Estimation and modeling of model errors with COSMO

Michael Tsyrulnikov and Dmitry Gayfulin

HydroMetCenter of Russia

Salzburg, 4 Oct 2018

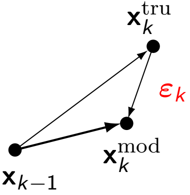
Outline

1. Model error: definition
2. Motivation
3. Approach
4. Numerical experiments:
   - Physical-parameterizations model errors
   - Numerical-approximation model errors

Model error: definition

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 model error:



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

The difference $\boxed{\varepsilon_k = \mathbf{x}_k^{\mathrm{mod}} - \mathbf{x}_k^{\mathrm{tru}} = \mathcal{T}_k^{\mathrm{mod}} - \mathcal{T}_k^{\mathrm{tru}}}$ 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.

Motivation

Tsyrulnikov and Gorin (COSMO Newsletter N 13, 2013) attempted to estimate model error by comparing the finite-time model tendency (started from the analysis) with the finite-time *observed tendency*.

They showed that in order to reliably estimate even the simplest constant-in-space-and-time model error, **every grid point** needs to be observed with currently unreachable accuracy (0.1 K for temperature and 0.02 m/s for winds).

Hence, reliable estimation of realistic model errors by comparing finite-time model tendencies with finite-time observed tendencies is **not possible** with the existing observation networks.

This has motivated the present research.

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.

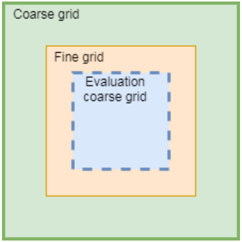
Numerical experiments with a “true model”

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.

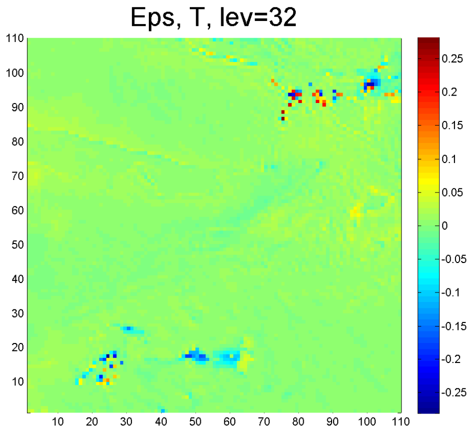


- The coarse-grid-model's domain: 250*250 points (greenish-grayish).
- The fine-grid-model's domain: 600*600 points (pinkish).
- Model errors are computed on the 2.2-km 110*110 sub-grid (bluish).
- 4 cases were studied (all 12 UTC):
  - 1 July and 29 July 2017 ("convective" days)
  - 17 July and 1 December 2017 ("non-convective" days).

Computing the model error

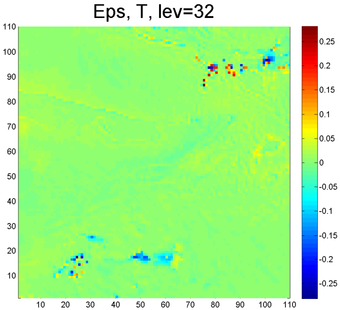
1. Run "the model" for 1 h lead time (to "spin it up"). The resulting 1h forecast is then used as the *starting point* $\mathbf{x}_0^{\mathrm{mod}}$.
2. Downscale $\mathbf{x}_0^{\mathrm{mod}}$ to the fine grid (on which "the true model" operates) (using the COSMO interpolator INT2LM). This is $\mathbf{x}_0^{\mathrm{tru}}$.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

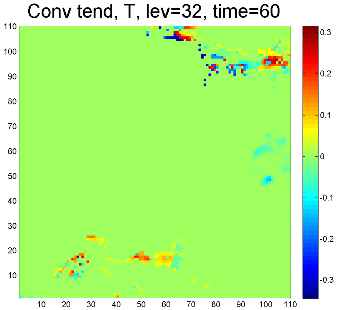
The model error field



Looks as a random field with *outliers*.

Model errors (left) and convective tendency (right)





The outliers are related to convection.

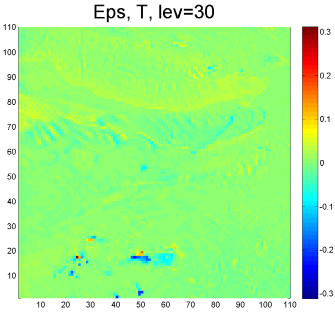
Convection

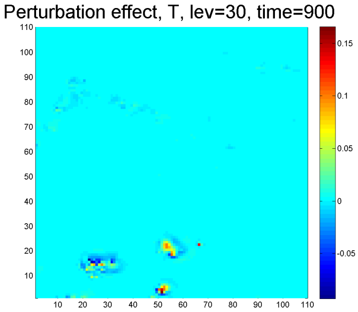
Convection

1. Predictors? Our attempts to relate convective model errors to CAPE and vertical lapse rate failed.
2. Given the *complexity* of the convection phenomenon, a purely stochastic approach looks unsuitable to model convective model errors. A physical model is needed.
3. Convection is *fast* so that the convective model errors we can measure are the **outcome** of convection, not its source. And it is the "convective source" that we would like to isolate, study, and model in this study (and then perturb in an ensemble forecast).

Perturbing the "convective source"

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



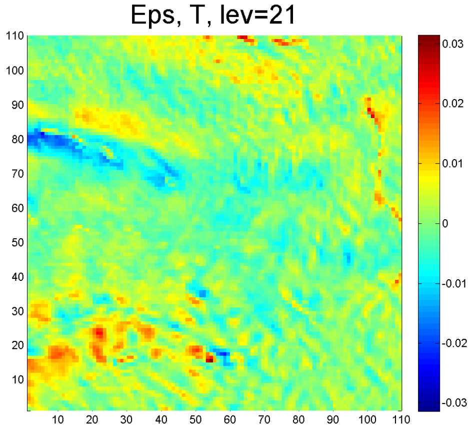


Conclusion on convection

1. Model errors we examine in this research are *not* useful in modeling convection in a stochastic way (a hidden "convective source" is to be studied, not the outcome of convection).
2. Arbitrary and tiny (but greater than a threshold) model error perturbations can trigger convection, albeit not in a perfect way.

So, we do *not* consider convection related model errors in this study. A *stochastic convection scheme* is best to be used to treat convection in generating an ensemble.

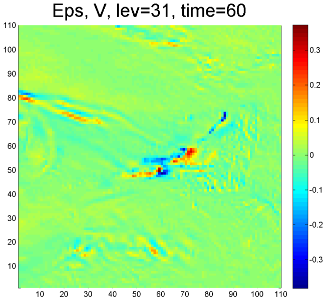
Non-convective model error

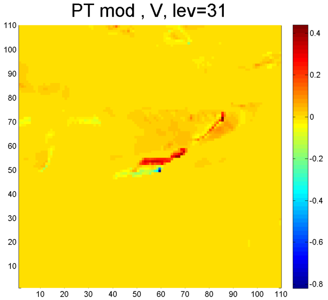


- Looks like a random field.

Physical tendency as a predictor for model error

Model error (left) and physical tendency (right)

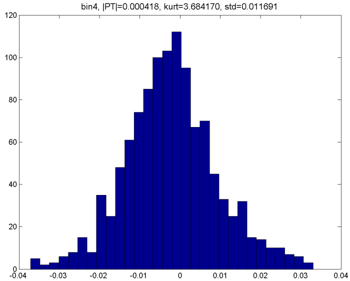




- Physical tendency is informative but not always.

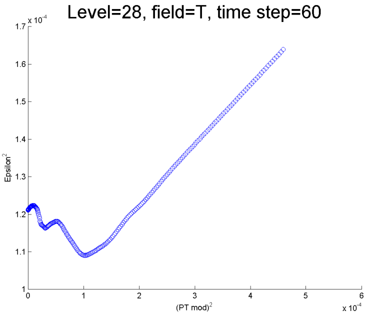
Gaussianity: $p(\varepsilon \,|\, \mathcal{P})$

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

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



y-axis: $\overline{\varepsilon^2}$, x-axis: $\mathcal{P}^2$

•The linear growth of ε^2 with $\mathcal{P}^2$ is the indicator of the "multiplicative" (physical-tendency dependent) model-error component.

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

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

Non-convective **model-error model**

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

where α and μ are Gaussian.

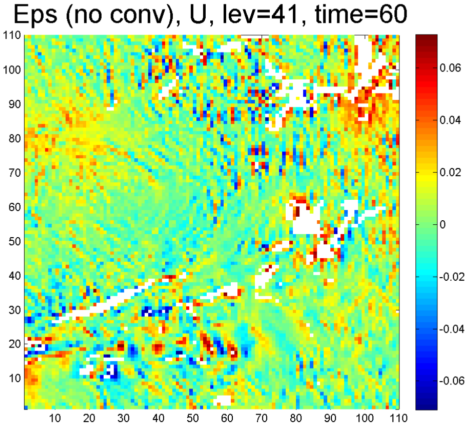
Vertically averaged ratio
of the multiplicative-error st.dev. to the additive-error st.dev.:

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

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

Need for *process-level* error treatment

Non-convective model error



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

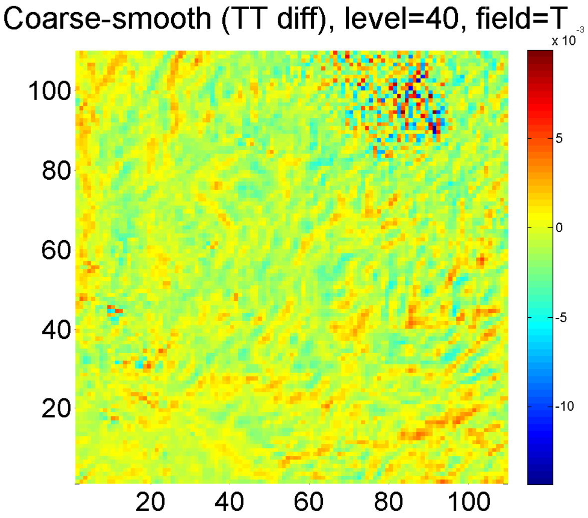
Process-level errors: model errors due to *numerics*

Numerical approximation (discretization) model errors: setup

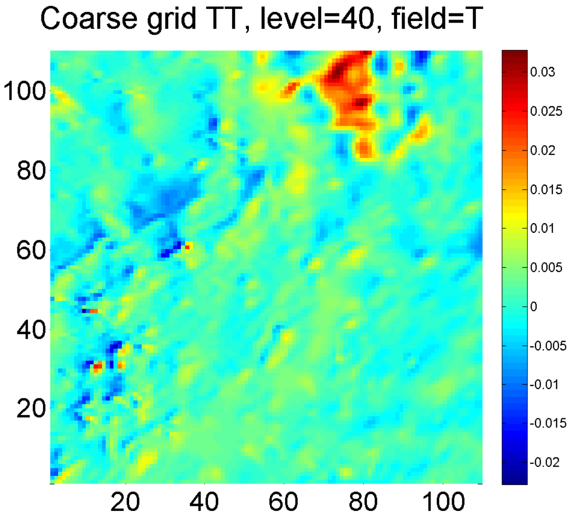
"Physics" is switched off in both "model" and "true model". Look at the difference in the 1-min tendencies (model errors due to the numerics only).

Cases:
– 1 July ("convective day")
– 17 July ("non-convective day")

Model-error field due to numerics (T, level 40)



The respective coarse-grid **total tendency** (T, level 40)



Conclusions on model errors due to numerics

1. "Numerical" model errors are an order of magnitude less than "physical" model errors.
2. When and where the total tendency is large, the numerical model error can be comparable with the physical model error.
3. The model-error field looks like the white noise with spatially variably intensity. The intensity grows with the growing magnitude of the **total** tendency $\mathcal{T}$:

$$\varepsilon_{\rm num}(\mathbf{s}) = \sqrt{\sigma_{\rm add}^2 + \sigma_{\rm mult}^2 \mathcal{T}^2(\mathbf{s})} \cdot \omega(\mathbf{s})$$

where ω is the white noise.

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:
  - 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 (for T, u, v).
  - Process-level treatment of model errors is needed.
- Numerical model errors can be modeled as the white noise whose intensity is proportional to the *total* tendency with an offset.

Further steps

- Extend the process-level model-error treatment from numerics to individual physical parameterizations.
- Multivariate and spatio-temporal aspects are to be addressed in an objective/justified way.

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

Thank you!