ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

On influence of NWP driver and NWP-CTM interface on dispersion ensembles

M.Sofiev, P.Siljamo, M.Prank, J.Soares

Contributors:

A.B.Hansen, M.Paales, L.Veriankaite, S.O.Jimenes, A.Svetlov, A.Gernega, E.Filatova, L.Grewling, E.Yakovleva, E.Khoreva

Content

- Introduction
- Ensemble AQ modelling: panacea or an expensive toy?
- Construction of an AQ ensemble
  - Multi-model, single-model, poor-man's
  - Special type of an AQ ensemble: single-CTM-multi-NWP
- Examples of AQ ensembles
- Summary

Ensemble modelling: why?

- Atmospheric processes are stochastic
  - The smaller scale and the shorter averaging the higher uncertainty
  - small-scale processes, as well as some chemical chains of reactions can be chaotic by nature
- Deterministic models work poor at small scales, with short averages and complicated chemical chains.
  - Reason is not (well, not only) model weaknesses but rather the stochastic nature of the atmosphere
- Right form of question: probability terms
- Ways to answer the probabilistic questions
  - make probabilistic models (what about physics?)
  - run ensembles of existing deterministic model(s)

VOEIKOV MAIN GEOPHYSICAL OBSERVATORY

Statistical plume model (F. Gifford, 1959) - 2

Results:

- Concentrations in the plume are stochastic variables;
- Frequency distribution of their logarithms is :

$$p(s) = A\frac{e^{m^2/\sigma^2}}{\pi\sigma^2}e^{-s^2/\sigma^2}I_0\left(\frac{m\sqrt{2s}}{\sigma^2}\right);$$

where

$$s = -\ln(\alpha_1 C/M);$$

$$\alpha_1 = 2\pi U \overline{Y}^2;$$

$$m = \sqrt{\frac{y^2+z^2}{4\overline{Y}^2}}$$

C is the concentration; M is the emission rate; A – normalizing factor; I_0 – modified Bessel function.

$$p(C/M) = A\frac{\alpha_1 e^{m^2/\sigma^2}}{\pi\sigma^2}(\alpha_1 C/M)^{0.5/\sigma^2-1}I_0\left(\frac{m}{\sigma^2}\sqrt{-2\ln(\alpha_1 C/M)}\right)$$

FINAL FREQUENCY DISTRIBUTION OF CONCENTRATIONS IN THE PLUME

VOEIKOV MAIN GEOPHYSICAL OBSERVATORY

Consequences for validation of dispersion models - 1

- The "traditional" model validations starts with stratifying the measurements into groups (gradations) with "insignificant scatter" of governing parameters;
- Indicators of performance of dispersion models

Indicators of performance (IP)

$$FBM = \frac{<M> - <P>}{<M> + <P>};$$

$$MFB = <\frac{M-P}{M+P}>;$$

$$FAa = \Pr ob\{\frac{P}{a} < M < aP\};$$

$$NMSE = \frac{<(M-P)^2>}{<P> \cdot <M>};$$

$$Corr = \frac{<(P-<P>)(M-<M>)>}{\sqrt{<(P-<P>)^2> \cdot <(M-<M>)^2>}}.$$

P is "prediction" (deterministic); M is "measurement" (stochastic); < > - symbol of averaging.

Consequences for validation of dispersion models - 2

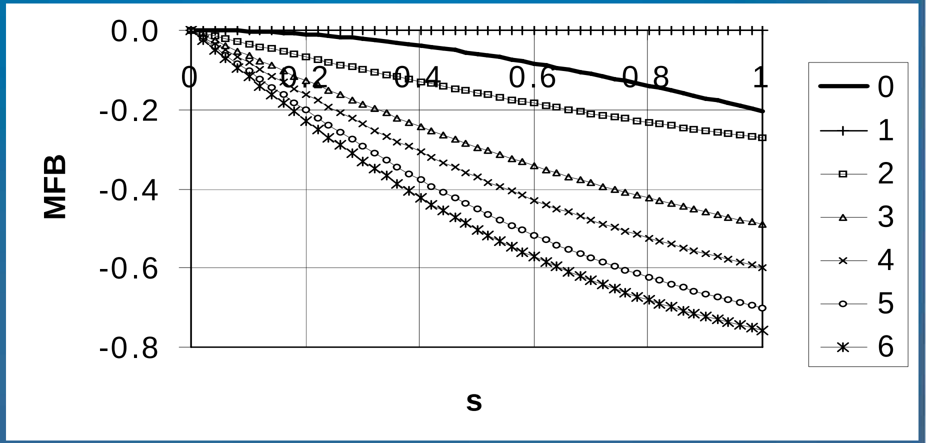
VOEIKOV
MAIN
GEOPHYSICAL
OBSERVATORY

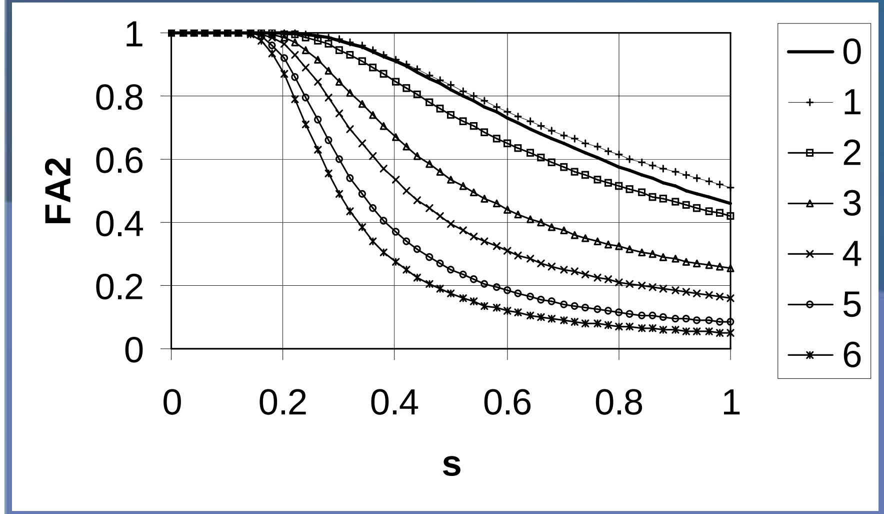
- The best ("ideal") values of IP correspond to an "ideal model" that exactly predicts for each gradation the characteristics of interest (e.g., mean value or upper percentile);
  - but only mean value can be reproduced exactly and only if the model is “perfectly” tuned to predict it.

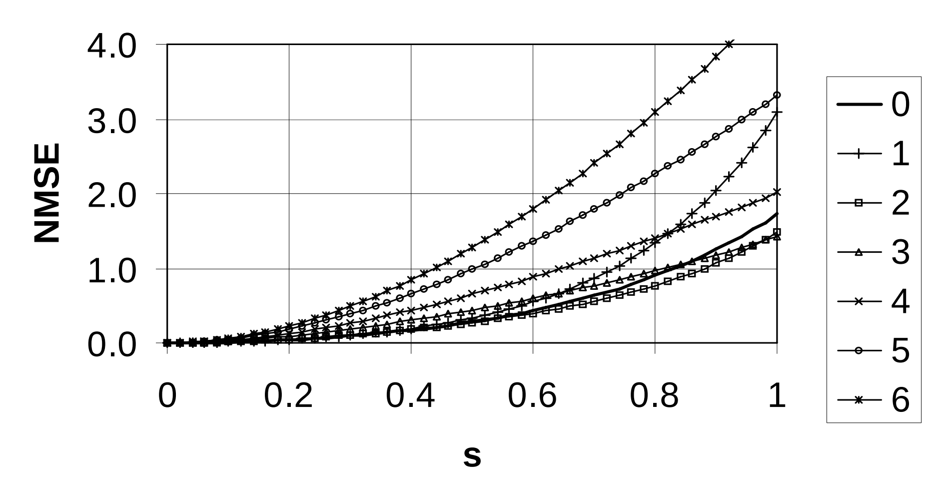
Consequences for validation of dispersion models - 3

Ind	Me-an	50 %	75 %	90 %	95 %	98 %	99 %
No	0	1	2	3	4	5	6

PDF of Kincaid data correspond to s varying between 0.6 and 1.2







Constructing the ensemble. 1

- Single-model multi-setup ensemble
  - One deterministic model
  - Input forcing, initial and/or boundary conditions are perturbed in a “reasonable” way or taken from several sources
  - Each perturbed set of data is computed in a normal way
  - Output datasets are considered as realizations of a stochastic process
  - Example: ECMWF ensemble weather forecast (operational !)
- Multi-model ensemble
  - Several deterministic (and/or other) models are used
  - Each model uses own input datasets and/or common set(s)
  - Output datasets are considered as realizations of a stochastic process
  - Example: EU FP5 ENSEMBLE project, NKS MetNet network, EMEP Pb-1996 model inter-comparison
- Poor-man’s “ensemble of anything“
  - In absence of computational possibilities to construct a representative ensemble, a set of ad-hoc picked members is used with a hope to get some hints on the actual uncertainty of the cases
    - Skeptics: all currently active ensembles are of that type

Constructing the ensemble. 2

- Statistical part
  - several models
  - several parameterizations of the same model (including the initial/boundary conditions)
  - several sources of input data
  - perturbations of the input data from a single source
- Deterministic part
  - Remaining part of the setup
  - The model(s) itself(themselves)
- Aggregating the ensemble: averaging, weight coefficients, …

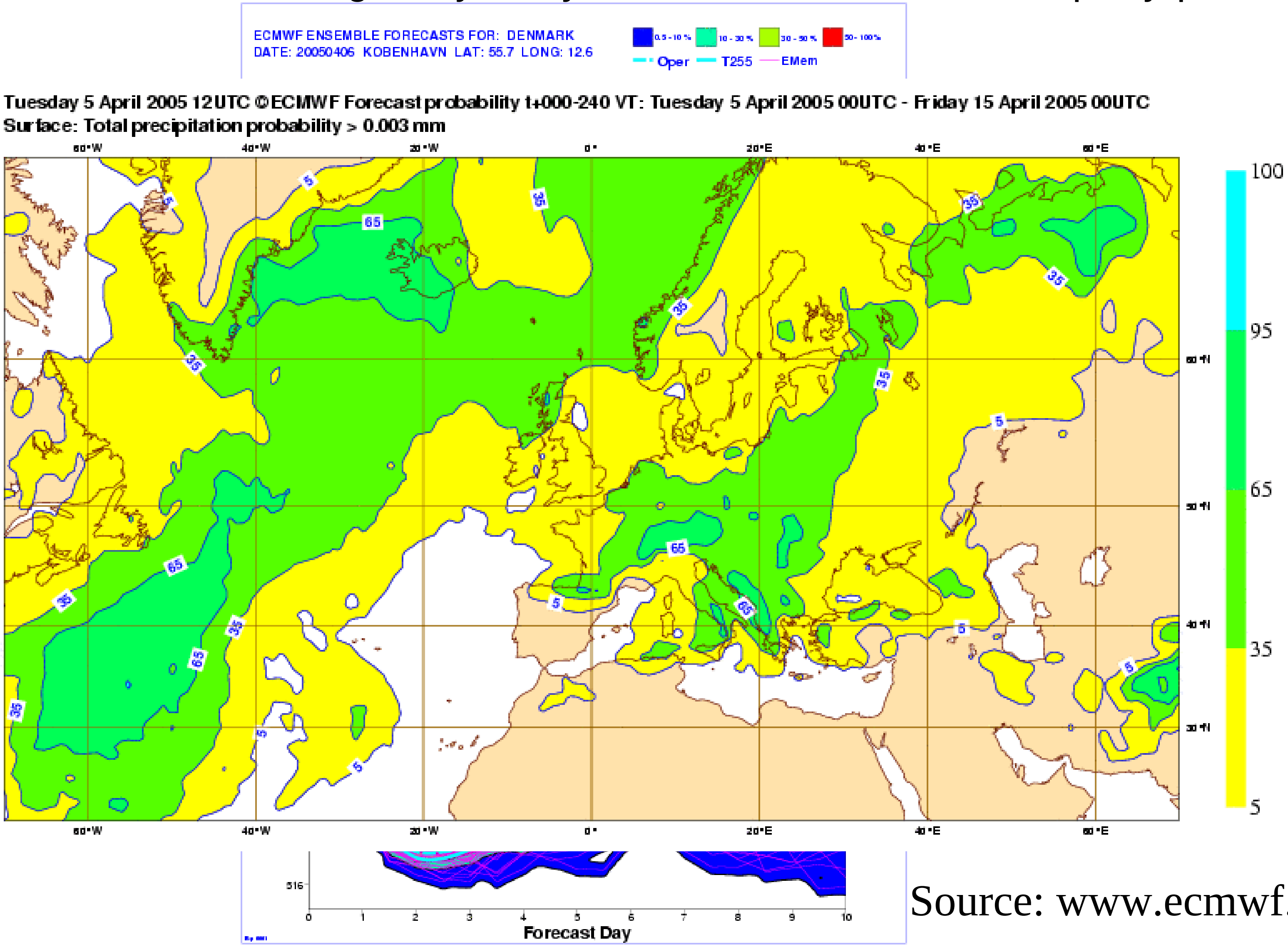
Problems of every ensemble

- The spread should be realistic: all probable situations should be reflected
- The probabilities for the specific perturbations should be estimated (or more members of the ensemble should be reflecting the more probable cases)

- Limited resources force selection of perturbations with max impact without any information about their probabilities
- A hope/belief is that the obtained set somehow represents the real uncertainties

Single-model ensemble: ECMWF

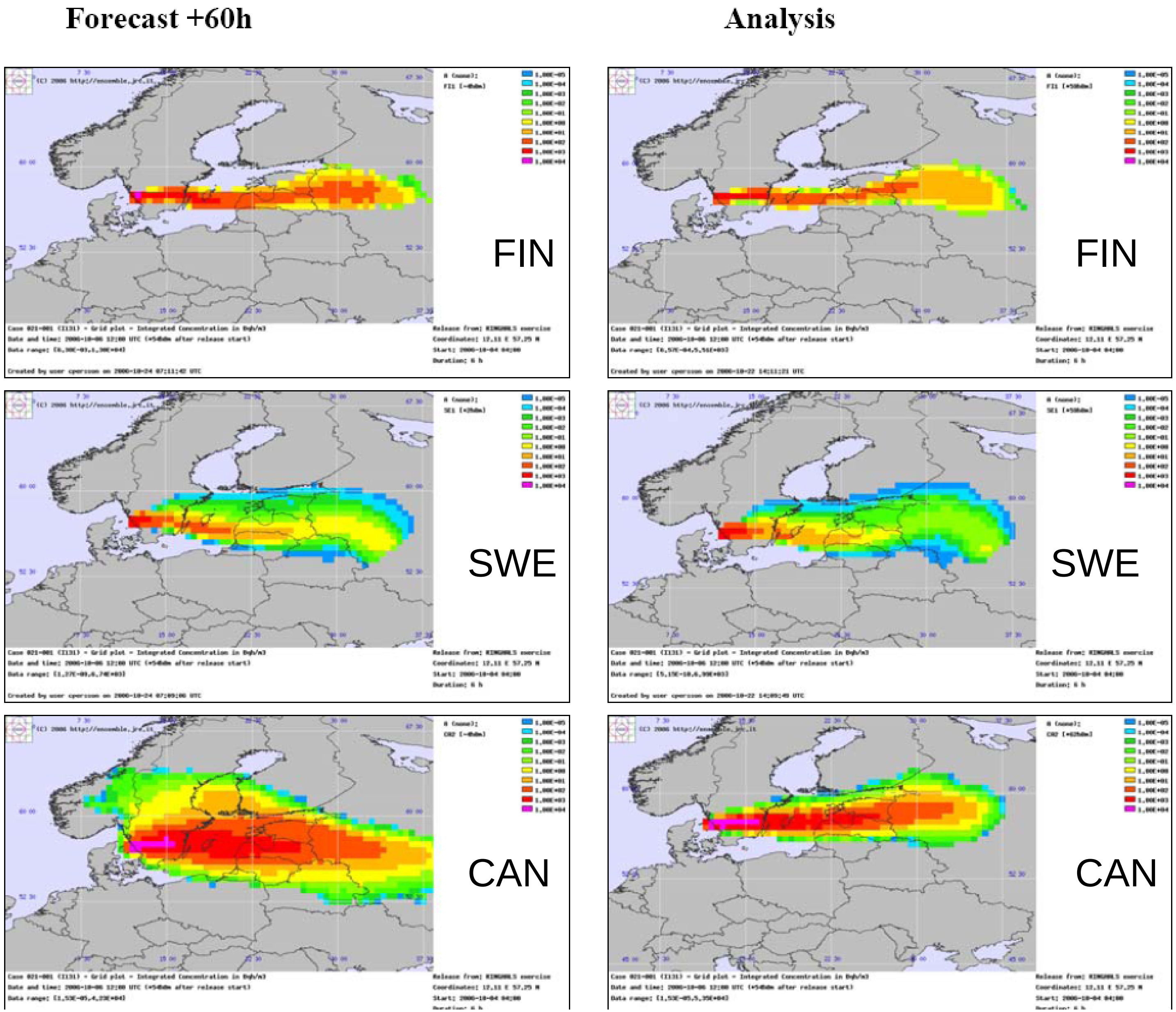
Ensemble construction: singluarity analysis, members considered equally-probable



Source: www.ecmwf.int

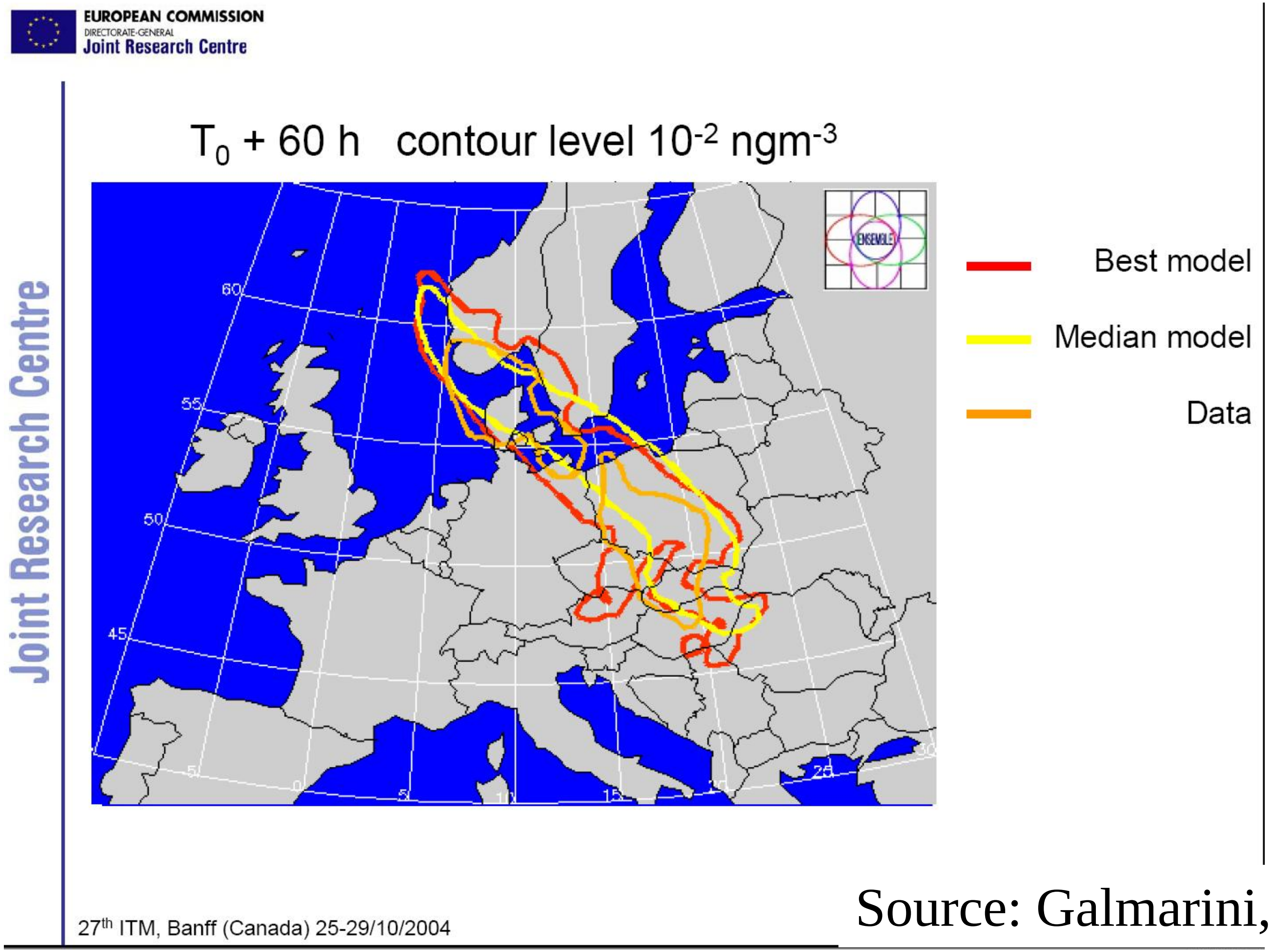
NWP forecast vs analysis (poor-man's ensemble)

- NWP +60hrs forecast vs same-CTM hindcast using analysis
- different NWPs react differently even in a simple case



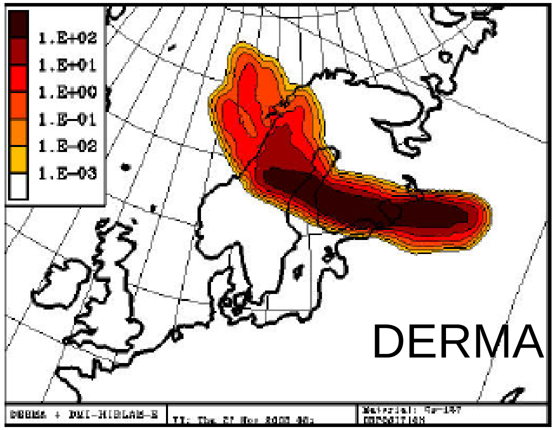
Source: report NKS-147, adapted from JRC-ENSEMBLE Web site

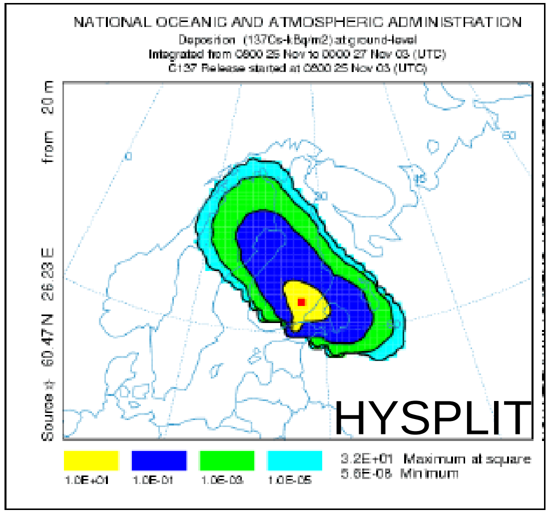
Multi-model ensemble: EU-ENSEMBLE project



Source: Galmarini, 2004

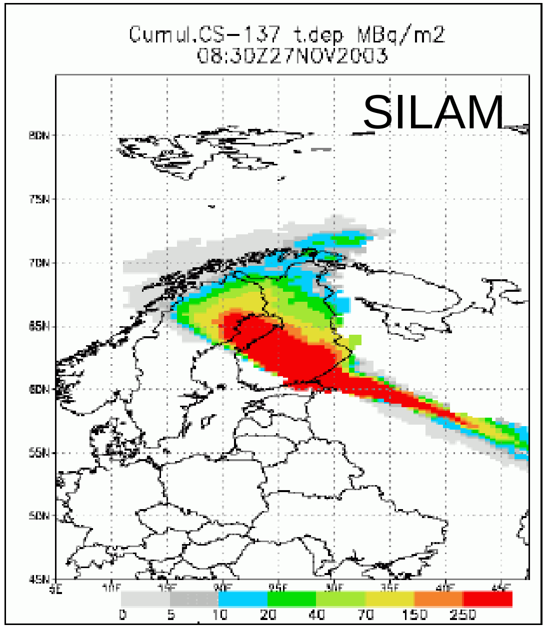
Artificial source, real meteo with front passing

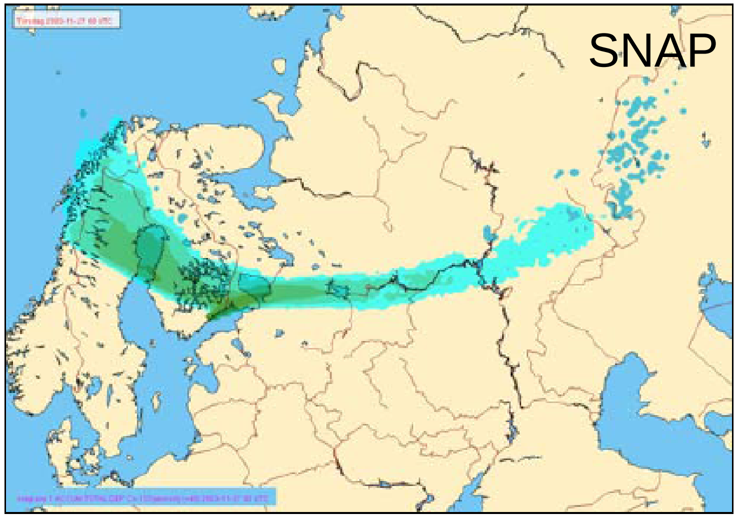


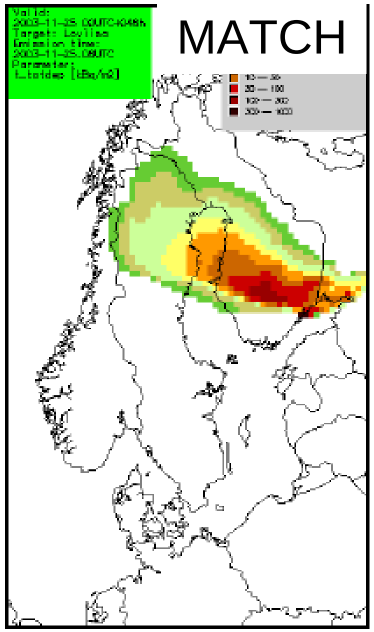


Total cumulative deposition shown, all model apart from HYSPLIT use versions of HIRLAM

Source: NKS report-147





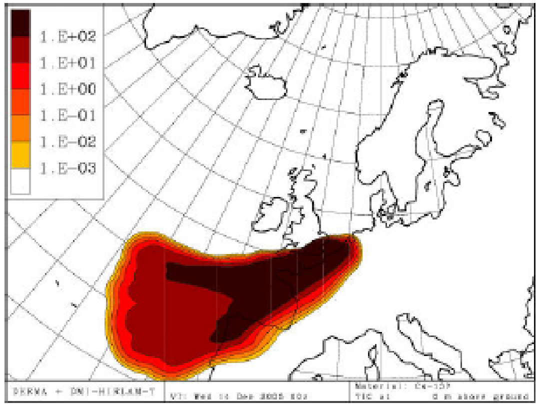


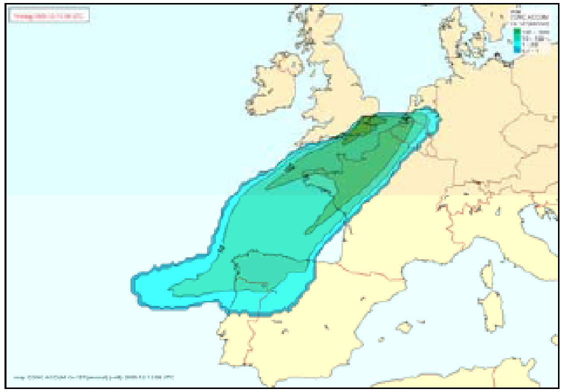
Real-life ensemble: Buncefield fire

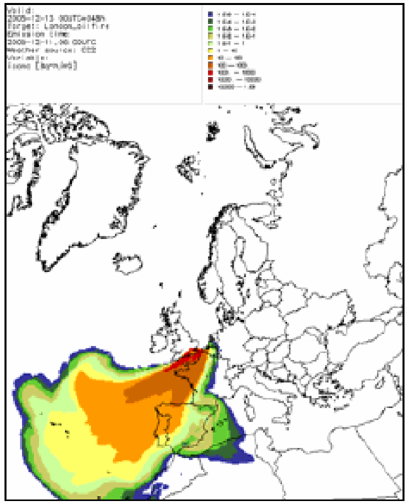
- 11.12.2005, near London, major explosion at oil refinery, (entirely demolished), burning for 4 days

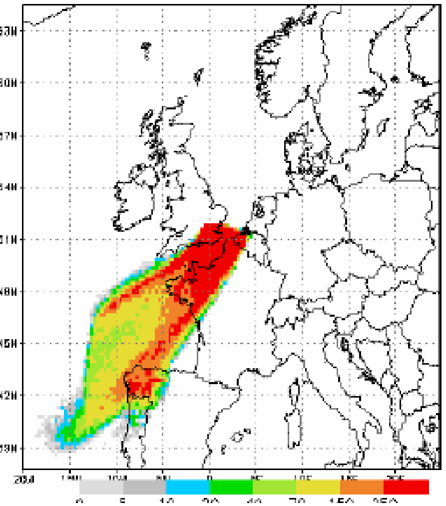
Source: http://www.buncefield-oil-fire-hemel-hempstead.wingedfeet.co.uk/

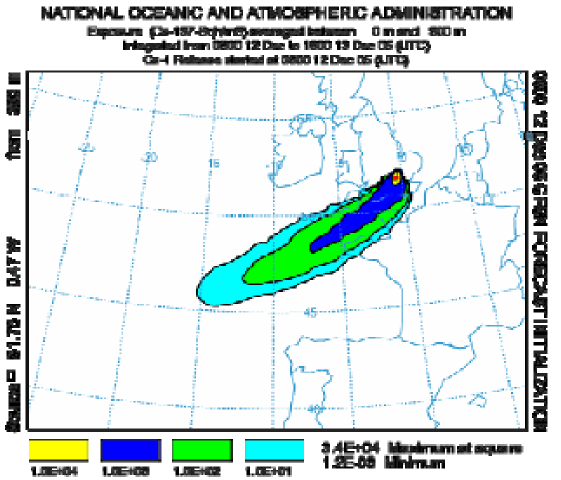
Buncefield fire: ensemble simulations

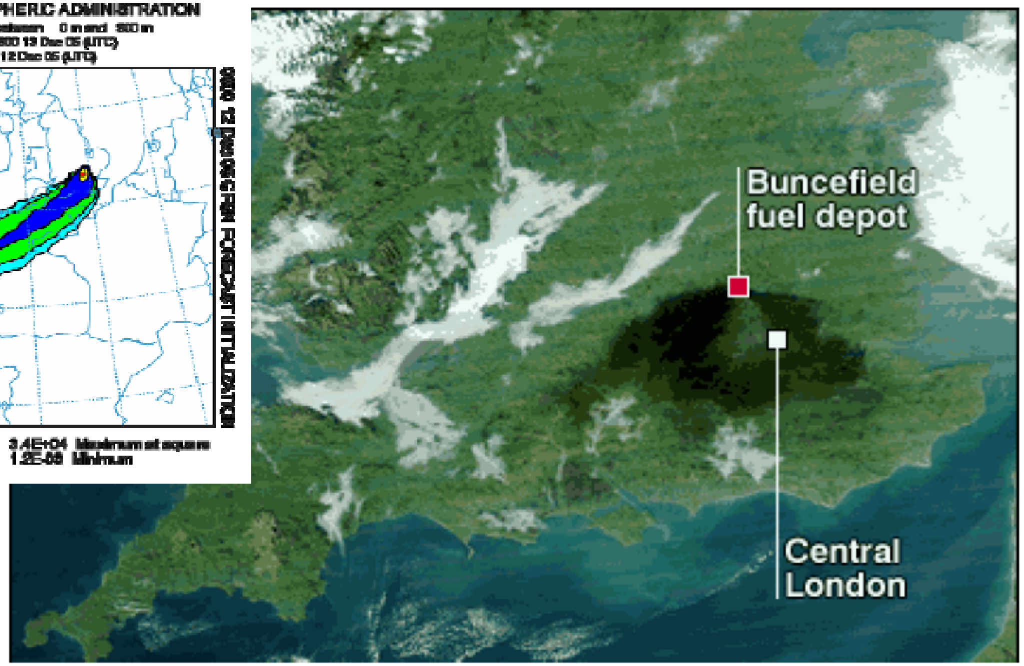












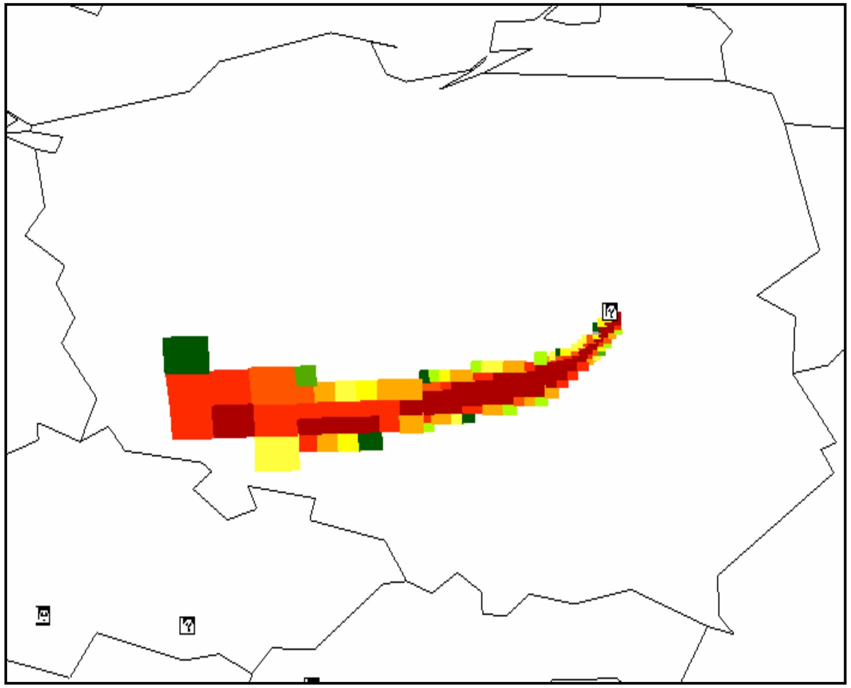
Multi-model ensemble: EMEP Pb model inter-comparison

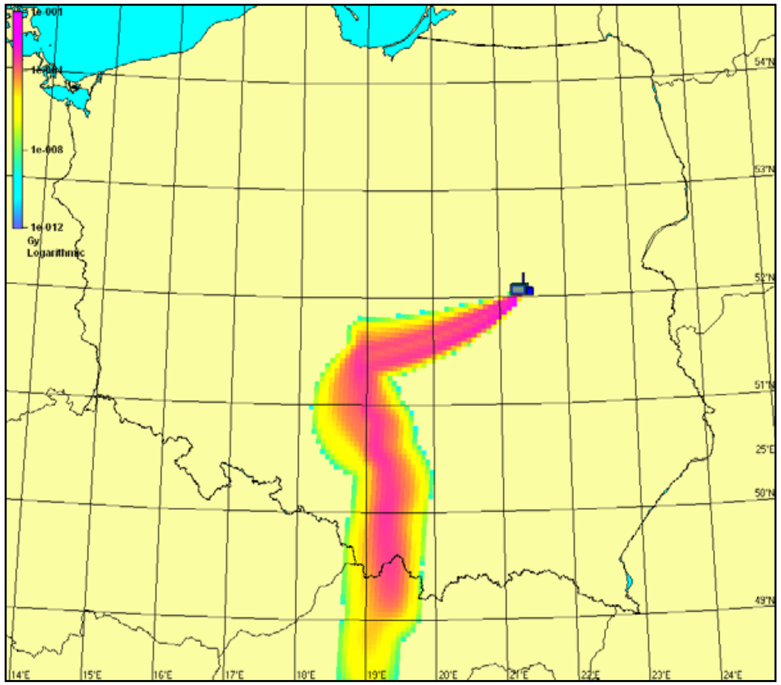
	Observed	Mdl1	Mdl2	Mdl3	Mdl4	Mdl5	Mdl6	Mdl7	**SAM**
Pb concentration in aerosol, ng / m3									
Mean	**32.2**	24.6	13.2	19.2	21.2	24.4	13.4	29.2	**21.0**
Correl		0.8	0.9	0.8	0.9	0.7	0.9	0.8	**0.9**
MLS slope		1.0	0.4	0.6	0.8	0.7	0.4	0.9	**0.7**
Pb concentration in precipitation, ug / l									
Mean	**3.5**	3.8	8.7	3.0	4.5	2.7	N/A	2.7	**3.5**
Correl		0.9	0.7	0.8	0.9	0.9	N/A	0.8	**0.99**
MLS slope		0.7	2.3	0.8	0.8	0.6	N/A	0.7	**0.7**
Pb wet deposition, mg / m2 year									
Mean	**2.7**	2.5	2.3	2.5	2.8	2.1	N/A	2.0	**2.3**
Correl		0.7	0.6	0.6	0.8	0.7	N/A	0.7	**0.7**
MLS slope		0.7	0.7	1.0	0.8	0.7	N/A	0.7	**0.7**

Source: Sofiev *et al.*, 1996

Multi-model debugging

- Source term are the same for both models
- Meteorological data are the same but:
  - Models used own meteo pre-processors



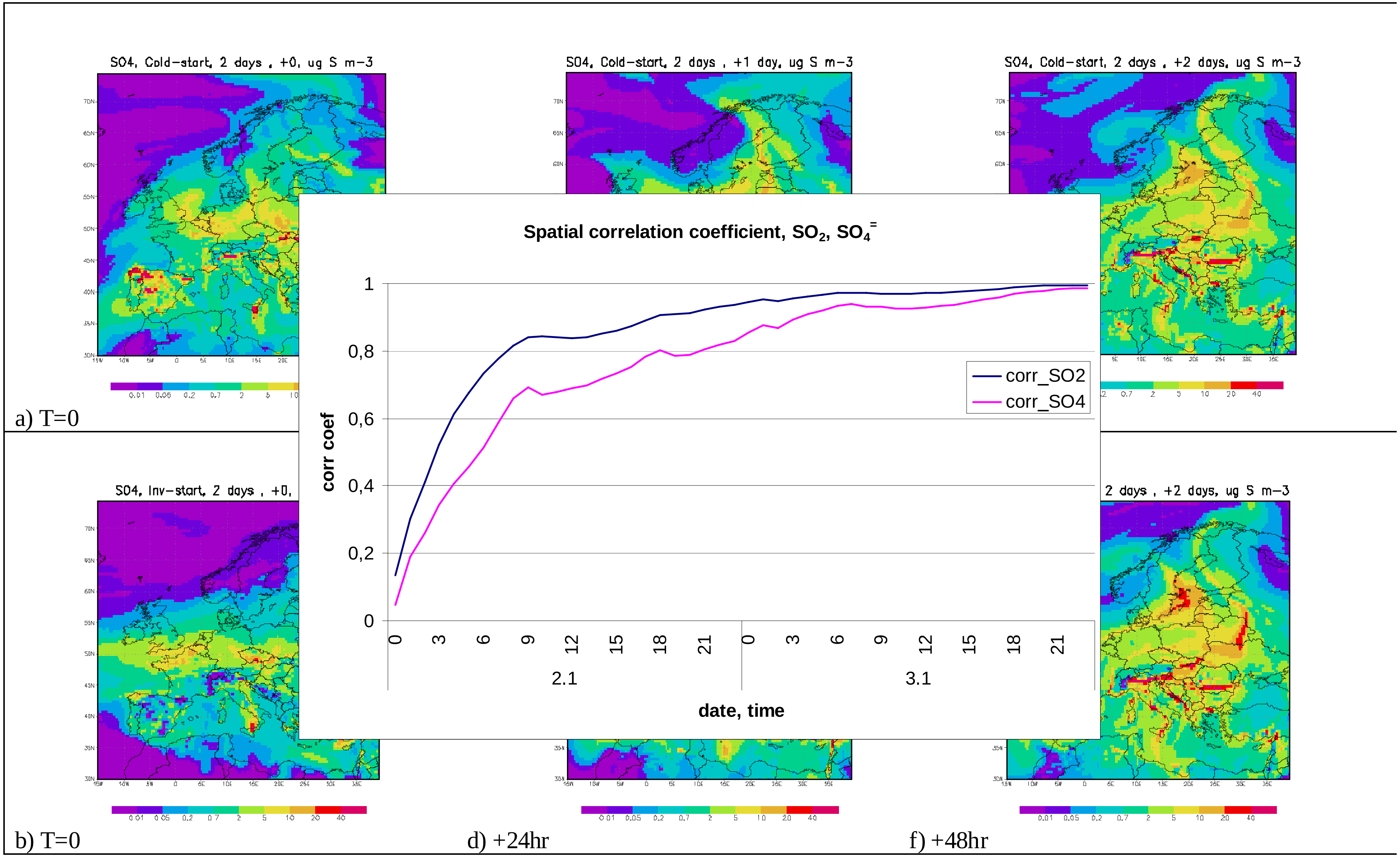


Source: Potempski, 2005

CTM specific: emission vs initial conditions.1

- For NWP: setting the starting state of the NWP model is sufficient to determine its following evolution
- For CTM: initial conditions and emission play both negligible and dominant roles depending on the time scale
  - close to start time initial conditions dominate
  - the longer the time scale the stronger the emission impact
  - characteristic time scale varies for different species and cases
- Consequences
  - perturbation of a “wrong” parameter does not generate any response from the model
  - same is true for data assimilation

CTM specific: emission vs initial conditions.2



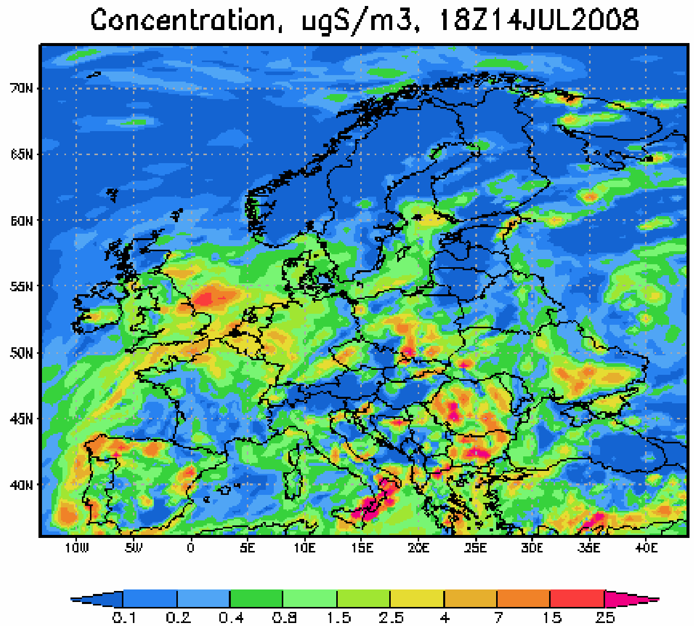
Example of a multi-NWP ensemble

- FMI AQ forecasting is based on routinely available ECMWF and HIRLAM meteorological fields
- Same SILAM setups (well, almost) allow for a poor-man's ensemble considerations
  - the most-evident places of potential bifurcations can be deducted
  - no quantitative analysis is possible

HIRLAM- vs ECMWF- based AQ forecast

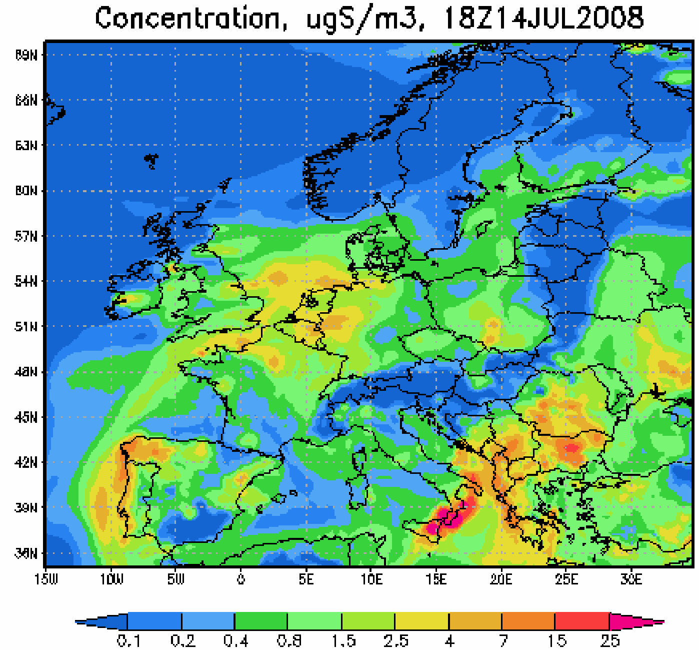
Forecast for SO2. Last analysis time: 20080714_00

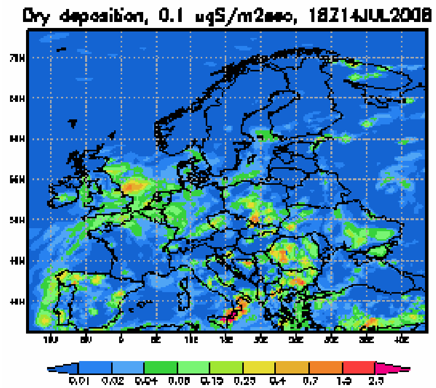
HIRLAM+SILAM v.4.0.1

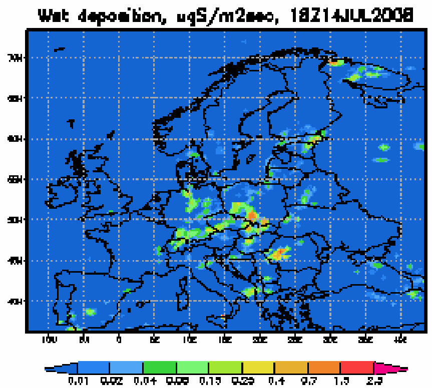


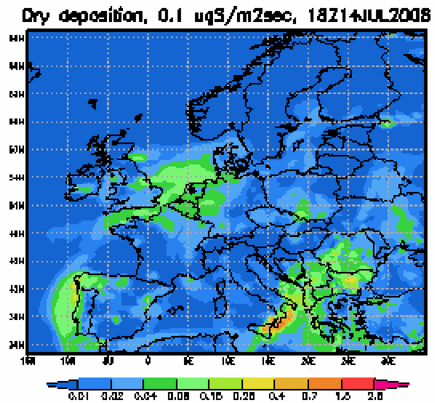
Forecast for SO2. Last analysis time: 20080714_00

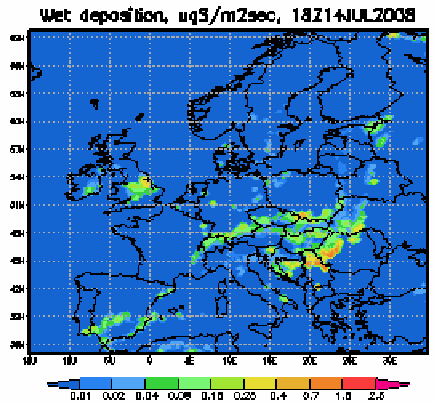
ECMWF+SILAM v.4.2





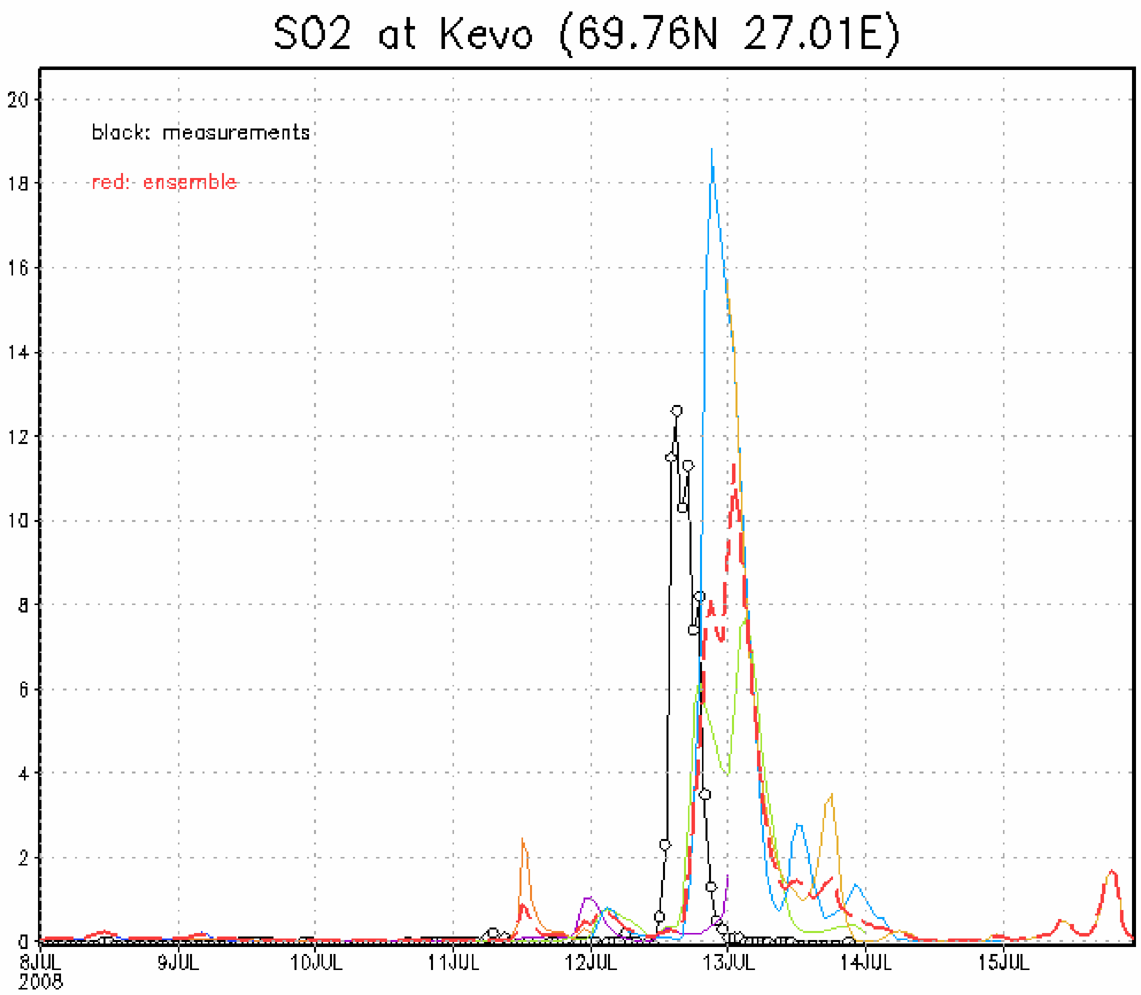






Example of single-NWP ensemble

- In … d time
  bu … st
  ho … e
- Re … asts
  wi … with
  ne … num
  un

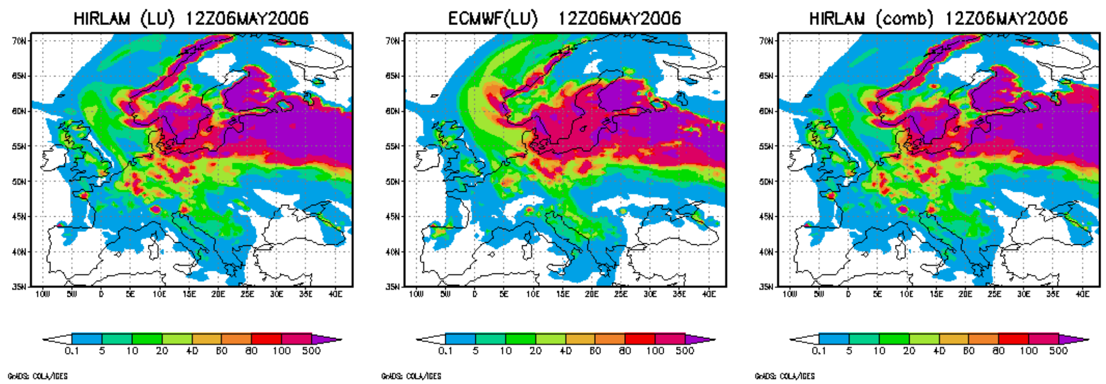


Quick lessons from the exercises: pollen .1

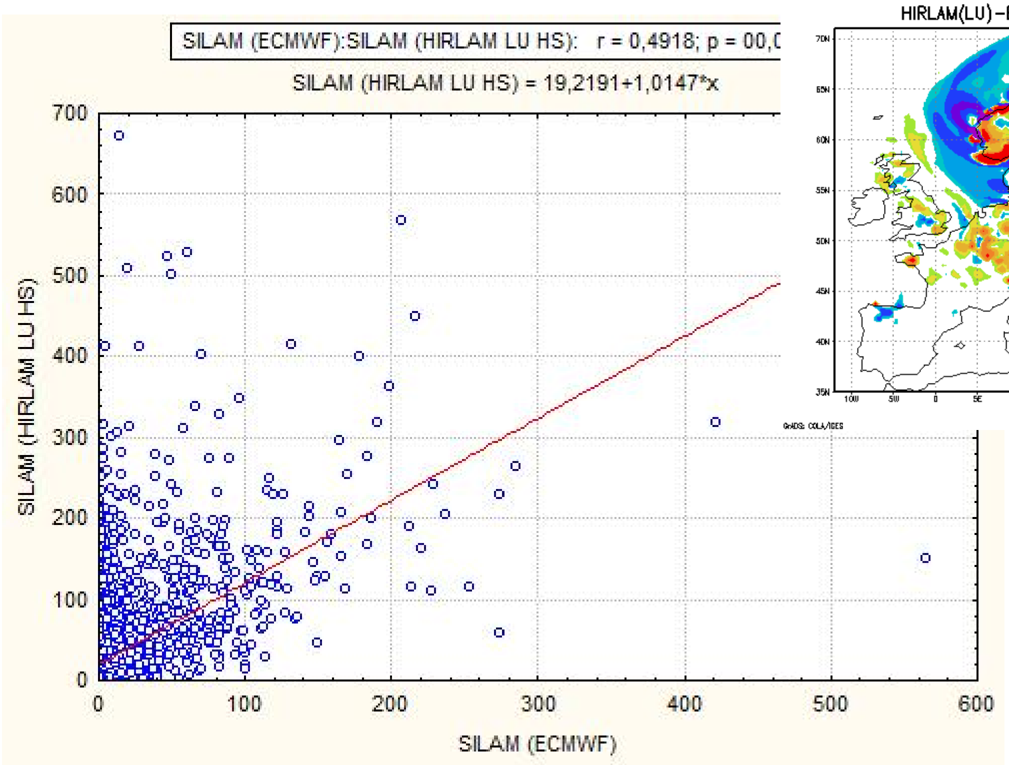
- Pollen exercise: single-model multi-NWP results can be drastically different
  - Half-a-degree bias (usually neglected in NWP model validation) for 2 months of integration means ~30 degree-days of accumulated heat sum, i.e. 25-50% of the flowering threshold
- Two scales of the problem: long- and short-term
  - Pollen season description is strongly dependent on regular bias in the input information, first of all, in temperature
    - Multi-NWP forecast helps revealing the potential problems before they turn into the incorrect season description
  - Short-term variability between the NWP drivers results in corresponding differences in the forecasts themselves
    - Treatment is similar to that of other AQ forecasts: hinting on possible bifurcations and other variability

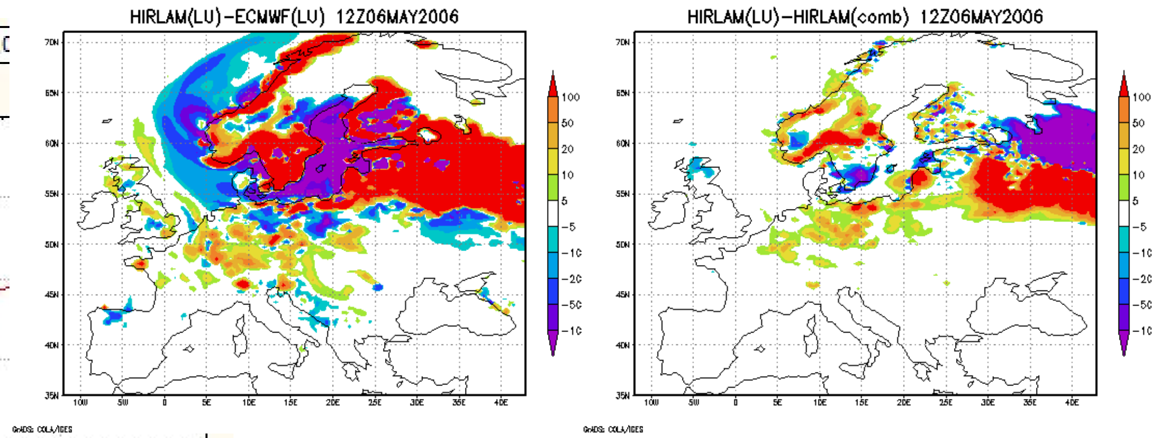
Quick lessons from exercises: pollen. 2

Long-term:



Short-term:

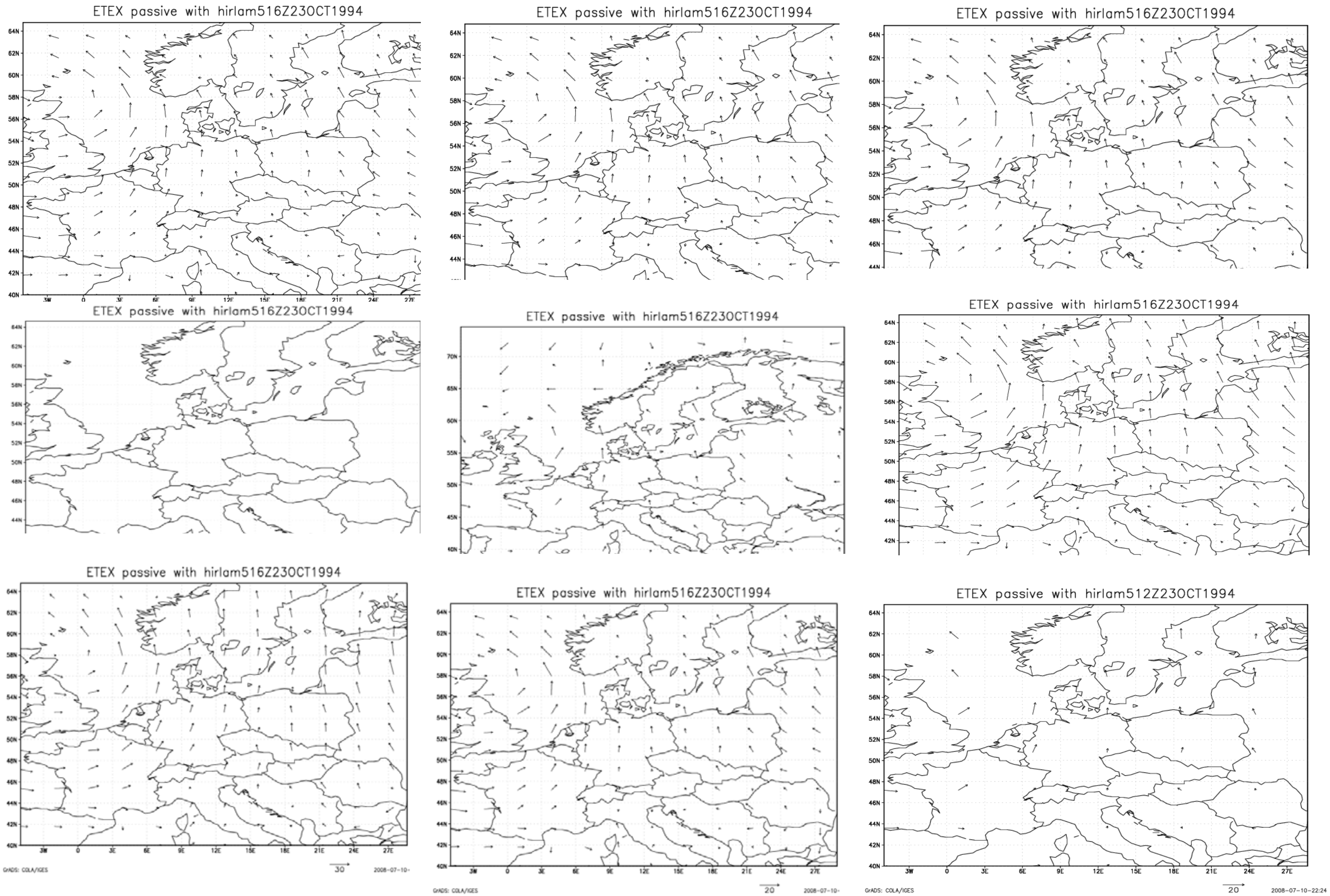




Quick lessons from exercises: ETEX. 1

- Known features from ETEX ensembles have been confirmed
- Generally stable pattern evolution and final distribution
- High uncertainty of the initial 1.5 days of transport, with the first arch of 5 stations being most-vulnerable
- No unequivocal answer whether the plume has split (but probability was evaluated low)

Quick lessons from exercises: ETEX. 2 (1, 2, 3, 4, 5, 8, 9, 10, 11)



Ensemble: a necessity or computerism? Discussion

- Computerism: a decease when scientists believe that a problem, which they can neither solve nor even formulate, can easily be handled if a sufficiently expensive computer is acquired.
- Ensembles: a necessity or computerism?
- Existing models are deterministic while processes to be described are stochastic
  - An “easy” way to describe probability distribution function using deterministic tools is a Monte-Carlo search (random picking): expensive but theoretically converging to a full PDF
  - Existing ensembles are not (and never will be) sufficiently rich to approach a full-PDF description
- In principle, propagation of stochastic processes through deterministic systems can be described too
- Full-PDF solutions are not needed: practically valuable questions require only a small part of it
- Substantial changes in the existing systems, regulations and people thinking is needed to accommodate the unavoidable switch to probabilistic way of AQ descriptions