Progress in objective estimation of COSMO model errors using coarse graining

Michael Tsyrulnikov and Dmitry Gayfulin

HydroMetCenter of Russia

Rome, 9 Sep 2019

Outline

1. Approach
2. Lessons from the previous year
3. Model error: a new definition
4. A new model-error estimation technique
5. Numerical experiments

Approach

For a low-resolution model in question, compute the model error with respect to a significantly higher-resolution model.

That is, start the two models from "the same initial data", compute the two short-term tendencies and claim that their difference is the model error.

Lessons from the previous year

1. Model error due to convection appears to be too complicated to be treated with a purely stochastic model. A physical model is needed for this purpose. Conclusion: a stochastic convection parameterization is to be used.
2. The estimated model error fields appeared to have a multi-scale and, likely, multi-component structure. So, process-level model errors are best to be treated.
3. The 550-m high-resolution model is not high-resolution enough to be regarded as the truth.
4. Soil fields and static model fields are to be more carefully treated.
5. The starting point for the high-resolution model's tendency forecast was generated from the **low-resolution** field. As a result it didn't contain sub-grid scales (i.e. the fields' components that are resolved on the fine grid but not resolved on the coarse grid).

The new setup

1. Convection: now we select winter cases and switch off the convective parameterization.
2. Process-level treatment: hardly possible because on the convective scales, turbulence, convection, gravity-wave drag become increasingly resolved and there is no filter to isolate a convection plume from a turbulent eddy, say.
3. The 550-m high-resolution model is not high-resolution enough: now, we work with a 220-m L130 (or L65) model.
4. Soil fields and static model fields are are now more carefully treated (upscaled).
5. The starting point for the high-resolution model's tendency forecast should contain the the sub-grid scales (w.r.t. the coarse grid): this is the hardest issue we have faced (described later).

Notation

The "model" $=$ the coarse-grid model (cgm)
The "true model" $=$ the fine-grid model (fgm)

cgm state variables: UPPER CASE $(X_k, F_k ...)$
fgm state variables: lower case $(x_k, f_k ...)$

Tendency forecasts of the same length Δt:
cgm model: $X^f = F(X_{\mathrm{start}})$

fgm model: $x^f = f(x_{\mathrm{start}})$

Model error:

$$\boxed{\varepsilon = F(X_{\mathrm{start}}) - f(x_{\mathrm{start}})}$$

where $X_{\mathrm{start}} \stackrel{?}{=} \tilde{x}_{\mathrm{start}}$, where $\tilde{x}$ denotes the *upscaling* which removes the sub-grid scales (the same start condition)

Generation of perfect model-error perturbations

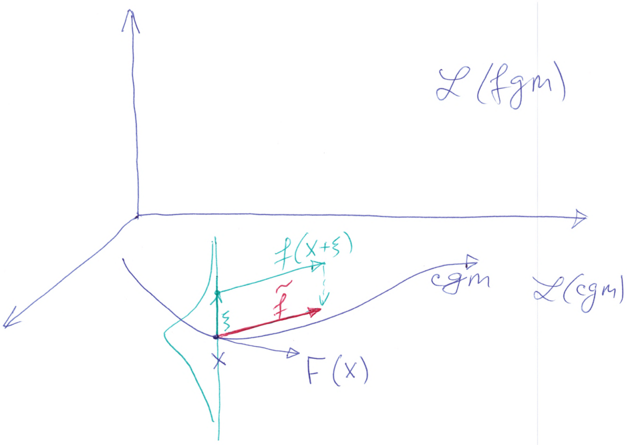
1. Take a cgm-forecast (an ensemble member) X. The cgm tendency at this point is $F(X)$.
2. Add sub-grid scales ξ.
3. Start fgm from $X + \xi$ and compute the fgm tendency $f(X + \xi)$.
4. Project f onto cgm-space $\mathcal{L}_{\mathrm{cgm}}$, getting $\tilde{f}$ – the ideal perturbed tendency.

Hence the new model-error definition:

$$\varepsilon = F(X) - \tilde{f}(X + \xi)$$

Note that on convective scales, model error becomes, largely, model uncertainty.

Generation of perfect model-error perturbations



Having this definition of model error, how can we estimate $\varepsilon = F(X) - \tilde{f}(X + \xi)$?

Estimation of perfect model-error perturbations

The major problem N1 is that a realistic stochastic model for sub-grid scales (multivariate, non-stationary, etc.) is not available.

All we can do is to take the sub-grid scales from an fgm field.

That is, start from an fgm field x, project it on $\mathcal{L}_{\rm cgm}$ (getting $\tilde{x}$ and the sub-grid scale field $\xi = x - \tilde{x}$), and compute

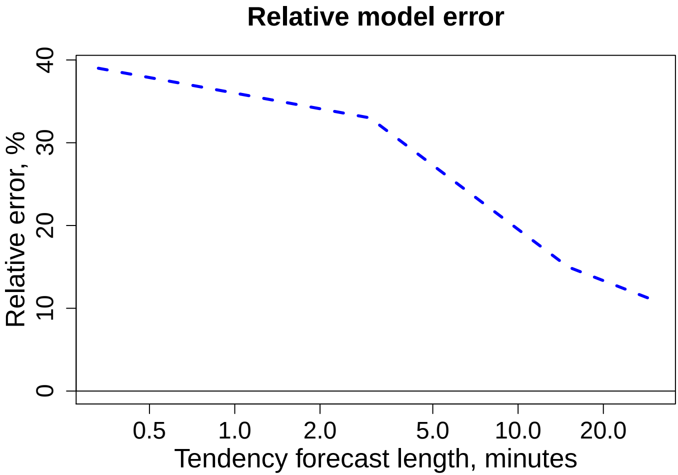
$$\widehat{\varepsilon} \stackrel{?}{=} F(x - \xi) - \tilde{f}(x)$$

instead of

$$\varepsilon = F(X) - \tilde{f}(X + \xi)$$

The major problem N2 is that $x - \xi$ appears to be not on the cgm attractor (i.e. not balanced) so that the small model error is invisible in the initial cgm shock, see the next slide.

Magnitude of the estimated model error



Fighting the imbalance: the proposed solution

1) Represent $\varepsilon = F(X) - \tilde{f}(X+\xi)$ as the sum of two components:

$$\varepsilon = \left[F(X) - \tilde{f}(X)\right] + \left[\tilde{f}(X) - \tilde{f}(X+\xi)\right] \equiv \varepsilon_{\mathrm{m}} + \varepsilon_{\xi}$$

The 1st term, ε_{m}, is due to the difference between cgm and fgm starting from the same cgm field without sub-grid scales.

$$\varepsilon_{\mathrm{m}} = F(X) - \tilde{f}(X)$$

can be estimated directly as we did it before.

The 2nd term, ε_{ξ}, is the uncertainty of the upscaled fgm tendency due to the presence of sub-grid scales in one of the two fgm starting fields.

2) Replace the unavailable 2nd term $\varepsilon_{\xi} = \tilde{f}(X) - \tilde{f}(X+\xi)$ by the available term

$$\widehat{\varepsilon}_{\xi} = \tilde{f}(\tilde{x}) - \tilde{f}(x) \equiv \mathcal{U}(f(\mathcal{U}x) - f(x))$$

3) Assume that ε_{m} and ε_{ξ} are stochastically independent.

Numerical experiments

The two models

The convection parameterization switched off in both models.

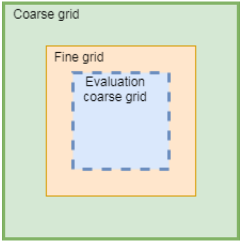
cgm is COSMO-L65 with the horizontal resolution 2.2 km and time step 20s.

fgm is COSMO-L130 (or L65) with the horizontal resolution 0.22 km and time step 2s.

Otherwise the two models are the same.

Domain and cases

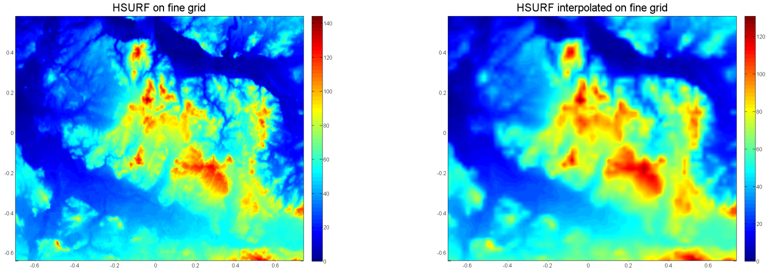
- Two domains: over land and over sea. The domains' centres are at (56N, 4E – North Sea) and (53N, 10E – Northern Germany).



- The coarse-grid-model's domain: 80*80 points (greenish).
- The fine-grid-model's domain: 851*851 points (187*187 km, pinkish).
- Model errors are computed on the 2.2-km 60*60 subgrid (bluish).
- 1 case was studied : 10 January 2019, 12 UTC.

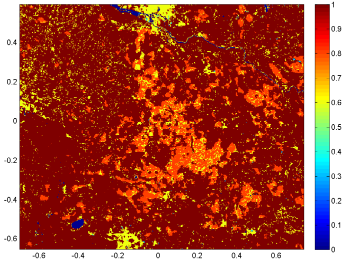
Static fields were taken from the fine grid and smoothed (upscaled, coarse-grained) before used in cgm

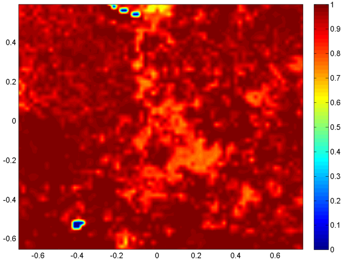
HSURF



Static fields were taken from the fine grid and smoothed (upscaled, coarse-grained) before used in cgm

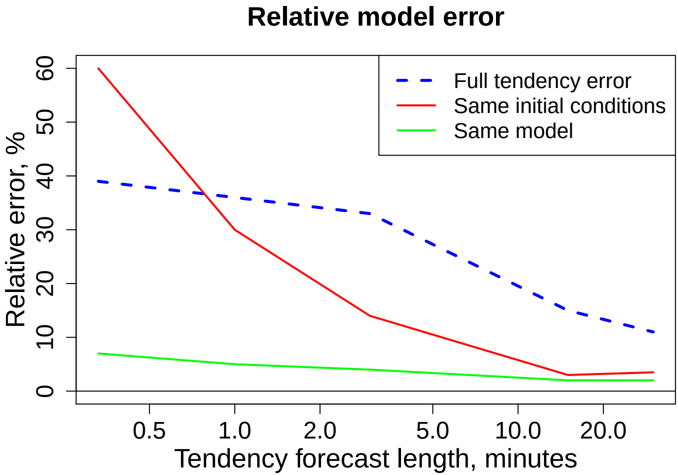
ROOTDP



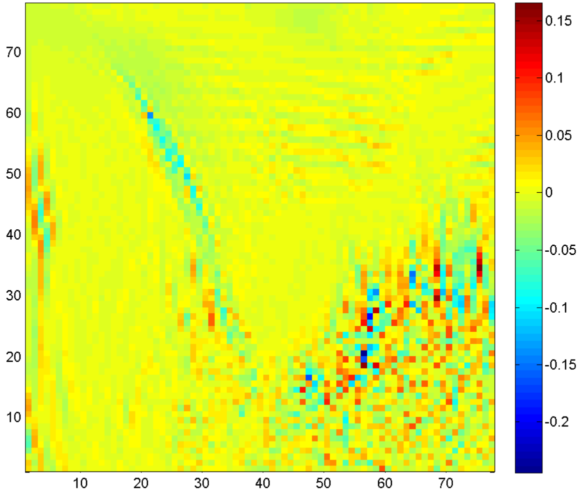


Magnitudes of the 2 estimated model-error components

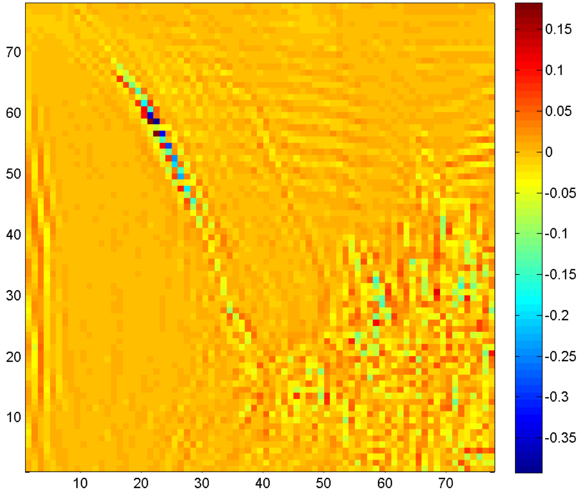
$$\boxed{\varepsilon = \varepsilon_{\mathbf{m}} + \varepsilon_{\xi}}$$



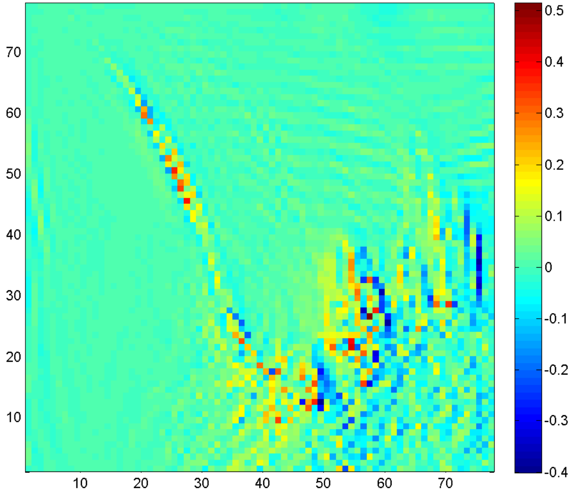
The model-error component due to the difference in the initial-field resolution. T. level 50



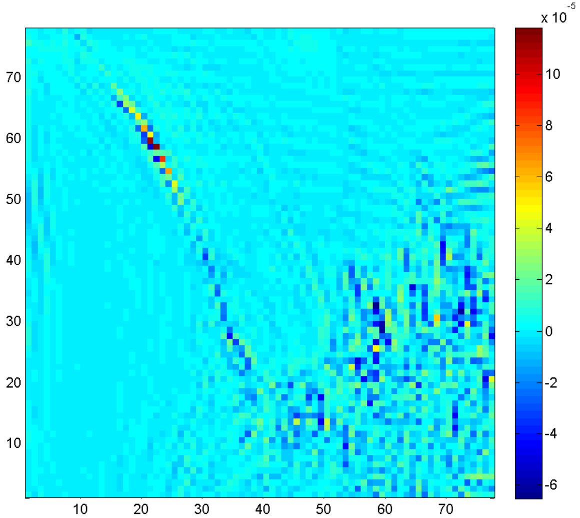
The model-error component due to the difference in the initial-field resolution. T. level 55



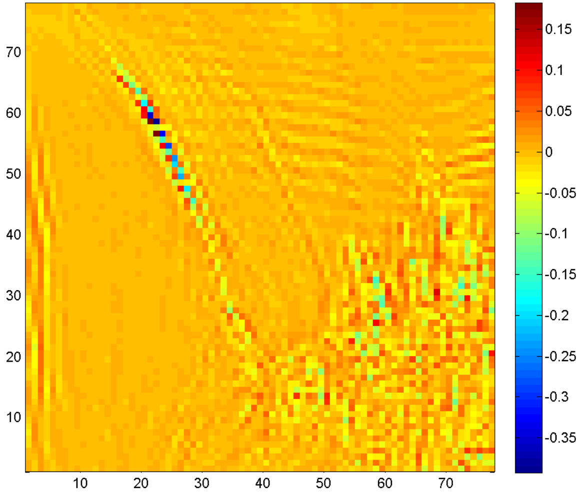
The model-error component due to the difference in the initial-field resolution. U. level 55



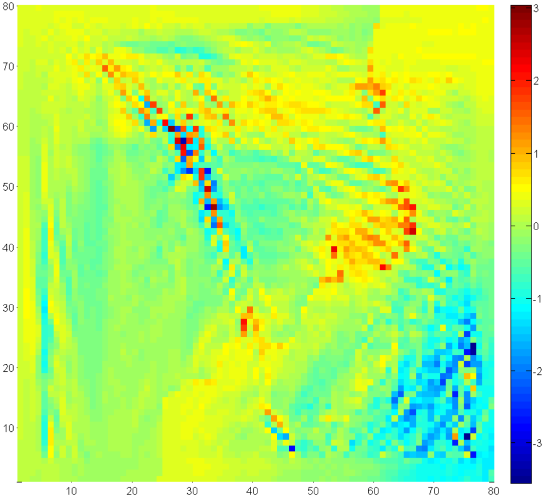
The model-error component due to the difference in the initial-field resolution. QV. level 55



The model-error component due to the difference in the initial-field resolution. T. level 55



The 20-s cgm (low-resolution) total tendency, ,
T, level 55



Conclusions

- A new definition for convective-scale model-tendency-errors estimator was introduced.
- A new, more precise, estimator for convective-scale model-tendency-errors was proposed.
- The model error has both multiplicative and additive components.
- The model error can be considered Gaussian in the first approximation.

Further steps

Stochastic modeling is to be done. The following aspects are to be addressed:

- Multivariate dependencies (balances, cross-correlations).
- 3D and 4D (spatio-temporal) correlations.
- Non-stationarity due to a link to the current state and/or current physical tendency is to be identified.
- An appropriate and as-simple-as-possible stochastic model is to be devised and estimated.
- Geographical and seasonal differences are to be accounted for.

The goal still being a justified practical convective-scale model-error model.

Thank you!

Many thanks to D. Blinov for his help with the COSMO model.