COMPARISON OF MODELLED AND MEASURED POINT SNOW PROFILES: A TOOL FOR VALIDATING SNOW-COVER MODELS OF THE NEXT GENERATION?

Charles Fierz[1]*, Franziska Gerber[1,2], and Michael Lehning[1,2]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2]CRYOS, School of Architecture, Civil and Environmental Engineering,
Ecole Polytechnique Fédérale de Lausanne, Switzerland

ABSTRACT: ProfEval, a method to objectively compare measured with computed snow profiles, was proposed back in 2001. The method includes algorithms to compare measured point profiles such as snow temperature profiles to model results. However, the behaviour and characteristics of the agreement score for point profiles has not been fully investigated yet. In future, outputs of instruments such as the SnowMicroPen will be increasingly compared to simulation results of next generation snow-cover models. As these objective measurements now provide data at the sub-centimeter scale just as we have them from the computed profiles, they are perfectly adapted to this kind of comparison. In addition, we ask whether we need to consider additional parameters to assess the goodness of fit. The results of our analysis show that more than one evaluator is needed to fully capture all relevant aspects of a model – measurement comparison. Moreover, we now can quantify the range within which the agreement score can be used.

KEYWORDS: snow, snow-cover modelling, snow profiles, model validation.

1. INTRODUCTION

Snow cover simulations are intensively used for avalanche forecasting. To calibrate such simulations a comparison algorithm for numerical model profiles versus snow pit profiles has been proposed by Lehning et al. (2001). The algorithm gives a quantitative overall agreement – disagreement score for various parameters measured in a snow pit such as snow temperature profiles. The method is designed for validating model output against measured data, but not to test how well models reproduce specific physical processes. While algorithms are available for both point and bulk profiles, here we only focus on point profiles as this part of the method may be mostly used in future with high-resolution objective measurements becoming more and more available.

In the past, the above score has been sometimes misinterpreted as a percentage agreement like in multi-variate model regression. Here we adapt, transform and evaluate the agreement – disagreement score proposed by Lehning et al. (2001) for point profiles and give a range above which the agreement score is meaningful, and below which it must be discarded. We then evaluate and discuss the helpfulness of our agreement score as compared to other well known scores that evaluate observation – model agreement mostly for time series (for example, see Legates and McCabe, 1999).

* *Corresponding author address:*
Charles Fierz, WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland;
tel: +41-81-417-0165; fax: +41-81-417-0110;
email: fierz@slf.ch

2. DATA

For our analysis we extracted observed snow profiles from the profile database of the WSL Institute for Snow and Avalanche Research SLF. In addition, we use SnowMicroPen (SMP) and snow density measurements from the Weissfluhjoch study plot located in the Eastern Swiss Alps above the town of Davos at an elevation of 2540 m. The measured profiles were compared to profiles calculated with various versions of the snow-cover model SNOWPACK (see models.slf.ch).

3. AGREEMENT SCORE

In this study we use an adapted version of the comparison algorithm by Lehning et al. (2001) where a calculated profile (slave) is compared to the obseved one (master). After adjusting and interpolating for mismatches in snow depth and heigths of measurements, respectively, the distance $d^X_{profile}$ is given as:

$$d^X_{profile} = tanh\left\{\frac{1}{N}\frac{\sum_{i=1}^{N}\left|X_i^M - X_i^S\right|}{max[1, \max(X^M) - min(X^M)]}\right\},\ N \geq 2 \quad (1)$$

where $X_i^{M,S}$ are the point values of the master (M) and slave (S) profile, respectively, N is the number of points in the master profile, and X^M runs over all values of the master. The agreement score a is then given by:

$$a = 1 - d_{profile}^{X}. \quad (2)$$

Finally we transform the agreement score to:

$$a_{trans} = \frac{\exp(a) a^{\exp(1)}}{\exp(1)} \quad (3)$$

4. RESULTS AND DISCUSSION

4.1 Application range for a

To investigate the application range of a we use the modified Nash-Sutcliffe Efficiency $mNSE$ given by (Legates & McCabe Jr., 1999; Nash & Sutcliffe, 1970):

$$mNSE = 1 - \frac{\sum_{i=1}^{N} \left| X_i^M - X_i^S \right|}{\sum_{i=1}^{N} \left| X_i^M - \overline{X^M} \right|} \quad (4)$$

where $\overline{X^M}$ represents the mean of the master profile. If $mNSE < 0$, the slave performs worse than $\overline{X^M}$ in representing the master profile.

The agreement score a gives information on the overall mean offset of the modeled profile (slave) with respect to the measured profile (master). Normalizing by the maximum range of the master profile, the offset is given as a measure of the averaged overall offset relative to the maximum range of the master profile. Comparing results of the agreement score a to results of $mNSE$ shows that for agreement scores a less than about 0.6, $mNSE$ is in most cases less than zero. This result clearly shows that the value of the agreement score cannot be thought of as a percentage agreement.

4.2 Does one score suffice?

The agreement score a is only a measure of the overall offset of a profile but does not give information on the form agreement of both master and slave profiles. Thus we use the Pearson correlation coefficient, c, as a second measure for the goodness of fit with respect to the shape of the profiles. However, as Legates & McCabe Jr. (1999) say, correlation measures may give a good agreement even though there is little agreement. In addition, c is not defined for profiles showing a constant value because the standard deviation is zero, which further reduces its usefulness. In summary, while we would like to use the correlation to characterize similarity in shape, the parameter is not well chosen from a numerical point of view. Nevertheless, we are convinced a second parameter is needed to fully characterize the goodness of fit of slave and master profiles.

This short summary of a work still in progress shows that the quantitative evaluation of snow-cover models remains a tricky task. Because of many deficiencies and problems associated with their calculation, score parameters are difficult to handle and interpret. For example, the more familiar Root Mean Square Error, $RMSE$, and Mean Absolute Error, MAE, are often used in model assessments. Note, however, that these scores are not normalized and this adds one more difficulty when applied, for example, to a time series of profiles.

REFERENCES

Legates, D. and G. J. McCabe Jr., 1999: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241.

Lehning, M., C. Fierz, and C. Lundy, 2001: An objective snow profile comparison method and ist application to SNOWPACK. *Cold Regions Science and Technology*, 33, 253–261.

Nash, J. E. and J. V. Sutcliffe, 1970: River flow forecasting trough conceptual models part I – A discussion of principles. *Journal of Hydrology*, 10, 282–290..