for annual extreme snow loads in the northeast United States and Canada have been examined. The lognormal and Type I Extreme Value distributions were both found to be acceptable models at numerous sites. Stations typified by either distribution appear to be grouped in regions rather than scattered randomly, suggesting a dependence of distribution on regional climatic conditions. The Type I distribution is preferable at most sites in the northeast United States and Canada, while the lognormal distribution is preferable at sites in the upper Midwest and plains states. Standard committees should consider the advisability of abandoning the notion that one distribution is sufficient to describe environmental loads such as snow over large areas with widely varying climates.

ACKNOWLEDGMENT

The assistance of Mr. William Nolan, student in the Department of Civil Engineering, The Johns Hopkins University, and Mr. Erik Hendrickson of the Center for Building Technology during the preparation of this paper is gratefully acknowledged.

REFERENCES

"American National Standard A58, Minimum Design Loads for Buildings and Other Structures (ANSI A58.1-1982)," American National Standards Institute, New York, 1982, 100 pp.

"Bases for Design of Structures - Determination of Snow Loads on Roofs (ISO 4355 - 1981)," International Organization for Standardization, Switzerland, 1981, 18 pp.

Boyd, D., "Maximum Snow Depths and Snow Loads on Roofs in Canada," National Research Council of Canada, Division of Building Research Paper 142, December 1961, Ottawa, Canada.

Ellingwood, B. and Redfield, R., "Ground Snow Loads for Structural Design," Journal of Structural Engineering, ASCE, Vol. 109, No. 4, April 1983, pp. 950-964.

Ellingwood, B. and Redfield, R., "Probability Models for Annual Extreme Water-Equivalent Ground Snow," Monthly Weather Review, 1984 (in press).

Filliben, J.J., "The Probability Plot Correlation Coefficient Test for Normality," Technometrics, Vol. 17, No. 1, February 1975, pp. 111-117.

Johnson, N.L., and Kotz, S., Continuous Univariate Distributions, John Wiley and Sons, Inc., New York, NY, 1970.

"National Building Code of Canada (NBCC)," National Research Council of Canada, Ottawa, Canada, 1977.

Newark, M., "A New Look at Ground Snow Loads in Canada," Proceedings, 41st Eastern Snow Conference, New Carrolton, MD, June 1984.

O'Rourke, M., Redfield, R. and VonBradsky, P., "Uniform Snow Loads on Structures," Journal of the Structural Division, ASCE, Vol. 108, No. ST12, December 1982, pp. 2781-2798.

Simiu, E. and Filliben, J.J., "Probability Distributions of Extreme Wind Speeds," Journal of the Structural Division, ASCE, Vol. 102, No. ST9, September 1976, pp. 1861-1877.

Steyaert, L., et al., "Estimating Water Equivalent Snow Depth from Related Meteorological Variables," U.S. Nuclear Regulatory Commission Report NUREG/CR-1389, Washington, DC, May 1980.

Taylor, D., "Roof Snow Loads in Canada," Canadian Journal of Civil Engineering, Vol. 7, No. 1, March 1980, pp. 1-18.

Thom, H.C.S., "Distribution of Maximum Annual Water Equivalent of Snow on the Ground," Monthly Weather Review, Vol.94, No. 4, April 1966, pp. 265-271.