Report from 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,
Heikki Järvinen (remote), Tomas Landelius, Magnus Lindskog,
Kristian Mogensen, Patrick Samuelsson, Michael Tjernström, Ole Vignes, Tomas Wilhelmsson, Xiaohua Yang

Organized in the context of the qCONDOR (a Quasi CONtinuous Data assimilation for nOwcasting and a very short Range forecasting) project of the Swedish National Space Board (SNSA) and is sponsored by the International Meteorological Institute (IMI) Visitors Program of The Meteorological Institute at Stockholm university.

Contents

Workshop scope

Scientific topics

Workshop recommendations

Agenda

In Memoriam of Nils Gustafsson

Workshop scope

This workshop is focusing on data assimilation for hectometric scales. How can we build on previous experiences from larger scales and what are the key aspects to address when approaching higher resolutions? Here we will try to take a variety of different aspects into account, including assimilation algorithms, resolution versus domain size, uncertainty estimates and modeling,
observation handling, coupling strategy, model physics, model spin-up, optimal length of assimilation window as well as computational aspects. How should the efficient data assimilation algorithm be designed? How can we profit in the best way from rapidly emerging technologies of data processing and analysis, known as Machine Learning (ML) or Artificial Intelligence (AI). Do we need the data assimilation at all on the hecto-metric scales? The workshop was organized in the context of the qCONDOR (a Quasi CONtinuous Data assimilation for nOwcasting and a very short Range forecasting) project of the Swedish National Space Board (SNSA) and was sponsored by the International Meteorological Institute (IMI) Visitors Program of The Meteorological Institute at Stockholm university. The primary activity IMI is to support scientific visits to Sweden within the atmospheric, oceanic and climate sciences and closely related fields. Roughly 50 scientific visits are made possible each year through funding from the institute.

Scientific topics

- Nonlinearity and non-gaussianity of the error evolution at hecto-metric scales and appropriate data assimilation algorithms
- Machine Learning and Artificial Intelligence
- Computational aspects
- Uncertainty estimates and modeling of these
- Predictability, model spin-up and adjustment processes
- Resolution, domain size, update frequency and length of the assimilation window
- Observability at hectometric scales, observation handling and quality control
- Process-oriented modeling: atmosphere – land surface – ocean interactions and challenges of coupled data assimilation
- Model development (vertical)+opt model
- Quality/success ratio

Workshop recommendations

Short summary of main findings

Here we provide a short summary of main findings. These could be addressed from the perspective of short (up to two years, S) and long (five years, L) time scales of the developments.

- There is room for innovation regarding improvements of observation quality control and bias correction procedures, possibly utilizing elements of machine learning. (S)
- Improvement of boundary information within the data assimilation procedure is needed, regarding both coupling procedures and use of observations. (S)
- It is seen as beneficial to focus development towards coupled atmospheric-land-ocean data assimilation systems. (S)
- 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. (S)
- Verification scores emphasizing high impact weather are needed. (S)
- 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. (S)
- 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. (S)
- Continuous data assimilation procedure with progressively increasing data assimilation window provides an opportunity to allow for late arriving observations in advanced data assimilation procedures. (S)
- 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. (S)

- Insufficient understanding and handling of spatially correlated observation errors hampers hecto-metric scale data assimilation. (L)
- 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. (L)
- Assimilate aggregated entities (e.g. moments) to reduce position errors and obtain easier correlated error structures. (L)
- Use of AI approaches like diffusion or normalizing flow to turn non-Gaussian errors into Gaussian ones to allow for efficient DA methods. (L)
- Employ AI auto-encoders to establish the physical balances at the hectometric scale and constrain the DA solution to obey these. (L)
- Exploit possibility of obtaining auto-differentiation when converting CPU codes to GPU ones. Opens up for having the full non-linear and the tangent-linear models in the cost function. (L)

Extensive report

Introduction

The workshop was organized as a three day event. The first and the last days were devoted to a limited group of invited experts. These experts were selected to provide knowledge in different areas relevant for the workshop and to complement each other. The second day was open to on-line participation and also contained a session devoted to Nils Gustafsson, who passed away Sunday 21 July 2024. Nils impressive contributions to limited area NWP was highlighted by the workshop participants.

There were 13 presentations covering various aspects of data assimilation at hecto-metric scales. Each of these were associated with follow on questions and discussions. In addition there was room for discussion at the end of each session. The last day onsite workshop participants were organized to, based on discussions and presentations, come up with workshop main findings and recommendations.

The focus of the qCONDOR project is the improved cloud analyses and the improved quality of the very short range forecasts. It is evident that hecto-metric modeling is important to adequately model cloud-related processes. At the same time the experience with the data assimilation at the hecto-metric scales is very limited and quite fragmented. The purpose of the workshop was to bring together available knowledge in that area addressing it from different angles of view and systemize the knowledge into a number of recommendations of how to tackle the data assimilation on hecto-metric scales. Other aspects of the hectometric modeling challenge are thoroughly discussed in a recent review scientific paper [1] where both NWP and climate applications were considered . One of the conclusions of the paper is that data assimilation on hecto-metric scales is in its infancy at the same time as hecto-metric modeling becomes more and more relevant. This workshop is a trial to outline strategic directions of the research to tackle data assimilation on hecto-metric scales.

The success story of HIRLAM 4D-Variational data assimilation has shown that hard systematic dedicated well-structured coordinated work is needed to achieve the result and that it takes time. Theoretical ideas have to be developed with a strong focus on operational applicability and using the best possible technological advances. The weaknesses of any methodology have always to be high-lighted. Only in this way can progress be achieved. In the last years the focus of LAM has been more and more shifted towards very high resolutions. It is important to ask ourselves to what extent the existing methodologies of data assimilation are applicable on hecto-metric scales, what are the limitations of those techniques and what are the promising directions of developments.

Verification and model development

We see a benefit of hectometric scale data assimilation and modeling by taking into account details in surface, orography and small spatial variations. This can be visualized by improved statistical distributions of clouds and improved description of high impact weather. Therefore procedures and measures of success for data assimilation at hectometric scales need to be addressed in verification and validation. In addition, it is important to realize that if the forecast model is not good enough in representing clouds or stable boundary-layer conditions and affected by systematic errors, it will not be possible for the data assimilation to provide initial state corrections that will remain during the forecast. There is a potential for validation as well as verification to use tools from the data assimilation procedure. It is important to realize that observability of the Earth Systems is limited. Currently remote sensing instruments are dominating and provide indirect information about the processes. A good understanding of the various mechanisms through physics-based modeling is essential in order to extract potential from the observations. One of the areas of particular interest is the Arctic where the number of direct measurements of the atmospheric quantities is extremely small and the observing system is dominated by satellite observations. Lack of the understanding of the processes may lead to large systematic errors and inappropriate decisions.

Another area where particular attention is needed is the modeling of surface atmosphere interactions. Processes on hecto-metric scales are affected to a larger extent by surface heterogeneity. The initial state of the atmosphere and of the surface model are constructed in an inconsistent manner. Imbalances introduce spin-up of processes that degrade the quality of the very short range forecasts. One fundamental challenge is that the aggregated fluxes over the heterogeneous surface conditions are not correct. Dynamical effects occur in a form of the horizontal advection and the turbulent fluxes and provoke artificial processes. Beside that, increasing vertical resolution and lowering the lowest model level violates assumptions of the blending height theory. Another challenge is that the observations of soil processes are very limited in number. The focus on a more efficient use of remote sensing measurements on a variety of frequencies is needed, both for validation, verification and the data assimilation processes.

Algorithmic developments and coupled data assimilation

The LAM community needs to tighten relations with global Earth system modeling. Quality of the forecast provided by LAM will always depend on the quality of the boundary conditions provided by the global models, including lateral boundary conditions, lower boundary conditions such as sea surface temperature and ice. Smaller size of the domain calls for more attention to the treatment of the lateral boundary conditions and the host model information. The Coupled Data Assimilation system introduces a lot of complexity, and more frequent update of the boundary fields passed from the global model to the LAM might be a more appropriate solution. Coupled Earth systems that use statistical emulators should also be explored.

Hecto-metric scale models are unaffordably expensive for larger size domains at the same time when too small size of the domain will hamper the possibility to capture the processes in a realistic way. This is a dilemma that has to be addressed. It might be an advantage to run the data assimilation on intermediate km-scales and perform model forecasting on hecto-metric scales on demand. Flow-dependent forecast error statistics are necessary for data assimilation on high resolutions. Variational data assimilation provides a flexible framework for progressive system while utilizing the advantage of global treatment of observations. Introduction of ensemble based covariances allows for initialisation of a large variety of model components, including couplings with ocean and soil. Innovation-based statistics provide an efficient tool for diagnosis and tuning of ensemble covariances. At the same time, generation of ensembles at high resolution is costly. Statistical (AI) methods utilizing probabilistic inference provide an attractive alternative for generation of high resolution ensembles at affordable cost. Diffusion models look to be promising for generation of ensemble reflecting error-of-the-day structures, however high-resolution and thus huge model space dimension is hampering the efficiency. Possibly the combination of auto-encoders to reduce the dimensionality of the model space projecting it to a lower dimensional Gaussian space and the diffusion models applied in the reduced space might provide an affordable solution to handle non-linearities and non-gaussianities. Multi-scale Monte-Carlo methods that allow optimal sampling of multi-scale features might significantly reduce the cost of ensemble generation. Multi-resolution ensemble is one possible realization of the concept. The quality of the ensemble members might be significantly improved if they are recentred around the high-resolution deterministic run. Small number of extremely computationally expensive high-resolution ensemble members can be augmented with a large number of low-resolution members obtained at much lower cost.

The optimal length of the data assimilation window is determined by several processes. The data assimilation window should be short enough to stay within predictability limits and to be long enough to assure adjustment processes to happen, depending also on employed initialisation methods. Incremental Analysis Update and Digital Filter Initialisation are well established techniques to tackle adjustment processes on larger scales. Efficiency of the DFI on hecto-metric scales still needs to be proven.

One perspective to hecto-metric modelling is to see it as an essential link in the chain that delivers with minimal delay the most recent observational information to the end users, processed into user-oriented geophysical quantities. Post-processing to the highest resolution should be seen as an integrated component of the forecasting system. For the sub-km, cloud-resolving modelling, frequent and quick delivery is ever more important to add value for users by capturing weather phenomena at cloud scale. Continuous update with observations is thus an essential property of the high-resolution data assimilation, striving towards better utilization of very late arriving observations, such as those behind weak data links or in re-submission services of remote sensing data. One possible way forward is the Continuous data assimilation (Cont-DA). Typically, data assimilation computations commence after the observational cut-off time. In optimum interpolation (and Kalman filtering in general), this is a necessity since observation weights are computed once with the maximum observation coverage. Variational data assimilation, in contrast, is iterative by nature, and a strict cut-off time is not necessary. The minimization can thus commence before the cut-off time with (1) gradually increasing observational coverage from one 4D-Var outer loop to the next, and (2) progressively extending outer-loop time-range to accommodate newly arriving observations. The first aspect is relatively straightforward to implement in 3D/4D-Var, provided the forecasting system is equipped with continuous observation processing capability. An augmented observation data base is thus exposed to assimilation at each outer loop – during each new trajectory run observation departures are computed for all observations, new and old. The second aspect is much more demanding. Selection of increasing time-ranges of outer loops is a matter of application-specific tuning. It also leads to a system that is technically harder to maintain. A poor-man's solution here is to maintain only one outer-loop time-range and have initially no observations towards the end of the range. Priority in hector-metric data assimilation is thus to commence minimization early with a single-range 4D-Var and let in a gradually increasing number of observations.

Potential benefits of Cont-DA remain to be demonstrated in hector-metric data assimilation. Capability to launch minimisation well ahead of the cut-off time is an advantage, offering potential to move parts of the time-consuming minimisation workload outside the time critical, peak-power path. This helps to prolong the available wall-clock time window for the data assimilation workload. Everything being equal, this helps on timely delivery. Hector-metric data assimilation relies on parametrization of physical processes, and it is thus a relatively non-linear problem. This calls for many 3D/4D-Var outer loops, which Cont-DA provides at an affordable cost. This helps to capture non-linearities and offers new potential to alleviate physical problems of adjustment processes, such as in water cycle. Hector-metric data assimilation lacks natural balance constraints. Cont-DA provides a potential benefit via better conditioning of the minimization task. Finally, 3D/4D-Var does not parallelize well, first, because of the sequential nature of minimization, and second, because the inner loop has a relatively small problem size. In small problems, parallel communications easily dominate over computations. Splitting the computations over a longer wall-clock time window affords solving the problem slower on a lower number of processors and thus offering better parallel performance of 4D-Var.

On the other hand, one must be also aware that such DA implementations may lead for instance to significantly increased I/O, corresponding to many large components which, instead of being simply stored in memory, need to be written, archived and read by each outer loop iteration of the same DA window : control variable, first-guess, tens of preconditioning vectors (in addition to the background and observations, with possibly associated statistics). Such I/O implications are important, they have to be accounted for in the strategy for considering associated experimentations and implementations.

Moreover, the relevance and efficiency of this Cont-DA approach are also likely to depend on the target cycling frequency, DA window length and observation timeliness. Effects may be expected to be more pronounced for long-window DA (e.g. 12h at ECMWF) in large-scale NWP, with long cutoff times. This may be less the case for short-window DA (e.g. 1h or 3h) in most mesoscale DA systems, with associated cutoff times that are relatively short but still sufficient for the target short DA window, allowing almost all observations to be available and assimilated over the LAM area.

Observations and observation handling

Observation handling will be an important aspect of data assimilation at hectometric scales. Although there is not a one-to-one correspondence between model grid distance and required grid-distance they are related. Increased model resolution requires observation density with increased spatial and temporal resolution. Status of observation impact and requirements for km scale modeling were recently summarized in the final report of the 8th WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction and Earth System Prediction [2]. We emphasize a dependence of observation usage on how advanced the data assimilation algorithm (and model quality) is. For instance, it has been shown recently that massive Mode-S aircraft data had much more clear positive impact in AROME with 3DEnVar than with 3D-Var, in connection with a better representation of flow-dependent vertical correlations related to the tropopause height. This has been also demonstrated for the impact of highly dense crowdsourced observations in AROME 3DEnVar, in relation with sharp correlations driven by the meteorological situation and for extension of model control variables to include hydrometeors for cloud and precipitation sensitive observations. For regional data assimilation at hectometric scales weather radar, satellite radiances and Aircraft based Observations (ABO) provides important existing data sources together with GNSS and synoptic surface measurements of two-meter temperature and moisture and surface pressure together with scatterometer winds over sea surfaces. There is a need for additional Planet Boundary Layer (PBL) wind observation, giving information of vertical profile. Such observations could be provided by wind-profilers. We also see room for use of emerging data types, such as UAVs, smartphones, wind farms, netatmo and wow etc.

There are some key aspects regarding handling of observations that need attention. A proper handling of satellite radiances in cloudy areas and close to the surface are needed. That involves model parameterizations, data assimilation algorithms and radiative transfer models. This is however an active area of research with several ongoing developments, including all-sky data assimilation, estimation of surface emissivities by dynamical approaches, handling of surface reflection assumptions, particle size assumptions and also algorithmic extensions with flow dependent error covariances and extent of model control variables.

Handling is needed of scale differences between model and observations by averaging operators or representation of errors due to sub-grid scale variations. Such developments have already been initiated. Furthermore, it will be increasingly important to take into account the slant of the path between satellites and mapping of signals at the surface. Last but not least handling of observation errors and quality control are crucial for data assimilation at hectometric scales. Currently observation error correlations are handled by spatial thinning and combined with inflation of observation errors. Such a procedure prevents use of high resolution observations and deriving small scale variations. An improved understanding and representation of spatial and also temporal observation error correlations are needed. An increased observation amount of diverse quality requires an advanced quality control and bias correction. We see scope for improvement of methods currently used, despite sometimes being flow dependent. Here we could clearly benefit from AI. The challenge is to reject observations affected by gross errors that can deteriorate the forecast, but retain observations indicating high impact weather within the quality control procedure.

Adopting the continuous data assimilation (Cont-DA) approach would require more flexible observation processing with incoming stream of raw observations continuously quality controlled and added to the observation database. If this step works correctly, every new outer loop of variational data assimilation has access to the latest arriving observations. Software with this objective was developed at ECMWF in their Continuous Observation Processing Environment (COPE) project. Adoption of continuous observation processing strategy implies need for corresponding adaptations at national weather centers.

Machine learning and artificial intelligence

At hectometric scales, accounting for non-hydrostatic dynamics is crucial as nonlinear effects become more pronounced. The breakdown of quasi-geostrophic balances in convective situations renders traditional Gaussian noise and linear data assimilation methods less effective. However, the evolution of the atmosphere is still constrained by physics at these scales, and the manifold hypothesis can be useful in addressing the curse of dimensionality.

One approach is to employ AI approaches like autoencoders followed by normalizing flows or diffusion models to transform high-dimensional non-Gaussian processes into lower-dimensional Gaussian counterparts. This nonlinear yet Gaussian latent space can then be explored to apply efficient data assimilation using Sequential Monte Carlo with locally optimal proposals or score-based data assimilation. Alternatively, the feasibility of applying Koopman neural operators to learn linear projections of the latent space dynamics can be investigated, allowing for the application of conventional Kalman filters. Neural ODEs could also be used to model the Perron-Frobenius operator for the system, ensuring solutions comply with the Liouville equation. Neural operators could enable direct modeling of probability density functions, providing efficient solutions to cost-loss problems related to extreme weather without relying on ensemble methods.

Nonlinearity and non-Gaussianity of error evolution at hectometric scales and appropriate data assimilation algorithms

AI can also be used in conjunction with variational data assimilation methods if a differentiable model is available, either as an AI emulator or a physical model written in a language that supports auto-differentiation. This connects to ongoing code adaptation projects where HPC code is translated to GPU code, potentially facilitating auto-differentiation as a byproduct. With a differentiable model, the variational cost function can be minimized directly using efficient AI libraries. On the other hand, the use of a differentiable model may need to be questioned with respect to strongly non linear moist processes such as deep convection. DA formulations relying on an ensemble of non linear forecasts may be considered as a possible alternative in this context. Still the linear nature of analysis increment computed from an ensemble of non-linear forecast runs might result in an inconsistent state for strongly non-linear moist processes. In all cases, the error statistics still need to be determined and ensemble approaches can provide information on non linear and non Gaussian aspects of error propagation in DA cycling. The forecast error structures can be estimated using probabilistic models and/or generative methods (e.g., as in the DE_371 project). For the R matrix, it will likely become increasingly important to include correlations between measurements, as satellite and radar data will be correlated on the kilometer scale and below when thinning is no longer an option.

Using traditional variational models, it is important to keep nonlinearities small. The background should be as accurate as possible (frequent updates, a sufficient number of outer loops in 4D DA), variances and covariances should be also accurate and flow-dependent, and observations/models should be aggregated to a matching scale (space and time). Assimilating features (e.g., local statistical moments) could help make the error less prone to position errors.

Both variational and Kalman filter techniques are developed under assumption of Gaussianity and linearity. Assimilation of aggregated quantities is more prone to fit these assumptions. Data assimilation is a statistical technique that optimally integrates different sources of information based on their probabilistic inference. For Gaussian distribution all statistical information is captured in the mean and the error covariance. Machine Learning techniques allow statistical inference under less restrictive assumptions. For highly non-linear and non-gaussian problems data assimilation becomes an inherited part of the training procedure. The workshop focused on atmospheric state estimation but revisited also model parameter estimation. On hecto-metric time scales, the large-scale dynamics - the driving geophysical flow - is largely linear and the motion spectrum of interest is non-linear and dominated by strong interaction with physical processes. These are parametrized and contain weakly constrained closure parameters. It is important to keep in mind opportunities provided by algorithmic model parameter estimation to ensure best forecast skill within the capabilities of the chosen model formulation (called model selection - there is an infinite number of options to specify model parameters). Also, it is important to strive towards parameterization schemes which are linear with respect to closure parameter values inside the range of plausible parameter values. This greatly facilitates model selection using manual and algorithmic methods.

Computational aspects

The time-critical production of high-resolution regional forecasts using NWP models run by national weather services is approaching computational limits. Additionally, chaos theory limits the accuracy of weather predictions at increasingly higher resolutions in both time and space. To address this, probabilistic forecasting using an ensemble of model runs is employed, though this adds to the computational load. Generating multi-resolution ensembles by physical causality-based traditional models and/or ensembles using AI-based emulators could be an efficient way to utilize computational power.

AI offers a potential solution to this dilemma. In recent years, the NWP community has been surprised by the capabilities of AI models, demonstrated by big tech companies, which can rival physics-based NWP systems at both global and regional scales. Unlike operational NWP, which struggles with computational limits, AI inference can be run on a few GPUs in minutes. Even with more complex, high-resolution probabilistic AI architectures, inference remains much cheaper compared to today's physics-based models.

Training large Earth system foundation models will be impossible without collaboration. The EU Weather Generator project is a first step towards collaborating on training such a system. With this foundation in place, tailored AI applications (e.g., regional data assimilation and postprocessing) can be developed and trained with fewer resources, such as those available to national institutes.

The ECMWF AI pilot project and the ECMWF Anemoi AI code repository facilitate member state collaboration on AI for NWP, both for global and limited area models. This could bring ECMWF member states closer to state-of-the-art developments at ECMWF and offer win-win collaborations. Member state contributions can enter the common system much faster, and member states, in turn, gain quicker access to ECMWF developments compared to the current Integrated Forecasting System (IFS).

Collaboration will be even more crucial when developing models for the hectometric scale. High-performance computing costs will be substantial, even with GPU adaptation. One option is to focus more on common physical models, while still maintaining different options to avoid the “all eggs in one basket” syndrome.

Collaborative efforts on tools for interfacing between Python and Fortran, and translating from Fortran to Python, will be essential.

Collaboration is also vital for accessing more observations at the hectometric scale. The NWP community could present a united front when negotiating access to citizen and company-owned data (e.g., from cars, air and sea traffic, non-public satellites).

Regardless of the model used, it is important that they are accompanied by an environment that makes them easy to configure and experiment with. It is essential with the modular and transparent code structure that is generic enough to be executed on different computer architectures. One way to achieve this is to separate the scientific part of the code from its technical implementation.

Uncertainty estimates and how to model them

Uncertainty estimates are crucial for both modeling error covariances in variational data assimilation and for assessing the uncertainty of the forecast itself. The flow-dependency has been largely proved in different DA systems (from low to high resolution) to be a successful aspect to get accuracy in the background errors estimation. On the other hand, techniques such as singular vectors can be utilized to exploit the tendency of chaotic systems to confine error growth within a smaller subspace than the full system's dimension. Tracking a relatively low-dimensional unstable manifold with a finite ensemble is more feasible than describing errors in the fully dimensional space. This approach is akin to using autoencoders and can similarly be combined with diffusion and normalization techniques to achieve Gaussian error structures.

Generative AI models can also be employed to enhance small ensembles or even a single deterministic forecast. This technique helps in better estimating dynamic background error covariances and generating a larger number of forecast ensemble members.

Predictability, model spin-up, and adjustment processes

As an initial step, we propose using 3D (FGAT) EnVar and incremental updates. The assimilation window can be gradually extended, introducing more observations in a manner that allows for issuing a time-critical forecast based on the available observations at that moment. Subsequently, another forecast can be issued with additional observations for less time-critical applications. This stepwise refinement, similar to outer loops, could potentially make the problem less non-linear.

References

[1] Humphrey W. Lean, Natalie E. Theeuwes, Michael Baldauf, Jan Barkmeijer, Geoffrey Bessardon, Lewis Blunn, Jelena Bojarova, Ian A. Boutle, Peter A. Clark, Matthias Demuzere, Peter Dueben, Inger-Lise Frogner, Siebren de Haan, Dawn Harrison, Chiel van Heerwaarden, Rachel Honnert, Adrian Lock, Chiara Marsigli, Valéry Masson, Anne Mccabe, Maarten van Reeuwijk, Nigel Roberts, Pier Siebesma, Petra Smolíková, Xiaohua Yang (2024). The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models. Quarterly Journal of the Royal Meteorological Society, 1–38. DOI: 10.1002/qj.4858.

*[2] Final report of the 8th WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction and Earth System Prediction. Available at:*https://community.wmo.int/en/meetings/8th-wmo-impact-workshop-home

Agenda

All presentations are available here:
https://drive.google.com/drive/u/0/folders/1xJmkokITZBtG_aaHuPa6ZSuPMwOPuqi_

Day 1
Session 1. 13.00-15.00

- Background and introduction and review of data assimilation developments and challenges for km-scale (Jelena)
- Advances and future challenges regarding use of observations (Magnus)
- Discussion
- Role of machine learning (Tomas)
- Scale-independent surface modeling and interaction with the atmosphere (Patrick)
- Discussion

Dinner 19.00

Day 2

Session 1. 9.00-11.00

- Welcome remote and reporting from day 1 (Jelena)
- Background error covariances for high resolution modeling (Loïk)
- Ensemble data assimilation and handling of moisture (Elias)
- Towards OOPS LELAM 4D-VAR (Pau)
- Discussion

Session 2. 11.30-13.00

- Continuous data assimilation in meso-scale (Heikki, 30 min)
- Why do we need an ocean model to predict the weather at high resolution? (Kristian, 30 min)
- High resolution re-analysis (Xiaohua, 30 min)

- Discussion (30 min)

Lunch

Session 3. 14.30-16.00

- Session devoted to Nils

Coffee break 16.00-16.30

Session 4. 16.30-18.30

- Observation systems for Arctic (Michael)
- Dynamics and computational aspects (Ole)
- Towards use of GPUs for hecto metric scale computing (Tomas W)
- Discussion and forming of groups for tomorrow
-

Day 3
Session 1. 9.00-11.00

- Discussions in groups and recommendations

Session 2. 11.30-13.00

- Reporting by groups
- Final discussion

In Memoriam of Nils Gustafsson

Our colleague, friend and source of inspiration Nils passed away on Sunday 21 July, 2024. With his enthusiastic style and great knowledge, he influenced several generations of researchers in data assimilation and NWP. Nils was an internationally respected researcher and a world leader in his field. He has contributed fundamentally to the development of regional NWP in Sweden, Europe and worldwide. In particular, Nils excelled in the development of novel methods for data assimilation, being at the forefront of the field for half a century. His international collaborative efforts contributed to the successful NWP consortia HIRLAM and ACCORD. He taught meteorology courses worldwide and ensured the transfer of knowledge to the next generation of researchers and forecasters.

Nils Gustafsson looking into the future with a positive smile on his face.

Nils became part of the research team that developed the first version of HIRLAM. Based in Copenhagen, Bennert Machenhauer headed the team that built the extremely successful collaborative consortium known as HIRLAM. Over several decades, Nils had many different leading roles in the HIRLAM community, He led the development of a variational data assimilation system. HIRLAM 3D-Var became operational at SMHI on 13 June, 2001, where it replaced the old data assimilation system based on optimal interpolation. Around the same time 3D-Var became operational in several other HIRLAM countries. Nils continued leading the intense work on the development of HIRLAM 4D-Var. HIRLAM 4D-Var became operational at SMHI on January 30, 2008 and about the same time also in other HIRLAM countries.

Despite the great achievements of Nils, exemplified above, Nils always looked for further potential enhancements. He actively took part in discussions and reviews regarding needs for km-scale and hecto-metric scale data assimilation and what can be learned from the successful meso-scale data assimilation developments. Nils actively took part in the planning of this workshop and he would have loved participating in fruitful lively discussions and open-minded ideas for future developments. The workshop provided an opportunity for the workshop participants to stake out future ideas and challenges for hecto-metric scale data assimilation based on experiences of Nils and other colleagues in the field of data assimilation. Even if it was not the purpose of the workshop, the workshop gave us an opportunity to pay tribute to Nils contribution to weather science.

"*Nils leaves an important legacy for our weather prediction in the basis he created for our regional data assimilation systems*" (Harald Schyberg). Nils had good communication and leading abilities and has managed to build a vital and powerful team around him. Nils was a genuine leader who trusted each person and gave them responsibilities that allowed them to grow. And he always was prepared to give support when this was needed. "*Nils was absolutely instrumental for bringing us all together and to make sure that we kept focus on the important things to work on*" (Kristian Mogensen) "..*It was Nils' enthusiasm, kindness and boundless energy that was key to us working so well together …*" (Erik Andersson) In addition to data assimilation Nils have had broad and deep knowledge on many different aspects such as computer-science, dynamics, physics, ensemble methods, verification, diagnostics. Nils was a very curious person who never hesitated to learn new subjects."*He was a marvelous scientist with many new ideas and always eager to learn about and understand new scientific developments. He wanted to learn about new things thoroughly, ready to dive deep into coding and mathematical derivations*" (Erland Källén). He had this fantastic ability to create a clear strategic view on a problem as a whole at the same time when he stayed involved in the details of code development. "*He is quite intolerant about bugs and won't hesitate to invest time and efforts to debug. This ultimately brought success of HIRLAM data assimilation including the 4DVAR.*" (Xiaohua Yang) . Nils was a very brave and infinitely honest person, to himself and to others and he was truly committed to his beliefs. " *He would vigorously argue for his opinions, but was happy to listen to and consider alternative views. He was a man of profound integrity, a truly noble guy* " (Peter Lynch).

Nothing could stop Nils addressing the challenges. “*Nils was brutally honest and made his position clear with the leadership, even when it came to selling "canned porridge"*” (Michael Tjernström). “*He was firm in his own standpoints, but also open to listen to other's opinions*.” (Ole Vignes) “*He is also quite tolerant about different opinions in scientific debate, as long as scientific method is applied*.” (Xiaohua Yang). Nils resisted all types of compromises but truly believed in synthesis as a way of progress. “*He was always open to other people’s opinions and competences and willing to reconsider political as well as scientific positions given really convincing arguments*” (Tomas Landelius). “*Nils was probably the most positive person I have ever had the pleasure to meet, with a tremendous zest for life, able to enjoy it intensely, and generous in sharing this joy with others*” (Jeanette Onvlee). Nils was very generous in sharing his experiences with others, either on his scientific investigations or showing “his personal face” by inviting to sailing, hiking, political discussions, discussing a recent novel, or just to share a bottle of good red wine. This was a combination of personal skills and deep scientific knowledge that made people around Nils trust him. Nils has passed away but his testimony through knowledge, dedication, commitment to science, passion experiencing life will stay with us forever. “Nils, You will be missed, but your legacy lives on in the countless lives you touched” (Lisa Bengtsson).