Multiscale Bayesian Fusion Approach using Geophysical and Remote Sensing Data for Characterizing Arctic Tundra Hydrogeochemical Properties

H.M. Wainwright, S.S. Hubbard, B. Dafflon, C. Ulrich & Y. Wu

Lawrence Berkeley National Laboratory, Berkeley, CA, USA

C. Gangodagamage, J. Rowland & C. Wilson

Los Alamos National Laboratory, Los Alamos, NM, USA

C. Tweedie

University of Texas at El Paso, El Paso, TX, USA

S.D Wullschleger

Oak Ridge National Laboratory, Oak Ridge, TN, USA

INTRODUCTION

Recent findings suggest that warming climate has a significant impact on the Arctic ecosystems, which could in turn cause feedback to the climate system. A new Department of Energy project, called the Next-Generation Ecosystem Experiments (NGEE Arctic; http://ngee.ornl.gov/), has been initiated to address how permafrost thaw and degradation– and the associated changes in landscape evolution, hydrology, soil biogeochemistry and plant community dynamics– affect this feedback.

Developing a predictive understanding of this feedback system requires characterization of various subsurface properties, such as active layer thickness (ALT), soil moisture, temperature and geochemical parameters. Such characterization is challenging due to the need to sample over large spatial areas in high resolution, and the spatial heterogeneity influenced by microtopography, soil texture and vegetation [Zona et al., 2011].

This work presents a multiscale data fusion method to estimate subsurface hydrogeochemical properties and state distributions, using datasets that sample different properties over various measurement support scales, such as point measurements, surface geophysical data and remote sensing data. The proposed method, based on a hierarchical Bayesian model, allows us to integrate multiscale, multi-type datasets consistently to quantify uncertainty associated with the estimates. We demonstrate our approach using co-located datasets collected at the Barrow Environmental Observatory, Barrow, Alaska.

METHODOLOGY

Data Collection Strategy

To improve predictive capabilities and process understanding, NGEE Arctic includes intensive subsurface-surface characterization efforts that involve various types of point, geophysical and remote-sensing data. At the Barrow site our datasets include (a) airborne high-resolution LiDAR data; (b) surface-based ground penetrating radar (GPR), electrical resistivity tomography, spectral induced polarization, and electromagnetic data; and (c) point-based soil temperature, moisture, geochemistry, texture and ALT measurements along and in the vicinity of several ~500m transects across the polygonal ground.

Each dataset offers both advantages and limitations; together, the datasets offer the possibility of providing high-resolution information over large spatial extents. The point-based measurements are invasive and typically sparse, yet they provide direct information about the subsurface properties and states. They are also used to establish the correlations among subsurface properties, geophysical attributes and surface features for calibrating geophysical and remote sensing data. The high-resolution surface geophysical datasets are non-invasive and spatially extensive; integration of these datasets increases the spatial coverage and potentially reveals the fine-scale property variability and spatial correlation structure. The remote-sensing data can further increase the spatial coverage and improve the estimation over a large area through the subsurface-surface correlations. Integration of these disparate datasets requires an understanding of relationships among different hydrogeochemical-geophysical-physical variables and a framework that can honor all datasets and their interdependencies.

Bayesian hierarchical model

The Bayesian hierarchical model consists of three main statistical models; (1) data model: $p(\text{data}|\text{process},\boldsymbol{\beta})$, (2) process model: $p(\text{process}|\boldsymbol{\alpha})$ and (3) prior model: $p(\boldsymbol{\alpha},\boldsymbol{\beta})$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the model parameters [Wikle et al., 2001]. The process model describes the spatial-temporal patterns of the subsurface properties, conditioned on $\boldsymbol{\alpha}$. Although it is usually not feasible to develop a complete physical model to create such patterns (e.g., depositional processes, polygonal-ground evolution), we can often approximate the patterns using simple functions such as polynomial and cosine functions. The data model connects the patterns created by the process model and the actual data for given $\boldsymbol{\beta}$ and measurement errors. The data can be a direct measurement or a function of the subsurface properties; for example, spatial averaging for low-resolution datasets.

The overall model – a series of conditional models – is flexible and expandable to include complex physical processes or observations. For example, the process model can include multiscale physical processes by having additional hierarchical structures [Ferreira et al., 2007]. Once all the conditional models are developed, we can estimate the joint posterior distribution of the parameters and process $p(\text{process},\boldsymbol{\alpha},\boldsymbol{\beta}|\text{data})$, using Bayesian estimation algorithms such as the Markov-chain Monte-Carlo (MCMC) method.

DEMONSTRATION

ALT Estimation

Here we show one example of our data integration; ALT estimation based on the point probe, surface GPR and LiDAR elevation data. For the GPR data, the arrival time of the signal

reflected at the bottom of active layer was converted to ALT using the velocity estimated by the common-midpoint surveys.

First, we developed a process model for ALT (m):

$$\text{ALT} = f(dh, w) + \varepsilon_a, \quad (1)$$

where f is a function, dh is the microtopography (i.e., deviation of elevation from the trend), w is the indicator for the presence of surface water, and ε_a is the residual random component representing additional spatial variability. Although the LiDAR data provides the elevation, we included it as an explanatory variable, since the topography is a dominant control over the subsurface properties [Zona et al., 2011], and also the elevation is quite accurate and spatially exhaustive in the domain of interest. The same argument can be applied for the surface water.

We define the data model for each point data z_p as:

$$z_p = \text{ALT} + \varepsilon_p, \quad (2)$$

where ε_p is the measurement error. The data model for each GPR data point z_g is:

$$z_g = g(\text{ALT}) + \varepsilon_g, \quad (3)$$

where ε_g is the measurement error, and g is a function connecting the actual ALT to the one from the surface GPR.

Based on the exploratory data analysis, we assumed a 2nd-order polynomial for $f(dh,w)$: $f(dh,w)=[dh^2\ dh\ 1\ w]\bullet[\alpha_2\ \alpha_1\ \alpha_0\ \alpha_w]^T$, and a linear model for g(ALT): g(ALT)=0.94ALT+0.04. We also assumed that ε_a is a multivariate Gaussian distribution with exponentially decaying spatial correlation (correlation length 11.0m, standard deviation (STD) 0.19m and nugget fraction 0.56). Table 1 shows the other fixed parameters.

Using MCMC-Gibbs sampling, we computed the posterior distribution of $\boldsymbol{\alpha}=\{\alpha_2\ \alpha_1\ \alpha_0\ \alpha_w\}$ and ALT at the prediction locations; $p(\{\text{ALT}\},\boldsymbol{\alpha}|\{z_p\},\{z_g\})$ (the curly bracket denotes ALT and data at multiple locations). Although the current model is fairly simple, we can extend it by adding more explanatory variables (e.g., surface vegetation), considering spatially variable $\boldsymbol{\alpha}$, and jointly estimating the currently fixed parameters.

Table 1. Assumed distributions and parameters.

	Value	Unit
Distribution of $\{\varepsilon_p, \varepsilon_g\}$	Independent normal	
STD of $\{\varepsilon_p, \varepsilon_g\}$	{0.01, 0.07}	{m, m}
Prior distribution of $\boldsymbol{\alpha}$	Independent normal	
Prior mean of $\boldsymbol{\alpha}$	{0.88,-0.25,0.26,0.10}	{m^{-1}, , m, m}
Prior STD of $\boldsymbol{\alpha}$	{0.10,0.10,0.10,0.10}	{m^{-1}, , m, m}

Results

Figure 1(a) shows the ALT data from the point probe and the surface GPR. Along this 100m transect, we had point data every 3m and GPR data every 0.2 to 0.4m. The surface GPR not only provides fairly accurate ALT but also captures the fine-scale heterogeneity missed by the point data. The multiple peaks of the ALT profile correspond to trough locations on the surface in the moderately high-centered polygons.

In Figure 1(b), the estimated ALT and uncertainty bounds from our estimation method are compared with the point data. Although the estimate is a simple function of microtopography and surface ponding, it shows good agreement to the point data. Figure 1(c) shows the estimated ALT in the 2-D domain including the point and GPR data, based on the estimated $\boldsymbol{\alpha}$ and LiDAR elevation data. It suggests that, by collecting sparse point and geophysical ALT data, this method can estimate ALT over a large area based on topography. We can see the polygonal-ground feature by large ALT at the troughs.

SUMMARY AND FUTURE WORK

In this paper, we described a multiscale hierarchical Bayesian method for integrating multiscale, multi-type datasets. As the first example, we illustrated how the method can be used with the GPR and LiDAR data to estimate ALT over a large area. Although the current example model is still simple, the results suggested that effective combination of point, geophysical and remote-sensing data can provide high-resolution information about important subsurface properties needed to parameterize both 2D/3D landscape and reactive transport models. As future work, we will improve the ALT estimation and extend the methodology to estimate other subsurface properties including ground ice using various combinations of datasets.

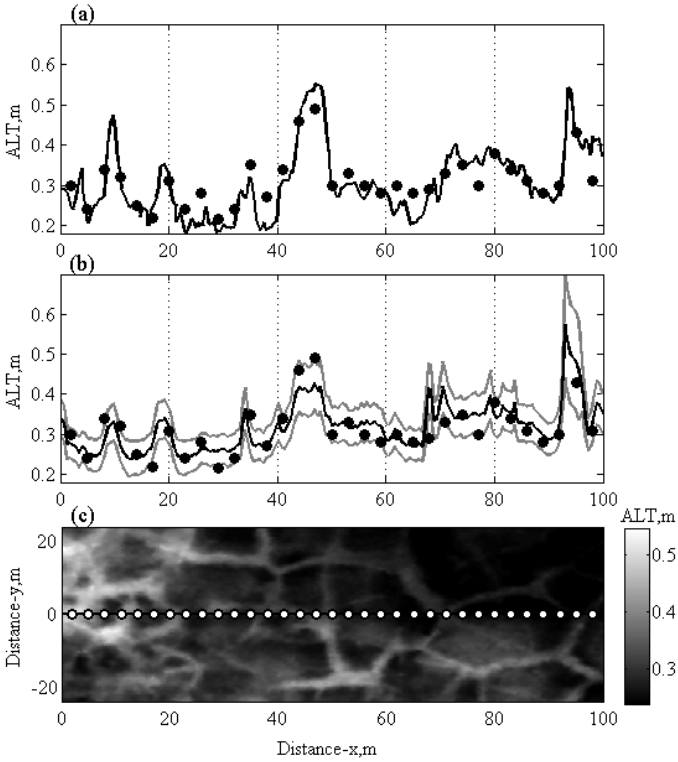

Figure 1. (a) ALT from the point probe (dots) and from the surface GPR (bold line), (b) Estimated ALT (black line), 99% confidence interval (gray lines) and point data (dots), and (c) Estimated ALT in the 2-D domain. The white dots in (b) are the point data locations, and the GPR line.

References

Ferreira, M.A.R. & Lee., H.K.H. 2007. Multiscale modeling: a Bayesian perspective, Springer.

Wikle, C.K., Milliff, R.F., Nychka, D. & Berliner, L.M. 2001. Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds, Journal of the American Statistical Association, 96(454):382-397.

Zona, D., Lipson, D.A., Zulueta, R.C., Oberbauer, S.F. & Oechel, W.C. 2011. Microtopographic controls on ecosystem functioning in the Arctic Coastal Plain, J. Geophys. Res., 116

Acknowledgement

Research sponsored by the U.S. Department of Energy, Office of Science, Biological and Environmental Research.