*Biogeochemistry of Seasonally Snow-Covered Catchments* (Proceedings of a Boulder Symposium, July 1995). IAHS Publ. no. 228, 1995.


# Distributed snowmelt modeling using a clustering algorithm

**ROBERT F. HARRINGTON**
*Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721, USA*

**KELLY ELDER**
*Institute for Computational Earth System Science, University of California, Santa Barbara, California 93106, USA*

**ROGER C. BALES**
*Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721, USA*

**Abstract** Methods for spatially distributed physical modeling of snowmelt are presented. Estimation of snow-water equivalence using regression trees, and distributed snowmelt modeling using iterative clustering are described; these methods are applied to Emerald Lake basin, a 120-ha seasonally snow covered basin in California. Modeled snowmelt and basin discharge substantially agree, yet modeled and observed snow covered area agree only over ~65% of the basin after 5 weeks of melt, highlighting the importance of using snow cover patterns to assess model performance.

## INTRODUCTION

Conceptual models have become the standard tool for operational snowmelt runoff modeling (WMO, 1986); however, with the increasing availability of faster computers, geographic information systems, and remote-sensing data, spatially distributed physical models are becoming attractive for some research applications (Bloschl *et al.*, 1991; Wigmosta *et al.*, 1994). For the purpose of calculating hydrologic inputs to a hydrochemical model, detailed specification of the hydrologic pathways is required (Moldan & Cerny, 1994), it is desirable to retain as much spatial detail as possible so that meltwater can be routed through the proper soil unit. In this paper, a method is presented for extending a one-dimensional physical snowmelt model into a spatially distributed model by classifying the basin into areas of similar physical characteristics, while retaining the spatial resolution of the spatially distributed input data. The method is applied to a small basin, and model performance is assessed by comparison to basin discharge, snow covered area (SCA), and snow-water equivalence (SWE).

## METHODS

### Field measurements

The snowmelt modeling undertaken here treats the snowpack at maximum accumulation

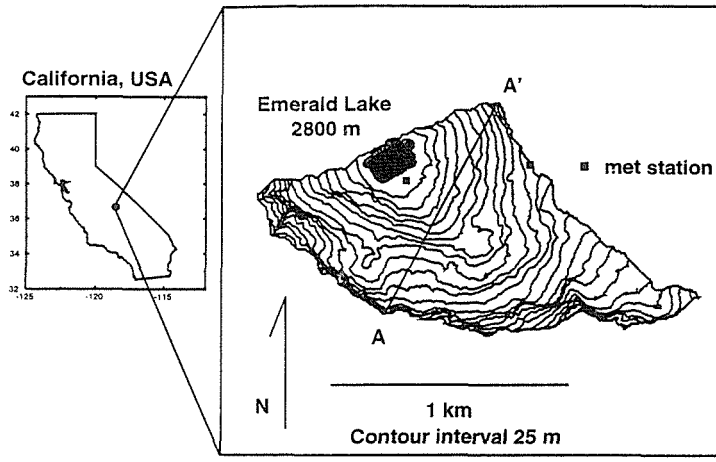


**Fig. 1** Emerald Lake basin. Transect A-A' refers to snow-water equivalences shown in Fig. 4.

as an initial condition for the model; no attempt to model the temporal evolution of the snowpack prior to maximum accumulation is made; therefore, the data required are snowpack properties at maximum accumulation, and a time series of the meteorological variables that drive subsequent snowmelt. A data set consisting of snow depth and snow density measurements (Elder *et al.*, 1991), stream discharge measurements (Kattelmann & Elder, 1991), and hourly meteorological measurements (Marks *et al.*, 1992) from Emerald Lake basin, Sequoia National Park, California (Fig. 1) for 1987 was used for this study (Dozier *et al.*, 1989). Emerald Lake basin is a 120-ha glacial cirque ranging in elevation from 2800 m a.m.s.l. at the Emerald Lake outlet to 3416 m a.m.s.l. on the southern boundary of the basin. The basin is sparsely vegetated, with poorly developed soils and much exposed granitic bedrock and talus. The topography of the basin is described by a 5-m resolution digital elevation model (DEM) derived from aerial photography.

Snow surveys were conducted on 18 April (day of year (DOY) 108), 22 May (DOY 142), and 5 June (DOY 156). The initial condition for the snowpack was derived from the 18 April survey; however, because the 18 April survey was slightly after the peak snow accumulation, the model was initiated with 7.6 cm additional snow-water equivalence (SWE) in the basin to account for ablation between peak accumulation (approximately 1 April) and 18 April. The locations of measurements made for the 18 April survey are shown in Fig. 1. The surveys on 22 May and 5 June were used to assess model performance. Water storage within the snowpack derived from the surveys is summarized in Table 1.

**Table 1** Summary of snowpack water storage.

| Date | SCA (ha) | SCA (%) | Volume (1000 $m^3$) | Mean SWE (m) |
|---|---|---|---|---|
| 18 April | 107.2 | 89.3 | 721.3 | 0.60 |
| 22 May | 59.5 | 49.5 | 287.3 | 0.24 |
| 5 June | 41.8 | 34.8 | 157.8 | 0.13 |

## Solar radiation

The spatial distribution of incoming radiation was calculated using the methods of Dozier & Frew (1990) as implemented in the Image Processing Workbench (Frew, 1990). The DEM was used to calculate slope, azimuth, solar illumination angle, horizons, and view factors for incident radiation from the sky and reflected from surrounding terrain, producing a map of incident solar radiation under clear-sky conditions for 1 May, producing a map of potential radiation at each pixel in the basin. The potential radiation map is used to determine the ratio of potential radiation at the meteorological station to the potential radiation at each pixel in the DEM, providing a factor by which the record of measured incoming solar radiation is multiplied to estimate the incoming solar radiation at each pixel. Changes in solar radiation due to cloudiness, changes in atmospheric properties, or changes in solar zenith angle are recorded in the hourly solar radiation measurements and treated as uniform across the basin; however, self-shading conditions underneath north-facing cliffs do not change over the course of the model run because the potential radiation map was calculated for 1 May only.

## Snow-water equivalence

The SWE distribution was modeled using the binary regression tree method (Breiman *et al.*, 1984), which operates as follows. Each independent variable is recursively divided by binary splits until the variable and the split are found that maximize the change in node deviance. This process is continued until a stopping rule is met, giving a set of terminal nodes. The independent variables used were based on physical parameters known to affect snow distribution (net solar radiation, elevation, slope, vegetation type, and soil or substrate type). SWE estimates based on maximizing the number of terminal nodes provide the best estimate based on several simple metrics, but also give complex spatial fields of SWE that are difficult to apply to snowmelt modeling, because of the large number of resulting classes; hence, trees with ten terminal nodes were used (i.e. ten discrete SWE values). SWE distributions were conditioned on snow covered area (SCA) determined by SPOT multispectral imagery (18 April) and aerial photography (22 May and 5 June), thereby ensuring that areas observed to be snow free in the imagery are snow free in the estimated SWE distribution. A complete treatment of SWE estimation using regression trees is given in Elder *et al.* (1995, this volume).

The snow density at peak accumulation was considered to be uniform across the basin and equal to the mean value observed in the snow pits (348 kg $m^{-3}$).

## Other meteorological variables

Air temperature was treated as a linear function of elevation, using a lapse rate of −4.0°C. Wind speed, relative humidity, and downward thermal radiation were considered uniform over the basin.

## Snowmelt simulation

The snowmelt simulation was carried out by classifying the basin into areas having similar SWE, incident solar radiation, and elevation, and modeling the one-dimensional (vertical) movement of meltwater for each class. The classification was carried out using the iterative clustering algorithm given by Richards (1993) and implemented in the geographic information system GRASS (Westervelt *et al.*, 1990). Each pixel in the DEM has an elevation, a SWE value, and a multiplier for incoming solar radiation associated with it. Each pixel location in the basin is a point in the three-dimensional space of pixels with coordinates defined by these three quantities. Each pixel is associated with a "cluster" of other pixels having nearby coordinates. Prior to applying the clustering algorithm to snowmelt modeling, digital maps of SWE, incoming solar radiation, and elevation were rescaled to the same range of values and histogram stretched such that the data were uniformly distributed over the rescaled range of the data. The algorithm is initiated by choosing a set of candidate mean vectors for a specified number of clusters (50 in this case); each pixel is assigned to the nearest candidate mean, and the mean vectors are recalculated based on the pixels assigned to each cluster. If the mean vectors of two clusters are too close to each other, the clusters are combined, and if a cluster becomes too small (425 $m^2$ in this case), the cluster is combined with the nearest cluster to it. The process is repeated until the classification is stable to a specified tolerance.

The result of the classification is a set of regions, each with a mean elevation, mean SWE, and mean incoming solar radiation. A one-dimensional finite difference model of heat, mass, and momentum transport in the snowpack (Jordan, 1991) was run for each class, producing snowmelt discharge from the bottom of the snowpack for each class. The discharge from each class was then summed into a basin-wide discharge. This procedure has the advantage that the classes need not be spatially contiguous, so a single one-dimensional model run is used to calculate snowmelt at multiple locations with similar attributes in the basin.

## RESULTS

The modeled basin-wide total snowmelt issuing from the bottom of the snowpack is compared to the discharge from the Emerald Lake outlet in Fig. 2. Because the basin is small and soils are thin or absent, subsurface storage is small, and daily volumes of modeled snowmelt and lake discharge are similar. However, the modeled snowmelt shows more variability than lake discharge because of the modulating effect that within-basin storage (such as subsurface and ponded water) has on the lake discharge relative to snowmelt (Fig. 2(a)). Using the efficiency criteria of model performance of Nash & Sutcliffe (1970), 0.52 of the variance in the lake discharge is explained by the modeled snowmelt. Applying a linear reservoir model to the modeled snowmelt (Chow *et al.*, 1988) produces a Nash and Suttcliffe efficiency of 0.88. Comparison of cumulative lake discharge and snowmelt indicates that the seasonal mass balance for the model and the observed stream discharge is good (Fig. 2(b)).

Modeled and observed SCA agree with respect to larger scale features such as the southeast facing slopes in the northern portion of the basin becoming snow free

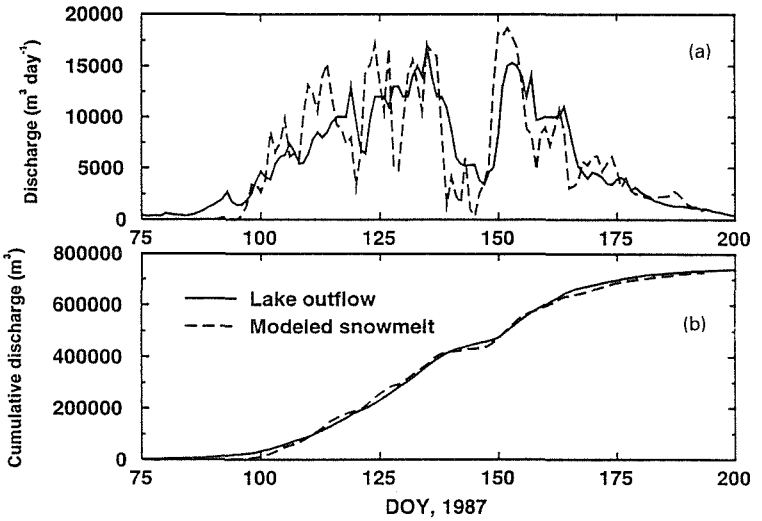


**Fig. 2** Comparison of lake outflow and modeled snowmelt. (a) Daily discharge; (b) cumulative discharge.

relatively early in the melt season. However, other features, such as the loss of snow cover from the ridge along the southern boundary of the basin, and the pattern of snow cover on the cliffs and talus fields on the northeast facing parts of the basin are not captured by the model (Fig. 3). Table 2 summarizes the percentage areal agreement and error between the observed and modeled SCA.

Three estimates of ablation were available for comparison: (1) change in snowpack storage as reflected by the sequential snow surveys; (2) change in snowpack water storage as reflected by modeled snowmelt and evaporation from the snowpack; and (3) estimated water exiting the basin by stream discharge and evaporation. These data are

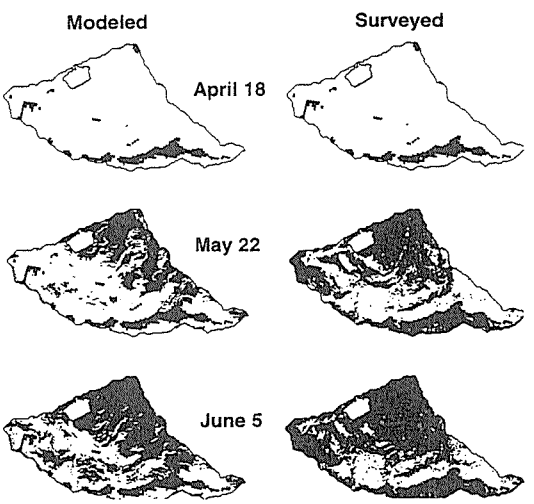


**Fig. 3** Comparison of observed and modeled SCA. White areas are snow covered; black are snow free.

**Table 2** Comparison of observed and modeled SCA as percent of basin area covered. Agreement was ~100% for April 18.

| Date | Model and observation agree | Model snow free; observed to be snow covered | Model snow covered; observed to be snow free |
|---|---|---|---|
| 22 May | 67% | 9% | 24% |
| 5 June | 65% | 11% | 24% |

compared in Table 3, where evaporation is calculated from the model, and the snow survey based ablation is from Elder (1995, this volume). Groundwater discharge from the basin was assumed to be negligible (Kattelmann & Elder, 1991); therefore, the total water exiting the basin in Table 3 is the sum of modeled evaporation and lake outlet discharge. The model overestimated SCA on 22 May and 5 June. However, the disagreement in ablation between model and survey for the periods 18 April-22 May and 22 May-5 June, were 15% and 4% respectively, suggesting that while basin-wide mass balance is being adequately modeled, pointwise ablation is not. Transects comparing modeled and surveyed SWE across the basin (Fig. 4) suggest that errors in modeled SWE were in part compensated for by errors in SCA; hence the mass balance given in Table 3 shows less error than the error in SCA given in Table 2. For example, on 5 June, the observed SCA is greater than predicted by the model, but the SWE is generally greater in the modeled snowpack in the areas where the model predicts snow coverage.

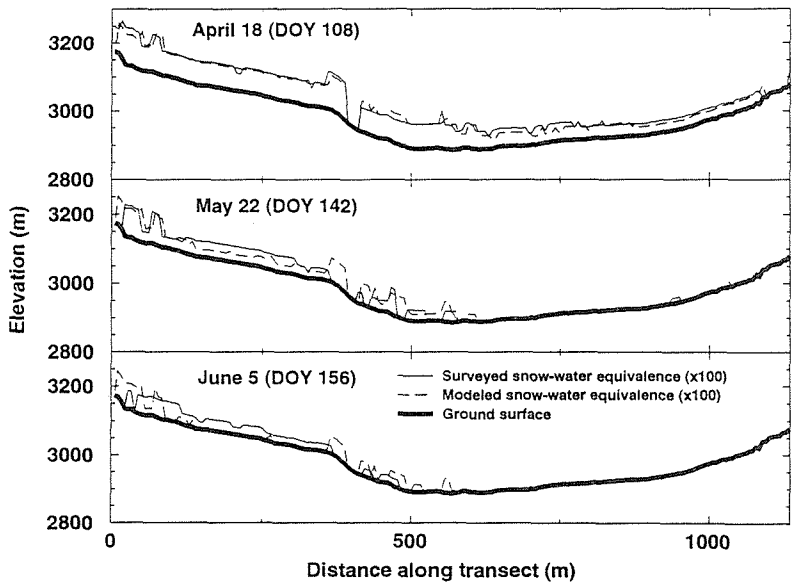


**Fig. 4** Cross sectional comparison of surveyed and modeled SWE along transect A-A'; zero distance along transect is at southwest end of transect (A)(see Fig. 1). SWE values are overlaid on the topographic profile.

**Table 3** Basin water balance. Change in water storage in snowpack estimated from snow surveys and calculated by model compared to water leaving the basin due to modeled ablation, and modeled evaporation plus observed discharge.

| | Change in snowpack storage (1000 $m^3$): | | Water exiting basin (1000 $m^3$): | | |
|---|---|---|---|---|---|
| Time interval | Surveyed | Modeled | Evaporation | Stream discharge | Total |
| 15 March-18 April | 92.8 | 103.7 | 35.2 | 80.4 | 115.6 |
| 18 April-22 May | 434.0 | 370.1 | 15.7 | 357.1 | 372.8 |
| 22 May-5 June | 128.5 | 138.0 | 4.6 | 129.7 | 134.4 |

## DISCUSSION

The water balance (Table 3) and snowmelt runoff (Fig. 2) simulated by the model were satisfactory, yet the changing pattern of snow cover was not well simulated (Fig. 3), reiterating the conclusion expressed by Kirnbauer *et al.* (1994) that snow-cover depletion patterns play a key role in assessing the performance of distributed snowmelt models. For example, Fig. 4 shows that on 5 June, the model predicted that steep northeast facing slopes would still be snow covered, whereas these areas were observed to be snow free. This suggests a problem in the initial regression-based SWE estimate, the reclassification by clustering, or the calculated ablation for these areas. Comparison of the regression-based SWE estimate and the reclassified SWE distribution shows that some of the largest discrepancies between the two are in these regions, with a loss of up to 24 cm SWE in going from regression-based SWE to reclassified SWE. This is expected, because these areas have the highest SWE accumulations in the basin; hence the clustering algorithm pulls some of the pixel values in toward smaller cluster means. Since the solar radiation distribution did not account for changes in self-shading over the course of the model run, this is a possible explanation for the discrepancy on steep northeast facing slopes. Lateral flow of melt-water, incorrectly modeled heat exchange from the substrate to the snowpack, or creep along rock slabs may also have contributed to the discrepancy.

The flatter areas on the south and southwest flanks of the basin that become snow-free earlier than the model predicted have very little discrepancy between the tree-based SWE estimate and the reclassified SWE distribution (1-3 cm). In these areas, the snowpack is underlain by talus, which imposes subpixel scale irregularities in snow depth, and local lateral thermal and hydraulic gradients. These conditions cannot be accounted for by a one-dimensional model, and their effect on ablation and on melt is unclear. The discrepancy in snow cover along the ridge along the southern boundary of the basin (Fig. 3) may be due to errors in modeling sensible heat exchange due to using a uniform wind speed across the basin. This idea will be tested in the future by applying an elevational gradient to the measured wind speed, to provide higher wind speeds on the ridgelines.

An additional improvement planned is to incorporate spatially distributed albedo into the energy balance calculations using the algorithm of Marks and Dozier (1992), and to use a more frequently sampled time series of SCA images for evaluation of model performance.

## CONCLUSIONS

The approach presented here is a general, flexible way to integrate spatially distributed data into a snowmelt model without sacrificing spatial resolution. The method also retains detailed information about the thermal and hydraulic profiles of the snowpack, which would be lost in a pixel by pixel application of a two-layer snowmelt model. The method will also be useful for incorporating high resolution remote-sensing data into snowmelt models. The model succeeded in reproducing the pattern of discharge from the basin and the general features of the change in snow cover. However, some details of the changing snow cover were not well reproduced, and the pattern of discrepancies between modeled and observed snow cover aids in identifying weaknesses in the model such as possible effects of substrate, changes in self-shading, or lateral transport effects.

**Acknowledgments** We thank Remigio Gallaraga-Sanchez for his help with radiation calculations. Research support came from the NASA EOS Interdisciplinary Science Program and an American Geophysical Union Horton Fellowship to the first author.

## REFERENCES

Bloschl, G., Gutknecht, D. & Kirnbauer, R. (1991) Distributed snowmelt simulations in an alpine catchment. 1, Model evaluation on the basis of snow cover patterns. *Wat. Resour. Res.* **27**(12), 3171-3180.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984) *Classification and Regression Trees*. Wadsworth, Belmont, California, USA.

Chow, V. T., Maidment, D. R. & Mays, L. W. (1988) *Applied Hydrology*. McGraw-Hill, New York.

Dozier, J. & Frew, J. E. (1990) Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. and Remote Sens.* **28**(5), 993-969.

Dozier, J., Melack, J., Elder, K., Kattelmann, R. & Marks, D. (1989) Snow, snowmelt, rain, runoff and chemistry in a Sierra Nevada watershed. *Final Report, Contract A6-147-32, California Air Resources Board, Sacramento, California.*

Elder, K., Dozier, J. & Michaelsen, J. (1991) Snow accumulation and distribution in an alpine watershed. *Wat. Resour. Res.* **27**(7), 1541-1552.

Elder, K., Michaelsen, J. & Dozier, J. (1995) Small basin modeling of snow water equivalence. In: *Biogeochemistry of Seasonally Snow-covered Catchments* (ed. by K. Tonnessen, M. W. Williams & M. Tranter) (Proc. Boulder Symposium, July 1995). IAHS Publ. no. 228.

Frew, J. E. (1990) The image processing workbench, Ph.D. Thesis, Department of Geography, University of California, Santa Barbara, USA.

Jordan, R. (1991) A one-dimensional temperature model for a snow-cover. Technical documentation for SNTHERM.89. *U.S. Army Cold Regions Research and Engineering Laboratory, Special Report 91-16.*

Kattelmann, R. & Elder, K. (1991) Hydrologic characteristics and water balance of an alpine basin in the Sierra Nevada. *Wat. Resour. Res.* **27**(7), 1553-1562.

Kirnbauer, R., Bloschl, G. & Gutknecht, D. (1994) Entering the era of distributed snow models. *Nordic Hydrol.* **25**(1/2), 1-24.

Marks, D., Dozier, J. & Davis, R. E. (1992) Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada. 1. Meteorological measurements and monitoring. *Wat. Resour. Res.* **28**(11), 3029-3042.

Marks, D. & Dozier, J. (1992) Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada. 2. Snow cover energy balance. *Wat. Resour. Res.* **28**(11), 3043-3054.

Moldan, B. & Cerny, J. (1994) Small catchment research. In: *Biogeochemistry of Small Catchments* (ed. by B. Moldan & J. Cerny), 1-29. Wiley, New York.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models, Part I – discussion of principles. *J. Hydrol.* **10**(3), 282-290.

Richards, J. A. (1993) *Remote Sensing and Digital Image Analysis, An Introduction*. Springer-Verlag, New York, USA.

World Meteorological Organization (1986) Intercomparison of models of snowmelt runoff. *WMO Operational Hydrology Report 23*. WMO, Geneva, Switzerland.

Westervelt, J., Shapiro, M. & Goran, W. (1990) Geographic Resource Analysis Support System version 4.0 users manual. *U.S. Army Construction and Engineering Laboratory, Champaign-Urbana, Illinois, USA.*

Wigmosta, M. S., Vail, L. S. & Lettenmaier, D. P. (1994) A distributed hydrology-vegetation model for complex terrain. *Wat. Resour. Res.* **30**(6), 1665-1680.