Geophysical Research Abstracts,
Vol. 11, EGU2009-5011-21, 2009
EGU General Assembly 2009

Statistical modelling in data assimilation

N.D. Smith (1), C.N. Mitchell (1), and C.J. Budd (2)

(1) Department of Electronic and Electrical Engineering, University of Bath, Bath, BA2 7AY, United Kingdom, (2) Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, United Kingdom

A data assimilation technique typically optimises a forward model to best replicate a sequence of observations. In geophysical applications, the forward model and underlying physical process yielding the observations are often driven. For example, Earth's ionosphere is influenced by solar radiation and Earth's magnetosphere. Both the forward model and underlying process may be linear or nonlinear. In optimising the forward model, an objective function is required. From a certain perspective, often the objective function makes implicit statistical assumptions, for example conditional independences between observations and the existence of Gaussian distributions. Even when these assumptions are incorrect, techniques based on them often prove remarkably robust. However with the anticipated future increase in availability of data, for example from satellites, it may be possible to begin to model more accurately the statistical dependencies and variation.

This presentation discusses some ideas from statistical modelling and pattern classification in the context of data assimilation for driven systems, where all variables are discretised. As known, data assimilation schemes may often be cast into a Maximum A-Posteriori (MAP) estimation framework. In this framework, the true statistical model associated with the underlying physical process minimises the 'overall risk' under a 'classification' or '0/1' loss function. Unfortunately the true statistical model is rarely known. Instead, in this framework, the objective function and forward model together imply a statistical model which is only an estimate, and often a poor one, of the true statistical model. Within the context of model selection from the field of statistical modelling, quantities such as overall risk may be used to evaluate and compare alternative statistical models implied by different objective functions and/or forward models. Mutual information may also be used. However these evaluation techniques require some knowledge of the truth (e.g. the true values of driver variables) and are in the context of classification. Also of interest is the effect on overall risk of a failure to adequately model variables or conditional dependencies.

However improving the statistical modelling is a challenging task. In geophysical applications, the underlying physical process is often nonstationary and it is expected that the statistical distributions will generally be difficult to estimate, particularly for infrequent extreme conditions. However an awareness of the above concepts may help us better understand the limitations of present data assimilation schemes.