Model Error and Parameter Estimation in a Simplified Mesoscale Prediction Framework, Part I:

Model Description and Sources of Uncertainty

Guillaume Vernieres, Josh Hacker, Montse Fuentes
Topics

- Mesoscale forecasting - some background.
- Data assimilation at mesoscales.
- Types of error in mesoscale models.
- A column model to emulate a full 3D mesoscale model, and experience with it.
- Some naive parameter estimation experiments.
Mesoscales

Horizontal wind spectra in the frequency domain.

Vinnichenko 1970

Vander Hoven 1957
Mesoscale prediction is fundamentally an initial condition problem. Nowcasting is typically done by extrapolating current conditions because dynamical models have less skill at these time scales. A place for better data assimilation and accounting for model error?
Fig. 2 compares the errors of the surface temperature and wind forecasts of the MM5 and WRF models, and the error evolution with forecast time.

Fig. 3. Domain average of BIAS errors of surface forecast, partially due to its more accurate Noah land in the forecasts of the surface variables. MM5 performs better on both surface wind and temperature, partially due to its modified MRH scheme that significantly improves daytime estimation (Liu et al., 2004). The forecast errors heating/cooling and PBL mixing. It can be observed that significant wind speed errors exist in the initial conditions of both models, as a result of interpolation. The error growth of temperature and wind generally grows with increasing forecast time. The observed in the error growth of temperature and wind that significantly warm bias in MM5 and a cold bias of a similar magnitude in WRF. The WRF model appears to handle these strong influences by the diurnal evolution of surface conditions (0-12 h forecasts) of the two models, although discrepancies can be seen between the initial conditions and the model predictions better in the upper levels. The WRF is better at predicting winds in the layer between 700 and 200 hPa, with a RMS error of the vector T_BIAS (C)

SPD_BIAS (m/s)
Ensembles are extremely underdispersive and show little intrinsic error growth near the surface in the short range, leading to experimentation with “multi-model” ensembles (FULL). From Hou et al. 2001 MWR.
Surface observations are relatively dense and inexpensive to gather.

Typically under-utilized in operational data assimilation. Model error? Constraints in the assimilation systems?

Potential to tell us something about the state of the underlying PBL.

Potential to tell us something about the model, including values of parameters.
Column Model Environment

- A 1-D PBL modeling framework: various land-surface and PBL parameterizations, forced. Original model development by Mariusz Pagowski, NOAA/ESRL.

- Internal dynamics for ageostrophic wind, diffusion equation, etc.

- Geostrophic and radiative forcing from a mesoscale model (e.g. RUC or WRF) or observations.

Cheap! Thousands of realizations possible with a quick turn-around
Prognostic in $U$, $V$, $\theta$, and $Q$ with parameterization providing closure. Parameterization is the same as in the Weather Research and Forecast (WRF) model.
Model Formulation

\[
\frac{\partial U}{\partial t} = f_c (V - V_g) - U (U, V, \theta, Q, P)
\]

\[
\frac{\partial V}{\partial t} = -f_c (U - U_g) - V (U, V, \theta, Q, P)
\]

\[
\frac{\partial \theta}{\partial t} = -T (U, V, \theta, Q, P)
\]

\[
\frac{\partial Q}{\partial t} = -Q (U, V, \theta, Q, P)
\]

Closure terms are functions of the resolved state (forcing and diffusion), and myriad parameters $P$. 
Model Formulation with Advection

\[
\frac{\partial U}{\partial t} = f_c (V - V_g) + V \cdot \nabla U - \frac{\partial u'w'}{\partial z}
\]

\[
\frac{\partial V}{\partial t} = -f_c (U - U_g) + V \cdot \nabla V - \frac{\partial v'w'}{\partial z}
\]

\[
\frac{\partial \theta}{\partial t} = V \cdot \nabla \theta - \frac{\partial w'\theta'}{\partial z}
\]

\[
\frac{\partial Q}{\partial t} = V \cdot \nabla Q - \frac{\partial w'q'}{\partial z}
\]

Advection acts to relax the column state toward an imposed 3D state.
Skill in PBL State Estimates: $T$

Using only screen-height (surface) observations, skillful profiles are estimated at all times of day: (a) 1PM LT, (b) 7PM LT, (c) 1AM LT, and (d) 7AM LT.
Using only screen-height (surface) observations, skillful profiles are estimated at all times of day: (a) 1PM LT, (b) 7PM LT, (c) 1AM LT, and (d) 7AM LT.
Skill in PBL State Estimates: $Q_v$

Using only screen-height (surface) observations, skillful profiles are estimated at all times of day: (a) 1PM LT, (b) 7PM LT, (c) 1AM LT, and (d) 7AM LT.
Data assimilation to estimate a discrete system state $Z$ at time $t$. 

$Z$ is a joint state, with both state variables and parameters. 

$X$ represents state variables. 

$x$ is a set of parameters, which may or may not be physical. 

Then $Z = (X, x)$. 

Given all observations up to the current time, $Y_t$, we want to estimate $p(Z_t|Y_t)$. 

These experiments are to estimate parameters in a land-surface scheme, given screen-height observations and an evolving model.
An exchange coefficient for moisture, $Q_c$, is computed:

$$Q_c = \frac{M \rho_1 \overline{w'q'}}{q_0 - q_1}$$

- $M$ is a moisture availability parameter \{0,1\}.
- $\rho_1$ is density at the first atmospheric model level.
- $q_0$ and $q_1$ are moisture contents at the surface and the first atmospheric level, respectively.
- $\overline{w'q'}$ is the parameterized kinematic moisture flux.

Provides a lower boundary condition (forcing) for the atmospheric model.
Single parameters (moisture availability) can be estimated when the true value is known.
Correlation coefficients of $T_2$ with parameters $M$ and $THC$, for 100 ensemble members integrated for 10 days.

- Parameter distributions are fixed.
- Distributions chosen as $\beta$ with $\sigma = 0.1M$ and $0.01THC$. 
Correlations With Assimilation

- Correlation coefficients of $T_2$ with parameters $M$ and $THC$, for 100 ensemble members integrated for 10 days.
- Parameter distributions are estimated while assimilating.
- Correlations change, transitions more pronounced.
Dependent Parameters

- $M$ and $THC$ are linearly dependent when estimated. Here is at 00 UTC for over 10 days, but this is true at any time.
- Cannot be distinguished, thus could be replaced by a single parameter.
Distribution Improves Assimilation

Compared to single fixed parameter values, distributed parameters result in a better fit to observations. The effect is particularly true during transitions.
Compared to fixed distributed parameter values, estimated parameters result in a better fit to observations.
Differences in error (estimated – fixed distribution) show the profile is generally improved, especially during the growth phase of the PBL.
State augmentation is a useful parameter estimation approach in observation system simulation experiments (OSSEs), but is much more difficult in real-data applications.

Much more work to do:

- How will a free bias parameter behave?
- Can we find distributions that make a better forecast in the face of other, unknown, model errors?
- Can we find appropriate stochastic processes to propagate the parameter distributions in time?