Geophysical Research Abstracts
Vol. 15, EGU2013-66-1, 2013
EGU General Assembly 2013

Information-Based Analysis of Bayes Filtering

Grey Nearing (1), Hoshin Gupta (1), Wade Crow (2), and Wei Gong (3)

(1) University of Arizona, Hydrology and Water Resources, Tucson, AZ, United States, (2) USDA-ARS, Hydrology and Remote Sensing Laboratory. Beltsville, MD, United States, (3) Beijing Normal University, College of Global Change and Earth System Science. Beijing, China

Data assimilation is defined as the Bayesian conditioning of uncertain model simulations on observations for the purpose of reducing uncertainty about model states. These techniques are widely used to improve model-based predictions of environmental systems. Practical data assimilation filters, such as the ensemble Kalman filter (Evensen, 2003), apply Bayes law independently at each simulation timestep and include simplifying assumptions about the prior, posterior and likelihood probability distributions. We propose a method to quantify the efficiency of these assumptions, and thereby measure the unused information in assimilated observations.

Uncertainty is quantified as the Shannon entropy of a discrete probability distribution p, $H(p)$ [nats], and the maximum reduction in entropy (uncertainty) of model states (X) possible due to Bayesian conditioning on observations (Y) is the mutual information between distributions over states and observations:

$I(p_X; p_Y) = H(p_X) - H(p_{X|Y})$,

p_X is the prior distribution over model states, p_Y is the observation uncertainty distribution and $p_{X|Y}$ is the hypothetical posterior conditional distribution of model states, which is the (usually unobtainable) objective of data assimilation. The efficiency of a filter is defined as the difference between entropies of the prior (p_X) and filter approximate posterior ($f_{X|Y}$) normalized by the mutual information between states and observations:

$E = \frac{H(p_X) - H(f_{X|Y})}{I(p_X; p_Y)}$.

Entropy in the filter posterior can be due to (i) non-injectivity of the mapping between states and observations, (ii) noise in the observations (a non-Dirac p_Y), and (iii) filter approximations of Bayes law. Metrics to assess the relative contribution to $H(f_X)$ of these three causes are:

$E_h = 1 - I(p_X; \delta_Y)$,

$E_Y = I(p_X; \delta_Y) - I(p_X; p_Y)$,

$E_f = I(p_X; p_Y) - H(p_X) + H(f_{X|Y})$.

where δ_Y is the Dirac delta function with mean taken to be the value of an observation. E_h is the fraction of posterior entropy in model states due to injectivity of the observation mapping, E_Y is the fraction due to noise in the observations, and E_f is the fraction due to filter inefficiency which is also the amount of wasted information. These metrics are estimated using an observing system simulation experiment (Arnold and Dey 1986).

These metrics were used to illustrate the propagation of information through a dynamic system model. The procedure was demonstrated on (1) the problem of estimating profile soil moisture from observations at the surface (top 5 cm) and (2) the problem of estimating agricultural productivity from observations of leaf area index; in both cases observations were assimilated by the ensemble Kalman filter. In the soil moisture experiment, information was lost due to inefficiency of the filter, and in the agronomy experiment almost no information was lost.

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics,53,5*

Arnold, C.P., Dey, C.H. (1986). Observing-systems simulation experiments - past, present, and future. *Bulletin of the American Meteorological Society,67*