A Bayesian approach to non-Gaussian model error modeling

Michael Tsyrulnikov and Dmitry Gayfulin

HydroMetCenter of Russia

München, 5 March 2018

Outline

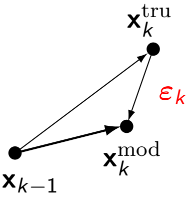
1. Model error: definition and motivation
2. Model error modeling: a Bayesian perspective
3. Numerical experiments with a “true model”
4. Preliminary results and further steps for this ongoing research

Model error: definition and motivation

Model error definition

- Model equation: $\mathbf{x}_k^{\mathrm{mod}} = F(\mathbf{x}_{k-1}^{\mathrm{mod}})$
- "True model": $\mathbf{x}_k^{\mathrm{tru}} = F^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}})$
- One-step error:

start the true model from $\boxed{\mathbf{x}_{k-1}^{\mathrm{tru}} = \mathbf{x}_{k-1}^{\mathrm{mod}}}$ (the "same start condition").



The difference $\boxed{\varepsilon_k = \mathbf{x}_k^{\mathrm{mod}} - \mathbf{x}_k^{\mathrm{tru}} = \mathsf{T}_k^{\mathrm{mod}} - \mathsf{T}_k^{\mathrm{tru}}}$ is the model error.

NB: Whenever the high-resolution true field is compared with the low-resolution model field, the true field is *upscaled*, so that only the *resolved* (grid-scale) field components are actually compared.

Model error: stochastic modeling

Model error is an unknown (and very complex) function of $\mathbf{x}_{k-1}^{\mathrm{tru}}$:

$$\varepsilon_k = F(\mathbf{x}_{k-1}^{\mathrm{tru}}) - F^{\mathrm{tru}}(\mathbf{x}_{k-1}^{\mathrm{tru}}) \qquad = G(\mathbf{x}_{k-1}^{\mathrm{tru}})$$

Because $G(\mathbf{x})$ is unkown, we introduce a stochastic model for it, thus assuming that ε is a random function of the spatial coordinates and time (i.e. a spatio-temporal random field).

Model error definition: motivation

The reason for the above definition of model error is the following.
It implies that, with the stochastic ε, the grid-scale **truth** satisfies the stochastic equation

$$\mathbf{x}_k^{\mathrm{tru}} = F(\mathbf{x}_{k-1}^{\mathrm{tru}}) - \varepsilon_k$$

actually solved in ensemble prediction.

Therefore, ensemble members are solutions to the stochastic equation for the (resolvable) **truth**.

State-dependent model error modeling: a Bayesian perspective

State-dependent model errors

In existing schemes, state dependence is *postulated*, e.g. in SPPT $\varepsilon = \boldsymbol{\mu} \cdot \mathsf{P}$, where P is the physical tendency and μ an independent random field. Other examples are STTP, SKEB, and SCB (Shutts 2015).

A more systematic way to account for various predictors P is to regard them as *data* in addition to what we have already computed at the previous and current time steps, $x_{k-1}^{\rm mod}$ and $x_k^{\rm mod}$.

Then we can write down the "target distribution" from which pseudo-random realizations of the current-time-step truth are to be drawn:

$$\pi(x) = p(x_k^{\rm tru}{=}x \,|\, x_{k-1}^{\rm tru}{=}x_{k-1}^{\rm mod},\, x_k^{\rm mod}, \mathsf{P}_k) \tag{1}$$

Transformation of the target density

The target density can be shown to be the product of three densities:

1. $p(x_k^{\mathrm{tru}} \,|\, x_{k-1}^{\mathrm{tru}})$ is the *transition density*. It ensures that the simulated values x_k^{tru} are within their meaningful range (important for humidity and hydrometeor fields). Can be approximated by $p(x_k^{\mathrm{mod}} \,|\, x_{k-1}^{\mathrm{mod}})$.

2. $p(\mathrm{P}_k \,|\, x_{k-1}^{\mathrm{tru}}{=}x_{k-1}^{\mathrm{mod}}, x_k^{\mathrm{tru}})$ describes the behaviour of the model, specifically, how the *model-generated* P is related to the truth.

3. $p(\varepsilon_k \,|\, \mathrm{P}_k)$ describes the model error itself and will be the focus in what follows.

The three densities constrain the current-time-step truth from the three different perspectives.

The truth x_k^{tru} is generated by sampling from the discretized target density (tested with the transition density function).

Numerical experiments with a “true model”

Approach

1. Take a model in question ("the model").
2. Select a significantly more sophisticated model ("the true model").
3. Start both models from the same point in phase space.
4. Compare the two short-time tendencies, compute their difference (the model error ε), and try to build a stochastic model for the ε field.

The general idea is to look for salient features of the model error field structure, so that the conclusions be not much model specific.

The two models

- "The model" is COSMO-L50 with the horizontal resolution 2.2 km.
- "The true model" is COSMO-L50 with the following differences from "the model":
  1. Horizontal resolution 0.55 km.
  2. Time step 5 s (vs. 20 s in "the model").
  3. Convection parameterization (vs. shallow Tiedtke in "the model") switched off.
  4. 3D turbulence scheme (vs. 1D).
  5. More sophisticated options in the cloud scheme, precipitation scheme, and radiation scheme.

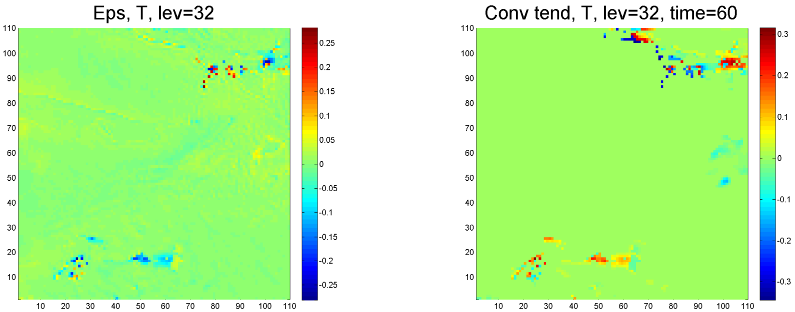
Domain and cases

- The models' domains are centered at 52N 25E.
- The model errors are computed on the 2.2 km 110*110 grid in the horizontal.
- 3 cases were studied: 1 and 29 July and 1 December 2017 (all 12 UTC).

Computing the model error

1. Run "the model" for 1 h lead time (to "spin it up"). The forecast is used as the *starting point* $\mathbf{x}_0^{\mathrm{mod}}$.
2. Downscale $\mathbf{x}_0^{\mathrm{mod}}$ to the fine grid (on which "the true model" operates) (using the COSMO tool INT2LM). This is $\mathbf{x}_0^{\mathrm{tru}}$. This procedure guarantees the "same start" condition.
3. Run "the model" for 3 time steps (60 s) starting from $\mathbf{x}_0^{\mathrm{mod}}$. Calculate the total tendency $\mathrm{T}_3^{\mathrm{mod}}$.
4. Run "the true model" for 12 time steps (60 s in total) starting from $\mathbf{x}_0^{\mathrm{tru}}$. Calculate the total tendency $\mathrm{T}_{12}^{\mathrm{tru}}$ and upscale it to the coarse grid.
5. Compute the model error as $\boxed{\varepsilon = \mathrm{T}_3^{\mathrm{mod}} - \mathrm{T}_{12}^{\mathrm{tru}}}$.

Model errors (left) and convective tendency (right)



The outliers are related to convection.

Outliers normally require a special treatment.

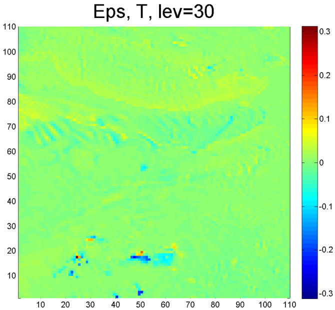
Convection

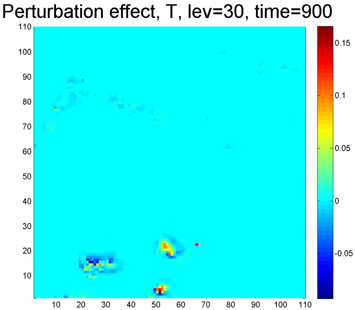
1. Our attempts to relate the convective model errors to CAPE and the vertical lapse rate failed.
2. With this strong instability (0.3 K/min) and complexity of the convection phenomenon, a purely stochastic approach looks unsuitable to model the convective **outcome**. A physical model is needed.
3. Convective model errors are the **outcome** of convection, not the source, which we would like to perturb.

Perturbing a "convective source" by applying tiny model-error perturbations easily give rise in a 15-min forecast to realistically looking convection (see the next slide):

Model error (left) and forecast perturbation (right)

in response to constant in space and time model-error perturbations, $5 \cdot 10^{-5}$ K per time step in T and 10^{-4} m/s in U, V (lead time is 15 min)

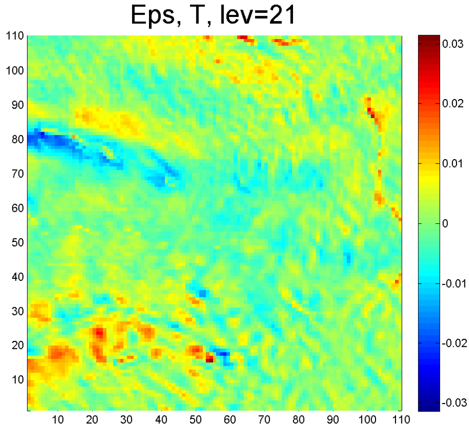




Conclusion on convection

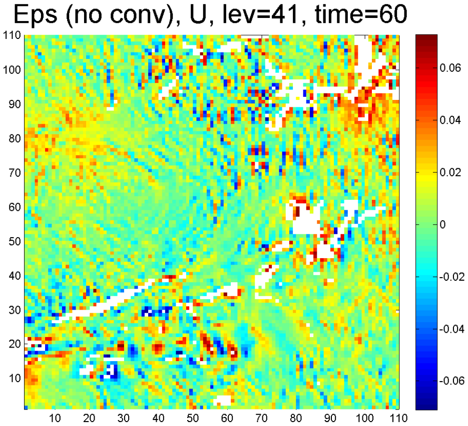
- Model errors we examine in this research are *not* useful in modeling convection (a hidden "convective source" is to be perturbed, not the outcome). So, we do *not* consider convection related model errors in this study.
- Arbitrary and tiny (but greater than a threshold) model error perturbations can trigger convection, albeit not in a perfect way.
- Thus, a stochastic convection scheme (e.g. Plant-Craig) is best to be used to treat convection in generating an ensemble.

Non-convective model error



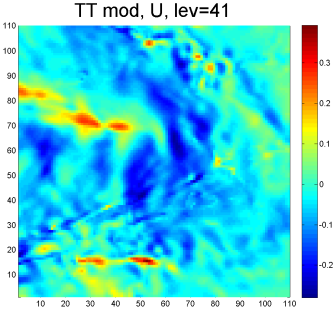
- Looks like a random field.

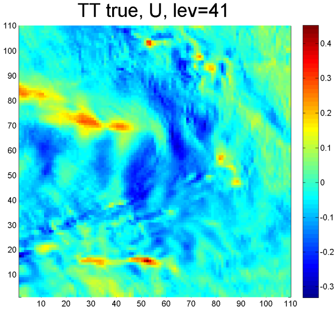
Non-convective model error



- Looks like a random field with complicated structure, with multiple scales, and, likely, with multiple components.

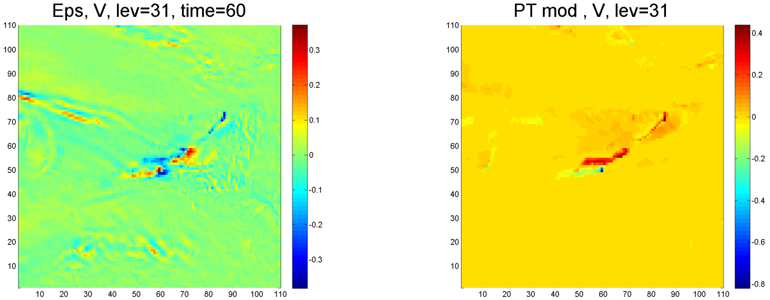
Total tendencies: model (left) and true (right)





- Look like smooth random fields, with the model tendency (left) being a bit smoother.

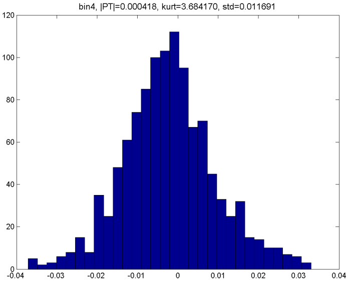
Model error (left) and physical tendency (right)



- Physical tendency is informative but not always.

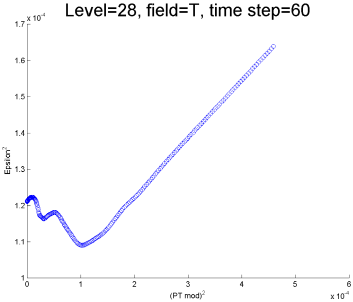
Non-convective model errors: $p(\varepsilon \mid \mathrm{P})$

After filtering out 2 percent largest $|\varepsilon|$ and $|\mathrm{P}|$, we estimated the $\boxed{p(\varepsilon \mid \mathrm{P})}$ density. Values of $|\mathrm{P}|$ were binned (with 10 equipopulated bins). As an example, below is the histogram of ε for the 4-th bin of $|\mathrm{P}|$ (V, level=36):



Kurtosis is normally not too far from 3, hence the non-convective ε can be reasonable modeled as being conditionally Gaussian (given P).

Non-convective model errors: Var $(\varepsilon \mid \mathrm{P})$



•The offset (the value of $\mathrm{E}\,\varepsilon^2$ for $\mathrm{P} = 0$) is the variance of the additive (physical-tendency independent) model-error component.

•The "multiplicative" (physical-tendency dependent) model-error variance is $\mathrm{E}\,\varepsilon^2 - \mathrm{E}\,(\varepsilon^2|\mathrm{P}{=}0)$.

•The most stable feature is that the additive component is always present and significant.

Non-convective model-error model

$$\boxed{\varepsilon(\mathbf{s}) = \alpha(\mathbf{s}) + \mu(\mathbf{s}) \cdot \mathrm{P}(\mathbf{s}) \quad \equiv \textbf{add} + \textbf{mult}}$$

Vertically averaged ratio
of the multiplicative-error st.dev. to the additive-error st.dev.:

	T	U	V
$\frac{\text{s.d. }(\textbf{mult})}{\text{s.d. }(\textbf{add})}$	0.5	0.5	0.8

(The difference between the values for U and for V is, perhaps, due to insufficient statistics.)

- The magnitudes of the additive error components are somewhat larger than the magnitudes of the multiplicative error components.
- The **mult/add** ratio is larger in the boundary layer.

Conclusions

- Model tendency error fields for a convective-scale model were computed (with respect to a more sophisticated model).
- Convection related model errors are found to be better treated with a stochastic convection parameterization.
- Non-convective model errors were studied for T, U, V:
  - They look as multi-scale random fields often with complex spatial structure.
  - They are much "patchier" than the tendency fields.
  - Both additive and multiplicative components are present in the model error. Additive errors have, on average, somewhat greater magnitudes.
  - Both the additive error component and the (SPPT's) multiplier field μ are approximately Gaussian.

Further steps

- Process-level model-error treatment may simplify the model-error modeling task.
- More/better model-error-predictors (uncertainty indicators provided by physical parameterization schemes) would be useful.
- Humidity, vertical velocity, and cloud fields are to be examined.
- The multivariate and the spatio-temporal aspects are to be addressed.

The goal is a practical convective-scale model-error model.

Thank you!