ARE MODEL COMPLEXITY AND TRANSFERABILITY ANTITHETICAL? INSIGHTS FROM VALIDATION OF A VARIABLE-COMPLEXITY SNOW MODEL ON NEW CONDITIONS

A.C. Lute[1] and Charles Luce[2]

EXTENDED ABSTRACT

The related challenges of predictions in ungauged basins and predictions in ungauged climates point to the need to develop environmental models that are transferable across both space and time. Hydrologic modeling has historically focused on modelling one or only a few basins using highly parameterized conceptual or physically based models. However, model parameters and structures have been shown to change significantly when calibrated to new basins or time periods, suggesting that model complexity and model transferability may be antithetical.

Spatial analog models provide a simple framework within which to assess model transferability and any tradeoff with model complexity. Using 497 SNOTEL sites in the western U.S., we develop spatial analog models of April 1 SWE (A1SWE) based on mean winter temperature and cumulative winter precipitation. Precipitation and SWE data is taken directly from SNOTEL sites for the years 1991-2011. Temperature data is from the TopoWx dataset [*Oyler et al.*, 2014] for the same SNOTEL sites. The spatial analog models are built using the locfit package [*Loader*, 1999] in R, following the methods of *Luce et al.*, [2014]. The models are local polynomial regressions of mean winter temperature and precipitation on A1SWE such that these three variables are related across space.

The transferability of the models to new conditions (in both space and time) is assessed using non-random cross-validation tests with consideration of the influence of model complexity on transferability. We designed seven non-random cross-validation tests which can be categorized as either temporal, spatial, or spatiotemporal tests as described below:

- Temporal- these tests split the data in time and each sample contains a subset of the period of record at all SNOTEL sites
    - Early/Late- the data was divided into the first ten years of record (Early sample) and the last ten years of record (Late sample)
    - Cold/Warm- the data was divided into the four coldest years west-wide (Cold sample) and the four warmest years west-wide (Warm sample)
    - El Niño/La Niña- data was divided into years with a mean winter Multivariate ENSO Index (MEI) [*Wolter and Timlin*, 1993] value greater than 0.4 (El Niño samples) and years with MEI less than -0.4 (La Niña sample)
- Spatial- these tests split the data in space and each sample contains the full period of record but at a subset of SNOTEL sites
    - North/South- data was divided into all years at sites North of 43°N latitude (North sample) and all years at sites South of 43°N latitude (South sample)
    - East/West- data was divided into all years at sites East of -112.5°W latitude (East sample) and all years at sites West of -112.5°W latitude (West sample)
- Spatiotemporal- these tests combine the above temporal and spatial tests such that each sample is a subset of years at a subset of SNOTEL sites
    - Late West/Early East- data was divided into the Late sample years at the West sample sites (Late West sample) and the Early sample years at the East sample sites (Early East sample)
    - Early West/Late East- data was divided into the Early sample years at the West sample sites (Early West sample) and the Late sample years at the East sample sites (Late East sample)
-

Transferability was evaluated by calibrating the model on the first sample and then applying it to the second sample in the pair. Then the model was calibrated on the second sample and validated on the first sample.

Paper presented Western Snow Conference 2017
[1] A.C. Lute, US Forest Service, Rocky Mountain Research Station, 322 E Front Street, Boise, ID, 82702, aby.lute@gmail.com
[2] Charles Luce, US Forest Service, Rocky Mountain Research Station, 322 E Front Street, Boise, ID, 82702, cluce@fs.fed.us

Calibration and validation performance was assessed at each step using the Nash-Sutcliffe Efficiency (NSE) and the Akaike Information Criterion (AIC; *Burnham and Anderson,* 2004).

The interaction of model transferability and model complexity was evaluated by varying the level of model complexity in the spatial analog model using the nearest neighbor (nn) setting in the locfit function. Specifically, nn was varied from 0.1 (most complex) to 2.0 (most simple) in increments of 0.1, corresponding to a range of roughly 76 to 6 degrees of freedom. Each of the seven, non-random cross-validation tests was repeated for each level of model complexity.

The results of the non-random cross-validation tests showed that in calibration, the most complex model performed best for all tests (Figure 1, orange bars). However, the AIC, which includes a parameter penalty, indicated that a less complex model was better for all cross-validations (green bars). In validation, the level of model complexity that did best was dependent on the type of cross-validation (purple bars). For the temporal cross-validations, the most complex model still had the highest NSE in validation. However, for the spatial and spatiotemporal cross-validations, less complex models had higher NSE values in validation.

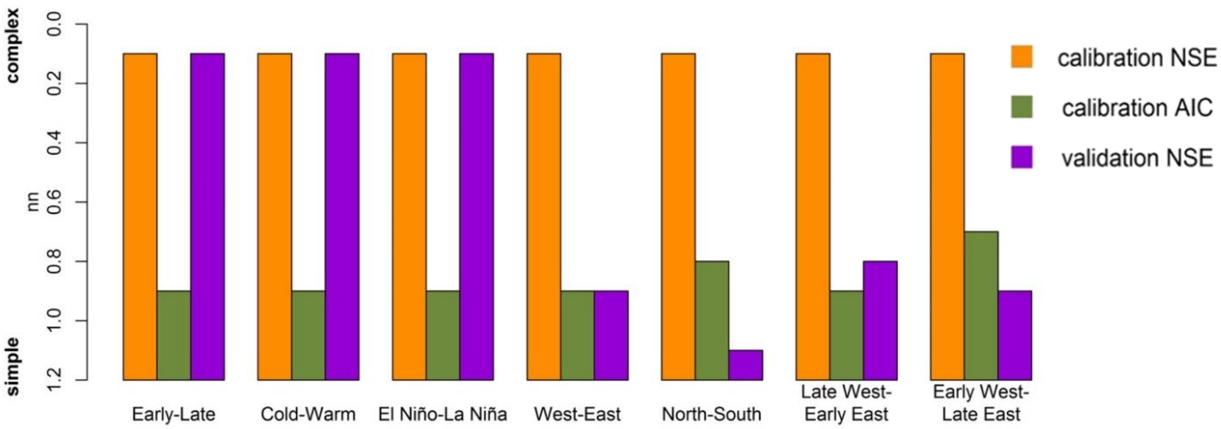

Figure 1. Complexity level of the model with the highest calibration NSE (orange), best calibration AIC (green), and highest validation NSE (purple) for calibration of the model on the Early, Cold, El Niño, West, North, Late West, and Early West samples and validation of the respective models on the Late, Warm, La Niña, East, South, Early East and Late East samples, respectively.

The fact that the most complex model did the best in both the temporal calibrations and validations despite obvious idiosyncrasies in the model (not shown), suggests that the site-specific characteristics present in the calibration datasets were also present in the validation datasets. This is further evidenced by the strong correlations between the calibration residuals across temporal samples. For example, the correlation between the warm model calibration residuals and the cold model calibration residuals was 0.76 (Figure 2). The strong dependence of errors across calibration and validation samples indicates that the samples are not truly independent samples. This can be attributed to the fact that site-specific characteristics persist over time, such as aspect, canopy cover, and windiness. The fact that the temporal samples were pseudoreplications [*Hurlbert,* 1984] resulted in the selection of overly complex models (Figure 1).

For the cross-validations that used truly independent, non-pseudoreplicated samples, we further evaluated their validation performance (Figure 3). The most complex model (orange bars) consistently had a lower NSE value in validation than the model with the best calibration AIC (purple bars) and the model with the best validation NSE (green bars). For some cross-validations the difference was substantial, such as for the Late West-Early East test. Furthermore, the model selected by the AIC in the calibration stage did as well or nearly as well as the best model in validation, suggesting that the AIC may be a useful model selection tool.

In summary, we found that temporal samples were pseudoreplicated, resulting in the selection of overly complex models. Several hazards are associated with overfit models, including the possibility of spurious relationships, greater data requirements in calibration and validation, difficulties related to parameter identifiability and validation, and poor transferability to truly different conditions. Furthermore, when applied to truly different

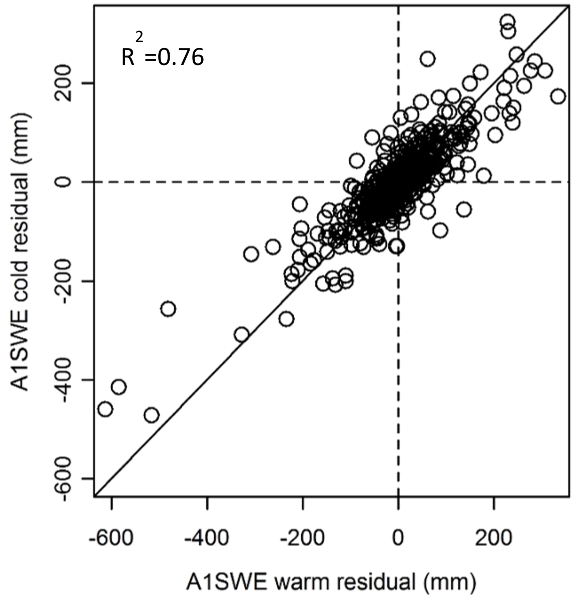

Figure 2. Residuals from the warm model calibration (x-axis) plotted against residuals from the cold model calibration (y-axis). Solid line marks the 1:1 line where the residuals are equal, dashed lines indicate where the respective residuals are 0 (no error).

conditions, less complex models transferred better than more complex models, providing an empirical confirmation of the parsimony principle. This suggests that improvements in hydrologic modeling may come from focusing on improved process understanding across scales. Finally, we saw that the AIC successfully selects a highly transferable model in the calibration stage. This result advocates for the use of the AIC as a model selection tool to avoid selecting overly complex models. (KEYWORDS: spatial analog models, pseudoreplication, SWE, model transferability)

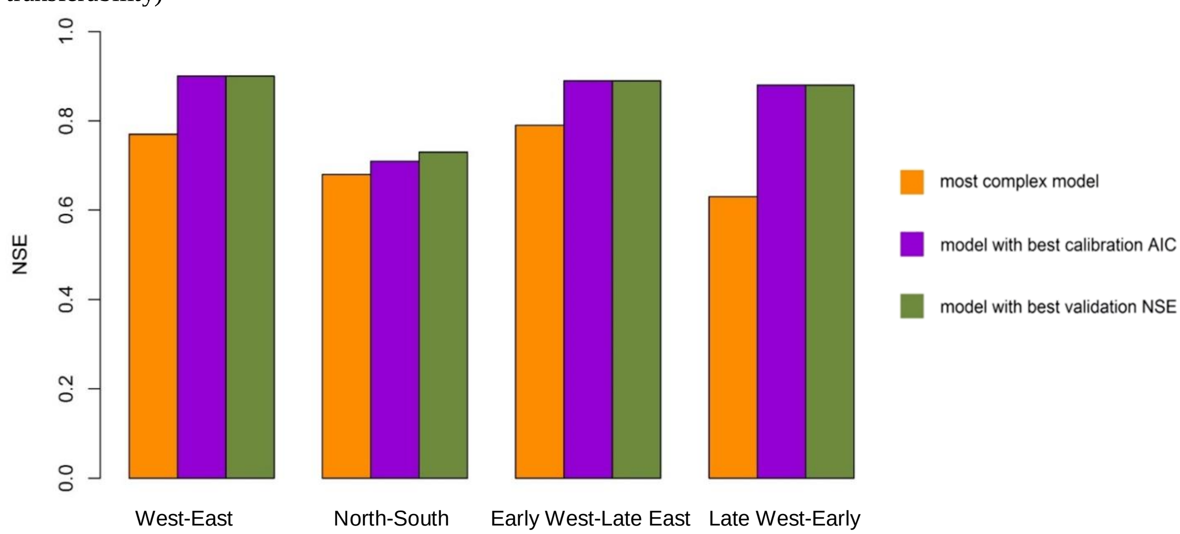

Figure 3. Validation NSE value of the most complex model (orange bars), the model with the best AIC in calibration (purple bars), and the model with the best validation NSE (green bars) across four cross-validation tests including the validation of the West model on the East sample, validation of the North model on the South sites, validation of the Early West model on the Late East sites, and validation of the Late West model on the Early East sites.

REFERENCES

Burnham, K.P. and D.R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in model selection, Sociological methods & research, 33(2), 261-304, doi: 10.1177/0049124104268644.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments, Ecological Monographs, 54(2), 187-211, doi: 10.2307/1942661.

Loader, C.R. 1999. Local Regression and Likelihood, Springer, N.Y.

Luce, C.H., V. Lopez-Burgos, and Z. Holden. 2014. Sensitivity of snowpack storage to precipitation and temperature using spatial and temporal analog models, Water Resources Research, 50, 9447-9462, doi: 10.1002/2013WR014844.

Oyler, J.W., A. Ballantyne, K. Jencso, M. Sweet, and S.W. Running. 2014. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature, International Journal of Climatology, 35, 2258-2279, doi:10.1002/joc.4127.

Wolter, K., and M.S. Timlin. 1993. Monitoring ENSO in COADS with a seasonally adjusted principal component index, in Proceedings of the 17th Climate Diagnostics Workshop, pp. 52–57, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Surv., CIMMS and the Sch. of Meteorol., Univ. of Oklahoma, Norman, Okla.