Abstract

Estimating snow water equivalent (SWE), the amount of water contained in snowpack, is essential for many applications. The infrastructure for automatic single-point SWE measurements is well-established. However, because SWE varies significantly over space, the estimation of SWE at the catchement-scale based on a single-point measurement is error-prone. It is therefore desirable to accurately model the spatial distribution of snowpack based on single-point measurements and other generally available data. Snowpack distribution is known to be highly nonlinear and to be determined by multiple simultaneous influences. We therefore propose genetic programming (GP), a nonlinear, white box, inductive machine learning algorithm, to model snowpack.

We conducted two experiment sets. First, we used a history of manual snow course data at eight sites across California to develop models that use a single-point SWE measurement and air temperature data to predict SWE over an area. While SWE is a product of snow depth (HS) and density, it has been shown that HS varies much more than density does, and that snow depth is therefore the determining factor of SWE variability. The essential requirement for estimating SWE over an area is an understanding of *snow depth* distribution. Our second experiment set therefore used HS measurements collected by wireless sensor network deployments in Norway and in California to develop models that use single-point HS measurements to predict mean HS over an area. In both experiment sets we introduced a time of year metric so that models can distinguish phases of the snow season.

We compared the performance of GP against linear regression (LR) and a basic method (BM) than naively assumes non-variable snowpack to estimate SWE. When estimating SWE in the first experiment set, the performance of the three methods was similar. However, when estimating HS in the second experiment set, GP outperformed LR. These results confirm the nonlinearity of snowpack distribution and suggest that GP can be used to improve snowpack estimates. Because it generates white box models, GP may contribute to an understanding of snowpack dynamics. Examination by domain experts of GP models may yield insight into the relationships governing snowpack distribution.