The work on Stochastic Physics in the NWP Models at DWD: Status Report

Ekaterina Machulskaya

DWD, FE14

Working Group 7
COSMO General Meeting
Eretria, Greece
8-11 September 2014

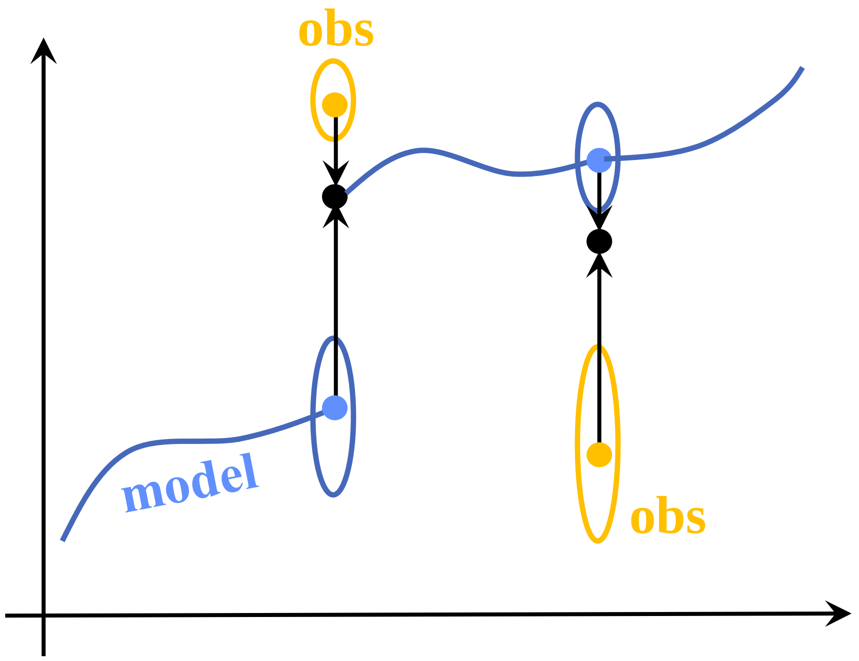
Outline

- Motivation: what we expect from a stochastic approach?
- A model for the model error
  - functional form
  - estimation of model parameters
- Results from COSMO-DE simulations
- Outlook

Motivation: estimation of the background error

In Kalman Filter, the weights for the interpolation between the observations and the model are inversely proportional to the corresponding uncertainties, or possible errors

The uncertainty of the first guess combines the propagated error of the last analysis and the model error



An estimate of the model error is needed in order to give an appropriate weight to the first guess. If the model error is underestimated, this weight will be too large and less regard will be paid to the observations than should be.

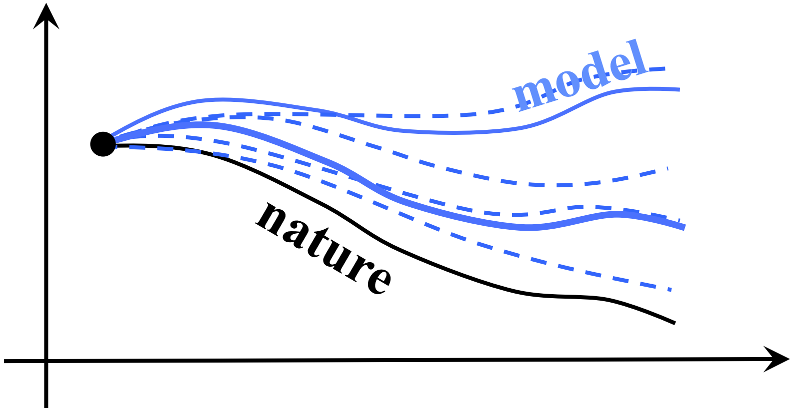
Motivation

Motivation 2:

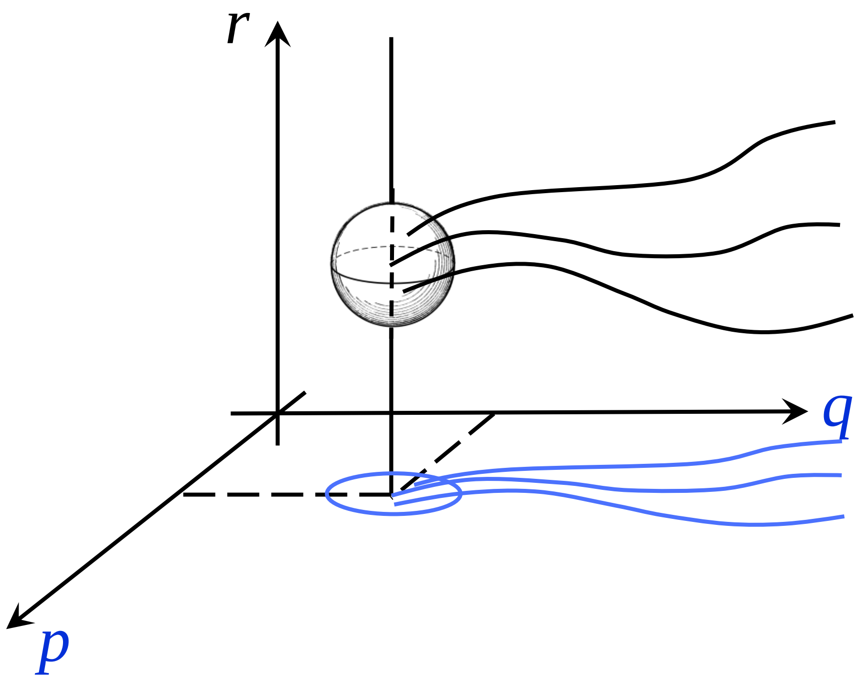
The end-users should be provided with the information how reliable/uncertain the forecast is.

Motivation 3:

If the perturbations are chosen correctly, the ensemble mean can be better then the deterministic forecast.



Formulation of the problem



$\{p, q, r\}$ –

full set of modes (= nature)

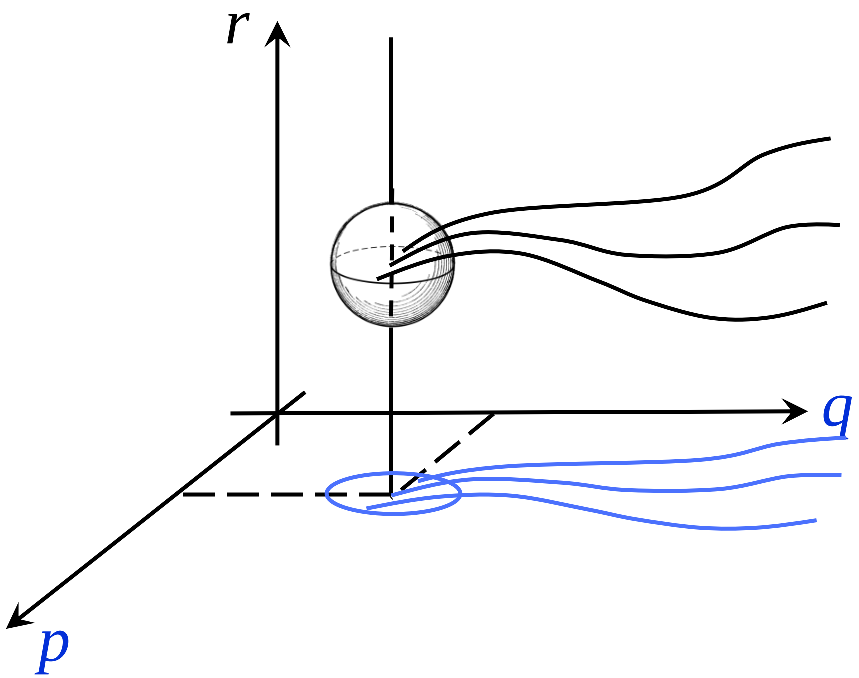
$\{p, q\}$ –

model variables

r – unaccounted degree of freedom

Usually, the exact initial condition is not known.

Formulation of the problem



$\{p, q, r\}$ –

full set of modes (= nature)

$\{p, q\}$ –

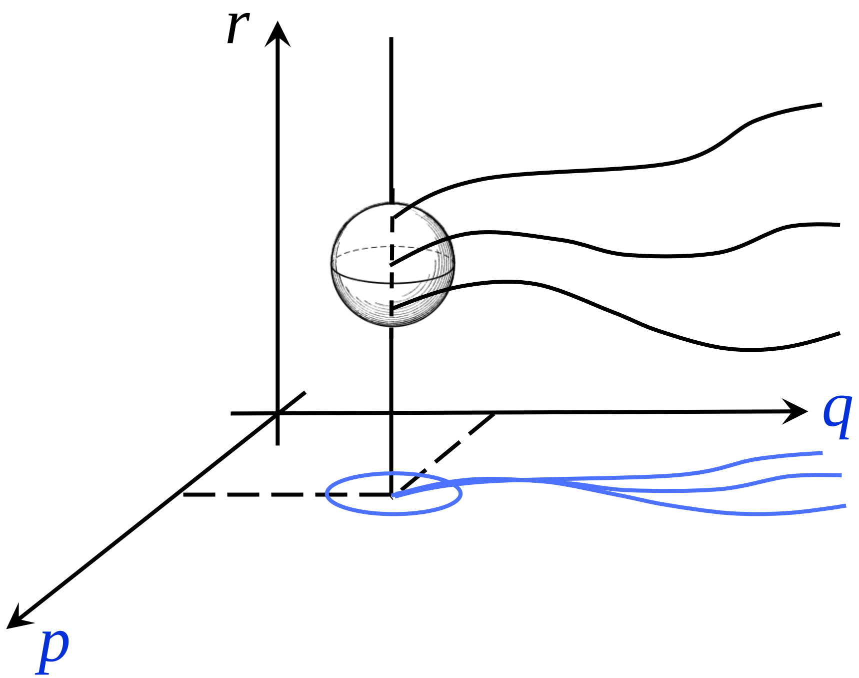
model variables

r – unaccounted degree of freedom

Usually, the exact initial condition is not known.

The lack of knowledge in the model variable's plane (p,q) = the uncertainty in the model's initial conditions.

Formulation of the problem



$\{p, q, r\}$ –

full set of modes (= nature)

$\{p, q\}$ –

model variables

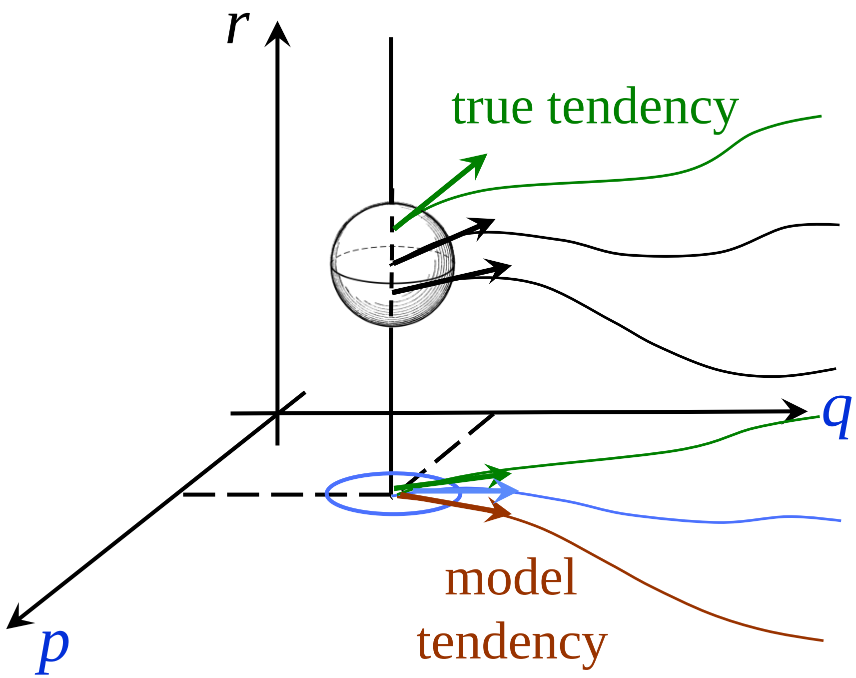
r – unaccounted degree of freedom

Usually, the exact initial condition is not known.

The lack of knowledge in the model variable's plane (p,q) =
the uncertainty in the model's initial conditions.

The lack of knowledge in the unresolved mode r =
the uncertainty in the model's physics.

Formulation of the problem



The deterministic model does not know the projection of the **true tendency**.

It choses one of many possible tendencies which all are realizable from the model point of view, i.e. if the knowledge about the unresolved modes is missing.

The objective of the stochastic forecast is to provide the spread in tendencies due to unresolved degrees of freedom (processes).

Only those errors can be represented by means of the stochastic approach.

A way to go

One way is to approximate the <u>empirically determined</u> entire model error by a random process with the <u>same statistical properties</u>.

<u>Disadvantage</u>: lack of the understanding of the physical processes

<u>Advantage</u>: the entire model error is represented, important for DA

<u>How to estimate model error</u>

Ideally it should be the series of the “one-step tendency error” to exclude the interactions between the model and model error.

As a proxy take the differences “forecast – analysis” as frequent as possible (3 hours).

A model for the model error

White noise $$\frac{\partial T}{\partial t} = \left[\frac{\partial T}{\partial t}\right]_{det} + \sigma\xi(t), \quad \xi(t) \sim N(0,1),$$

with no correlations in space and time

Bad approximation of the model error.
+ the model does not feel those perturbations

The noise should be **red**, i.e. correlated in space and time.

$$\frac{\partial T}{\partial t} = \left[\frac{\partial T}{\partial t}\right]_{det} + \eta(t)$$

The only equation that describes stationary Markov continiuos random process with non-zero time correlation is the Ornstein-Uhlenbeck equation

$$\frac{\partial \eta}{\partial t} = -\gamma\eta + \sigma\xi(t)$$

A model for the model error

$$\frac{\partial \eta}{\partial t} = -\gamma\eta + \sigma\xi(t)$$

random component, $\xi(t) \sim N(0,1)$

persistence in time, $\frac{1}{\gamma}$ – characteristic time scale

Adding spatial correlations

$$\frac{\partial \eta}{\partial t} = -\gamma\eta + \gamma\lambda^2\nabla^2\eta + \sigma\xi(x,t)$$

diffusion measures spatial influence

σ, γ, and λ should be flow-dependent
and can be determined from the available statistics

A model for the model error

Why $\xi(t) \sim N(0,1)$, i.e. Gaussian? (Why not e.g. a uniform as in SPPT?)

Central Limit Theorem:
sum of many independent identically distributed random variables is Gaussian

→ the normally distributed independent increments are the only increments that consist of many smaller increments with the same distribution

→ the process that stands on the right-hand side of a SDE can have a Gaussian distribution only

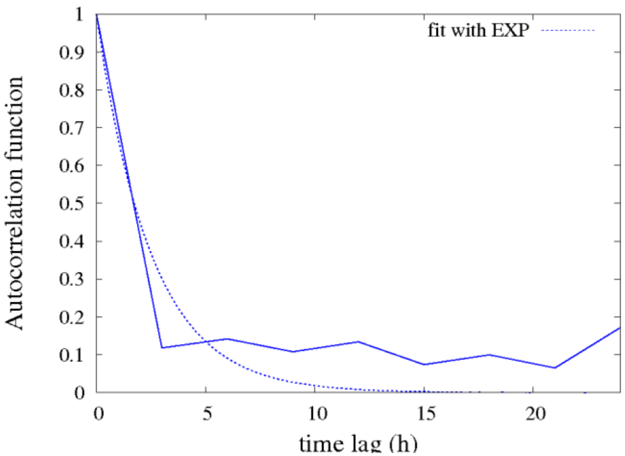
A numerical scheme with any other process
does not correspond to the discretization of any analytical SDE.

Determination of the parameters

By definition $\gamma = \lim_{\Delta t \to 0} \left\langle \frac{\eta(t+\Delta t) - \eta(t)}{\eta(t)\Delta t} \right\rangle$ $\sigma = \lim_{\Delta t \to 0} \left\langle \frac{\left(\eta(t+\Delta t) - \eta(t)\right)^2}{\Delta t} \right\rangle$

In practice, $\gamma = \frac{1}{N} \sum_{i=1}^{N} \frac{\eta_i(t+\Delta t) - \eta_i(t)}{\eta_i(t)\Delta t}$ $\sigma = \frac{1}{N} \sum_{i=1}^{N} \frac{(\eta_i(t+\Delta t) - \eta_i(t))^2}{\Delta t}$

where $\Delta t = 3$ hours and the parameters are determined for each bin of a predictor that characterizes the flow



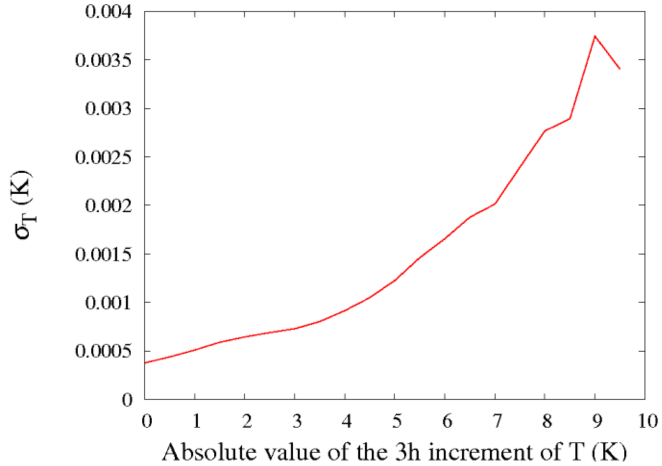
γ: approximation of the exponent of the autocorrelation function

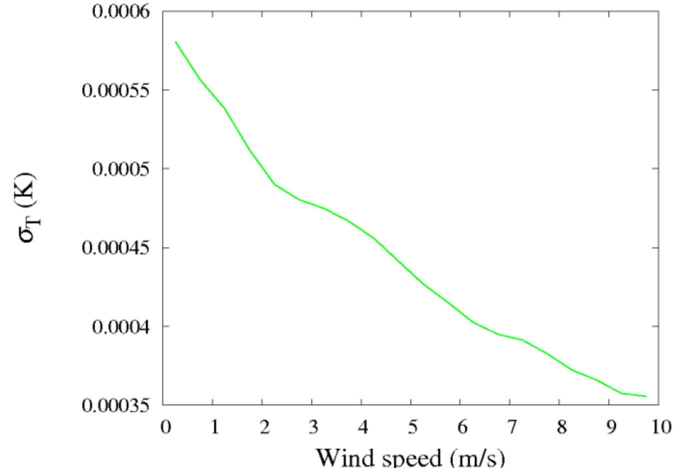
Spatial autocorrelation function (from data)

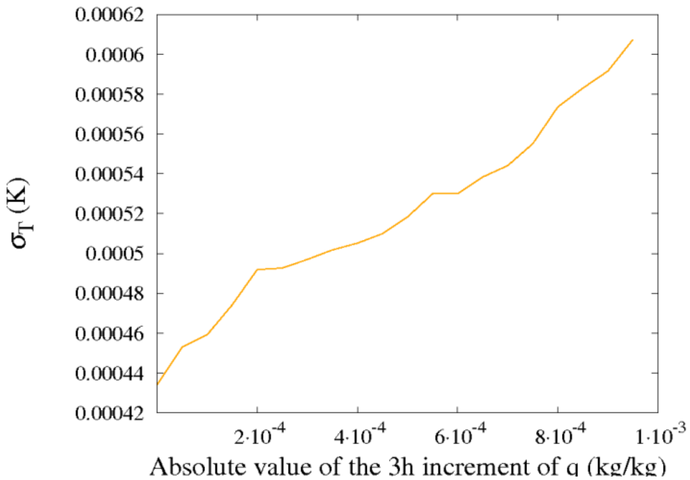
$$G(\vec{r}) = \sum_{\vec{k}} \frac{\cos(\vec{k} \cdot \vec{r})}{1 + \lambda^2 \vec{k}^2}$$

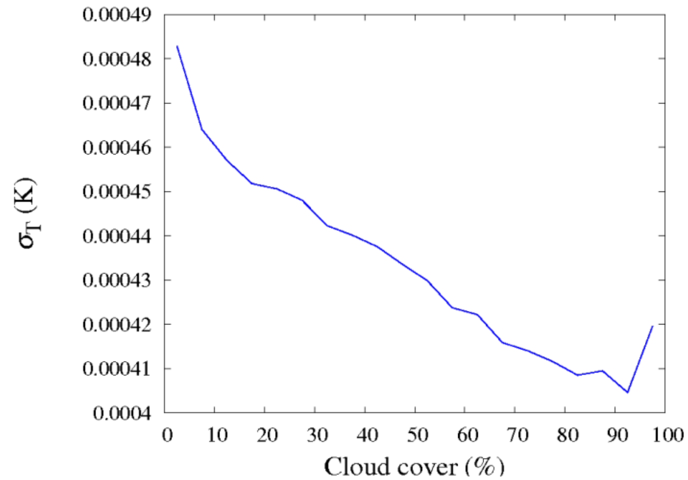
Implicit equation with respect to λ for certain $\vec{r}$

Temperature error: variance







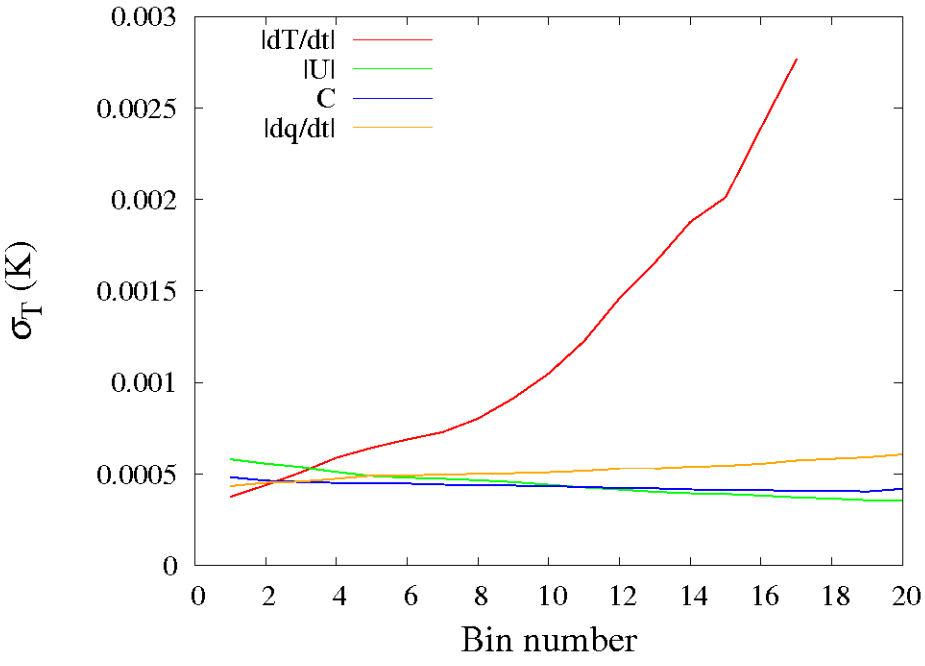


There is clear dependences of the variance on some quantities
→ they may serve as predictors…

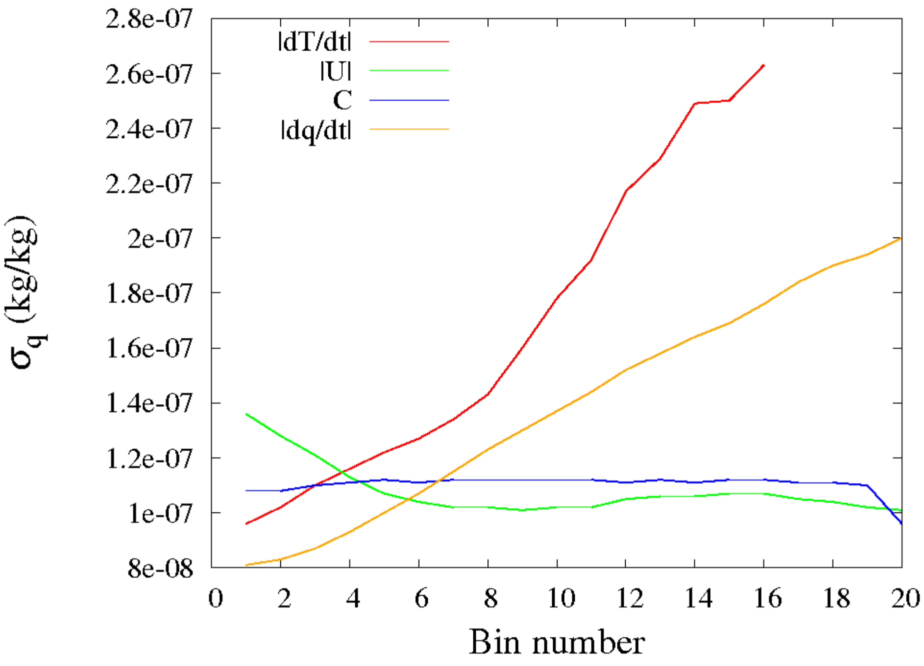
Variances: summary

…but their relative importance is different

temperature



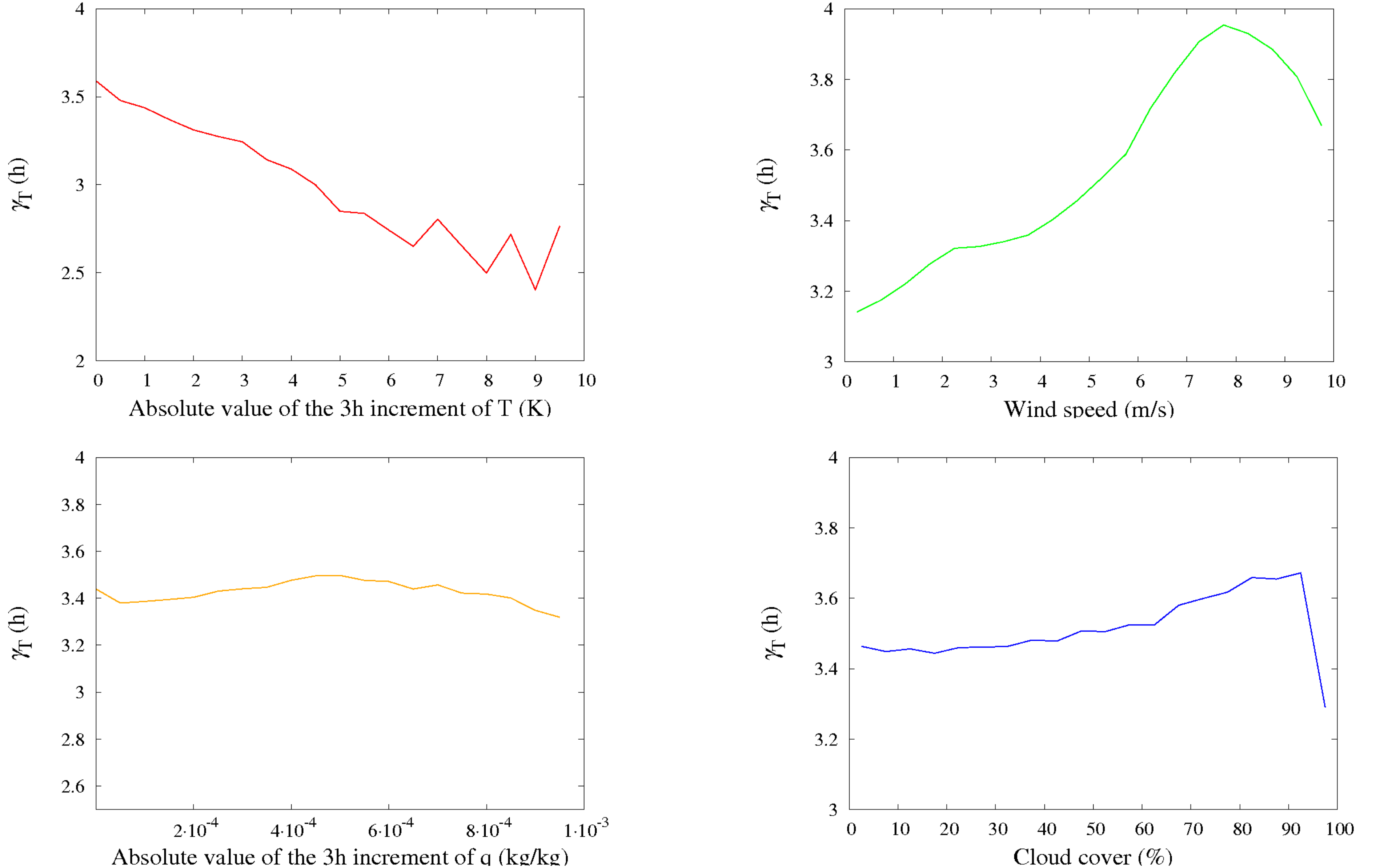
specific humidity



The most important is the dependence on |dT/dt|

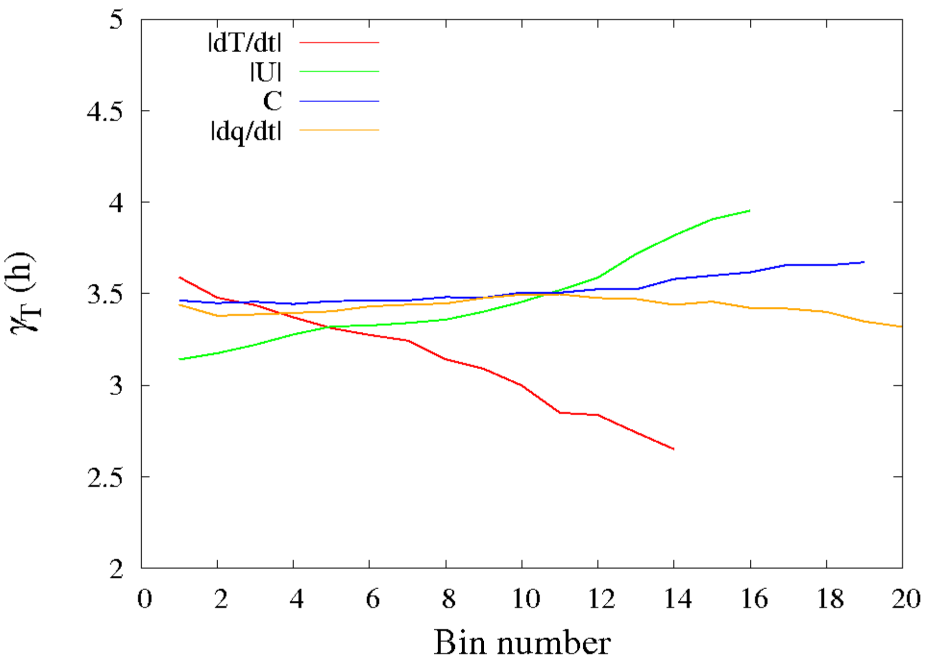
The most important are the dependences on |dT/dt| and |dq/dt|

Temperature error: time correlations

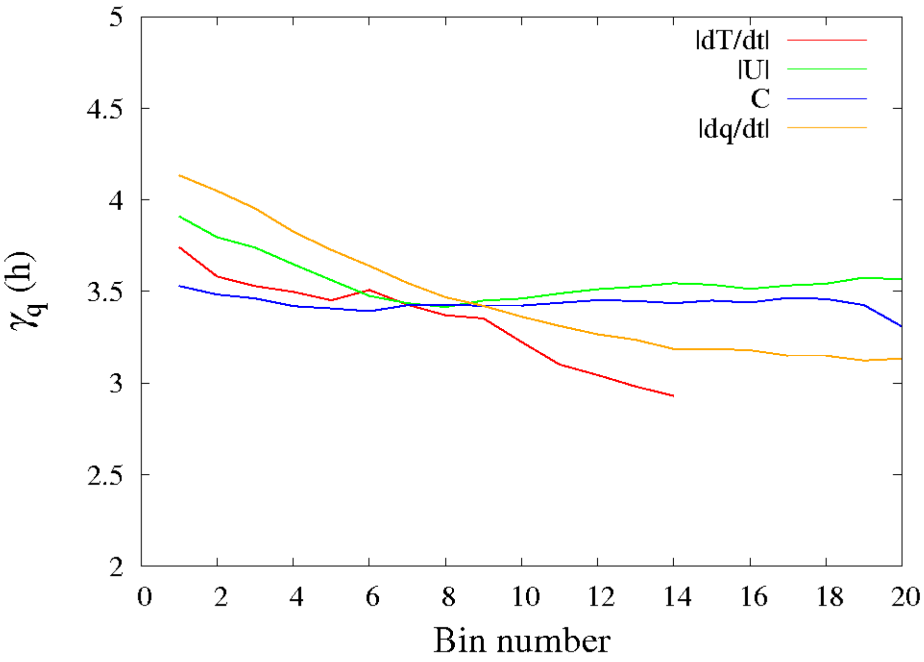


Time correlations: summary

temperature



specific humidity

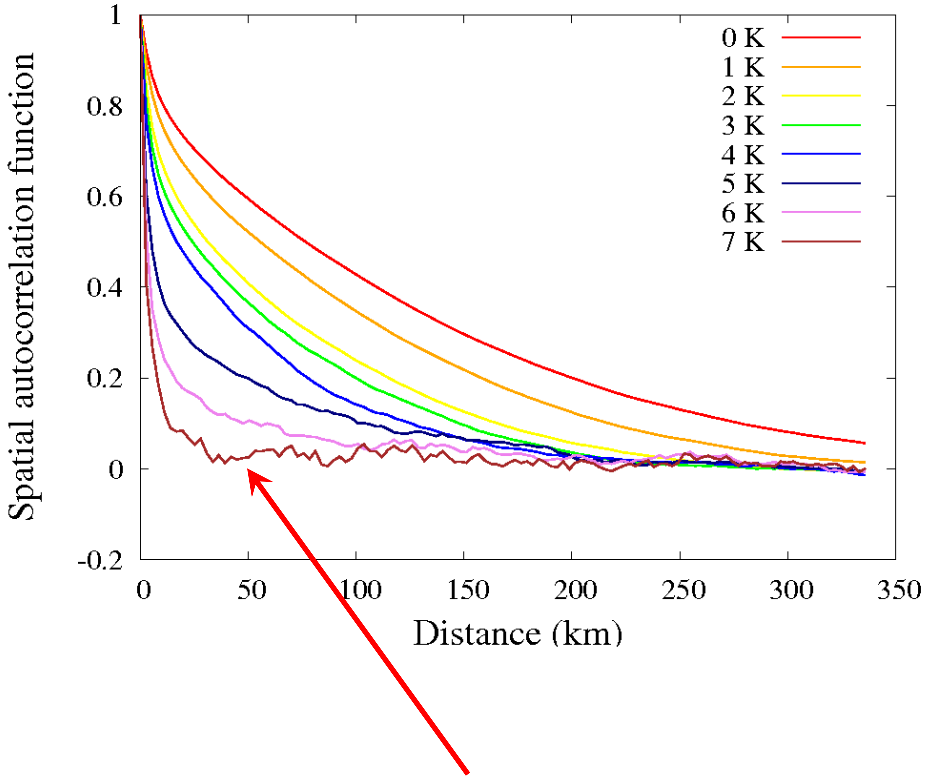


The most important are the dependences on |dT/dt| and wind speed

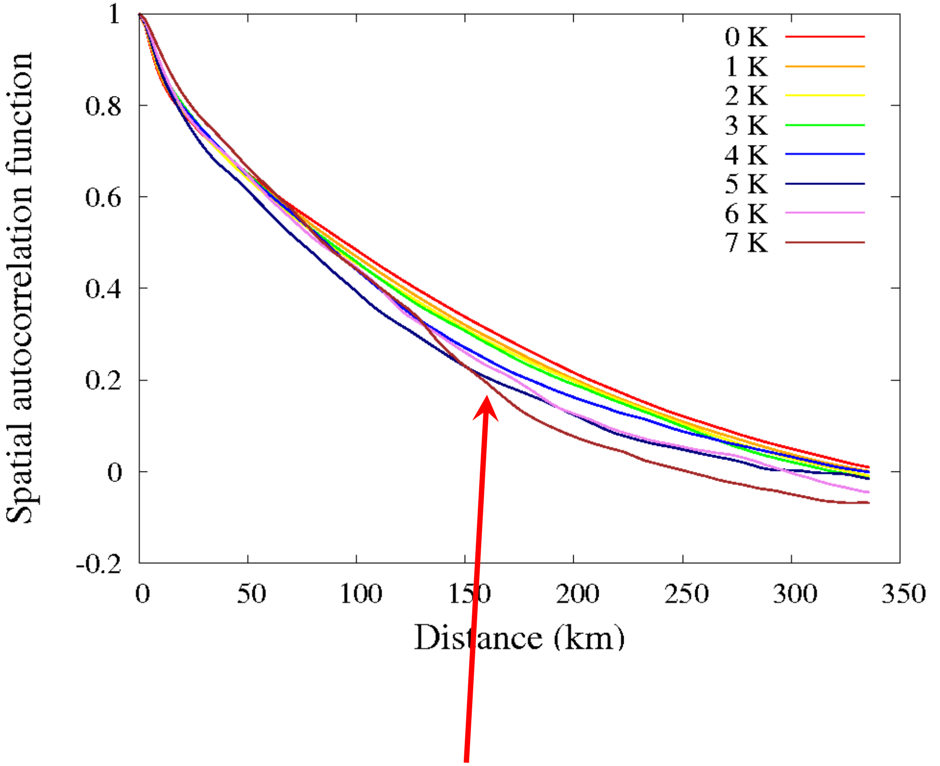
The most important are the dependences on |dT/dt| and |dq/dt|

Spatial correlations

near the surface



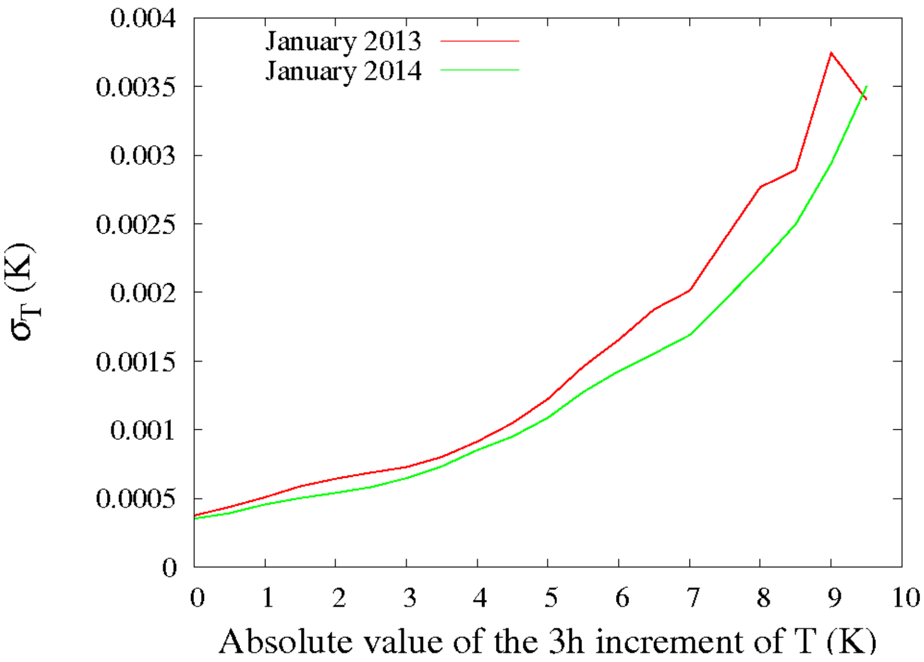
at 5 km height

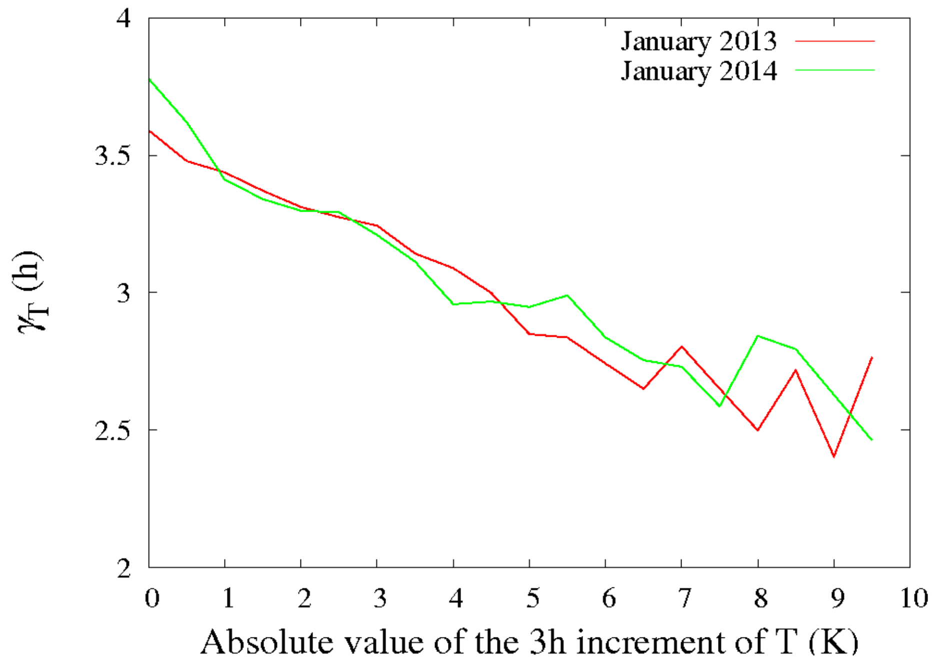


Near the surface the errors at large |dT/dt| are localized

In the free atmosphere all errors have long correlations

Persistence over time: 2013 vs 2014





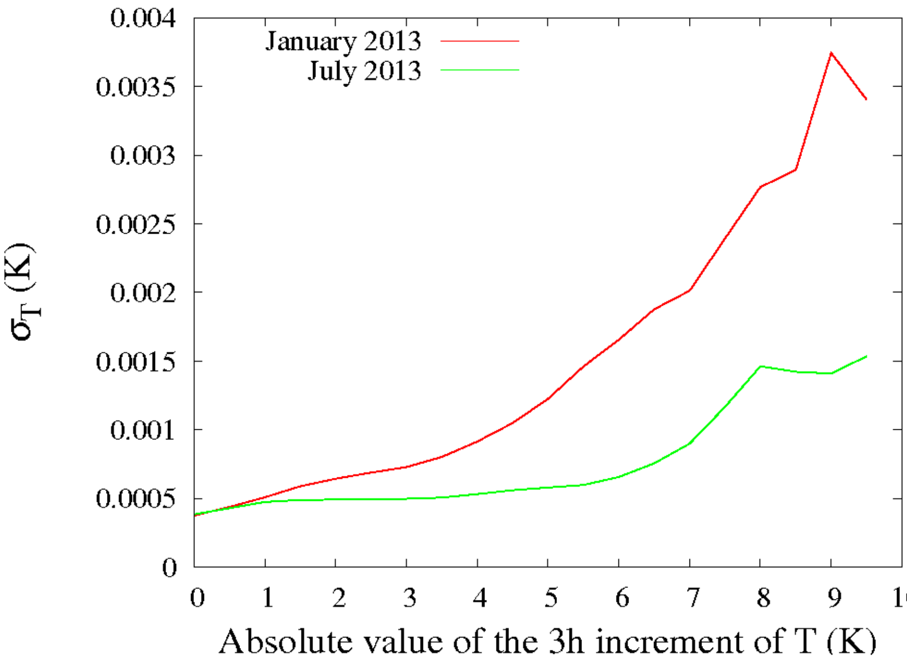
Variance: <u>similar</u> dependences, although the error variance in 2014 is slightly smaller than in 2013

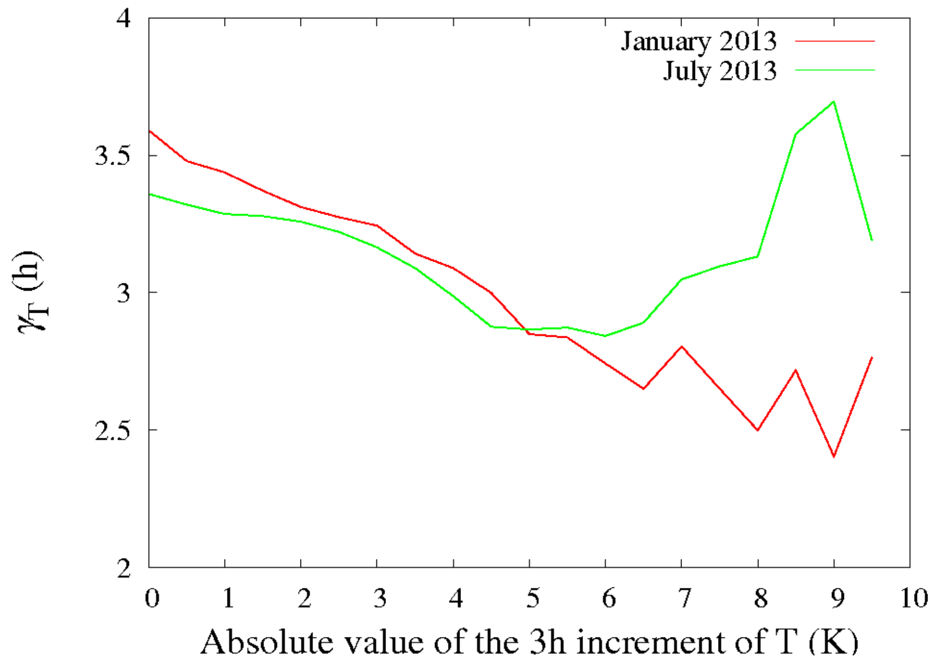
Time correlation: <u>similar</u> dependences

Spatial correlations are also similar (not shown)

Parameters determined during the training period may be used for the forecasts (probably with a slight adjustment each year)

Persistence over time: January vs July

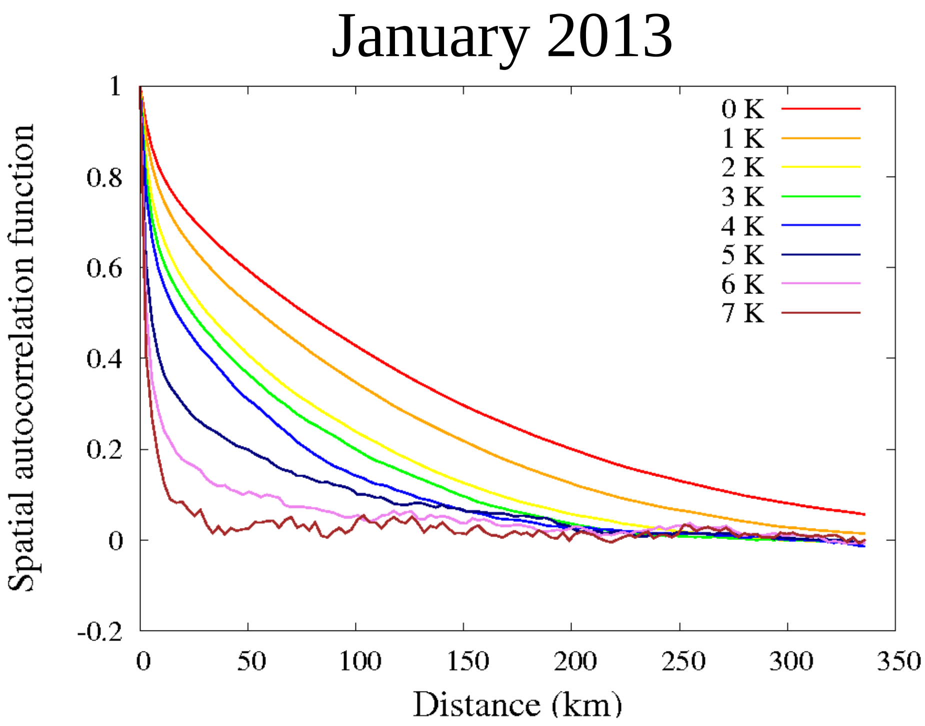


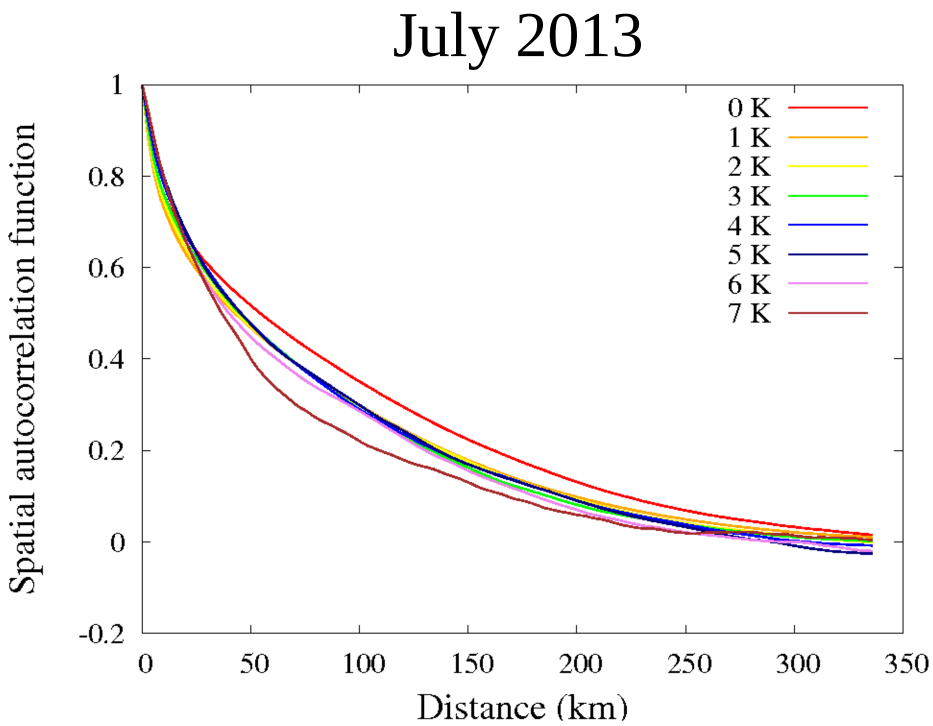


Variance: there is a difference between various seasons (in summer the error variance is smaller)

Time correlation: presumably there is a difference

Persistence over time: January vs July





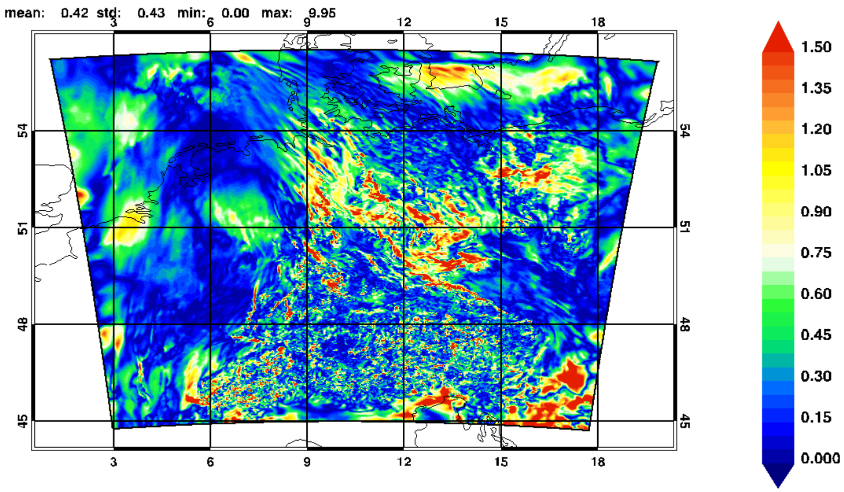
Spatial correlations: there is a difference between various seasons.

In winter the errors at large temperature tendencies are localized; in summer all errors have long correlations

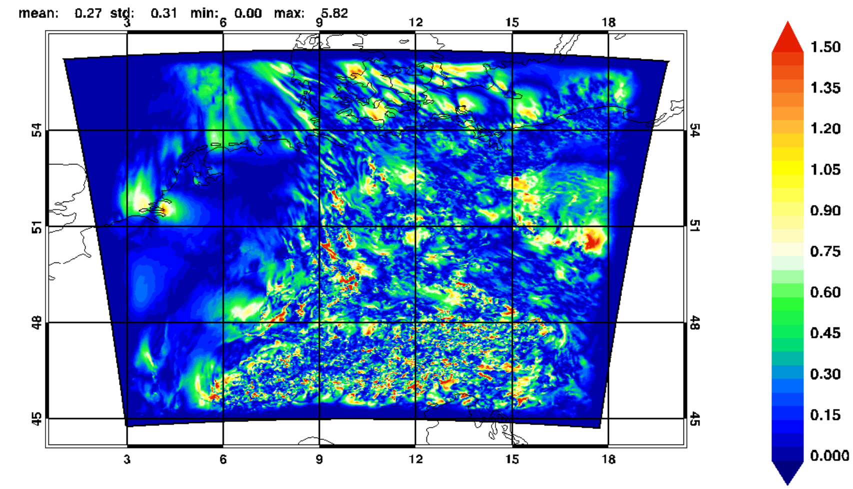
Experiments with 3d COSMO

T51, 02.01.2014, 00 UTC

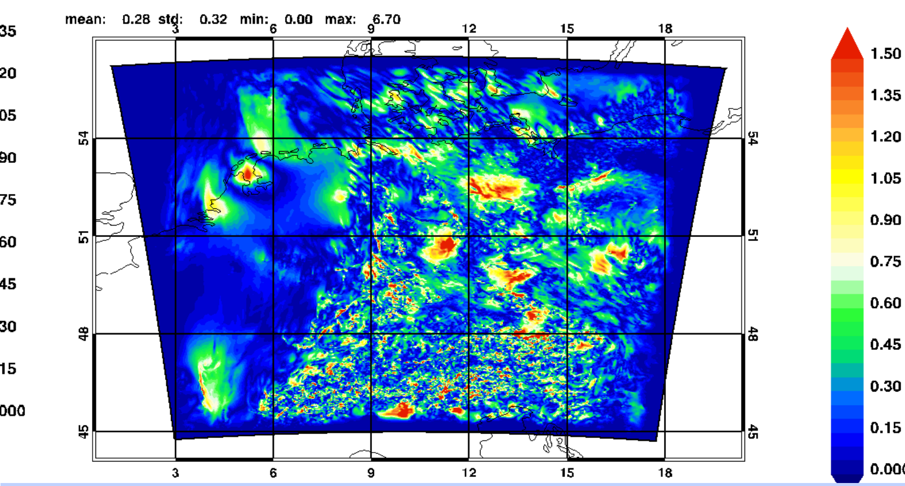
3 h for – ana



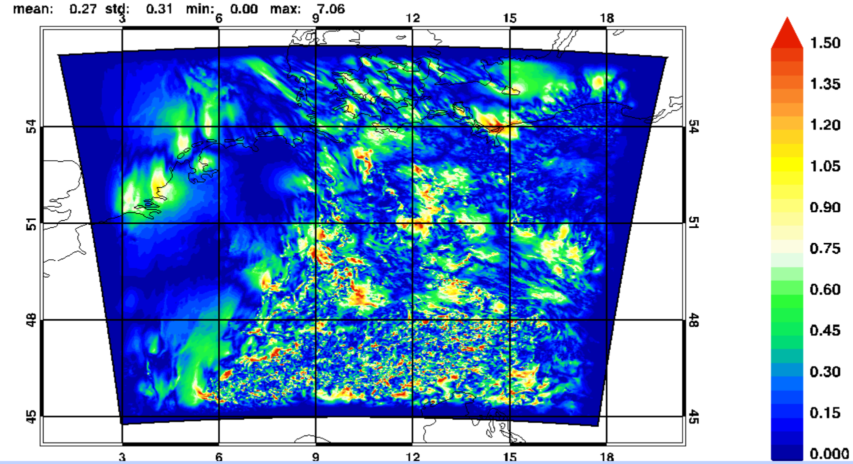
member 1 – ens mean



member 7 – ens mean



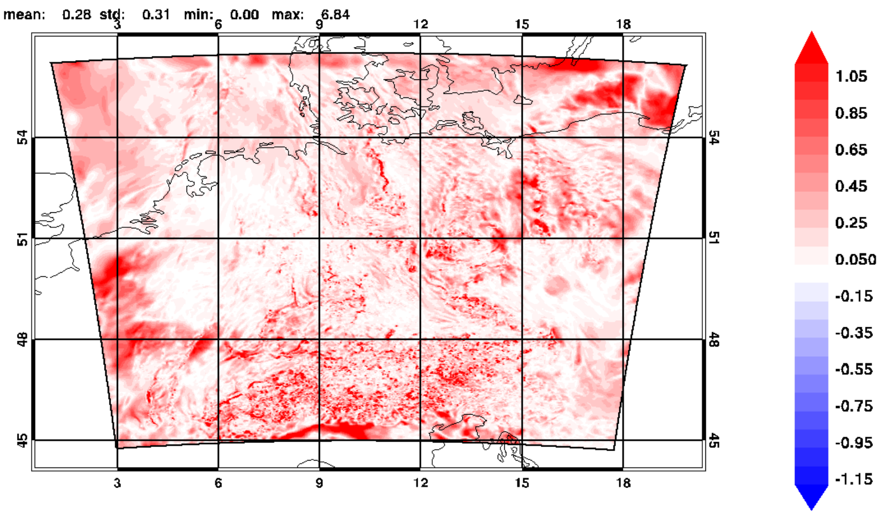
member 12 – ens mean



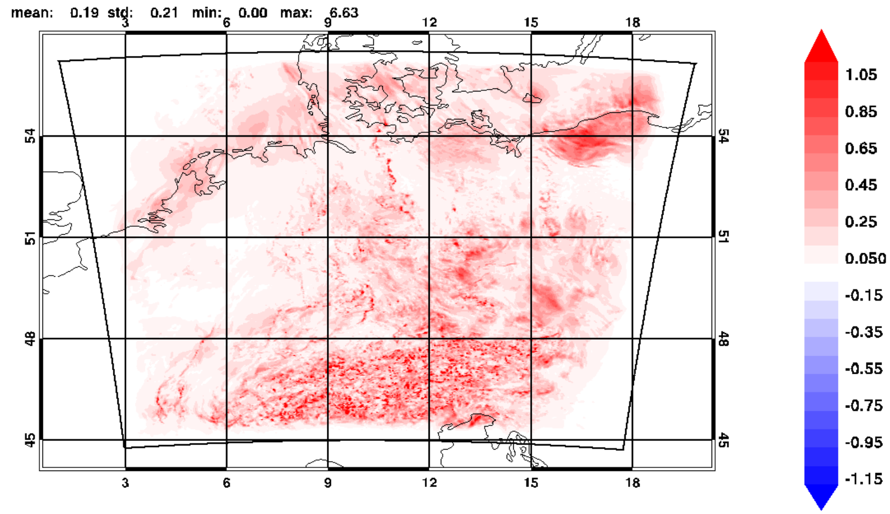
Experiments with 3d COSMO

T51, 01.01.2014, 03 UTC

3 h for – ana

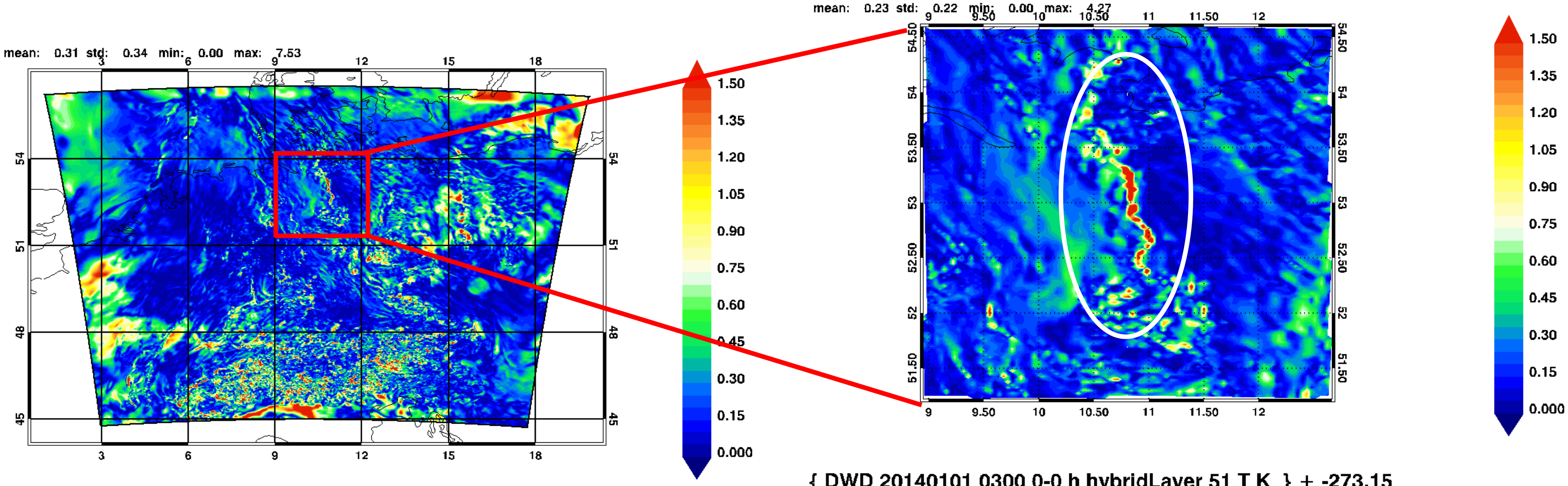


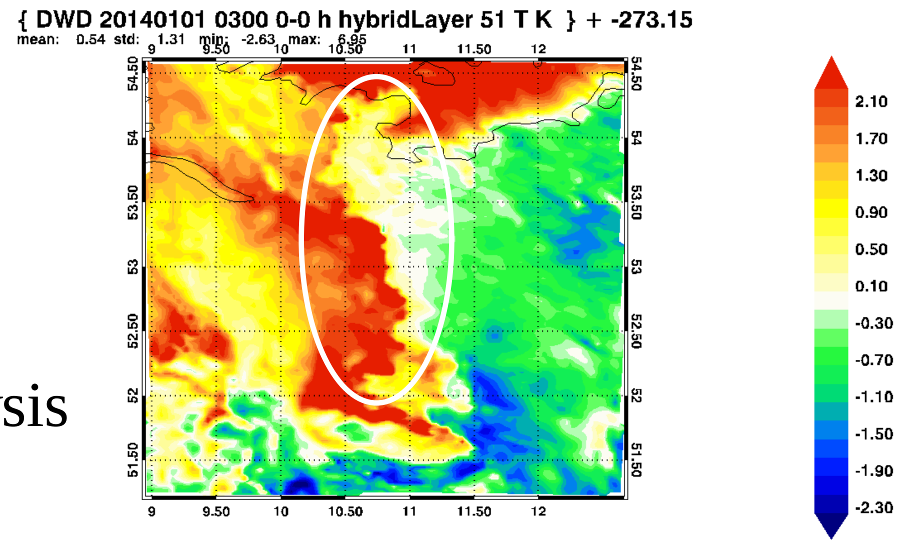
ensemble spread



Experiments with 3d COSMO

T51, 3 h for – ana, 01.01.2014, 03 UTC



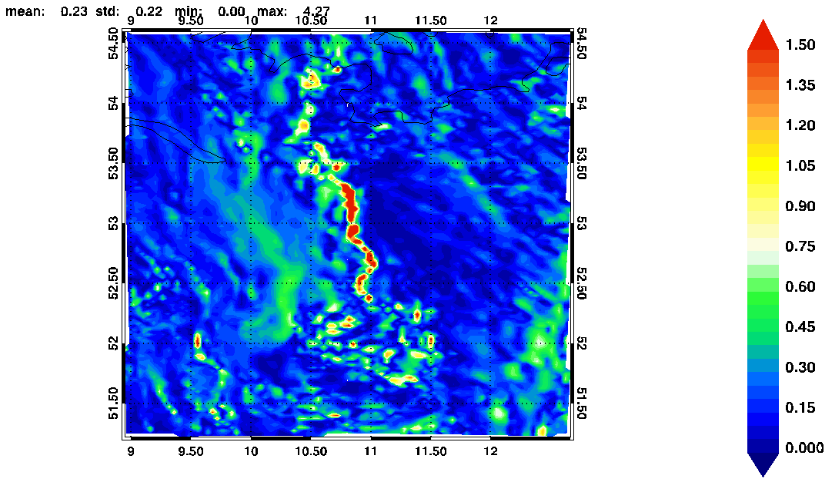


T51, analysis

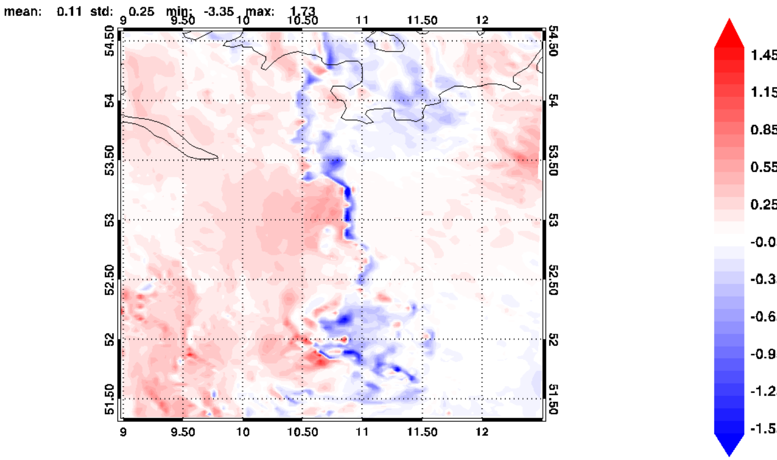
Experiments with 3d COSMO

T51, 01.01.2014, 03 UTC

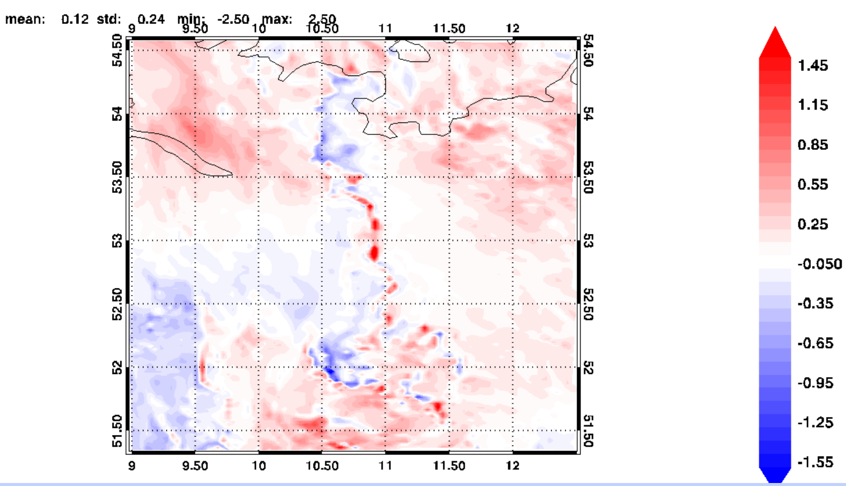
3 h for – ana



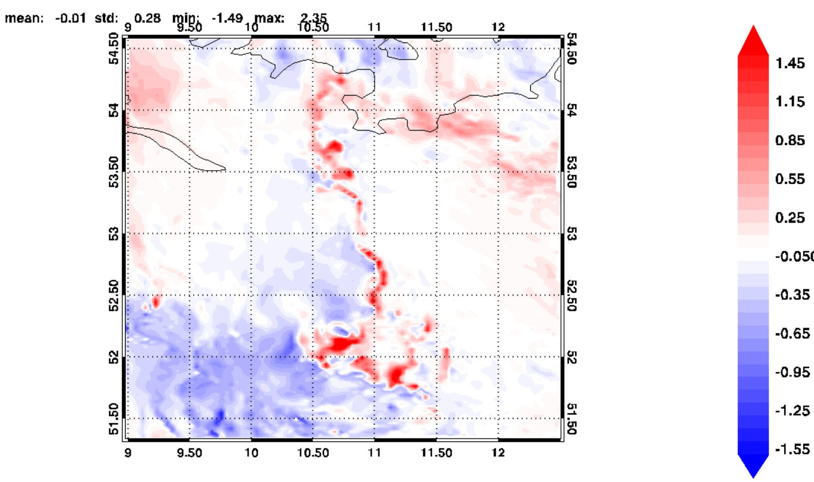
member 11 – ens mean



member 15 – ens mean



member 17 – ens mean



Summary

- The time series of the model error estimates are analysed
- A functional form for the model error is proposed
- An approach for the determination of the necessary parameters is developed
- The approach is implemented into COSMO-DE, parallel experiments are being performed, results look promising

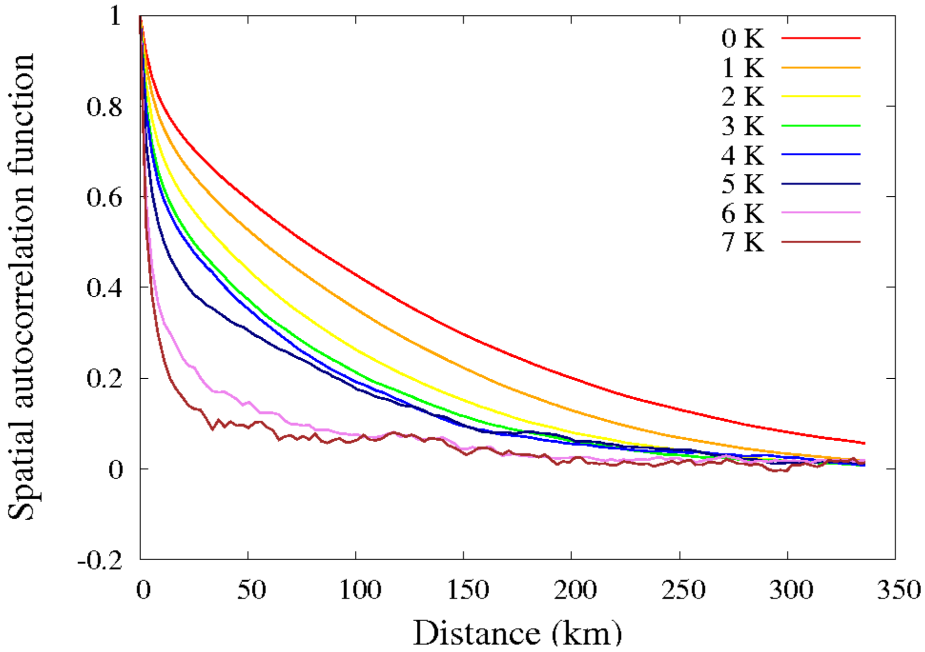
Outlook

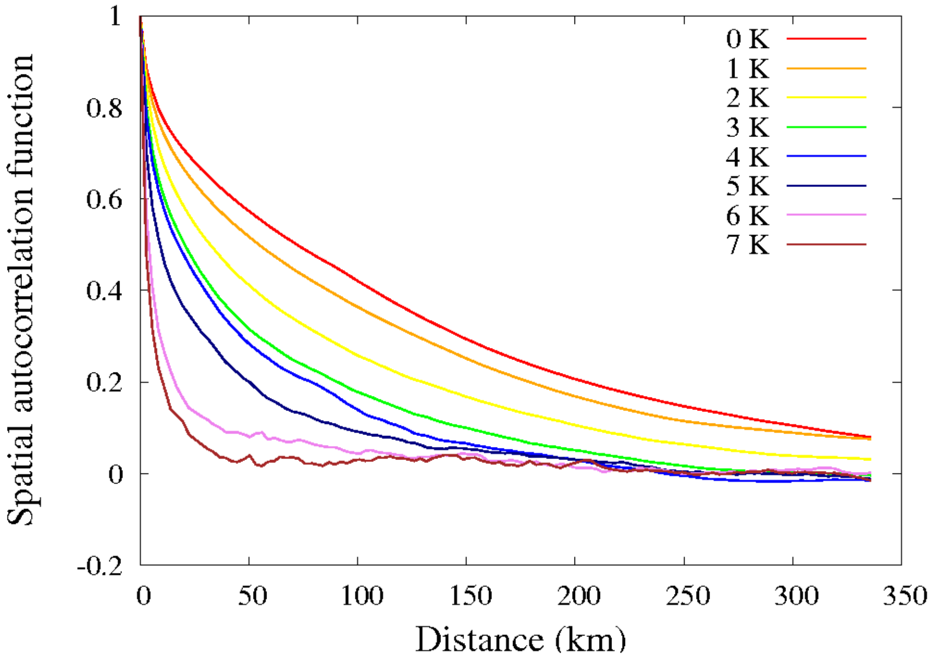
- Further testing of the implemented approach within the COSMO-DE (longer period, other seasons, behaviour of the ensemble mean vs. deterministic forecast, etc.)
- Verification of results by means of various ensemble prediction scores
- Development of a more physically plausible approach

Thank you for your attention!

Thanks to Jochen Förstner and Thomas Hanisch for technical support,
and Dmitrii Mironov and Bodo Ritter for fruitful discussions!

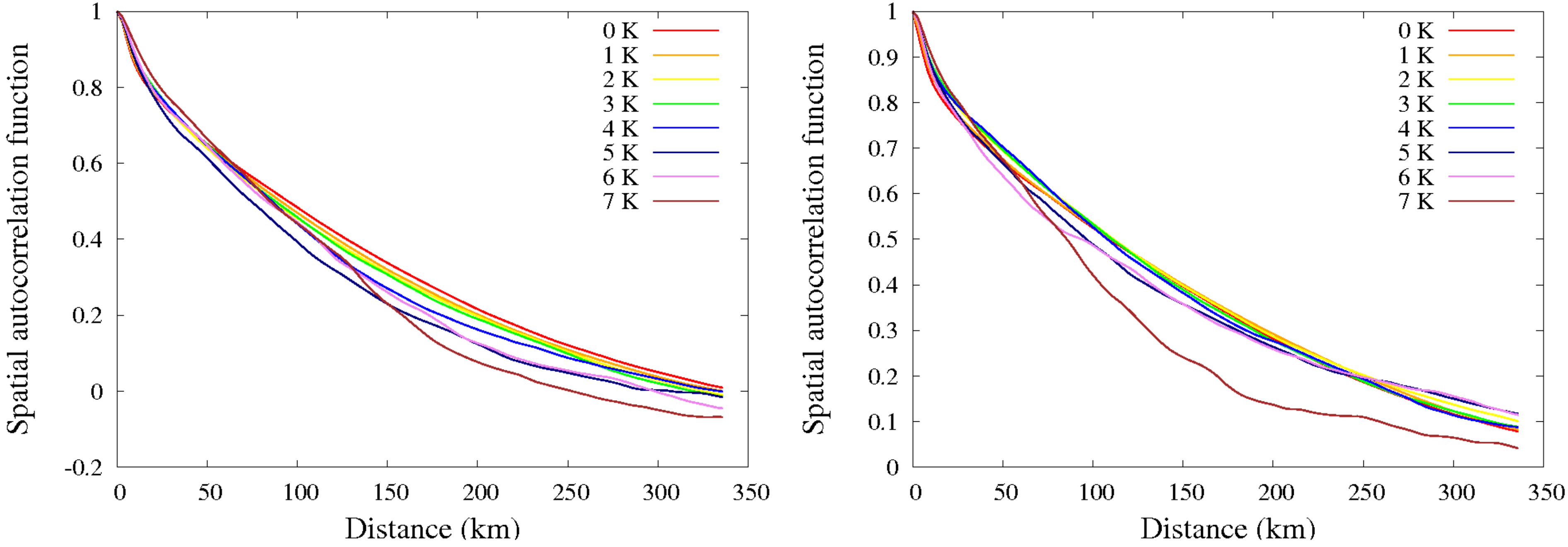
lambda





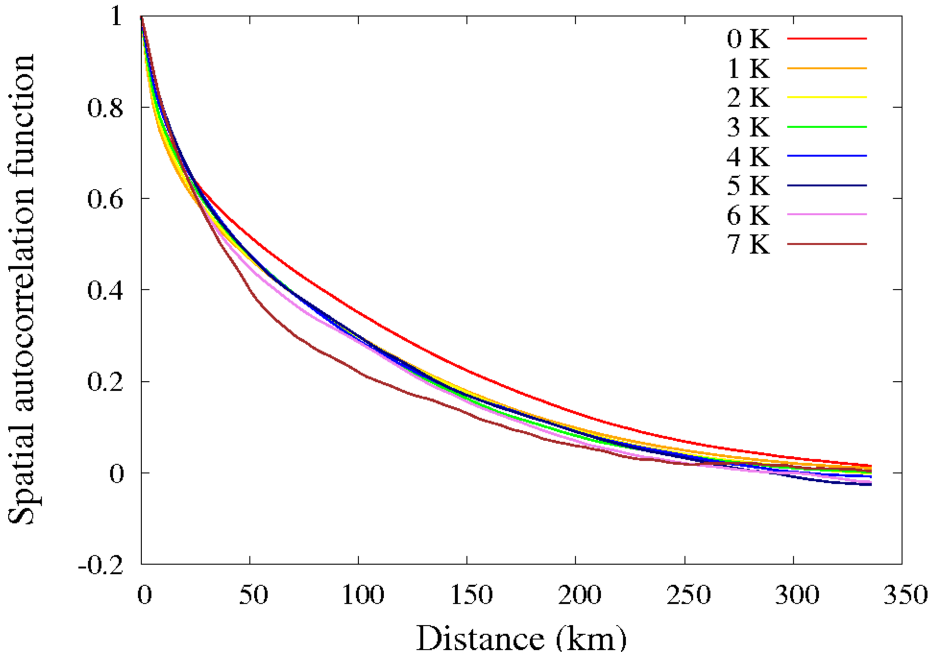
January 2013, x, y

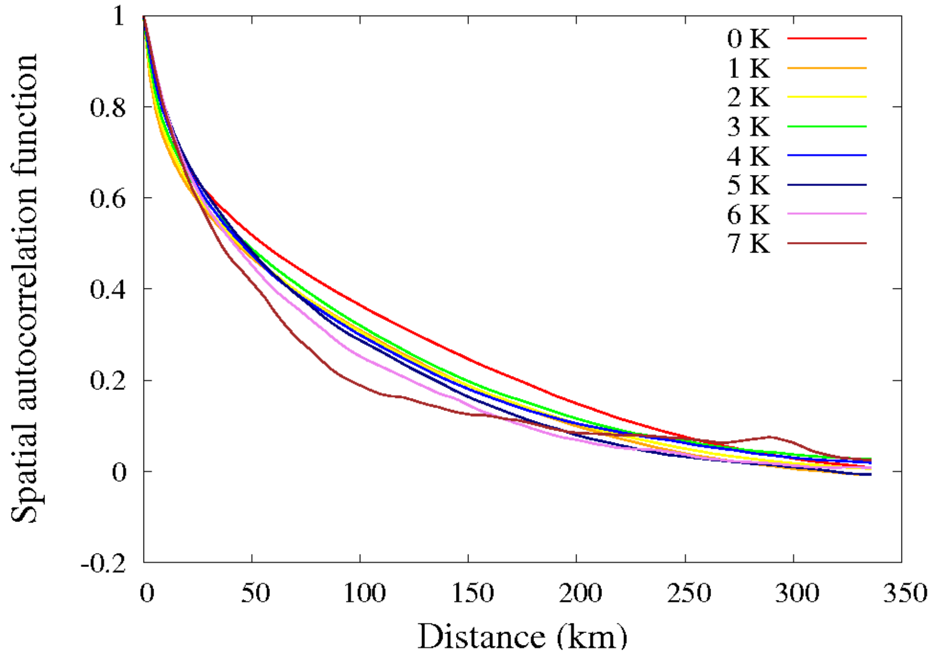
lambda



January 2013, 5 km, x, y

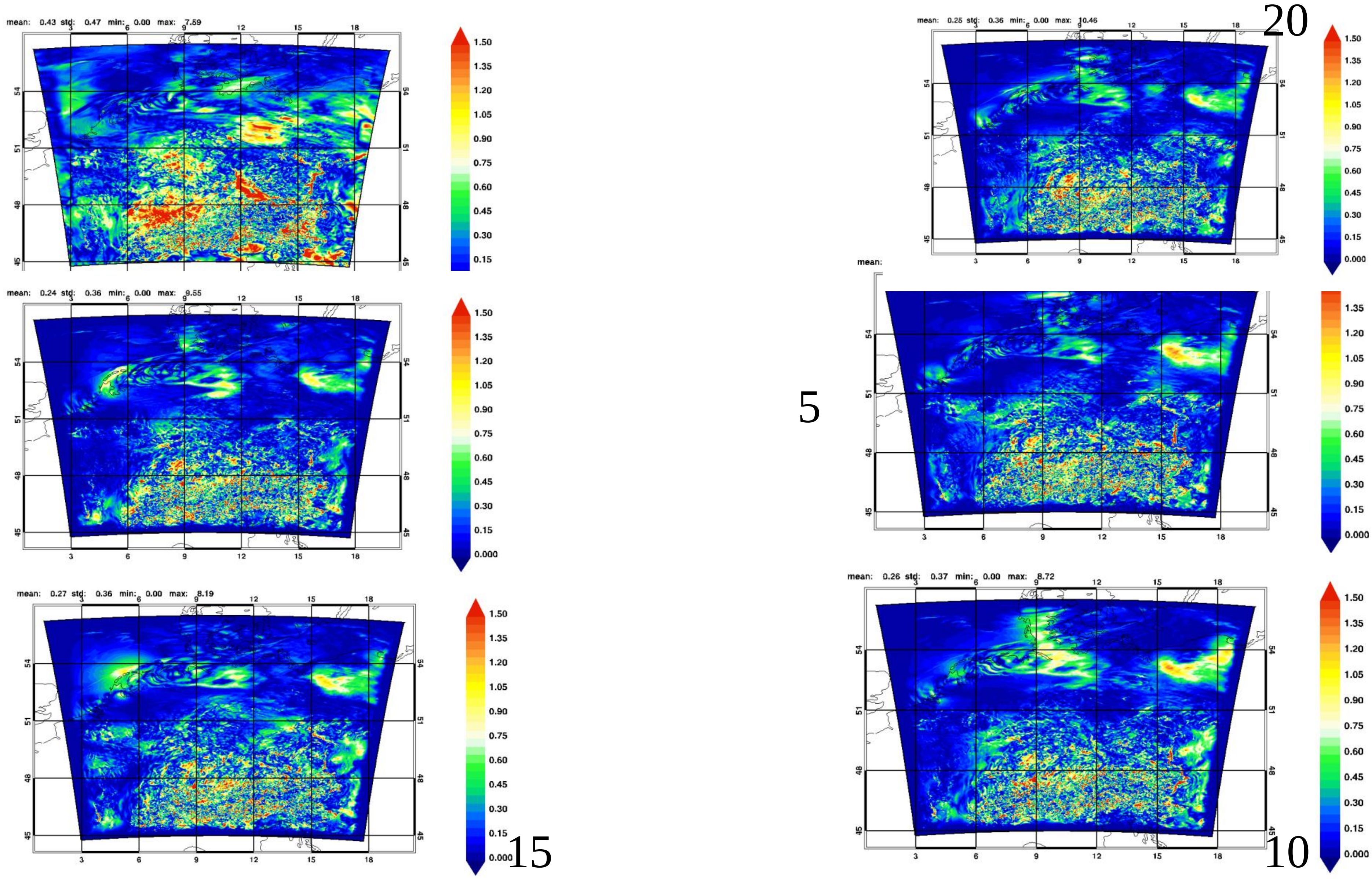
lambda



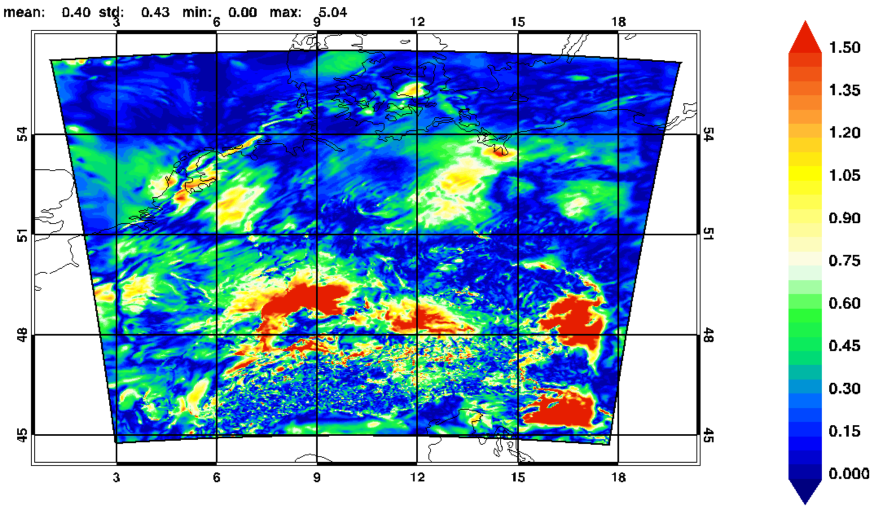


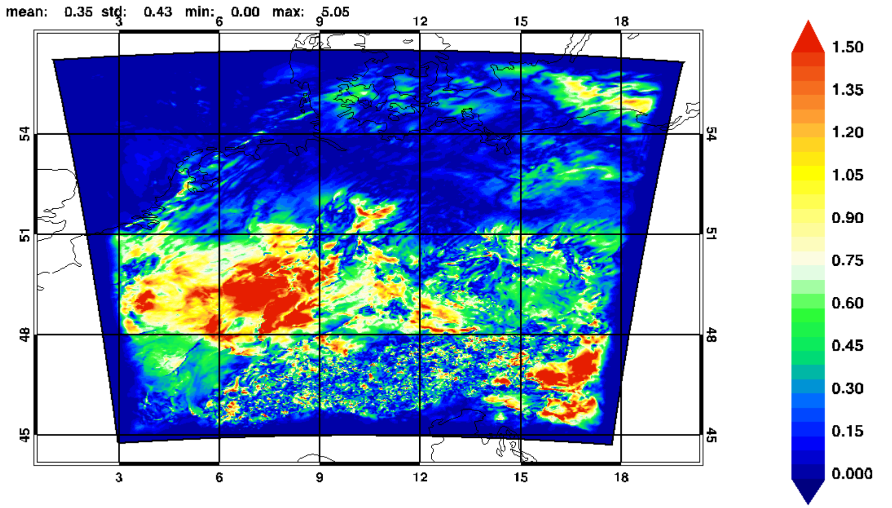
July 2013, surface, x, y

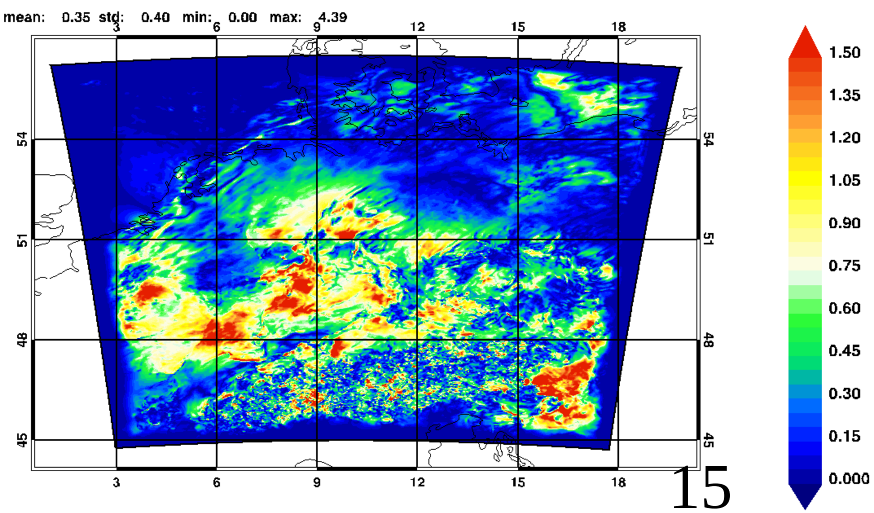
08-18 # Experiments with 3d COSMO



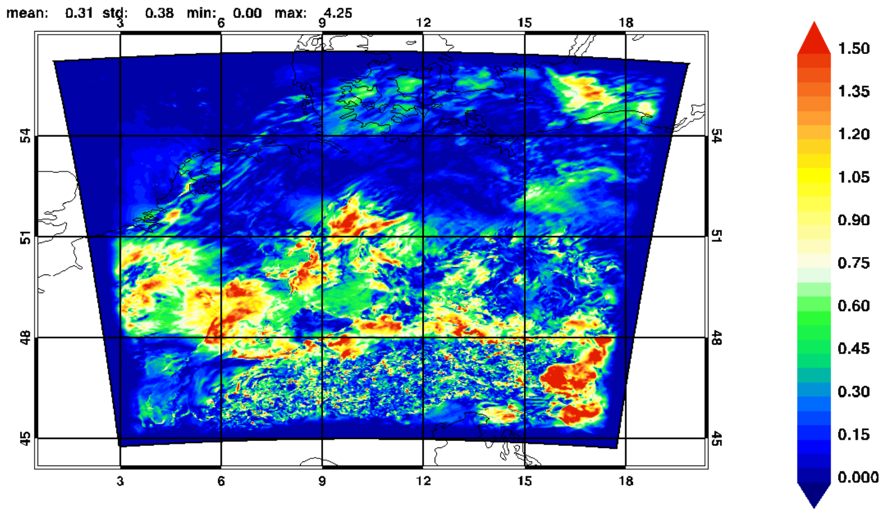
Experiments with 3d COSMO





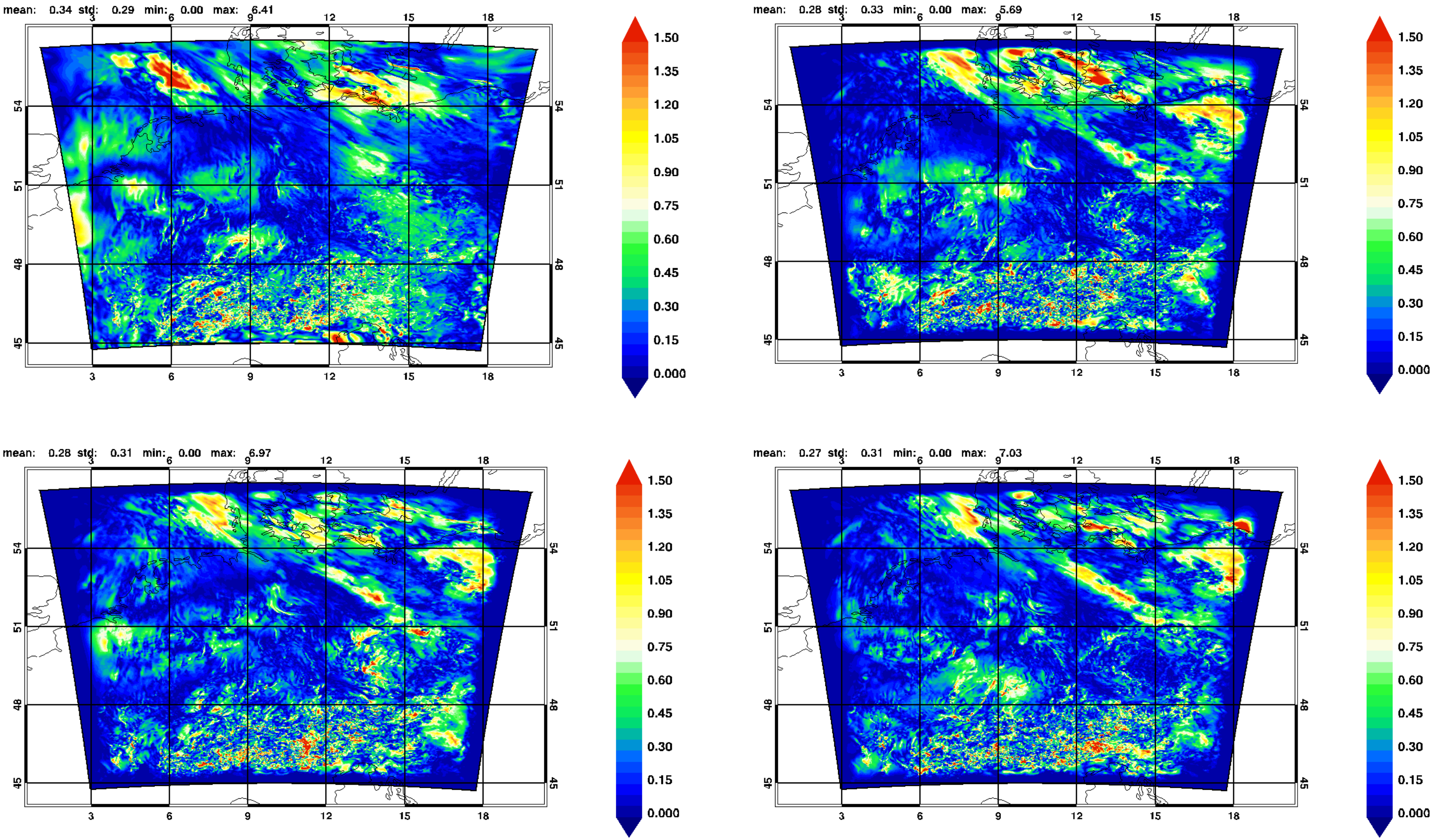


15



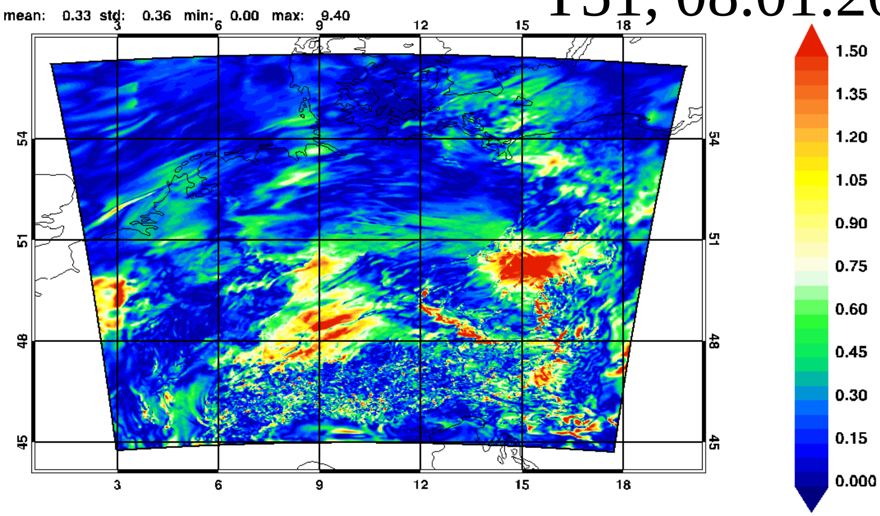
10

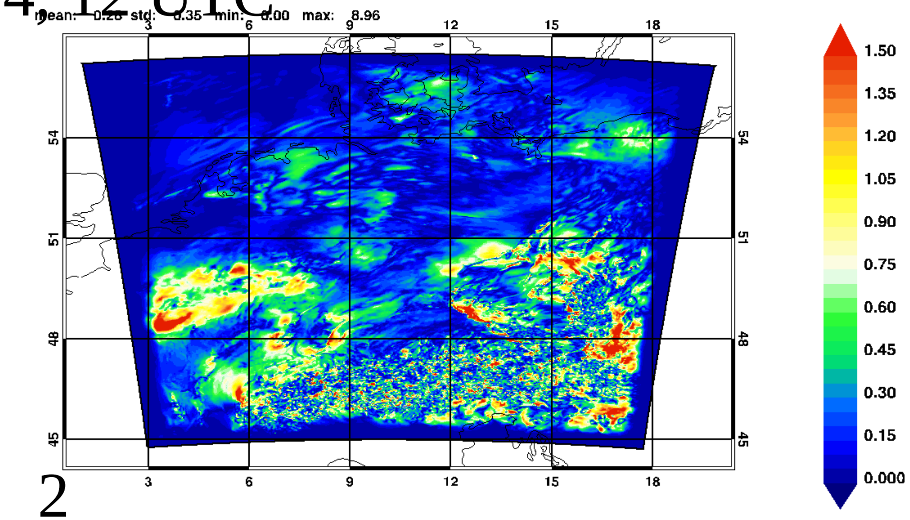
15-21 Experiments with 3d COSMO



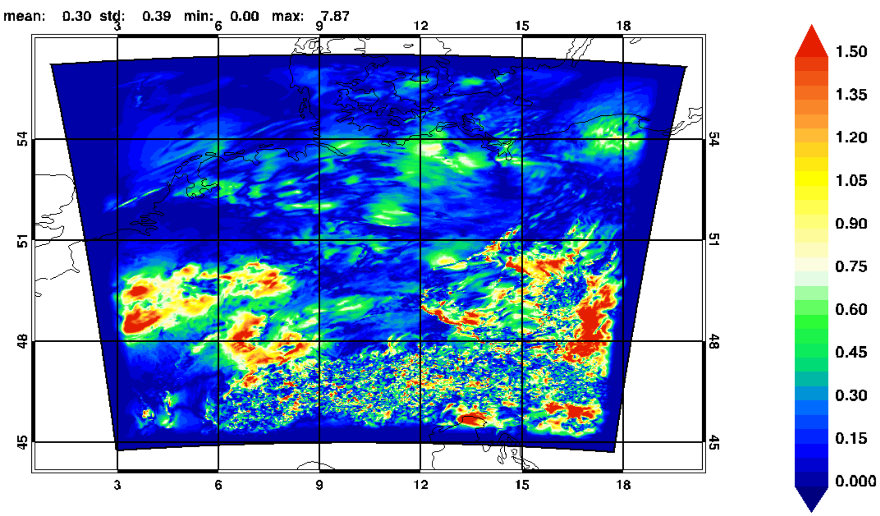
Experiments with 3d COSMO

T51, 08.01.2014, 12 UTC

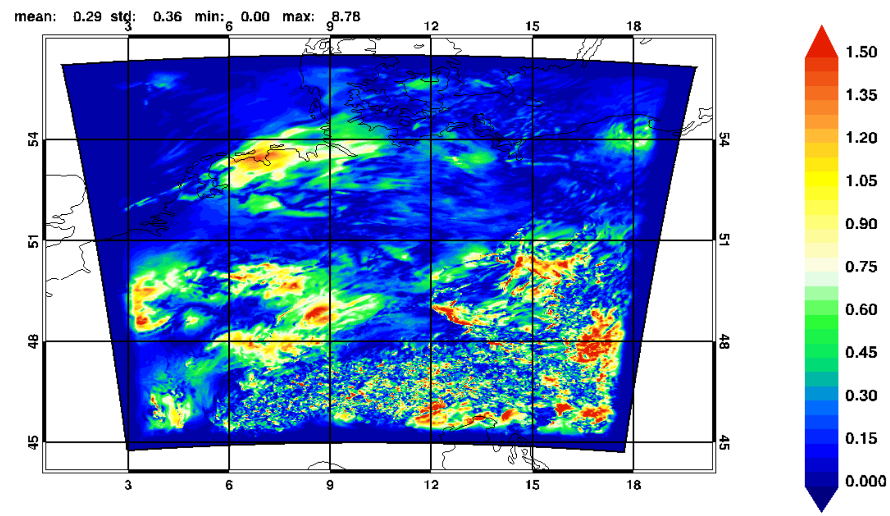




2



18

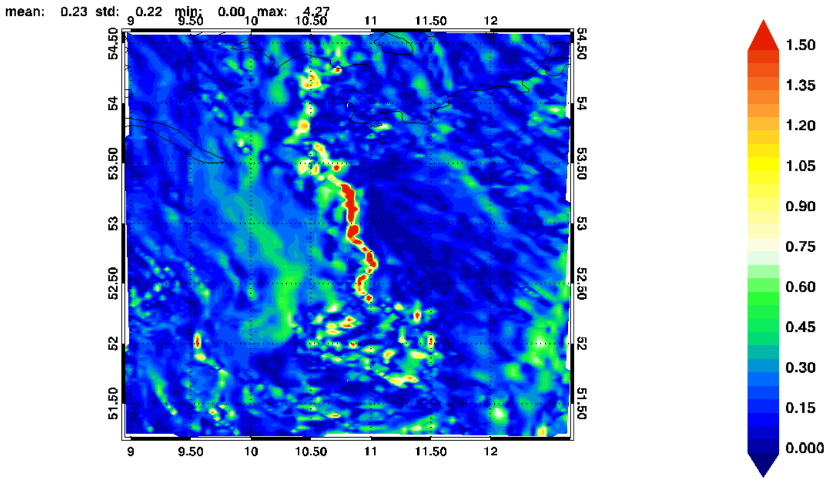


11

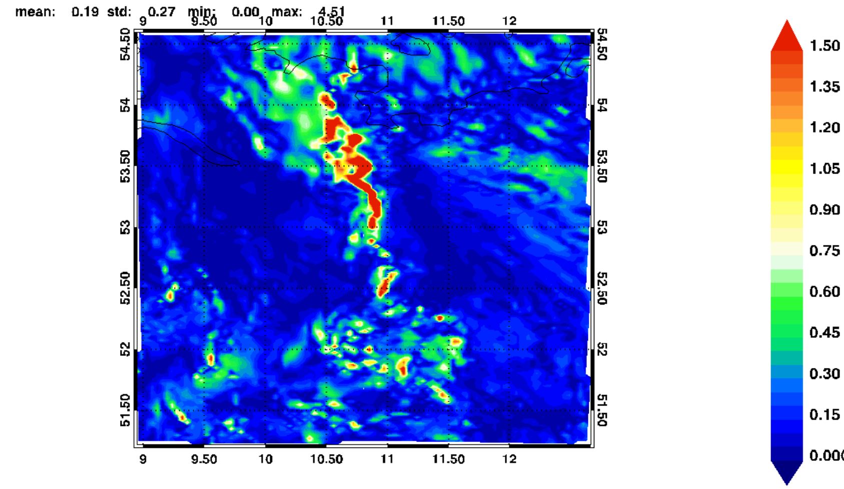
Experiments with 3d COSMO

T51, 01.01.2014, 03 UTC

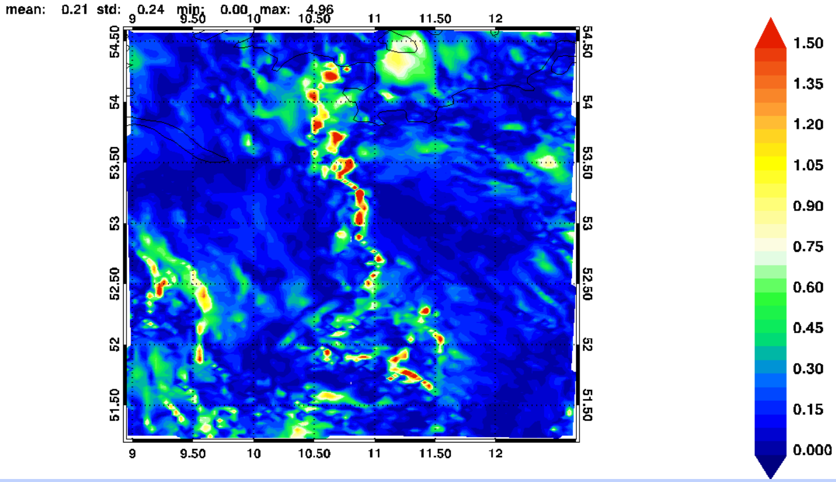
3 h for – ana



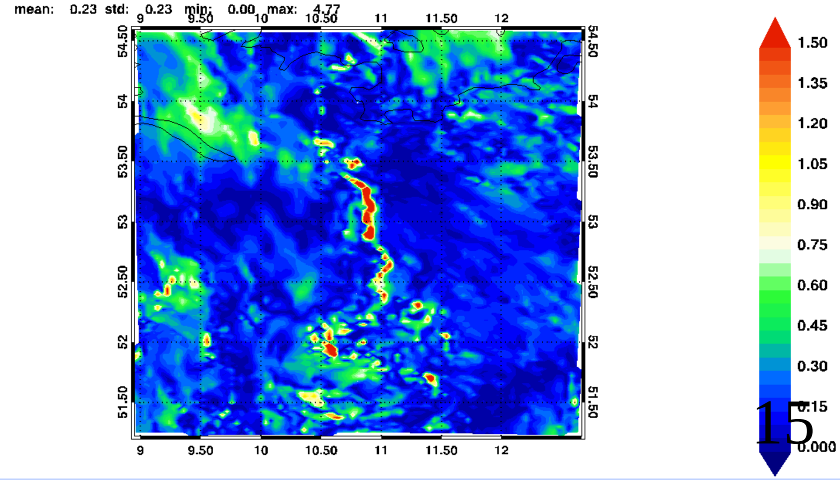
member 1 – ctrl forecast



member 5 – ctrl forecast



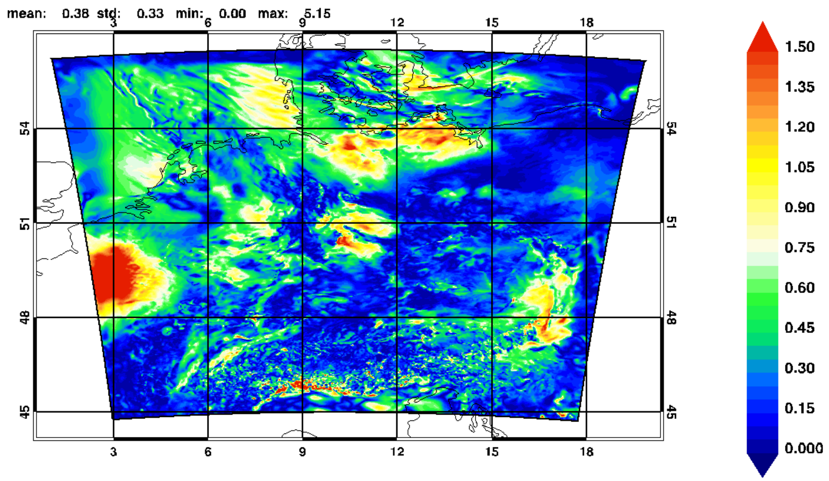
member 15 – ctrl forecast



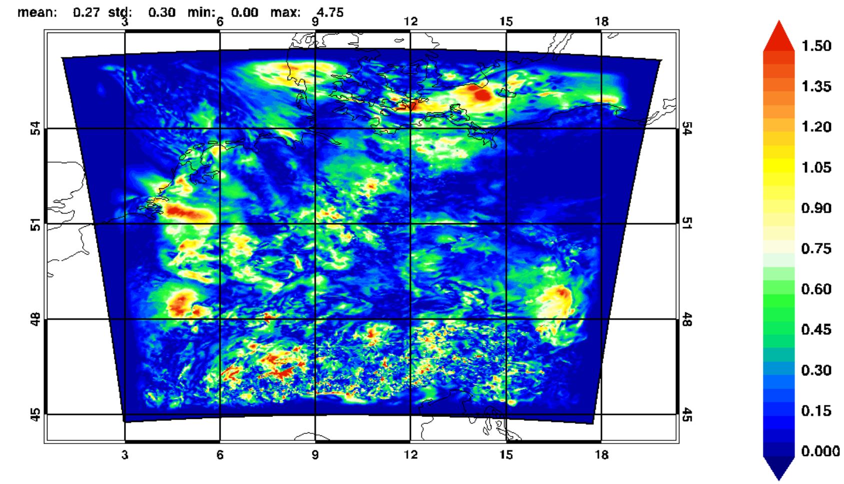
Experiments with 3d COSMO

T51, 14.01.2014, 00 UTC

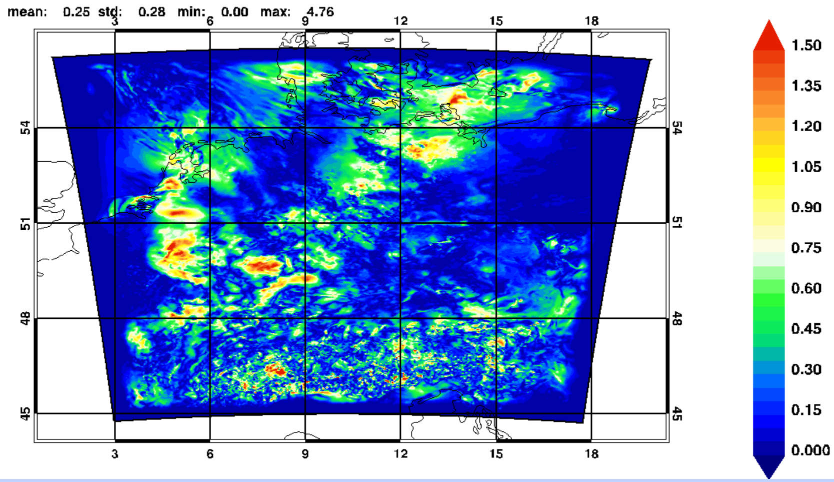
3 h for – ana



member 1 – ctrl forecast



member 7 – ctrl forecast



member 12 – ctrl forecast