Strategy 2020-2025
Task Team for Data assimilation

encompassing algorithms, observations, coupling and ensemble aspects

TT composition

- Task Team(TT) Leader: Benedikt Strajnar
- TT Members: Alena Trojakova, Eoin Whelan, Florian Meier, Jean-François Mahfouf, Loik Berre, Maria Monteiro, Roel Stappers, Roger Randriamampianina, Stefan Schneider, Tomas Landelius, Xiaohua Yang
- Further inputs to the draft document provided by: Jan Barkmeijer, Magnus Lindskog, Jelena Bojarova and Carlos Geijo, Ekaterina Kurzeneva

Needs (WHY?)

- Realistic quantitative forecasts at convective scale, especially at the nowcasting and very short time range
- Data assimilation (DA) requires efficient and accurate algorithms, but also a strategy to ease maintenance and development of these algorithms
- Increased use of ensemble information in variational methods to incorporate flow-dependent features
- It is needed to propagate spatially the observed information in a way that accounts for the ongoing meteorological situation, as it strongly influences the amplitudes and structures of background errors.
- There is also a need to implement four-dimensional DA algorithms in order to use time-distributed observations in an optimal way, and also for reducing spin up effects.
- Specific approaches for simulating observation and model errors in the ensemble

Needs (WHY?)(2)

- Optimize and increase use of observations
- To build the basic level of data assimilation (that all Members should be able to afford) in the ALH System and to provide a minimum of support to the Members to handle their local data streams.
- Need for consistent inputs to numerical models of different geophysical systems at several temporal scales, including atmosphere, land/soil, snow, lake, hydrology and photo-chemistry

Main strategic goals (WHAT?) (1)

- Algorithms
  - **Implement and switch to the OOPS version of the DA system.**
  - Assessment and implementation of flow dependent algorithms.
  - Further evolution/implementation of ensemble generation methods and uncertainty quantification.
  - Definition of DA CSC's and technical validation tools/unit tests for DA components.
  - Advance surface assimilation methodology.
  - Provide effective methods for nowcasting.
  - Successfull combination of LAM with large-scale information
  - Progressive integration of DAsKIT

Main strategic goals (WHAT?) (2)

- Observations
  - Successful implementation of forthcoming satellite mission (MTG IRS, EPS-NG)
  - Increased use of EUMETSAT SAF/Copernicus Services
  - Increased use of EUMETNET (E-OPERA, E-ABO, E-GVAP)
  - Manage increased observation volumes.
  - Unification of bias correction, blacklisting, monitoring & diagnostic tools.
- Coupling
  - Support for weak coupling (except atmosphere-surface)

Detailed strategic goals (WHAT/HOW?)

Object-Oriented Prediction System (OOPS)

- OOPS is considered as a suitable and flexible framework for implementing new DA algorithms
- Enables unit-tests to be built as separate applications.
- The general evolution towards the OOPS approach will require specific training for the consortium members and close coordination with ECMWF.
- Current ongoing (preoperational) algorithmic developments in the pre-OOPS environment
- how to ensure smooth transition towards OOPS-ified versions?

OOPS DA Timeline

2021 | 2022 | 2023 | 2024

Validation of LAM OOPS components at MF

CY46,47
validation and optimization of elementary DA units, e.g. obs. operators, 3D/4D variational minimization.

E-suites at MF including OOPS features
CY ??

Validation in ALADIN/HIRLAM
Based on export CY46,47??
ALADIN/HIRLAM
Design of applications.
Capacity building in local teams.

Operational implementation ECMWF, MF
CY49??. Includes global 4D-Var and regional 3D or 4D EnVar

Operational implementation in ALADIN/HIRLAM members
Main configurations of ALADIN/HIRLAM centres under OOPS framework.

DAsKIT
Implementation of DasKIT (basic/lightweight DA) in OOPS framework.

Challenges related to OOPS-ification

- The general evolution towards the OOPS approach will **require specific training** for the consortium members and close coordination with ECMWF.
- There are ongoing (preoperational and R&D) algorithmic developments in ‚h' cycles (LETKF, 4D-Var, hybrid 4D-Var/Ens, DA for NWC). A procedure to secure smooth transition towards their OOPS-ified versions needs to be defined.

Assessment and implementation of flow dependent algorithms

- Several types of flow-dependent methods to be considered:
  - Ensemble-variational (EnVar)
    - 3D-EnVar shows good potential wrt. 3D-Var.
    - Extension to 4D without TL/adjoint model.
    - Depends on quality of an ensemble
  - Four-dimensional variational methods (4D-Var)
    - Depending on tangent linear and adjoint versions of the model are considered to be sufficiently realistic 4D-Var.
    - Possibly including a hybrid combination of static covariances and ensemble localized covariances for 3D background errors.
    - Further development of simplified physics within a multi-incremental 4D-Var framework will also be considered.
  - Various types of hybrid methods
- Ability to reduce spin-up is important

Further evolution of ensemble generation methods and uncertainty quantification

- DA algorithms require the use of ensemble of background perturbations
  - The Ensemble of Data Assimilations (EDA)
    - Attractive scientific consistency with DDA and also for development and maintenance purposes.
  - Local Ensemble Transform Kalman Filter (LETKF)
    - Ensemble-based representation of the Kalman gain matrix, in order to update background perturbations into analysis perturbations.
- Comparison to identify best method for generating initial perturbations associated to background error covariance estimation and EPS initialisation.
- More attention is needed in specification and representation of correlated observation errors for km-scale data assimilation, in order to be able to assimilate more data.

Move to enhanced surface assimilation methods

- Progressively move to enhanced surface assimilation techniques, with respect to the current static Optimal Interpolation (CANARI or GridPP).
- First possible approach: derive OI coefficients from the ensemble simulations (which include weak atmosphere/surface coupling in their cycling) to provide cross-covariance between screen-level variables and surface parameters.
- Options for surface analysis: Simplified Extended Kalman Filter (SEKF), the Extended Kalman Filter (EKF), the Ensemble Kalman Filter (EnKF), or the ensemble-variational (EnVar) algorithm.
- Those methods can (and should) be equally applied to town or lake/sea/ice tiles in case representative observations are available.
- Methods accounting for horizontal correlations should be considered to replace current pointwise algorithms.

Definition of DA CSC's and technical validation tools/unit tests for DA

- Wish: possibility to activate different options for observation processing, quality control, blacklisting, assimilation scheme and for observation usage monitoring.
- The goal is to build a common repository which has the capability to feature different system configurations with reasonable interoperability, and to allow, at the same time, experimentations by partners to explore different flavours of configuration features and options.
- One of the challenges is that this requires a common modular (scheduler driven; e. g. ECFlow) scripting system which can be used for different applications.
- Technical validation tools and unit tests for DA components also need to be built (such as DAVAI).

Nowcasting (1)

- High resolution observations to be available within a short time (10-15 minutes).
- Very promising results with overlapping observation windows; this approach can be considered as a potential solution.
- The NWP system should be relatively fast to support the nowcasting applications.
- The goal is then to build a flow-dependent DA system that allows reduced spin up and is able to process a large range of high temporal and spatial resolution observations with frequent (hourly or sub-hourly) update.
- The system might be applied to relatively small target areas.
- A more appropriate observation operator with supermoding will be explored, and correlated observation errors should be considered.

Nowcasting (2)

- The use of crowdsourced observations processed with adapted machine learning technique will be investigated.
- Initialisation methods (DF, IAU, VC etc…) and approaches that are able to correct phase error (e.g. field alignment) will be also explored.
- For initialization of very rapidly updated systems, pragmatic alternatives to variational/ensemble data assimilation are approaches based on nudging, which currently offer more flexibility to use 2D rapid observations.
- The codes for nudging, which is realized during the model integration, needs to be maintained as it is likely that the desired 4D assimilation methods will not be feasible for all the consortium members in the near future.
- The strategy should allow for a broad variety of implementations as part of the common code releases.

Large scale information

- Several methods (large scale mixing LSM, BlendVar and Jk) are currently used by several countries to transfer large scale information into the LAM DA.
- It has to be kept in mind when switching to OOPS and new DA algorithms such as EnVar that this approach is further supported.
- On the long term, an extension to more flow dependency in large scale correlation consistent with aforementioned background error correlation development might also be necessary.

DAsKIT

- Currently it provides a basic set of DA infrastructure to ease developments of the DA from consortium members with no prior experience.
- Should be flexible and easy to handle, even with low manpower.
- The target DA solution has to account for small HPC resources and/or particularities of regional observation networks (not sufficiently covered by remote sensing for instance)
- Locally available high density conventional data streams, as well as radar volumetric data, are the observation types to be handled initially.
- Specific training on DA science and coding should complement the DAsKIT implementation.
- On the longer term, this functionality is to be ensured as part of a flexible DA system realized through the Object-Oriented Prediction System (OOPS).

Satellite observations (1)

- The most important evolution will come from the availability of radiances from the hyperspectral infrared sounder **IRS** on board the MTG-S (2023).
  - The data volume increase will require adaptations of transfers, storage and pre-processing.
  - Radiances to be provided in PC scores. The specific features of reconstructed radiances will have to be accounted for.
- EPS-SG will be exploited for infra-red sounders (IASI-NG), microwave sounders and imagers (MWS, MWI, ICI) and scatterometers (SCA).
- Studies shall be pursued to better extract information in cloudy conditions and over continental surfaces.
- The use of EnVar systems with hydrometeors in the control vector shall allow an easier assimilation of cloudy radiances.
- Radiances over continents shall provide information both on the lower atmosphere and on surface temperature, to be used land DA.
- Moving from 3D to 4D schemes) will allow to better handle observations.

Satellite observations (2)

- The Flexible Combined Imager (FCI) on board MTG-I (successor of SEVIRI, to be launched by the end of 2021) will provide radiances every 10 min to be exploited in data assimilation.
- The lightning imager (LI) on board MTG-I, should also be exploited for model initialisation with information on severe convection in regions where radars do not operate (oceans, mountains).
- Possibility to assimilate observations from passive microwave sensors like AMSR2 and SSMI on board of US DMSP and JAXA satellites as information on the snow water equivalent.

EUMETNET

- Non-satellite data to be provided by EUMETNET programmes shall be extensively used for data assimilation purposes.
- Radar radial winds and Doppler velocity + polarimetric data from **OPERA**, the extension of the control vector to allow direct assimilation of hydrometeors.
- Data from ground based GNSS receivers through the **E-GVAP** programme shall be further exploited - slant delays and tropospheric gradients with suitable observation operators to be developed.
- The EUMETNET aircraft programme **E-ABO** shall provide additional data from MODE-S and ADS/B receivers to enhance the number of wind and to some extent temperature measurements to be assimilated in LAMs.

EUMETSAT SAF/Copernicus Services

- Exploitation of satellite products from the SAFs of EUMETSAT and Copernicus Services shall be enhanced
- Mainly information on initial conditions of continental surfaces (superficial soil moisture, snow on ground extent, surface temperature (land and water), albedo, vegetation properties) via using these data in the Kalman Filter-based soil assimilation, in combination with optimal interpolation for spatialization.
- The already ongoing initiatives in the consortia should be better coordinated.

Crowd-sourced observations

- It is likely that during the period the amount of meteorological data provided by **IoT** will increase significantly (personal weather stations, smartphones, wind farm and road observations, ...).
- Their potential shall be assessed with specific challenges in terms of quality controls, data exchanges policy exchanges, data volumes.
- Such evolution cannot be ignored despite the fact that the backbone observing system will continue to rely on WMO, EUMETSAT and EUMETNET.
- Other new observations to be exploited (e.g. microlinks)

Data volumes

- The huge amount of additional data will raise specific issues in terms of quality controls, bias corrections and data thinning (spatially and temporally).
- A number of them could be addressed at the algorithm level: spatially and spectrally correlated observation errors, temporal correlation errors.
- Questions related to the scale differences between model grid and data density shall be examined in terms of superobbing or supermodding within the observation operator definition.

Bias correction, blacklisting, monitoring & diagnostics

- Current algorithms rely on unbiased hypothesis, the biases must be removed prior to or during the assimilation process.
- Offline methods alone become impractical. A goal is thus to **implement the Var-BC approach to all observation types**.
- Selection of good quality observations is essential. Decisions currently based on either ARPEGE or ECMWF. In both cases, the updates of the blacklisting choices should be regularly shared with the partners.
- Dedicated monitoring tools like Obsmon (HIRLAM) and Obstat (ECMWF) for research and operations to be further maintained
- Suite of dedicated diagnostics (innovations, energy norms or covariances of residuals) require central system maintenance
- Ensemble assimilation approach will also allow for computation of adjoint-based metrics such as FSOI.

Coupled surface/atmosphere assimilation

- The long-term goal is to move from currently used weak towards strong coupling, which includes cross-covariances of forecast/observation errors between the systems and immediate impact of observations in both systems.
- Two concrete first steps:
  - Improve consistency by ensuring that satellite assimilation uses updated skin temperature and other low level fields.
  - Next, the atmosphere and surface assimilation can be linked by estimation of screen level- surface covariances using ensemble members.

Coupling with other Earth systems

- Coupling (2-way) with ocean/wave, photo-chemistry and hydrological models is not implemented in a typical LAM.
- There are some capabilities to run those systems in a weakly-coupled mode through the SURFEX and OASIS coupler.
- Challenge: the assimilation cycling requires a compatible mesoscale analysis/observations for the ocean as it probably has the most significant feedback on the atmosphere.
- Further enhancement towards a quasi-strong coupling can be achieved within 4D assimilation algorithms like 4D-Var or 4DEnVar by coupled trajectory (outer loop) runs.
- The main challenge in such a coupled system would be maintenance of (compatible) LAM versions of all those components given the limited manpower.
- Potential added value of such weakly/quasi-strongly coupled systems needs to be demonstrated in the next five years.

Beyond 2025

- use of increasingly higher resolution models (e.g. LES models) to become common in areas of high risk (e.g. highly populated areas, airports).
- Therefore it will be necessary for the longer term perspective to develop algorithms that can treat efficiently wide range of scales.
- This would require to include adjustment processes within the data assimilation procedure.
- The following directions, such as non-hydrostatic data assimilation, assimilation of structures instead of point quantities, non-convex constraints need to be exploited.
- A closer connection to non-exact optimization techniques that would allow treatment of on-off observations used in machine learning algorithms such as momentum and ensemble based search techniques needs to be achieved.