

A STATISTICAL VALIDATION OF THE SNOWPACK MODEL IN A MONTANA CLIMATE

Christopher C. Lundy*, Robert L. Brown and Edward E. Adams
Department of Civil Engineering, Montana State University-Bozeman

Karl W. Birkeland
USDA Forest Service National Avalanche Center

Michael Lehning
Swiss Federal Institute for Snow and Avalanche Research

ABSTRACT: Recently, a computer model has been developed by the Swiss Federal Institute for Snow and Avalanche Research that simulates the evolution of a natural snow cover. Using common meteorological parameters as input, SNOWPACK predicts characteristics such as snowpack temperature and density, in addition to snow microstructure and layering. An investigation was conducted to evaluate the effectiveness of SNOWPACK in a Montana climate. A weather station was constructed in the Bridger Mountains near Bozeman, Montana, to provide the meteorological parameters necessary to run SNOWPACK. Throughout the 1999-2000 winter, weekly snow profiles were performed in undisturbed snow to provide a benchmark for the model output. Density, grain size, and crystallography were recorded on 10 cm intervals over the full snow depth, and the temperature profile was monitored with a thermocouple array. Finally, the meteorological parameters were input into SNOWPACK, and a statistical comparison was performed comparing the predicted snowpack to the observational data. Snowpack temperatures are predicted reasonably accurately by SNOWPACK. The modeled and observed densities correlated well, but the model typically underestimates snowpack settlement. Comparison of grain size and shape was problematic due to different definitions utilized by the model and observer, but still demonstrated some agreement.

KEYWORDS: snowpack modeling, snow metamorphism, model validation, snowpack evolution

1. INTRODUCTION

Recent advances in snow research, as well as the increasing availability of powerful computer systems, have led to the development of computer models that are becoming better at predicting the evolution of a mountain snowpack. The more advanced models have already been used operationally and provide avalanche hazard forecasters and other mountain safety experts with yet another tool for evaluating the alpine snowpack.

The focus of this paper is SNOWPACK, a numerical snow cover model developed by the Swiss Federal Institute for Snow and Avalanche Research. SNOWPACK is a predictive finite element model that computes the heat and mass transfer, stresses, and strains within the snowpack

(Lehning *et al.*, 1998). SNOWPACK uses common meteorological parameters as inputs and provide as output predicted snowpack temperature, density, grain and bond size, and coordination number. The model also characterizes the expected crystal type based on the grain shape parameters sphericity and dendricity (Figure 1). Currently, wet snow metamorphism and surface hoar prediction are still under development, and the models for dry snow metamorphism are being improved (Lehning *et al.*, 1998).

One problem associated with efforts to model the evolution of a mountain snowpack is the lack of a meaningful comparison of the model predictions to an actual snowpack. Plots of simulated and measured snowpack parameters are abundant in the literature, but simply comparing these graphs visually does not provide a meaningful or consistent evaluation of the similarity between observed and predicted values. Introducing statistical methods using well-established measures provides a means of quantifying the accuracy of the models.

*Corresponding author address: Christopher C. Lundy, Civil Engineering Dept., Montana State University, Bozeman, MT 59717; tel: (406) 763-4179; email: lundy@mcn.net

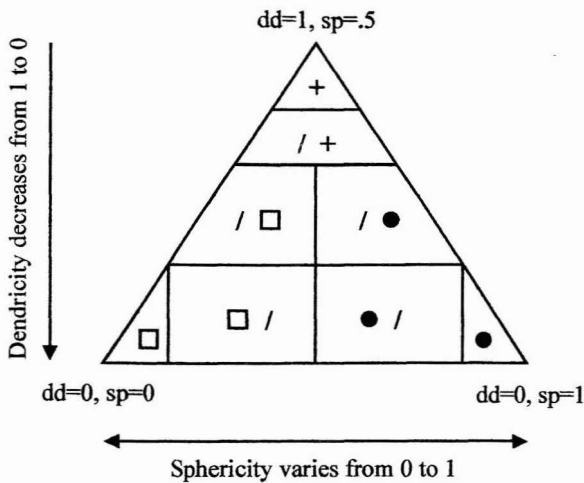


Figure 1. Relation of the model parameters sphericity and dendricity to the ISCI symbols (Colbect *et al.*, 1990).

2. PURPOSE

To date, validation of the snowpack simulation models has not been adequately addressed using an objective and numerical approach. This will continue to limit their improvement and acceptance, since few users are willing to expend the resources necessary to operationalize a snowpack model without first having access to an extensive evaluation of the program.

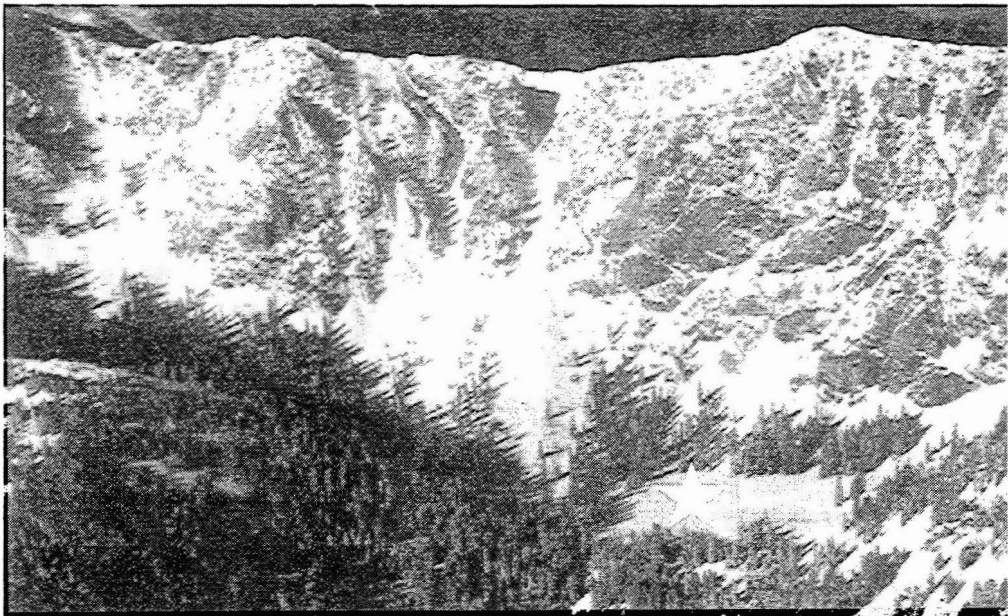


Figure 2. Aerial photo showing location of Wolverine Basin weather station.

To complete an objective validation of SNOWPACK, a weather station was constructed to provide the meteorological parameters necessary to run the model. During the 1999-2000 winter, regular snow profiles were conducted to provide a benchmark for the model output. The meteorological data was then input into SNOWPACK, and a "predicted" snowpack was computed. Finally, the use of statistical methods allow a thorough and objective comparison of the model output to field observations.

3. METHODS

3.1 *Description of Field Site*

The field research site is located approximately one kilometer north of the Bridger Bowl Ski Area near Bozeman, Montana, in an area known as Wolverine Basin. This region falls within the intermountain or transition climate regime as defined by McClung and Schaerer (1993). An average annual snowfall of approximately 6.5 m is measured at the adjacent ski area. The site is a large, open, relatively flat meadow situated at an elevation of 2240 m (Figure 2).

A weather station was constructed at the site to obtain the necessary input data for the SNOWPACK model. From 17 November 1999 to 6 April 2000, the following were collected on 30 minute intervals:

- Air temperature
- Snow surface temperature
- Relative humidity
- Wind speed
- Reflected shortwave radiation
- Total snow depth.

Additionally, a temperature-measurement array was utilized to obtain a real-time temperature profile within the snowpack. It was constructed from a 3 m PVC tube fitted with thermocouples on 5 cm intervals, standing vertically with the bottom thermocouple at ground level. The PVC tube is filled with foam so that the entire unit has a low thermal conductivity.

3.2 Collection of Snow Profile Data

On a weekly basis, snowpits were excavated in undisturbed, fenced-off snow near the instrument tower. Density was measured with a triangular density box of known volume, and weighed on a portable digital scale. Grain diameter and the ISCI grain classification (Colbeck *et al.*, 1990) were determined with a 20× Pentax hand lens. Observations were recorded every 10 cm through the full depth of the snowpack.

3.3 Comparison of the Predicted Snowpack to the Observed Snowpack

While SNOWPACK includes a graphical user interface that presents a visual description of the snowpack predicted by the model, a simple visual comparison of the model results to the snowpit data is not adequate to objectively evaluate the model. A better method is to employ familiar statistical measures to evaluate the level of agreement between the modeled and observed parameters:

- Temperature
- Density
- Grain size
- Grain type.

The model itself was configured to output the snow profile data for each day at 1100 hours, which corresponded with the time at which the snowpits were typically performed in the field. It is this profile data that is compared to the weekly snowpit observations.

Before any comparison of the numeric data can be undertaken, predicted model data must be calculated at the same depths within the snowpack as the snowpit observations (i.e. on 10 cm intervals). To accomplish this task, a technique was devised by (Lehning *et al.*, this volume) to obtain model results at desired locations. First, to develop a basis for comparing the predicted and observed snowpack heights, even if the total depths differ, a normalization of depth is performed. Next, linear interpolation is applied between normalized heights to obtain model results at depths that correspond with the observed measurements. Since grain type is not measured on a continuous numeric scale, interpolation is not possible and the only feasible technique is to use the grain type value that occurs at the height closest to the observed location.

3.4 Statistical Measures of Model Performance

A detailed summary of statistical descriptors that evaluate a model's ability to match an observational dataset is provided in Imam *et al.* (1999). These goodness-of-fit indicators fall loosely into one of two categories: residual-based and statistical association-based. By employing statistical measures from both categories, a more complete description of the model's performance is obtained.

In the residual-based category, the mean bias (B) and the root mean square error ($RMSE$) are commonly used for model verification (Sorooshian *et al.*, 1983; Imam *et al.*, 1999). The mean bias indicates the direction of the expected model error, and is a useful measure of a model's tendency towards overestimation or underestimation. In contrast, the $RMSE$ estimates the expected magnitude of error associated with a model's prediction.

Among indicators of statistical association, Pearson's correlation coefficient, r , is perhaps the most common. The correlation coefficient has an upper bound of 1, indicating perfect positive linear correlation, and a lower bound of -1 , corresponding to negative linear correlation. Although Pearson's r is familiar among many scientists and often used for validation purposes, other researchers have provided arguments against its use for model verification (Imam *et al.*, 1999; Imam, 1994; Fox, 1981; Willmott, 1981). Another measure of association is the Nash-Sutcliffe coefficient of efficiency, E , which is upper

	<i>n</i>	<i>B</i>	<i>RMSE</i>	<i>r</i>	<i>E</i>	<i>d</i>
Temperature	196	0.10 °C	0.97 °C	0.90	0.77	0.95
Density	177	-48.15 kg/m ³	69.15 kg/m ³	0.85	0.30	0.76
Grain Size	179	-0.08 mm	0.42 mm	0.30	0.05	0.38

Table 1. Statistical measures comparing the predicted and observed snowpack parameters.

bounded by 1 and yields higher magnitudes with increasing model accuracy. The coefficient of efficiency can also assume negative values, which have a less intuitive interpretation. To address this, Willmott and Wicks (1980) proposed an index of agreement, *d*, which is bounded by [0 1] so that it does not produce negative values.

The preceding statistical indicators are applicable to numeric parameters such as temperature, density, and grain size, but not to grain type which is measured on a categorical scale. Instead, Cramer's Phi (*V*) and Sakoda's adjusted contingency coefficient (*C**) (Agresti, 1996) are used which quantify the degree of association between two categorical variables. These Chi-Square (χ^2)-based coefficients are convenient since they range from 0 to 1 with higher values indicating better association. Both *V* and *C** are computed twice; once for the majority grain type *F*₁, and again for the minority classification *F*₂.

The preceding discussion serves to emphasize that there is no single statistical descriptor that will effectively assess a model's ability to predict observed data. Only by combining the merits of several different measures can a complete model evaluation be obtained.

4. RESULTS AND DISCUSSION

Table 1 summarizes the results of the descriptive statistical analysis for the numerical parameters temperature, density, and grain size. While there are no hard-and-fast rules for interpreting these measures, the utilization of several different statistical descriptors allows meaningful conclusions to be drawn regarding the performance of the SNOWPACK model. Additionally, the analysis provides a consistent framework for comparing different versions of the model, or evaluation of the model using datasets from different seasons or geographical locations.

4.1 Temperature

By all statistical measures, the SNOWPACK model predicts the snow cover temperatures

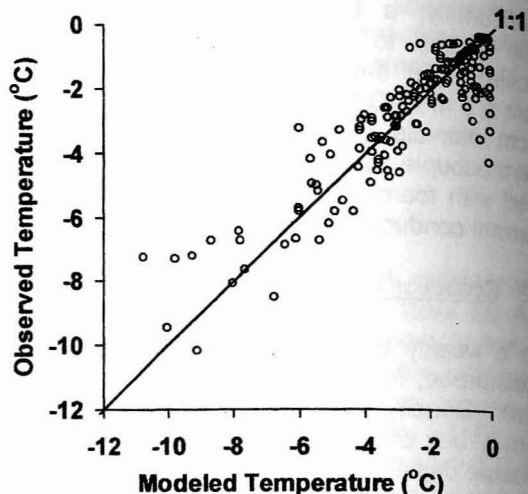


Figure 3. Observed versus modeled temperature.

reasonably well. The *RMSE* is only 0.97 °C, and the mean bias *B* shows very little tendency toward over or underestimation. Since the temperature measurements are accurate to within ± 0.5 °C, the *RMSE* demonstrates that the model does a good job of predicting temperature.

The measures of association *r* and *d* are nearly one and the coefficient of efficiency *E* is fairly high, suggesting a high degree of correlation between the observed snowpack temperatures and those predicted by the model. Referring to Figure 3, the observed-predicted data pairs lie very close to the 1:1 line which represents perfect agreement. It should also be noted from Figure 3 that there are several instances where SNOWPACK predicts isothermal temperatures, but colder temperatures were measured. The accuracy of the model also tends to diminish as the temperatures decrease.

4.2 Density

It is apparent from the *RMSE* of 69.15 kg/m³ that SNOWPACK has some difficulty predicting the snow cover density. A mean bias *B* of -48.15 kg/m³ confirms that the majority of the error

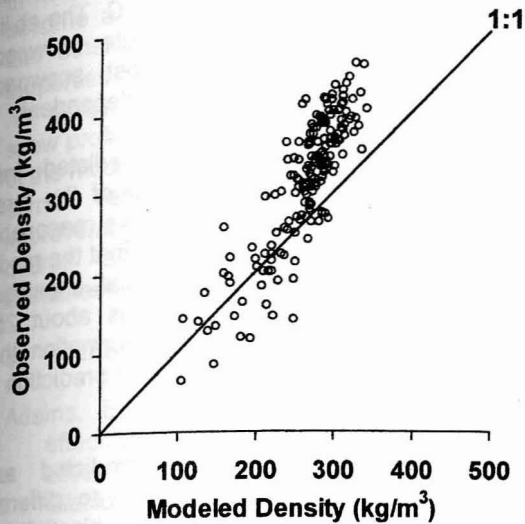


Figure 4. Observed versus modeled density.

present in the model's prediction is due to a consistent underestimation of the measured snowpack density. Figure 4 illustrates that the observed-predicted data pairs are somewhat centered on the 1:1 line until the observed density approaches 250 kg/m³. For densities greater than 250 kg/m³, the data diverges markedly from the 1:1 line as the model increasingly underpredicts the measured density.

Despite the discrepancy in the magnitudes of the predicted and observed density, the correlation measures yielded more positive results. The coefficient of efficiency E is fairly low at 0.30, but r and d are fairly high at 0.85 and 0.76, respectively. The high r shows that there is a tendency toward a linear relationship; however, a look at Figure 4 reveals that the linear relationship does not follow the 1:1 line. This means that the general theory of the model may be sound, but that an adjustment of parameters may be necessary to improve the accuracy of density prediction.

4.3 Grain Size

The comparison of predicted and observed grain size was difficult and does not supply conclusive information about the accuracy of the model. All of the statistical measures for grain size given in Table 1 give low values. The $RMSE$ is large at 0.42 mm, but shows little or no bias with B equaling -0.08 mm. The statistical association scores r , E , and d are well below acceptable values.

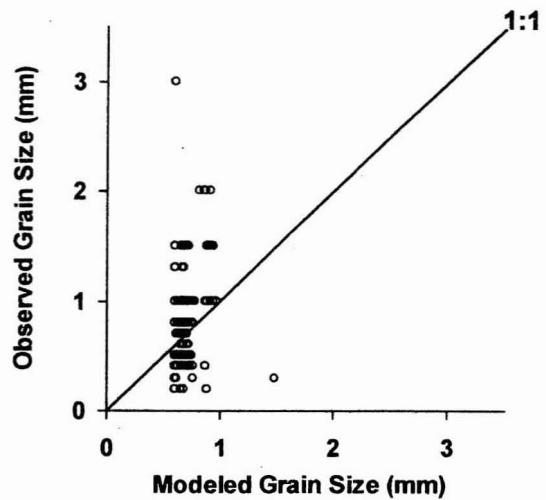


Figure 5. Observed versus modeled grain size.

The problem in this comparison lies partly in the differing definitions of grain size used by the model and the field observer. The model chooses 0.6 mm as the size of new snow particles, and allows only the growth of these grains. As a result, it is evident from Figure 5 that SNOWPACK never predicts a grain diameter less than 0.6 mm, but on several occasions, grain sizes less than this were reported in the field observations. Furthermore, grain sizes greater than 1 mm are seldom predicted by the model, but were routinely observed in the field. SNOWPACK uses a grain diameter that is independent of the crystallography; essentially all grains are treated as spheres. This is not true of the field observer, whose estimation of grain size is often tied to the shape of the snow grain, especially for faceted crystals. The estimation of grain size in the field is unavoidably a subjective measurement. Despite these contrasting definitions of grain size and the possibility of observer error, the low statistical measures suggest the need for model refinement in this area.

4.4 Grain Type

Cramer's V for both $F1$ and $F2$ are both low and indicate only a weak correlation between the predicted and observed values (Table 2). The adjusted contingency coefficients C^* of 0.71 and 0.66 suggest a somewhat larger amount of association is present, but the relationship is still not very strong.

	<i>n</i>	<i>V</i>	<i>C*</i>
Majority (<i>F1</i>)	206	0.41	0.71
Minority (<i>F2</i>)	206	0.34	0.66

Table 2. Statistical association measures for the majority and minority grain types.

Similar to grain size, SNOWPACK uses an alternate grain classification scheme, namely dendricity and sphericity. Since there is little basis for estimating these parameters during field observation, the model must choose a standard ISCI grain shape according to various combinations of dendricity and sphericity. Obviously, this is not an ideal or particularly accurate procedure. A more desirable technique would be to develop a common classification system employed by both the SNOWPACK model and the observer. Another alternative is to focus less on the crystallography, and more on the microstructural parameters that actually define the physical properties of the snow. Crystal type is observed in the field primarily as an indicator of the degree of bonding and strength possessed by the snow. For instance, when rounded grains are encountered, a high degree of bonding and strength is assumed; the opposite is true when faceted crystals are observed. If a reliable field method was developed for measuring bond size, and perhaps bond density, the classification of grain shape may be less important.

5. CONCLUSION

From 17 November 1999 to 6 April 2000, meteorological data were collected from a mountain weather station adjacent to Bridger Bowl Ski Area near Bozeman, MT. During the same period, full snowpack profiles were performed on a weekly basis within a short distance from the weather station. By running the SNOWPACK model using the collected weather data and comparing the output to the snow profiles, a thorough evaluation of the predictive capabilities of the model was possible. Statistical tests were utilized to make the comparison objective and consistent.

The statistical measures utilized in the analysis indicate that the SNOWPACK model predicts the temperature profile within the snow cover fairly accurately. However, inspection of the plot of predicted and measured temperature reveals a diminishing model accuracy with colder

temperatures, and additional difficulty when the modeled temperatures approach 0 °C. The ability of the model to effectively simulate snowpack temperature is crucial since most snowpack processes are strongly temperature dependent.

Though snowpack density is predicted less successfully than temperature, two of the three measures of statistical correlation give reasonable values. The data also demonstrates that the model significantly underpredicts the actual snowpack density when the densities exceeds about 250 kg/m³. Still, our results provide information that might be useful for improving density prediction in future versions of the model.

A meaningful comparison of predicted and observed grain size is difficult due to different definitions of grain size and the subjectivity of human measurement. The statistical measures gave generally poor results and indicate little correlation between the simulated and observed values. Therefore, the results of the comparison indicate not only the need to improve the model, but also the utility of a more standardized observation technique using a definition of grain size similar to that employed by SNOWPACK.

Since snow crystallography is not measured on a continuous or numeric scale, the use of alternative statistical measures is required. The results of the statistical association measures demonstrate a weak correlation between the modeled and observed grain types. Like grain size, crystal shape is a subjective observation; therefore developing a classification scheme that can be both utilized by the model and accurately measured in the field would be advantageous.

This analysis does not evaluate the model's prediction of surface hoar since we inactivated the routine during model execution. This portion of the model is still under development (Lehning *et al.*, 1998) and reportedly overpredicts surface hoar occurrence (Pielmeier *et al.*, 2000). Another area of the model needing improvement is wet snow metamorphism. Currently, SNOWPACK has only rudimentary provisions for simulating wet snow metamorphism. Since this routine influences the grain size, crystal shape, and density, its effects are included in the analysis. Future work on the wet snow capabilities of SNOWPACK will be important for ablation prediction, hydrological purposes, and for applying the model to warmer, maritime climates.

In its present form, the SNOWPACK model can become a useful tool for avalanche forecasters and other practitioners who need to know the properties and structure of the snow cover, but do not always have the ability to conduct frequent snow profiles in a given location. Of course, users of the model must understand the limitations of the current version of SNOWPACK. Future model improvements and validation should increase the overall accuracy of SNOWPACK and its usefulness as a tool for snow practitioners.

5. REFERENCES

- Adams, E. E. and Sato, A. 1993. Model for effective thermal conductivity of a dry snow cover composed of uniform ice spheres. *Ann. Glaciology* 18, 300-304.
- Agresti, A. 1996. *Introduction to Categorical Data Analysis*. John Wiley and Sons, New York.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. 1978. *Statistics for Experimenters*. John Wiley and Sons, New York.
- Colbeck, S. C., Akitaya, E., Armstrong, R., Gubler, H., Lafeuille, K., Leid, D., McClung, D., and Morris, E. 1990. The international classification for seasonal snow on the ground. *International Commission on Snow and Ice*, 23 pp.
- Fox, D. G. 1981. Judging air quality model performance: A summary of the AMS workshop on dispersion model performance. *Bull., Am. Meteorol. Soc.* 62, 599-609.
- Imam, B., Sorooshian, S., Mayr, T., Schaap, M., Wosten, H., and Scholes, B. 1999. Comparison of pedotransfer functions to compute water holding capacity using the van Genuchten model in inorganic soils. *IGBP-DIS Working Paper #22*.
- Imam, B. 1994. Non-linear uncertainty analysis for multiple criteria natural resource decision support system. Ph.D. Dissertation, University of Arizona.
- Lehning, M., Bartelt, P., Brown, R. L., Russi, T., Stockli, U., and Zimmerli, M. 1998. A network of automatic weather and snow stations and supplementary model calculations providing snowpack information for avalanche warning. *In Proc. 1998 International Snow Sci. Workshop*, Sun River, Oregon, 225-233.
- McClung, D. and Schaerer, P. 1993. *The Avalanche Handbook*. The Mountaineers, Seattle, Washington. 271 pp.
- Nash, J. E. and Sutcliffe, J. V. 1970. River flow forecasting through conceptual models, Part I – A discussion of principles. *J. Hydrol.* 10, 282-290.
- Pielmeier, C., Schneebeli, M., and Stucki, T. 2000. Snow texture: a comparison of empirical versus simulated texture index for alpine snow. *Ann. Glaciol.* 32, in press.
- Sorooshian, S., Gupta, V. K., and Fulton, J. L. 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall runoff models: influence of calibration data variability and length on model credibility. *Water Resources Research*, 19(1), 251-259.
- Willmott, C. J. 1981. On the validation of models. *Phys. Geography* 2, 184-194.
- Willmott, C. J. and Wicks, D. E. 1980. An empirical method for the spatial interpolation of monthly precipitation within California. 1, 59-73.