Data assimilation in numerical weather prediction models: An introduction and an example with MM5 and 4DVAR during IHOP_2002

Hans-Stefan Bauer
Institute for Physics and Meteorology, Hohenheim University, Germany

Content

- What is data assimilation ?
- Why is it necessary ?
- Methods used in data assimilation
  - Nudging
  - Variational assimilation
- Assimilation of DIAL data
  - Scientific questions
  - Case study
  - Methodology
  - First results
  - Summary and Outlook

What is data assimilation ?

Data assimilation is the method of producing an accurate description of the atmospheric state from raw observations and a prior knowledge given by the forecast model ("first guess").

It provides the most probable state of the system, given the uncertainties in the observations and the model forecast. The produced analysis represents the best fit to the observations as well as to the model background field.

The central axiom used in nearly every data assimilation scheme is the method of the least-squares developed by Gauss in the 19th century. It is used to minimize the differences between model state and observations.

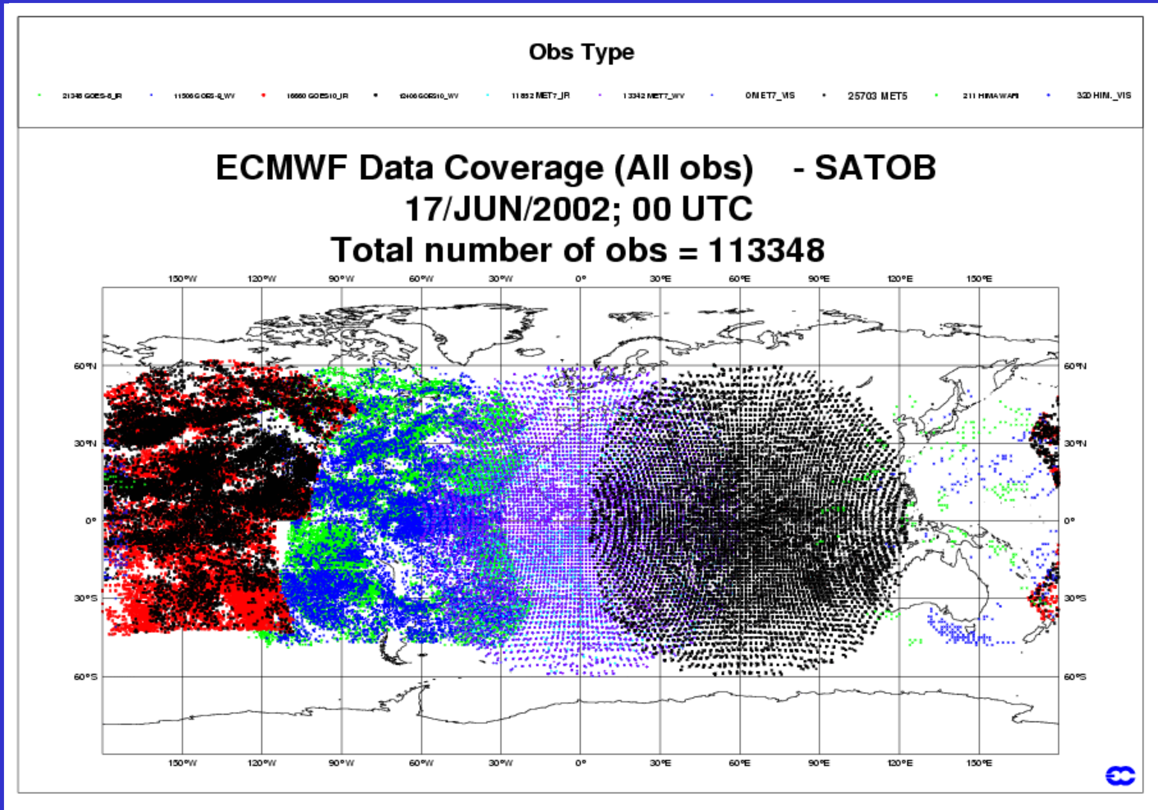
Every days data assimilation ☺

As the following example (Wergen, 2003) shows – data assimilation is part of our daily life.

- To cross the road you first need observations about approaching cars.
- From the observation of the traffic you can estimate their speed.
- Then you use your knowledge about an "average car driver" (your forecast model) to decide whether it is safe to cross the road.
- This can fail either when the observations or your forecast model ("average car driver") are incorrect.

Errors in the initial conditions or the forecast model are the most important reasons for failures of model forecasts.

Why is it necessary ?



Observations are unevenly distributed in space and time.

To ensure the availability of initial information at each grid point is therefore necessary to use as additional information a background field from an earlier model forecast.

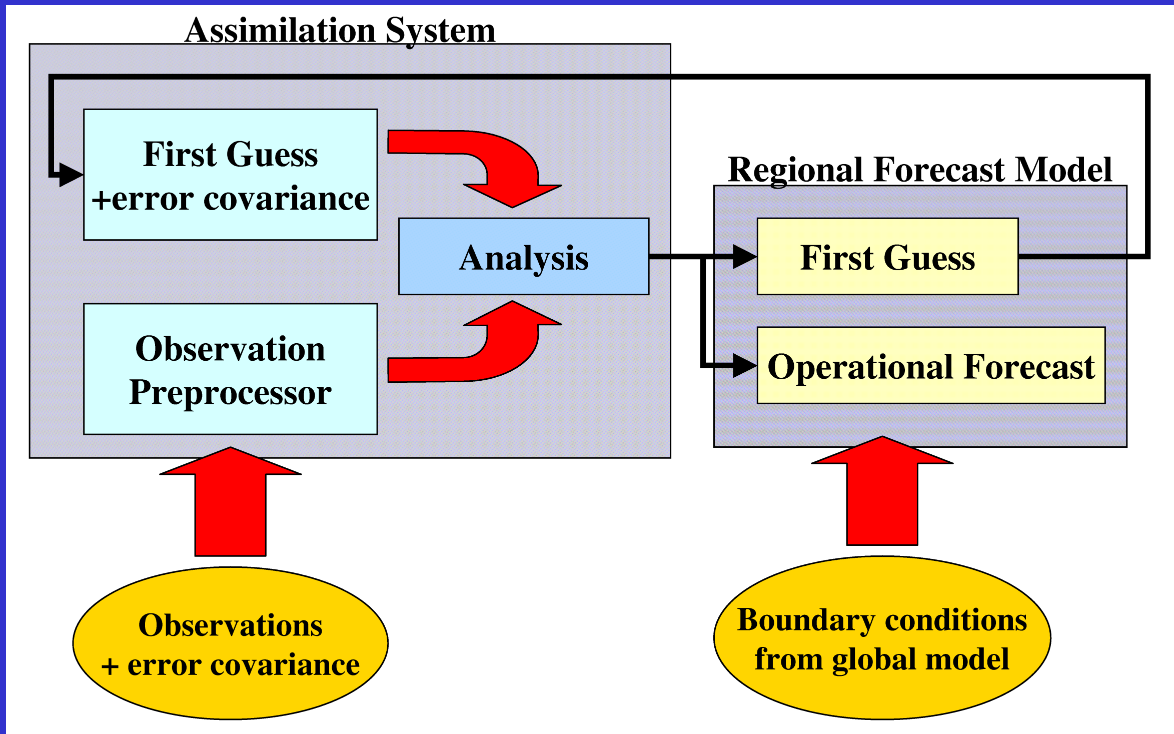
Methods

Many assimilation techniques have been developed for meteorology and oceanography. They differ in their numerical cost, their optimality, and in their suitability for real-time data assimilation (Bouttier and Courtier, 1999).

e.g. the whole data assimilation cycle in the operational LM setup has to be finished within 15 min.

Coarse 4DVAR run:

- forward run 15 s
- data assimilation needs an hour or more



Nudging

Nudging is done by adding a non-physical relaxation term (depending on observations) to the model equations at each time step. Nudging can be done with analysis (e.g. ECHAM) or with real observations (e.g. LM)

$$\frac{\partial}{\partial t}\Psi(x,t) = F(\Psi,x,t) + G_\Psi \cdot \sum_{k_{Obs}} W_k \cdot \left[\Psi_k - \Psi(x_k,t)\right]$$

Model forcing — Nudging term

G_Ψ nudging coefficient

W_k observation dependent weight for a grid point depending on distance and observation quality.

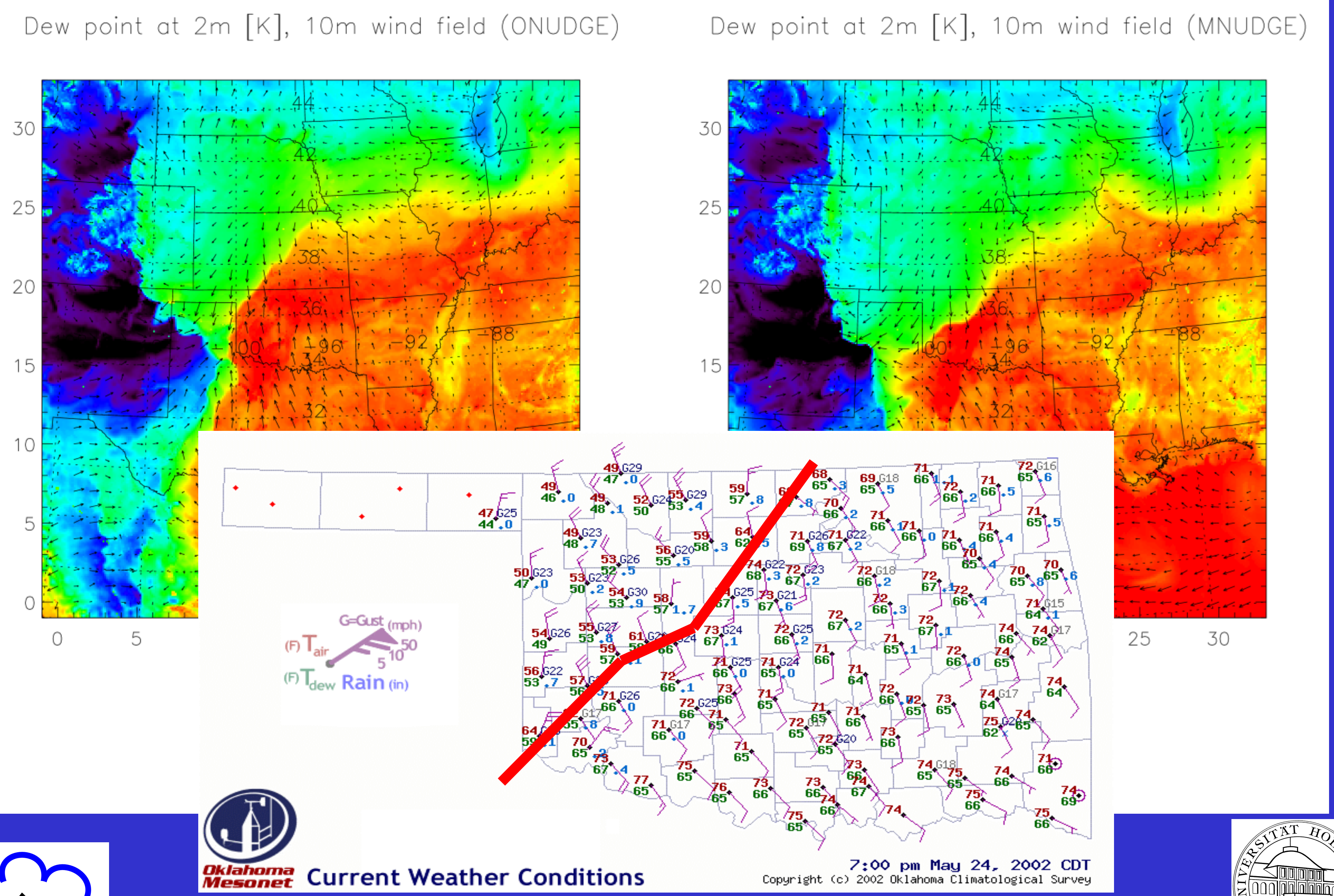
$$W_k = \frac{w_k}{\sum_j w_j} \cdot w_k \quad ; \quad w_k = w_t \cdot w_{xy} \cdot w_z \cdot \varepsilon_k$$

time — horizontal — vertical — quality

Weight depending on the temporal and spatial distance as well as the quality of the observations.

$G_\Psi = 6 \cdot 10^{-4}\ s^{-1}$ for all assimilated quantities

Example: LM and IHOP



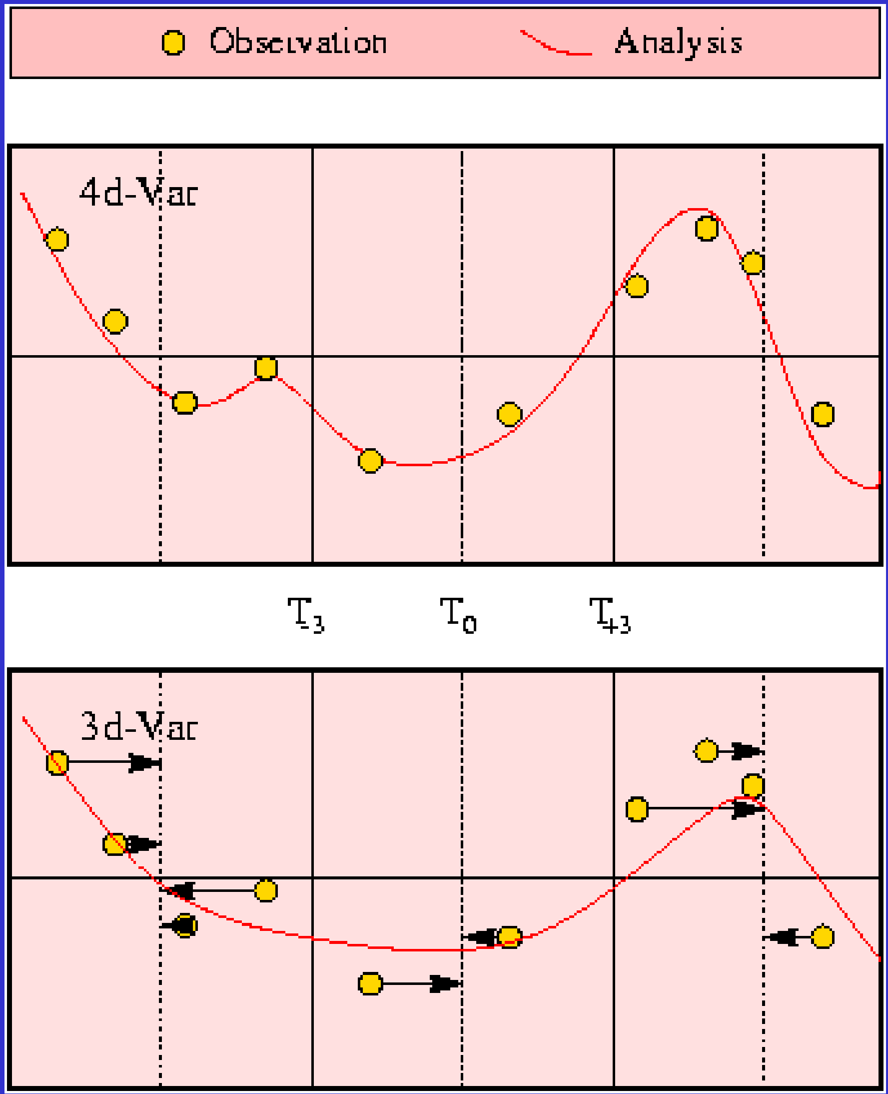
Variational data assimilation

This method minimizes the square of the difference between the analysis and the observations/background in an iterative way. It was developed in the 70ies and allows the assimilation of quantities that are not variables of the model (e.g. data from remote sensing systems).

$$J(\mathbf{x}) = (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{n} [\mathbf{y}_i - \mathbf{H}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{H}_i(\mathbf{x}_i)]$$

- Cost function: $J(\mathbf{x})$
- State vector: $\mathbf{x}_0$
- Background field vector: $\mathbf{x}_b$
- Background error covariance matrix: $\mathbf{B}$
- Observations: $\mathbf{y}_i$
- Forward operator: $\mathbf{H}_i$
- Observation error covariance matrix: $\mathbf{R}_i$

3DVAR versus 4DVAR



3DVAR uses all observations within a time window as simultaneous observations.

The holy grail of variational data assimilation in the so-called 4DVAR, where the observations are assimilated at the same time they were collected. This is clearly more realistic than 3DVAR, but it is bought dearly by a much larger need of computing time.

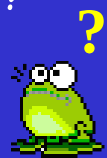
Assimilation of LASE water vapor data into MM5 using 4DVAR

H.-S. Bauer, V. Wulfmeyer, M. Grzeschik, A. Behrendt,
F. Vandenberghe, and E.V. Browell

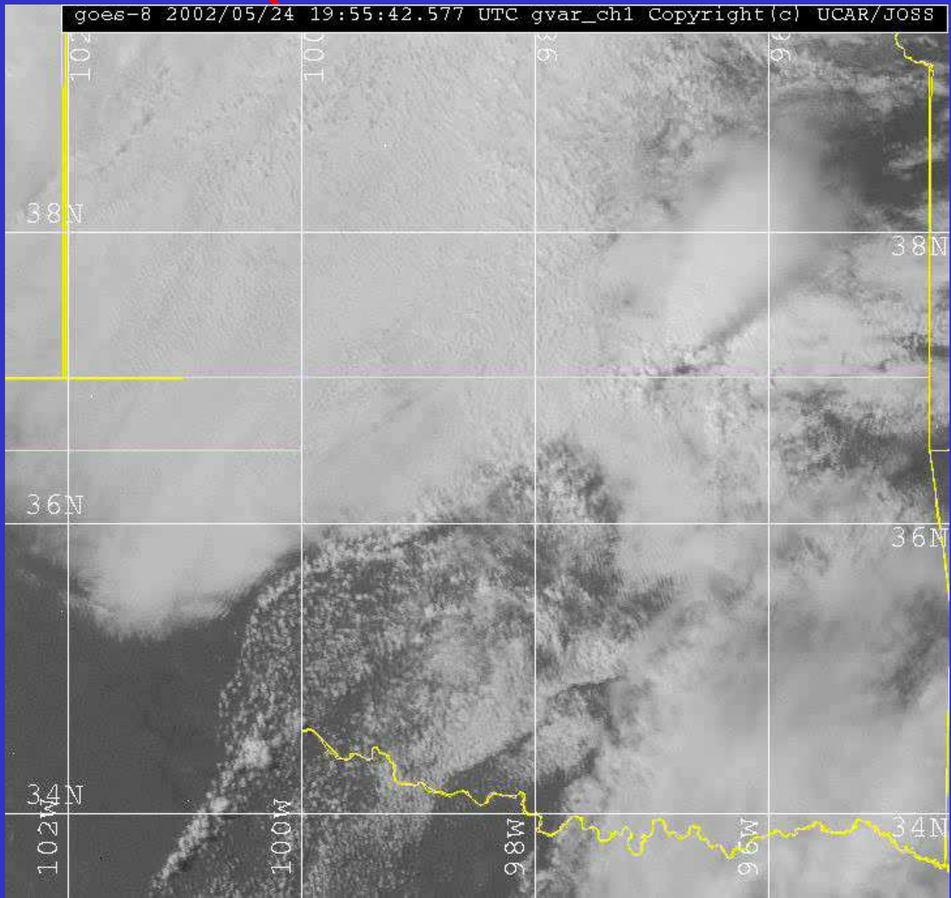
International H_2O Project
IHOP_2002

Scientific Questions:

- Is there a significant influence on the water vapor field and its temporal development when only data from one flight track is assimilated ?
- Is the precipitation forecast influenced by the assimilation ?
- Does the signal propagate in the model forecast even if data is no longer assimilated ?
- Are the DIAL data suited for data assimilation into mesoscale models or are there some secondary effects which erroneously push the model into a wrong direction.



Case study during IHOP_2002: May 24

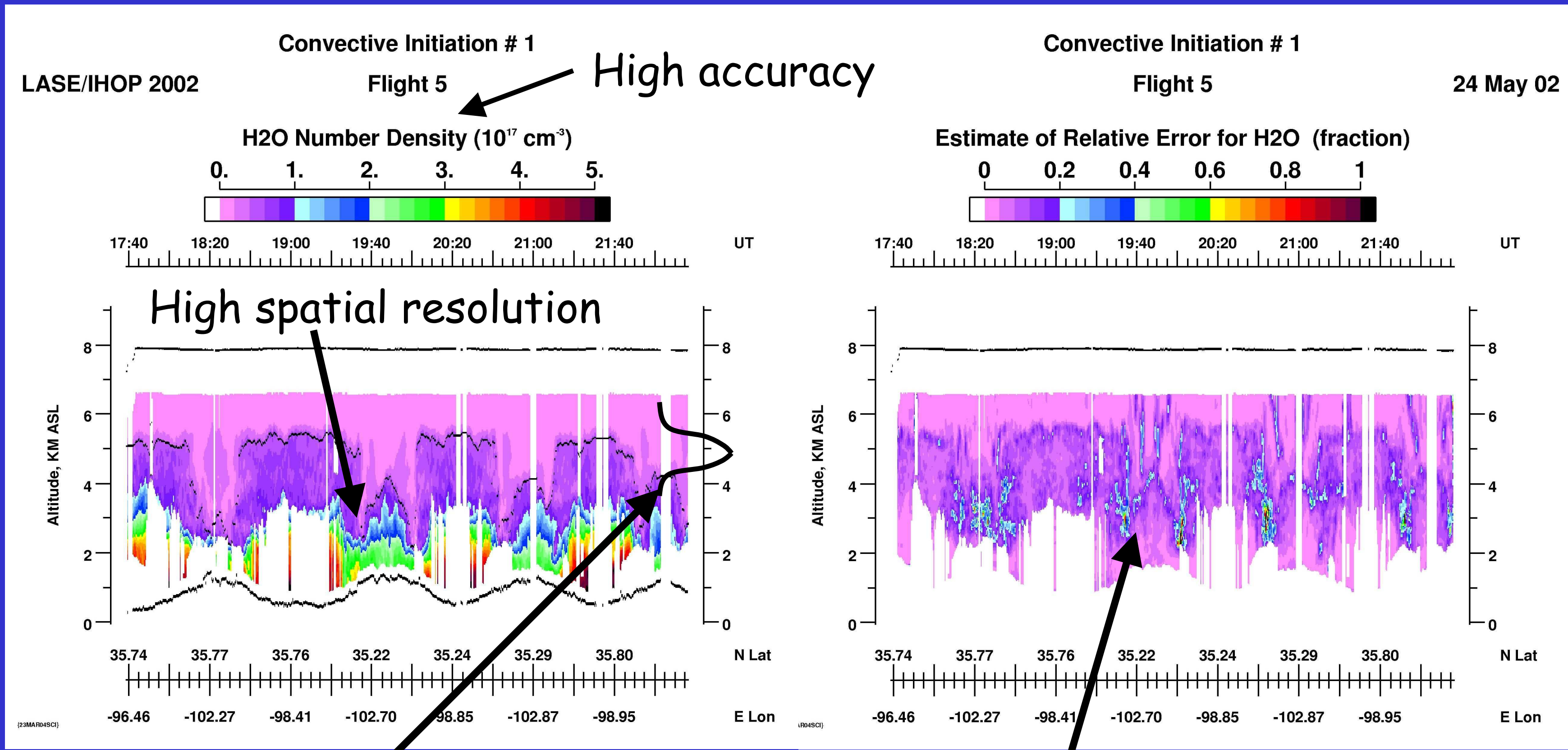


- Study of the role of the water vapor field in the initiation of convection
- Combination of state-of-the-art water vapor instrumentation applied
- Huge amount of supporting background data and model runs available

Ideal opportunity for studying the impact of water vapor data assimilation on the forecast quality of mesoscale models.

Initiation of convection

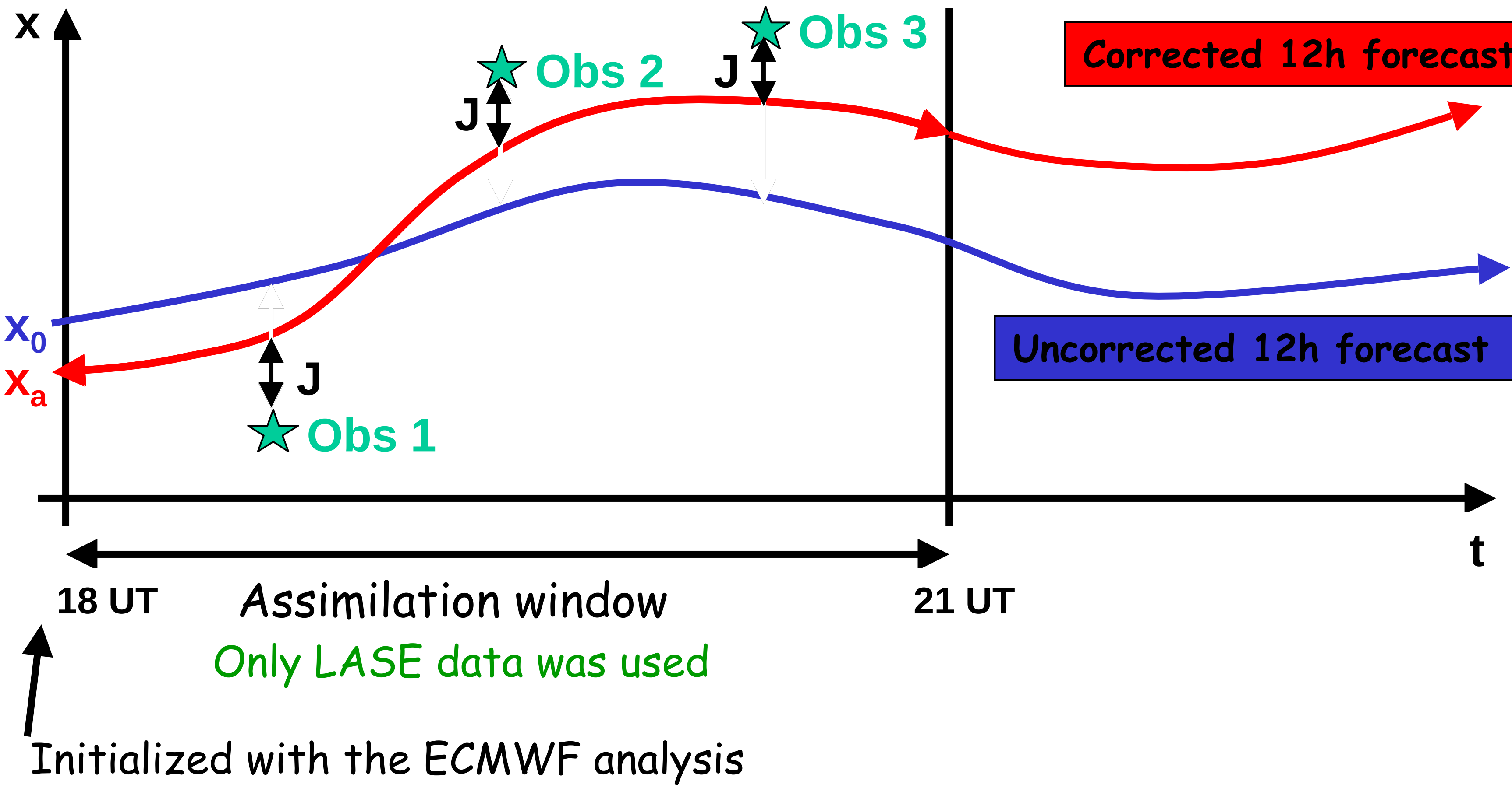
Why are DIAL systems well-suited for data assimilation?



Well-known error correlation (e.g. autocor. of 300-m triangle function in the vertical)

Low and well-specified errors for each profile using gliding autocovariance analysis of water vapor time series

4-D variational data assimilation (4DVAR)



4DVAR uses the LASE data in an optimal way

Preparation of IHOP 4DVAR analysis at IPM

Installation of MM5 with data assimilation tools

Prepare boundary conditions from ECMWF analysis

Development of the "LASE decoder":
- **Use MM5 radiosonde operator**
- **Incorporate corresponding LASE error profiles**
- **Construct error covariance matrix**
- **Interpolate between the MM5 model grid and the LASE data**

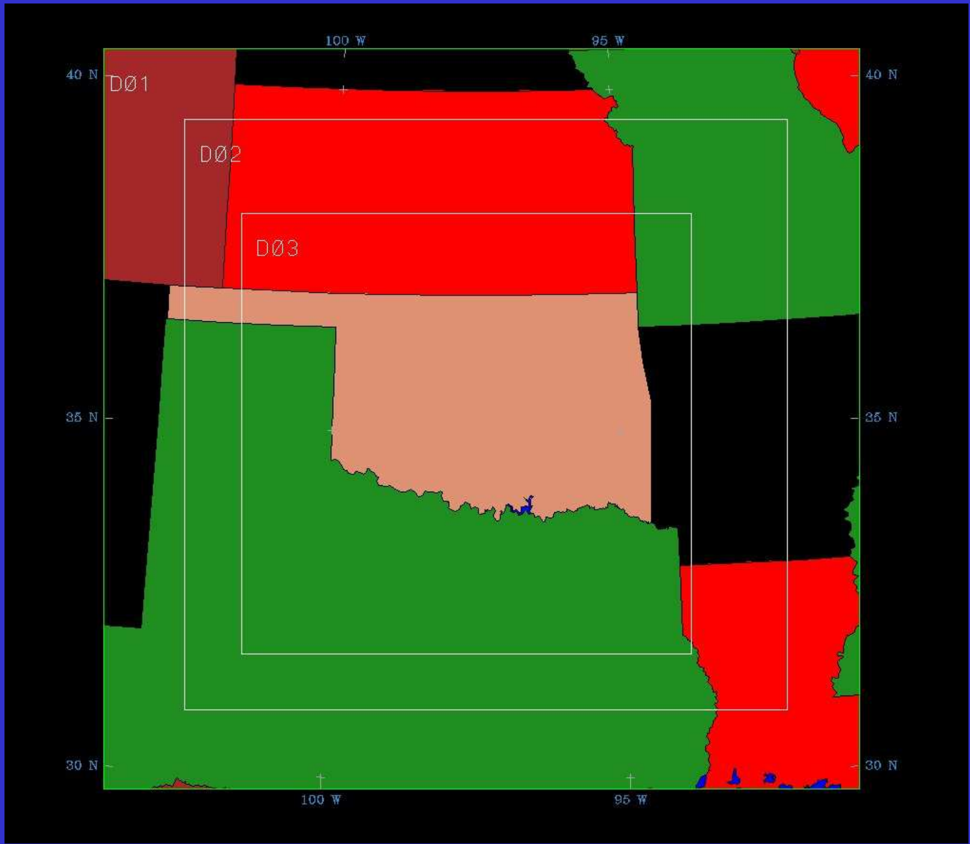
Data assimilation run with low resolution and simplified physics

Interpolation of improved initial field X_a to high-resolution grid

Rerun MM5 with high-resolution initial field and optimized initial conditions

First 4DVAR DIAL operator and data assimilation effort

MM5 configuration



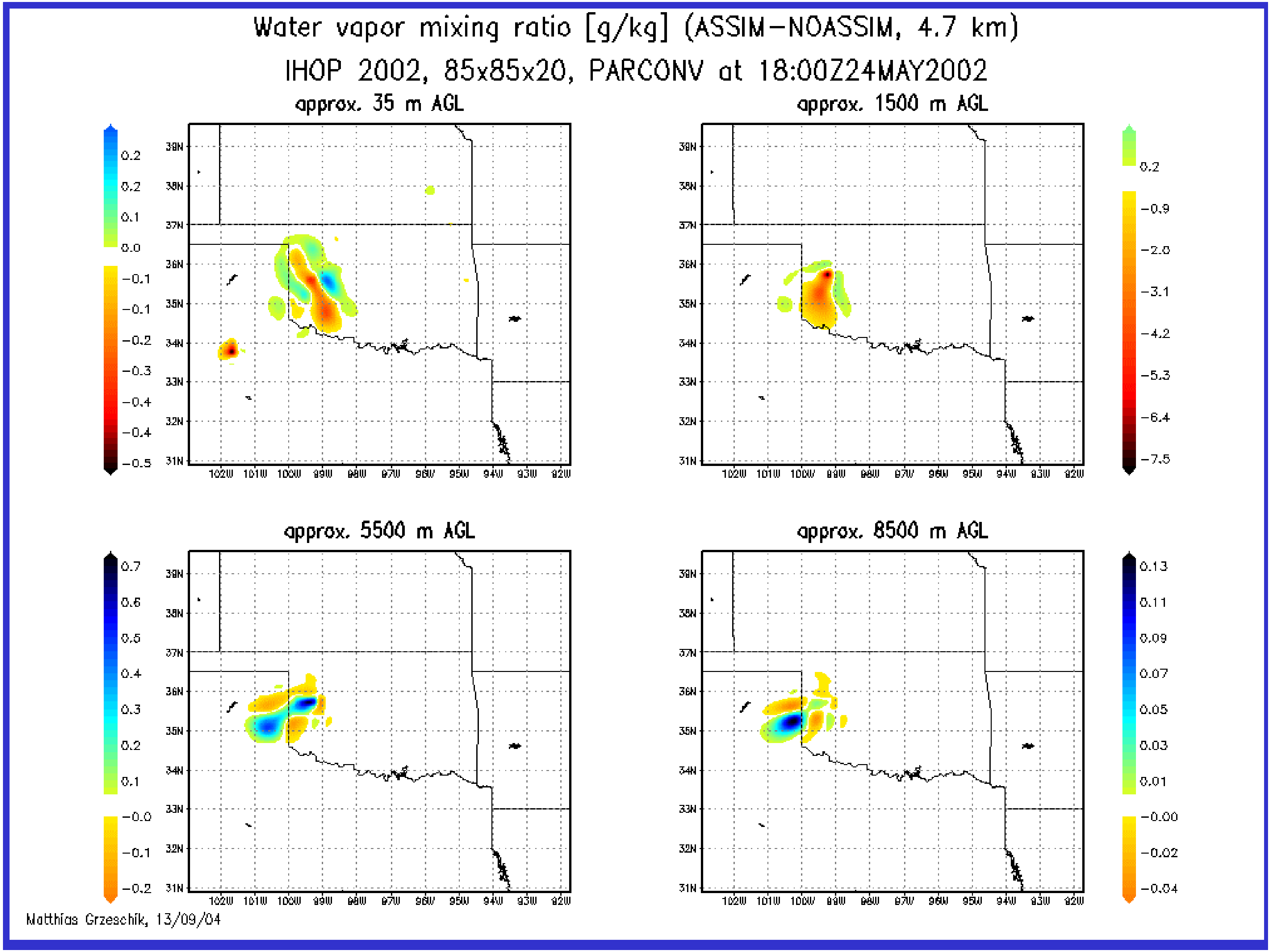
D1: 85x85x20 grid points
D2: 202x202x36 grid points
D3: 451x451x80 grid points

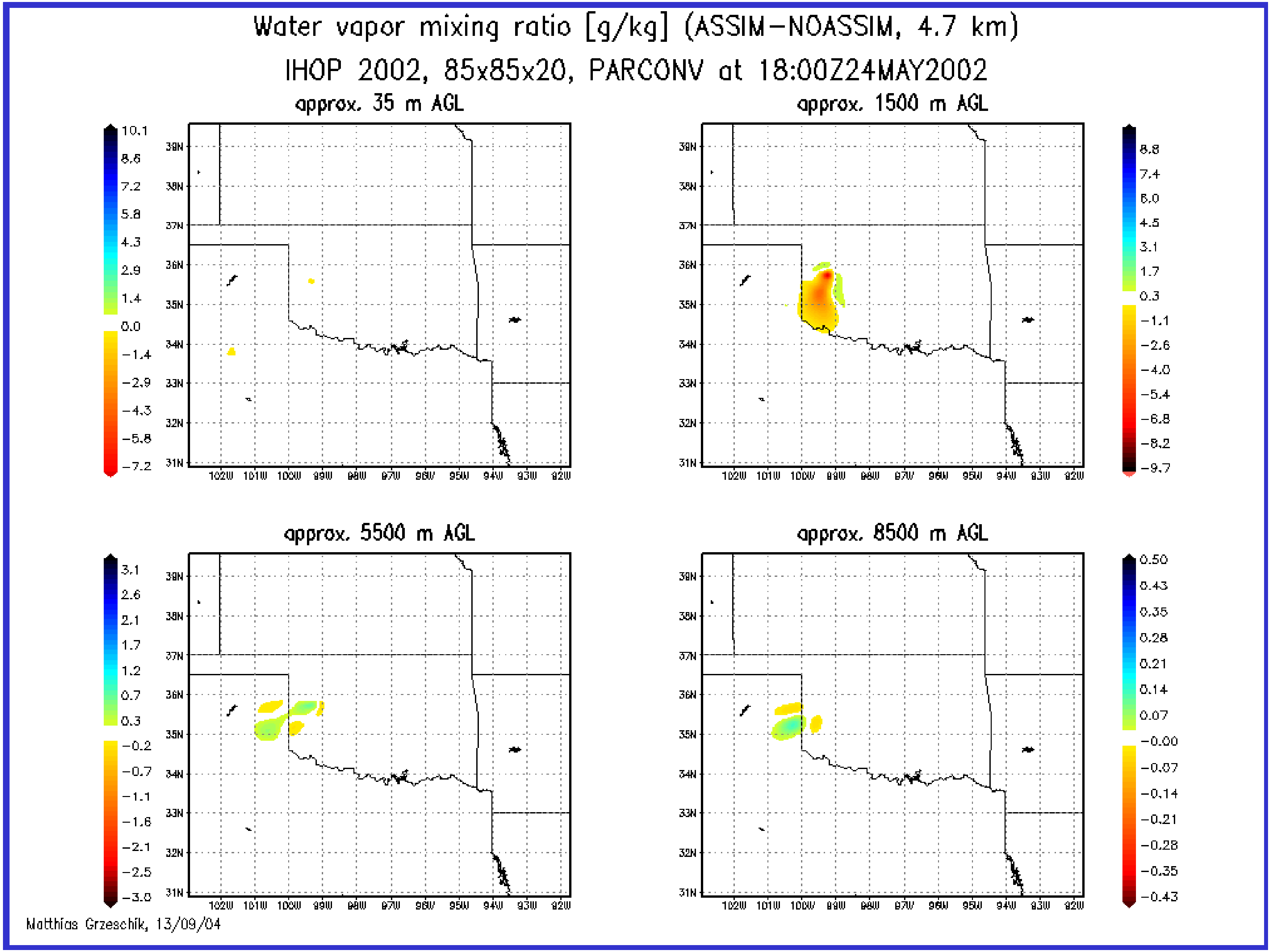
Assimilation run:

- Anthes-Kuo convection
- MRF PBL scheme
- stable precipitation scheme
- Simple radiation scheme

Free forecasts:

- Kain-Fritsch 2 convection
- MRF PBL scheme
- GSFC graupel scheme
- RRTM radiation

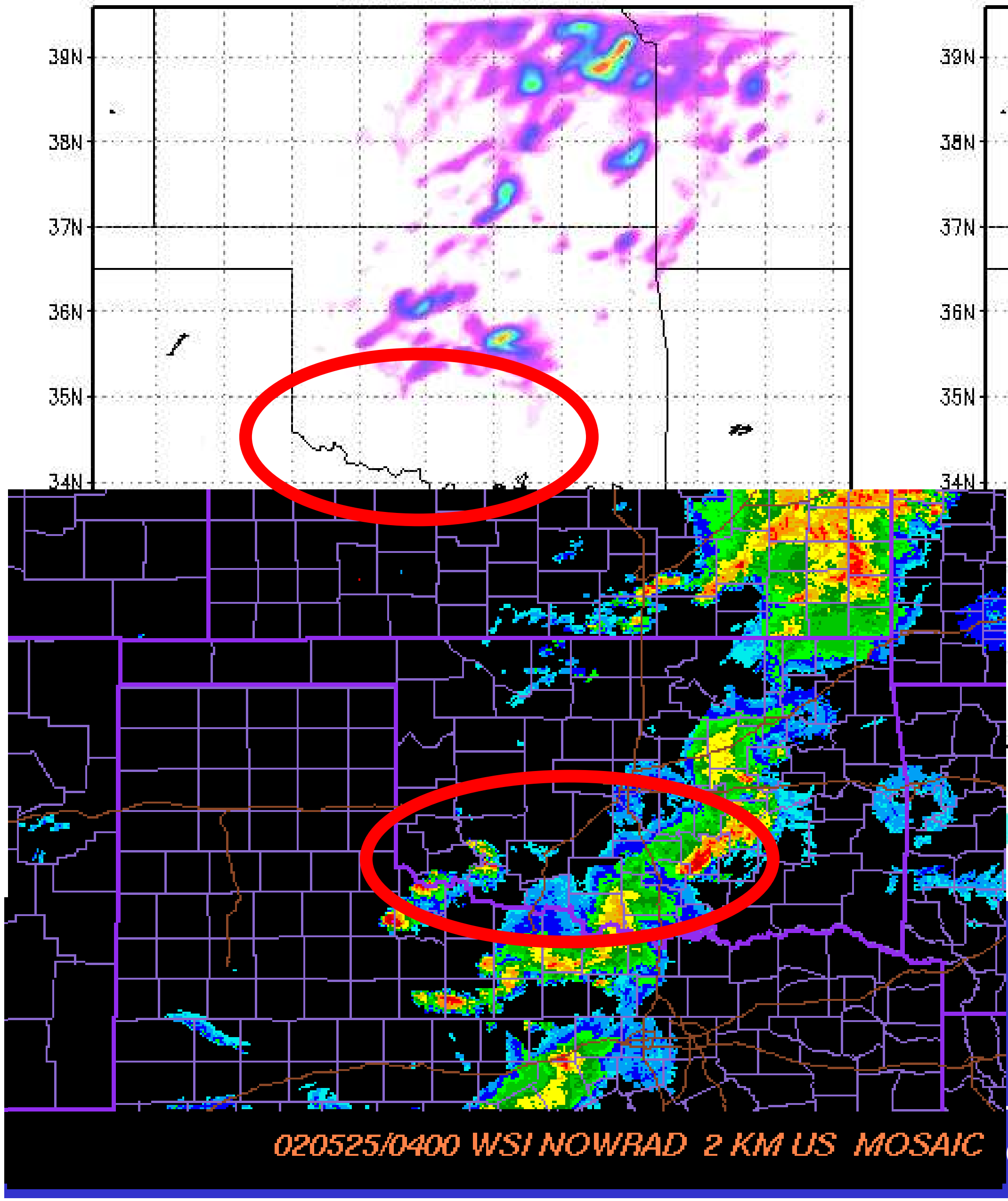

Water vapor mixing ratio [g/kg] (ASSIM−NOASSIM, 4.7 km)
IHOP 2002, 85x85x20, PARCONV at 18:00Z24MAY2002
approx. 35 m AGL
0.2 0.2 0.1 0.0 −0.1 −0.1 −0.2 −0.3 −0.4 −0.4 −0.5
approx. 1500 m AGL
0.2 −0.9 −2.0 −3.1 −4.2 −5.3 −6.4 −7.5
approx. 5500 m AGL
0.7 0.6 0.5 0.4 0.3 0.2 0.1 −0.0 −0.1 −0.2
approx. 8500 m AGL
0.13 0.11 0.09 0.07 0.05 0.03 0.01 −0.00 −0.02 −0.04
39N 38N 37N 36N 35N 34N 33N 32N 31N
102W 101W 100W 99W 98W 97W 96W 95W 94W 93W 92W
Matthias Grzeschik, 13/09/04



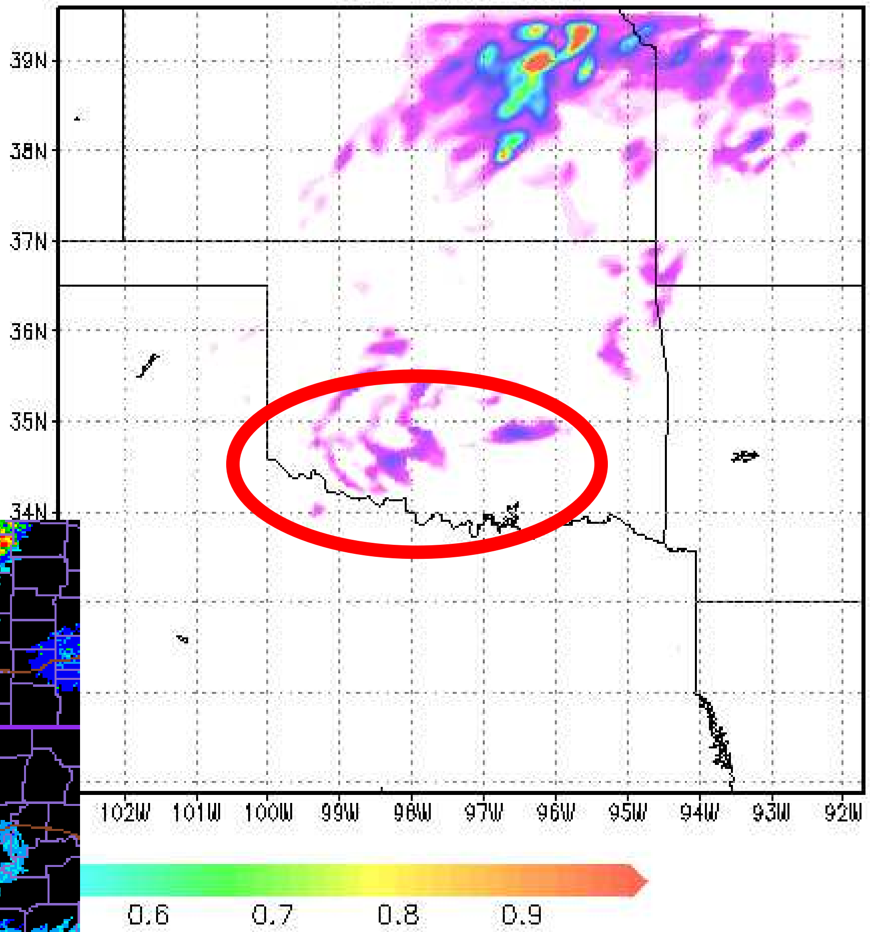
Water vapor mixing ratio [g/kg] (ASSIM-NOASSIM, 4.7 km)
IHOP 2002, 85x85x20, PARCONV at 18:00Z24MAY2002
approx. 35 m AGL
10.1 8.6 7.2 5.8 4.3 2.9 1.4 0.0 -1.4 -2.9 -4.3 -5.8 -7.2
approx. 1500 m AGL
8.8 7.4 6.0 4.5 3.1 1.7 0.3 -1.1 -2.6 -4.0 -5.4 -6.8 -8.2 -9.7
approx. 5500 m AGL
3.1 2.6 2.1 1.7 1.2 0.7 0.3 -0.2 -0.7 -1.1 -1.6 -2.1 -2.5 -3.0
approx. 8500 m AGL
0.50 0.43 0.35 0.28 0.21 0.14 0.07 -0.00 -0.07 -0.14 -0.21 -0.28 -0.35 -0.43
39N 38N 37N 36N 35N 34N 33N 32N 31N
102W 101W 100W 99W 98W 97W 96W 95W 94W 93W 92W
Matthias Grzeschik, 13/09/04

Total Precipitation (4.7 km horizontal resolution)

without assimilation

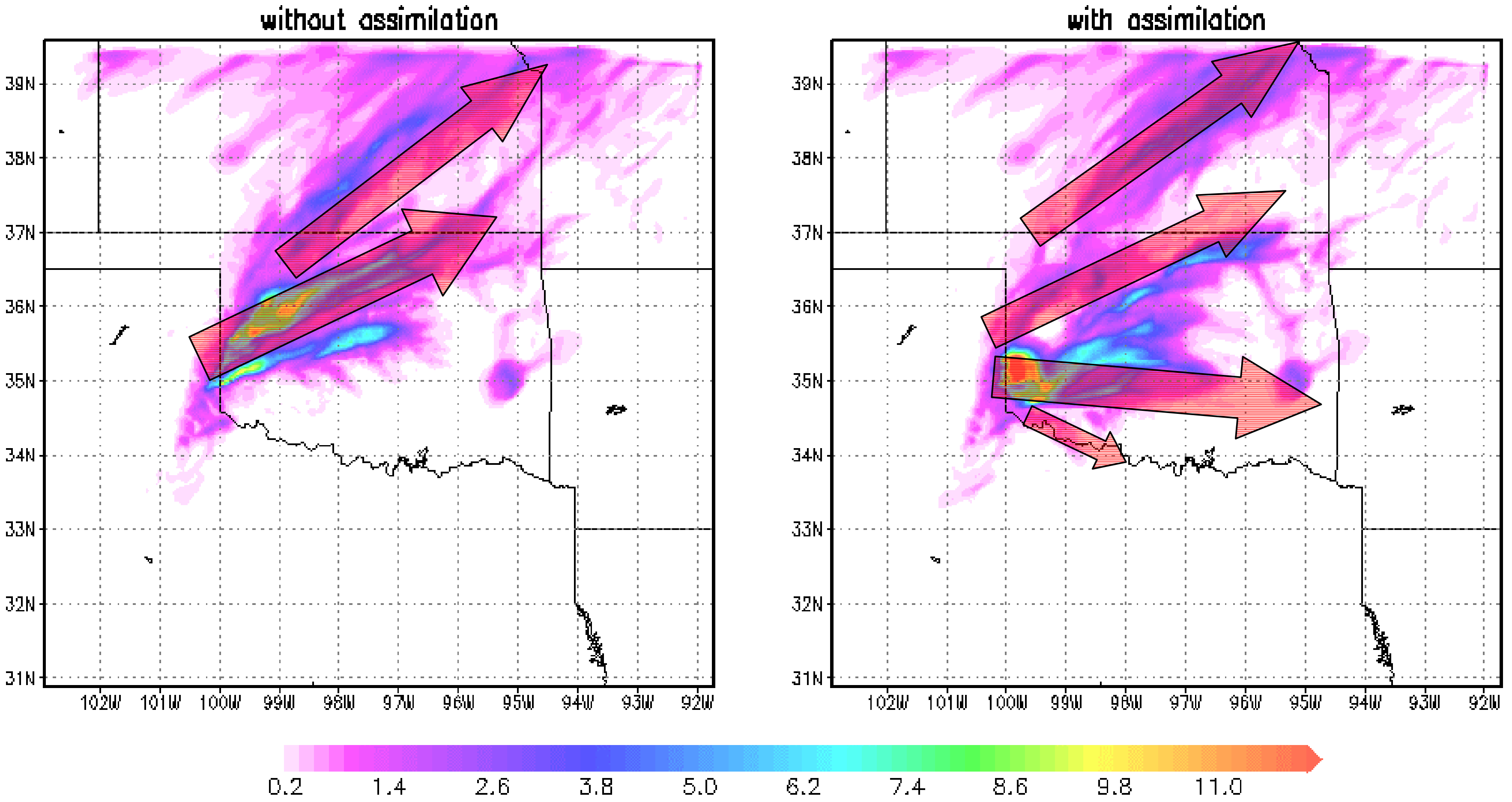


with assimilation



Motion direction of the main precipitation centers

Accumulated total precipitation [cm] (4.7 km horizontal resolution)



Summary

- The experiments show a significant influence of the LASE data onto the water vapor field
- Improved forecast of the distribution of precipitation
- The influence is clearly seen even hours after the assimilation window.
- The LASE data is of high quality and well suited for data assimilation.

Outlook

- Inclusion of additional observations (e.g. further Lidar data, TEMP, SYNOP, radar refractivity, dropsondes, …)
- Detailed quantitative validation of the results and the used ECMWF analysis (e.g. DLR DIAL, GPS, IHOP data base)
- Comparison with other assimilation systems (e.g. 3DVAR, nudging).
- Use of the assimilation system during the field campaign COPS within the DFG priority program “Quantitative Precipitation Forecast”.

Many thanks for your attention !