Statistical analysis for high resolution data assimilation

Laura Quinn*[1], K.V. Mardia[1], S.D. Mobbs[1] and S.B. Vosper [1,2]

[1] University of Leeds
[2] Met Office

1 Introduction

Over the past 15 years, there has been a gradual shift away from a purely deterministic approach to weather forecasting towards a more integrated probabilistic approach. In particular, methods such as meteorological data assimilation, ensemble forecasting and post-processing of numerical weather prediction model output have been influenced by theories from statistics and control theory.

Like any mathematical modelling of a complex dynamic system, the traditional deterministic approach to weather forecasting seeks to solve nonlinear partial differential equations. This is not possible by precise analytical methods, but is achieved by numerical approximation, integrating forward in time simplified versions of these differential equations. Limitations in this arise due to the deterministic chaotic nature which such geophysical processes exhibit (i.e. a system that is ordered and predictable but in a highly mathematically complex fashion), together with limitations in computing power for such high order systems.

Data assimilation improves on numerical weather forecasting accuracy through the combination of deterministic modelling and observational data. In short, data assimilation inserts spatially weighted observational data into a deterministic model to constrain the model to more accurately represent the “true” atmospheric state. Usually data assimilation proceeds sequentially with time, while the model organises and propagates forward information from previous observations. The information from new observations is then used to modify the model state, to be as consistent as possible with the observations and the previous information. Lorenc (1988, 1995) discusses the Bayesian Analysis behind this methodology.

2 Motivation

Data assimilation may be considered an umbrella term for a variety of different techniques applied to achieve the objectives outlined above. A variety of such techniques are currently being utilised in most operational global and regional weather forecasting models worldwide with demonstrable success; at present these are being utilised up to mesoscale resolution (model mesh spacings in the order of 1km), currently no robust operational assimilation schemes are routinely applied to finer scale models.

At such high resolution, a numerical weather prediction model must ideally take into account a number of other factors, namely the effect that topography, that is the surface features of a region, has on atmospheric conditions. A number of atmospheric phenomena are known to occur as a result of atmospheric flows passing over hilly or mountainous terrain. In particular, atmospheric flows combined with certain regimes of temperature profile within the atmosphere can lead to quite disruptive turbulent events which are extremely difficult to predict via purely deterministic means.

With such high resolution models (model mesh spacings $\sim$100s metres), the assimilation process aims to constrain the model on scales where normal balances do not apply, via the inclusion of a number of observationally derived nudging terms in the model equations.

Vosper (2003) has developed a 3-dimensional numerical weather prediction model for application at such high resolution, and our aim is to develop a data assimilation scheme for this model in order to more successfully predict the occurrence of such weather phenomena.

A simple representation of the observational data is necessary for the faciliation of this assimilation step.

3 Experimental data

During a year long field campaign, an array of 20 automatic weather stations were deployed on East Falkland island in the South Pacific to record near-surface flow conditions.

Most of these were situated nearby to Mount Pleasant Airfield (MPA), which is well-known for the occurrence of low-level turbulence. MPA lies to the south of two east-west oriented mountain ridges, each approximately 600m in height. During periods of stable northerly flow turbulence thought to be associated with mountain wave activity is experienced around the airfield. These periods of extreme turbulence pose a very real threat to aviation safety.

The remainder of the stations were arranged in a rough north-south transect through MPA. The stations recorded 30s averaged data almost continuously for the duration of the campaign.

The diagram in figure 1 shows the final positions of the weather stations in relation to the runway at Mount Pleasant Airfield.

Prior to the field experiment it was not known what the optimum positions of sites would be in order to maximise the amount of information captured. The positions of the stations are not topographically equivalent with some being placed on the top of ridges and others being situated at the bottom of valleys between ridges. Several stations are situated in extremely close proximity to the runway itself (shown).

The 20 automatic weather stations recorded data as 30 second averages over a period of approximately one year, meaning a potential maximum of around 1,051,200 records per station over the duration of the experiment. However, due to the nature of the experiment, some of these stations performed better than others and data sets between stations range from sporadic to near uninterrupted.

The high resolution data assimilation methodology being considered envisages that via statistical interpretation of the spatial and temporal data, appropriate weightings may be attributed to observations collected at certain weather stations. In turn these weighted observations are expected to allow development of statistically derived “nudging constraints” to be incorporated into the three-dimensional weather prediction model developed by Vosper (2003), to improve its forecast accuracy.

4 Exploratory data analysis

As a first step in the assimilation problem, it is essential to understand the general underlying temporal and spatial behaviour and distribution of the data. Some exploratory statistical analysis performed on the data set brings out interesting features.

In a similar vein to work carried out by Fiorino and Correia (2002), a principal component analysis methodology has been extensively applied to the data in order to uncover key spatio-temporal patterns. The insights offered via this analysis have formed the basis for the derivation of the added model constraints and thus the representation of observational data in the NWP.

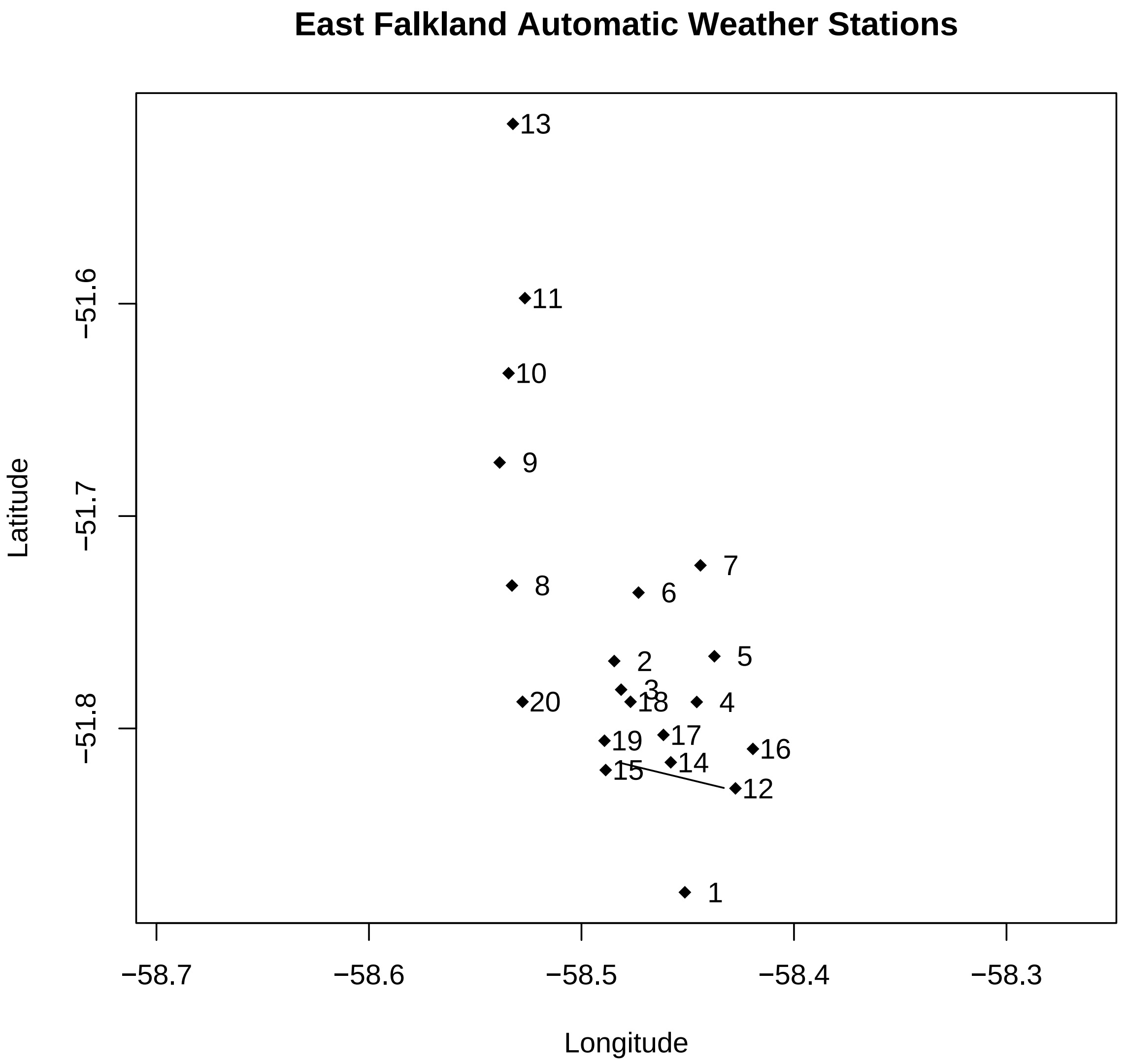

Figure 1: The final positions of the weather stations in relation to the runway at Mount Pleasant Airfield.

A number of case studies from the Falkland's campaign will be presented in light of their role in developing these constraints. In particular, the appearance of "fingerprint" patterns of PCA loadings representative of severe turbulence events as opposed to steady flow conditions, together with their temporal dependency will be presented. These distinctive fingerprints can be used diagnostically to pinpoint turbulent events.

The cross comparison of observational PCA with model output PCA is also being undertaken as a means of detecting model weakness. Some results from this will also be included.

There are particular issues related to the high resolution data assimilation process problem given a spatially sparse data set, where interpolation and smoothing considerations are key. From the work of Mardia *et al.* (1998), Sahu & Mardia (2005) and Bengtsson *et al.* (2003), the application of Kriging and Kalman filtering techniques are under investigation. The Kriged-Kalman filter approach of Mardia *et al.* (1998) and Sahu & Mardia (2004) as well as the non-linear ensemble approach of Bengtsson *et al.* (2003) have great potential for this problem.

References

Bengtsson, T., Snyder, C., Nychka, D. (2003). Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research-Atmospheres*, **108 (D24)**, art no. 8775.

Fiorino, S.T. & Correia Jr., J. (2002). Analysis of a mesoscale gravity wave event using empirical orthogonal functions. *Earth Interactions*, **6**, paper 1.

Lorenc, A. & Hammon, O. (1988). Objective quality control of observations using Bayesian methods. Theory and a practical implementation. *Quart. J. Royal Meteorological Soc.*, **114**, 515-543.

Mardia, K.V., Goodall, C., Redfern, E.J., & Alonso F.J. (1998). The Kriged Kalman Filter (with discussion). Test, **7**, 217-252.

Mobbs, S.D., Vosper, S.B., Sheridan, P.F, Cardoso, R., Burton, R.R., Arnold, S.J., Hill, M.K., Horlacher, V. & Gadian, A.M. (2004). Observations of downslope winds and rotors in the Falkland Islands. *Quart. J. Royal Meteorological Soc.*, **131**, 329-351.

Sahu, S.K. & Mardia, K.V. (2004). A Bayesian Kriged-Kalman model for short-term forecasting of air pollution levels. *J Roy. Stat. Soc. Series C.*, **54**, pp223-244.

Vosper, S.B. (2003). Development and testing of high resolution mountain-wave forecasting system. *Meteorol. Appl.* **10**, 1-100.