ALH Strategy 2021-2025

Table of Contents

1. Introduction and General objectives

This first Strategy for the ALH Consortium has been prepared by combining (i) the ambitions expressed by Members of the current ALADIN, LACE and HIRLAM consortia and (ii) the objectives proposed by the staff working for these consortia. Detailed scientific and technical objectives originate from a specific Strategy Workshop held in Toulouse on 3-4 February, 2020, based on the input of seven task teams that have worked in December 2019 and January 2020.

Further to the Strategy Workshop, the Convergence Working Group has worked to design an initial management structure for the ALH Consortium. This is explained at the end of this document.

The driving factors for the ALH consortium and the current trends of limited-area, short-range NWP

A defining motivation for Members in this intensive collective effort on short range NWP is to enable our organisations to deliver the highest quality weather services to the whole of society.

Members adhere to the ALH consortium in order to develop and share expertise in NWP and mobilize the most advanced computing and meteorological science to create world-leading short-range NWP suites remaining entirely under their control.

Most Members are also part of ECMWF and wish to keep a strong synergy between the codes developed by the Consortium and by ECMWF. As the global NWP systems of ECMWF are increasing fast in resolution, it is compulsory for the Consortium to remain ahead of ECMWF, both in resolution and in meteorological quality of the short-range forecasts, on the limited domains of national interest for Members.

Another essential reason for Members to maintain an intensive collective effort on short-range NWP is to remain ahead of large private companies increasingly trying to capture the market of short-range weather prediction, an evolution that may ultimately deprive the NMSs of their audience.

The market for high-performance computing is evolving fast and is not driven by the needs of meteorological services. It is virtually impossible to know which types of processors and interconnects will be proposed by vendors in five years from now, so our codes should be highly adaptable to guarantee that the quality of forecasts will not suffer from unforeseen evolution of the computing market. The evidence is that the time of Moore’s law is over, opening the door to a diversity of processing units. Another evidence is that the energy-to-solution is becoming a driving parameter in the choice of the best combination of technologies and algorithms. Investing in new software organization and layering, separating scientific algorithms from their actual implementation, the latter exposing all the vectorisation or parallelisation potential of the former, is compulsory to continue increasing the horizontal resolution of NWP systems and ensure the success of future procurements of HPC systems by Members.

Beside resolution, the quality of future forecasts will depend on the use of observations and the correct treatment of all physical processes. This requires continued investment into these two particular areas. The observations available for weather forecasting undergo rapid evolution, e.g. dual polarisation Doppler radars and other surface-based remote sensing instruments, crowd-sourced data, Mode-S aircraft data, GNSS data, etc… Physical processes to take into account now include the coupling with land-surface, ocean and atmospheric chemistry. Members will need to acquire more expertise on these domains by establishing cooperation with new science communities. More realistic and richer models offer the opportunity of richer derived parameters and diagnostics, both in deterministic and probabilistic forecasts. It will be necessary to work closely with users to co-construct the decision models exploiting fully this wealth of information.

The increased realism of high resolution models is also posing some new challenges. Like reality, they have an extraordinary potential for variability. All members of a high resolution ensemble appear as equally possible, offering a wealth of realistic small scale details, but they are more different from one another as well as from an image of reality (as from a radar at full resolution) than with a low resolution ensemble and require more advanced techniques to be turned into usable probabilistic forecasts.

Other recent developments such as artificial intelligence and cloud computing can greatly influence the future of short-range forecasting. The Consortium must closely follow the state of science in these domains and launch actions to exploit these new advances. These techniques can be applied in several areas of a modern forecasting suite, serving different purposes. One such area that is bound to grow in importance is that of post-processing and its probabilistic dimension.

The extent of the challenges faced by the Consortium to increase the computing and meteorological performance of the common codes is such that new gains of productivity are judged compulsory. This pleads in favour of developing more integrated and efficient working methods, and minimizing duplication of work by achieving a high degree of interoperability of the common codes.

The high-level objectives of the Consortium

In the next five years, we will:

- Deliver codes ready to provide forecasts of world-leading quality, with a focus towards higher resolution (sub-km), multiple time scales (from nowcasting to 3-day), and more explicit assessment of predictability (ensembles for both data assimilation and forecast).

In order to enable this prime objective, we will also:

- Ensure regular evolution of the codes towards more interoperability, portability and flexibility,
- Modernize working methods and increase the communication between teams from all Members, to accelerate the preparation and adoption of code upgrades,
- Extend progressively the perimeter of the collaboration to the scripts of the NWP suites and other relevant aspects (such as content and format of database/files, tools for compile/build, etc…),
- Further develop the capacity building activities of the Consortium.

Explanation of the objectives

- *Deliver codes ready to provide forecasts of world-leading quality, with a focus on higher resolution (sub-km), multiple time scales (from nowcasting to 3-day), and more explicit assessment of predictability (ensembles for both data assimilation and forecast),*

As global models are rapidly increasing in resolution, pushing LAMs to hectometric resolutions to maintain their added-value has become the first scientific priority of the Consortium. A similar trend, also already considered by a growing part of members, applies to the development of ensemble forecasts. Their consequences are developed under all scientific themes of this Strategy, and are more visible under the sections on Dynamics, Physics, Surface, Ensembles and Verification. The trend towards nowcasting systems based on the NWP tools is also well visible. It commands specific work, e.g. in the section on Data Assimilation, where the ensemble approaches are also becoming a central concept.

- *Ensure regular evolution of the codes towards more interoperability, portability and flexibility*

Interoperability refers to the diversity of configurations currently supported by the Consortium. It must be recognized that developing a single configuration of the codes for all Members is impossible because of the differences in computing power, computer architectures and environment, available observations and operational constraints.
Portability refers to the facility to deploy the common codes on various computers in order to benefit from computing resources available throughout the Consortium and at ECMWF or other european computing infrastructures (Prace, EuroHPC).
Flexibility refers to the capacity of the codes to run on computers having diverse or hybrid architectures (processors and interconnect). This objective will help the Members to run their computer procurements by enhancing competition between vendors offering various technologies.

This high-level objective is reflected in all chapters of this Strategy, but especially in the Software infrastructure, Dynamics, Physics, Data assimilation and System sections.

- *Modernize working methods and increase the communication between teams from all Members, to accelerate the preparation and adoption of code upgrades*

The Consortium is well aware that developing an efficient collaboration between 26 services is a challenge. The most modern collaborative tools will need to be experimented to facilitate an efficient communication and share the short-term objectives, especially concerning the integration, phasing and validation of codes upgrades. The technical validation tools will need to be further developed and deployed on several computers, in order to streamline the work of integration. This is reflected in specific objectives under the System section.

- *Extend progressively the perimeter of the convergence to the scripts and other aspects of the computing environment (such as content and format of database/files, tools for compile/build, etc…)*

This new objective will offer a real opportunity for more efficiency if the operational computing environments of Members can be aligned. In the short term however, it will represent a challenge because it is a new activity for the Consortium. Objectives are described in the System section.

- *Further develop the capacity building activities of the Consortium*

The pace of scientific and technological change demands highly motivated, skilled and flexible personnel. The expertise of Consortium Members in NWP and computing is heterogeneous. The Consortium should identify areas of knowledge deficits and strive to increase its level of expertise by organizing training sessions, short term scientific missions, exchange programmes and various types of events in order to up-skill and share experiences and best practices. Special attention should be paid on human resources related to critical work areas to monitor single point of failures and adequate mitigation measures should be put in place.

A specific effort is also necessary to ensure the highest quality scientific and technical documentation for all common codes. All Members are also keen to increase their expertise through exchanges with other communities (e.g scientific computing; surface parametrizations; ocean models; climate models; atmospheric chemistry models). The Consortium should play a leading role in these contacts.

The Strategic goals are laid out across eight areas and elaborated in detail in the following chapters of this document:

- Transversal: Addressing future evolutions of software infrastructure
- Dynamics
- Physics parametrizations
- Surface (model and data assimilation)
- Ensemble Forecasting
- Data Assimilation
- Meteorological quality assurance
- System

Objectives by topics

2. Transversal: Addressing future evolutions of software infrastructure

Introduction

The evolution of computing technology is one of the major driving forces of NWP, but, at the same time, it is one that is strongly driven by economic market mechanisms that take place outside the NWP community. The latter has no or very little influence to steer its underlying scientific and technical developments. This implies large uncertainties on the short and long time frames. In the 5-year time frame, the main uncertainty comes from different known technologies being available. In the 10-year time frame, what platforms will provide effective exascale or more computing power is a virtually complete unknown, not for lack of possibilities, but too many rather.

As a consequence, many issues have to be addressed today:

- The adaptation of the common code to new emerging computing technologies calls for major rewrites
  - *a one-sided focus on parallelism is not sufficient anymore.* This has been true as long as only “CPU” based technologies were available. Adapting a code to, for example, GPU is yet another matter, although it benefits from a high level of parallelization;
  - *GPU adaptations need to be performed on most parts of the code*, since performance simulator-based studies shows that significant benefits are achieved only if a very large portion of the code (80 to 90%) is adapted;
  - *so, it is the overall design that has to be addressed.* So far, actual limitations came from other areas than the dynamical core, in particular from input/output handling. When coming to making the code “flexible”, fit for diverse computing architectures, all components are directly involved;
  - *making the code “flexible” is also a data assimilation issue*. There are strong time-critical constraints bearing on data assimilation, complicated from both the sides of spatial resolution of the output, and that of the number of observations. It must be noted here that the traditional strong constraint 4DVar algorithm is intrinsically sequential. Furthermore, very much like atmospheric dynamics, phase space minimisation is intrinsically a non-local problem. There are important and difficult adaptation questions in this area as well.
- At the same time, it should be stressed that *the existing spectral semi-implicit semi-Lagrangian dynamical core is, as of today, still outperforming explicit Eulerian gridpoint dynamical cores in terms of time to solution, a feature from which the ALH community may still benefit in terms of computing cost for a long time*. From the technological viewpoint, having solved the I/O problem, parallelization is hindered by the repeated transposes of the state variable requested prior to performing efficient transforms and, to a lesser extent, by the further need for communicating model fields across several processors to fill so-called “halos” needed for interpolations or, in non-spectral contexts, for spatial operators. Nevertheless, improvement in interconnect technologies in standard parallel computers may enable the approach to continue to be efficient, given the huge advantage of being able to

reach a given forecast range by making about 10 times less time steps than with other time-stepping schemes. [run at large Courant numbers (in other words, being able to run long time steps with respect to a given grid mesh size);

- Lastly, *making the most out of future computing systems is a not compute problem only*. Given the increase of incoming data and the even larger increase of output data that will result from running hectometric scale ensemble data assimilations and forecasts, all the efforts to adapt the compute part of a production suite can easily be lost by keeping an outdated overall data flow. State-of-the-art forecast suites are not only compute intensive. They also are data intensive. The consortium needs to keep an eye on this component and on these aspects which, requiring increasing investment, both scientifically and technologically (because they are fully included in the critical path), may soon be beyond the capabilities of individual members.

The term "making the code flexible" will be used in the wide sense including all the above listed issues, throughout this document.

The approaches to these issues in the international NWP community is contrasted. Some institutions have seriously invested in redesigning forecasts systems from scratch: NCEP, the Met Office and DWD offer different examples of this. NCEP and DWD mostly had parallelization in mind, which does not in fact prepare for mixed architecture computing. The Met Office had the flexibility property in mind at the beginning, but, as time passes, has found that it is too huge a task to make a full modern forecast suite compliant with new software standards.

ECMWF has seriously invested into these topics these past 6 years. This activity took place under the umbrella of the so-called "Scalability Programme". The related investment amounts to a 9 Meuros money stream from European projects, it has enabled to have about 12 project funded FTEs, supervised by 5 core FTE from ECMWF employees, so about a 15 to 20 people strong team. The findings and directions chosen by ECMWF are presented in a comprehensive note "*Scalability programme progress and plans*" (published in Feb. 2020). Selected members from the consortium have been able to take part in a few of these projects. As a result, the know-how that ECMWF has acquired does not really exist in the consortium. *The consortium, having a yearly amount of 4 FTE for flexibility developments, is not in a position to devise its own fully comprehensive scientific strategy, so it is crucial to increase and strengthen its collaboration with ECMWF on this topic.*

Strategic goals

The key concept for addressing the scalability issue is called *separation of concerns*. The software ends up being separated between high-level algorithms that are expressed in a very abstracted way and low-level, hardware dependent, implementations of these algorithms. At a high level, a small number of scientific "objects" are defined and acted upon by low-level flexible software codes. The adaptations of the low level code to the local computing platform are not visible to the high-level scientific developer, thus separating the scientific concerns from the computing ones. The objects and methods can be represented using a standard language. However, it is even better to move towards a very flexible language, easy to play with, with extensions specific to the field, NWP as it is. Some of these ideas are already present in the OOPS framework for data assimilation, and they are generalized in what are called "*domain specific languages*" (DSL). At that level, there should be little or no difference between a mock-up or "toy" system and the full-size problem. The full-size problem requires to properly take into account problem dimensions, memory layout and distribution, multi-standard parallelization (CPU distributed, CPU shared, GPU), vectorization, etc. The way to shape state variable arrays, to organize loops hierarchies is strongly hardware dependent. *The challenge is therefore to develop new layers of software that generates an efficient but specific hardware code starting from the high-level abstract code.* In this way, a technological breakthrough does not require everything to be rewritten. Instead, new “backends” for the new hardware need to be developed. These backends do not have to deal with the full problem complexity but only with a predefined, reduced set of object implementation aspects and methods.

This separation of concerns is the basis of ECMWF’s long-term scientific strategy and is strongly inspired by what Switzerland has been able to achieve through a strong and long-term cooperation between its national meteorological service, one of its most renowned research institution (ETH Zurich), its national computing centre, and, not to forget very strong links established with vendors such as NVIDIA and Cray. It appears that our Swiss colleagues however seem to have been a long way into making the views above an actual working environment. The consortium will seek inspiration to directly learn more about these approaches from them.

ECMWF proposes to stick with the well established tradition of step transformations of the existing code. Its roadmap roughly plans two phases. The first phase aims at a flexible enough code version so that mixed architectures tenders could answer the calls that will be published by ECMWF or by some members around 2024-25. The second phase ambitions to fully develop and implement a numerical environmental prediction “domain specific language”.

Parts of this language are still undefined, but some of its key components are known. One of them is the *Atlas library*, one of the major achievements of the initial years of ECMWF's Scalability Programme. One or two members of the consortium have taken part in its development. Atlas manages and implements high-level constructors holding geometries, structured and unstructured grids, meshes and connectivity information for model state variables. It will also handle its partitioning, therefore their distribution and parallelisation. It also enables the availability, within a single run, to have grids or meshes at several resolutions and provides a range of interpolation methods. ECMWF has the ambition to implement Atlas in *all* Earth system components and not only the atmospheric, wave and continental surface models. Atlas having all the information on state variables and their distribution, having interpolation capabilities further becomes a coupling tool. For non-spectral discretization schemes, Atlas will also provide all the necessary differential and integral operators. Needless to say, the development of Atlas is ongoing, albeit far from being complete at this stage.

There are two other important components that the consortium needs to become familiar with. One is a much expanded I/O server called *MultIO*. For NWP applications of the common code, ECMWF and members use the I/O server provided in the early 2010 by Météo-France. MultIO appears to start from it but adds functionalities, in particular the ability to be used by other components such as the continental surface and wave models. It should also be smoothly integrated with Atlas. The last component is meant to provide flexibility to code pieces that perform calculations along a single dimension, primarily the vertical one: physical parametrizations, surface models, etc. It performs the “single column abstraction” of these components. “Horizontal” dimensions, loop ordering and boundaries and of course the exact memory layout of the state variables are abstracted, so the “SGA” code itself only exposes a compact form of the schemes or codes, with the “vertical” operations only. Two things must be noted: *(i)* the related software, called *Claw*, has been developed by Swiss institutions and will be imported from them, *(ii)* to avoid future rewrites, it

needs to be adapted to Atlas at the same time as the existing representation of the state variables (GFL, GMV) needs to be made Atlas compatible.
In short, a first implementation of Atlas will soon take place in the common code to enable a first level of flexibility that will target the current spectral, semi-implicit semi-Lagrangian models together with their “columnwise” schemes. The CPU performances should not be affected, meaning that Atlas should be used at high-level and precompile stages only.

Assuming the consortium agrees to follow ECMWF's invitation to cooperate on its roadmap, the following directions must be taken. Some of them are somewhat urgent and will be addressed first (starting already in 2020):

- **Decide on the approach to introduce flexibility in the 1D components, namely physical parametrizations and surface models**. One may rely on introducing OpenACC directives (or the latest OpenMP standard, already a choice to make). Alternatively, one may follow ECMWF and implement single column abstraction, using the Claw-derived software. Both approaches have pros and cons, they need to be analysed and a consortium wide decision must be taken. There seems to be a related decision about which form of the software the development of physics should continue to be performed on, once Claw is introduced.
- **Train people and start implementing the chosen approach**, while ECMWF moves on to adapt the setup and dynamics. This will be based on being able to use the transform library that has been optimized for GPU (possibly MPI-aware GPUs). Of course, the consortium will have to take over and extend the work to limited area.
- **Train people to Atlas, to MultIO**. Both combined raise questions on the future of the diagnostic post-processing code called “fullpos” (at least as it exists now), that will have to be addressed. It is highly recommended that the consortium contributes, as soon as possible, to the development and implementation of both.

These early steps, if they are properly managed, will enable members to gather the necessary know how. These will be useful to enter the second phase with a significant consortium contribution to defining a limited area version of a numerical environmental prediction “domain specific language”. The longer-term issues, such as the implementation of Atlas or the development of alternative discretization schemes to LAM configurations, are taken into account in the area’s described in the next chapters.

3. Dynamics

Introduction

The dynamical core is the model part that numerically solves the dynamical equations of the atmosphere. Our current spectral semi-Implicit, semi-Lagrangian (SISL) approach yields a high accuracy and allows for very large time steps, which provides the most computationally cost-effective solution. However, the long-term future of this approach can be questioned, for mainly two reasons. Firstly, at very high resolutions, the treatment of steep slopes in the orography will become unstable or inaccurate with the current formulation of the semi-Implicit scheme. The second is the so-called scalability problem discussed in chapter 2.

Over the past decades, several NWP communities developed new dynamical cores, largely driven by the need to move from a hydrostatic to a high-resolution non-hydrostatic model. The ALADIN-NH dynamical core has been very successful, so our community did not have such an urgent need. There are also strong indications that the spectral SISL solution will remain reliable for our applications for the next 10 years. Developing a new dynamical core is a long and tedious task and requires a very specific expertise that is usually rare to find. Most of our dynamical core developments in the past decade have therefore been focused on refining the existing solutions and adapting them to higher resolutions. Over the past five years, the ALH community has nevertheless started to investigate alternative grid- or mesh-like solutions and intends to proceed with this in the coming five years, taking into account the dynamical core developments at the ECMWF.

The two main strategic objectives for the Dynamics Area are:

- *Continue with improvements of the present dynamical core towards the hectometric scale*
- *Invest in a long-term (~10 yr) evolution of an Atlas based dynamical core.*

Strategy goals

Continue with improvements of the present dynamical core towards the hectometric scale

When adapting the model to higher resolutions, parts of the existing dynamics may have to be revisited. This is a permanently recurring process and will still be needed within the next five to ten years. Specifically, the role of the horizontal diffusion, the time and space aspects of the lateral-boundary conditions (LBC) and the treatment of the orography at the vertical boundaries will have to be addressed. Recent work on reformulating the variables of the dynamics opens the way for a more stable and accurate treatment of the orography. This leads to several quick wins in the code and this will be finalized. Successful work was carried out on the vertical discretization with Vertical Finite Elements (with even successful tests within the IFS) and will continue. There are still various ways to improve conservativeness and the cost-effectiveness of the semi-Lagrangian transport and these efforts will continue. Additionally, the effectiveness of the LBC data transfer from the global to the LAM models will be further investigated. At a more fundamental level, making the current Iterative Centered Implicit (ICI) scheme less expensive will be addressed by applying the corrector step only when needed.

Invest in a long-term (~10 yr) evolution of an Atlas based dynamical core

The dynamical core handles the adiabatic aspects of the atmospheric forces and transport. Since the geographical locations of the grid boxes are usually spread over different cores across the HPC platforms, this requires interconnect communications which may decrease the time-to-solution of the model execution. Our spectral methods make back-and-forth transforms from gridpoint to spectral space at every step. The transforms include a step of global communications that decreases the computational efficiency of the model when it is run on massively parallel machines, so far not to the point that it reduces its performance.

This problem is currently investigated in the ALH community by implementing an alternative SISL finite difference solver within the existing code. The advantage of this method is that it is minimally invasive in the current code since it only replaces the current discretizations of the horizontal derivatives and the solver of the Helmholtz problem and it leaves the rest of the so-called gridpoint computations untouched. This solver is an excellent scientific tool to perform clean intercomparisons tests and may provide an alternative to the spectral method at a longer time scale. However, it has only been outlined and many properties have to be assessed (handling of key mathematical identities, conservations, degrees of freedom and spurious modes, stability, etc).

Our code is shared with the global models ARPEGE and the IFS of ECMWF. ECMWF is now developing a new gridpoint-like dynamical core, called the Finite Volume Module (FVM), with a conservative advection scheme. The FVM is a non-spectral NH system that uses a different set of equations on the same horizontal grid and with the same vertical levels as the IFS. Its implementation is based on the novel framework to handle the data structures called Atlas, discussed in the previous section. The FVM is currently under evaluation. In the short term, up to 2024 at least, ECMWF intends to keep the historical spectral IFS, possibly using the ALADIN-NH options, while still improving many aspects concerning code structure, optimizations and portability. A large part of the source code should be ready for running efficiently on GPUs by 2024 and the Atlas-based structures should replace the current structures meaning that IFS might operationally work as well with the Atlas library. The LAM codes should closely follow these developments.

The long-term dynamical core strategy for the LAMs is based on a twofold approach:

- **Develop a LAM solution based on a finite-volume approach following the FVM developments of ECMWF.**
- **Finalize a gridpoint (finite difference) dynamics solver as a scientific testbed, as a backup solution and as an alternative to the spectral dynamics.**

Both approaches will be based on the implementation of the LAM aspects within the Atlas framework.

4. Physics parametrizations

Introduction

Currently, the system can be configured with a choice of three physics packages. This has led to the concept of the CSCs (ALARO, AROME-Météo-France and Harmonie-AROME). The way the current options for physics parametrizations are organized in separate CSCs has often been misinterpreted as if three separate models are being developed. However, these physics packages originally were designed only to run the model at different resolution ranges and to deal differently with the grey zone of deep convection. When going to hectometric resolutions the validity of the model formulations regarding grey zone issues is not yet fully known, and this validity issue should be readdressed. This may be a good opportunity to document the available parameterizations and to investigate what are the blocking points to reorganize the calling routines into a common calling structure.

The computations of all physics packages are presently limited to vertical columns. However when going to hectometric scales, 3D representations of the physics will be required, which may more broadly affect the code structure.

The continuous improvements are routinely being validated. The ancillary tools and environments for this should be shared as much as possible. There is ongoing common work on the single column application MUSC and the making of an inventory of model output postprocessing parameters, organized transversely across CSCs.

Strategic goals

For the physics parametrizations we will:

- **Work towards a greater level of interoperability (enabling exchange of individual parameterizations across CSCs)**
- **Develop the model physics to be fit to represent the hectometric scales.**
- **Improve the model forecast performance through (i) the introduction of more realistic physics descriptions and (ii) assessment of the underlying causes of systematic model errors.**

Greater level of interoperability for the physics

In the future we should aim towards further convergence at the level of individual parametrizations to have a greater choice in combining different parametrizations. This will require the following actions:

1. First, list scientific and technical blocking points for convergence and analyze to what extent the dataflow can be reorganized to interface individual physics parametrizations from one CSC to another. The conclusions will be delivered in 2023.
2. Next , develop the interfaces for exchanging individual physics parameterizations for the same physical process within the physics calling sequence.
3. Perform the technical refactoring and restructuring of the code that should improve the parametrization calls,
4. Homogenize the availability of the model post-processing parameters over different physics packages.

The resulting increased choice will facilitate a broader resolution range to configure the models, e.g. for climate applications or perturbations of EPS systems. The three physics packages have been used in EPS applications, either separately or jointly (multimodel). The variety of physics parametrizations and the options within them provides a valuable tool for perturbing the models and is already used in several ensemble prediction systems.

Model Physics towards the hectometric scales

The aim is to support the increase of resolution to hectometric scales. To this end,

- we will explore the range of validity of individual physics parametrizations from all available physics packages, the possibility of extending their range and determine the range of the grey zones for different parameterized processes. The result of this research should reveal which options for a given physical process (for example turbulence or radiation) represent viable options to be further developed and used at hectometric scales.
- We expect a need for accounting for some 3D effects in the representation of turbulence (at large: moist and dry, homogeneous and structured) and radiation at the hectometric scales.
- To begin with we will start to develop affordable technical solutions in the code to represent 3D effects in turbulence and radiation.
- We need to explore which data set and what methodology to use to validate at the hectometric scales.
- We expect that the complete development of these 3D parameterizations schemes will continue beyond this MoU as their finalization and the delivery into operations will require a longer time frame.

The new code developments needed for the underlying computations for 3D effects will be done such that they are interoperable between de CSCs.

Improve the model forecast performance

The models are continuously improved and validated. This will be done through:

- Introduction of more complex and more realistic parametrizations for radiation, clouds and aerosols, involving the interactions between radiation, clouds and aerosols, the use of near-real-time aerosol information and the two-moment microphysics scheme.
- investigation of higher order turbulence schemes.
- Starting the development of EPS-oriented physics with statistically based parametrizations and built- in perturbation methods.
- Exploration of the use of machine learning techniques for specific problems, for example but not limited to parametrizations of triggering the deep convection or specific application in tuning of different physics parameters in the model. This will require increasing the level of expertise on these techniques.
- Investigation of the known systematic errors and biases in different physics packages (or individual parametrizations) and identification of their causes. Here, there should be a focus on common problems.
- Building and further developing a common set of diagnostic and validation tools, such as:
  - a common single column model testing framework MUSC,
  - complete the developments of the flexible DDH diagnostic tool flexible DDH.

5. Surface (model and DA)

Introduction

The Surface area covers the scientific and development strategy concerning the surface model, the methods for surface data assimilation, as well as the use of input data, such as the physiography data and surface observations. A detailed description of the surface properties and their exchanges (water, energy, momentum fluxes) with the planetary boundary layer is vital to improve the forecast skill of related atmospheric processes, e.g. diurnal cycle of the boundary layer, fog formation and dissipation, initiation of convective storms. A great opportunity for the surface area is that we have a common platform in SURFEX across the CSCs and that the NWP activities can benefit from model development done in climate research and for process studies. Besides these external developments, it will be important to streamline SURFEX for NWP and to simplify the phasing, as well as the planning process for the next version in a new cycle.

The higher resolution requires also higher details in the input data both for the physiography and surface data assimilation. Satellite products, crowd-sourced data, national databases and other sources provide new opportunities where we need to prepare our code for utilization and assess the impact of the new data. For the surface data assimilation, a modernisation of the code is required to allow for new data, improved methods and better code design for OOPS and scalability requirements. The long-term aim is a strongly coupled atmosphere-surface assimilation system.

Strategic goals

For the surface model, we will:

- **ensure that all members switch to SURFEX, while some explore more advanced options of that software,**
- **streamline the code for phasing and coupling with other parts of the NWP system.**

SURFEX provides many detailed processes for the surface physics that can be beneficial for NWP. Our aim is that SURFEX will be used in all CSCs and by all members. For continuous operational usage, it is required to contribute to SURFEX code improvements for simpler phasing and to streamline it for coupling with the atmospheric models, data assimilation and diagnostics. Some advanced SURFEX options, for example concerning soil, urban and snow physics, will be explored. The goal is to achieve a more consistent modelling of the Earth System including the relevant processes for the modelled scales. As advanced SURFEX options are partly only approved in SURFEX v8.1 or later, it is desirable that new cycles of our system are based on as recent SURFEX versions as possible. In order to capture the different ambitions, we need to define a shared approach for deciding on the next common SURFEX version in a new cycle, as well as a common validation and documentation of the code.

For physiography, we will:

- **update physiographic databases in advance of/anticipating model resolution increases,**
- **assess the impact of the new databases on forecasts and perform the required tuning of surface modules.**

Whenever model resolution increases, physiographic databases need to be improved accordingly, as these parameters are one of the key sources of fine scale added value of limited-area forecasts relative to their larger scale coupling model. New alternatives continuously appear, like e.g. ECOCLIMAP-SG based on the ESA-CCI land cover product, but with new databases, discrepancies occur in quality between different surface characteristics such as land-sea mask, land cover, urban characteristic, etc. We need to spend effort on the identification of such discrepancies and their handling as well as corresponding tuning of the surface parametrizations, such as roughness formulations and surface heat capacities. Expertise on the required tools and methods have to be spread among the members, e.g., in order to introduce local changes in physiographic databases and provide feedback on these changes to the ECOCLIMAP team.

For surface data assimilation, we will:

- **assimilate high resolution satellite data and crowd-sourced data,**
- **work on the interoperability and modernization of the code for the spatialization and the adaptation to the OOPS framework,**
- **progressively move to a coupled surface-atmosphere data assimilation system.**

The increased spatial and physical details in the surface modelling have to be initialized appropriately. The number of relevant non-conventional observations, such as remote sensing products, radiances, backscatter, and in-situ measurements, is constantly increasing but our current operational system setups cannot utilize these data. This is also true for the potential impact from crowd-sourced data such as amateur weather stations. To assimilate the data from these sources and assess their impact, the short term goal is the development of current and new surface data assimilation methods, such as Optimal Interpolation and Kalman-filter based methods. This will also provide a first step for assimilation with more advanced surface physics, e.g. diffusion soil scheme. Furthermore, a modernization of the software used for horizontal spatialization is needed. All development and modernization of the surface DA code have to aim at interoperability and adaption as well as contributions to the OOPS framework, subject to discussions with ECMWF. To capture the relevant balances of water and energy, the above mentioned consistent modelling of the Earth system is aimed for the medium term. For the long term, we aim at a coupled surface-atmosphere assimilation system that will allow a consistent initialization of the surface and atmosphere. The algorithm is envisioned to rely on the ensemble-based methods in accordance with the DA strategy.

6. Ensemble Forecasting

Introduction

Limited area ensemble prediction systems are a crucial NWP component needed for the timely prediction of the probability of occurrence of local phenomena, particularly high impact weather. With increasing model resolution, the importance of using an ensemble approach becomes even stronger, as predictability is more limited at smaller scales.
EPS systems face challenges related to their high computing costs and data volumes for storage and transfer that pose severe limitations for ensemble generation and calibration. Keeping its computational cost manageable is one of the main challenges for research and development.
A second issue is that developments in this area often are strongly dependent on development strategies in all other R&D areas. The development of a comprehensive set of perturbation methods which provide a realistic representation for the various sources of uncertainty cannot be achieved by the EPS staff alone. Stronger and more direct collaboration between EPS staff and researchers working in other areas needs to be fostered, including also experts on calibration, statistics and machine learning methods, post-processing, and system and computer science aspects.

Strategic goals

The main strategic goal for the Ensemble forecasting area is to achieve a seamless, well-calibrated high-resolution ensemble prediction system from nowcasting to 2-3 days ahead. More specifically, we will:

- **Further develop the perturbation methods towards:**
  - More physically based perturbations
  - Mutually balanced perturbations of different model components (e.g. physics parametrizations and initial conditions)
  - Creating an ensemble assimilation system which is of mutual benefit for EPS and data assimilation, with the same uncertainty estimates used in both
- **Explore ensemble performance for extreme weather**, in nowcasting and at sub-km resolution.
- **Further develop an ensemble calibration approach** which pays attention to extreme weather as well as the usual day to day weather, enhances EPS for any grid, accounts for weather-type dependencies, and uses the latest statistical/machine learning methods.
- **Develop user-oriented approaches of common interest** that facilitate use of ensemble outputs

Further development of perturbation methods

There is a need for a stronger focus on perturbation methods which represent uncertainty close to their source and/or which respect model balances. Presently model perturbations are generated by perturbing either total physics tendencies or parameters in physics schemes, with little or no

attention to the balance between different perturbations. Several new sources of uncertainty, e.g. in dynamics, remain to be modelled. Options should be explored for developing fully stochastic versions of individual physics parametrizations.
The ensemble system should be suited for use in, and be of benefit to, the data assimilation system. It is aimed to achieve a consistent approach which leads to the same uncertainty estimates being used for both EPS and data assimilation. The impact of ensemble size, model resolution and the possible need for a compromise between size and resolution on error sampling, on convective scales should be investigated.

Explore performance for extreme weather, nowcasting range and at sub-km resolution

Investigations will be done to assess the benefits of, and trade-off between, higher EPS resolution, larger ensemble size and more frequent EPS computation for various forecast ranges. In particular, the consequences for choices made here for the representation of extreme weather in the EPS should be considered.

Calibration

Ensemble calibration is important to achieve reliable ensemble predictions, but so far has been developed for only a very limited set of near-surface model parameters. One aim is to extend this set. Methods are sought which are capable to enhance ensemble reliability for the entire grid, not only at grid points where observations are available. A second objective is to ensure that calibration methods can improve on ensemble performance particularly for high impact weather. The relative rarity of such events will likely require long training data sets. It should be investigated to what extent machine learning techniques may be able to more effectively draw information from the available training data, e.g. in the accounting for weather-type dependencies.

Use of ensemble output

To make the use of ensemble output more attractive to forecasters and end users, it will be important to devote attention to the development of inline or offline dedicated postprocessing methods. The aim here is to facilitate a swift interpretation of the vast amount of EPS output by the user (e.g. visualization of upscaling to skilful scales, synthesis combinations of ensemble members, time-junction of ensembles).

Collaboration and working environment aspects

It is important to achieve a more effective cooperation in the EPS area and a higher level of interoperability that should lead to a larger number of EPS components (not just the perturbation methods) available for individual tests in the common code. While methodologies and code for generating perturbations and calibration are being exchanged and shared, effective cooperation across the subgroups is limited in practice due to great differences in the ensemble configurations in regard particularly to their use of scripting systems. In the System area (see that section), a number of investigations will be carried out to assess how to converge more, among others on scripting systems; in these investigations, the ensemble forecasting aspects should be specifically considered.

7. Data Assimilation

Introduction

The Data Assimilation (DA) Area will address challenges of innovation in the assimilation methods, code framework changes, move towards higher resolution and high density observations, while ensuring aspects of capacity building. The focus will be on atmospheric DA, with close coordination with the Surface Area regarding the work on surface assimilation.

At the start of the strategy implementation period (2021 and after), mostly 3D-VAR and 4D-VAR configurations will exist in the operational systems of Consortium Members (albeit not tested across all CSCs), and the mid-term goal (by 2025) will be to bring solutions based on ensemble variational algorithms (EN-VAR) as close as possible to scientific (and operational) maturity. In parallel, the DA codes will progressively move towards OOPS-based versions, in which EN-VAR already exists as prototype code at the beginning of the Strategy period.

Over the full Strategy period 2021-2025, a number of Consortium DA systems will reach the 1km resolution scale, and both the number and density of observations will steadily increase. In addition, DA solutions for very short range forecasting (nowcasting) may require specific adaptation, and it will be important to ensure that any of those specific developments remain compatible with the overall VAR and EN-VAR solutions and codes.

A specific aspect of capacity building will be to ensure that any Member of the Consortium can implement a baseline DA system configuration, which is believed to be an upper-air 3D-VAR complemented by a surface OI (aka “DAs-KIT”). A range of GTS observations is then supposed to be available via local reception and throughput at the Member site.

At the forefront of new or improved observation types, one may list Satellite data (EPS-SG, MTG, GPS derived obs), ground-based radar data (Eumetnet/OPERA), aircraft (Mode-S), lidar (Aeolus), crowd-sourced data (probably the most challenging and novel source of observations, in terms of volume, density, frequency and quality assessment and furthermore, essentially near surface based). Specific issues of interoperability and capacity building will have to be addressed at Consortium-level, like the potential share of a common pre-processing tool.

Strategic goals

Main goals listed for the upper-air DA Area (with regular shape font indicating the tasks of the DA Area Leader*, while italic shape indicates topics on which the DA Area Leader should be ready to provide expertise or guidance*):

Efficient, accurate, maintainable algorithms

- **Implement OOPS**: the OOPS framework offers an opportunity for convergence of algorithms and codes, through the systematic implementation of object-oriented features (classes) following the OOPS design for IFS. OOPS will further enable DA component testing, a feature which is believed to become instrumental in systematic technical validation of non-regression of Common Code changes.

- **Develop a DA algorithmic solution based on the ensemble variational approach**. EN-VAR algorithms as formulated in the OOPS framework enable a number of choices for combining ensemble-based trajectory and variational-based operators in the evaluation of the cost function and its gradient. A specific aspect of 4D-EN-VAR solutions is that they do not require the forecast model tangent linear and adjoint codes (and thus their maintenance), while offering new directions for studying scalability properties of the overall minimization algorithm.
- **Designing DA scientific tests**: the performance assessment of any specific DA solution may require careful design and experimentation in order to compare solutions, with an assessment of the relative weighting of meteorological performance with respect to cost, interoperability and potential for future innovation.
- **Optimize DA for nowcasting**: the added value of specific nowcasting solutions, with respect to a baseline VAR or EN-VAR DA configuration, should be assessed. Several techniques for initializing cloud variables, or for moderating spin-up effects, should be compared (while avoiding a "zoo" of options). Any retained specific technique should be implemented in the Common Codes while keeping the compatibility with the formulations of VAR and EN-VAR, as well as the design of OOPS.
- ***DA CSC's definition and testing****: the specific design of one DA configuration, for any given CSC, is of the responsibility of the CSC Leader. He/she will organize the implementation and testing of new features from the DA Area within the appropriate system environment (which may not be a commonly shared environment for feasibility reasons).*

Observations

- **Assess the possibility for the common observations pre-processing and monitoring tool SAPP to be used across the full Consortium**. Many Members do not have the means to continue independent development of observations pre-processing tools. In the recent past, the ECMWF SAPP software has become a prominent tool among several Member States, taking advantage of the ECMWF's role as intensive user of many new observations from various international Programs.
- **Liaison with Eumetnet and Eumetsat Programs**, in close coordination with other national representatives from Consortium Members in the Eumetnet/OBS Programs and Eumetsat SAF. Lobbying for the use of observations in high resolution DA LAM systems will also be undertaken.
- ***Continuous effort on using high-resolution observations.*** *One novelty is the massive number of data from various sources (satellite, radar, aircraft, crowd-sourced data etc.) and their quality control exploring Machine Learning approaches. Members or groups of Members who will have an early access to massive sets of new observations are expected to engage first into this activity. Note: the daily monitoring of routine observations is entirely of the responsibility of each Member (operational-type), but mechanisms for sharing monitoring information, such as sudden changes, should be set up.*

Surface assimilation:

This topic is part of the Surface Area of the Strategy.

8. Meteorological quality assurance

Introduction

Within the ALH consortium, meteorological quality assurance and verification are generally used in three contexts:

- Monitoring the quality of the operational NWP systems, routinely and from cycle to cycle;
- Verification in support of model development;
- Feedback between developers and users

In view of the multitude of relevant model parameters to verify, LAM verification systems need to be able to work with a wide range of observational data, and should be very flexible in their ability to deal with different data formats and resolutions. Both spatial and EPS verification demand a high efficiency in the handling of large data amounts.

Historically, a diversity of tools have been used in the ALH community for standard point verification against in-situ observations. The functionality of the jointly developed HARP tool, has gradually been extended from ensemble verification to spatial, point and conditional verification.

Strategic goals

In the Meteorological quality assurance area we will:

- Make the jointly developed HARP verification system attractive as a common verification tool.
- Further develop common methods/metrics, with a focus on methods for high density/resolution spatial-temporal verification and high impact weather.
- Enhance the verification of 3/4D physical processes to aid model development, including the necessary observations
- Consider greater synergies with the DA team on observation uses and quality control
- Enhance the user-developer interaction

HARP as a common tool

The aim is to assess to what extent the functionality of the HARP system is sufficient for HARP to become an attractive replacement for existing, operationally established verification tools. An iterative consultation process (estimated duration ~1 year) will be arranged, in which aspects will be considered like methods and metrics, observations (which ones and why), maintainability (as cycle-independent as possible) and distribution. On the basis of the outcome, follow-on actions for HARP development will be identified and carried out.

Further develop common verification methods/metrics

The development of verification methodologies will be focussed on the introduction of spatial-temporal methods for ensemble forecasts, and on verification of a wider range of parameters (cloud characteristics, lightning, etc.). Machine learning techniques may contribute in the context of automated pattern detection. Dedicated methods will be developed to verify phenomena related to extreme or high impact weather. In this context the potential of crowd-sourced data to provide detailed information on structure and intensity of high impact weather events will be explored. At very high-resolution, observations to verify against may be scarce and the verification against analyses or reanalyses data could be considered.

Verification of 3/4D processes to aid model development

For model development, it is crucial to know how well the models are able to represent complex physical processes and their interactions. To assess this, relevant model parameters should be compared as directly as possible against observations. It is aimed to increase the use of available remote sensing data and retrieval products which offer good potential for process verification. The goal is to set up a verification environment, in which a variety of relevant observations and retrievals are included in a stepwise manner.

Consider greater synergies with the DA team on observation use and quality control

Like data assimilation, the development and use of verification methods depends very strongly on the availability and fitness for use of observations. The acquisition, use and quality control of observations is a topic which is very much of shared interest between the verification and data assimilation teams. Therefore, it is aimed to create more direct cooperation and greater synergies between these two groups.

Enhance the user-developer interaction

It is desirable to establish a feedback culture, both from model developers to operational users (in particular duty forecasters) and from users to developers. This interaction is mainly expected to take place locally. However, there are some activities by which the consortium can aid in the transmission of knowledge on new model capabilities and performance improvements, e.g. by organizing consortium-wide user meetings or the development of tools which can present the strengths and weaknesses of the model in a concise manner.

9. System

Introduction

The System Area encompasses a variety of aspects from code management, through testing tools and compile tools, to more or less complex or comprehensive scripting. Most of these technical items are closely associated with working practices and platforms for development. They all are part of the broad scope of the topic “working environment”.

At the start of the Strategy planning period, various Members and grouping of Members will use very specific working environments, and it is being acknowledged that convergence will be very progressive, possibly not reaching a fully common work environment of all Members by the end of the Strategy period 2021-2025. For some aspects, like code management, the IFS/Arpege collaboration on the global models codes is likely to provide boundary conditions to Consortium-internal decisions. For other aspects like scripting, local specifications of IT or the link between Research and Production environments also are potential barriers for defining Consortium-wide common (and identical) choices. Nevertheless, the System Area work should undertake a number of investigations and actions in order to identify and exploit potential areas for convergence of tools and work methods.

Strategic goals

Develop a more distributed, efficient and continuous process for the integration and validation of new developments for the T-codes.

At the start of the Strategy period, MF has already taken a few preliminary steps for changing the working practices locally, i.e. within their own NWP work. These steps include the development of a new testing tool (“DAVAÏ”) along with the implementation of a wider range of test configurations, taking advantage of the outcomes of the OOPS Project (e.g. testing components of the data assimilation codes, separate from a full test of 3D/4D-VAR). In the first years of the Strategy planning period, the testing and integration methods at MF therefore will evolve, and this will have consequences on the Code collaboration with the Members of the Consortium. The System Area Leader will analyze and propose efficient work practices throughout Members or grouping of Members, in order to complement the possibilities for collaborative work, in close coordination with the Integration Leader. In this task, he/she must ensure that the proposed tools and work process abide to evolving ECMWF and MF integration constraints.

- **The source code repository (SCR) should be shared and easily accessible**, likely to be a mirror of the MF central repository of the T-cycles. Such an environment should enable specific R&D work on the codes or “pre-integration” work, ahead of a commitment to the MF central repository.
- The above SCR should be complemented by **a shared information environment** (e.g. wiki/ ticketing) in order to facilitate the interaction between all consortium members. Relevant information from the presently separate environments should be accumulated into the shared

one. This will provide a basis for exploring new common communication and working practices. The possibility to link such a shared information environment with a MF equivalent environment should be evaluated.

- The SCR requires **a testing tool** (for the technical validation of a code contribution, or while integrating contributions for a new cycle) **with an increased level of portability and flexibility** as compared to the current situation, i.e. one that can also be used outside the MF environment in at least one or two other locations, and in the future also on the more heterogeneous HPC architectures which are expected to emerge. The primary target would be to assess and ensure the portability of the testing tool DAVAI between the ECMWF and MF platforms. In addition, or alternatively, the level of consistency of the test definitions between the test environments located at ECMWF and MF should be carefully addressed.

Explore what could be the elements of a more common working environment,

Through the collection of information from Members in order to map their current scripting systems, their functionalities and their dependencies on IT elements that may constitute barriers to convergence on the scripting system. Assess where opportunities, or barriers, exist for convergence on scripting systems or tools. Specific aspects related to the link between Research and Production, in any local organization of the NWP system, also could be listed (like database access, archiving etc.), especially when such aspects have triggered specific local work on the scripting.

Assess the potential of the VORTEX library as the basis for a future common system.

VORTEX has been designed at Meteo-France with the aim to mitigate some of the usual shortcomings of scripts, like the need to overload the scripts with a long list of local explicit details that would not apply to another partner. Furthermore, it has been developed in close coordination with the MF Operational Sections, which likely gave it a somewhat “general” flavour. However, whether the level of generality reached by VORTEX makes it a valuable candidate as a base ingredient for a Consortium common scripting tool remains to be evaluated.