SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	Jan 2023 – Jun 2023……………………………….……
Project Title:	HIRLAM-C 4th phase (2023-2025) Special Project
	…………………………………………………….……
Computer Project Account:	Spsehlam…………………………………………………
Principal Investigator(s):	J. Onvlee…………………………………………………
	…………………………………………………….……
Affiliation:	KNMI
Name of ECMWF scientist(s) collaborating to the project (if applicable)	…………………………………………………….……
	…………………………………………………….……
Start date of the project:	1 January 2023……………………………………………
Expected end date:	31 December 2025………………………………………

Computer resources allocated/used for the current year and the previous one

(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used (by 21 June)
High Performance Computing Facility	(units)	40 MSBU	10.7 MSBU	40 MSBU	0 MSBU
Data storage capacity	(Gbytes)				

Summary of project objectives (10 lines max)

Special project resources are primarily used for full integration testing Main areas of attention in 2023-2025:

- Complete the refactoring and code adaptation of Harmonie-Arome for alternative HPC architectures
- Integrate all assimilation algorithms developed for Harmonie-Arome in the OOPS framework. Introduce and assess 3/4D EnVar and all-sky radiance assimilation, optimize assimilation for nowcasting, and work towards coupled data assimilation.
- Include new (esp. surface) remote sensing observations and increase use of high resolution third-party data.
- Develop and assess hectometric scale and nowcasting (ensemble) setups, with focus on urban aspects
- Develop, and assess the impact of, more realistic descriptions of the microphysics, radiation, use of real-time aerosol, stochastic physics, new or improved surface model components and surface physiography.

Build experience with and training datasets for machine learning applications throughout Harmonie system.

Summary of problems encountered (10 lines max)

Running longer experiments for integration testing of new model components has been rather problematic for our researchers in 2022 and 2023, due to the many changes in the research working environment in Bologna. The unfamiliar and unstable environment has resulted in crashes and delays in many of our longer experiments and has caused users to switch to local facilities for longer model runs. Relatively slow retrievals from MARS have also been causing problems with some regularity.

Summary of plans for the continuation of the project (10 lines max)

At the end of this special project (Dec. 2025), the HIRLAM programme will be terminated. After that, a new Special Project for similar purposes might be initiated by the United Weather Centers, but this is still uncertain.

List of publications/reports from the project with complete references

De Bruijn, E.I.F., F. C. Bosveld, S. de Haan, G.-J. Marseille, A.A.M. Holtslag, 2023: "Wind observations from Hot-air Balloons and their application in an NWP model", Met. Appl., doi: 10.1002/met.2128

Calvo-Sancho, C., et al., 2023: "Assessing the performance of the HARMONIE-AROME and WRF-ARW numerical models in North Atlantic Tropical Transitions". Atmospheric Research 291, 10680

Kähnert, M., Sodemann, H., Remes, T.M. et al., 2023: "Spatial Variability of Nocturnal Stability Regimes in an Operational Weather Prediction Model". Boundary Layer Meteorol 186, 373–397. https://doi.org/10.1007/s10546-022-00762-1

Lindskog, M., R. Azad, S. de Haan, J. Blomster, M. Ridal, 2023: "Impact of Mode-S Enhanced Surveillance Weather Observations on Weather Forecasts over the MetCoOp Northern European model domain", accepted in Journal of Applied Meteorology and Climatology.

Mile, M., 2023: "Advanced assimilation of satellite observations in a limited-area numerical weather prediction model over the Arctic region", PhD Thesis submitted to Univ. Oslo (defense: Sept.2023)

L. Quitián-Hernández et al., 2023: Assessment of HARMONIE-AROME in the simulation of the convective activity associated to a subtropical transition using satellite data. Atmospheric Research 290, 106794

Ridal, M., Sanchez-Arriola, J., Dahlbom, M., 2023: "Optimal use of radar radial winds in the HARMONIE numerical weather prediction system", submitted to Journal of Applied Meteorology and Climatology.

De Rooij, W, 2023: "Lateral mixing in shallow cumulus convection", Ph.D. Thesis TU Delft.

Savazzi, A., Nuijens, L., de Rooij, W., Janssens, M., Siebesma, P., 2023: "Momentum transport in (un)organized precipitating shallow convection", submitted to Journal Atm. Sci.

Solomon, A., et al., 2023: The winter central Arctic surface energy budget: A model evaluation using observations from the MOSAiC campaign. Elementa: Science of the Anthropocene 5 January 2023; 11 (1): 00104. doi: https://doi.org/10.1525/elementa.2022.00104

Juuso Suomi, et al., 2023: "Evaluation of surface air temperature in the HARMONIE-AROME weather model during a heatwave in the coastal city of Turku, Finland", submitted to Urban Climate

Summary of results (from January 2023 to June 2023)

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

The HIRLAM-C research programme (January 2016 - December 2025) is a research cooperation of the national meteorological institutes of Denmark, Estonia, Finland, Iceland, Ireland, Lithuania, Netherlands, Norway, Spain and Sweden. Research efforts are focused on the scientific development and implementation of the mesoscale analysis and forecast system Harmonie, and its associated ensemble prediction system HarmonEPS. A Harmonie Reference system is being maintained on the ECMWF HPC platform. Below, the main R&D and testing activities in the fields of data assimilation, the atmospheric forecast model, surface analysis and modelling, ensemble forecasting, and code efficiency and system aspects during the first half year of this special project period 2023-2025 are outlined.

The computational resources for the HIRLAM-C Special Project at ECMWF are primarily used for integration and pre-operational testing and meteorological performance assessment of new cycles of the Reference System for several domains. For scientific evaluations before this integration phase, the HIRLAM service create every year a pool of SBU resources at ECMWF. In 2023, until June most experiments which have been done to prepare for the release candidate for the next operational Harmonie Reference cycle, Cy46h1, have been using these pool resources (>40MSBU), and not yet the spsehlam special project resources; these have been kept reserved for the actual pre-operational testing of Cy46h1, which has been delayed but will start soon. The release of the Cy46h1 release candidate is expected to take place at the end of June, after which a large set of already planned month-long test runs will be carried out for three domains and up to 4 seasons in the summer and fall, both in deterministic and ensemble mode. It is expected that these will consume the spsehlam resources, and presumably pool resources as well, before the end of the year.

A) Data assimilation

A1: Development, operationalization and optimization of flow-dependent data assimilation methods

Work in 2023 has focussed on finalizing the upgrade of 4D-Var to Harmonie Cy46h and to enable it to run stably and efficiently with all available observations types in both Cy43h and Cy46h. In the past months, most issues with computational performance and observation types in combination with 4D-Var have been solved on local computers. In the remainder of 2023, extensive integration and performance testing of 4D-Var vs 3D-Var in Harmonie Cy46h1 will be done on the ATOS system.

Preparatory work has started on bringing Harmonie algorithmic developments into the new LAM OOPS data assimilation framework. Experiments in 2023 are focussing on expanding the OOPS 3D-Var system with the full available set of non-conventional data, and on exploiting OOPS 3D-Var for introducing new control variables, e.g. for hydrometeors, and assessing their impact. In 2024, prime attention will be given to exploring the OOPS 3DEnVar and consider how to implement the 4D-Var TL/AD setup in OOPS. Ongoing activities on coupled atmosphere-surface assimilation are described in section C (Surface).

A2: Optimal use of (high-density) atmospheric observations

At present, the Harmonie 3D-Var- and 4D-Var assimilation systems are capable of assimilating conventional data, cloud-free IR and microwave radiances from e.g. AMSU-A, AMSU-B, MHS, CRiS and IASI, radar radial wind and reflectivity volume data, GNSS ZTD, GPS-RO, SEVIRI IR and WV observations, Mode-S data, AMV's and scatterometer winds. In addition, work is ongoing on improving the quality control and bias correction for e.g. radar radial winds, GNSS STD and gradient products, and observations from private weather stations and smart phones.

In the past year, the amount of radiance information used in local operational suites has increased significantly, mostly through a better handling of surface-sensitive microwave and IR channels by means of dynamic emissivity maps. An improved treatment for correlated spectral channels from IASI has been developed. A footprint operator has been introduced for microwave radiances, tested successfully for use with low-peaking channels, and adapted to 4D-Var. Use of this footprint operator is expected to be beneficial primarily in areas with great transitions in the surface, e.g coastal zones and near the sea ice edge. First observing system experiments over Svalbard showed neutral impact. The option to assimilate all-sky radiances has been implemented in Cy46h1, and activities have begun to shift towards assessment of all-sky microwave radiances.

Work has started on a Eumetnet/E-GVAP study to compare and combine GNSS ZTD, STD and gradients products, and assess their individual and combined impact over several areas. For radar data, parallel studies have been carried out to intercompare the quality control and assimilation of OPERA OIFS and Baltrad radar reflectivity and wind data. Baltrad data generally have shorter latency and thus are preferred for nowcasting purposes. This intercomparison work is now temporarily on hold, awaiting new data files from OPERA OIFS that will become operational during 2023.

EMADDC Mode-S-EHS wind and temperature data are now routinely available and used over a large part of Europe. Because of their quality and high spatial and temporal density over Europe, their usefulness is being explored for the verification of sub-km resolution runs.

In summer and fall 2023, an extensive tuning exercise will be carried out for the observation system as a whole for the Cy46h1 3D-Var on ATOS, revisiting background and observation error characteristics and rejection limits for quality control. Diagnostic tools to be used in this tuning have been collected, including newly developed methods e.g. for improved cloud detection and observation fit statistics.

A3. Optimization of data assimilation setups at sub-km resolutions and for the nowcasting range

Efforts have become more focused on the use of data assimilation on nowcasting time scales and hectometric spatial scales. Technically, sub-hourly cycling has been enabled and first experiences have been gained. The main aim this year is to experiment extensively with sub-hourly cycling setups and assess the challenges for assimilation on sub-hourly and hectometric spatial scales, particularly the issue of model spinup. The focus is on assimilating data sources with sufficiently short latency (10-20min): SYNOP, EMADDC Mode-S, radar (Baltrad), NetAtmo/WOW stations and various sources of (geo and polar) satellite data. A first implementation has been made of an improved cloud ingest system, and this is being tested in the nowcasting range. 4D-Var is being assessed in a 500m resolution local pre-operational configuration.

Problems with correlated observations become more pronounced at higher temporal and spatial resolution. For 4D-Var especially, temporal correlations have been shown to be an issue to be tackled. For the handling of spatially correlated observations, error modelling approaches are being considered as a more sophisticated alternative to standard thinning or superobbing practices.

B) Atmospheric forecast model

B1. Studies to eliminate systematic model errors for clouds, fog and (very) stable boundary layer conditions:

A long-standing forecasting problem has been the prediction of fog under stable boundary conditions, especially over sea: fog generally building up too quickly, having a too large extent (especially over sea), too much cloud water, and starting a too quick and too strong cooling over sea. To tackle this, a CCN density profile has been introduced in the ICE3 microphysics in Cy43h2.2, and this is being tested in combination with near-real-time observed aerosol values from CAMS in Cy46h1. For future experimentation, a set of fog cases plus associated data is being collected for 1- and 3-D testing for a variety of domains and fog conditions. The idea is that this set of cases may triggered from the Harmonie testbed to quickly provide experimental data when testing new ideas for potential improvements. Earlier studies have shown that modelled fog development can be as sensitive to the assumed shape parameters of the cloud droplet size distribution and/or terminal fall speeds of solid hydrometeors as it is to the underlying aerosol or cloud droplet number concentration assumptions (Contreras et al. 2022). These sensitivities are being explored further in the context of SPP perturbations for HarmonEPS (see section D).

A long-standing problem in all NWP models is the treatment of the weakly stable and very stable boundary layer (WSBL, VSBL). Especially in the VSBL, the assumptions of horizontal homogeneity underlying the turbulence scheme are no longer valid, and the details of actual surface heterogeneity have a crucial impact on turbulence behaviour and heat transport. This problem is especially important for the representation of highly heterogeneous urban areas, and will require new modelling approaches for tackling the roughness sublayer (see section C1). Both an intrinsically stochastic turbulence scheme and a data-driven approach to derive a stochastic stability correction to model turbulence will be used to study the transition of turbulence from weakly stable to very stable conditions, and its impact on energy balances and profiles close to the surface.

B2. Improved description of the cloud-radiation – microphysics- aerosol interaction:

A dataflow for the initialization and consistent handling of near-real-time CAMS aerosols in the forecast model has been prepared for introduction in Cy46h1: the ingest of CAMS fields and computation of mass mixing ratios, the handling of cloud droplets in the processes of auto-conversion, sedimentation and collision, and the handling of effective radius and extinction in the radiation. Full integration testing will take place in summer and fall 2023. Research is ongoing on a parametrization for the activation of aerosols, depending on e.g. TKE and vertical velocity. Impact studies for both near-real-time and climatological CAMS aerosol information have shown the added value of the NRT aerosol information in specific weather situations such as fog and Sahara dust intrusions. Also, using near-real-time aerosol values rather than aerosol climatologies has been shown to lead to significant improvements in model cloud behaviour on days with thick and low clouds (see fig.1 below).

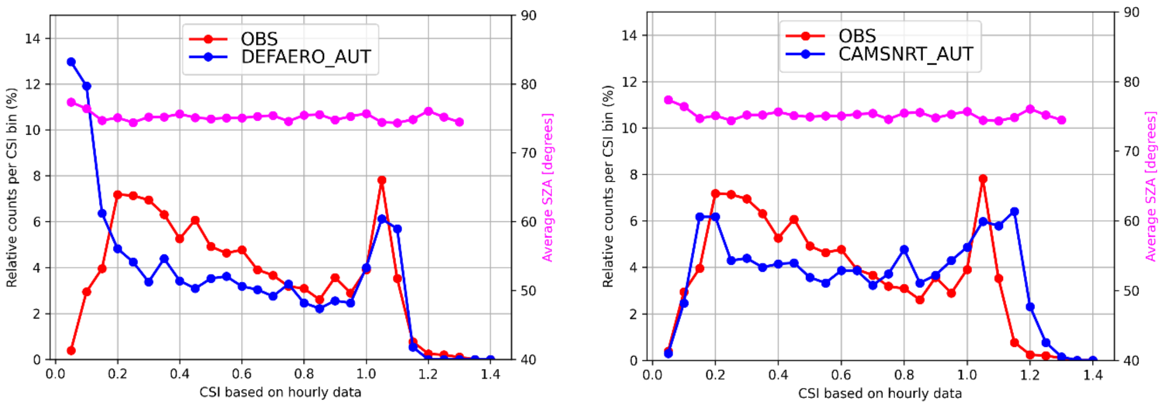

Fig.1: The figure on the left shows the outcomes of a comparison of the model clear-sky irradiance (CSI, based on hourly data) using standard aerosol climatology (blue curve) vs observed CSI values (red curve). The figure on the right shows the same, but for a model with aerosol values initialized by near-real-time CAMS data. It can be seen that for low CSI values (thick clouds) the behaviour of the model with aerosols from CAMS is much more realistic than when using aerosol climatology.

After having provided a realistic initialization of aerosols, the next step is to consider the introduction of dynamically evolving condensating nuclei density concentrations by means of a second moment microphysics scheme, LIMA, which will become available for Harmonie in Cy49. LIMA studies by the mesoNH team have shown the need for very careful stepwise testing and tuning of this more realistic but also much more complicated scheme. In parallel, work will need to be done to assess how the additional computational cost of this scheme can be reduced, or met by speedups elsewhere.

In Cy49, it is intended to replace the present old IFS radiation scheme by a LAM version of the ECRAD scheme, which represents a much more modern, maintainable and faster code. Work on the development and scientific validation of this scheme has started between DMI and Meteo-France.

B3. Sub-km resolution modelling

Most HIRLAM services have implemented, and are running routinely, sub-km (150-750m) horizontal resolution model setups. In the first half of 2023, attention has focused on the impact of increasing the horizontal and vertical resolution on boundary layer dynamics and thermodynamics, resolved versus unresolved processes as a function of resolution, the impact of nesting and coupling hydrometeors, and the development of quasi-3D turbulence and radiation.

Vertical resolution increase has been shown to contribute significantly to a good performance of several hectometric setups. Experiments have also confirmed the importance of a large domain size and a great sensitivity to the use of all available hydrometeors in the nesting model in the LBC coupling. At resolutions of 200m and finer, dynamical instabilities are observed near steep orography in the Alps for which no good solution has been found yet. In other domains, 100-200m resolution models appear to be stable enough at least for dry conditions, but some deficiencies have been seen in moist atmospheric processes. For stable boundary layer conditions, first results indicate that a stochastic turbulence parametrization may be needed to create realistic boundary layer clouds.

Sub-gridscale 3D cloud shading effects can be represented in the SPARTACUS scheme in ECRAD, which is intended to replace the present Arome/Harmonie radiation scheme in Cy49. To represent 3D cloud shading effects in hectometric models, it is intended to run ECRAD on a coarser grid (and timestep) than the rest of the forecast model, and use the fine-grid cloud fields to compute SPARTACUS inputs related to cloud variability. The handling of two grids in the forecast model will require externalization of the radiation from the rest of the physics.

A scale-aware convection parametrization has been developed to handle the grey zone where shallow convection is still only partially resolved. The scheme has been tested successfully in a near-real-time double-nested sub-km model setup over the Netherlands, against both an LES model and remote sensing observations from supersite Cabauw.

C) Surface analysis and modelling

C1: Improving the sophistication of surface model components

In the past two years, many tests have been done with the new many-layer diffusion soil, extended snow and snow-over-vegetation schemes (DIF, ES, MEB) and the new SEKF assimilation scheme for soil and snow. This combination of schemes, with some fixes, has been implemented in Cy46h1 and pre-operational testing is taking place on ATOS this summer. One problem which still remains to be solved is that unrealistically low snow temperatures are occasionally seen in the Alps which appear to originate from the snow analysis.

In the context of CARRA, the sea ice model SICE has been tested both offline and inline, still without sea ice assimilation. This setup, although far from perfect, has proven to outperform ERA-5 in the Arctic region. Suggestions have been made how to better represent the mixed snow-ice layer over sea and lakes, which should be taken into consideration in the near future.

A new parametrization for surface fluxes over the ocean, ECUME-6, has been tested. In climate runs, the scheme was shown to overestimate the latent heat flux over sea, leading to cloud and radiation biases there. This overestimate, with respect to both observations and the original ECUME scheme, was confirmed by a direct comparison of the fluxes over sea from ECUME-6 and ECUME against EUREC4A field experiment flux and eddy covariance observations, the differences being largest for strong winds. Therefore it has been decided to retain ECUME as sea flux scheme in Cy46h1.

Online simulations with the roughness sublayer (RSL) scheme of Harman and Finnegan (2007) for canopy-induced turbulent mixing, are ongoing over Iberia and Sweden. This parametrization includes the concept of displacement height, and a more advanced computation of the roughness length z_{0m} as a function of leaf area index (vegetation sparseness) and atmospheric stability. Extensive offline validation experiments have shown improved performance when applying the RSL scheme against tower observations.

The move to hectometric models has given impetus to studies of high-resolution urban physiographic maps and more advanced options in the TEB scheme in urban environments, e.g. over Turku and especially the PANAME observation campaign over Paris for the Paris Olympics. Updates have been provided to the ECOCLIMAP-SG physiography for Greenland and Iceland. The physiographic workflow which is being developed to create 60m resolution physiographic data over Europe is progressing as planned. At present, in Harmonie/Surfex there is a strict interface between the atmosphere and surface at the lowest atmospheric model level, where state variables and fluxes are interchanged. Work has recently started on implementing and testing a new approach (Schoetter et al 2020) where urban (high) buildings may interact with a number of the lowest model levels, depending on their height. This approach may be more realistic for the treatment of heterogeneous urban landscapes in hectometric scale models.

C2. Enhanced use of satellite surface observations and more advanced surface assimilation

The new SEKF assimilation scheme which will be tested pre-operationally in summer and fall of 2023 enables increased use of satellite surface products. Preparations are being made for the subsequent progressive inclusion and testing of satellite retrieval data in the SEKF surface assimilation after, are soil moisture from Sentinel-1 C-band SAR and Metop ASCAT, OSI-SAF and MODIS sea ice concentration and sea ice surface temperatures, snow depth from Sentinel-1, and retrievals from various satellites in lake data assimilation. Most attention in this respect has been given to the assimilation of satellite snow extent, which is important in the cold season in areas where SYNOP snow cover stations are sparse or missing. To avoid errors in the screening/thinning and interpolation of raw satellite data, the so-called “snow barrel” method has been developed by FMI for the assimilation of satellite snow extent data. Regular production of “snow barrel” data has started at FMI, and experimentation with their assimilation is ongoing.

In the coming years it is intended to move stepwise towards a more advanced EnKF or EnVar algorithm which can be used as a basis for a strongly coupled atmosphere-surface data assimilation system and which would permit a more direct assimilation of satellite surface radiances and other surface remote sensing observations. An EnKF setup has been developed for testing these ideas at SMHI while a 3D-EnVar approach is being pursued by MeteoFrance.

D) Probabilistic forecasting

In the EPS area, scientifically the focus has remained on introducing and optimizing model perturbations using stochastic perturbation of model parameters (SPP). Presently, a set of 18 atmospheric parameters has been identified which has shown some appreciable impact, six of them providing most of this. The computational cost of adding this reduced SPP setup of 6 parameters (with a mix of lognormal and uniform distributions) to the default initial and LBC perturbations has been shown to increase only marginally, after the decision to generate and apply the perturbation pattern only once per hour rather than every time step. The reduced setup will be tested for operational introduction in Cy46h1 in the summer and fall. Experiments studying the behaviour of correlated or anti-correlated perturbed parameters are ongoing. Initial experiments using the same perturbation pattern for two correlated parameters resulted in less bias and better spread.

New parameters which are being introduced in SPP for testing are the size distributions of solid water species and hydrometeor terminal fall velocities. In the microphysics, the values adopted for these distributions and speeds have large uncertainty, and under certain circumstances strongly affect CDNC and convection evolution, cloud dynamics, radiation and precipitation types and amounts. Additionally SPP is being extended with surface parameters. Surface parameters which are presently undergoing testing are the vegetative heat capacity (CV) and minimum surface resistance (RSMIN); the framework for introducing surface perturbations has been technically tested and found to function properly for CV. Also, more advanced LAI perturbations are being considered.

E) Computational efficiency and machine learning aspects

Various tests have been carried out for Harmonie-Arome in single vs double and mixed precision, for the forecast model but also for the 3- and 4D-Var data assimilation and ensemble systems, see sections A and D above. The studies on single precision (SP) in the Harmonie forecast model so far have generally shown no significant loss of accuracy, with runtime savings of ~35-40%, mainly because data volumes get smaller (memory cache can hold larger arrays, data involved in MPI passing is halved). For 4D-Var, some problems remain with using SP in minimization. A recent study by MetEireann on the use of SP vs double precision DP in HarmonEPS showed a slight but consistent degradation when using SP in terms of scores, a significant impact of SP vs DP on the perturbation patterns of SPP model perturbations in the ensemble, and occasional crashes in SP. Modifications are being made to tackle these issues.

The variety of applications of machine learning within the HIRLAM community is expanding: in surface physiography (the development of a workflow for deriving O(60m) resolution physiographic datasets for European domains), data assimilation (the development of observation operators, quality control of observations, tuning of assimilation parameters) and the forecast model (emulation of 1D and 3D radiation, optimization of parameter settings in the forecast model and their perturbations in the EPS system), and model postprocessing (ensemble calibration, improved fog/visibility forecasting, more localized forecasts). Staff at DMI have considered how to transfer the use of machine learning for speeding up radiation computations for the ECRAD scheme in the global IFS model to Harmonie-Arome in Cy49t, and how to best apply this in combination with hectometric-scale setups with quasi-3D radiation (see section B). Several other options for ML emulation for computational efficiency have been identified but need further exploration.