

A Consortium for COnvection-scale modelling
Research and Development

HARMONIE-AROME data assimilation algorithmic work

5th ACCORD All Staff Workshop
Zalakaros, 31 March – 4 April, 2025

Based on contributors from many colleagues and presented by Magnus Lindskog

- 4D-Var
 - Parallel experiments and daily runs
 - Preparation of cy49 OOPS and MASTERODB
- ENVAR and HYBRID EN/VAR
- Other algorithmic activities for data assimilation and initialisation
- Background errors in observations space
- Exploiting machine learning for DA components
- Concluding remarks

HARMONIE-AROME MASTERODB 4D-Var

- **Daily running at UWC-W (cy43)**
 - United Weather Centers West - Denmark, Iceland, Ireland, the Netherlands.
 - Red frame shows model domain. 2 km horizontal grid distance, 90 vertical model levels.
 - 3-hour assimilation window, 2 outer loop iterations.
 - On-going work with thinning of Mode-S EHS.
- **AEMET 4D-Var parallel runs and case studies (cy46)**
 - Parallel 3D-Var versus 4D-Var runs over AEMET domain (blue frame). The aim is to optimise 4D-Var settings.
 - Valencia flooding case studies.
- **4D-Var in cy49h2 HARMONIE repository**
 - Now functioning with almost the same obs usage as in cy46
 - Will be used for comparison with HARMONIE-AROME cy49h2 OOPS 4D-Var.

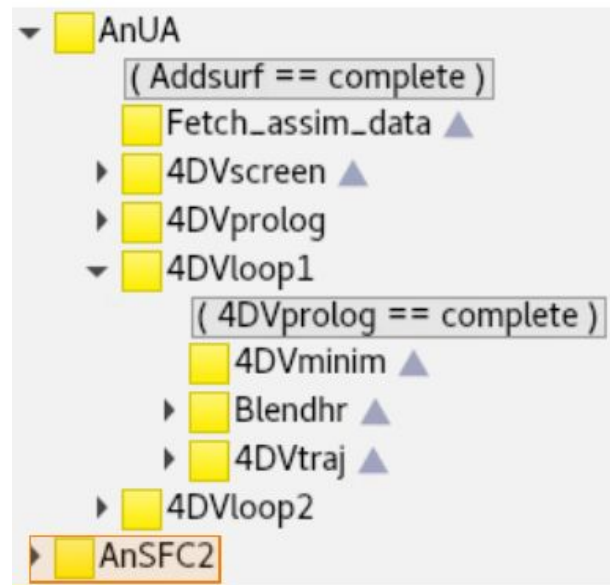


Towards a weak surface/upper-air coupling in HARMONIE-AROME 4D-Var

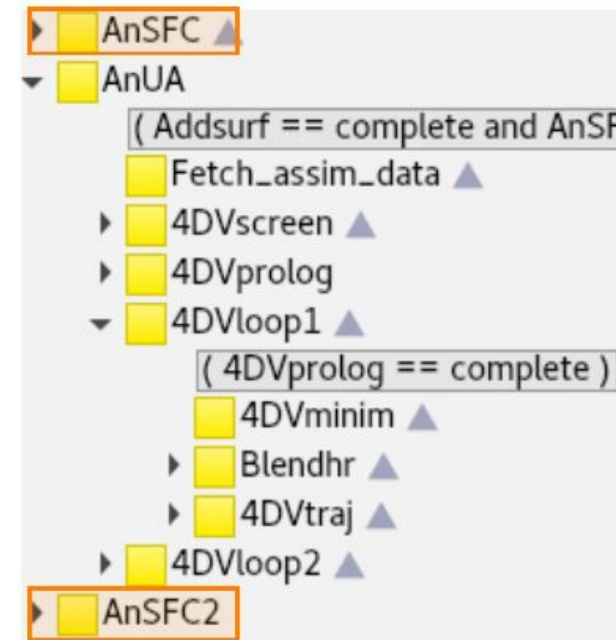
- Exploring a two-way coupling between surface and upper-air analyses, taking advantage of the 4D-Var so-called trajectory runs.
- 3 configurations: one reference case and two different ways of coupling



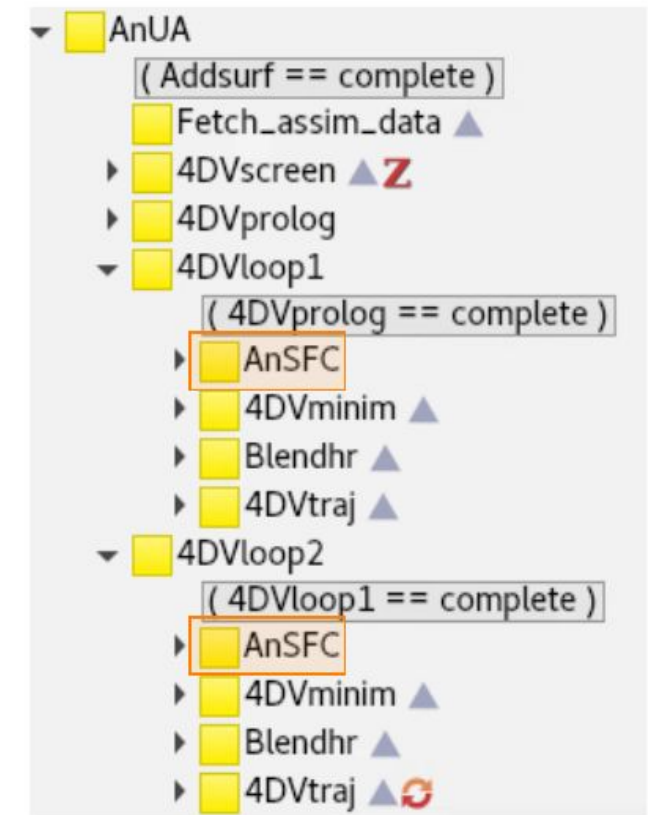
Reference: SA after 4DVar



Before and after 4DVar



Inside outer loops

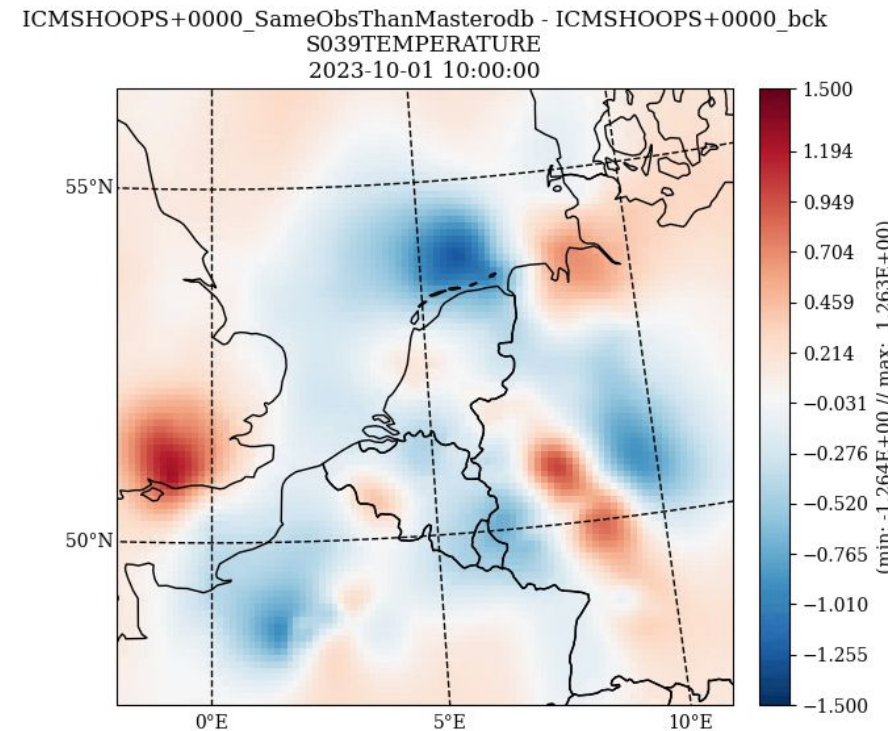


HARMONIE-AROME OOPS 4D-VAR

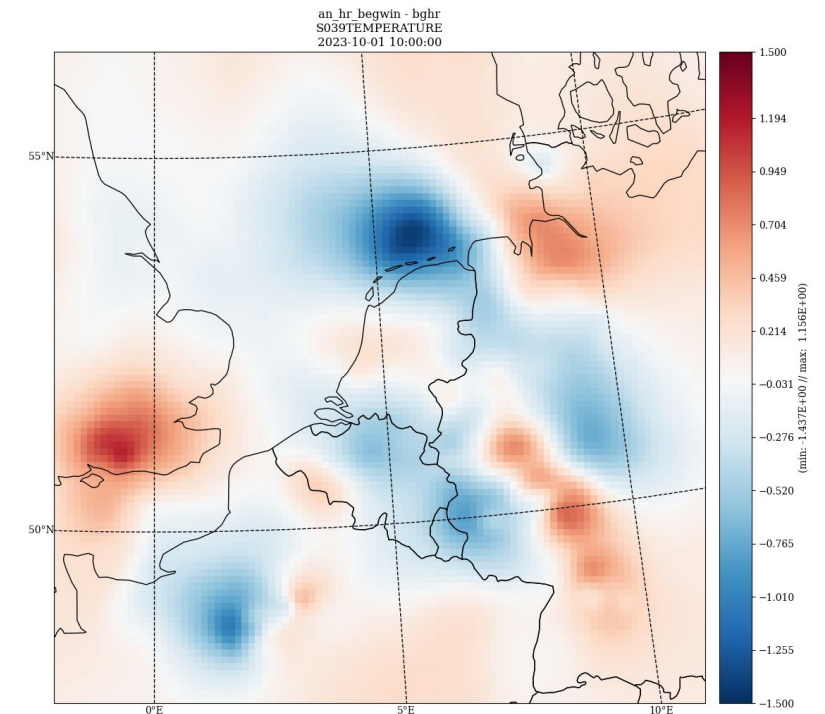
- Development in cy49h2 HARMONIE-AROME repository and low resolution setup.
- Exploring one and two outer loop iteration LAM 4D-Var minimisations and also the change of resolution.
- Comparison with MASTERODB 4D-Var confirms basic functionality.
- Increments are overall similar and we have an idea of the reason for small differences.

Status

OOPS LAM 4D-VAR



MASTERODB LAM 4D-VAR

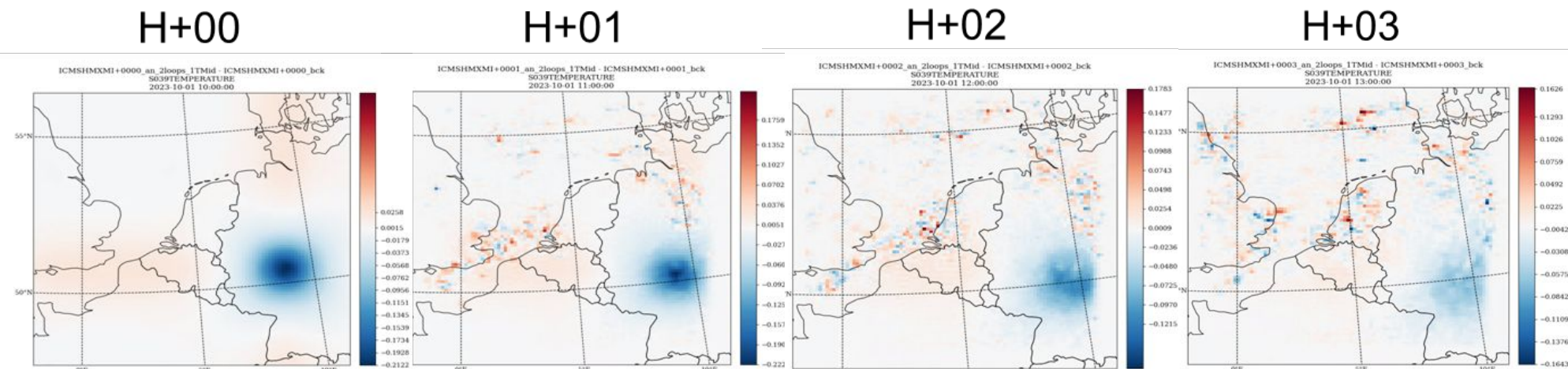


Temperature analysis increments at model level 39 (~820 hPa) for one single assimilation cycle using conventional types of observations.

HARMONIE-AROME OOPS 4D-VAR

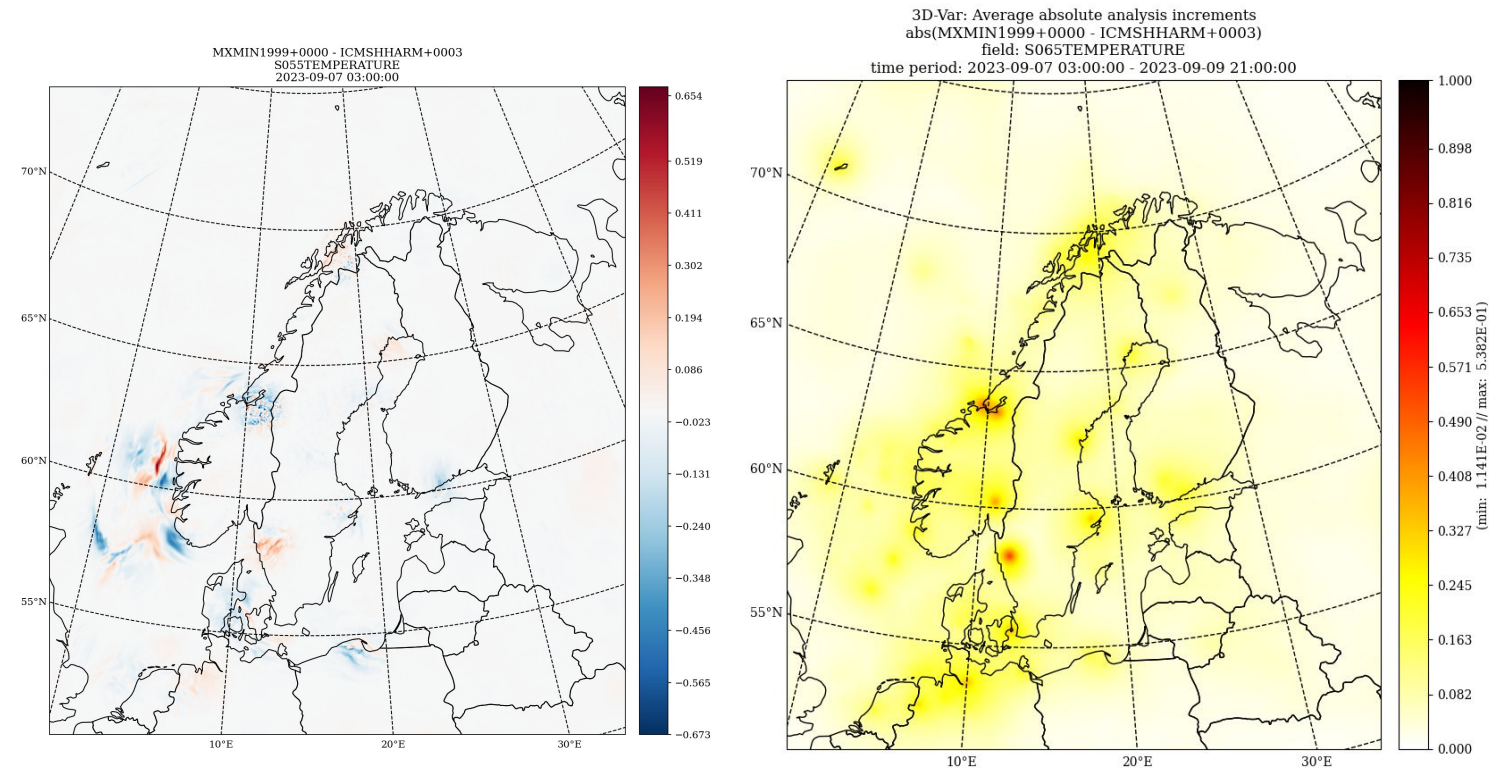
Ongoing and future work

- Introduction of the AROME/SURFEX model setup in the json/C++ layer to be used only in the high resolution non-linear trajectory run. The general model setup for minimization is the ALADIN setup. This 2 model setup mimics the MASTERODB 4D-VAR MASTERODB.
- Find the reason for noise found in OOPS LAM propagation of increments (below).
- Apply in full scale data assimilation configurations and introduce remote sensing observation types.
- OOPS paves the way for research on fully coupled Earth system data assimilation (atmosphere/land/ocean).



Temperature increment evolution within 3-hour assimilation window at model level 39 (around 820 hPa) with only single AMDAR temperature observation assimilated.

- Ongoing exploitation of ENVAR approaches in HIRLAM.
- Quality of the ENVAR heavily dependent on ensemble.
- Ensemble data under investigation
 - ECMWF global ensemble forecasts
 - MetCoOp limited area ensemble and various flavours of this.
- First assimilation experiments, but further extended studies planned.
 - Some work on localisation tuning in cy49



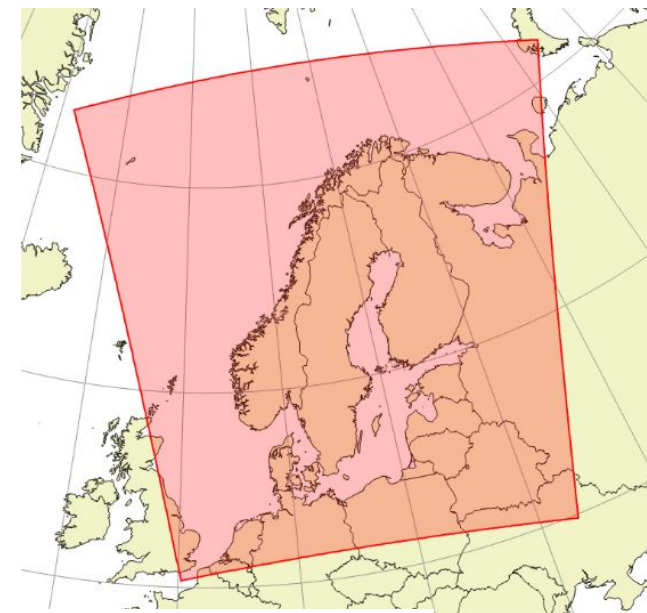
ENVAR Temperature assimilation increments (K) at lowest model level (65) for (one particular assimilation cycle (left) and absolute increment value accumulated over three days of assimilation (right).

MetCoOp Ensemble

Ensemble settings

- Member zero is left unperturbed and considered as the control member.
- Initial uncertainties
 - Perturbation of upper air observations
 - Perturbation of the following surface fields: SST, soil temperature and soil moisture.
- Boundary perturbations
 - Using IFSENS boundaries member 1-28 for the perturbed members and IFS HRES for the control (member zero). The mapping between the MEPS member and the IFSENS member used as boundary conditions corresponds to $n:2n-1$ and $n:2n$ for 00/01/02 and 03/04/05 cycles respectively. And so on for the other cycles.
- Model uncertainty
 - Perturbation of the following static surface fields: albedo, roughness.
 - Stochastically Perturbed Parameterizations (SPP) perturbs CLDDPTHDP, VSIGSQSAT, ICE_CLD_WGT, RZC_H and RZL_INF with a spatial/temporal scale of 200km and 12h respectively.

The surface field perturbations mentioned above have a typical length scale of 150 km and may be either multiplicative or additive. The member distribution over time is given in the table below.



Member distribution			
Time (UTC)	CIRRUS	STRATUS	VINHA
00,03,...,21	0	1,2,12	9
01,04,...,22	7	3,4,13	10
02,05,...,23	8,14	5,6	11

5 new members
each hour
(run on 3 HPCs)

Collect ensemble members for extended period at different forecast ranges to gain experience in ENVAR and also investigate various selection of perturbations.

HYBRID EN/VAR

Formulation

$$J(\delta x_{\text{var}}, \alpha) = \beta_{\text{var}} J_{\text{var}}(\delta x_{\text{var}}) + \beta_{\text{ens}} J_{\text{ens}}(\alpha) + J_0$$

$$\frac{1}{\beta_{\text{var}}} + \frac{1}{\beta_{\text{ens}}} = 1. \quad \delta x = \delta x_{\text{var}} + \kappa \sum_{k=1}^K (\alpha_k \circ \delta x_k^{\text{ens}})$$

$$J_{\text{ens}} = \frac{1}{2} \alpha^T \mathbf{A}^{-1} \alpha$$

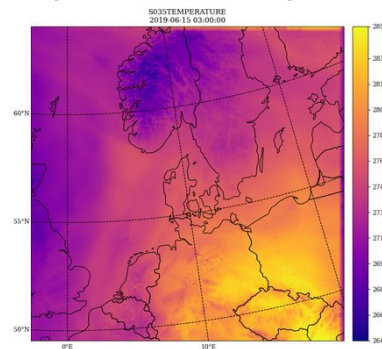
$$\delta x_k^{\text{ens}} = x_k^{\text{ens}} - \overline{x_k^{\text{ens}}}$$

$$\mathbf{B}_{\text{ens}} = \mathbf{A} \circ \mathbf{B}_{\text{raw-ens}}$$

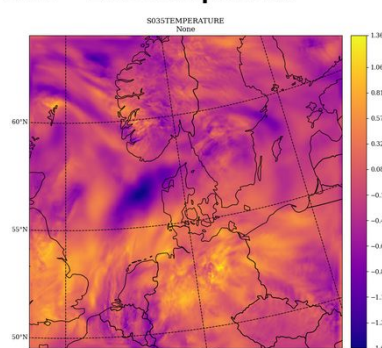
A constructed perturbation framework was implemented in the HARMONIE-AROME Hybrid EN/VAR system to gain a better insight into the error growth mechanisms.

Cloud analysis (q-CONDOR project)

Temperature Level 35 (500hPa)

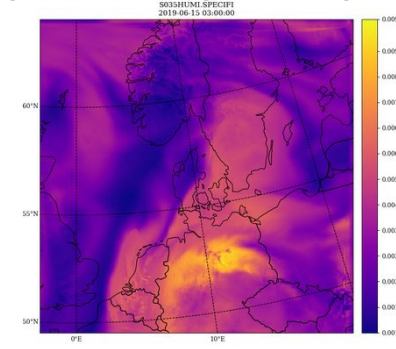


Clim + Ens component



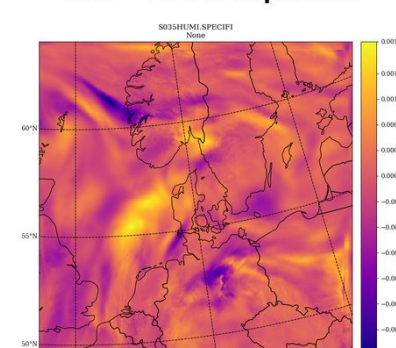
Background state (+03h forecast)

Specific Humidity Level 35 (500hPa)

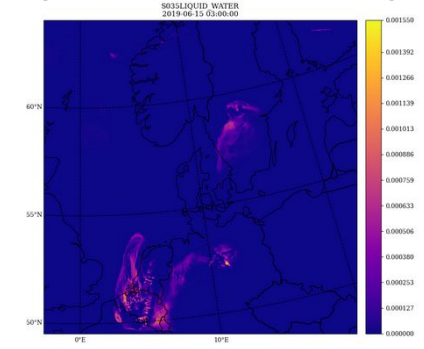


Analysis increment (Hybrid 3DEnVar scheme)

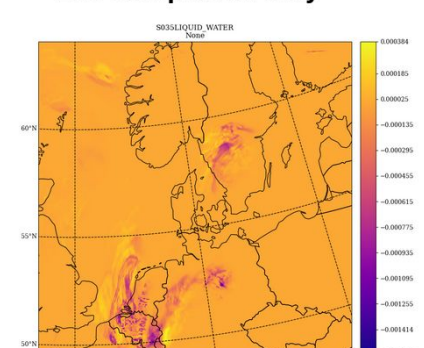
Clim + Ens component



Liquid Cloud Water Level 35 (500hPa)



Ens component only



Other algorithmic activities for data assimilation and initialisation

- Centered incremental analysis update
- Variational constraints
- Incorporating host model information in the data assimilation
- Cloud data assimilation by additional cost function penalty term
- Exploiting gaussian integrals for modelling of background error covariances

Cloud data assimilation

Observation operator defined by the Penalty Function

Observations:

- 1) Ceilometer observation of cloud base altitude (p_b)
- 2) Satellite observation of cloud top altitude (p_t)

Penalty Function definition:

$$P(vl) = \begin{cases} \frac{RH(vl) - RH_{crit}(vl)}{\beta}, & \text{for } vl \in [h_b, h_t] \\ \frac{RH(vl) - \alpha(vl)}{\beta}, & \text{for } vl \notin [h_b, h_t] \text{ and } RH(vl) > \alpha \\ 0, & \text{otherwise} \end{cases}$$

where vl – vertical model level; RH – model relative humidity;
 RH_{crit} – critical relative humidity threshold above which cloud starts to form; α - threshold for the humidity level below saturation, β -uncertainty in RH_{crit}

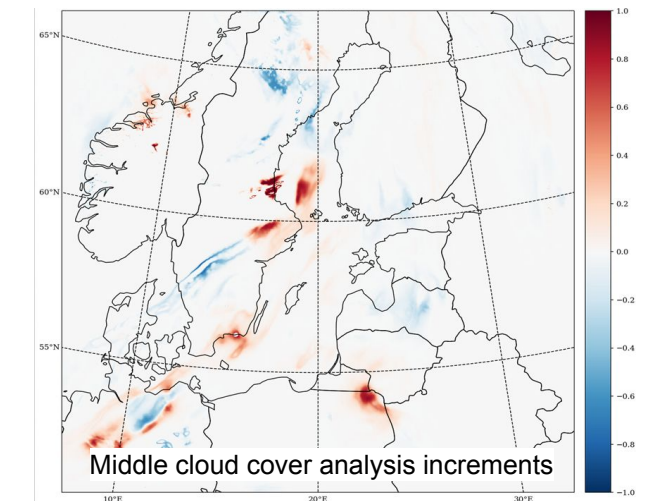
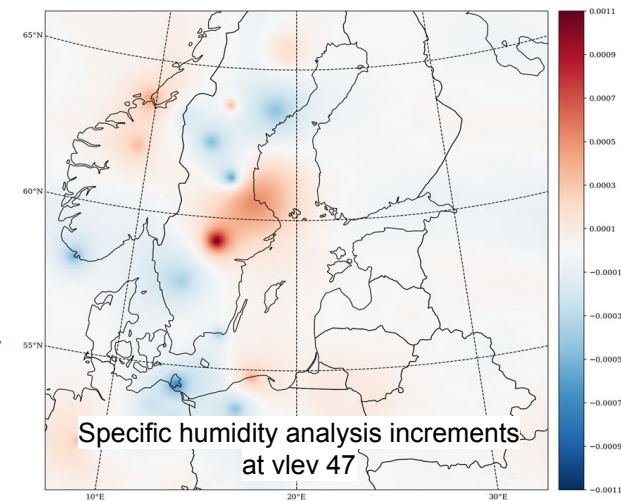
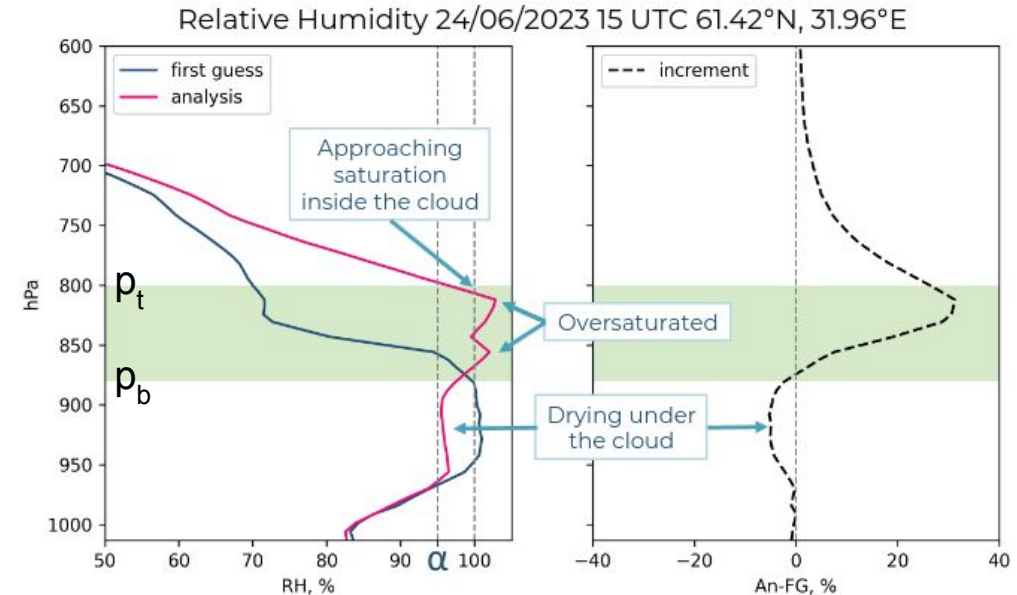
$$J = J_b + J_o + J_c$$

$$J_c = \frac{1}{2} (\mathbf{y}_c - (\mathbf{y}_c + \mathbf{P}))^T (\mathbf{y}_c - (\mathbf{y}_c + \mathbf{P})) = \frac{1}{2} \mathbf{P}^T \mathbf{P}$$

Real observations experiment 20240816_09
 SYNOP (ceilometers) + NWCSAF (CTTH)



Single pseudo-observation illustration



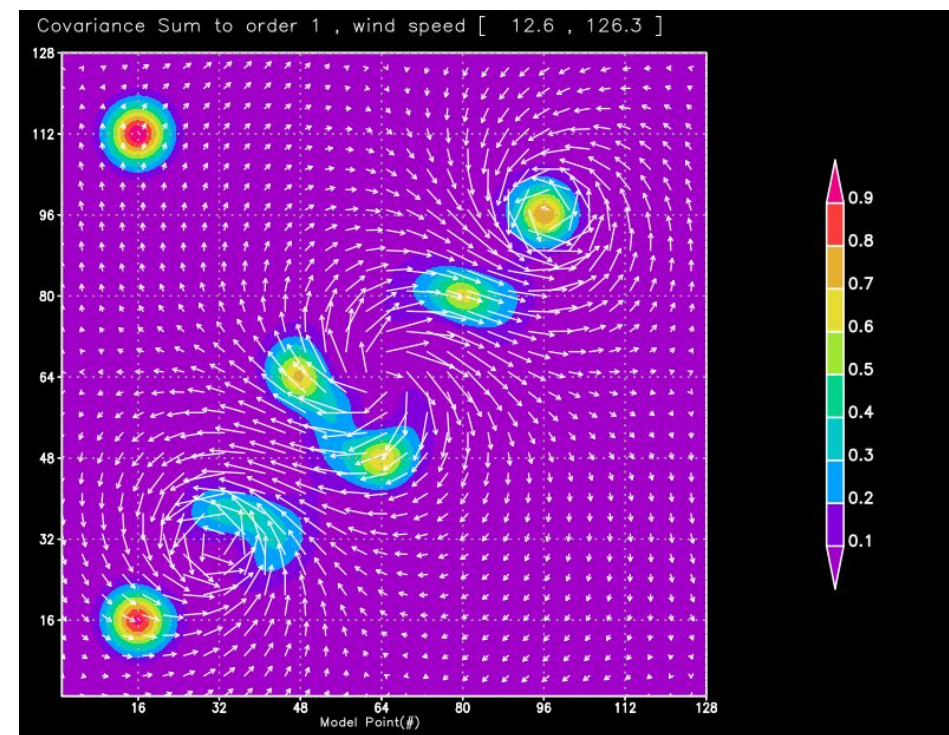
Modelling Background Error Covariances with Gaussian Integrals

- It is possible to give flow-dependency to the \mathbf{B} by introducing a positive quadratic form that depends on the \mathbf{V} field:

$$Z(S) = \int d\Delta_1 \dots d\Delta_N e^{-\frac{1}{2} \left(\Delta^T B^{-1} \Delta + \mu \operatorname{tr} \left[(\mathbf{V} \cdot \vec{\nabla} \Delta) (\mathbf{V} \cdot \vec{\nabla} \Delta)^T \right] \right) + \Delta^T S}$$

- Implementation of prototype and “Proof of Concept” successful.
- Computing performance analysis on CPUs and standard MPI and OpenMP parallelization APIs carried out.
- The idea is appealing, in the first place, for its potential utility in ENVAR, $\mathbf{B} \rightarrow \mathbf{B}[\mathbf{V}]$.

(see Carlos Geijo Guerrero’s presentation in today’s late afternoon session)



BackGround error statistics in Observation Space (BGOS)

Application developed in the OOPS framework to compute background error standard deviation in observation space:

$$\sigma_b^o = \sqrt{\frac{1}{N} \sum_{i=1}^N (HU\chi_i)^2}$$

where

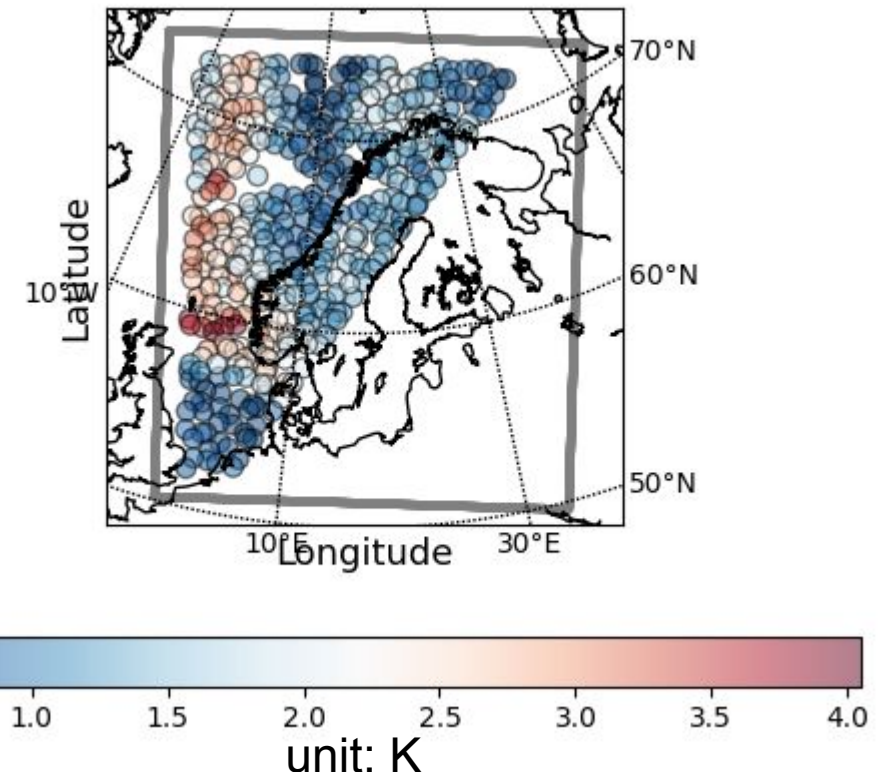
σ_b^o bg error std in ob space
N sample size

U is the series of transform applied to get a unit B matrix in minimization, H is the observation operator (tl or nl) and χ_i is the control vector (containing Gaussian errors) for the individual member i.

$$\left(\begin{array}{l} \delta x = U\chi \\ J_b = \frac{1}{2} \delta x^T B^{-1} \delta x = \frac{1}{2} \chi^T \chi \end{array} \right)$$

Example

MHS channel 3 background error standard deviations in observation space 15 July 2022, 00 UTC. MetCoOp domain.



(corresponding observation error standard deviation is 1.8 K)

(see Jana Sanchez-Arriola's presentation in today's late afternoon session)

Towards use of ML/AI for data assimilation components

- ML-based observation operators for satellite radiances.
- Introduction of hybrid ML/DA approach for improving the MW observation operator over sea ice.
- Enhancing quality of products to be used for assimilation.
- Bias correction of satellite radiances.
- Observation quality control procedures (crowdsourced observations).
- Ensemble generation techniques of relevance for ensemble-based data assimilation.
- Constraining the to a "balanced" manifold in latent-space.
- First prototype for an EnKF using ensemble from AIFS/Bris.

(see Alice Abrahamovic's presentation in Friday's morning session)

Concluding remarks

- Ongoing 4D-Var work towards operational use and porting to OOPS.
- Research on flow-dependent ensemble-based data assimilation algorithms.
- Weakly coupled surface/upper-air data assimilation is being explored.
- Machine-learning provides an opportunity in various aspects of data assimilation.
- Important to have verification systems and metrics relevant for high-resolution data assimilation and to emphasize high-impact weather.

(see Eoin Whelan's presentation in next session)

Thank you for your attention!