HIRLAM data assimilation: current status and vision

SMHI

Meteorologisk institutt

Roger Randriamampianina, Magnus Lindskog
and
The HIRLAM DA Team (see next slide)

High Resolution Hirlam Limited Area Model

ACCORD
A Consortium for COnvection-scale modelling Research and Development

ECMWF Anniversary, Workshop on data assimilation: initial conditions and beyond, 9-10 April 2025

The HIRLAM DA Team:

AEMET: Jana Sanchez; Pau Escriba; Carlos Geijo; Maria Diez; Joan Campins

DMI: Kasper Hintz; Emy Alerskans; Mats Dahlbom; Fabrizio Baordo; Simon Kamuk Christiansen ; Claus Petersen

ESTEA: Ahto Mets; Alina Lerner

KNMI: Sibbo van der Veen, Siebren de Haan, Michal Koutek; Gert-Jan Marseille, Isabel Monteiro, Chris Romick; Erik Dedding, Alice Abramowicz

MET Eireann: Eoin Whelan; Rónán Darcy

MET Norway: Stephanie Guedj; Maté Mile; Jostein Blyverket; Ole Vignes; Roohollah Azad; Benjamin Menetrier; Åsmund Bakketun, Per Dahlgren, Roel Stappers, Roger Randriamampianina

FMI: Erik Gregow; David Schönach ; Reima Eresmaa

SMHI: Günther Haase; Martin Ridal; Swapan Mallick; José Faúndez; Nina Håkansson; Tomas Landelius; Sandra Sandra Andersson; Jelena Bojarova; Magnus Lindskog; Bengt Rydberg; Ronald Scheirer

Outline

- Introduction
- Current status of HARMONIE-AROME data assimilation
  - Algorithms
  - Observation handling
  - Artificial Intelligence and Machine Learning
  - Diagnostics and verification
- Challenges and vision
- Summary and conclusions

Tribute to Nils Gustafsson

Nils Gustafsson (1942-2024)

SMHI Rymdstyrelsen Swedish National Space Agency

Workshop
on
Perspectives of data assimilation on hecto-metric scales

In Memoriam of Nils Gustafsson

Tofta, Gotland, Sweden, 10- 12 September, 2024

Loïk Berre (remote), Jelena Bojarova, Pau Escribà, Elias Holm, Elias Holm, Heikki Järvinen (remote), Tomas Landelius, Magnus Lindskog, Kristian Mogensen, Patrick Samuelsson, Michael Tjernström, Ole Vignes, Tomas Wilhelmsson, Xiaohua Yang

Longer term Planning and Vision

- Recent workshop report

- ACCORD-cooperation longer term strategy document

Introduction

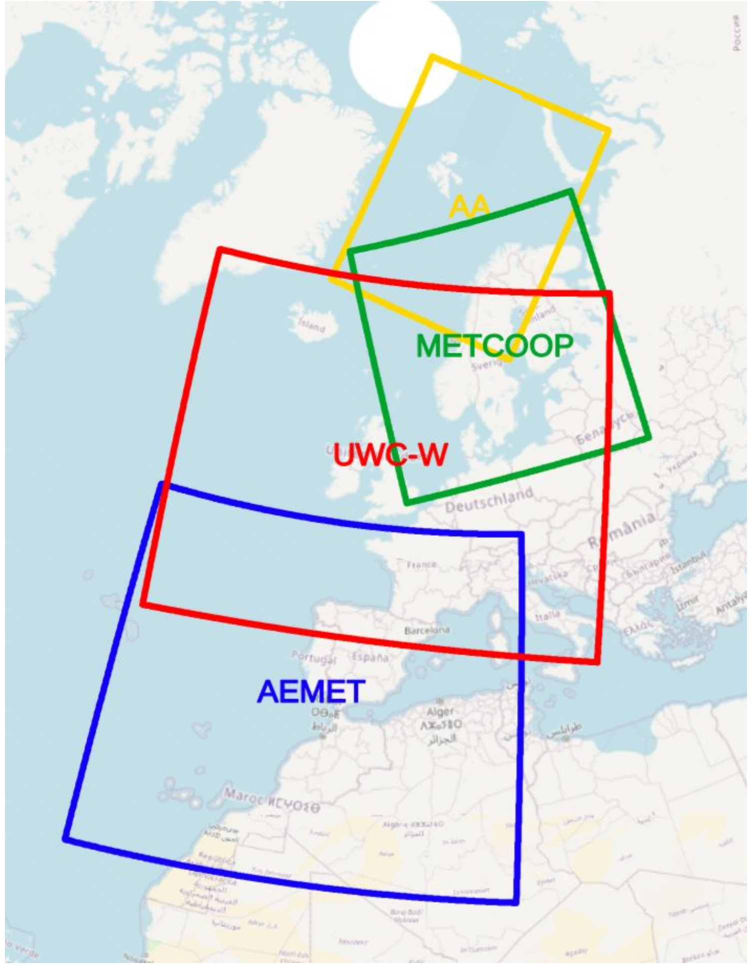
Brief History

- The HIRLAM cooperation started in 1985.
- Close collaboration with ALADIN consortium on km-scale modelling started in 2004.
- Sub-consortium of ACCORD was formed in 2020.
- Transition to UWC (United Weather Centers) from 2025.

Introduction

- ACCORD research and development consortium consisting of 26 countries for convection-scale limited-area modelling.
- Sub-consortia of ALADIN, LACE and HIRLAM.
- HIRLAM flavour of common modelling framework is referred to as HARMONIE-AROME. It consists of HIRLAM quality assured modelling framework containing source-code and scripts prepared for operational use.
- HARMONIE-AROME used by several operational centres, including MetCoOp, UWC-W, AEMET and AROME-ARCTIC.

HARMONIE-AROME in operations



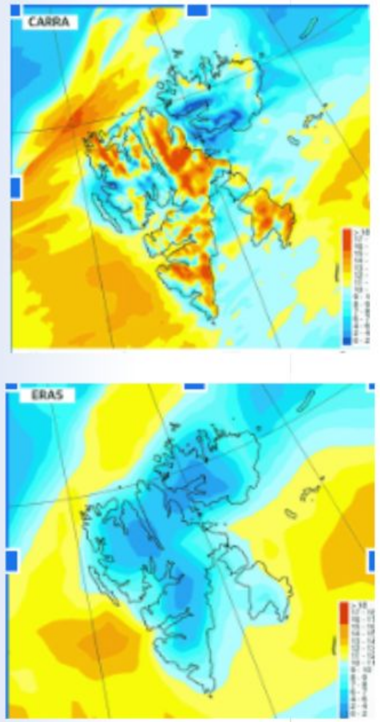
- Regional system (HARMONIE-AROME) for high resolution NWP. Horizontal grid-distance 2.0-2.5 km, 65-90 vertical levels.
- MetCoOp Ensemble-system with 30 ensemble-members (UWC-W in preop).
- Forecasts are produced each 1-3 h up to ~70 h.
- NWP-based nowcasting suites for shorter range forecasts (~up to 12 h).
- 3D-Var with conventional types of obs., radar, (GB) GNSS ZTD, Mode-S EHS, GPS RO, satellite based ASCAT, MW and IR radiances used (still clear-sky approach).

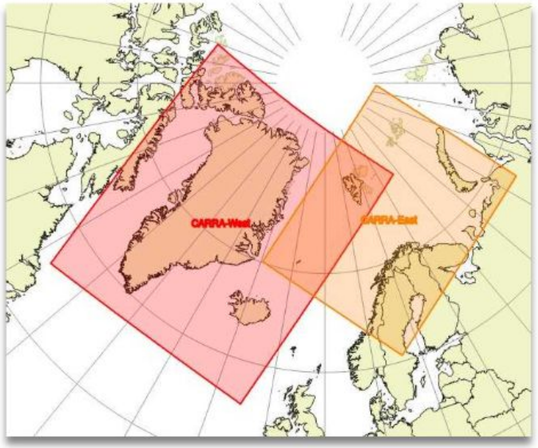
HARMONIE-AROME in several re-analysis (and other) projects

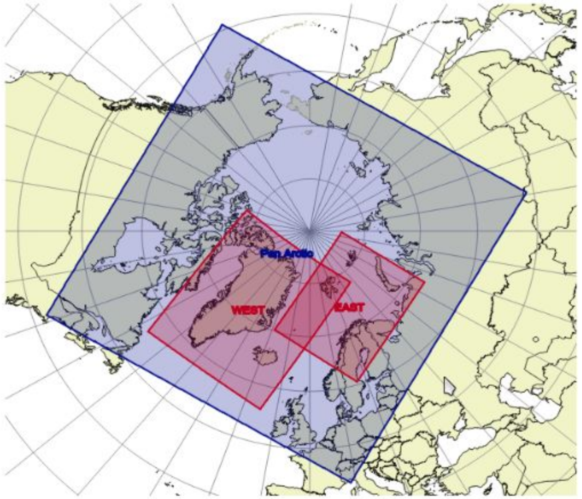
Example:
CARRA (right)
and
CERRA (HARMONIE)
reanalyses

Climate Change

The CARRA: a km-scale reanalysis data set: an Arctic magnifying glass of ERA5







- (1991-2021, finishing now)
- Near real time stream to follow
- data available on Copernicus Data Store (CDS)

One-year Pan-arctic "Pilot" for YPP to test for the next generation Arctic reanalysis (3,75 km, hydrostatic, AROME)

DMI
Norwegian Meteorological Institute
ECMWF
opernicus
Europe's eyes on Earth
European Commission

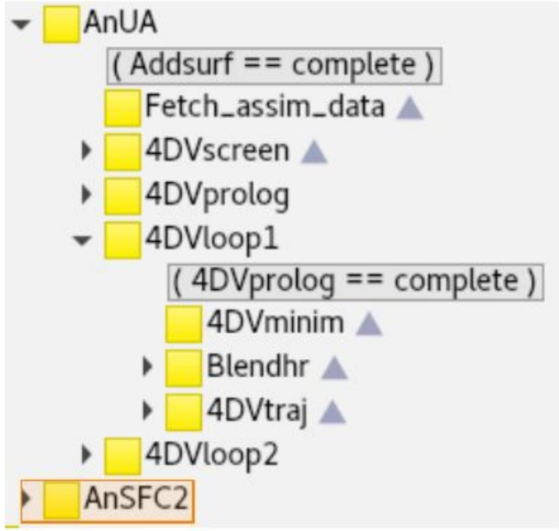
HARMONIE-AROME 4D-Var

- **Running daily at UWC-W**
  - United Weather Centers West - Denmark, Iceland, Ireland, the Netherlands.
  - Daily running suite.
  - 2 km horizontal resolution, 90 levels, 2 outer loop iterations
  - HARMONIE Cy43. 3-hour assimilation window.
  - On-going work with thinning of Mode-S EHS.
- **AEMET 4D-Var parallel runs and case studies**
  - Parallel 3D-Var versus 4D-Var runs aiming at optimising 4D-Var settings.
  - Valencia flooding case studies.
- **Porting to into object oriented OOPS framework**
  - Will pave the way for future research on fully coupled data assimilation.

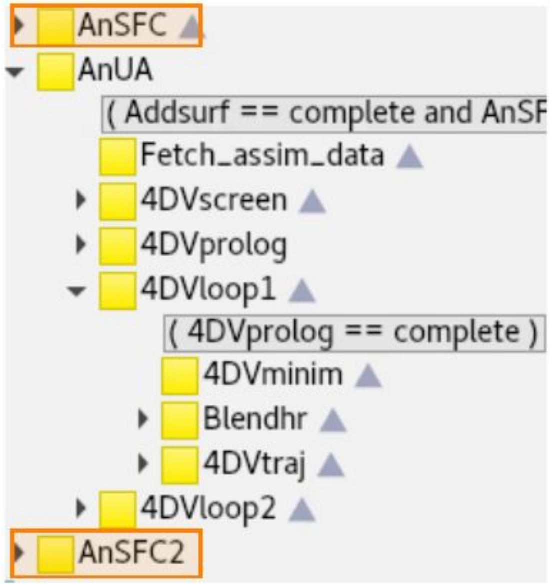
Towards a weak surface-upper air coupling in HARMONIE-AROME 4D-Var

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

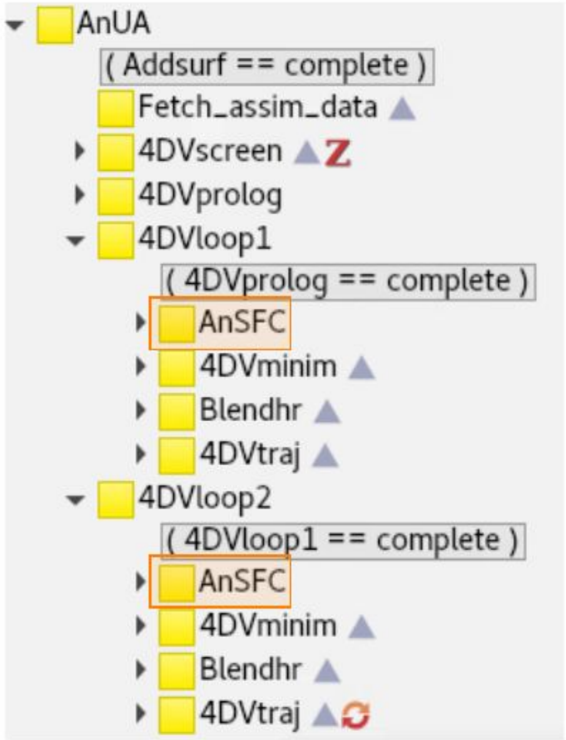
Reference: SA after 4DVar



Before and after 4DVar

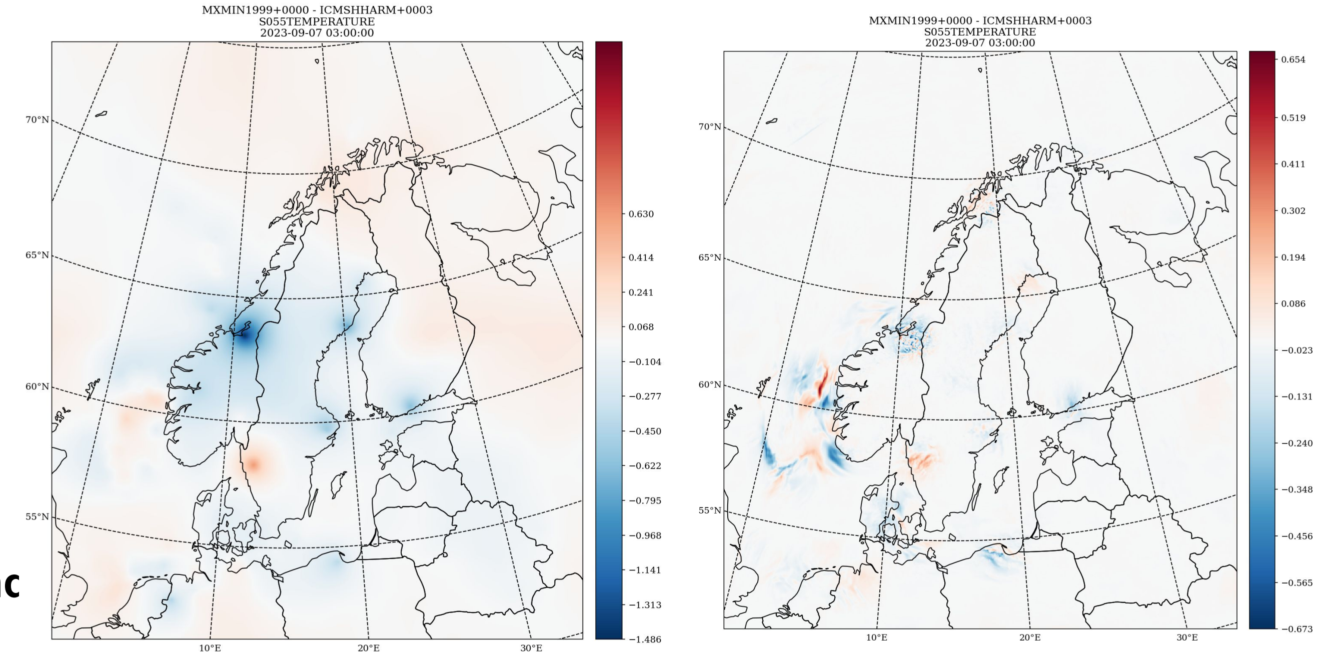


Inside outer loops



Ensemble Variational (ENVAR)

- **Ongoing exploitation of ENVAR approaches in HIRLAM**
- **No need for tangent linear and adjoint nor static B-matrix.**
- **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 extenc studies with refined system planned.**
  - Some work on localisation tuning in cy49



Temperature assimilation increments (K) at lowest model level (65) for one particular assimilation cycle. 3DVAR (left) and ENVAR (right).

HYBRID-Combining static B with an ensemble

$$J\left(\delta \boldsymbol{x}_{\mathrm{var}}, \boldsymbol{\alpha}\right)=\beta_{\mathrm{var}} J_{\mathrm{var}}\left(\delta \boldsymbol{x}_{\mathrm{var}}\right)+\beta_{\mathrm{ens}} J_{\mathrm{ens}}(\boldsymbol{\alpha})+J_{\mathrm{o}}$$

$$\frac{1}{\beta_{\mathrm{var}}}+\frac{1}{\beta_{\mathrm{ens}}}=1. \qquad \delta \boldsymbol{x}=\delta \boldsymbol{x}_{\mathrm{var}}+\kappa \sum_{k=1}^{K}\left(\boldsymbol{\alpha}_{k} \circ \delta \boldsymbol{x}_{k}^{\mathrm{ens}}\right)$$

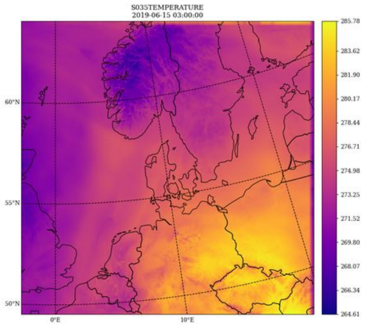
$$J_{\mathrm{ens}}=\frac{1}{2} \boldsymbol{\alpha}^{T} \mathbf{A}^{-1} \boldsymbol{\alpha} \qquad \delta \boldsymbol{x}_{k}^{\mathrm{ens}}=\boldsymbol{x}_{k}^{\mathrm{ens}}-\overline{\boldsymbol{x}^{\mathrm{ens}}}$$

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

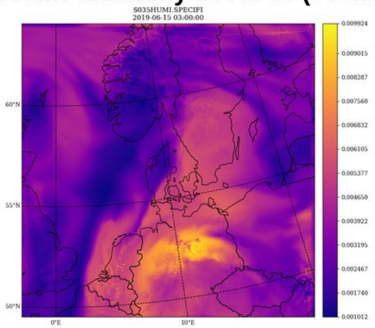
A constructed perturbation framework was implemented in the Harmonie HybridEnVar system to gain a better insight into the error growth mechanisms.

Background state (+03h forecast)

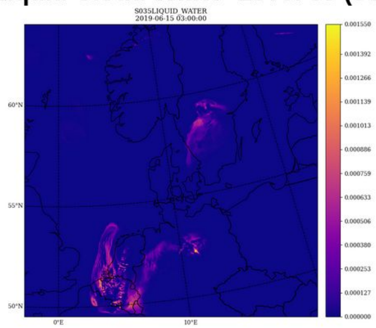
Temperature Level 35 (500hPa)



Specific Humidity Level 35 (500hPa)

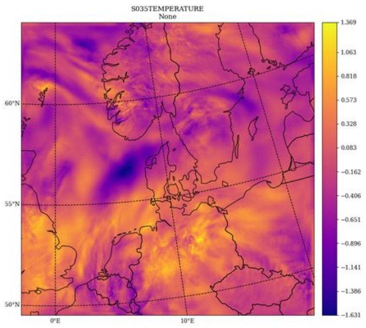


Liquid Cloud Water Level 35 (500hPa

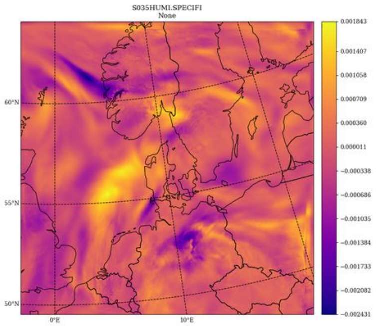


Analysis increment (Hybrid 3DEnVar scheme)

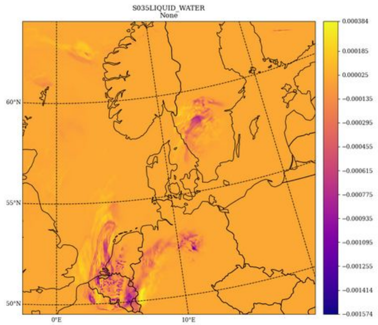
Clim + Ens component



Clim + Ens component



Ens component only



Other ongoing algorithmic activities for assimilation and initialisation

- 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

Observation usage

- synop, ship, aircraft, buoys, radiosondes.
- crowd-sourced observations (netatmo, wow, smartphone).
- Weather radar (Doppler winds and reflectivities).
- ground based GNSS, Mode-S EHS, hot-air balloons, drones.
- Satellite MW radiances, satellite IR radiances from polar and geostationary satellites, GPS RO, Aeolus HLOS.
- Satellite based products (AMV, ASCAT and cloud products).

Observation handling

Observation handling components

- Get access to observations (in time for limited-area data assimilation)
- Produce Model counterparts of observations, e.g. for Slant total delay GNSS
- Handling of observation errors
  - Estimate and model observation error characteristics
  - Correct for biases in observations, e.g. VarBC for LAM
  - Reject observations affected by gross errors by quality control

Observation collection by Kasper S. Hintz

Getting SPOs accessible for NWP applications in Europe

The SMAPS framework can be included as a module. SMAPS computes the observation and returns the observation when a new one is ready. The module is being used in the DMI app operationally.

- SDK (Software Development Kit) accesses barometer with consent from user
- Barometer is there, to track altitude changes (height curves after biking/running)
- Unfortunately, GPS altitude has low accuracy, so determining the altitude of the phone is the hard part.

Observation density heat map
2023-08-07
07:50 - 08:00 UTC

7785 observations

External S3 Bucket
- Access the data

- Data shared via an S3 bucket since 6th September 2023
- Data stored in 15 days, before being deleted
- To use command-line, use AWS-CLI (pypi.org/project/awscli/)
  But many S3 interfaces exist.
- You will need credentials to access. Contact me to get those.

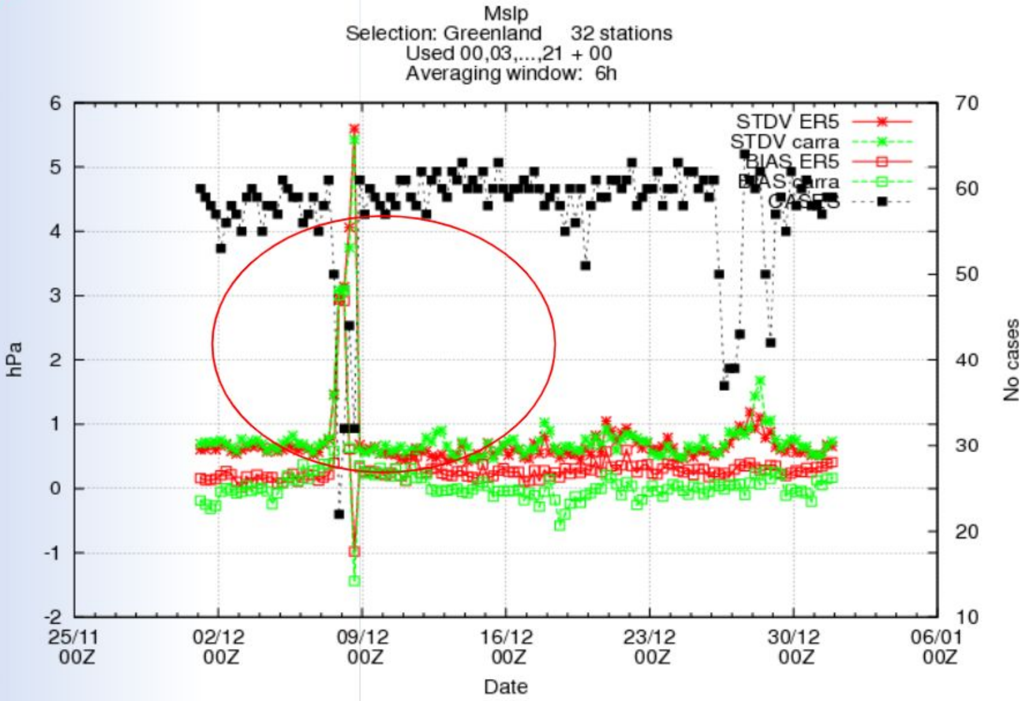
Observation handling: (rejection connected to) large model error

Importance of Observation Quality Control Procedures

- Tuning of current system
- Introducing new approaches

Climate Change

Known deficiency: analysis bust cases



CARRA data series contain occasionally cases with large deviation to the verifying observations for key parameters (analysis bust cases), where substantial amount of good observation data got rejected due to large model error.

Future phase of CARRA reanalyses shall hopefully improve

DMI Norwegian Meteorological Institute ECMWF opernicus European Commission

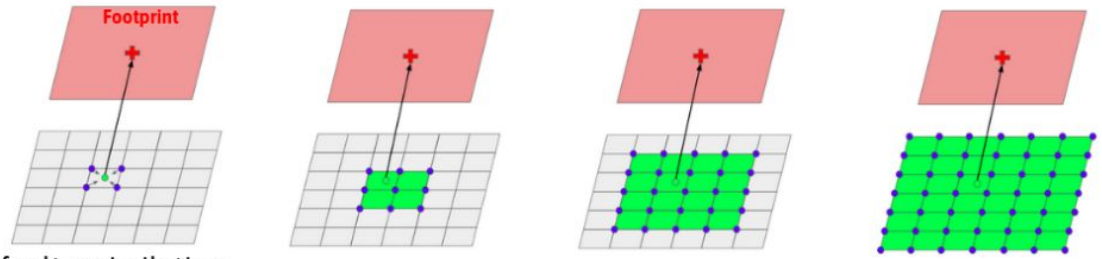
Observation handling: Satellite observations

Satellite-based observations

- Better handling of currently used instruments
  - Near surface channels (surface emissivity and reflection)
  - Taking scale differences between observation and model into account.
- Improved satellite based products and optimal use of them (Cloud Products and AMV).
- Towards All-sky radiance data assimilation.

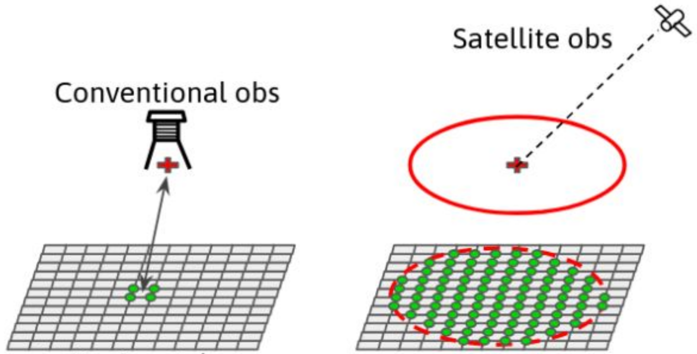
Observation handling: accounting for low resolution observations

Scatterometer supermodding and the radiance footprint operator



Default assimilation as **Point observation**

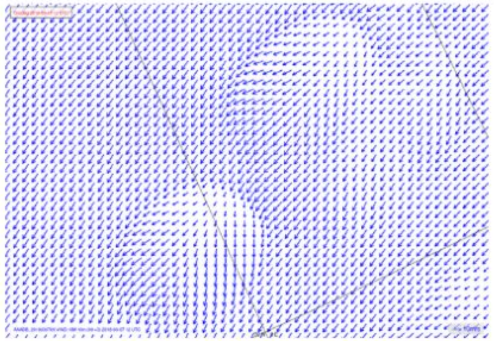
Averaging in model space **Supermodding**



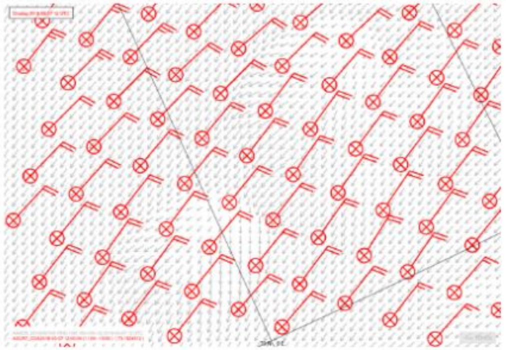
Horizontal interpolation

Averaging in observation space

AROME model on full resolution (too strong convection, wind downburst)



ASCAT cannot provide such small-scale information to constrain AROME wind fields



AMSU-A single obs + footprint

Simulated Tb

Retrieved emissivity

Artificial intelligence and Machine Learning

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

(see also talk by Tomas Landelius)

Diagnostics and verification

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

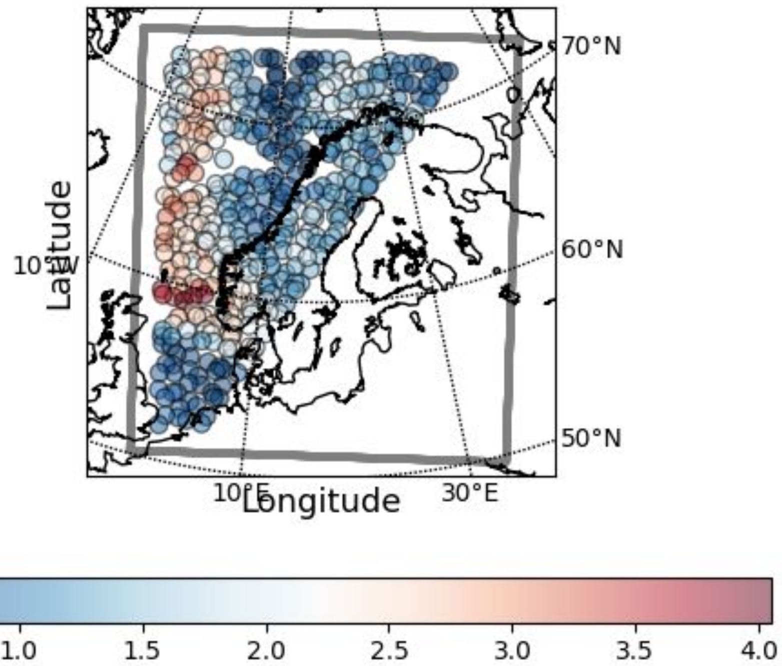
N sample size

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.

($\delta x = U\chi$ $\quad J_b = \frac{1}{2}\delta x^T B^{-1}\delta x = \frac{1}{2}\chi^T \chi$)

Example

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



unit: K

(corresponding observation error standard deviation is 1.8 K)

Diagnostics and verification

Data assimilation based verification

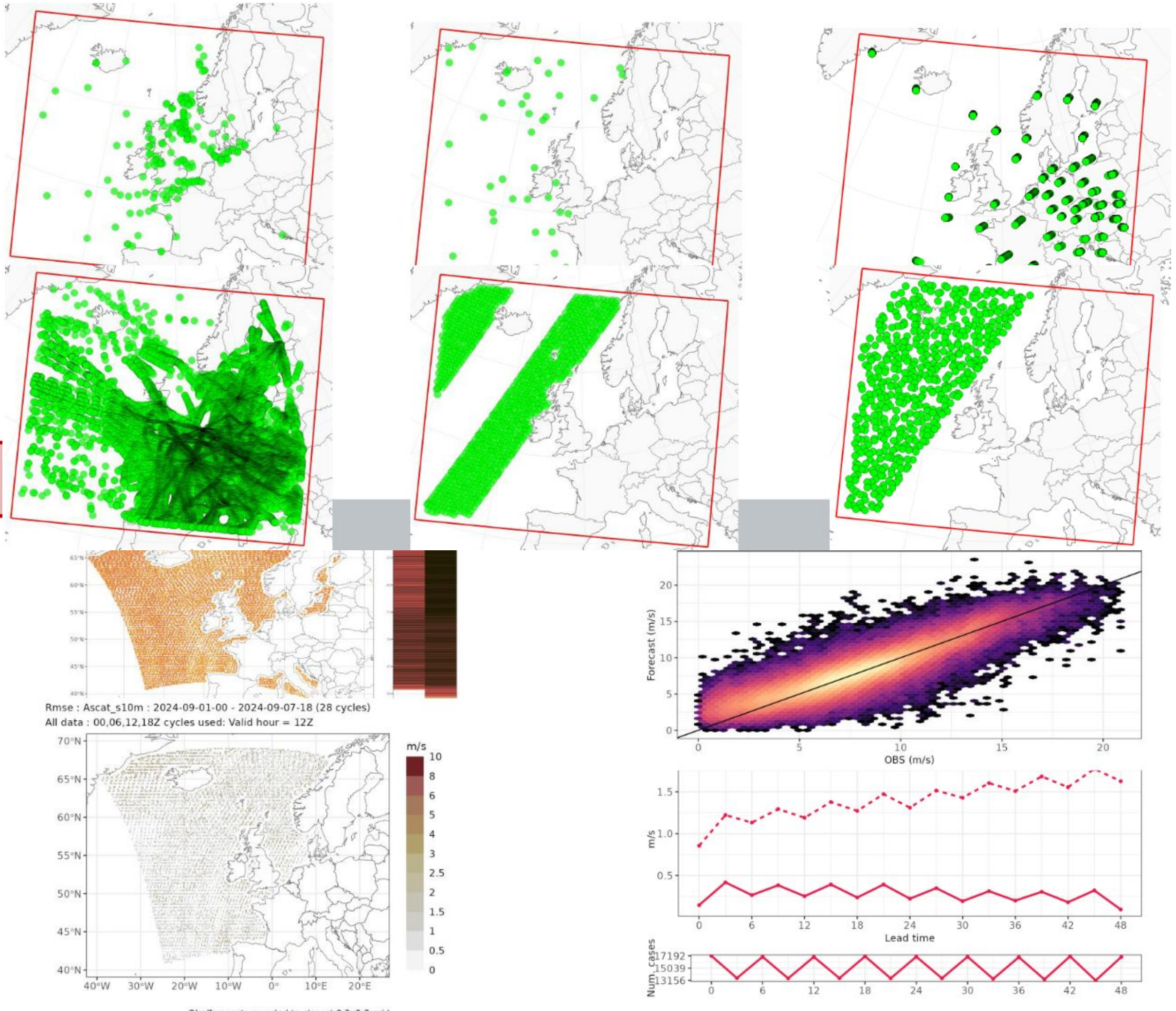
In DA we apply a variational formalism that consists in finding the best possible initial state, $\mathbf{x^a}$, by **minimising** a penalty function, J:

$$J = J_b + J_o = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(H\mathbf{x^b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(H\mathbf{x^b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y})$$

J_b measures the distance of the state to be derived, $\mathbf{x}$, to the background state $\mathbf{x^b}$
J_o measures the distance to the different types of observations.
$\boldsymbol{\delta}_\mathbf{x}$ represents the assimilation increments added to the background state, $\mathbf{x^b}$, to form the

Observation operator, *H*, used in DA can also be used for model verification!

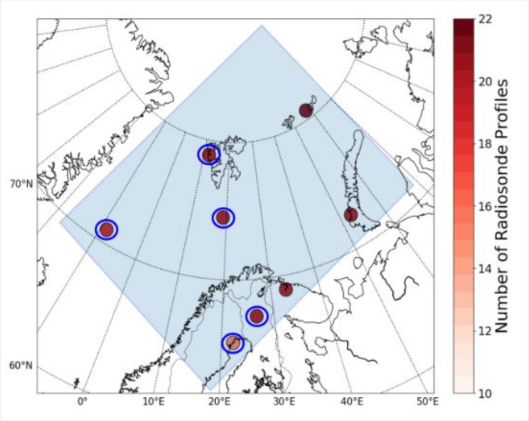
The observation operator, *H*, provides the link that projects the model state to all types of observations used in the data assimilation to enable comparison between the model state (background or analysis) and the observations.



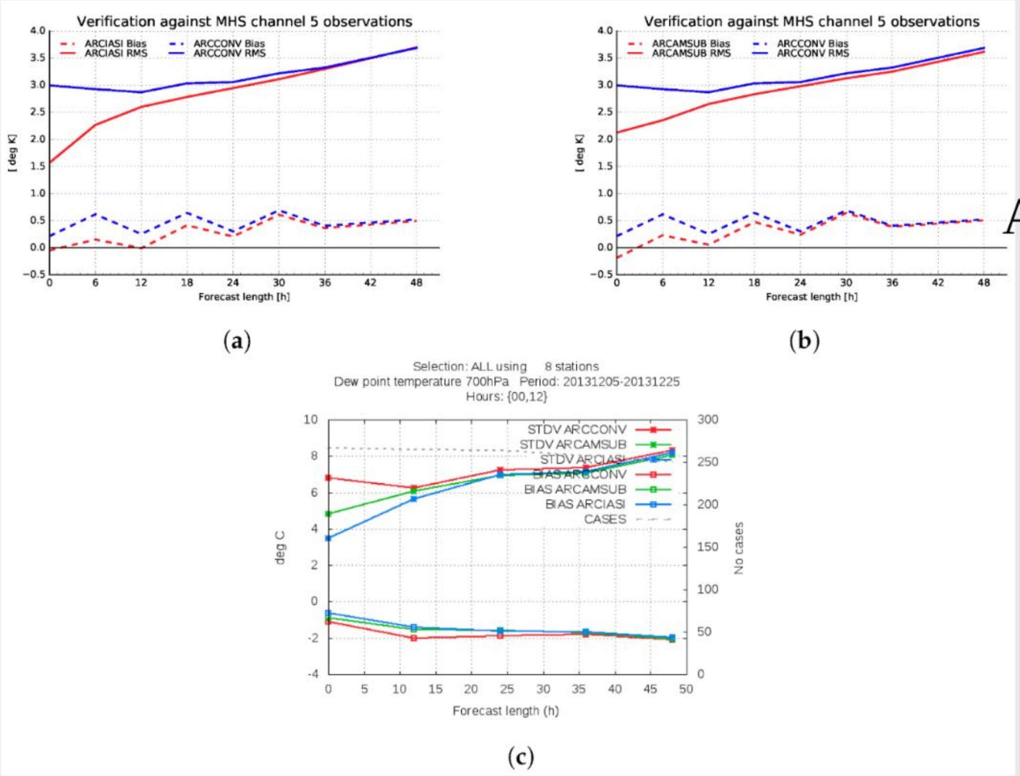
Diagnostics and verification

Verification against all available observations

Same verification period in both cases



AROME-Arctic and radiosonde locations



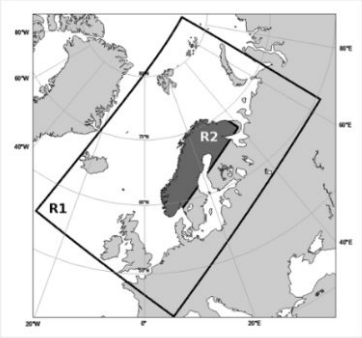
Against radiance, AMSU-B/MHS

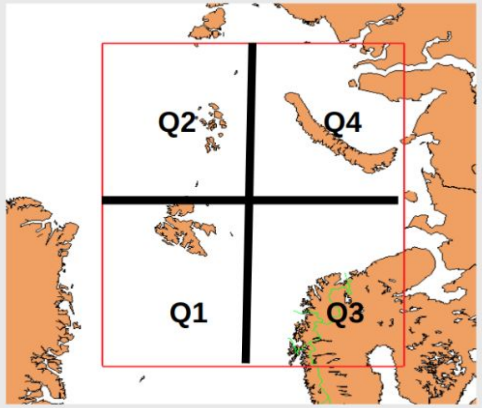
Against radiosonde

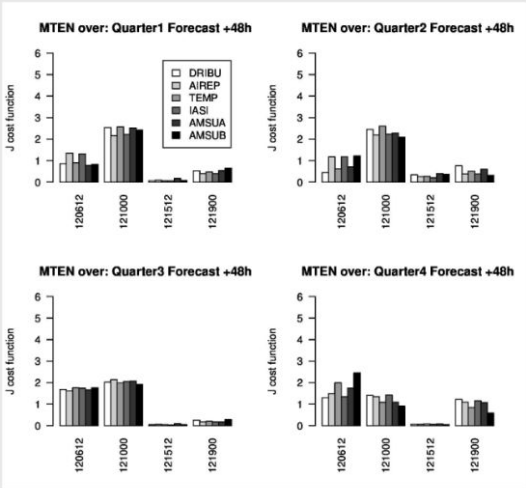
Diagnostics and verification: MTEN– Moist Total Energy Norm

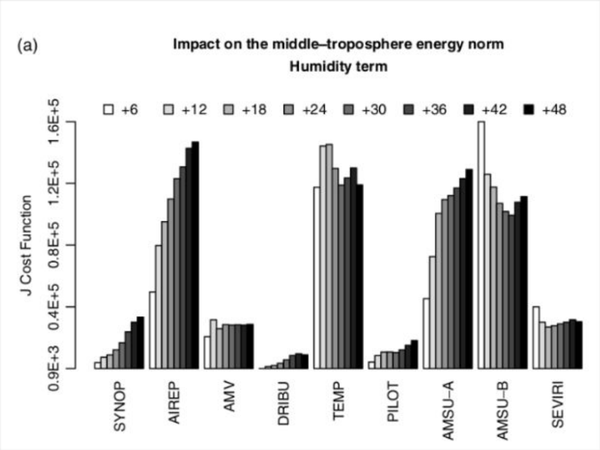
Sensitivity of the forecast model to the assimilated observations

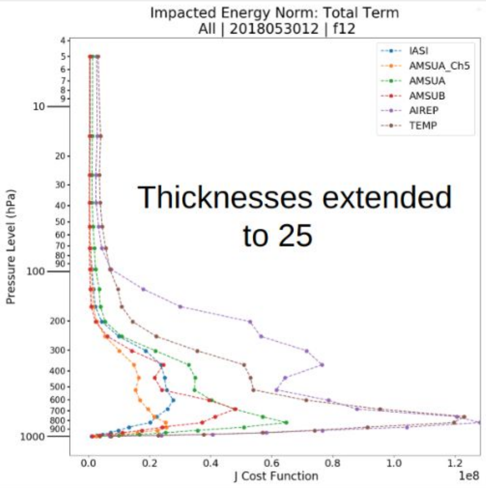
Vertical region	Region Bottom	Region Top
Low-troposphere	850 hPa	600 hPa
Middle-troposphere	600 hPa	350 hPa
High-troposphere	350 hPa	150 hPa
Stratosphere	150 hPa	20 hPa





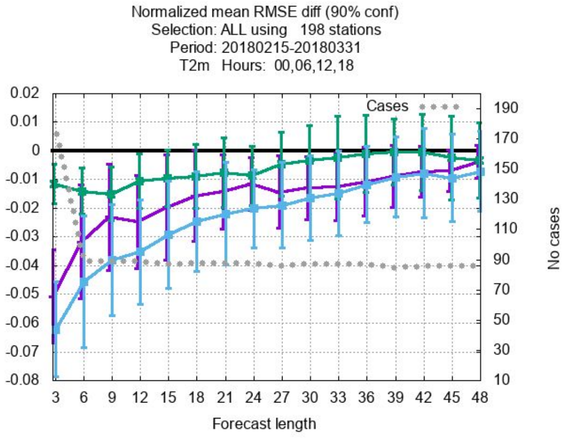






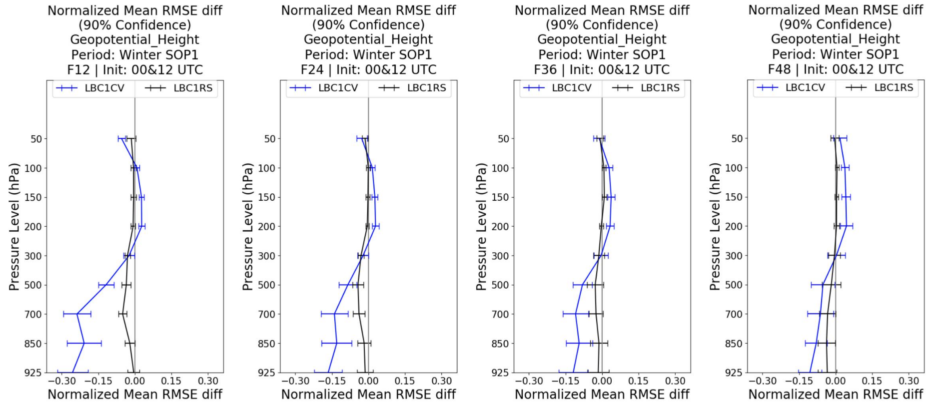
Impact of observations in LAM 3D-Var

Impact of observations through regional DA and LBCs



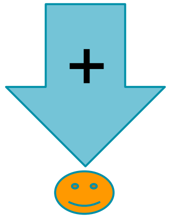
Surface

Conventional



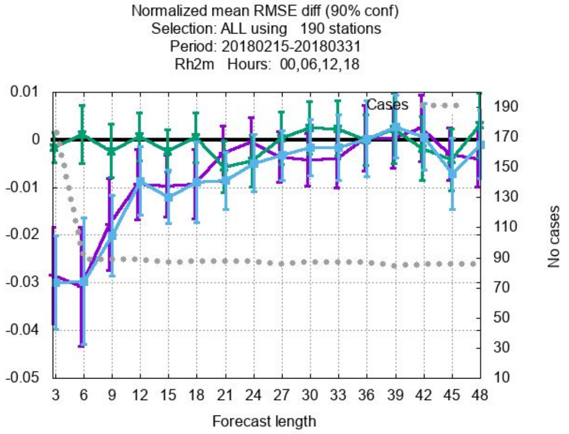
Upper air
Impact through
LBC

Conventional
Radiosonde

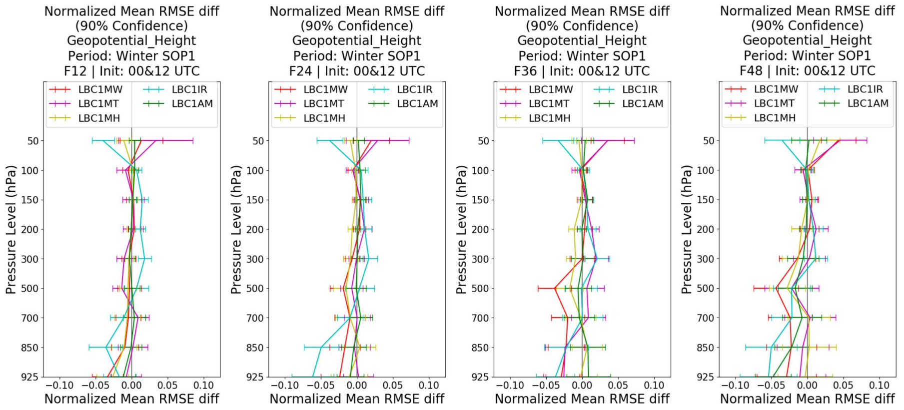


SOP1

Impact through:
Regional DA
LBCs
and
Total impact



Microwave
ATOVS



ATOVS, IASI,
AMSU-A, MHS,
AMV

Challenges and vision

Nils Gustafsson (1942-2024)

SMHI Rymdstyrelsen Swedish National Space Agency

Workshop
on
Perspectives of data assimilation on hecto-metric scales

In Memoriam of Nils Gustafsson

Tofta, Gotland, Sweden, 10- 12 September, 2024

Loïk Berre (remote), Jelena Bojarova, Pau Escribà, Elias Holm, Elias Holm, Heikki Järvinen (remote), Tomas Landelius, Magnus Lindskog, Kristian Mogensen, Patrick Samuelsson, Michael Tjernström, Ole Vignes, Tomas Wilhelmsson, Xiaohua Yang

Longer term Planning and Vision

- Recent workshop report

- ACCORD-cooperation longer term strategy document

Challenges and vision

Shorter term (2-3 years time scale)

- **Ensemble methods allow non linear effects of observation and model uncertainties to be simulated in data assimilation and forecasts.** They are important to consider in order to optimise the impact of highly dense observations with flow-dependent error covariances in high resolution DA and also in order to initialise associated ensemble predictions.
- **Improvement of boundary information within the data assimilation procedure is needed, regarding both coupling procedures and use of observations**.
- The Limited Area Modelling (LAM) community should tighten relationships with global Earth Modelling. The quality of LAM forecasts will always depend on the quality of boundary conditions provided by the host model.
- It is seen as beneficial to focus development towards coupled atmospheric-land-ocean data assimilation systems.
- **There is room for innovation regarding improvements of observation quality control and bias correction procedures, possibly utilizing elements of machine learning**.
- To benefit from the data assimilation also the model improvements are needed of particularly vertical profiles and in stable conditions. Such model improvements could benefit from objective model parameter tuning.
- **Verification scores emphasizing high impact weather are needed.**
- **Continuous data assimilation procedure with progressively increasing data assimilation window provides an opportunity to allow for late arriving observations in advanced data assimilation procedures.**
- Modular and transparent code structure that is generic enough to be executed on different computer architectures is needed in order to utilize the best technological advances, for example to combine physics-based and statistical models.

Challenges and vision

Longer term (longer than 3 years time scale)

- Insufficient understanding and handling of spatially correlated observation errors hampers hecto-metric scale data assimilation.
- While ensemble approaches are already operational for km-scale NWP, ensemble forecasting at higher resolution implies high computational cost, depending on future available computer power, target resolution and domain sizes. Variety of approaches can be explored to optimize ensemble generation, including multi-resolution ensemble, multi-scale Monte Carlo, generative AI models to enlarge ensemble size.
- Assimilate aggregated entities (e.g. moments) to reduce position errors and obtain easier correlated error structures.
- Use of AI approaches like diffusion or normalizing flow to turn non-Gaussian errors into Gaussian ones to allow for efficient DA methods.
- Employ AI auto-encoders to establish the physical balances at the hectometric scale and constrain the DA solution to obey these.
- Exploit possibility of obtaining auto-differentiation when converting CPU codes to GPU ones. Open up for having the full non-linear and the tangent-linear models in the cost function.

Summary and conclusions

- Both algorithmic developments and an extended observation usage are important for improving high resolution data assimilation system.
  - Flow-dependent data assimilation algorithms are under evaluation and require high quality ensemble system.
  - A large amount of various types of crowd-sourced observations provide an opportunity for future use in data assimilation.
  - All aspects of the data assimilation system, including quality control of observations and feedback from the regional reanalysis activities.
- Machine-learning provides an opportunity in various aspects of data assimilation.
- Several remaining challenges in data assimilation
  - e.g. representation of observation error correlations.
- Important to have verification metrics relevant for high-resolution data assimilation and to emphasize high-impact weather.

Thank you for your attention