Efficient uncertainty estimation of atmospheric chemical forecast due to model parameters

Annika Vogel[1,2]* (av@eurad.uni-koeln.de) and **Hendrik Elbern**[1]

[1]Rhenisch Institute for Environmental Research (RIU) at the University of Cologne [2]IEK-8, FZ Jülich *now also at: Institute for Geophysics and Meteorology, University of Cologne

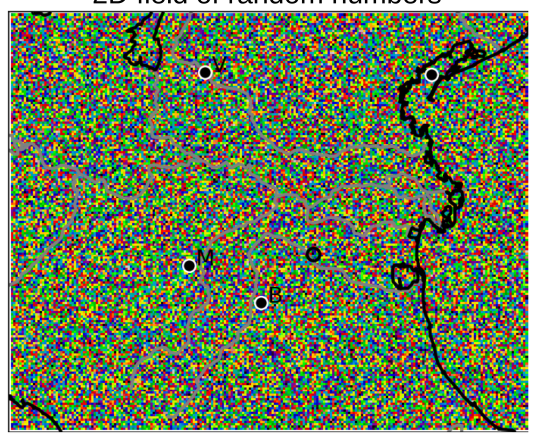
Introduction:

Forecasts of atmospheric chemical composition rely highly on various poorly-known model parameters, e.g. representing emission or deposition processes. However, a feasible estimation of resulting uncertainties by a limited ensemble of forecasts has to deal with the high-dimensionality of the system.

A new algorithm is presented to efficiently estimate forecast uncertainties caused by leading uncertainties in model parameters.

The algorithm is based on the idea, that the dynamical system induces multi-variational coupling of model states and uncertainties. For the application to biogenic emissions, case-dependent uncertainties are considered in form of sensitivities to local atmospheric and terrestrial conditions.

2D field of random numbers



ALGORITHM

Requirements:

1) Efficient calculation of sensitivities
 - consideration of multiple uncertainties, low computational effort for calculation of sensitivities

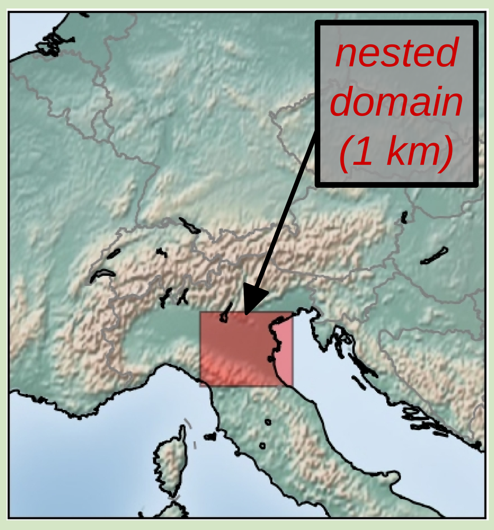
2) Efficient estimation of largest uncertainties
 - representation of largest uncertainties, suitable for high-dimensional systems $O(10^5 - 10^7)$ $\rightarrow O(10^8)$

3) Efficient physical-based ensemble generation $O(10-10^2)$
 - sampling in reduced space, physical-based uncertainty estimation

RESULTS

Modeling setup:

- ***WRF-ARW***[4] v.3.8.1 meteorological driver
- ***EURAD-IM***[1] Eulerian chemistry transport model
- ***MEGAN 2.1***[5] estimate biogenic emissions as function of vegetation distribution, leaf area index, soil moisture, solar radiation, air temperature



- Po valley
- initialization: 11.07.2012 00 UTC
- time of interest: 12.07.2012 00 – 10 UTC
- 5 biogenic gases → here: isoprene

1) Sensitivity estimation:

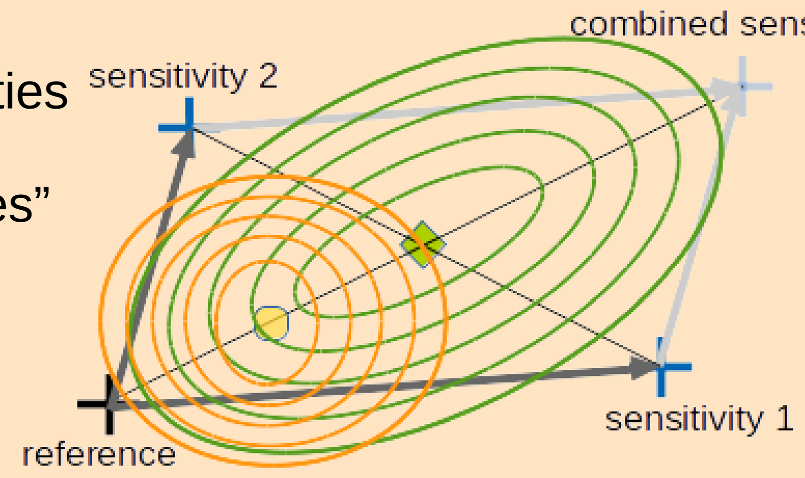
- estimate uncertainties of parameters induced by different input options
- translation to averaged ***factors***[1] w.r.t. reference values:

$$f_{sens}(j,s) := \frac{1}{T}\sum_{t=t_0}^{t_1}\frac{q(j,s,t)}{q(j_{ref},s,t)}$$

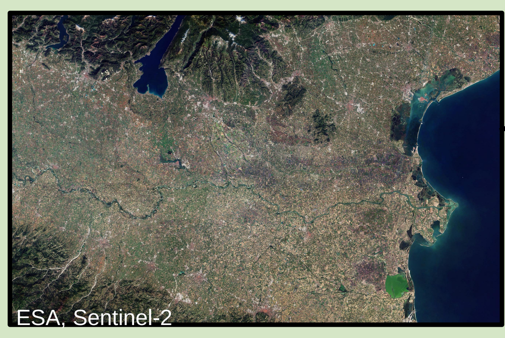
"***independent sensitivities***":
- assume tangent linearity of sensitivities
- approximate combined sensitivities by single "independent sensitivities"

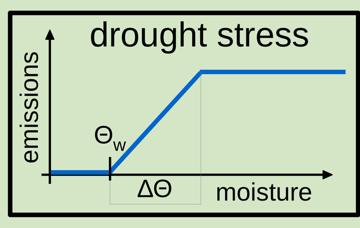
→ calculate only single sensitivities

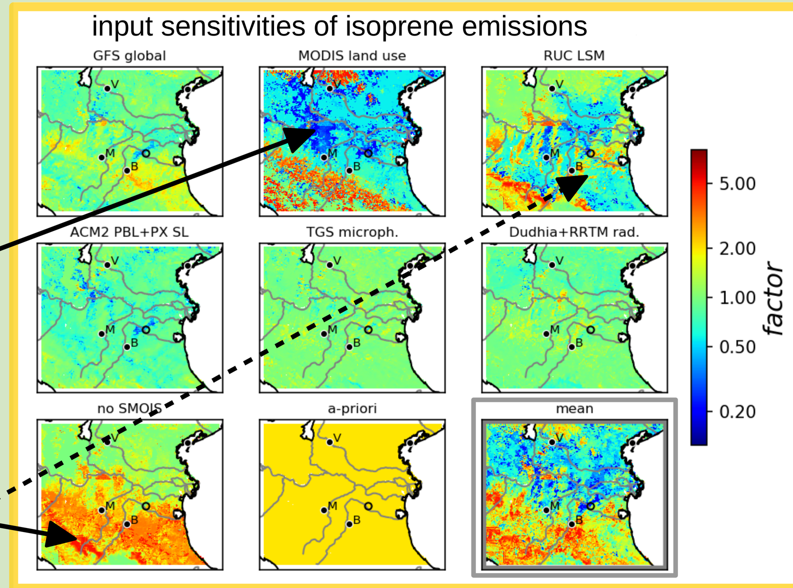
→ include additional uncertainties (with / without direction)



- independent sensitivities to global meteorology, land use information, land surface model, boundary layer-, microphysics- and radiation parameterizations
- additional uncertainties from formulation of drought stress, emission module







→ high sensitivity to vegetation and soil dryness

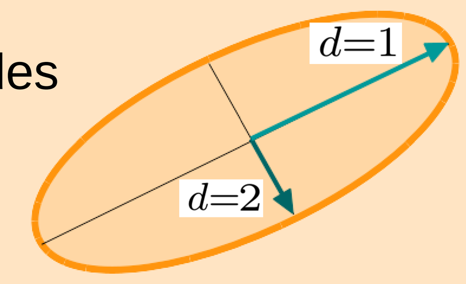
2) Eigenmode decomposition:

- describe uncertainties as covariance matrix
- extract dominating uncertainties as leading eigenmodes

Leading eigenvectors indicate dominant directions of uncertainties

$$\int_S C(s,s')\varphi_d(s')\,ds' = \lambda_d\varphi_d(s)$$

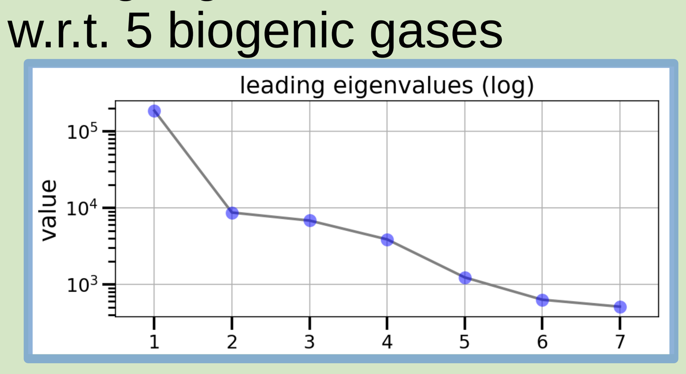
(eigenvalues = length, eigenvectors = direction)

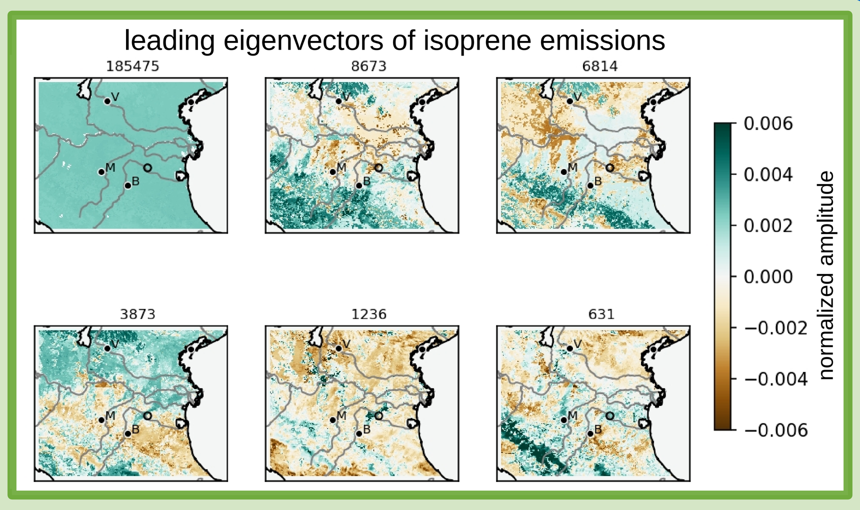


→ numerical solution with PARPACK[2] (sufficient for $C \in \mathbb{R}^{10.000 \cdot 10.000}$)

→ low number of perpendicular uncertainties for high correlations

- leading eigenmodes w.r.t. 5 biogenic gases





→ high correlations in regional domain

3) Generation of perturbations:

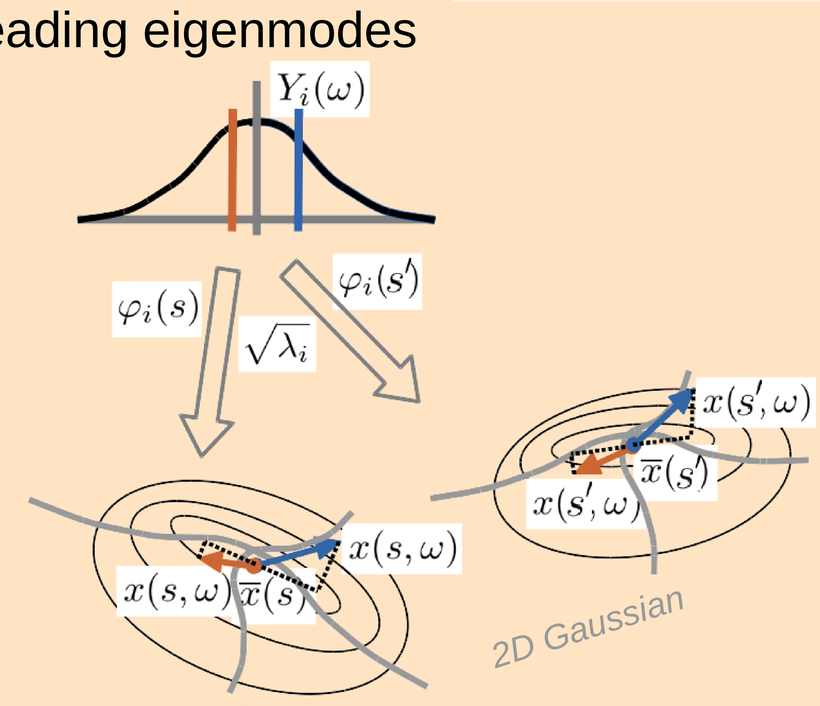
- generate perturbations according to leading eigenmodes
- ***Karhunen-Loève expansion***[3]

$$x_{KL}(\omega,s) = \sum_{d=1}^{D}\sqrt{\lambda_d}\,\varphi_d(s)\,Y_d(\omega)$$

- optimal approximation by finite sum
- generation of different members from global sampling

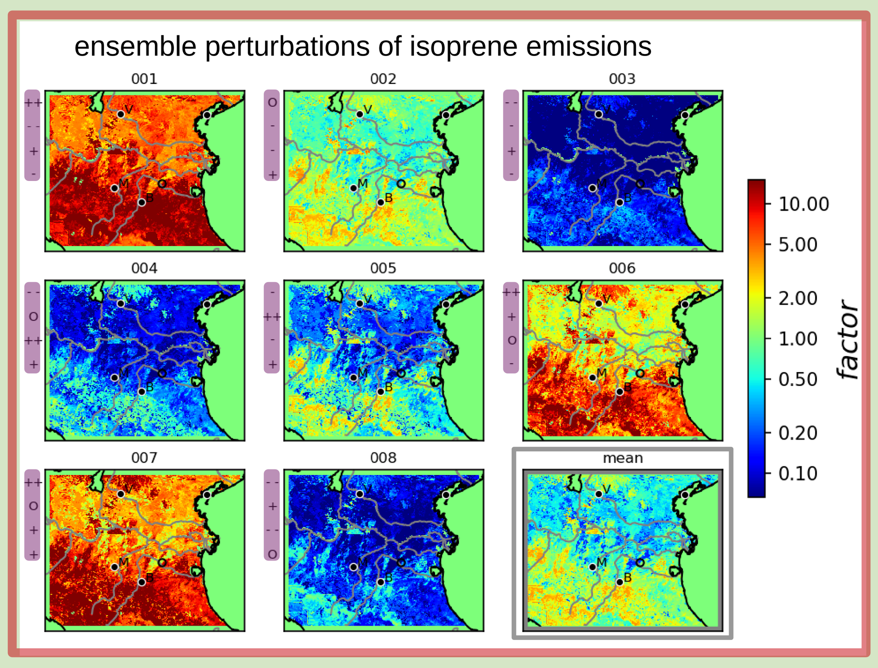
$$q(\omega,s,t) = q(j_{ref},s,t)\cdot f_{KL}(\omega,s)$$

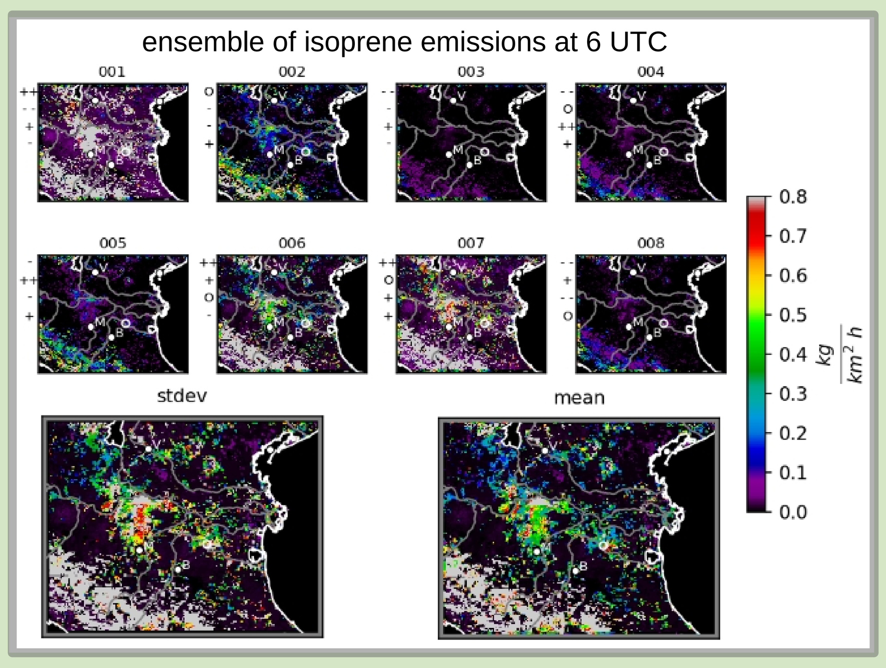
with $f_{KL}(\omega,s) = exp\left[x_{KL}(\omega,s) + \mu(s)\right]$



→ reduce degrees of freedom for random sampling to directions of largest uncertainties according to covariances

- perturbation factors and resulting emissions for 8 ensemble members





→ ensemble spread ≥ mean emissions

Conclusions:

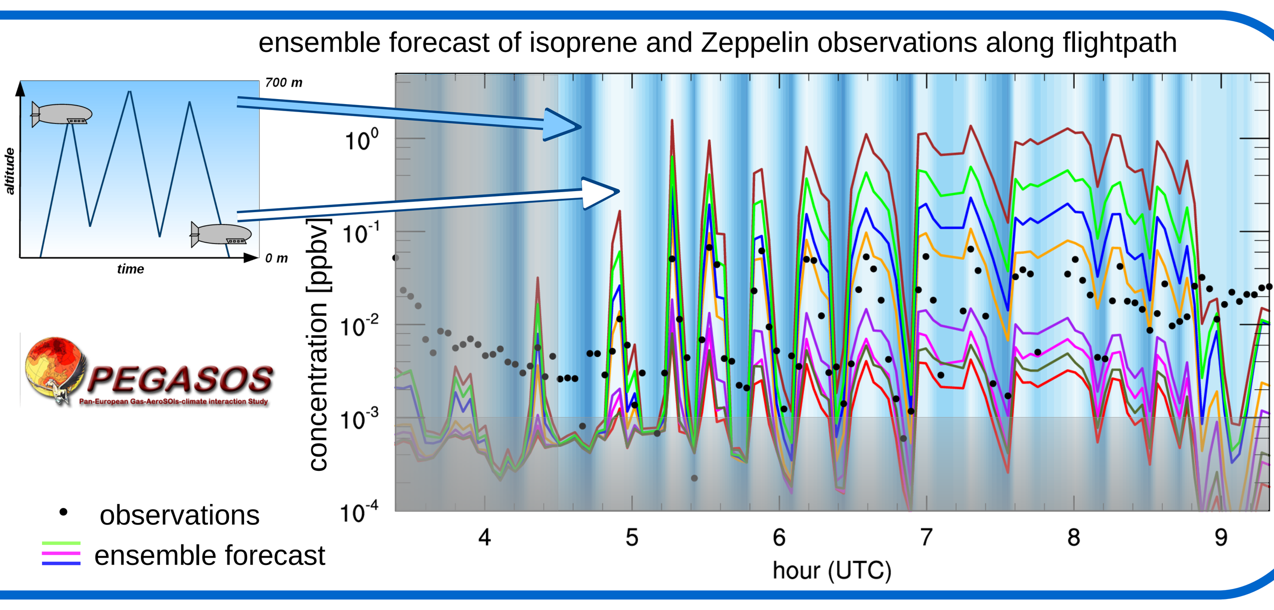
- combination of different kinds of uncertainties
- representation of largest uncertainties with low ensemble size
- high sensitivity to generation of covariances

sufficient estimation of forecast uncertainties

- high sensitivity of vegetation and soil dryness
- high cross-correlations between gases
- ensemble spread ≥ mean emissions
- high uncertainty confirmed by observations

predictability of biogenic gases limited to regions, not forecast time

ensemble forecast of isoprene and Zeppelin observations along flightpath



[1] Elbern et al. (2007). Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmospheric Chemistry and Physics
[2] Lehoucq et al. (1997) .Arpack users guide: Solution of large scale eigenvalue problems by implicitly restarted arnoldi methods
[3] Xiu, D. (2010). Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press
[4] Skamarock et al. (2008). A description of the advanced research WRF version 3. NCAR technical note.
[5] Guenther et al. (2012). The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1). Geoscientific Model Development

for further details see PhD thesis A.Vogel, University of Cologne, 2020

The authors gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC).