



ALADIN - HIRLAM NEWSLETTER

No. 1, September 2013



The Blue Lagoon, Iceland, by Jozef Vivoda

The joint HIRLAM All Staff Meeting 2013 and ALADIN 23st Workshop took place in Reykjavik, Iceland, from April 11th till 15th 2013.

HIRLAM-B Programme, c/o J. Onvlee, KNMI, P.O. Box 201, 3730 AE De Bilt, The Netherlands

ALADIN Programme, c/o P. Termonia, IRM, Avenue Circulaire 3, 1180 Bruxelles, Belgium

Contents

Introduction, Tilly Driesenaar and Patricia Pottier.....	2
---	----------

Papers presented at the combined ALADIN 23th Workshop / HIRLAM All-Staff Meeting 2013 15 - 19 April, Reykjavik
--

Mode-S Enhanced Surveillance derived observations from multiple Air Traffic Control Radars and the impact in hourly HIRLAM, Siebren de Haan.....	4
State and parameter estimation with the extended Kalman filter for soil analysis: a feasibility study, Rafiq Hamdi, Alberto Carrassi, Stéphane Vannitsem, Piet Termonia.....	9
GLAMEPS and HarmonEPS: LAM ensemble prediction systems under development, Inger-Lise Frogner, GLAMEPS team and HarmonEPS team.....	16
Finite elements used in the vertical discretization of the fully compressible forecast model ALADIN-NH, Jozef Vivoda, Petra Smolíková.....	31
Status of ARPEGE and AROME physics, Yves Bouteloup.....	47
Experiences with HARMONIE at IMO, Bolli Palmason, Hróbjartur Thorsteinsson, Sigurdur Thorsteinsson, Gudrun Nina Petersen.....	52
Can HARMONIE be accelerated with GPU's or coprocessors?, Enda O'Brien and Adam Ralph....	64
FROST-2014 – First experiences of deterministic Harmonie in the Sochi region, Sami Niemelä Evgeny Atlaskin, Sigbritt Näsman and Pertti Nurmi.....	72

Other papers

New model cycle 38t1, Joël Stein and Marc Tardy.....	79
---	-----------

Annex A	Programme and participant list of the ASM 2013/ALADIN 23th Wk
----------------	--

Introduction

Tilly Driesenaar and Patricia Pottier

This is the first edition of the combined Newsletter of the HIRLAM and ALADIN consortia. We decided to join forces because the researchers from both consortia are working closer and closer together. Furthermore, the number of contributions is continuously decreasing because we publish increasingly in peer reviewed literature.

This Newsletter is devoted to presentations made by the ALADIN and HIRLAM staff in the joint HIRLAM/ ALADIN Workshop/All Staff Meeting (ASM) that took place on 15-19 April 2013 in Reykjavik, hosted by the Icelandic Meteorological Organisation (IMO). We thank the staff of IMO for providing the pleasant and high tech meeting facilities at the Hilton hotel in Reykjavik. For the first time there we had live internet broadcast, which was appreciated very much by quite a number of distant followers who were not in the opportunity to travel to Iceland.

As usual the scientific part of the programme took place in fully plenary sessions. After a warm welcome by the Dr. Arni Snorrason, director general of IMO, there was a short opening session on Tuesday morning, in which overviews were given of the progress and status of the ALADIN, HIRLAM and LACE programmes. This was followed by six sessions. As usual we had the thematic sessions on data assimilation and use of observations, probabilistic forecasting and LAMEPS, dynamics, model physics, and system aspects and verification. And this year a session operational experiences was added. The programme is listed in Appendix A and the presentations can be found at <http://www.cnrm.meteo.fr/aladin>.

We hope you enjoy the first ALADIN-HIRLAM Newsletter!





Mode-S Enhanced Surveillance derived observations from multiple Air Traffic Control Radars and the impact in hourly HIRLAM

Siebren de Haan

1 Introduction

Upper air wind is one of the most important parameters to obtain a good analysis using a data assimilation method. High resolution observations are beneficial for Numerical Weather Prediction with a rapid update cycle [1, 6]. Local observations have a localized and short term effect [4, 5]. For numerical nowcasting, rapid availability is also a crucial factor. Furthermore, numerical nowcasting will become more accurate on longer forecasting time-scales when more data from a larger area becomes available. Since 2008, KNMI is receiving Mode-S Enhanced Surveillance (EHS) data from the Air Traffic Control (ATC) radar at Schiphol airport. This dataset is made available by the ATC The Netherlands (LVNL). The coverage of this dataset is the Dutch airspace, retrieving information from all aircraft within a 200 NM range of the tracking radar. This data can be used to derive wind information with good quality when compared to NWP and AMDAR [2, 3]. However, one should note that temperature information derived from Mode-S EHS is of less quality. Assimilation of wind and temperature observations derived from data received from LVNL in the HIRLAM model showed a positive impact up to four hours into the forecast [6].

This short report addresses the first attempts to assimilate Mode-S EHS derived wind information from a larger area exploiting the Dutch, German and Belgium Mode-S EHS radar information. A very straightforward thinning scheme is applied. When assimilated in the hourly update cycle of HIRLAM7.4, an improvement in wind forecast is observed up to a forecast time of 9 hours for wind speed and wind direction. Especially, below 700 hPa a large improvement is observed up to 18 to 24 hours in range for wind direction forecasts.

2 Mode-S EHS

The Mode-S project started in 2008. Since then several studies have been performed on the quality of the data [2, 3] and the usage in numerical weather prediction [6, 4].

The concept of Mode-S EHS is created to become the next generation of air traffic management systems. A Mode-S EHS system consists of a user segment and a ground segment by exploiting (amongst others) the GNSS positioning technique. The user segment is an aircraft equipped with a Mode-S transponder; the aircraft determines its position using GNSS satellites. The ground segment consists of a Mode-S Enhanced Surveillance tracking radar which is capable of interrogating the Mode-S transponders for specific registers.

Within the European designated EHS airspace, all fixed wing aircraft, having a maximum take-off mass greater than 5,700 kg or a maximum cruising true airspeed in excess of 250 kts, intending to fly instrument flight regulation (IFR) must be Mode-S EHS compliant. An aircraft is compliant with Mode-S EHS when it provides basic functionality features (such as position and flight number) plus eight downlinked aircraft parameters. The downlinked parameters are selected altitude, roll angle, track angle rate, true track angle, ground speed, magnetic heading, true airspeed, and Mach number.

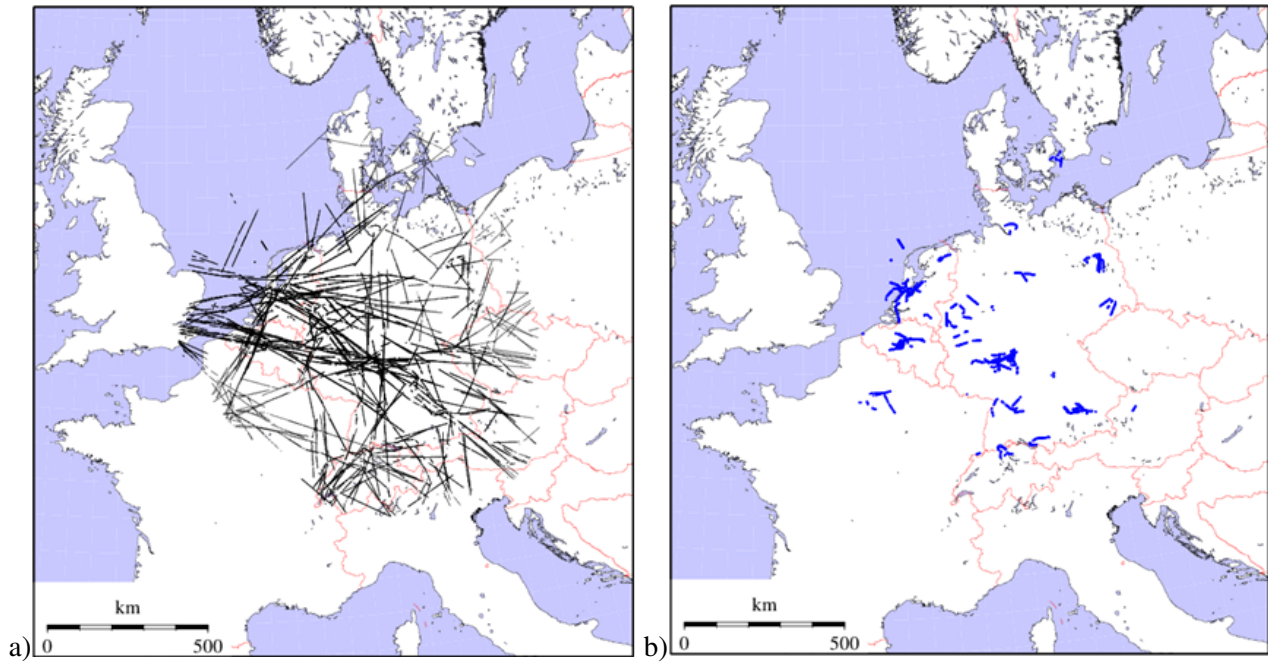


Figure 1: An example of the quality controlled and corrected wind observations in a 15 minute interval. Panel a) shows all observations; panel b) shows the observations below FL100.

An aircraft observation closely linked to Mode-S EHS is Mode-S MRAR. It uses the same ground and user segment, but the ATC radar interrogates a different register, the so-called Meteorological Routine Air Report. This register contains already wind and temperature information of good quality. Unfortunately, although aircraft are obliged to respond to a Mode-S radar request, only around 5% of the aircraft transmit this requested information since it is not obligatory [8].

2.1 Observations from Mode-S EHS

From the downlinked information wind and temperature can be derived.

Wind information is inferred from the vector difference between the air vector (airspeed and heading) and the ground vector (ground speed and track angle). Temperature is calculated using the Mach number and the observation of the airspeed. After heading correction and airspeed correction the wind information derived from Mode-S EHS information has good quality compared to AMDAR and NWP. Details of the derivation and quality can be found in [2] and [3]. In short, the wind vector \mathbf{V} is the difference between the ground vector \mathbf{V}_g and the air vector \mathbf{V}_a .

$$\mathbf{V} = \mathbf{V}_g - \mathbf{V}_a, \quad (1)$$

and

$$T = K(V_a/M)^2, \quad (2)$$

where K is a constant, V_a is the airspeed and M is the Mach number.

Since the heading is reported with respect to the magnetic north, a heading correction must be applied to obtain a heading with respect to true north. Additionally, an aircraft dependent heading correction needs to be employed [2]. On top of the heading correction, an airspeed correction is deployed. The accuracy of an airspeed observation should be within 3%, according the FAA regulations (FAA). The error in airspeed and error in wind are directly related (see 1). Airspeed correction can improve the wind observation by roughly 2%, see [3]. Therefore a dynamic heading correction lookup table and a static airspeed correction lookup table were constructed. Both tables were based on continuous comparison of the measured heading and airspeed



Figure 2: Model domain used in this study.

using Numerical Weather Prediction (ECMWF) data. These lookup tables are aircraft dependent and the heading correction table needs to be updated regularly.

2.2 Mode-S EHS MUAC observations

The Mode-S EHS radar dataset used in this study is collected by EUROCONTROL Maastricht Upper Area Control (MUAC). The data set consists of all Mode-S EHS radars from the Dutch, German and Belgium ATC organizations. A copy of the operational radar dataset is made available in 15 minute batches with a delay of approximately 10 minutes. Figure 1 shows an example of the locations of all quality controlled and corrected wind observations in a 15 minute interval. In total more than 73 thousand observations are available 1a), with more than 6.6 thousand observations below Flight Level 100 (approximately 3km) 1b).

3 Numerical Weather Prediction

A hourly cycle of HIRLAM (v7.4) is used to show the impact of assimilating Mode-S EHS. The resolution of the model is 11 km with 40 atmospheric levels. The hourly cycle uses a re-forecast every hour in order to be able to assimilate observations that generally have a long latency (scatterometer and AMSU-A observations from polar orbiting satellites, and radiosonde). The one hour forecast of the delayed run is then used as the first guess in the "real-time" run. The delayed run has (obviously) a cycle of one hour. The assimilation method deployed here is the HIRLAM 3D variational method, see for example [7].

Figure 2 shows the domain of the hourly run. A first order, rather basic, thinning procedure is applied to the MUAC derived observations. The model domain is separated into 50 squared kilometre boxes each with a thickness of 300 m and only one single observation is selected per box; the observation closest to the assimilation time.

4 Impact

For a three week period two H11 runs have been run in parallel. Both are run in operational mode and use exactly the same boundaries and observations, except of course the MUAC Mode-S EHS derived observations. All observations used in the assimilation are also used to compare previous model forecasts at different

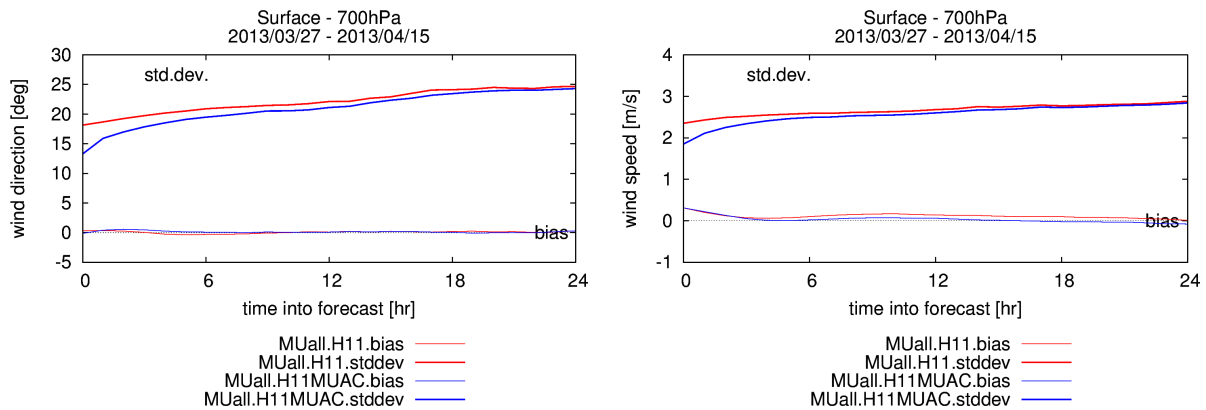


Figure 3: Impact on wind direction and wind speed. Comparison of Mode-S EHS MUAC observations between surface to 700hPa with independent forecasts ranging from +00 (analysis) to +24.

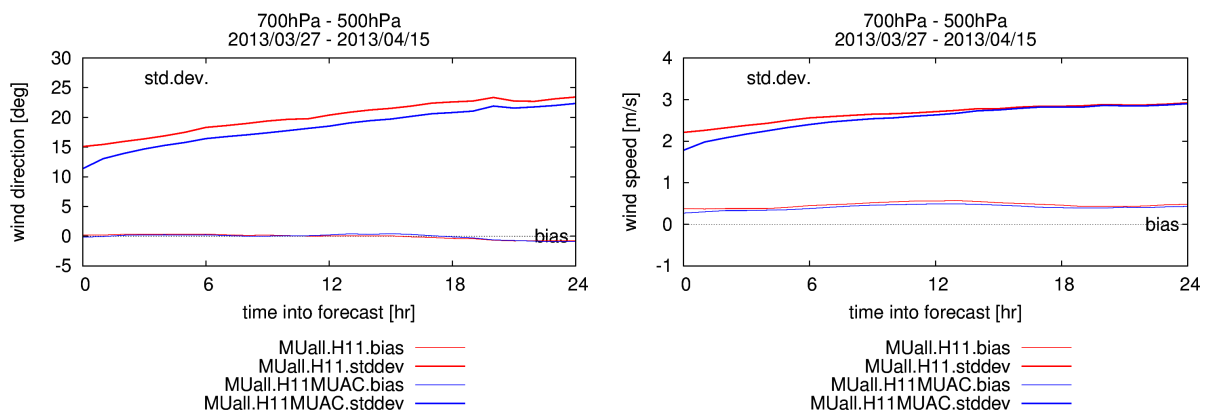


Figure 4: Impact on wind direction and wind speed. Comparison of Mode-S EHS MUAC observations between 700hPa and 500hPa with independent forecasts ranging from +00 (analysis) to +24.

forecast lengths. The standard HIRLAM 3DVAR routine is (slightly) adapted to be able to perform this dynamic forecast comparison.

Figure 3 shows the impact on wind speed and direction from the surface to 700hPa and Figure 4 shows the impact between 700hPa and 500hPa. Both model runs show comparable and small biases over the whole forecast range for both parameters indicating that there is no gross error in the observations. One remark has to be made on the small wind speed bias in the lowest level, which could be related to not optimal airspeed correction tables for lower airspeed; this will be investigated in future research. The positive impact on wind direction is most pronounced in the 700hPa to 500hPa layer and is visible over the whole forecast range. Also, the wind direction forecast in the lowest layer shows an improvement when MUAC data is assimilated, however after 12 to 15 hours the impact is neutral. Wind speed forecasts improve up to 6 hours in the lowest level and up to 9 hours in the level between 700hPa and 500hPa.

5 Conclusion and outlook

In this short paper it was shown that high resolution numerical weather prediction models benefit from high resolution observations of good quality. Impact on wind direction and wind speed forecast is observed in an operational comparison of two identical runs, one with Mode-S EHS MUAC derived observations assimilated

and one without this data set.

The positive impact is still visible after 24 hours for wind direction in the level between 700hPa and 500hPa. For the surface level the impact on wind direction forecast is positive up to 12 hours.

The wind speed forecasts improve up to at least 9 hours into the forecast.

Despite the cruel thinning procedure a positive impact is observed. Only a fraction of the wind information is used in the assimilation. Next research steps, with HARMONIE within SESAR WP11.2, will be undertaken to improve the usage of Mode-S EHS derived observations.

5.1 Acknowledgement

This study was funded by the Knowledge and Development Centre Mainport Schiphol in The Netherlands (KDC, <http://www.kdc-mainport.nl>). The author thanks Maastricht Upper Area Control Centre of EUROCONTROL for providing the raw Mode-S EHS data set.

References

- [1] S. G. Benjamin, B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schlatter. Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR and mesonet observations via the RUC hourly assimilation cycle. *Monthly Weather Review*, 138(4):1319 – 1343, 2010.
- [2] S. de Haan. High-resolution wind and temperature observations from aircraft tracked by Mode-S air traffic control radar. *J. Geophys. Res.*, 116:D10111–, 2011.
- [3] S. de Haan. An Improved Correction Method for High Quality Wind and Temperature Observations from Mode-S EHS. *KNMI-WR*, 2013. in preparation.
- [4] S. de Haan. Assimilation of GNSS-ZTD and radar radial velocity for the benefit of very short range regional weather forecast. *Q. J. Roy. Met. Soc.*, 2013. accepted.
- [5] S. de Haan, G.J. Marseille, J.C.W. de Vries, and J.P.J.M.M de Valk. Impact of ASCAT Scatterometer Wind Observations on the High Resolution Limited Area Model (HIRLAM) within an Operational Context. *Weath. Forec.*, 28(2):489–503, 2013.
- [6] S. de Haan and A. Stoffelen. Assimilation of High-Resolution Mode-S Wind and Temperature Observations in a Regional NWP model. *Weath. Forec.*, 27:918–937, 2012.
- [7] M. Lindskog, H. Järvinen, and D. B. Michelson. Assimilation of radar radial winds in the hirlam 3d-var. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 25(10-12):1243 – 1249, 2000. First European Conference on Radar Meteorology.
- [8] B. Strajnar. Validation of mode-s meteorological routine air report aircraft observations. *Journal of Geophysical Research: Atmospheres*, 117(D23), 2012.

State and parameter estimation with the extended Kalman filter for soil analysis: a feasibility study

Rafiq Hamdi, Alberto Carrassi, Stéphane Vannitsem, Piet Termonia

Introduction

In numerical weather prediction (NWP) model, soil initialization plays an important role [Viterbo and Courtier, 1995; Fischer et al., 2007]. However, the lack of accessible and reliable observations has been one of the primary causes for the employment of simplified strategies for soil initialization in operational NWP. Currently, in most NWP centers, soil moisture analysis systems are based on analyzed or observed screen-level variables. An optimal interpolation technique (OI) developed by Mahfouf (1991) based on short-range forecast errors of 2m temperature (T2m) and relative humidity (RH2m) is operational in a number of centers: Météo-France [Giard and Bazile, 2000], Environment Canada [Belair et al., 2003], and the European Center for Medium-Range Weather Forecasts [Douville et al., 2000] among others.

However, some fundamental limitations are inherent with the use of the OI [Mahfouf et al., 2009]. Among these, the fact that it does not allow for the use of nonlinear observation operator (note that the physical link between near-surface observations and soil variables is highly nonlinear), the OI coefficients are usually derived using very simple assumptions and/or model configurations and that, once the OI scheme has been optimized for the assimilation of the observation at synoptic times, it is not easily adaptable to incorporate observations available at different times (a typical circumstance when dealing with satellite on-board measurements).

A particularly promising approach to address the above mentioned shortcomings is the use of an Extended Kalman Filter (EKF). The EKF and a simplified version known as simplified extended Kalman filter (SEKF), in which a static forecast error covariance matrix is adopted, have been successfully applied to the assimilation of direct and/or indirect soil observations [Mahfouf et al., 2009; Drusch et al., 2009; Draper et al., 2009; Albergel et al., 2010; Draper et al., 2011; Mahfouf and Bliznak, 2011]. Recently, Carrassi and Vannitsem [2011] introduced an alternative formulation of the EKF where the uncertain model parameters are estimated along with the system state variables (STAEKF, Short Time Augmented Extended Kalman Filter). The algorithm combines the state-augmentation approach [Jazwinski, 1970] with the deterministic formulation of the model error dynamics [Nicolis, 2003]. With the aim of contributing to the current discussion on soil data assimilation strategies, we undertake a set of numerical "twin" experiments designed to test the STAEKF for the assimilation of screen-level observations using an offline version of the land surface model, Interactions between Surface, Biosphere, and Atmosphere (ISBA) [Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996] where errors in three important land surface parameters, the LAI, the albedo, and the RS_{min} (minimal stomatal resistance) are introduced. The performance of the STAEKF is studied in comparison with the EKF and the SEKF.

The land surface model ISBA

The soil temperature and moisture contents are evolved by the two-layer land-surface scheme Interactions Soil-Biosphere-Atmosphere (ISBA) based on the force-restore method [Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996]. The ISBA scheme evolves four prognostic variables: (1) T_s the surface soil temperature, (2) T_2 the deep soil temperature, and two prognostic variables to

describe water transfers in the soil: (3) a surface reservoir defined by its volumetric water content w_g ($\text{m}^3 \text{m}^{-3}$) representative of a depth d_1/C_1 where d_1 is an arbitrary depth of 1 cm and C_1 is a coefficient depending upon soil texture and soil water content, (4) a deep reservoir (including the root zone) defined by its volumetric water content w_2 ($\text{m}^3 \text{m}^{-3}$) representative of a depth d_2 that varies between 1 m and 3 m depending upon vegetation type.

The equations can be formally written as a dynamical system:

$$dx/dt = g(x, \lambda) \quad (1)$$

With the state vector $x = (T_s, T_2, W_g, W_2)$. The vector λ is taken to represent the set of model parameters. A detailed description of ISBA can be found in the study by Noilhan and Mahfouf [1996].

Surface assimilation schemes

Extended Kalman filter (EKF)

Let us write the model equations as a mapping from time t_k to t_{k+1} :

$$x_{k+1}^f = F(x_k^a) \quad (2)$$

where x^f and x^a are the forecast and analysis state respectively, of dimension I , and M is the nonlinear model forward operator. A set of M observations of the true system, stored as the components of an M -dimensional observation vector y^o , is supposed to be available at the regularly spaced discrete times $t_k = t_0 + k\tau$, $k = 1, 2, \dots$, with τ being the assimilation interval,

$$y_k^o = H(y_k) + \epsilon_k \quad (3)$$

with ϵ_k being the observation error, assumed to be Gaussian with known covariance matrix R and uncorrelated in time, and y is taken to represent the unknown true state; H is the (possibly nonlinear) observation operator mapping model variables to observation space.

The EKF analysis state update equation reads:

$$x^a = [I - KH]x^f + Ky^o \quad (4)$$

The analysis error covariance, P^a , is updated through:

$$P^a = [I - KH]P^f \quad (5)$$

The $I \times M$ gain matrix K is given by:

$$K = P^f H^T [H P^f H^T + R]^{-1} \quad (6)$$

with P^f being the $I \times I$ forecast error covariance matrix and H the linearized observation operator. The matrix P^f is obtained by linearizing the model around its trajectory between two successive analysis times. The model error is assumed to be random uncorrelated noise of which the effect is modeled by adding an error covariance matrix, P^m , at the forecast step so that:

$$P^f = M P^a M^T + P^m \quad (7)$$

with M being the tangent linear model. Note that while P^f is propagated in time and is therefore flow dependent, P^m is defined once for all and it is then kept constant.

Short time augmented extended Kalman filter (STAEKF)

The dynamical system [Eq. 1], is augmented with P model parameters that will also be estimated:

$$z^f = \begin{bmatrix} x^f \\ \lambda^f \end{bmatrix} = F z^a = \begin{bmatrix} M x^a \\ F^\lambda \lambda^a \end{bmatrix} \quad (8)$$

where $z = (x, \lambda)$ is the augmented state vector.

The augmented dynamical system F includes the dynamical model for the system's state, M , and a dynamical model for the parameters F^λ . In the absence of additional information, a persistence model for F^λ is often assumed so that $F^\lambda = I$ and $\lambda_{t_{k+1}}^f = \lambda_{t_k}^a$. The same choice has been adopted here.

The forecast/analysis error covariance matrix, $P_z^{f,a}$, for the augmented system reads:

$$P_z^{f,a} = \begin{pmatrix} P_x^{f,a} & P_{x\lambda}^{f,a} \\ P_{x\lambda}^{f,a^T} & P_\lambda^{f,a} \end{pmatrix} \quad (9)$$

where the $I \times I$ matrix $P_x^{f,a}$ is the error covariance of the state estimate $x^{f,a}$, $P_\lambda^{f,a}$ is the $P \times P$ parametric error covariance and $P_{x\lambda}^{f,a}$ the $I \times P$ error correlation matrix between the state vector, x , and the vector of parameters λ . These correlations are essential for the estimation of the parameters. In general one does not have access to a direct measurement of the parameters, and information are only obtained through observations of the system's state. At the analysis step, the estimate of the parameters will be updated only if they correlate with the system's state, that is $P_{x\lambda}^{f,a} \neq 0$.

The forecast error propagation in the STAEKF is given by:

$$P_z^f = C P_z^a C^T \quad (10)$$

with C being the STAEKF forward operator defined as:

$$C = \begin{pmatrix} M & \frac{\partial g}{\partial \lambda_{\lambda^a}} \tau \\ 0 & I_P \end{pmatrix} \quad (11)$$

where I_P is the $P \times P$ identity matrix. Equation (11) embeds the key feature of the STAEKF; the

presence of the term $\frac{\partial g}{\partial \lambda_{\lambda^a}} \tau$ allows for accounting for the contribution of the parametric error to the

forecast error as well as to the error correlation between model state and parameters.

As for the standard EKF, we need the observation operator linking the model to the observed variables. An augmented observation operator is introduced, $H_z = [H \ 0]$ with H as for the standard EKF. Its linearization, H_z is now a $M \times (I + P)$ matrix in which the last P columns contain zeros. The augmented state and covariance update complete the algorithm and are equivalent to Eqs of the EKF except that they refer now to the augmented system and the gain matrix has dimension $(I + P) \times M$ [Carrassi and Vannitsem, 2011].

Experimental set-up

We undertook a set of numerical “twin” experiments to develop the method to account for the model parameter error. The “observation” and evaluation data used in this study were generated from model output. The “twin” experiment is an important first step in the development of a data assimilation system. These types of experiments allow us to determine the feasibility of the assimilation approach under known conditions, namely the “truth”.

The forcing data for the land surface model consist of 1-hourly observations of air temperature, specific humidity, atmospheric pressure, incoming global radiation, incoming long-wave radiation, precipitation rate and wind speed derived from 10 summer seasons (1990-1999) ECMWF Re-analysis [ERA40, Uppala et al. 2005] that was dynamically downscaled to 10 km horizontal resolution over Belgium [Hamdi et al. 2012] using the numerical weather prediction limited area model ALADIN developed by the ALADIN international team [1997]. These data are then temporally interpolated to get data with the time resolution of the integration scheme of ISBA (300 s). These drivers are taken at a height of 20 m above ground level to ensure that they are representative of the local scale (10^2 - 10^4 m). We run ISBA in one offline single column mode for a 90 day period, and the forcing parameters are derived from the downscaling of the ERA-40 for the grid point that is closest to Brussels. Screen-level variables, namely 2 m temperature (T_{2m}) and relative humidity (RH_{2m}) used here were generated by the land surface model ISBA using the parameter specification, forcing data, and initial conditions described above. This simulation provided the “truth” data for the evaluation of our numerical experiments (hereafter referred to as the REF run) and the “observation” data for assimilation.

The data used to evaluate the assimilation included T_s , T_2 , w_g , and w_2 .

The REF run is contrasted with a degraded simulation to illustrate initial state and parameter error impact on state variables. This simulation (hereafter referred to as the “Open Loop” run) used the same model, parameters, and atmospheric forcing as the Ref run, but with degraded initial condition and parameters. The extended Kalman filter experiments started with the same degraded initial conditions as the “Open Loop” run assimilating 2m temperature and relative humidity for fixed times at 00, 06, 12, and 18 UTC.

The assimilation interval is $\tau = 6$ h, while the observational noise is drawn from a Gaussian, $N(0, R)$, with zero-mean and covariance given by the diagonal matrix R with elements:

$$\text{diag}(R) = (\sigma_{T_{2m}}^2, \sigma_{RH_{2m}}^2) = (1 K^2, 10^{-2}) \quad (12)$$

As explained in Mahfouf et al. (2009), the observation operator H , relating the state vector to the observation includes the model integration. The initial P^a and P^m required by the EKF are set as diagonal with elements:

$$\text{diag}(P_0^a) = (\sigma_{T_s}^2, \sigma_{T_2}^2, \sigma_{w_g}^2, \sigma_{w_2}^2) = (1 K^2, 1 K^2, 10^{-2}, 10^{-2}) \quad (13)$$

and

$$\text{diag}(P^m) = (\sigma_{T_s}^2, \sigma_{T_2}^2, \sigma_{w_g}^2, \sigma_{w_2}^2) = (25 \times 10^{-2} K^2, 25 \times 10^{-2} K^2, 4 \times 10^{-4}, 4 \times 10^{-4}) \quad (14)$$

(Mahfouf, 2007).

Parametric errors are introduced by perturbing either alternatively or simultaneously the LAI, the albedo and the RSmin. These parameters strongly influence the surface energy balance budget and partitioning, which in turn regulate the circulation patterns and modify the hydrological processes. According to the STAEKF setup, we compute (not shown) the functional dependence of the model equations on the uncertain parameters, the term $\partial g / \partial \lambda_k$, required in Equation (11), relative to the ISBA model for the LAI, the albedo, and the Rsmin.

For each summer in the period 1990–1999, a reference trajectory is generated by integrating the model with $LAI = 1 \text{ m}^2/\text{m}^2$, albedo = 0.2 and $RS_{\min} = 94 \text{ s/m}$. Around each of these trajectories, Gaussian samples of 100 initial conditions and uncertain parameters are used to initialize the assimilation cycles. The initial conditions are sampled from a distribution with standard deviation

$(\sigma_{T_s}, \sigma_{T_2}, \sigma_{w_g}, \sigma_{w_2}) = (5 \text{ K}, 5 \text{ K}, 1, 1)$, whereas LAI, the albedo, and the RSmin, are sampled with standard deviations, $\sigma_{LAI} = 0.5 \text{ m}^2/\text{m}^2$, $\sigma_{\text{albedo}} = 0.05$ and $\sigma_{RS_{\min}} = 50 \text{ s/m}$ respectively (Ghilain et al., 2011). The initial P_{λ}^a in the STAEKF read $P_{LAI}^a = 1 (\text{m}^2/\text{m}^2)^2$; $P_{\text{albedo}}^a = 10^{-4}$;

$$P_{RS_{\min}}^a = 5000 (\text{s/m})^2 \quad P_x^a \text{ is taken as in the EKF while } P_{x, \lambda}^a \text{ is initially set to zero.}$$

Results

Leaf Area Index

In the first set of experiments, only LAI is uncertain. Figure 1 displays the time series of the RMS estimation error in the four state variables for the EKF (red) and the STAEKF (blue). The root-mean-square (RMS) in the estimate of LAI obtained with the STAEKF is shown in the bottom-left panel, while the corresponding estimated error variance, P_{LAI}^a , in the bottom-right panel. The error is averaged over the ensemble of 100 experiments and over the ten summers. Results relative to an experiment without assimilation (referred to as Open Loop) are also shown for completeness. Note that in view of the rapid fluctuations of T_s and T_2 , to improve visualization the corresponding error curves are displayed every 150 time-steps (equivalent to 12.5 h) in all the figures.

For surface and deep temperature, T_s and T_2 , the EKF and STAEKF give similar results for the first

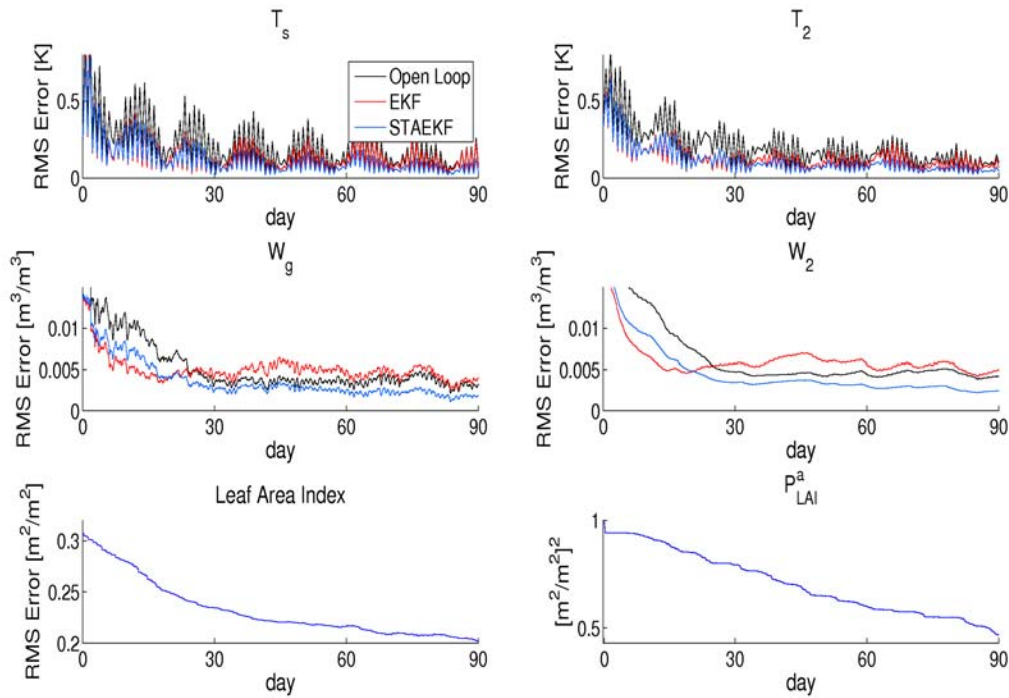


Figure 1: RMS estimation error in the four state variables for the EKF (red) and the STAEKF (blue). The RMS error in the estimate of LAI obtained with the STAEKF is shown in the bottom-left panel, while the corresponding estimated error variance in the bottom-right panel.

month with a slight advantage for the STAEKF at the error peaks, while a larger improvement is visible afterward. For the water contents, w_g and w_s , the STAEKF takes a longer time to exhibit a benefit over the EKF, but after about 20 days the STAEKF estimation error attains a lower level. After a sufficiently long time the estimation error essentially comes from the parametric error and the STAEKF becomes advantageous. The estimation of LAI with the STAEKF is also satisfactory as depicted by the error reduction observed in the bottom-left panel of Figure 1. The reduction of the actual error is tracked by the associated estimated variance, as it can be observed in the bottom-right panel.

By looking at the error level for the open loop experiment, the benefit of implementing a data assimilation technique is evident; then the slight advantage of the STAEKF over the EKF comes from its additional parameter estimation feature. After around 20 days the component of the prediction error coming from the growth of the previous analysis error, becomes progressively smaller and the error mainly originates from the misspecification of the parameters.

Minimum Stomatal Resistance

In analogy to Figures 1, Figure 2 refers to experiments with error in RS_{min} . In this case the overall performance of the STAEKF is less satisfactory. The estimation of RS_{min} appears more difficult (see bottom-left panel) and although a trend of error reduction is visible, it is now lower than in the two previous cases, with a parametric error reduction of about 40 and 35% in the LAI and albedo (Carrassi et al. 2012), respectively, and a reduction of 14% for RS_{min} now. The error is also subject to significant fluctuations. As a consequence, the STAEKF is not able to reach the skill of the EKF, although it improves over the open loop (Carrassi et al. 2012). The different behavior of the STAEKF when dealing with different parameters was already observed in the study by Carrassi and Vannitsem (2011) in the context of a low-order atmospheric model, and it was found connected to the sensitivity of the model dynamics to the uncertain parameters. In the STAEKF, the effect of parametric error is

accounted for on the basis of a first-order approximation, a truncation which turns out to be poorly accurate in certain cases, such as the RSmin here, affected by very large errors.

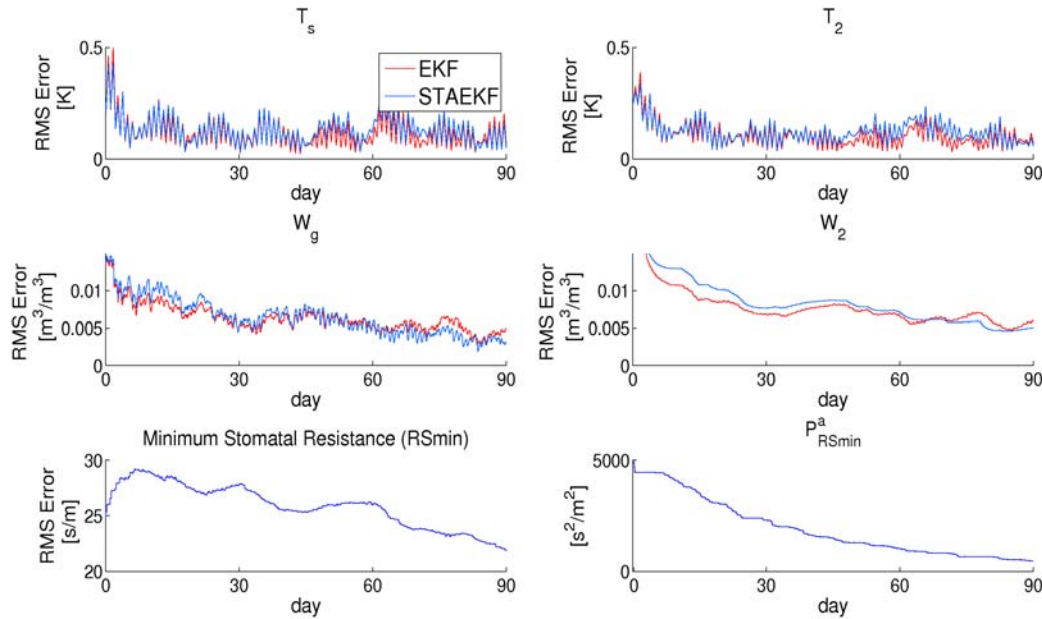


Figure 2: Same as Figure 1 but for Rs_{min} .

Conclusions

This paper describes a first application of the STAEKF scheme to soil analysis, for the assimilation of screen-level observations into an off-line version of the land surface model, ISBA. A twin experiment approach has been adopted for the summers for the decade 1990–1999. Model error is simulated by perturbing either independently or simultaneously the LAI, the albedo, and the minimum stomatal resistance, three key parameters in soil modeling, usually affected by significant errors. The effect of the model error is implicitly accounted for in the EKF forecast error covariance, while in the STAEKF the uncertain parameters are estimated at the analysis time along with the system's state. Other schemes such as an EKF including a bias correction scheme also exist in the literature (Dee and da Silva, 1998).

Results indicate that the STAEKF is able to retrieve the true value of the uncertain parameter and progressively reduces the associated estimation error. The accuracy of these estimates is inherently related to the type of parameter to be estimated, given the model sensitivity to the specific parameter and the accuracy of the short-time approximation at the core of the STAEKF. These conditions have to be verified case by case. The present results suggest that the on-line estimation of LAI and/or the Albedo (Carrassi et al. 2012) helps in improving the soil analysis skill to a larger extent than the estimate of RSmin.

The rate of error convergence in the STAEKF is related to the initial parametric error variance. Such a rate could be accelerated, to a certain extent, by an optimal tuning of the initial error covariance, a case-dependent procedure which has not been implemented in this feasibility study. By reducing the parametric error a better guess for the system state can be obtained and this in turn improves the analysis field and again the accuracy of the parameter estimate. Moreover, given that this feature is incorporated using the short-time formulation (Carrassi and Vannitsem, 2011a), the additional computational cost with respect to the standard EKF is almost negligible.

This proof-of-concept study is a first step in the validation of the STAEKF scheme and encourages its use in a more realistic model and observational scenario, possibly in the context of a full three-dimensional NWP model. This research is currently on-going.

References

- Albergel C, Calvet J-C, Mahfouf J-F, Rudiger C, Barbu AL, Lafont S, Roujean J-L, Walker JP, Crapeau M, Wigneron J-P. 2010. Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France. *Hydrology and Earth System Sciences. Discussion 7*: 1705–1744.
- Bélair S, Crevier L-P, Mailhot J, Bilodeau B, Delage Y. 2003. Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part I: Warm season results. *Journal of Hydrometeorology* 4: 352–370.
- Carrassi A, Vannitsem S. 2011. State and parameter estimation with the extended Kalman filter: An alternative formulation of the model error dynamics. *Quarterly Journal of the Royal Meteorological Society* 137: 435–451.
- Carrassi, A., R. Hamdi, P. Termonia, S. Vannitsem. Short time augmented extended Kalman filter for soil analysis: a feasibility study. *Atmospheric Science Letters*, DOI: 10.1002/asl.394, [Volume 13, Issue 4](#), pages 268–274, October/December 2012.
- Dee D, da Silva A. 1998. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society* 117: 269–295.
- Douville H, Viterbo P, Mahfouf J-F, Beljaars ACM. 2000. Evaluation of optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Monthly Weather Review* 128: 1733–1756.
- Draper C, Mahfouf J-F, Walker J. 2009. An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *Journal of the Geophysical Research* 114: D201004, DOI: 10.1029/2008JD011650.
- Drusch M, Scipal K, de Rosnay P, Balsamo G, Andersson E, Bougeault P, Viterbo P. 2009. Towards a Kalman filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophysical Research Letters* 36: L10401. DOI: 10.1029/2009GL037716.
- Ghilain N, Arboleda A, Sepulcre-Cantò G, Batelaan O, Ardö J, Gellens-Meulenberghs F. 2011. Improving evapotranspiration in land surface models by using biophysical parameters derived from MSG/SEVIRI satellite. *Hydrology and Earth System Sciences. Discussion 8*: 9113–9171.
- Giard D, Bazile E. 2000. Implementation of a new assimilation scheme for soil and surface variables in a global NWM model. *Monthly Weather Review* 128: 997–1015.
- Hamdi R, Van de Vyver H, Termonia P. 2012. New cloud and micro-physics parameterization for use in high-resolution dynamical downscaling: Application for summer extreme temperature over Belgium. *International Journal of Climatology* DOI: 10.1002/joc.2409.
- Mahfouf J-F. 1991. Analysis of soil moisture from near-surface parameters: a feasibility study. *Journal of Applied Meteorology* 30: 506–526.
- Mahfouf J-F. 2007. Soil analysis at Météo-France, Part I: Evaluation and perspectives at local scale (in French), *Note de Cent. CNRM/GMME*, 84, 58 pp., Groupe de Mtorol. Moyenne Echelle, Center National de la Recherche Meteorological Service, Météo- France, Toulouse, France.
- Mahfouf J-F, Bliznak V. 2011. Combined assimilation of screenlevel observations and radar-derived precipitation for soil moisture analysis. *Quarterly Journal of the Royal Meteorological Society* 137: 709–722.
- Mahfouf J-F, Bergaoui K, Draper C, Bouyssel F, Taillefer F, Taseva L. 2009. A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *Journal of the Geophysical Research* 114: D80105, DOI: 10.1029/2008JD011077.
- Nicolis C. 2003. Dynamics of model error: Some generic features. *Journal of the Atmospheric Sciences* 60: 2208–2218.
- Noilhan J, Mahfouf J-F. 1996. The ISBA land-surface parameterization scheme. *Global and Planetary Change* 13: 145–159.

GLAMEPS and HarmonEPS:

LAM ensemble prediction systems under development

Inger-Lise Frogner, GLAMEPS team and HarmonEPS team

Introduction

GLAMEPS is a (semi)-operational limited area ensemble system developed as a part of the cooperation between HIRLAM and ALADIN. It aims at predicting atmospheric features on spatial scales intermediate between the synoptic, covered by leading global EPS, and the convection-permitting scales. GLAMEPS_v1 has run as Member state time-critical option 2 (TCF_2) at ECMWF since 20 January 2012 and in this paper the system is described together with the latest developments. Verification compared to IFS ENS is shown and future plans are presented.

In parallel to the development of GLAMEPS a convection-permitting ensemble prediction system is also being developed, called HarmonEPS, for the very short range (<36h). The basic model tool is the non-hydrostatic Harmonie with Alaro and Arome physics. The basic set-up for the first experiments with the system is described together with some very first verification results. Future plans are also presented.

GLAMEPS_v1

Description of system

The challenge for GLAMEPS is to construct and deliver a well-calibrated, pan-European ensemble for short-range NWP by accounting for both initial state and model inaccuracies. Model uncertainties are presently taken into account by using a small number of different models and versions, two versions of the HIRLAM model (HirEPS_S and HireEPS_K) and AladEPS as well as stochastic physics in the two HIRLAM versions. Initial state uncertainties are taken into account in two ways: Ensemble perturbations are imported from the global ECMWF 51-member EPS. This system also provides perturbations at the lateral boundaries during the prediction period. Additional initial state perturbations are included by running three different assimilation cycles in parallel with different models and model versions (the two HIRLAM control members run 3d-Var, and IFS ENS EDA, Alaro is down scaling), and by perturbing the surface observations in the two HIRLAM versions. All LAM-members also run with a separate data-assimilation cycling for the ground surface, yielding a unique surface analysis per ensemble member. GLAMEPS is set up for producing a 54-member hydrostatic multi-model EPS on a pan-European integration domain for 54h forecasts with a grid mesh width of around 11 km. All ensemble members from EC EPS are used as well as the deterministic EC forecast in the following way, yielding 54 GLAMEPS members:

1. GLAMEPS members 01 - 12: HirEPS_S based on IFS ENS 01-12
2. GLAMEPS members 13 - 24: HirEPS_K based on IFS ENS 13-24
3. GLAMEPS members 25 - 36: AladEPS based on IFS ENS 25 -36
4. GLAMEPS members 37 - 50: IFS ENS 37 – 50 are added to the GLAMEPS ensembles as they are

5. GLAMEPS member 51: AladEPS control, based on IFS ENS control
6. GLAMEPS member 52: ECMWF DET.
7. GLAMEPS member 53: HirEPS_K control, based on IFS ENS control
8. GLAMEPS member 54: HirEPS_S control, based on IFS ENS control

IFS ENS was formerly known as EC EPS, or just EPS.

The integration domains for GLAMEPS_v1 are shown in figure 1. In March 2010, a test version of GLAMEPS (“version 0”) was set up to run twice daily (00 and 12 UTC). This was later updated to GLAMEPS version 1, which has replaced version 0 and is now being run daily at 06 and 18 UTC. The scheduling at ECMWF will only allow GLAMEPS to start around 8 hours after the main observation times 00 and 12 UTC, and since observations from 06 and 18 UTC (respectively) will be available at the time GLAMEPS production can start, we changed to start the forecasts from those times (06 and 18), even if the perturbations at the lateral boundaries and at the initial time will be 6 hours older, and thus have higher amplitudes.

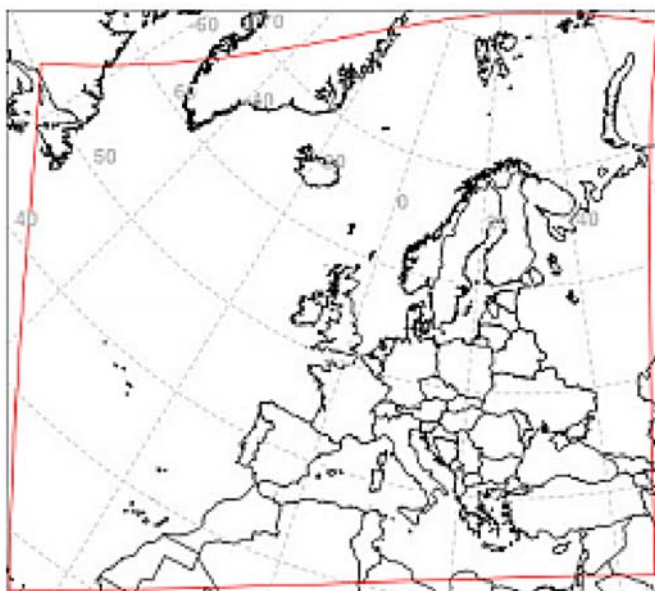


Figure 1: GLAMEPS domains. Black domain is the Aladin grid (Lambert conical), red domain is the Hirlam domain and common output domain (rotated lat-lon).

Performance

Over the last months the “Hirlam-Aladin R Package” (HARP) has come to a stage where it is now possible to use for verifying the performance of GLAMEPS_v1 (as well as other systems). The verification is done for stations and compared to IFS ENS. In figure 2 the spread-skill relationship is displayed for mean sea level pressure (MSLP) as a function of lead time for the period of December 2012 to March 2013. The spread-skill relationship is quite good for both systems. It is clear that for this parameter GLAMEPS has larger spread than IFS ENS, but also slightly larger error after 24h. This increased error is something we have not seen in earlier verification of GLAMEPS and that we need to keep an eye on. In figure 3 the continuous ranked probability skill score (CRPSS) is shown for mslp for four different lead times. Here IFS ENS is used as a reference forecast, hence it lies on the zero line in the figure, meaning that if GLAMEPS is above the zero line it is better than IFS ENS and if it is lower than the zero line it is worse. From figure 3 it is clear that for this parameter and for this verification period GLAMEPS is not able to beat IFS ENS as it is fluctuating around the zero line for

the first two lead times (upper left and upper right in figure 3), while it is more below the zero line than over for the two last lead times (lower left and lower right in figure 3). This is in line with what has been seen in earlier verification of GLAMEPS; it is hard to beat IFS ENS for mslp.

In figure 4 CRPSS for the same period and same lead times as in figure 3 is shown for two meter temperature (T2m). A clear improvement in the scores is seen in the beginning of October. On 11 October a bug in the way SST was handled was corrected, and the improvement is clear. After this date GLAMEPS is mainly well above the zero line, and hence performing better than IFS ENS. In figure 5 a flavour of other verification scores are shown for T2m. From all these scores GLAMEPS (in black) performs better than IFS ENS (in green).

Also for 10 meter wind speed (S10m) the scores for GLAMEPS are in general better than those of IFS ENS, see figure 6 and 7 (corresponding to figure 4 and 5 for T2m).

Scores for 12 h accumulated precipitation are shown in figure 8 and figure 9. As can be seen from figure 8 (left) GLAMEPS is in general performing better than IFS ENS for the first lead time (+24h), while the results are more mixed for the second lead time (right). Looking at BSS, reliability, ROC curves and Value curves, the improvement of GLAMEPS over IFS ENS is confirmed for the early lead time (figure 9). The scores for GLAMEPS are generally better than the scores for IFS ENS also for the longer lead time, see figure 10.

All in all GLAMEPS is performing well in comparison with IFS ENS for the weather parameters looked at (T2m, S10m and 12 h accumulated precipitation).

Possible updates to GLAMEPS_v2

Research and development for potentially further improved GLAMEPS are ongoing. There are several possible configuration experiments that have started or are planned starting later this year or next year, they are briefly mentioned here.

Calibration

Calibration experiments using ELR (extended logistic regression) have started in GLAMEPS. Three different approaches have been proposed for ELR:

1. Calibrate per lat/lon box of stations and interpolate regression coefficients to grid points if necessary.
2. Perform calibration over the whole model domain using lat/lon and/or elevation as predictors.
3. Regionalize over regional types: sea, land, coast, mountain.

Blending with ETKF or ETKF-Hybrid./Blending with Hirlam EDA or EDA-Hybrid

Experiments on these aspects have so far proven promising in the sense that perturbations that are directly related to present analysis uncertainties are taken into account. Preliminary experiments show improvements in ensemble spread during the first 12-24h of the forecasts. Experiments have tentatively also been made in which the ensemble is used to estimate the actual flow dependant model error covariance matrix to be used in the 3D-Var analysis for Hirlam. This is called ETKF-Hybrid. Initial experiments have only been made with 12+1 ensemble members, and blending between ETKF-perturbations and EC-EPS51 boundary data are implemented. Experiments in full GLAMEPS mode have been performed over the last year for the experiment periods, and evaluation is currently ongoing.

Blending with Hirlam CAPE singular vectors.

An experiment in full GLAMEPS mode has been run over the last year for the experiment periods before mentioned, where ensemble perturbations based on Hirlam Singular Vectors that maximize convective available potential energy over a target domain after 12h (CAPE SVs), are blended with perturbations from IFS ENS. Results are being evaluated at the moment, and first results indicate an improved skill for large precipitation events and neutral scores for other parameters.

Increase the number of Aladin ensemble members at the expense of the IFS ENS members.

It is planned to increase the number of Aladin members to the level of HIRLAM members, this experiment will start this autumn.

GLAMEPS 4 times per day (lagged ensembles). Increased resolution (~8 km)

There has been several requests from users of GLAMEPS to have it available four times a day. An experiment is presently being set up to investigate the skill of a system that runs four times a day, but with half the number of members and to introduce lagging to keep the number of members at the same level as today. At the same time increased resolution to 8 km will be tested.

HarmonEPS

HarmonEPS is the name of the ensemble prediction systems for the very short range (<36h) on so-called convection-permitting scales that is currently being developed. The basic model tool is the non-hydrostatic Harmonie with Alaro and Arome physics. The development towards a cloud-permitting, meso-scale model system (Harmonie) has had considerable progress, and a prototype system for HarmonEPS is developed. Experiments are done on a few selected sub-European domains, thus gaining experience that can eventually lead to a prototype system that member countries can install at their home computers for their own purpose. Also the area of Sochi at the Black sea is such a test area since the Hirlam consortium is engaged in the FROST project (Winter Olympic games in Sochi 2014) and HarmonEPS is one component of this engagement. Some experience has been gained in a few HIRLAM and ALADIN member countries, as well as in other European consortia, but a lot more is necessary, e.g. in order to develop close links to meso-scale data assimilation, physics parameterizations, and the description of the land surface.

During the last year the general system for running HarmonEPS in experimental mode has been finished, and a set of first experiments have been run. First test are run with a horizontal resolution of 2.5 km, +36 h lead time, full data assimilation and 6 hourly cycling for the control members, HarmonEPS run every 12 h, surface assimilation included for every member and with 22 members (10 + 1 with Arome physics and 10 + 1 with Alaro physics). In view of the cold-start surface nesting problems it was decided to run HarmonEPS directly with surface DA for each member. It is believed that including surface assimilation in the HarmonEPS runs will be crucial, especially when nesting in the very much coarser operational EPS. Running separately surface assimilations for each member around the same observations is likely to reduce the spread, so ideally one should also include observation perturbations. Even if this was not technically possible for the first experiments it is believed that including surface assimilation will be beneficial. Also, it was decided to cycle the control for a couple of weeks with full DA until the start of the experiment period, and continue this full cycling for the control forecast of HarmonEPS through the experiment period. This HarmonEPS control analysis is then used as a basis for all the members. The members are then perturbed with the perturbations from the host model IFS ENS like this: $pert = member - control$, and add that to the HarmonEPS control.

ECMWF has kindly produced and made available test sets of IFS ENS to be used as boundary conditions for LAM EPS, both at the present operational resolution of about 32 km, and at 16 km. The first HarmonEPS experiments have been performed in order to assess the impact of the increased horizontal resolution of the boundary files, and to assess if it is possible to nest directly in the present IFS ENS. The evaluation of these experiments is ongoing. Preliminary findings show a small increase

in scores of HarmonEPS when nesting in high resolution IFS ENS as compared to when nesting in operational resolution IFS ENS, but it is too early to conclude if this minor increase in scores is worth the extra cost of running it. More evaluation will be performed before any conclusion can be drawn, including looking at high impact events like heavy precipitation. Figure 11 shows an example of verification (area under ROC curve) where HarmonEPS nested in high resolution EPS is compared to HarmonEPS nested in operational resolution EPS .

For models uncertainties we have so far started on the following three approaches:

- 1) Multi physics approach. The current set-up of HarmonEPS consists of 10+1 members with Alaro physics and 10+1 members of Arome physics, thus continuing the multi-physics approach that has been seen to be so beneficial in GLAMEPS;
- 2) Include SPPT in HarmonEPS. The first experiments have finished and are currently being evaluated, with some preliminary findings indicating that the perturbations used in this first trial might have been too large;
- 3) Include cellular automata (CA) in Alaro part of the ensemble. A summer period has been run with CA included and is currently being evaluated as well.

HarmonEPS experimentation has just started, and will likely increase in the time to come. There are many challenges when designing an ensemble prediction system on the convection-permitting scales, and a close cooperation with experts on data assimilation, physical parameterizations, land surface and dynamics will be crucial. In addition a close cooperation with system experts and experts on diagnostics and verification is also important. More emphasis should also be given to the work on developing products from the system and to promote the use both to duty forecasters and to advanced users, as well as the general public.

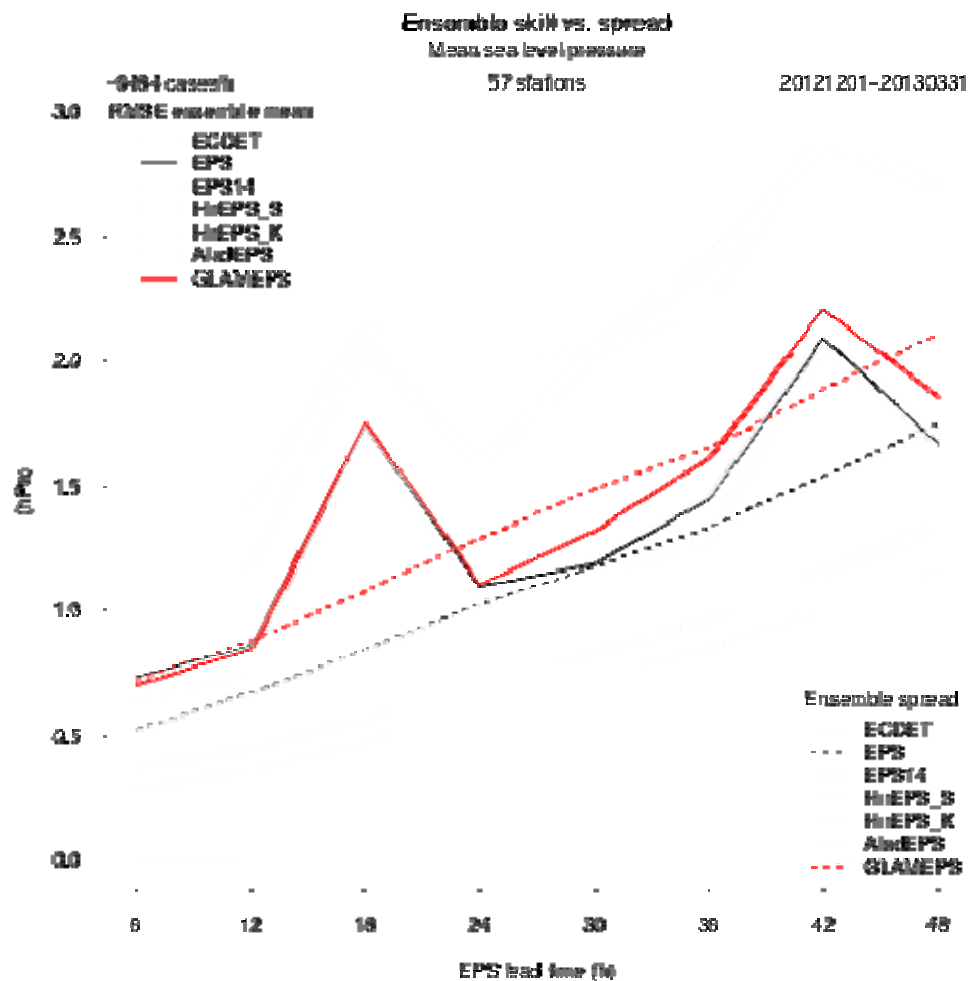


Figure 2: Spread-skill relationship for GLAMEPS (red) and IFS ENS (black) as a function of lead time for the period December 2012 to March 2013. Solid lines are RMSE of ensemble mean dashed lines are ensemble spread.

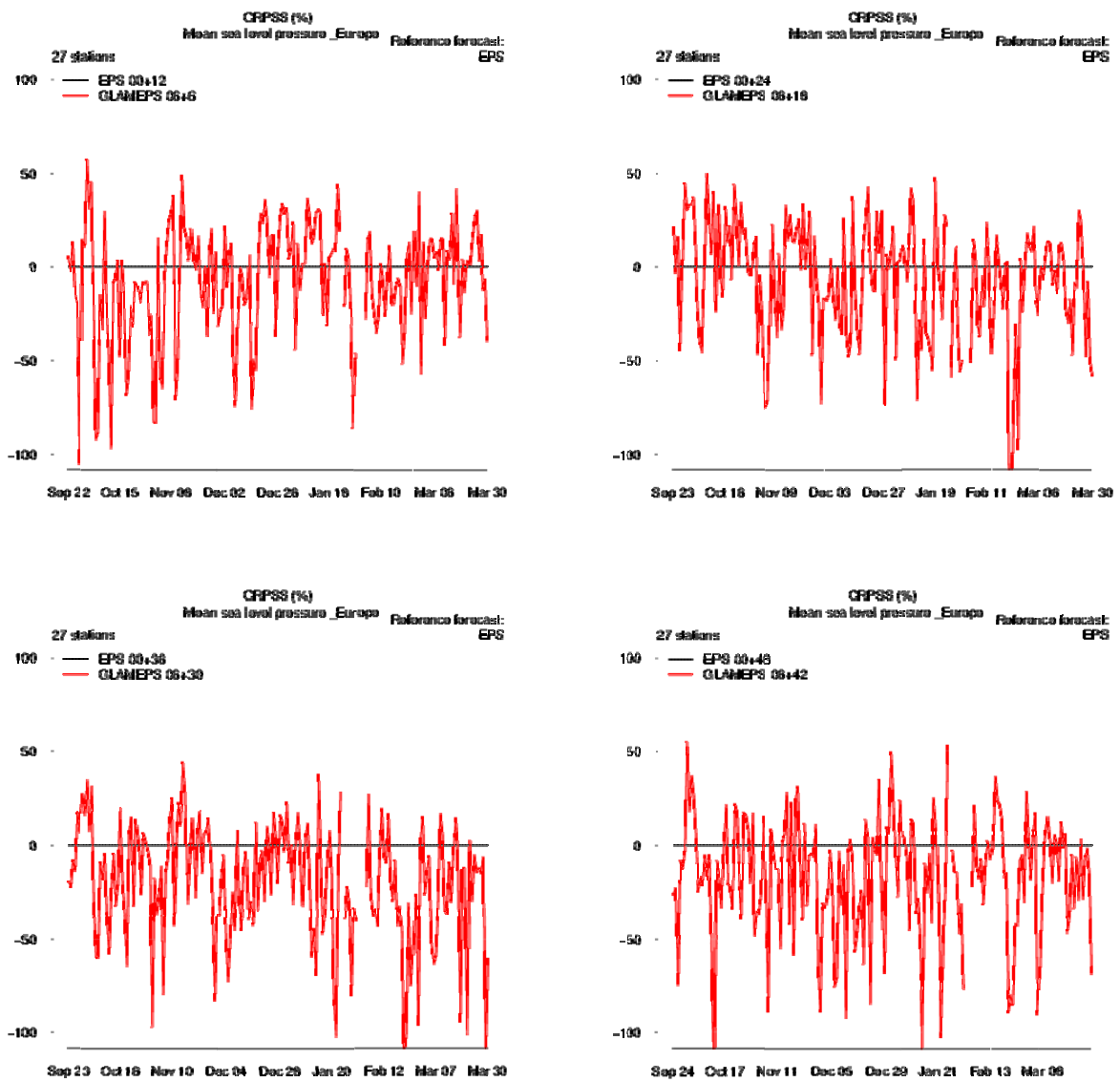


Figure 3: CRPSS for mslp for four different lead times; upper left +12h, upper right +24h, lower left +36h, lower right +48h. IFS ENS is used as reference, hence it lies on the zero line. GLAMEPS is in red. For the period from 22 September 2012 to 31 March 2013.

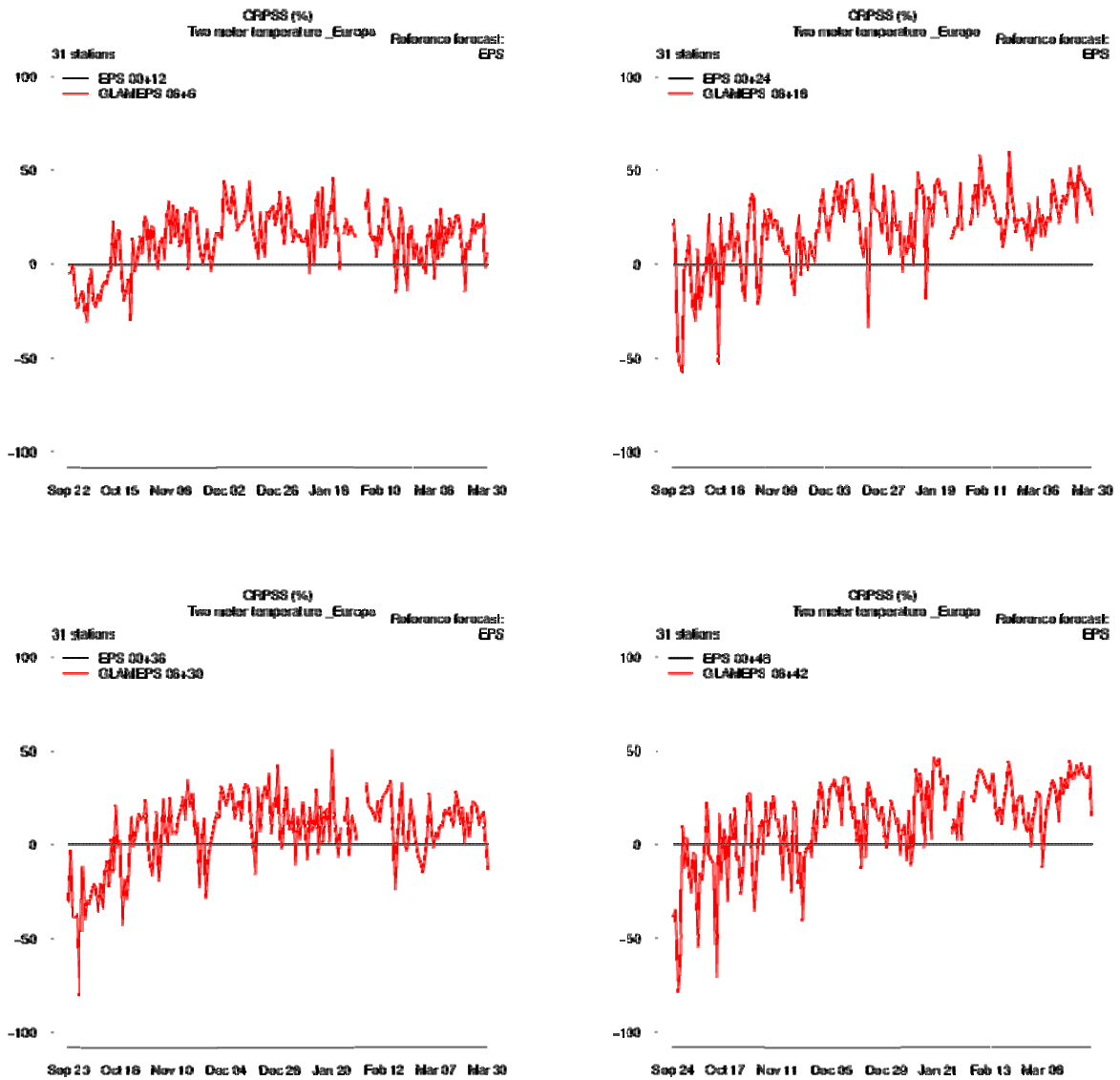


Figure 4: CRPSS for T2m for four different lead times; upper left +12h, upper right +24h, lower left +36h, lower right +48h. IFS ENS is used as reference, hence it lies on the zero line. GLAMEPS is in red. For the period from 22 September 2012 to 31 March 2013.

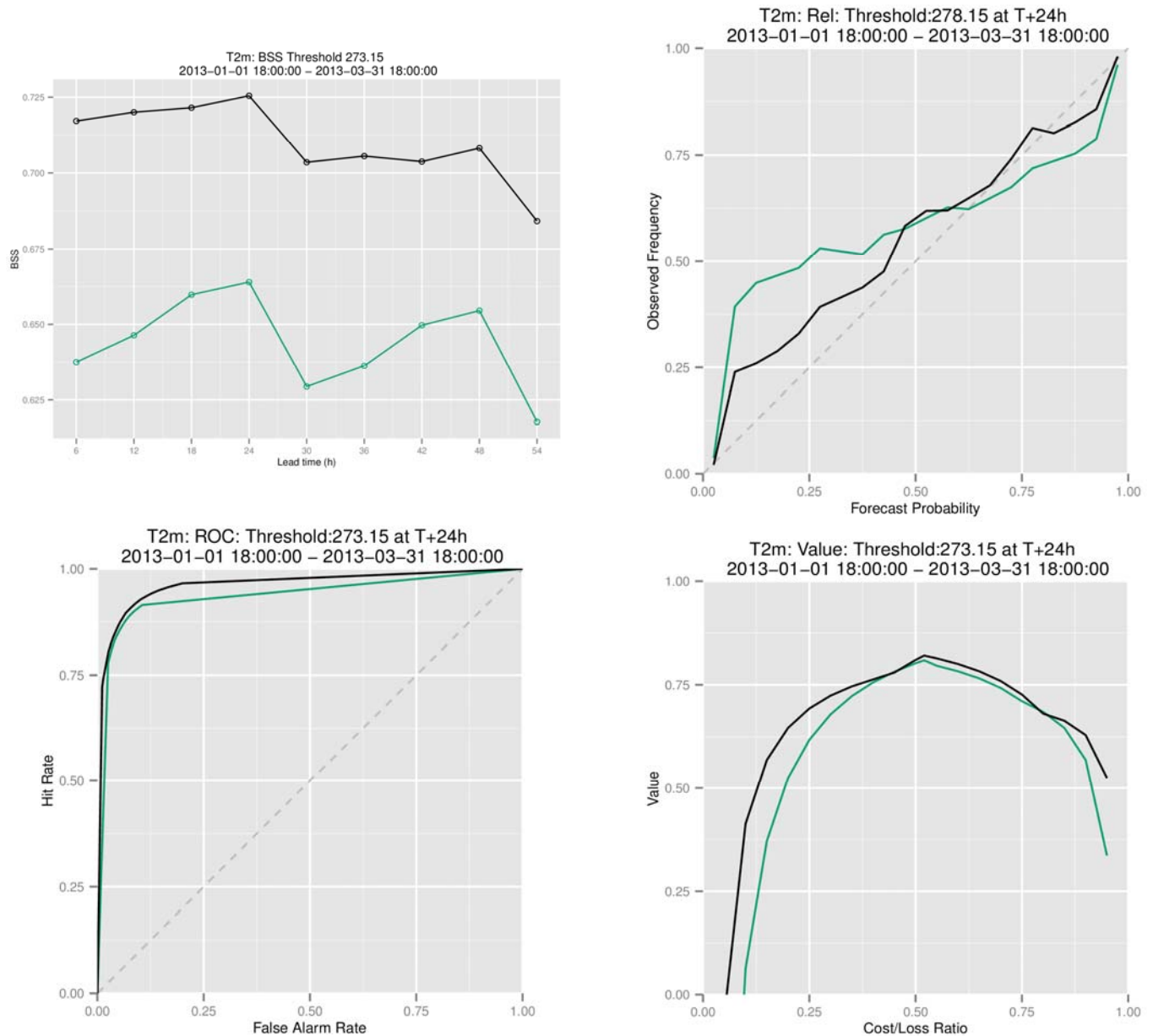


Figure 5: Scores for T2m, GLAMEPS is in black, IFS ENS in green. Upper left: Brier Skill Score (BSS) for a threshold of 0 degrees, upper right: reliability diagram for a threshold of 5 degrees and 24 hour lead time, lower left ROC curve for a threshold of 0 degrees and a lead time of 24 hours, lower right: value curve for a threshold of 0 degrees and a lead time of 24 hours.

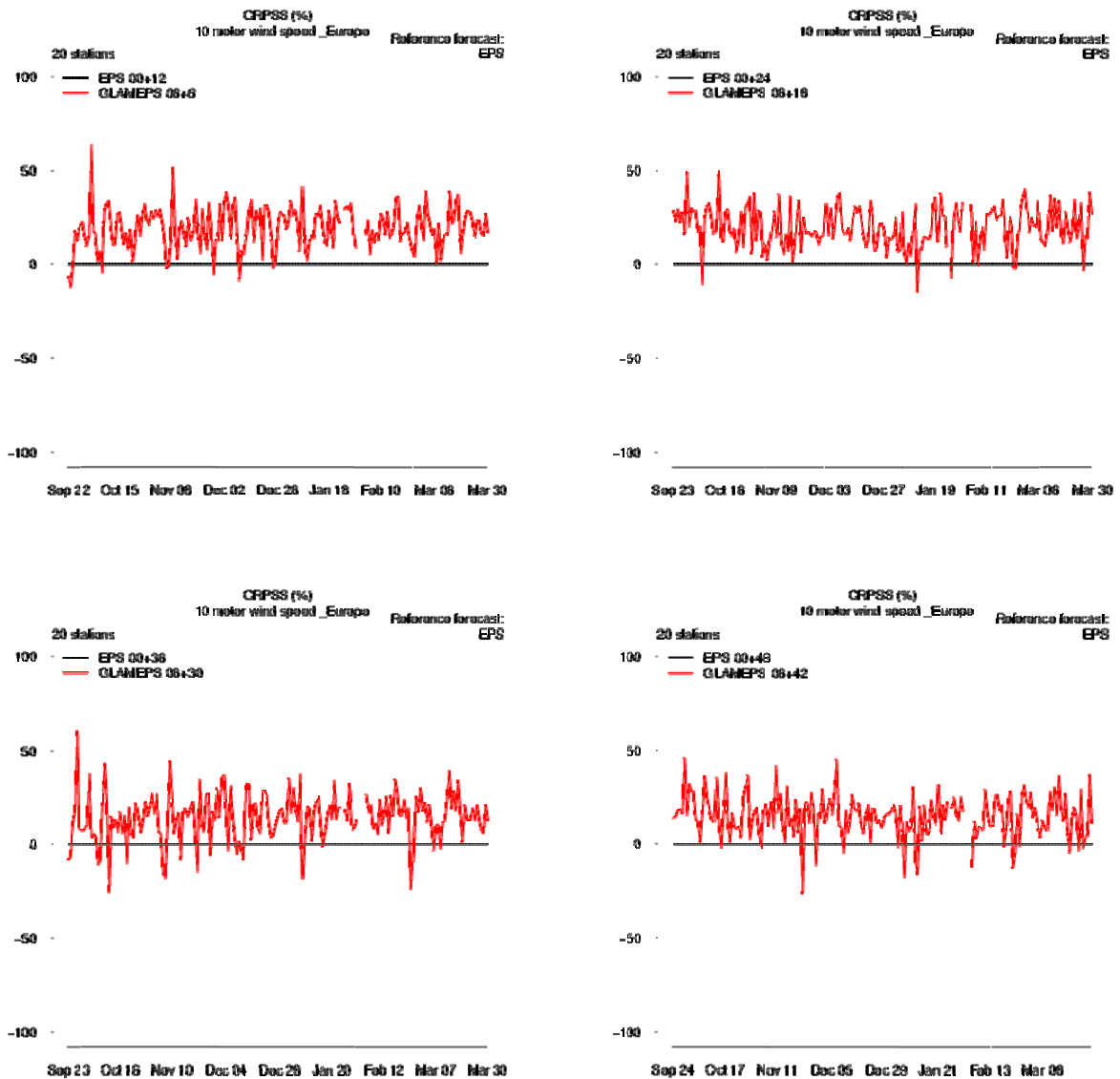


Figure 6: CRPSS for S10m for four different lead times; upper left +12h, upper right +24h, lower left +36h, lower right +48h. IFS ENS is used as reference, hence it lies on the zero line. GLAMEPS is in red. For the period from 22 September 2012 to 31 March 2013.

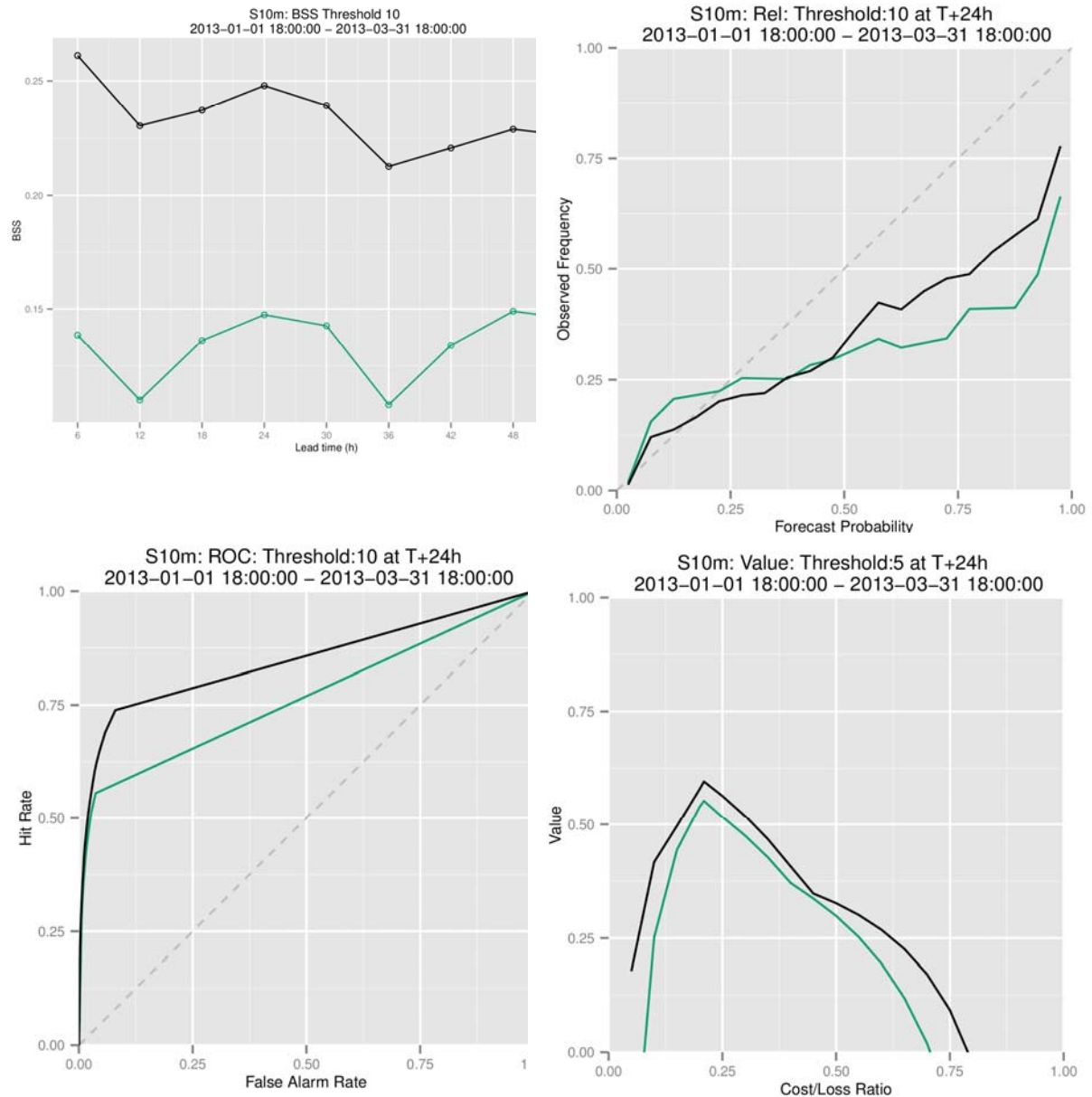


Figure 7: Scores for S10m, GLAMEPS is in black, IFS ENS in green. Upper left: Brier Skill Score (BSS) for a threshold of 10 m/s, upper right: reliability diagram for a threshold of 10 m/s and 24 hour lead time, lower left ROC curve for a threshold of 10 m/s and a lead time of 24 hours, lower right: value curve for a threshold of 5 m/s and a lead time of 24 hours.

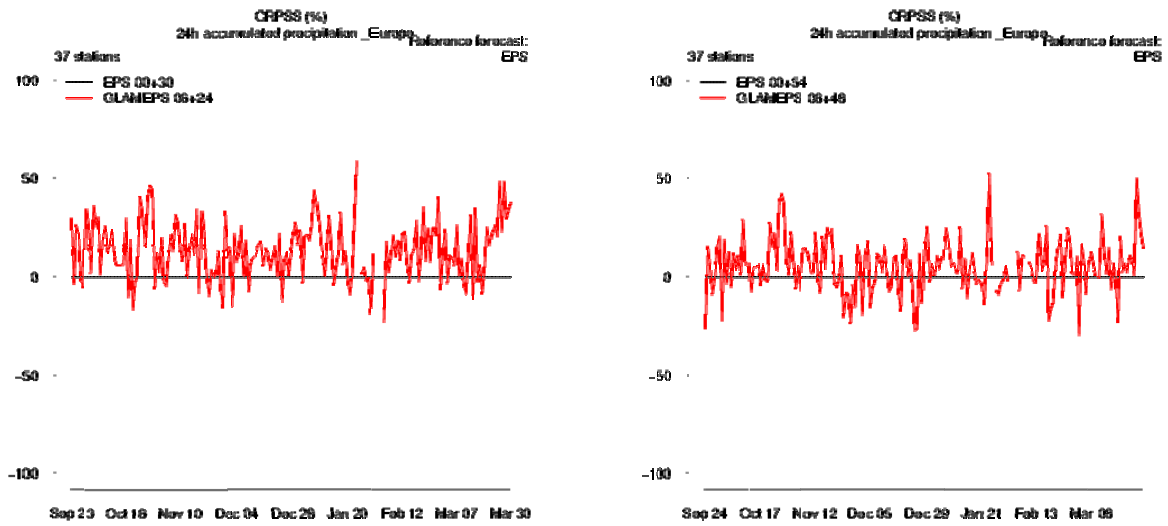


Figure 8: CRPSS for 12h accumulated precipitation for two different lead times; left +24h, right +48h. IFS ENS is used as reference, hence it lies on the zero line. GLAMEPS is in red. For the period from 22 September 2012 to 31 March 2013.

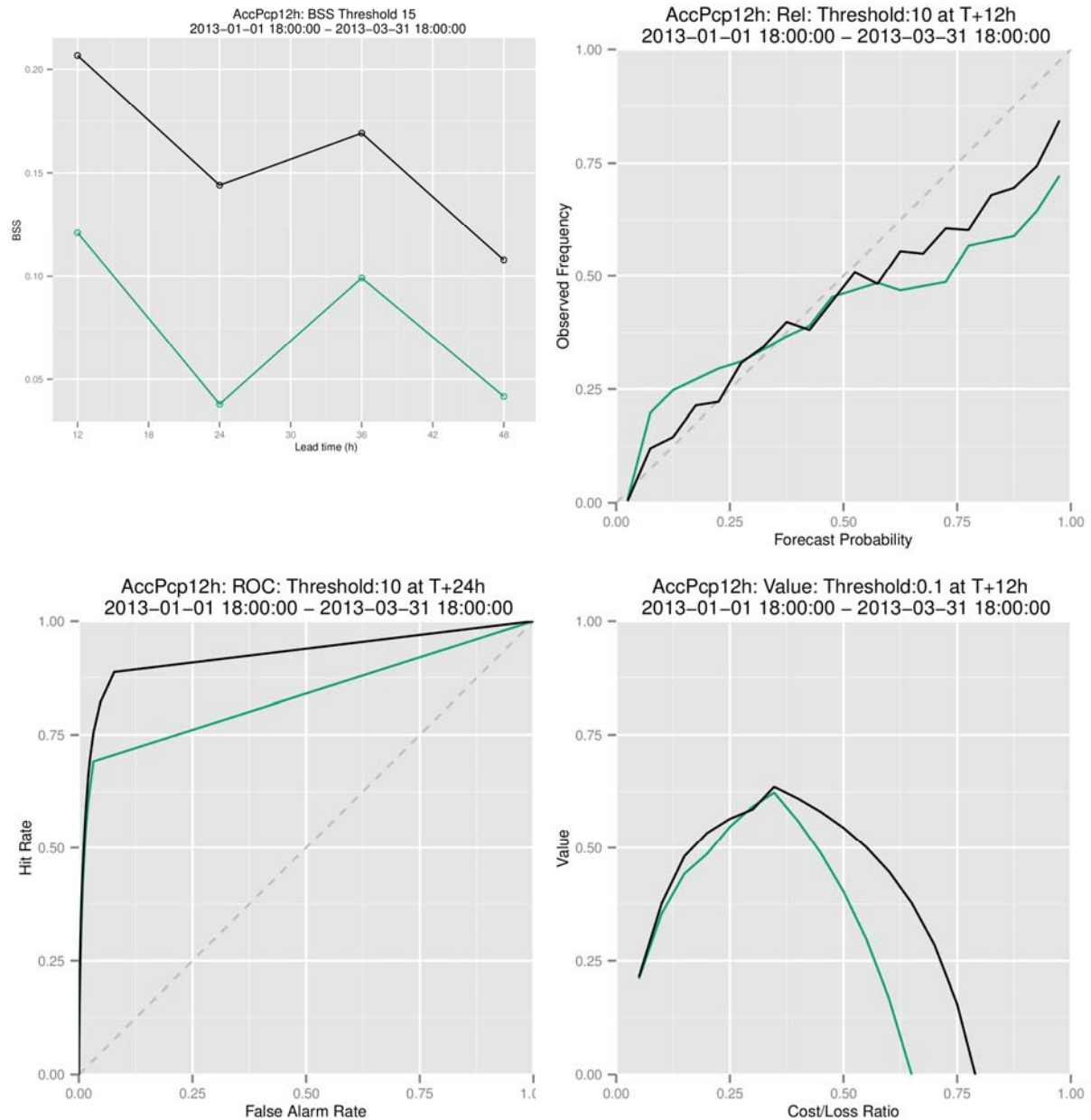


Figure 9: Scores for 12h accumulated precipitation, GLAMEPS is in black, IFS ENS in green. Upper left: Brier Skill Score (BSS) for a threshold of 15 mm/12h, upper right: reliability diagram for a threshold of 10 mm/12h and 12 hour lead time, lower left ROC curve for a threshold of 10 mm/12h and a lead time of 24 hours, lower right: value curve for a threshold of 0.1 mm/12h and a lead time of 12 hours.

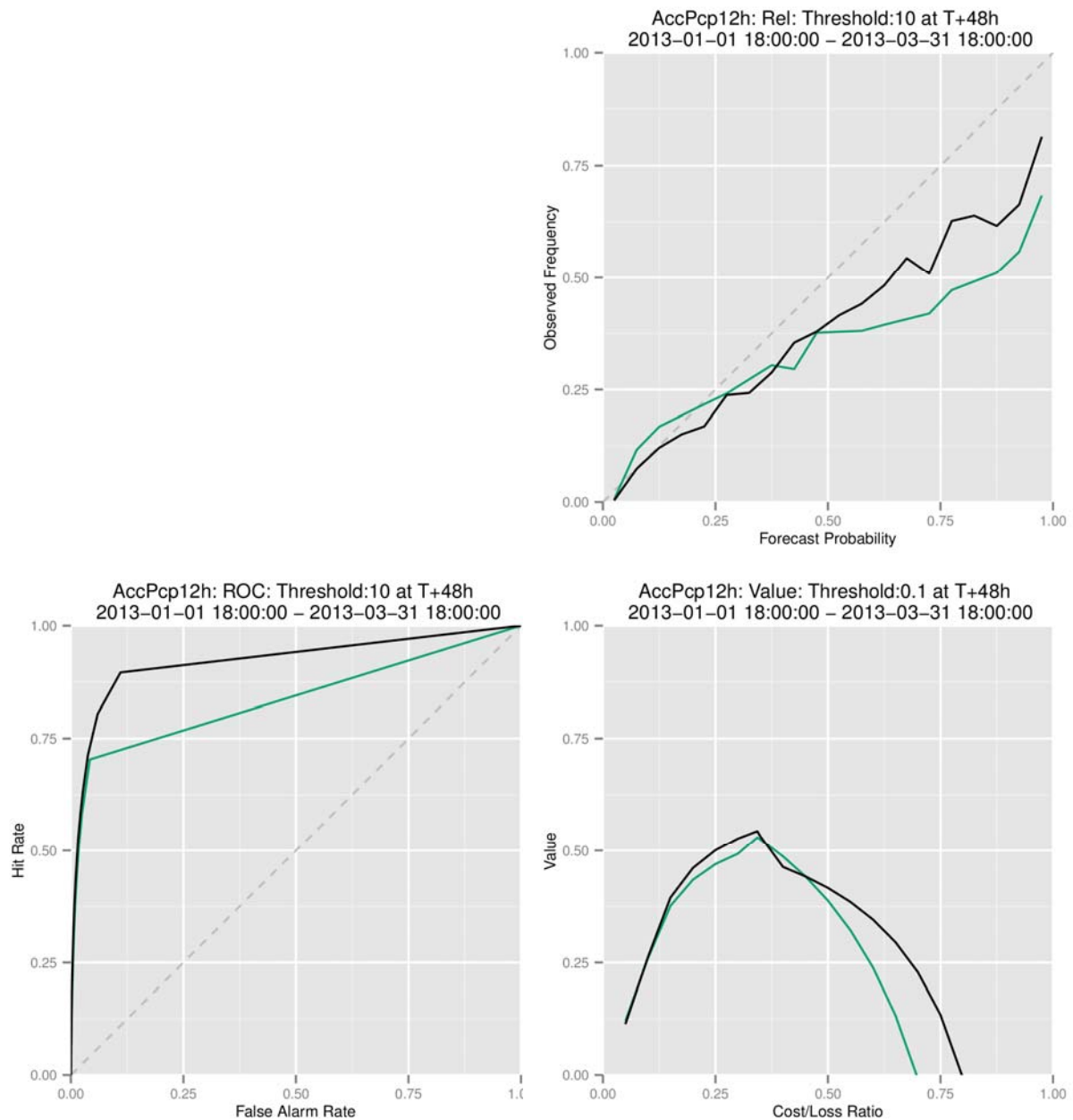


Figure 10: Scores for 12h accumulated precipitation, GLAMEPS is in black, IFS ENS in green. Upper right: reliability diagram for a threshold of 10 mm/12h and 48 hour lead time, lower left ROC curve for a threshold of 10 mm/12h and a lead time of 48 hours, lower right: value curve for a threshold of 0.1 mm/12h and a lead time of 48 hours.

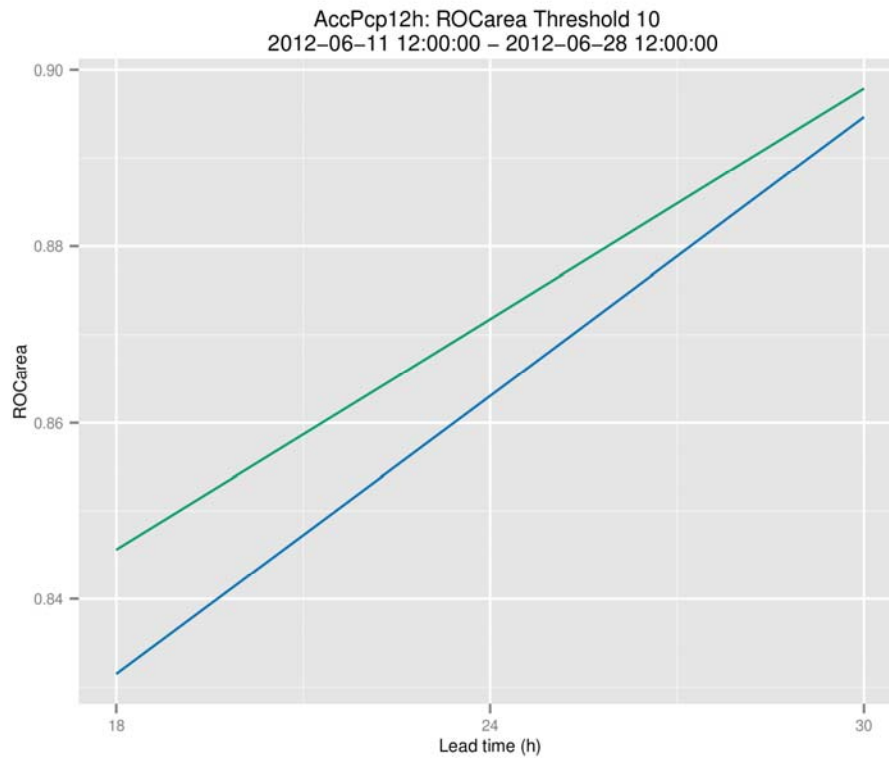


Figure 11: Area under ROC curve for 12 h accumulated precipitation for a threshold of 10 mm/12h, for 18 days in June 2012. Green is HarmonEPS nested in high resolution IFS ENS, blue the same but nested in operational resolution IFS ENS.

Finite elements used in the vertical discretization of the fully compressible forecast model ALADIN-NH

Jozef Vivoda, jozef.vivoda@shmu.sk

Petra Smolíková, petra.smolikova@chmi.cz

1 Introduction

While approaching higher horizontal and vertical resolutions and especially when reaching the limit where both the horizontal and vertical scales of motion become of comparable dimension, typically with grid size around 2 km, the non-hydrostatic effects start to play a significant role. Aiming to those scales, non-hydrostatic dynamics were introduced in the originally hydrostatic limited area NWP model ALADIN. The arisen dynamical kernel of ALADIN-NH was designed following the rule that it keeps as much features of its hydrostatic version as possible. The basic choices made for both models thus are: the spectral technique used for the horizontal spatial discretization, semi-implicit time stepping and semi-Lagrangian advection. (See [6] for details.)

For vertical discretization, the finite-difference (FD) scheme of Simmons and Burridge ([15]) was applied since the beginning of the project, innovated with the semi-Lagrangian vertical advection according to [17]. This scheme is only first order accurate for non-uniform spacing of vertical levels. An alternative finite-element (FE) vertical discretization was adapted for the hydrostatic version of ALADIN (as first implemented in the ECMWF global forecast model IFS by Untch and Hortal, see [19] for the method description). With the hydrostatic approximation and semi-Lagrangian advection the only vertical operations evaluated are vertical integrals. An integral operator was derived based on the Galerkin method using cubic B-splines with compact support as basis functions. Piecewise linear (hat) basis functions have been implemented as an alternative option. It was shown that the FE scheme gives more accurate phase speeds of most of the linear gravity waves than the FD scheme. In addition, the cubic FE scheme has proven to be eight order accurate for integrating smooth functions, this to be compared to only first order accuracy of FD method. The computational cost of such an improvement in accuracy is negligible. Furthermore, the FE scheme reduces the level of vertical noise in forecasts with the full hydrostatic model of ECMWF, reduces the cold bias in the lower atmosphere and improves the transport in the stratosphere. Consecutively, the finite-element vertical discretization has been tested in several hydrostatic applications of the LAM model ALADIN with detected positive impact on the objective verification scores. On top of that, the FE method has proven beneficial in 2D vertical plane idealized tests of a resting, hydrostatically balanced state, see [12].

Not surprisingly, the need emerged to extend the FE method to the vertical discretization of the fully compressible limited area model ALADIN-NH. This task has shown to be intricate and troublesome. Contrary to the hydrostatic equations where only integral vertical operators appear, and to the fully compressible system of equations cast in the height based coordinate where only derivative vertical operators occur (see [14]), in the fully compressible Euler equations of ALADIN-NH designed for the mass based vertical coordinate of Laprise (see [10]), both the vertical integral operator, and the derivative vertical operators appear. Moreover, a set of constraints fulfilled by continuous vertical operators was found to be crucial for stability and those relations were carefully introduced in the FD formulation of vertical discretization (see [6]). The fact that these constraints are not satisfied in FE approach has consequences in the design of the implicit part of the time scheme and it was necessary to check its influence on the stability of the whole model. In particular, it is not possible

to eliminate all the variables but one and get a single structure Helmholtz equation in the implicit part of the time scheme. We abide with the system of linear equations for two variables which we solve with a stationary iterative method with predictor being solved as if the given constraint was satisfied. We show that our solution converges in all tested idealized and real cases and one iteration of the implicit solver is enough to reach satisfying results. The computational price to be paid remains in percents units and does not penalize the whole integration significantly, see section 3.5 for details.

When implementing the FE method in the non-hydrostatic model ALADIN-NH, we aimed at the comparable stability properties as for the FD method in both idealized experiments and real simulations. Moreover, similarly as in the hydrostatic case we expect enhanced accuracy. To reach the theoretically possible degree of accuracy, we need to replace the FD discretization wherever it appears in the whole system with the FE version. This goal is not fulfilled up to now since the transformations between the vertical divergence variable \mathbf{d} and the vertical velocity w used in the explicit calculations are still finite difference ones (see section 3.4 for explanation). The accuracy reached in real forecasts may thus be limited by this fact. For each application of a vertical operator we need to fulfill distinct top and bottom boundary conditions depending on the properties of the term it is applied on. Hence, we are forced to define several distinct vertical operators for each operation varying in dependence on the boundary conditions to be satisfied.

The paper is organized as follows. In section 2, model variables and the set of equations used are presented. In section 3, the vertical discretization based on finite elements is described together with the semi-implicit time scheme and the consequences of the usage of the FE method in the vertical on its design are given. Results of idealized test cases are summarized in section 4 and the comparison with reference cases using the finite difference method in vertical is shown. Real case experiments are presented in section 5 and section 6 summarizes the results achieved, outlines future directions and concludes the paper.

2 Model variables and equations

For the sake of simplicity we restrict ourselves to dry shallow atmosphere and omit here the Coriolis, frictional and diabatic forcing. On top of the prognostic variables used in the hydrostatic version of the model ALADIN which are the horizontal velocity vector \mathbf{V} , temperature T and logarithm of hydrostatic surface pressure $q_s = \ln(\pi_s)$, there are two additional non-hydrostatic variables in ALADIN-NH, i.e. pressure departure $\hat{q} = \ln(\frac{p}{\pi})$ and the variable \mathbf{d} representing vertical motion. Lets denote the reduced vertical divergence d and the orographic component X as

$$d = -g \frac{p}{mRT} \left(\frac{\partial w}{\partial \eta} \right), \quad (1)$$

$$X = \frac{p}{mRT} \nabla \phi \cdot \left(\frac{\partial \mathbf{V}}{\partial \eta} \right), \quad (2)$$

where true pressure p is calculated with the use of the hydrostatic pressure $\pi = A(\eta) + B(\eta) \pi_s$ as $p = \pi \exp(\hat{q})$, with $A(\eta)$, $B(\eta)$ being sets of constants. (The details may be found in [5] and [19].) Further, ∇ is the gradient along the constant η -surfaces, R is the perfect gas constant for dry air, g the acceleration of gravity, $m = \frac{\partial \pi}{\partial \eta}$ is the vertical metric factor, ϕ geopotential and w vertical velocity. (Notice that d is defined with the perfect gas constant for dry air even in the moist atmosphere. Detailed consequences go beyond the scope of this paper.) Then $\mathbf{d} = d + X$ is the prognostic variable used in spectral space (and implicit calculations) of ALADIN-NH while in the explicit part the vertical velocity w remains (under the option LGWADV=T). We describe first the fully compressible Euler equations of ALADIN-NH in the hybrid mass-based vertical

coordinate η for the set of prognostic variables ($\mathbf{V}, w, T, \hat{q}, q_s$):

$$\frac{d\mathbf{V}}{dt} = -\frac{RT}{p}\nabla p - \left(\frac{p}{\pi} + \frac{p}{m}\frac{\partial\hat{q}}{\partial\eta}\right)\nabla\phi, \quad (3)$$

$$\frac{dw}{dt} = \frac{g}{m}\frac{\partial(p-\pi)}{\partial\eta}, \quad (4)$$

$$\frac{dT}{dt} = -\frac{RT}{C_v}D_3, \quad (5)$$

$$\frac{d\hat{q}}{dt} = -\frac{C_p}{C_v}D_3 - \frac{\omega}{\pi}, \quad (6)$$

$$\frac{\partial q_s}{\partial t} = -\frac{1}{\pi_s}\int_0^1 \nabla \cdot m\mathbf{V}d\eta, \quad (7)$$

where C_p, C_v are the specific heat capacities of dry air at constant pressure and volume respectively, $\omega = \frac{d\pi}{dt}$ indicates the mass-based vertical velocity diagnosed from

$$\omega = \mathbf{V} \cdot \nabla\pi - \int_0^\eta \nabla \cdot m\mathbf{V}d\eta, \quad (8)$$

and $D_3 = \nabla \cdot \mathbf{V} + \mathbf{d}$ stands for the local tridimensional divergence. The geopotential ϕ is obtained through an upward integration of atmospheric depths from the surface geopotential ϕ_s using

$$\phi = \phi_s + \int_\eta^1 \frac{mRT}{p}d\eta. \quad (9)$$

The total derivative operator on the left hand sides of given equations is defined as

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla + \frac{d\eta}{dt} \frac{\partial}{\partial \eta}. \quad (10)$$

Thus, to complete the system we need the following diagnostic relations

$$m \frac{d\eta}{dt} = B(\eta) \int_0^1 \nabla \cdot m\mathbf{V}d\eta - \int_0^\eta \nabla \cdot m\mathbf{V}d\eta, \quad (11)$$

$$w_s = \mathbf{V}_s \cdot \nabla\phi_s. \quad (12)$$

All horizontal derivatives are evaluated spectrally. See [5] for more details.

3 Finite element scheme

To discretize vertical operators we apply the finite element procedure as described for example in [11]. We use B-spline functions as the basis functions for the FE method in order to keep consistency with the hydrostatic version of the ALADIN dynamical core.

3.1 Vertical discretization

In order to solve numerically the equations from the previous section, the model domain is divided into L layers and the full model levels are located inside these layers while the half model levels are located at material boundaries and layer's interfaces. The model variables are staggered in the vertical direction (w being defined on the half levels, while all the other variables are on the full levels). Boundary conditions are defined for each prognostic quantity on material boundaries on the top and on the bottom of the model domain.

3.2 Interpolation with B-spline curve

Values of a prognostic quantity f on the model full levels $\eta_j, j = 1, \dots, L$, can be represented as a vector of data points $\langle \eta, f \rangle = ((\eta_i, f(\eta_i)), i = 1, \dots, L)$, supplemented with two boundary conditions defining value $f(\eta_0 = \text{top})$ and $f(\eta_{L+1} = \text{bottom})$. We would like to interpolate these data points with the parametric B-spline curve S as

$$S(\eta, f(\eta)) = \sum_{i=0}^{L+1} (\hat{\eta}_i, \hat{f}_i) \cdot \mathbf{a}_i(\eta). \quad (13)$$

The vector $\langle \hat{\eta}, \hat{f} \rangle = ((\hat{\eta}_i, \hat{f}_i), i = 0, \dots, L+1)$ represents the control points of the interpolating B-spline curve S of order C . (Notice that the spline order is not the degree of piecewise polynomials that form the B-spline. For example the spline order of a cubic B-spline is $C = 4$). The choice of B-spline basis functions $\mathbf{a}_i(\eta)$ is not unique and it affects the properties of the resulting interpolating B-spline curve S . In order to determine B-splines, we have to define η points used in the B-spline construction. These points are referred to as *the knots*, and they differ from the values of η at the model levels. Moreover, B-spline basis functions are fully determined by knots since knowing knots the B-splines can be constructed uniquely with the use of de Boor's algorithm. Description of the algorithm is out of scope of this paper and can be found in [8]. Once the distribution of the model full levels is given, the choice of knots is not arbitrary, because B-spline basis functions must be distributed in such a way that there is at least one model full level in the support of each B-spline function.

In the theory of curve interpolation the choice of knots is dependent on the values of input function f . More information can be found in the computer graphics theory literature, see for example [9]. Such solution is impossible to adopt in our case. This would lead to a very inefficient FE method implementation as B-spline basis functions would become temporally and horizontally dependent and consequently FE operators would inherit this property. Therefore, we construct our operators with respect to the shape of SI background pressure π^* which is time and horizontal space independent. Its value on half levels is determined from known values of A and B as $\pi_i^* = A_i + B_i \pi_s^*$.

The knots k_i are calculated in the following procedure.

1. Set $\eta_{L+1}^{\sim} = \eta_0 = 0$ and define the remaining half level values of η parameter with *the centripetal method* (where $\alpha = 0.5$) as

$$\eta_i^{\sim} = \frac{\sum_{l=1}^i (\pi_l^* - \pi_{l-1}^*)^\alpha}{\sum_{i=1}^L (\pi_i^* - \pi_{i-1}^*)^\alpha} \quad \text{for } i = 1, \dots, L. \quad (14)$$

2. Compute estimate of η values on full levels as simple average of surrounding half level values

$$\eta_0 = 0, \eta_i = \frac{\eta_i^{\sim} + \eta_{i-1}^{\sim}}{2} \quad \text{for } i = 1, \dots, L, \eta_{L+1} = 1. \quad (15)$$

3. Define the knots k_i . The number of knots is $L+2+C$ with multiplicity order C on the model boundaries. This yields

$$k_i = \begin{cases} 0 & \text{for } 1 \leq i \leq C \\ \eta_{i+1-C} & \text{for } C < i \leq L+2 \\ 1 & \text{for } L+2 < i \leq L+2+C. \end{cases} \quad (16)$$

An example of cubic B-spline basis functions for 7 vertical full levels is given in Fig. 1.

In order to be able to express the spline curve (13), we need to determine the vector of control points $\langle \hat{\eta}, \hat{f} \rangle$. We set

$$(\eta_k, f(\eta_k)) = \sum_{i=0}^{L+1} (\hat{\eta}_i, \hat{f}_i) \cdot \mathbf{a}_i(\eta_k) \quad \text{for } k = 0, \dots, L+1, \quad (17)$$

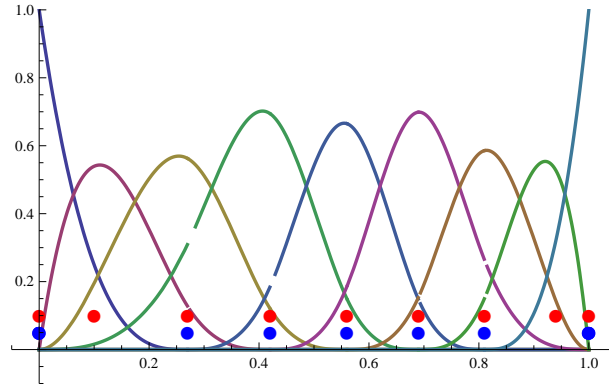


Figure 1: Cubic B-spline basis functions constructed with de Boor's algorithm for 7 full levels and 2 material boundaries. All together there are 9 functions. The red dots (upper row) represents the values of η on full levels. The blue dots (bottom row) are the knots. The knots are multiple on model boundaries and the vector of knots is given by $(0, 0, 0, 0, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, 1, 1, 1, 1)$.

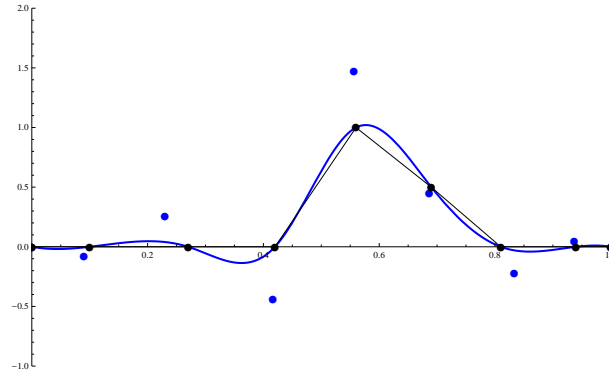


Figure 2: Cubic B-spline curve computed from 7 full levels η values and 2 boundary conditions. The black dots are the data points $\langle \eta_i, f_i \rangle$ where η_i are calculated with the centripetal method and the blue dots are the control points $\langle \hat{\eta}_i, \hat{f}_i \rangle$. In blue is the B-spline curve and in black the piecewise linear interpolation of the data points.

where $k = 0, L + 1$ represent two a priori known boundary conditions. We define the projection matrix \mathbf{A} as the matrix of rank $L + 2$ with elements $\mathbf{A}_{ki} = \mathbf{a}_i(\eta_k)$, $i, k = 0, \dots, L + 1$. Then the unknown control points may be determined as

$$\langle \hat{\eta}, \hat{f} \rangle = \mathbf{A}^{-1} \langle \eta, f \rangle \quad (18)$$

providing \mathbf{A} is regular. Our choice of basis $\mathbf{a}_i(\eta)$ ensures this property and \mathbf{A}^{-1} represents the projection matrix from the physical space to the FE space. There is an example of an interpolated B-spline curve in Fig. 2.

Let us demonstrate the importance of η parameter definition. Keeping the B-spline basis functions unchanged, we can redefine the η values as the model full levels. We adopt here the method used in hydrostatic version of the model ALADIN (LREGETA=.FALSE.) and set

$$\eta_i = \frac{A_i}{p_{ref}} + B_i \quad \text{and} \quad \eta_i = \frac{\eta_i + \eta_{i-1}}{2}, \quad (19)$$

for $i = 1, \dots, L$ and the reference pressure $p_{ref} = 1013.25 \text{ hPa}$ instead of (14), (15). The B-spline curve obtained with such definition of η parameter is depicted on Fig. 3. One can see that the interpolating curve is more oscillatory in this case and that the distance of control points is larger than with the centripetal definition (Fig. 2). There are other possible definitions of parameter η . We have tested several of them and concluded that the centripetal definition (14) provides the most reasonable results.

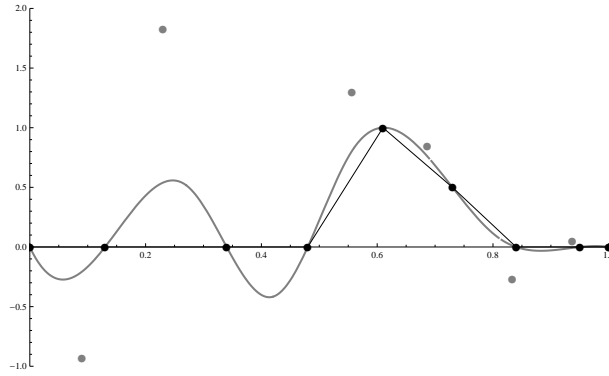


Figure 3: The same as on Fig. 2, but with the definition (19) of full η levels.

3.3 Vertical operators definition

We can apply traditional FE procedure on vertical operator Ψ given by

$$\Psi(f(\eta)) = g(\eta). \quad (20)$$

The operator input is a function f sampled on the model full levels (the model half levels respectively), supplemented with boundary conditions on the material boundaries. We denote it $\langle f \rangle$ and omit the η levels specification here for the sake of simplicity. Function g is unknown and we would like to determine it with FE procedure. Once we have continuous g , we can evaluate it anywhere inside our domain. It is appropriate to evaluate it on the same levels (η values) as f is sampled, because in this case the average error of FE operator is of very high order, due to superconvergence properties of the FE method. A FE operator with cubic B-spline basis that computes a full level quantity from the full level values is therefore eighth order accurate, while a FE operator that computes half levels quantities from the full level ones is fourth order accurate only since one can not assume superconvergence in this case.

We discretize (20) with

$$\sum_{i=0}^{L+1} \hat{f}_i \cdot \Psi(\mathbf{a}_i(\eta)) = \sum_{i=0}^{L+1} \hat{g}_i \cdot \mathbf{b}_i(\eta), \quad (21)$$

where B-spline basis functions $\mathbf{a}_i(\eta)$ and $\mathbf{b}_i(\eta)$ are generally not the same. The boundary conditions of g must be controlled by implicit properties of $\mathbf{b}_i(\eta)$ basis functions. For example, if Ψ is an integral operator from the model top, it is natural that its value on the model top is $\Psi(f(\eta_{L+1})) = 0$. To ensure this property, we have to use basis functions with zero value at the model top. There is no such limitation imposed on basis functions $\mathbf{a}_i(\eta)$, because boundary conditions are prescribed a priori in the vector $\langle \hat{f} \rangle = (\hat{f}_i, i = 0, \dots, L+1)$.

The mean weighted residual approach is applied to (21) with the weighting functions $\mathbf{w}_i(\eta) = \mathbf{a}_i(\eta)$ in order to minimize the residual with respect to the basis $\mathbf{a}_i(\eta)$. We get

$$\sum_{i=0}^{L+1} \hat{f}_i \cdot \left[\int_0^1 \Psi(\mathbf{a}_i(\eta)) \mathbf{w}_j(\eta) d\eta \right] = \sum_{i=0}^{L+1} \hat{g}_i \cdot \left[\int_0^1 \mathbf{b}_i(\eta) \mathbf{w}_j(\eta) d\eta \right]. \quad (22)$$

We can express the same equation in matrix form as $\mathbf{S} \cdot \langle \hat{f} \rangle = \mathbf{M} \cdot \langle \hat{g} \rangle$ using $\langle \hat{g} \rangle = (\hat{g}_i, i = 0, \dots, L+1)$. Here the expression in brackets on the left hand side of (22) are elements of the stiff matrix \mathbf{S} and the expression in brackets on the right hand side of (22) are elements of the mass matrix \mathbf{M} . Both matrices are of rank $L+2$. On top of that they are regular and invertible. The vector $\langle \hat{g} \rangle$ of FE coefficients of g is obtained from

$$\langle \hat{g} \rangle = \mathbf{M}^{-1} \mathbf{S} \cdot \langle \hat{f} \rangle. \quad (23)$$

The backward projection from the FE space to the physical space may be done by evaluating g at K arbitrary locations defined by η_k parameters, $k = 1, \dots, K$. Usually this locations are full, resp. half, model levels. In the case of the integral operator there is the model surface as well. This yields

$$g(\eta_k) = \sum_{i=0}^{L+1} \hat{g}_i \cdot \mathbf{b}_i(\eta_k). \quad (24)$$

We define the projection matrix \mathbf{B} from the FE space to the physical space with elements $\mathbf{B}_{ki} = \mathbf{b}_i(\eta_k)$ and get $\langle g \rangle = \mathbf{B} \cdot \langle \hat{g} \rangle$ for $\langle g \rangle = (g(\eta_i), k = i, \dots, K)$. The matrix \mathbf{B} has dimension $K \times (L+2)$ and does not need to be regular.

Combining (18), (23) and (24) together, we can express the FE operator Ψ applied on $\langle f \rangle = (f(\eta_i), i = 0, \dots, L+1)$ as

$$\Psi(f(\eta)) \approx \Psi \cdot \langle f \rangle = \mathbf{B} \mathbf{M}^{-1} \mathbf{S} \mathbf{A}^{-1} \cdot \langle f \rangle = \langle g \rangle. \quad (25)$$

One can see that each vertical operator Ψ is represented with one single matrix Ψ of rank $K \times (L+2)$. This matrix is time and space independent.

When we describe a FE vertical operator Ψ in the following text we always mention two sets of boundary conditions. The first set represents the *explicit* boundary conditions of the input vector $\langle f \rangle$. They are reflected in the projection matrix \mathbf{A} . The second set represents the *implicit* boundary conditions required on the output vector $\langle g \rangle$. This set of conditions is enforced via the proper choice of B-spline basis functions $\mathbf{b}_i(\eta)$ during the operator construction.

3.4 Linear system

For stability reasons described in details in [1, 3, 4], in the linear model the vertical velocity w is replaced by the vertical divergence based variable $\mathbf{d} = d + X$. Equation (4) should be then replaced by

$$\frac{d\mathbf{d}}{dt} = -g^2 \frac{p}{mRT} \frac{\partial}{\partial \eta} \left(\frac{1}{m} \frac{\partial(p - \pi)}{\partial \eta} \right) + \frac{p}{mRT} \frac{\partial \mathbf{V}}{\partial \eta} \cdot g \nabla w - \mathbf{d}(\mathbf{d} - X) + \frac{dX}{dt} - g \frac{p}{mRT} \frac{\partial \mathcal{W}}{\partial \eta} \quad (26)$$

with $g \nabla w$ being diagnosed from

$$g \nabla w = g \nabla w_s + \int_{\eta}^1 \nabla \left((\mathbf{d} - X) \frac{mRT}{p} \right) d\eta.$$

The total derivative of the residual X is evaluated along the semi-Lagrangian trajectory (see [5]). A transformation from w to \mathbf{d} and vice versa is performed at the beginning and at the end of the explicit computations. Starting from grid-point space computations, an explicit guess for w is first computed using (4). Then, the transformation is applied to obtain an explicit guess for \mathbf{d} . The implicit part of the SI linear system is computed for \mathbf{d} in order to obtain the total forcing term of the linear implicit system. Back to spectral space, the inverse transformation is applied to recover w at the new time level. Having atmosphere in steady state we have

$$\frac{\partial w}{\partial t} = 0.$$

To fulfill this relation the transformation from w to \mathbf{d} and back must be inverse of each other. Unfortunately, this can not be guaranteed by finite-element operators. To ensure this property w is defined on half levels and w to \mathbf{d} transformation and \mathbf{d} to w backward transformation are done with finite-difference operators \mathbf{T}_d , \mathbf{T}_w respectively.

The linear system used to solve implicitly the evolution of fast (gravity and elastic) modes is directly derived from the complete system (3), (26), (5)–(7), through a linearization around a stationary, horizontally homogeneous reference state following principles as for example in Simmons and Burridge [15]. Classically, the

chosen reference state is resting, hydrostatically balanced, dry, isothermal, and with no orography. The value of the surface hydrostatic pressure for this reference state is denoted π_s^* . Two background temperature values T_e^*, T^* are used with the constraint $r = \frac{T_e^*}{T^*} \leq 1$ ensuring satisfactory stability properties. See [2] for details. Instead of horizontal wind \mathbf{V} , the divergence $D = \nabla \cdot \mathbf{V}$ of horizontal velocity is used. We denote $\Delta = \nabla^2$. The linear system for the perturbed variables $(D, \mathbf{d}, T, \hat{q}, q_s)$ may thus be written as

$$\frac{\partial D}{\partial t} = R\mathcal{G}^* \Delta T + RT^* (\mathcal{G}^* - 1) \Delta \hat{q} - RT^* \Delta q_s - \Delta \phi_s, \quad (27)$$

$$\frac{\partial \mathbf{d}}{\partial t} = -\frac{g^2}{RT_e^*} \mathcal{L}^* \hat{q}, \quad (28)$$

$$\frac{\partial T}{\partial t} = -\frac{RT^*}{C_v} (D + \mathbf{d}), \quad (29)$$

$$\frac{\partial \hat{q}}{\partial t} = \mathcal{S}^* D - \frac{C_p}{C_v} (D + \mathbf{d}), \quad (30)$$

$$\frac{\partial q_s}{\partial t} = -\mathcal{N}^* D, \quad (31)$$

where $\mathcal{G}^*, \mathcal{S}^*, \mathcal{N}^*$ and \mathcal{L}^* are linear vertical operators defined similarly as those introduced in [15].

The discretization of vertical integral operators $\mathcal{G}^*, \mathcal{S}^*$ and \mathcal{N}^* is straightforward. We denote \mathbf{K}, \mathbf{P} the integral FE operators from the model top, and from the model surface respectively, up to the given model level l with parameter η . Then we replace

$$\int_0^\eta \psi d\eta \quad \text{with} \quad (\mathbf{K}\psi)_l, \quad (32)$$

$$\int_\eta^1 \psi d\eta \quad \text{with} \quad (\mathbf{K}\psi)_1 - (\mathbf{K}\psi)_l = (\mathbf{P}\psi)_l, \quad (33)$$

where boundary conditions applied on ψ and \mathbf{K} are given in the Table 1. Then the discrete form of the linear model integral operators may be expressed using \mathbf{K}, \mathbf{P} as

$$(\mathcal{S}^* \psi)_l = \frac{1}{\pi_l^*} (\mathbf{K} m^* \psi)_l, \quad (34)$$

$$(\mathcal{G}^* \psi)_l = \left(\mathbf{P} \frac{m^*}{\pi^*} \psi \right)_l, \quad (35)$$

$$(\mathcal{N}^* \psi)_l = (\mathcal{S}^* \psi)_{L+1}, \quad (36)$$

while the discretization of vertical Laplacian term is more cumbersome. We reformulate in the continuous form

$$\mathcal{L}^* \psi = \frac{1}{m^*} \frac{\partial}{\partial \eta} \left(\frac{\pi^{*2}}{m^*} \right) \frac{\partial \psi}{\partial \eta} + \left(\frac{\pi^*}{m^*} \right)^2 \frac{\partial^2 \psi}{\partial \eta^2} \quad (37)$$

and substitute vertical derivations using discrete FE operators $\mathbf{D}_1, \mathbf{D}_2, \mathbf{DD}$ with boundary conditions given in table 1. Finally we replace

$$\mathcal{L}^* \psi(\eta) \quad \text{with} \quad \frac{1}{m_l^*} \left(\mathbf{D}_1 \left(\frac{\pi^{*2}}{m^*} \right) \mathbf{D}_2 \psi \right)_l + \left(\frac{\pi_l^*}{m_l^*} \right)^2 (\mathbf{DD}\psi)_l. \quad (38)$$

3.5 Implicit problem

In the semi-implicit scheme of the ALADIN-NH model with the finite-difference method used in the vertical, a single Helmholtz equation is solved in spectral space. This equation is reached by a suitable algebraic elimination of all variables but one (namely \mathbf{d}). The mathematical constraint (C1) providing a necessary condition

Table 1: Boundary conditions for discrete FE operators.

Operator	Input	Output
$\mathbf{D}_2\psi$	$\psi_0 = 0, \left(\frac{\partial\psi}{\partial\eta}\right)_{L+1} = 0$	$(\mathbf{D}_2\psi)_{L+1} = 0$
$\mathbf{DD}\psi$	$\psi_0 = 0, \left(\frac{\partial\psi}{\partial\eta}\right)_{L+1} = 0$	$(\mathbf{DD}\psi)_{L+1} = 0$
$\mathbf{D}_1\psi$	$\psi_0 = 0, \psi_{L+1} = \psi_L$	-
$\mathbf{D}_h\psi$	$\psi_0 = 0, \psi_{L+1} = \psi_L$	-
$\mathbf{K}\psi$	$\left(\frac{\partial\psi}{\partial\eta}\right)_0 = 0, \left(\frac{\partial\psi}{\partial\eta}\right)_{L+1} = 0$	$(\mathbf{K}\psi)_0 = 0$

for such elimination writes $\mathcal{A}_1 = 0$ for a matrix operator $\mathcal{A}_1 = \mathcal{G}^*\mathcal{S}^* - \mathcal{S}^* - \mathcal{G}^* + \mathcal{N}^*$. Unfortunately, the constraint (C1) is not necessarily fulfilled for finite-elements used in the vertical discretization. For example for the current Czech operational configuration with 87 vertical levels, the spectral radius of \mathcal{A}_1 is 0.198.

On the other hand, the implicit problem in discrete form is a linear inversion, and could be performed with two (or more) variables in the vector. If we adopt a solution of the implicit problem for the couple (D, d) , then the constraint (C1) does no longer need to be fulfilled. We define several matrix operators to simplify the notation

$$\begin{aligned}
\mathbb{A} &= (1 - \delta t^2 c^2 \Delta), \\
\mathbb{B} &= \delta t^2 \Delta (-RT^* \mathcal{G}^* + c^2), \\
\mathbb{C} &= \delta t^2 \Delta RT^* \mathcal{A}_1, \\
\mathbb{E} &= \left(1 - \delta t^2 c^2 \frac{\mathcal{L}^*}{rH^2}\right), \\
\mathbb{F} &= \delta t^2 \frac{\mathcal{L}^*}{rH^2} (-RT^* \mathcal{S}^* + c^2).
\end{aligned}$$

The rank of all these operators equals the number of vertical levels L . Denote d^\bullet, D^\bullet the explicit guess of d, D variables. In the matrix form we write

$$\begin{pmatrix} \mathbb{E} & -\mathbb{F} \\ -\mathbb{B} & \mathbb{A} + \mathbb{C} \end{pmatrix} \begin{pmatrix} d \\ D \end{pmatrix} = \begin{pmatrix} d^\bullet \\ D^\bullet \end{pmatrix}. \quad (39)$$

This system of $2L$ equations could be solved. It is twice as large as for the hydrostatic case, but in terms of computing time it does not constitute a problem because the inversion of the coefficient matrix may be performed only once in the model's setup. However, this solution has particular memory requirements. (We invert indeed a set of matrices, one for each pair of spectral wave numbers, in the setup and store them during the whole integration.) Furthermore, such inversion can not be done for the case of horizontally varying map factor linearization. In this case, the linearized map factor m becomes a spatially-variable operator and multiplication by m and by the spatial operator Δ does not commute. See [20] for details. So we precondition (39) by multiplying with $\begin{pmatrix} \mathbb{A} & \mathbb{F} \\ 0 & \mathbb{I} \end{pmatrix}$. Then, as the predictor step, we guess an approximation of the solution by assuming $\mathcal{A}_1 = 0$, i.e. $\mathbb{C} = 0$. Afterwards, we divide the left hand side matrix operator between the part independent of \mathbb{C} and the rest and look for a solution in a stationary iterative procedure evaluating the \mathbb{C} -independent part by the use of values from the previous iteration step on which the \mathbb{C} -dependent part is applied. Now, we may finish the elimination of variables and invert the left hand side operators separately, similarly as if $\mathbb{C} = 0$. The convergence of the method is ensured if the spectral radius of the iteration matrix is not larger than 1. This condition was fulfilled in all tested cases, both idealized and real simulations.

The i th residual r_i can be defined as the usual l^2 -norm of the difference between the left hand side matrix operator of (39) applied to the solution reached after i steps of the iterative algorithm and the right hand side

of (39), calculated over the whole domain. Then the convergence of the method ensures that $\lim_{i \rightarrow \infty} r_i = 0$. We can formulate the termination criterion on threshold τ as

$$\frac{r_i}{r_0} < \tau.$$

Then the threshold $\tau = 0.5$ is reached with only 1 iteration in all tested cases, see Fig. 4 for details. We conclude that the speed of convergence of the proposed method is satisfactory.

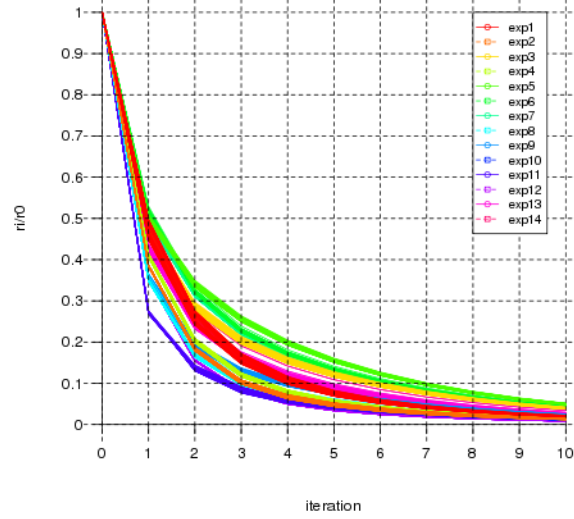


Figure 4: The speed of convergence in tested real cases.

3.6 Non-linear system

The vertically discretized fully compressible Euler equations write

$$\begin{aligned} \left(\frac{d\mathbf{V}}{dt} \right)_l &= -\frac{RT_l}{p_l} \nabla p_l - \left(\frac{p_l}{\pi_l} + \frac{p_l}{m_l} (\mathbf{D}_1 \hat{q})_l \right) \cdot (\nabla \phi)_l, \\ \left(\frac{dT}{dt} \right)_l &= -\frac{RT_l}{C_v} (D_3)_l, \\ \left(\frac{d\hat{q}}{dt} \right)_l &= -\frac{C_p}{C_v} (D_3)_l - \frac{1}{\pi_l} \cdot \omega_l, \\ \frac{\partial q_s}{\partial t} &= -\frac{1}{\pi_s} \cdot (\mathbf{K} \nabla \cdot m \mathbf{V})_L. \end{aligned}$$

These equations are evaluated at the model full levels, while the equation for vertical velocity is evaluated at the model half levels, since gw is a half level quantity. It is the reason why another FE operator denoted \mathbf{D}_h is needed which gives values on the half levels when applied on a full level variable ψ . We have

$$\left(\frac{dw}{dt} \right)_{\tilde{l}} = \frac{g}{m_{\tilde{l}}} (\mathbf{D}_h (p - \pi))_{\tilde{l}} \quad \text{and} \quad m_{\tilde{l}} = \frac{\pi_{l+1} - \pi_l}{\eta_{l+1} - \eta_l}.$$

Terms $\nabla \phi$, D_3 and ω further include vertical operators and are hence affected by the choice of vertical discretization. They are discretized as follows:

$$\begin{aligned} (\nabla \phi)_l &= \nabla \phi_s + \left(\mathbf{P} \nabla \left(\frac{mRT}{p} \right) \right)_l, \\ (D_3)_l &= D_l + \frac{p_l}{m_l RT_l} (\mathbf{T}_d (gw))_l - \frac{p_l}{m_l RT_l} (\mathbf{D}_1 \mathbf{V})_l \cdot (\nabla \phi)_l, \\ \omega_l &= \mathbf{V}_l \cdot \nabla \pi_l - (\mathbf{K} \nabla \cdot m \mathbf{V})_l. \end{aligned}$$

4 Sensitivity in idealized experiments

A set of test cases has been run in the 2D vertical plane version of ALADIN-NH, including the potential flow and the non-linear non-hydrostatic flow over idealized orography according to [6] and the density current test published by Straka in [16].

First, we examine flows with a constant uniform velocity U in x direction in a dry atmosphere with a constant Brunt-Vaisälä frequency N over a bell-shape mountain characterized by its height H and its half-width a . The surface geopotential is defined by

$$\Phi_s(x) = g \cdot H \frac{a^2}{a^2 + x^2}.$$

Potential flow The first test is given by $U = 15m \cdot s^{-1}$, $N = 0.02s^{-1}$. The atmosphere is isothermal with $T_0 = 239K$. The mountain has maximum height $H = 100m = 5\Delta x$ and half-width $a = 100m$. The horizontal resolution is $\Delta x = 20m$ with 128 points, while in the vertical we have 120 vertical levels with nearly regular spacing of $\Delta z = 20m$ up to 1.8km and geometric increase of Δz over this limit up to 36km. We use the time step of 1s. In this regime no wave is generated, the flow only ascends on one side of the mountain and descends on the other side. Both vertical schemes give satisfactory results in this test, see Fig. 5 for the fields of vertical velocity at time 500s.

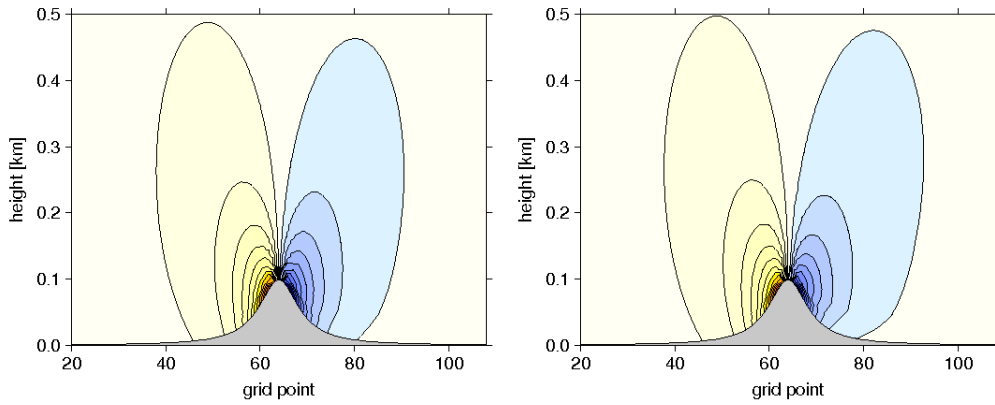


Figure 5: Potential temperature at time 500s, the contour interval is 1K. On the left: finite differences; on the right: finite elements with 1 iteration of the SI solver.

Non-linear non-hydrostatic flow For this regime we set $U = 4m \cdot s^{-1}$, $N = 0.01s^{-1}$, bottom temperature $T_0 = 285K$, and the mountain characteristics to $H = a = 400m = 5\Delta x$. The atmosphere is not isothermal in this case. We have 256 grid points in x direction with $\Delta x = 80m$ and regular vertical spacing with 150 levels and $\Delta z = 200m$ up to the top of the atmosphere placed at 30km height. We use the time step $\Delta t = 4s$ and integrate up to 8000s. Results for the vertical velocity field for the two vertical discretizations are shown in Fig. 6. The additional wave pattern pronounced in VFD experiment (left) is partially smoothed in the VFE experiment (right).

The question has arisen how the stability of the designed FE discretization scheme depends on the vertical levels distribution. To answer this question, we set a general distribution of η -levels with parameter α by

$$\begin{aligned} \sigma_l &= \frac{l}{L}, \eta_l = \alpha \sigma_l + 3(1 - \alpha) \cdot \sigma_l^2 - 2(1 - \alpha) \sigma_l^3, \\ B_l &= (3 - 2\eta) \cdot \eta_l^3, A_l = 101325 \cdot (\eta_l - B_l). \end{aligned}$$

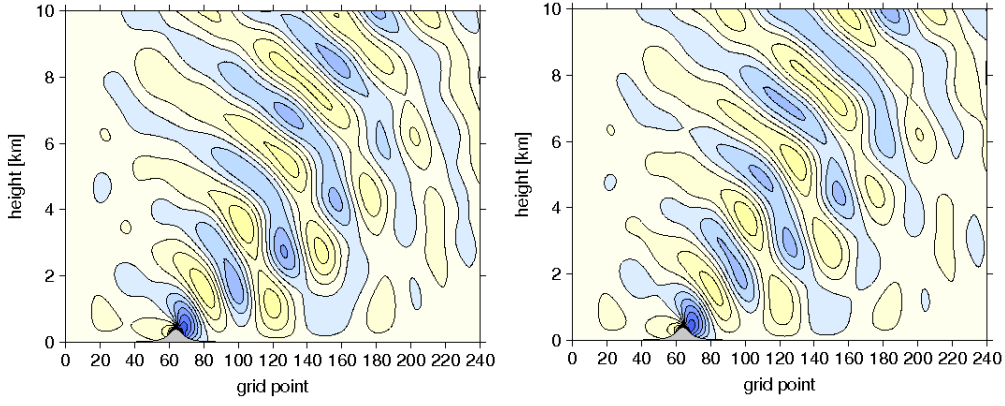


Figure 6: Vertical velocity at time 8000s, the contour interval is 0.4ms^{-1} . On the left: finite differences; on the right: finite elements with 1 iteration of the SI solver.

Then we let α vary in the set of values $\{0., 0.5, 1., 1.5, 2.\}$. Thus we got various vertical levels distributions and prepared the non-linear non-hydrostatic flow test with them. None of them has apparent stability problems, although the solutions slightly vary for the respective cases.

Density current We apply the same initial configuration as in [16] for comparison. We have 1024 points in the x direction with $\Delta x = 50\text{m}$ and 200 vertical levels with $\Delta z = 50\text{m}$. To initiate a density current, the temperature field is specified as a sum of the background value calculated from the constant potential temperature $\theta_0 = 300\text{K}$ and a temperature perturbation $T'(x, z)$ symmetric around a central point $[x_c, z_c]$ with maximum value of $T_0 = 15\text{K}$ according to the following definition

$$T(x, z) = \theta_0 \left(\frac{\pi_s}{\pi_0} \right)^{\frac{R}{C_p}} + T'(x, z),$$

$$T'(x, z) = \begin{cases} 0 & \text{iff } L > 1 \\ T_0 (\cos(\pi L) + 1) / 2 & \text{iff } L \leq 1 \end{cases}$$

$$\text{for } L = \sqrt{\left(\frac{x - x_c}{x_r} \right)^2 + \left(\frac{z - z_c}{z_r} \right)^2}$$

with $\pi_0 = 1000\text{hPa}$, $x_c = 2560\text{m}$, $x_r = 4000\text{m}$, $z_c = 3000\text{m}$ and $z_r = 2000\text{m}$. We use the time step of 3s. Periodic boundary conditions are imposed in the horizontal direction. The results of the experiments with the two vertical discretizations used are shown in Fig. 7. The plotted field is the potential temperature at time 900s, i.e. after 300 steps. Both results are in good agreement with the results shown in Fig.1 of [16].

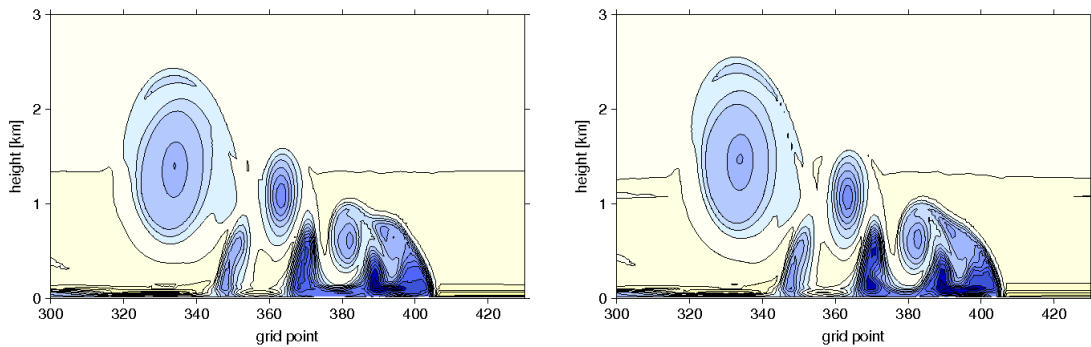


Figure 7: The potential temperature field at time 900s of the Straka test, the contour interval is 1K. On the left: finite differences; on the right: finite elements with 1 iteration of the SI solver.

5 Sensitivity in real-case experiments

A set of forecasts starting from the ARPEGE analysis at 00 UTC was run in 2.2km horizontal resolution, with 87 Czech operational vertical levels over the Central Europe domain covering Alps, see Fig. 9, for two 10 days time series. One serie covered a winter time period starting from 21 December 2011 to 31 December 2011, the second serie was a summer one starting from 1 July 2012 to 10 July 2012. The time step used was 90s, the forecast integration range 24 hours with 3 hours coupling interval for coupling to ARPEGE LBC files. We use a non-extrapolating two-time level iterative centered-implicit time scheme with 1 iteration and semi-Lagrangian advection. As described in Section 2, we use the w variable in the explicit part of the model (LGWADV=T), and d in the semi-implicit part. The additional prognostic non-hydrostatic variable $\hat{q} = \ln\left(\frac{p}{\pi}\right)$ is employed for pressure estimations. As parametrization package, the ALARO physics was applied with non-parametrized deep convection, microphysics applied to resolved clouds and precipitation (LSTRAPRO=T, L3MT=F).

No sign of instability is apparent in any of the experiments. The iterative semi-implicit solver converges as indicated by the spectral radius test of the iteration matrix calculated in the setup part of the integration. Objective scores of the results with 1 iteration and with 8 iterations coincide in all parameters except time evolution of bias for precipitation cumulated for 24hours in winter period, where a small advantage of further iterations may be observed, see Fig. 8. We conclude that one iteration of the semi-implicit solver is enough for an accurate solution.

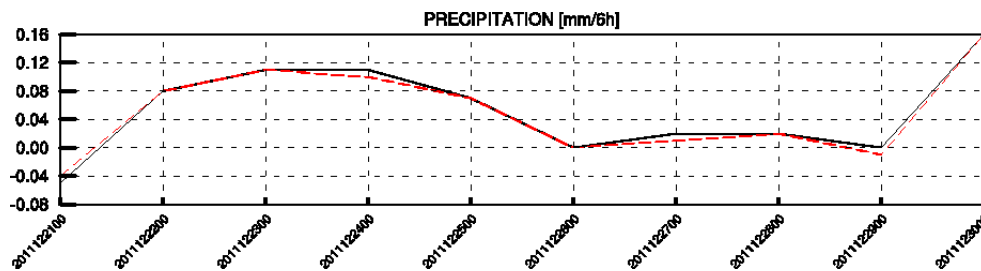


Figure 8: Bias time evolution of individual model runs for cumulated precipitation with the forecast range of 24 hours, winter time serie, run with 1 iteration in semi-implicit solver (black) and with 8 iterations (red).

Furthermore, objective scores characteristics are neutral to the change of vertical discretization (from VFD to VFE). The phenomenon which can be identified in the results is an interaction of vertical discretization with the resolved convection. The fields of precipitation cumulated for 1 hour are shown in Fig. 9 for the 22nd hour of integration starting from 6 July 2012 00 UTC. When using finite elements the weak precipitations occur

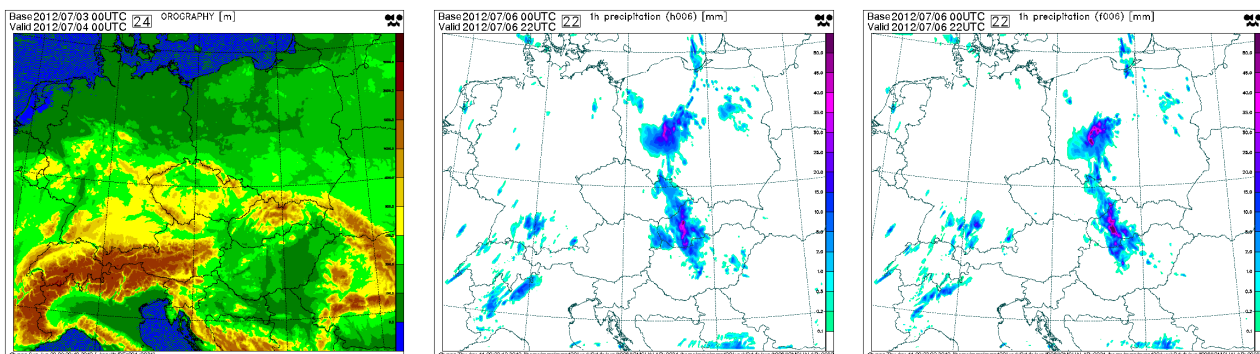


Figure 9: Orography used for the real-case experiments (left) and the precipitation fields cumulated for 1 hour: vertical finite differences (middle) and VFE, 1 iteration of SI solver (right).

less often, while there is a shift to more intense precipitations. Histograms for 1 hour cumulated precipitations

calculated for each hour of the 24 hours integration for all cases of the winter and the summer time serie are depicted in Fig. 10.

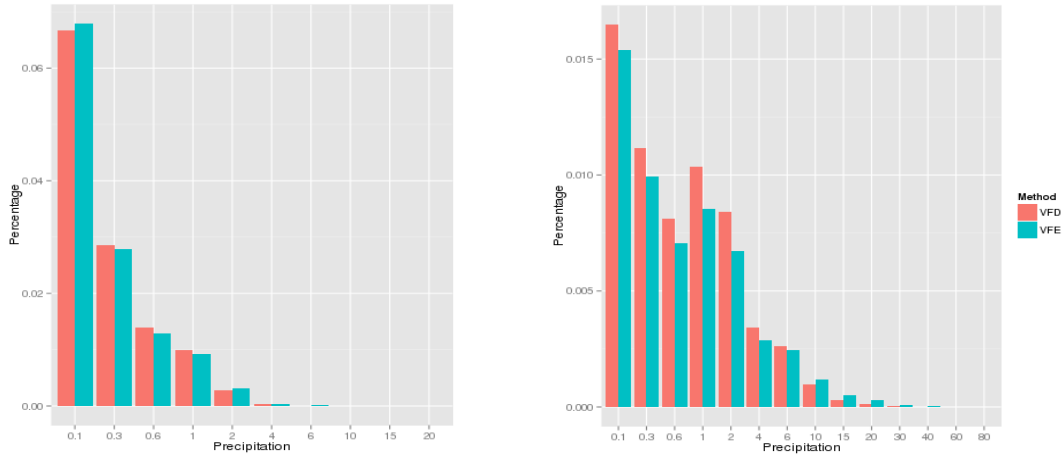


Figure 10: Precipitation density estimates divided into intervals used in radar measurements: winter (left), summer (right).

6 Summary and discussion

We describe here an extension of the existing finite element vertical discretization of the hydrostatic dynamical core of the LAM model ALADIN to its non-hydrostatic version ALADIN-NH. We redefine the way of interpolating prognostic quantities with B-spline curves. The centripetal definition of parameter η is used and the B-spline bases are constructed with de Boor's recursive algorithm.

The stability properties of the NH dynamical core require to keep vertical staggering of the model variables for FE discretization. Hence unlike all the other prognostic variables the vertical velocity is defined on the model half levels. The vertical derivative of w then requires an application of a staggered FE operator. This requirement limits the theoretical accuracy of the proposed FE method, because staggered FE operators are not superconvergent. Moreover, the transformations between w and the vertical divergence variable d needed in the implicit calculations keep the finite-difference approach. This may again limit the reached overall accuracy of the whole model. This is left for further investigation.

The treatment of boundary conditions was changed as well. FE vertical operators are explicitly defined on the material boundaries, not only on the model full resp. half levels. The boundary conditions of prognostic variables are prescribed explicitly. The boundary conditions of the vertical operator itself are implicitly build into the B-spline basis functions. The whole set of vertical discrete operators was defined in order to satisfy various boundary conditions required by different terms in the system of prognostic equations of ALADIN-NH. In addition compared to FD spectral space computations, we had to implement a stationary iterative solution of the Helmholtz structure equation. The proposed method appears to be convergent and it was shown that one iteration provides a sufficient accuracy.

We performed a set of standard idealized tests, like the density current (Straka test) and various flow regimes over an Agnesi shaped mountain. These experiments proved the satisfactory accuracy properties of the proposed FE scheme and they showed that the FE NH dynamical core is as stable as the finite difference one when using the iterative centered implicit time scheme with one iteration (predictor - corrector scheme). Moreover, 3D diabatic experiments were performed with model horizontal resolution of $2.2km$ over the Central Europe domain covering Alps. The objective scores were neutral to the change of vertical discretization in all tested cases. A slight shift of the precipitation amounts to higher intensities was observed with the FE method used,

especially in the summer period.

We have implemented the FE method with the general order of B-splines. So far all tests were restricted to the cubic B-splines only. Nevertheless, we plan to study the influence of the B-spline order on the accuracy and the time stepping stability of the whole system. Further, we plan to extend the designed FE vertical discretization scheme into the dynamical core of the global model ARPEGE/IFS. This task requires to overcome some further difficulties. It has been shown that using non-hydrostatic equations in stretched version of ARPEGE (and similarly using ALADIN-NH with large domains) where the mapping factor varies significantly from the value of 1 can lead to instabilities if the model is linearized around a constant reference mapping factor. In this case, the real value of the mapping factor has to be taken into account in the semi-implicit scheme. This option has been coded in ARPEGE (hydrostatic and non-hydrostatic versions) under the key LSIDG. However, a harmonization with the proposed FE method for vertical discretization with its iterative solution to the implicit problem needs further investigation.

The current implementation is expected to enter the official ALADIN code at the beginning of year 2014.

Acknowledgments We wish to thank our colleagues from AEMET in Spain for stimulating discussions. We would like to thank also all colleagues at CHMI in Prague and at ZAMG in Vienna for their hospitality during stays of the first author at these institutions. Finally, we thank the RC LACE for financially supporting the stays.

References

- [1] Bénard P., 2003: Stability of semi-implicit and iterative centered-implicit time discretizations for various equation systems used in NWP. *Mon. Wea. Rev.*, **131**, 2479–2491.
- [2] Bénard P., 2004: On the use of a wider class of linear systems for the design of constant-coefficients semi-implicit time-schemes in NWP. *Mon. Wea. Rev.*, **132**, 1319–1324.
- [3] Bénard P., Laprise R., Vivoda J., Smolíková P., 2004: Stability of leapfrog constant-coefficient semi-implicit schemes for the fully elastic system of Euler equations: Flat-terrain case. *Mon. Wea. Rev.*, **132**, 1306–1318.
- [4] Bénard P., Mašek J., Smolíková P., 2005: Stability of leapfrog constant-coefficient semi-implicit schemes for the fully elastic system of Euler equations: Case with orography, *Mon. Wea. Rev.*, **133**, 1065–1075.
- [5] Bénard P., Vivoda J., Mašek J., Smolíková P., Yessad K., Smith Ch., Brožková R., Geleyn J.-F., 2010: Dynamical kernel of the Aladin–NH spectral limited-area model: Revised formulation and sensitivity experiments, *Q. J. R. Meteorol. Soc.*, **136**, 155–169.
- [6] Bubnová, R., Hello G., Bénard P. and Geleyn J.-F., 1995: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP System. *Mon. Wea. Rev.*, **123**, 515–535.
- [7] Côté J., Béland M., Staniforth A., 1983: Stability of vertical discretization schemes for semi-implicit primitive equation models: Theory and application. *Mon. Wea. Rev.*, **111**, 1189–1207.
- [8] de Boor C., 2001: A practical guide to splines. Applied mathematical sciences 27, Springer-Verlag: New York.
- [9] Hoschek J., Lasser D., 1993: Fundamentals of Computer Aided Geometric Design, A. K. Peters, Wellesley, Mass, USA.
- [10] Laprise R., 1992: The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Weather Rev.*, **120**, 197–207.
- [11] Lynch D. R., 2005: Numerical partial differential equations for environmental scientists and engineers. A first practical course, Springer-Verlag: New York.
- [12] Mašek J., 2011: Discretization of horizontal pressure gradient force in model ALADIN. *Report from RC LACE scientific stay*, internal document of RC LACE.

- [13] Phillips, N. A., 1966: The equations of motion for a shallow rotating atmosphere and the “traditional approximation”. *J. Atmos. Sci.*, **23**, 626–628.
- [14] Simarro J., Hortal M., 2012: A semi-implicit non-hydrostatic dynamical kernel using finite elements in the vertical discretization. *Quart. J. Roy. Meteor. Soc.*, **138**, 826–839.
- [15] Simmons, A., and Burridge D., 1981: An energy and angular momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109**, 758–766.
- [16] Straka J.M., Wilhelmson R.B., Wicker L.J., Anderson J.R., Droegemeier K.K., 1993: Numerical solutions of a non-linear density current: A benchmark solution and comparisons, *Int. J. Numer. Methods Fluids*, **17**, 1–22.
- [17] Ritchie H., Temperton C., Simmons A.J., Hortal M., Davies T., Dent D., Hamrud M., 1995: Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, 489–512.
- [18] Tanguay M., Robert A., Laprise R., 1990: A semi-implicit semi-Lagrangian fully compressible regional forecast model., *Mon. Wea. Rev.*, **118**, 1970–1980.
- [19] Untch A., Hortal M., 2004: A finite-element scheme for the vertical discretization of the semi-Lagrangian version of the ECMWF forecast model. *Q. J. R. Meteorol. Soc.*, **130**, 1505–1530.
- [20] Yessad K., 2006: Semi-implicit spectral computations in the NH version OF ARPEGE/ALADIN: Specific problems when the constraint "C1" is relaxed., internal note of Météo France.

Status of ARPEGE and AROME physics

Yves Bouteloup

with contributions from

Eric Bazile, François Bouyssel, Rachel Honnert, Jean-Marcel Piriou and Yann Seity

1 Introduction

Three parts in this short paper. The first one is the description of the physical modifications in the present experimental suite for AROME and ARPEGE. The second one is about PCMT, the new deep convection scheme of ARPEGE and some experiments with PMMC09 the shallow convection scheme used in AROME. In the third part some "on going work" is rapidly described.

2 Current e-suite of ARPEGE and AROME

There are not a lot of modifications concerning the physical parametrization in this e-suite.

2.1 AROME

- Correction of negative humidity at the end of analysis.
- New version of SURFEX (v7.2)
- Deactivation of CANOPY over sea to avoid instability in case of strong wind.

2.2 ARPEGE

- Modification in the deep convection scheme to avoid unexpected cyclogenesis.
- Modification of the thermal inertia, albedo and z0 for glacier.
- Modification in the KFB shallow convection scheme.

3 PCMT and PMMC09

3.1 PCMT: The new deep convection scheme of ARPEGE

- Better diurnal cycle of precipitation especially in region with dry medium troposphere.
- Better coastal behaviour with precipitation over land and not only over sea (Figure 1)

- Improvement of synoptic scores of wind.
- But some problems with horizontal variability of precipitation over France.

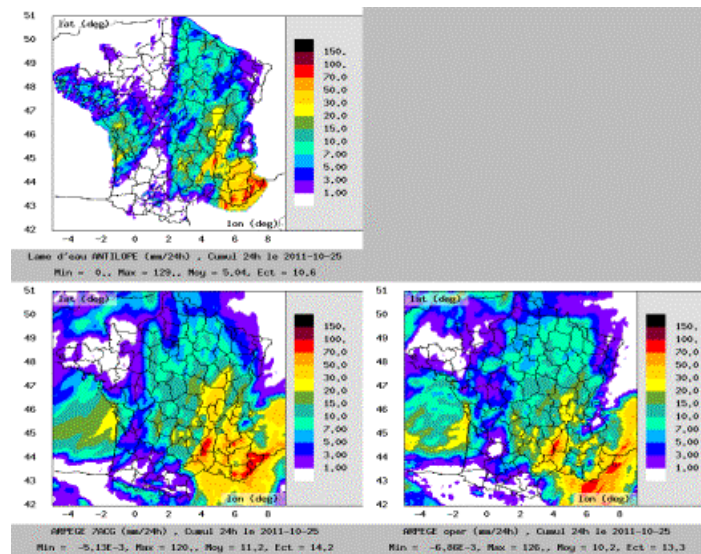


Figure 1: Upper left: Observation of 24h precipitation accumulation for the 25 october 2011. Lower right: Operational model, in the south east of France precipitation are over sea. Lower left: Experiment with PCMT. One can see the better behaviour in the south-east of France.

3.2 PMMC09: The shallow convection scheme of AROME

Tests of PMMC09 with AROME's tuning in ARPEGE has shown a too strong activity. The use of the formulations of entrainment and detrainment from Rio et al (2010) has significantly improved the simulation of the 1D Arm cumulus case (Figure 2). Tests in ARPEGE of these formulations lead to an improvement of the wind in the tropical area but there is still a too strong cold bias of temperature around 850hpa. It is expected that this problem is due to a bad interaction with deep convection scheme. Tests with PCMT are planned.

4 On going work

4.1 High resolution in AROME

A 1.3km resolution of AROME is running daily with a 720x720 points domain over France. It uses 90 vertical levels with the lowest level at 5m and 33 levels below 2000m (only 21 levels below 2000m in operation). At this resolution the maximum slope is 39° against 23° in the 2.5 km operational model. The future operational domain (Figure 3) will be larger than the one used during the benchmark (1500x1440 against 1350x1440)

A very interesting non-hydrostatic feature (orographic waves) has been simulated by the high resolution model (31st january 2013 + 14TU Figure 4). The horizontal resolution is the key to access that kind of feature. A test with the hydrostatic dynamic and a 1.3km resolution doesn't resolve these structures (not shown).

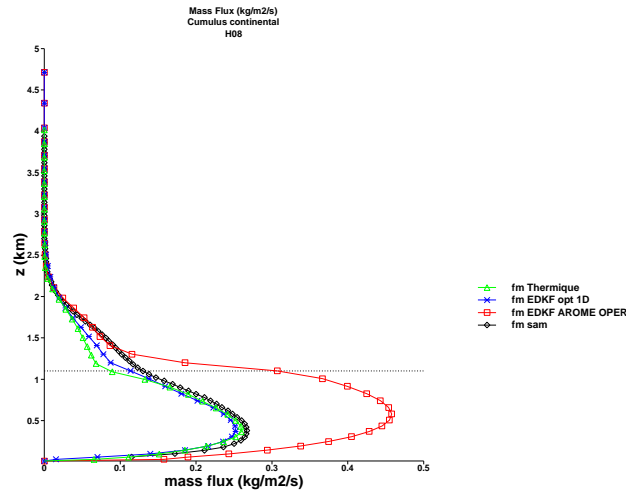


Figure 2: Mass flux as a function of height after 8 hours of simulation for the Arm cumulus case. LES is in black, current operational formulation in red, PMMC09 with Rio et al (2010) formulations is in blue, mass flux scheme from IPSL is in green.

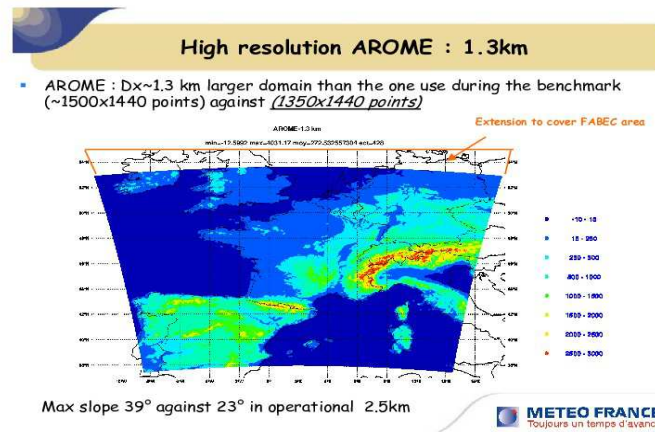


Figure 3: The benchmark 1350x1440 points domain and the north FABEC extension of the future operational 1500x1440 domain

4.2 Grey zone of turbulence

Turbulence is well represented by atmospheric models at very fine grid sizes, from 10 to 100m for which turbulent movements are mainly resolved, and by atmospheric models with grid sizes greater than 2km, for which those movements are entirely parameterized. Honnert et al (2011) provide a new diagnostic by calculating the subgrid and resolved parts of five variables at different scales: turbulent kinetic energy, heat and moisture flux, and potential temperature and mixing ratio variances. This new diagnostic gives a reference of evaluating current and future parameterizations at kilometric scales (Figure 5) In its standard version, the model produces too many resolved movements, as the turbulence scheme does not sufficiently represent the impact of the subgrid thermal. This is not true when a mass flux scheme (PMMC09) is introduced. However in this case, a completely subgrid thermal is modelled leading to an overestimation of the subgrid part (not shown)

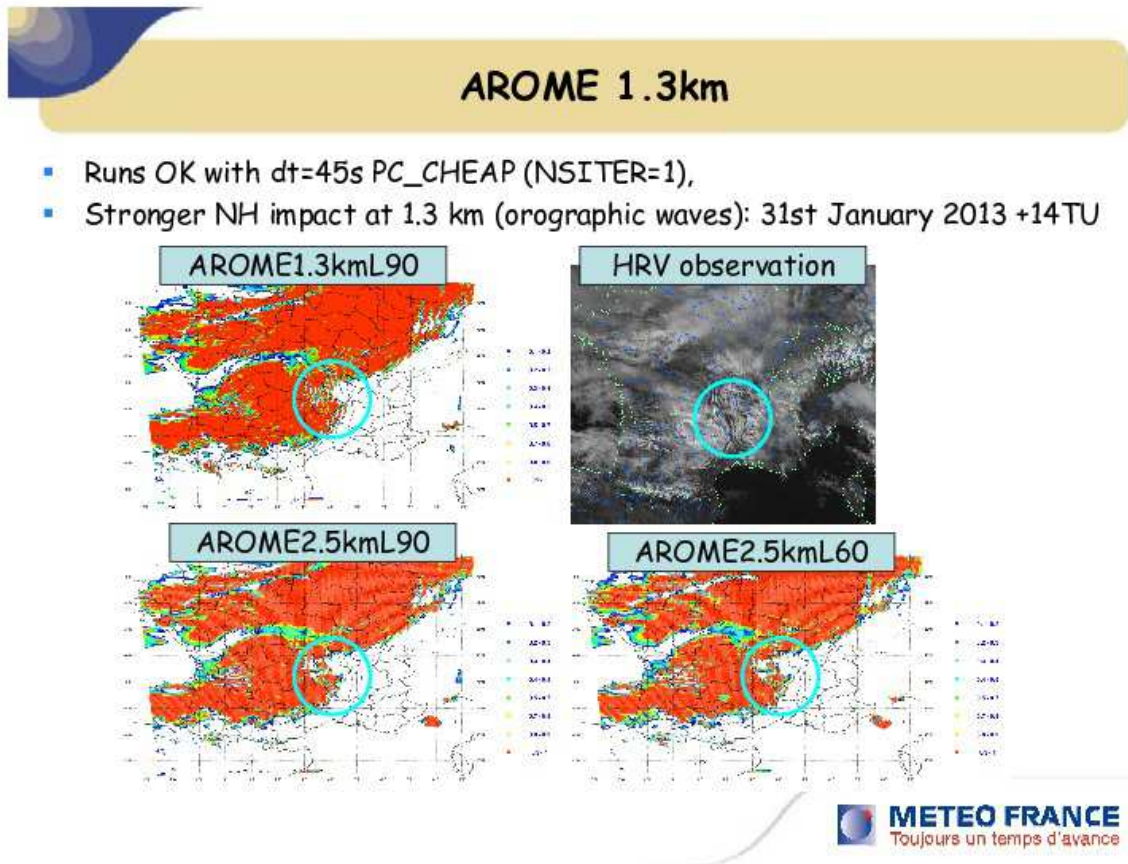


Figure 4: Orographic waves over Massif-Central are only present in non-hydrostatic 1.3km simulation. Not in 2.5km non-hydrostatic simulation nor in 1.3km hydrostatic test (not shown)

(b) $0.05 \leq \frac{z}{h} \leq 0.85$

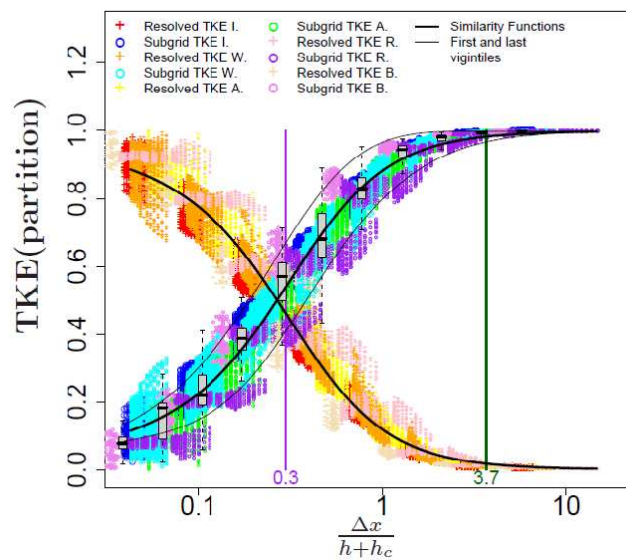


Figure 5: Resolved (crosses: yellow to red) and subgrid (circles: blue, green and purple) TKE as a function of the dimensionless mesh size

References

Honnert R., V. Masson and F. Couvreux, A diagnostic for Evaluating the representation of Turbulence in Atmospheric Models at the Kilometric Scale, *Journal of the Atmospheric Sciences*, 68, 3112-3131, 2011

Rio, C., F. Hourdin, F. Couvreux and A. Jam, Resolved Versus Parameterized Boundary-layer Plumes. Part II: Continuous Formulations of Mixing Rates for Mass-Flux Schemes, *Boundary Layer Meteorologie*, 135, 469-483, 2010.

Experiences of HARMONIE at IMO

Bolli Palmason, Guðrún Nína Petersen, Hróbjartur Thorsteinsson, Sigurður Thorsteinsson

1 Introduction

In the autumn of 2011 the Icelandic Meteorological Office (IMO) started running the HARMONIE NWP for the Icelandic domain at 2.5 km horizontal resolution. In September 2012 we reached a milestone when we upgraded to version 37 and pre-operational status.

In this article we'd like to share with you some of our experiences of the HARMONIE model. Iceland is an island close to the Arctic circle with varying landscape, fjords, mountainous areas and glaciers as well as agricultural regions and wet lands. Extreme weather events are frequent, as Iceland is located in the North Atlantic storm track. Consequently Icelandic presents HARMONIE with a real challenge in simulating and predicting various weather phenomena. The model setup is described in the next section, followed by sections describing our experience with forecasts of wind (Section 3), temperature (Section 4), precipitation (Section 5) and snow accumulation (Section 6) before ending with a short summary of the findings (Section 7).

2 Model setup

The current setup and the one used in the experiments in this article is version 37h1.2 of the HARMONIE model. The model grid has a horizontal resolution of approximately 2.5 km and the Icelandic domain has 300×240 horizontal grid points in the horizontal, see Fig. 1, and 65 vertical levels with model top at roughly 10 hPa. The lateral boundary conditions are from ECMWF and are updated every 3 hours. The forecast model is run in non-hydrostatic mode utilizing AROME physics parameterizations. The forecasting system consists of CANARI + OI_Main surface analysis and no upper-air analysis, i.e. blending mode. Around 230 SYNOP observations from the Icelandic observational network are included in the surface analysis. The model runs 4 times a day with a forecasting range of 48 hours.

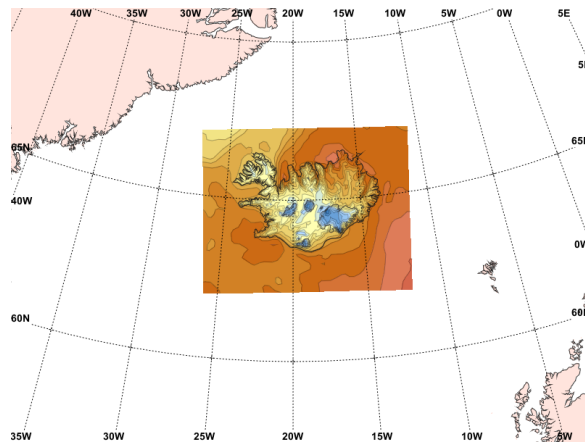


Figure 1: The horizontal extent of the Iceland model domain is illustrated by the coloured area.

3 Winds

A rather extreme winter storm hit Iceland on 2 November 2012. The HARMONIE had difficulties, as well as all other NWP models available to IMO, to forecast the high wind speeds observed. To try to improve HARMONIE forecasts of such wind storms an experiment was carried out investigating the impact of the sub-grid parameterization on the 10-metre wind speed, and if it was possible to calibrate the parameterization better. In particular it was our understanding that it was the sub-grid scale orographic frictional parameterization that damped the surface winds too much. In these experiments the pre-operational setup of the model was applied and the model run for the whole of July and November 2012 in forecasting mode.

3.1 Impact of sub-grid orographic drag on 10 m wind speed

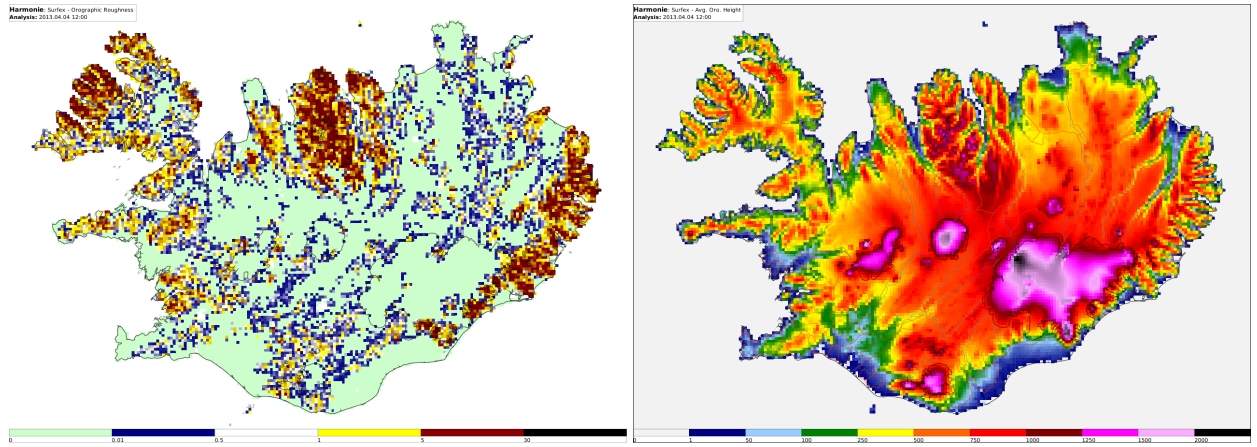


Figure 2: The sub-grid orographic roughness length (left) and average orographic height (right) in HARMONIE-37h1.2.

There are two orographic drag options in HARMONIE: Z01D and BE04. The default drag parameterization is the Z01D parameterization:

$$drag_{Z01D} = \rho^2 \left(\frac{0.4}{\ln \frac{H}{Z_0}} \right)^2 U; \quad Z_0 = \min \left(Z_0, \frac{H}{XFRACD} \right)$$

where Z_0 is orographic roughness length, ρ air density, U wind speed and H the height of the atmospheric forcing level. XFRACD is chosen to minimize the bias and RMSE (See Y. Seity, C. Lac, V. Masson: About orographic drag options in SURFEX. Tech. Report.)

The alternative is the Beljaars parameterization:

$$drag_{BE04} = 2\alpha\beta C_{md} C_{corr} C_a S_{st}^2 H^{-1.2} \left(e - \frac{H}{1500} \right)^{1.5} U$$

where H is the altitude, S_{st} is the sub-grid orography standard deviation and other parameters are constants described in Beljaars (2004).

3.2 Experiment design

Two parallel experiments were carried out for July and November 2012. The parallel experiments, summarized below, were different only in the value of XFRACD which where the aim was to minimize the bias and RMSE.

- Experiments using the default parameterization, Z01D, with XFRACD=15 for both months.

- Experiments using the alternative parameterization BE04 for November.
- Experiments using Z01D with XFRACD=1000 for both months.
- Experiments using Z01D but with two values of XFRACD, i.e. reducing Z_0 for high wind speed (tunewind): XFRACD=2500 for $U < 13$ m/s and XFRACD=10000 for $U \geq 13$ m/s for November.

We used lateral boundary conditions from +6 h old ECMWF forecasts and all SYNOP observations, except snow depth, were sent externally from IMO in a 6 h cycle. For the 6 h data assimilation experiment an observation time window of ± 3 h was used. Fig. 2 shows the sub-grid orographic roughness length and orographic height for Iceland in HARMONIE-37h1.2.

3.3 Verification against observations

For all experiments forecasts with a forecast range of 48 hours were run from 00 and 12 UTC. These forecasts were verified against SYNOP observations. Fig. 3 shows scatter-plots of 10 m winds for November 2012, forecasts compared to observations. It is clearly seen from the figure that the default Z_0 XFRACD=15 scores (upper left) systematically underestimated high wind speeds during storms and had a large negative bias while the BE04 (upper right) systematically overestimated wind speed for lower wind speeds and had a positive bias. We chose to tune up Z01D rather than tune down BE04. Using XFRACD=1000 (not shown) gave some improvement but it turned out that it gave better results to use two values of XFRACD, depending on the forecasted wind speed (tunewind). Here the Z_0 is reduced for higher wind speeds applying XFRACD=2500 for $U < 13$ m/s and XFRACD=10000 for $U \geq 13$ m/s. Note that there is one drawback, by reducing Z_0 in this way results in cases where the forecasted wind speed is much too large at low wind speed, $U_{obs} < 3$ m/s, i.e. there are cases with forecasted wind being about 20 m/s when only 3 m/s is observed. This needs some consideration.

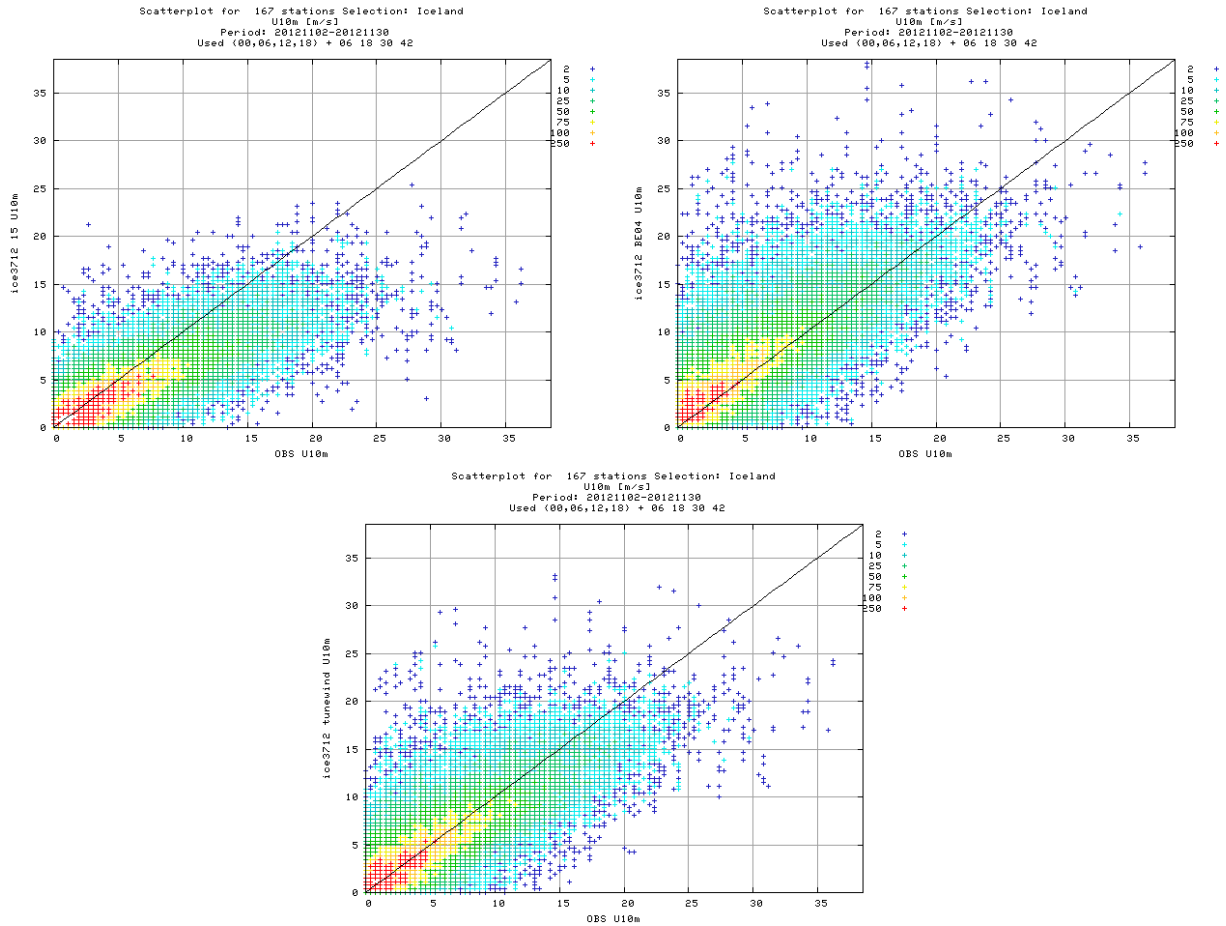


Figure 3: Scatterplots of observed and forecasted 10 m wind speed during November 2012 for Z_0 with $XFRACD=15$ (top left) for BE04 (top right) and for reduced Z_0 , applying tunewind (bottom).

Fig. 4 shows the bias and root mean square error (RMSE) verification scores averaged over November 2012 (upper left). Reduced Z_0 with $XFRACD=1000$ and tunewind leads to improved bias and RMSE fit for the 10 m wind speed. The frequency distribution for this month (upper right) reveals that the reduced Z_0 follow best the frequency distribution of the observations. The Kuiper Skill Scorea (lower left) reveals the drawback that reducing Z_0 gives too strong forecasted winds at $U_{obs} < 3$ m/s. For July 2012 the reduced Z_0 also leads to improved bias and RMSE fit.

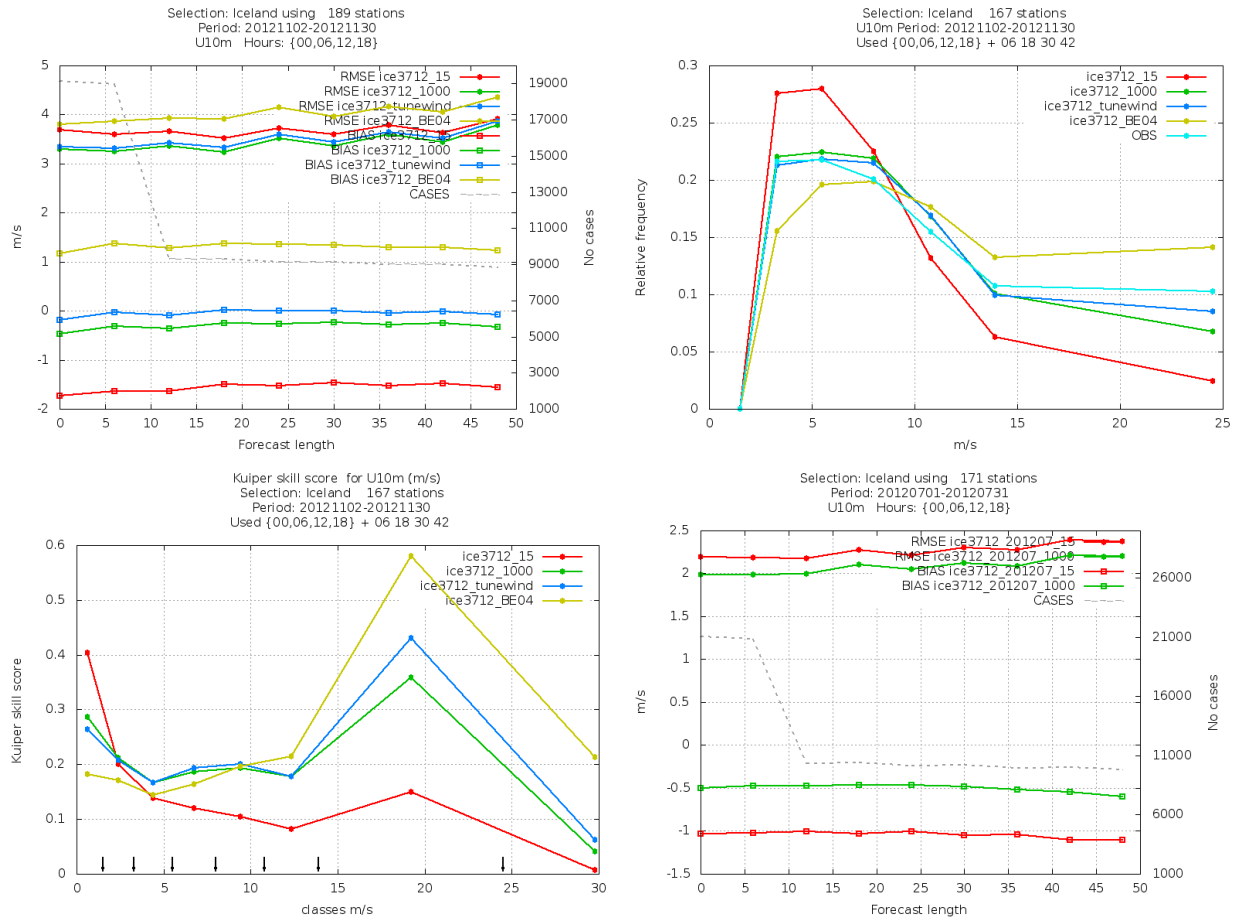


Figure 4: A few error statistics for 10 m wind forecasts: The RMSE and bias for November 2012 (top left), relative frequency for November 2012 (top right), Kuiper Skill Score for November 2012 (bottom left) and RMSE and bias for July 2012 (bottom right).

3.4 Conclusions

The HARMONIE system has been set up for use with two orography drag options and varying value of orographic surface roughness and evaluated for July and November 2012. When applying tunewind the orographic surface roughness is depended on the forecasted wind speed. Future roughness experiments with HARMONIE cycle 38 will also evaluate the roughness for different surface types as well as running with and without sub-grid orography drag.

The parameterization of wind speed at 10 m seems to be a general problem for all of the NWP models available to us. For the first time we have an opportunity to tune the orographic roughness drag such that it better fits the Icelandic 10 m wind observations.

4 Temperature

4.1 Comparing the forecasting performance of HARMONIE to that of other NWP models

The HARMONIE forecasts, as well as other NWP forecasts, are routinely compared to observations and the error statistics calculated monthly for a number of stations. A comparison of the performance of forecasts gives a good picture of how the models are performing and where the largest errors in the forecasts occur.

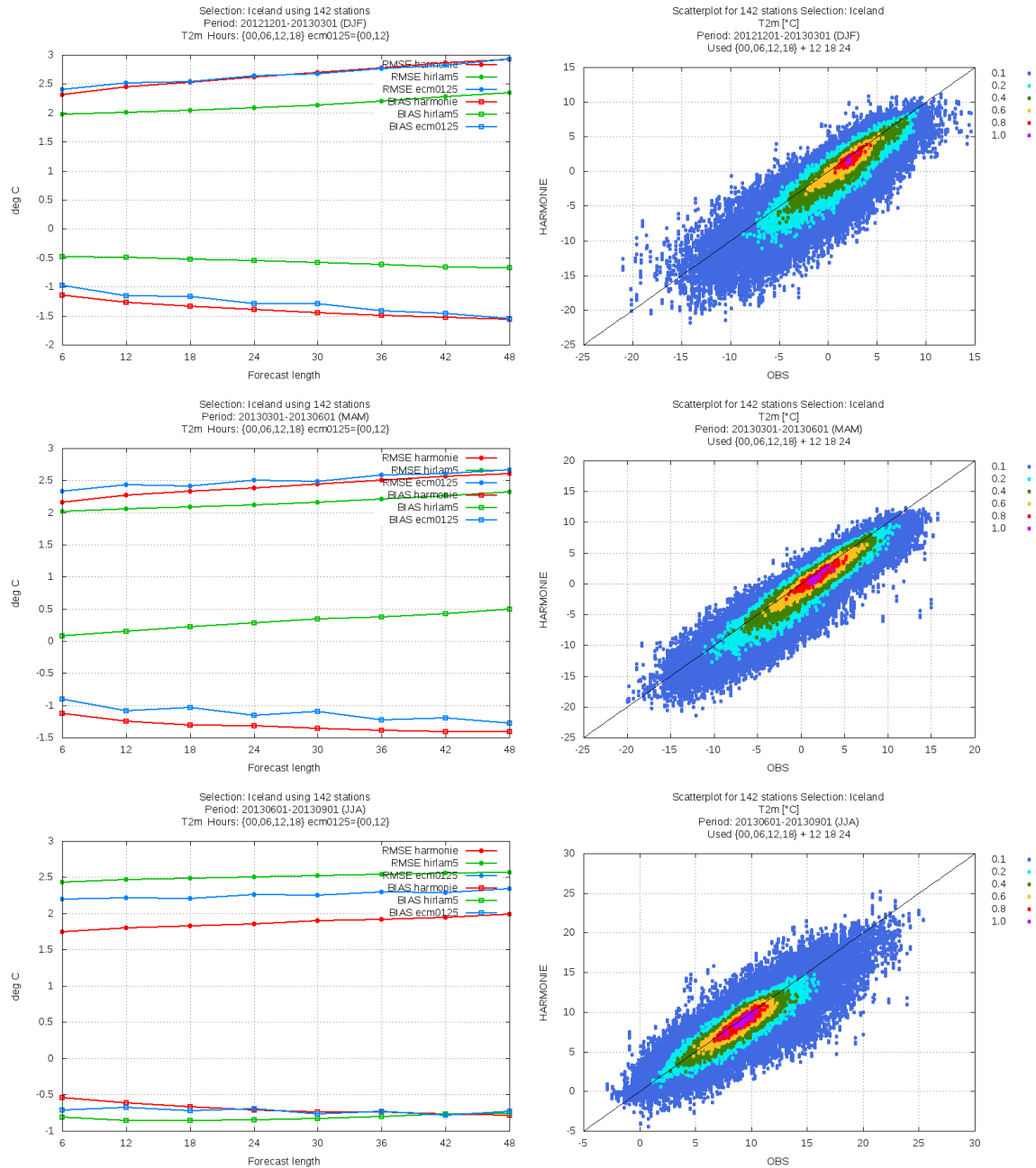


Figure 5: Left: A comparison of 2 m temperature RMSE/bias scores for ECMWF 0.125° (blue), HARMONIE 2.5 km (red) and DMI K05 Hirlam 5 km (green) for 142 stations for the periods DJF (top), MAM (centre) and JJA (bottom) 2012–2013. Right: Scatterplots of observed and forecasted by HARMONIE 2 m temperature for the same periods.

Figure 5 shows a comparison of error statistics for HARMONIE, ECMWF and Hirlam K05 as well as a comparison of observations and Harmonie forecasts. There was a negative bias of 1-1.5°C in winter temperature in the HARMONIE forecasts, similar to the one for the ECMWF forecasts. On the other hand the Hirlam forecasts performed clearly better during this period. This concurs with previous experiences from other Nordic countries. The situation improves slightly in spring of 2013 and drastically for the summer months. We see improvements in bias and RMSE for HARMONIE and it outperforms other models during the summer of 2013.

4.2 Large negative bias in 2 m temperature near inland lakes

Figure 6 shows a comparison of 48 hours forecasts at the stations Mývatn for HARMONIE, ECMWF at 0.125°, Hirlam at 3 km and 5 km horizontal resolution and MM5 at 3 km resolution to observed 2 m temperature during February 2013. The station Mývatn is located by the shore of Lake Mývatn and the figure shows clearly there is a much larger negative bias in the temperature in the HARMONIE forecast than for the other models. This bias is around -2.5 to -3°C in spring but decreases rapidly in June when the lake seems to warm drastically. From around 12 June 2013 the temperature at the station becomes realistic both in scale and variation. This problem of large negative bias is also seen at other stations close to inland lakes and seems to be directly related to the treatment of lakes in Surfex.

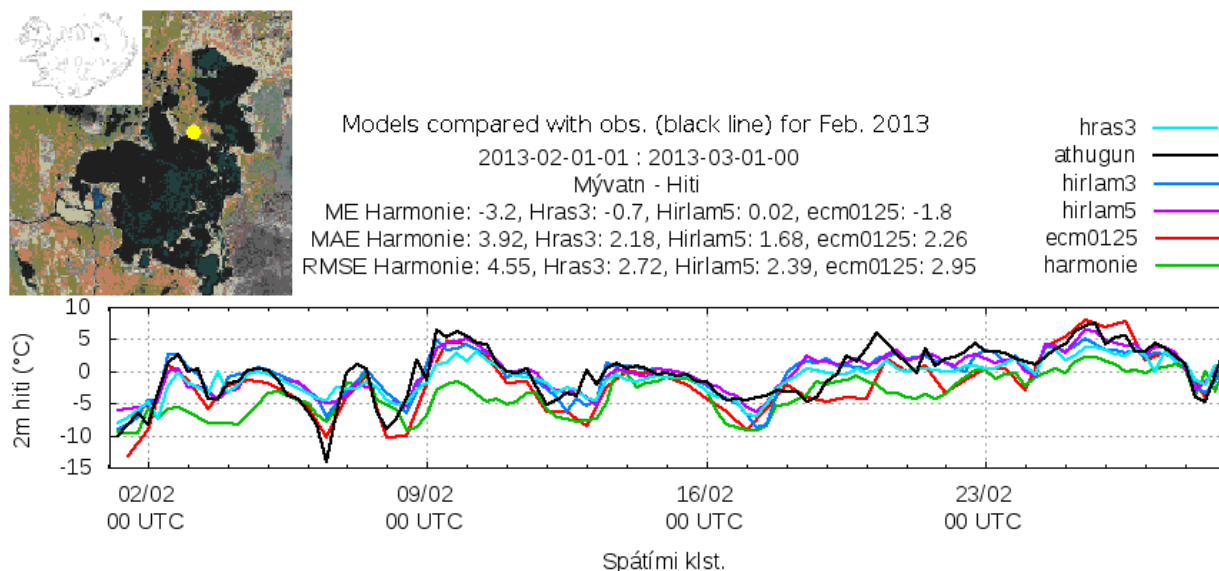


Figure 6: A comparison of T+48h forecasts for ECMWF 0.125° (red), HARMONIE 2.5 km (green), Hirlam 3 km (blue), Hirlam 5 km (purple) and MM5 3 km (cyan) to observations at the weather station Mývatn (see yellow dot on the small map) for February 2012.

4.3 Conclusions

Improving the bias in 2 m temperatures during winter months should be a priority for us and our neighbouring Nordic countries in the HIRLAM-B programme. The FLake lake model might be the solution to our problem with temperatures over lakes. FLake is included as an option in newer versions of Surfex and will soon be made available in HARMONIE.

5 Precipitation

Large scale precipitation and convective precipitation over land seems to be simulated adequately in Harmonie, but experience shows that marine convection is poorly simulated, see Fig. 7. Domain size might be a factor allowing more time for convective development. However reanalysis experiments at IMO with a slightly larger domain don't indicate increased convectivity. Furthermore, The Norwegian Meteorological Institute has also been experiencing the same problem with their much larger domain. It is possible that in our case the blending mode and a small domain might be the cause, but further investigation is needed to conclude anything about this. Currently we use EDMFm shallow convection parameterization in our setup.

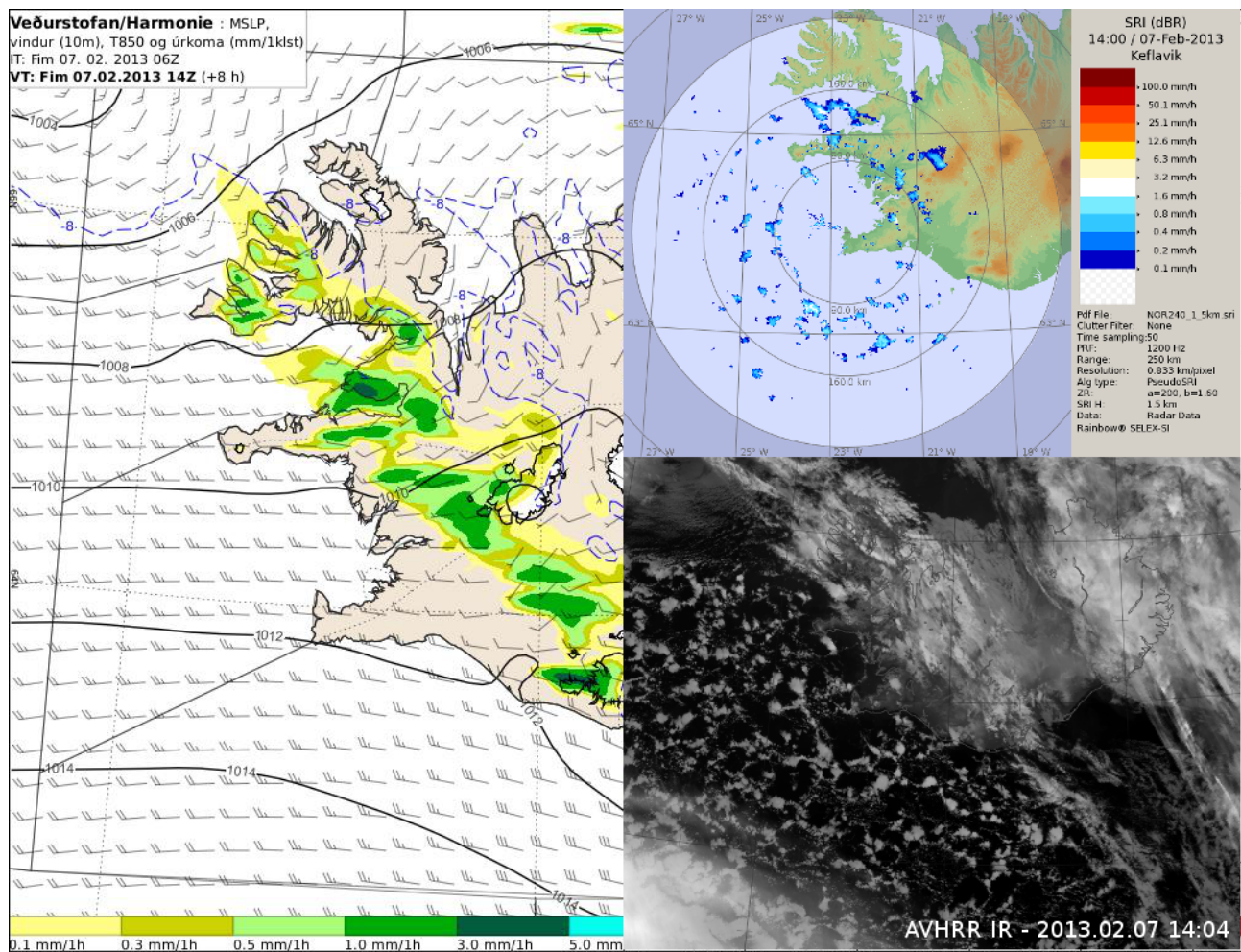


Figure 7: A forecast map (T+8h) for MSLP, T850 and total precipitation valid at 14:00 on the 7 February 2013 and images from the IMO radar and AVHRR (NOAA19 IR10.8) weather satellite from approximately the same time. The images show well how observed maritime convective precipitation is missed by the HARMONIE forecast. This problem has been seen repeatedly.

6 Snow accumulation

During the winter of 2012–2013 for the first time the IMO's pre-operational HARMONIE forecasts for Iceland accumulated snow. We used the simple one layer scheme in Surfex but all snow depth observations were excluded in the data assimilation. The results have been encouraging.

6.1 Snow extent verification with MODIS snow-maps

Visual comparison with satellite imagery has so far revealed that HARMONIE predicts the snow extent in Iceland quite well despite repeated periods of thawing during the winter. Here are presented some initial attempts to numerically quantify and monitor the accuracy of the HARMONIE snow extent by comparing with a satellite based reflectance snow mapping product recently developed by the IMO within the Snow, Ice and Avalanche Applications (SNAPS) programme (<http://www.snaps-project.eu>).

6.1.1 Verification methodology

As HARMONIE does not contain a fractional snow cover parameter, comparison of fractional snow cover observations with HARMONIE output are somewhat subjective. However, we can imagine that large enough snow water values in HARMONIE must in most conditions saturate to 100% snow cover. Based on this premise the methodology adopted was to test if relatively high HARMONIE snow water amounts coincide with 100% snow cover estimates from the MODIS snow cover observations. A fixed threshold on the snow water amount from HARMONIE was used and chosen to be 15 mm (kg/m^2), approximately 2–3 cm of snow. We note that the choice of threshold is fairly arbitrary, and that differences in ground 'smoothness', snow age and snow drift will bias our snow extent comparison.

For each day in the period from the beginning of February until the end of May, the IMO daily mosaic fractional snow cover product has been map projected to the HARMONIE-Iceland grid. Cloud-mask, land-mask, water proximity and permanent snow-cover masks were applied to both forecasts and the daily satellite snow map. Within these reduced observational windows, forecasts and observations were tested for agreement (snow or not snow). Over- and underestimates were then counted as fraction of total useable grid points. See Fig 8 depicting this work flow, and the resulting time series of snow extent comparisons in Fig. 9.

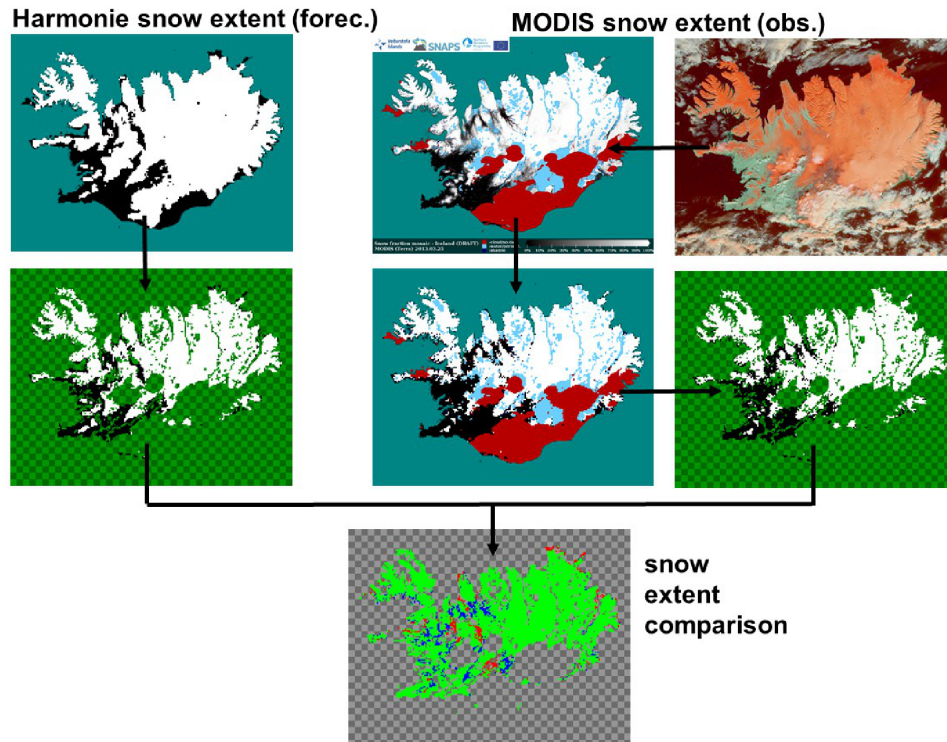


Figure 8: The processing chain for performing HARMONIE snow extent verification. The resulting comparison map indicates agreement between forecast and satellite observations, i.e. green grid points indicate where forecast and observations correspond, while blue suggest that forecast overestimates snow extent and red suggests underestimate in snow extent.

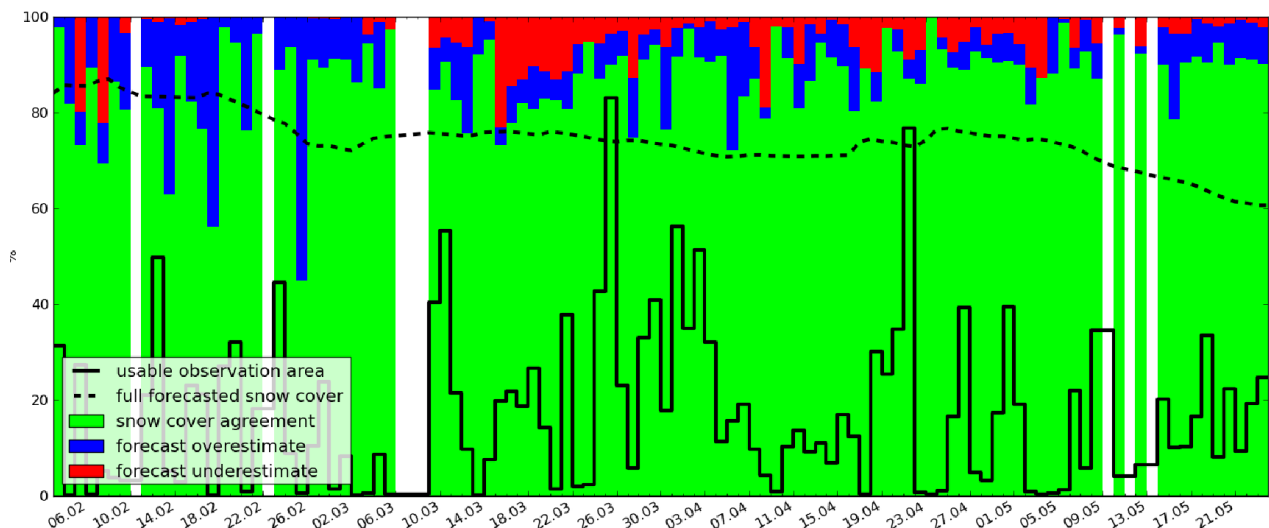


Figure 9: The snow extent comparison from the beginning of February until the end of May 2013. Comparisons were only performed in the available observational window. This means that estimates based on small observation windows are less representative of the total Iceland snow cover.

6.1.2 Conclusions

Visually and numerically, HARMONIE snow cover compares very well with satellite observations. The snow extent statistics used in this methodology indicate that HARMONIE snow extent accuracy is somewhere in the range of 80-90%. Furthermore, no clear long term bias trends are visible in the forecast, however, short period fluctuations in bias are visible, seemingly associated with snow melt. It is pleasantly surprising that the HARMONIE snow precipitation and snow melting model maintain such a close comparison in snow extent throughout the Icelandic winter. These results give some indication that HARMONIE is predicting precipitation quantities reasonably well. We hope that continued verification throughout the next winters will shed more light on these results. Also, snow extent combined with actual in situ snow water measurements will be important to gain a more complete picture of the snow forecast accuracies. Currently satellite based snow extent verification is also proving useful in verifying and calibrating assimilation of snow depth observations into the Iceland HARMONIE runs.

6.2 Verification with snow amount measurements on Hofsjökull glacier

Measurements of snow amount in water equivalence (SWE) on the Icelandic glaciers are done several times a year by glaciologists from IMO. This gives us an opportunity to verify the total precipitation and snow accumulation in HARMONIE for the period Sept–May. The Hofsjökull glacier is located in the centre of the Icelandic highlands and is approximately 1800 m high at the top (Fig. 10). Several measurements were made on 14 May 2013 at different places on the glacier. Those as well as values from the HARMONIE simulations are shown in Table 1.

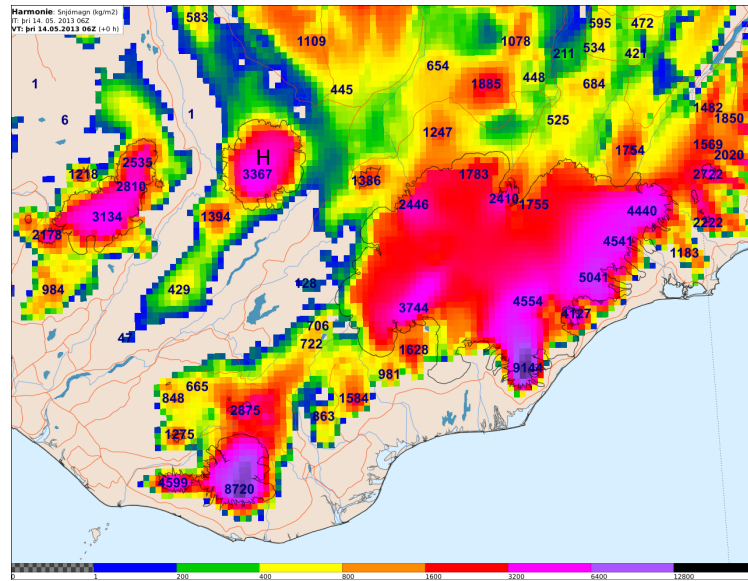


Figure 10: The amount of snow in HARMONIE accumulated field on 14 May 2013. Hofsjökull glacier is marked with an H and the numbers show the min/max values of SWE (kg/m^2).

Table 1: The measurements from Hofsjökull glacier on 14 May 2013 and SWE values from HARMONIE analysis for the same day. Credit goes to Þorsteinn Þorsteinsson and his team for the measurements.

Place (°N/°W)	Height (m)	Density (kg/m^3)	SWE (kg/m^2)	HARMONIE height (m)	HARMONIE SWE (kg/m^2)
64.770083 18.542850	872	410	990	899	780
64.838967 18.708483	1452	460	2230	1427	2530
64.722783 19.057883	1272	470	1680	1163	1350
64.967617 18.863967	981	510	650	1021	700
64.892850 18.858783	1344	440	1120	1412	2050
64.808983 18.869583	1792	460	2550	1661	3070

6.2.1 Conclusions

HARMONIE is overestimating the total snow amount by approximately 14%. Although we believe this to be promising results there is need for multi-year data for any conclusion to be drawn from the comparison. We recently began a re-analysis project with HARMONIE and should soon we have several years of data to compare with the annual measurements made.

7 Summary

The HARMONIE NWP model has been operated by IMO with surface data assimilation since September 2012. The experiences during this first year show promising results, however improvements are needed in several key parameters. We've already made some progress with near surface wind speed and our current efforts are focussed on improving the underlying surface in Surfex, find ways to improve 2 m temperatures during winter months and analyse the problem with maritime convective precipitation. We believe that improvements can be made with proper calibration/setup for Iceland.

References

- Beljaars, A. C. M., A. R. Brown, N. Wood, A new parameterization of turbulent orographic form drag, Quart. J. Roy. Met. Soc., 130, 1327—1347, 2004
- Seity, Y., C. Lac, V. Masson: About orographic drag options in SURFEX. Tech. Report.

Can HARMONIE be accelerated with GPUs or coprocessors?

Enda O'Brien, Adam Ralph

1 Introduction

One of the characteristics of numerical weather prediction is that it has an insatiable appetite for better numerical performance. This desire must come to terms with the reality that, in the famous words of Herb Sutter, “The free lunch is over” [1]. In other words, individual CPU cores are not getting any faster, so the only possible way to improve performance now is through the exploitation of parallelism.

Thus conventional high-performance computing (HPC) processors (now dominated by x86_64 architecture) have evolved from single-core CPUs, which were prevalent until 2003, into today's multi-core processors with 12 or more processing “cores” per chip.

Some time around 2007 a completely new class of parallel computing hardware became available when Nvidia and ATI provided a programming interface to their graphics processing units (GPUs). Along with some modifications to the GPUs themselves, this transformed graphics cards from closed “black boxes” into “General Purpose Graphics Processing Units”, or GP-GPUs. Programmers were quick to take advantage of the massive hardware parallelism offered by GPUs and soon word began to spread of applications ported to GPUs achieving performance speedup factors of 10 or even 100 over the standard host CPU. Many such amazing performance gains are documented on Nvidia's CUDA web-site (www.nvidia.com/cuda). A closer evaluation of GPU performance (i.e., relative to a modern host node rather than an old PC) reveals that more realistic speedup factors are typically in the range 2-10 – more modest perhaps, but still impressive, and worth pursuing!

In February 2013 Intel released the Xeon Phi coprocessor, which provides an alternative to GPUs as a massively parallel hardware accelerator. As an architecture, the Phi coprocessor is in a class known as “many integrated core”, or “MIC”. It does not offer as much parallelism as a GPU, but its 61 cores (each providing 4 threads) have standard x86 architecture and so are more versatile than GPU cores, and can be programmed with “standard” software tools. On the other hand each individual core in the Phi coprocessor is slower and weaker than the dozen or so cores in the conventional “multi-core” Xeon processor. In comparison with GPU accelerators, the Phi coprocessor is somewhat less parallel but arguably easier to program.

HARMONIE is already well adapted to run on multi-core processors. This article takes the first steps to address the questions of: how can HARMONIE be adapted to run on GPU or MIC accelerators? And what (if any) performance gains can be expected for HARMONIE from such devices?

2 GP-GPUs

Figure 1 shows a schematic of a system comprising a “host” node (with standard X86_64 processors) and a GPU accelerator (in this case, a Fermi GPU from Nvidia). The complete system may be thought of as having just four main parts. These are the host CPUs (“X86 host”) and “host memory” on the left side, and the array of GPU cores and the main GPU or “device” memory on the right side. In the early days of GPU-programming, programmers needed to know some details about the GPU cores (such as how many there were, and how they were arranged). Such knowledge is still useful but thankfully no longer necessary.

In principle, standard Fortran or C source code that compiles and runs on the host can also be compiled to run on the device (i.e., the GPU). The adaptation of source code instructions to run on new architecture is primarily a job for compilers, not application programmers.

The main concern for programmers is the fact that host and device memories are physically distinct (as in Fig. 1). For each executable line of source code, the programmer must ask, if this is running on the host, is the data it needs available on the host memory? If running on the device, is the data it needs available on the device memory? As computational work is “offloaded” to the device, the input data for that work must first be moved to the device memory, and any output from that work must be moved back to host memory afterwards. These data movements are both tricky (e.g., pointers themselves should not be copied between host and device, only the data they point to) and slow (since data transfer is over the “DMA” or “direct memory access” link in Fig. 1, which uses the PCIe bus). Control of data movement is primarily the responsibility of the programmer, not the compiler. Very similar considerations apply to MPI programming; the main difference being that while MPI tasks are usually “symmetric” (e.g., peers of each other in the SIMD sense), the relationship between CPU host and GPU device is highly asymmetric.

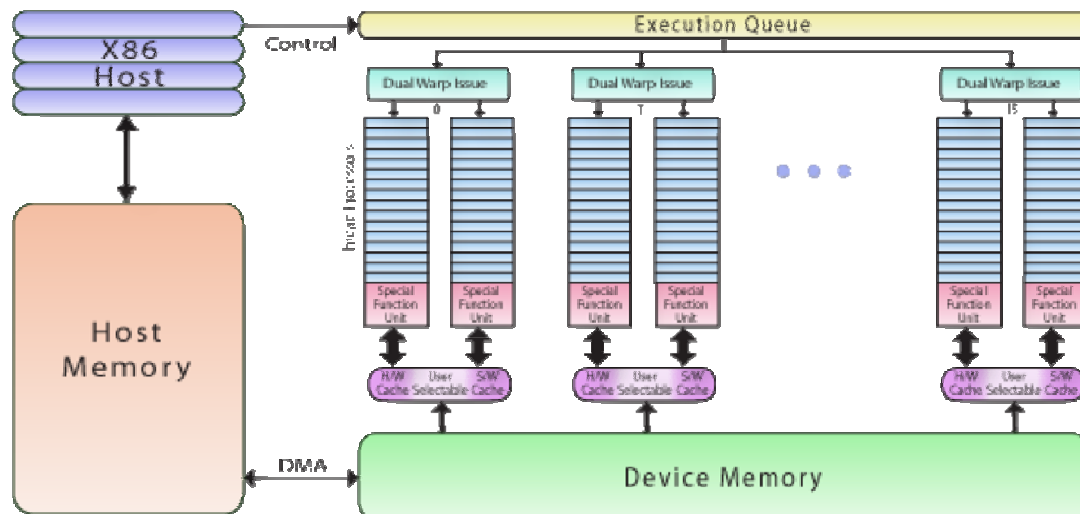


Figure 1 Schematic of Host System (left side) with GPU “Fermi” accelerator device attached (right side).

Given that data movement between the host and device over the PCIe bus is relatively slow, one basic principle of programming for GPUs is to minimize such traffic. In the limit, the ideal is to run the entire application on the device, perhaps only moving output data back to the host for final storage. In the more usual situation where only parts of the code can be run on the device, some data transfer between host and device is unavoidable (e.g., at every time step during a forecast run). One option that can then be used to avoid the transfer becoming a performance bottleneck is to make the transfers *asynchronous*. Asynchronous transfers, once started, are only required to complete by the time some specified point is reached further along in the algorithm. In this way, either (or both) the host and device can continue calculating while the transfer is taking place. Synchronous transfers, on the other hand, must start and complete before the code can move on to the next instruction.

Despite a programmer’s best efforts, it is a very common experience when running on GPUs that the performance gained by GPU acceleration is lost again during data transfer. The code sections running on the GPU may finish a lot faster, but overall run-time may decrease by very little, or even increase because of time lost moving data between the host and device.

Apart from data transfer considerations, it must be remembered that GPUs are designed primarily for displaying streaming frames of images whose pixels are usually independent of each other and can be processed in parallel. The more closely the flow of data through a program can mimic the flow of digital images, the better that program will perform on a GPU. In practise this means that routines with uniform stride through large “do” loops and a lot of calculations during each trip through the loop will perform well on GPUs, while routines with a lot of branches (“if” conditions) or short loops with little computational work will perform poorly.

3 Weather Models on GPUs

Some NWP or related models have already been ported to GPUs.

One very impressive example is the ASUCA “next generation” non-hydrostatic weather prediction model from the Japanese Meteorological Agency. Prof. Takayuki Aoki translated the entire model into CUDA (the extension to the C language provided by Nvidia for programming their GPUs), and so the model runs completely on GPUs. A single GPU runs this model 80 times faster than a single CPU core, while a sustained performance of 15 TFlops in single precision was achieved using 528 Nvidia GT200 Tesla GPUs, over a computational domain with 6,956 x 6,052 x 48 grid-points.

The dynamical core of the Non-hydrostatic Icosahedral Model (NIM) was ported to GPUs by Tom Henderson and his group in NOAA. They first used a “Fortran to C” translator, and then used HMPP directives from CAPS (analogous to OpenMP directives) to control data and code offloads to the GPU. They found that 2 “Kepler” K10 GPUs performed approximately twice as fast as a dual-processor (16-core) Sandybridge node. That speedup was achieved despite the model reaching only 2% of the 2.2 TFlop peak performance on the K10s.

The COSMO model from DWD has been ported to GPUs by Oliver Fuhrer and his team at CSCS and Meteo Swiss. The dynamics parts of the code were translated to CUDA, while the physics parts used OpenACC directives (again, closely analogous to OpenMP directives). Input data is transferred to the GPU at the start of each model run, and output is transferred back to the host for storage, but otherwise the entire code runs on the GPU.

4 Acceleration Options for HARMONIE

As evidenced by work on the models described above, there are two main options for porting a code like HARMONIE to GPUs: one is to translate it to C, and thence to CUDA (or OpenCL); the other is to insert accelerator directives into the source, similar to the OpenMP directives already there. There is now a standard for accelerator directives called OpenACC (see www.openacc.org for details).

The main appeal of CUDA is that it is the most developed and most powerful programming interface to GPUs; e.g., it supports features such as direct GPU-to-GPU communication that are not possible in any other way. While CUDA is not a formal standard, it is the de-facto programming standard for GPUs, with a large developer community and a wide range of CUDA-enable applications.

A manual translation of HARMONIE to C is certainly impractical, but automatic “Fortran to C” translators are available that might do the job. Indeed, PGI’s implementation of CUDA-Fortran invokes such a translator behind the scenes. However, translation to CUDA (whether from C, as is most common, or from Fortran, as is currently possible only with PGI) raises some major issues: Translation to CUDA is a lot of work that presumably would need to be re-done – or at least re-visited – for each new cycle of HARMONIE.

A C or CUDA version of HARMONIE would probably be largely unrecognizable to most HARMONIE developers and users – it could become a “black box”.

CUDA is proprietary software, owned by Nvidia, and is not a “standard” in the way that Fortran, OpenMP and MPI are industry-wide standards. An emerging standard called OpenCL is very closely related to CUDA and could possibly be used to overcome this objection.

The pros and cons of a directives-based approach to accelerating HARMONIE are almost the mirror opposites to those of CUDA. The main drawback to using directives is that they are more restrictive than CUDA and do not support the full CUDA feature-set. Moreover, currently only 3 compilers support such directives, namely: PGI, CAPS/HMPP, and Cray.

On the plus side:

HARMONIE already contains OpenMP directives (at a very high-level!), so the addition of another set of directives should not be too disruptive. Directives are treated as comments unless specifically invoked with a compiler option, so they can easily be either ignored or implemented, as desired, at compile-time.

Directives can be added incrementally, one loop or one subroutine at a time. So it should be possible to make, and evaluate, initial progress relatively quickly.

OpenACC directives are emerging as the standard for GPUs, and in future may well merge with OpenMP into a single standard directives package. Intel has its own set of directives to support its Xeon Phi coprocessor, and eventually these too may be absorbed into a single standard with OpenACC and OpenMP.

5 Building HARMONIE with OpenACC directive from PGI

5.1 Basic PGI Build

Given the considerations above, the most natural approach to accelerating HARMONIE was to use OpenACC directives. High profile and naturally parallel routines such as LAITRI or ACRANEB can easily be instrumented with simple OpenACC directives to offload the main loops (and whatever data they need) to any attached GPU.

As mentioned above, only 3 compilers currently support OpenACC directives in Fortran source code. These are HMPP from CAPS, PGI, and Cray. In practise, Cray is only an option for centres with Cray hardware. Initial tests with HMPP (in early 2012) revealed that its OpenACCFortran support at the time only extended to Fortran-77 syntax, which made it impractical for use with HARMONIE. (Later versions of HMPP do have OpenACC support for many features of the current Fortran standard, but we have not tested these). PGI, on the other hand, had no problem compiling OpenACC directives in Fortran-95 code, so we chose this one for our effort to port at least part of HARMONIE to GPUs.

So far, we have been unable to get even the basic build of HARMONIE for conventional host processors to work properly with PGI compilers – at least, when using AROME physics (runs with ALARO physics completed smoothly). This is true for both HARMONIE cycle 37h1.1 and the “benchmark” distribution based on cycle 38 (h1.alpha.2). Despite much debugging, there just seem to be too many incompatibilities between HARMONIE and the basic PGI compiler.

One characteristic of some HARMONIE code that is definitely incompatible with PGI is the use of “unassociated pointers”. These can appear as arrays that are declared as “allocatable”, but never actually allocated (because of some logical switches). Nevertheless they are passed to subroutines where they are declared as if they had a finite size. This is the case with the RUCONVCA and RNLCONVCA arrays, which are declared as “allocatable” in yoe_cuconvca.F90, but only conditionally allocated. For HARMONIE built with PGI, it was necessary to make that allocation unconditional to prevent segmentation faults associated with them.

One helpful error message from PGI was:

Null pointer for sptndsi_vor (etransinvh.F90: 123)

The solution in this case was to make the call to SUALDN_DDH in sy0yomb.F90 unconditional instead of conditional. The SUALDYN_DDH routine allocates memory to over a dozen arrays, which are already declared as “allocatable”. Those arrays in turn are passed down to various other subroutines, where they may take on different names and may even be assumed to have finite size. With the original code, the arrays may not be allocated, in which case they cause failure with the HARMONIE PGI-build if they are “touched” in any way during a run.

Another coding practise in HARMONIE (especially in SURFEX) that causes trouble with the PGI-build is the use of the SIZE() function in declarations. The resolution of the SIZE function depends on the rest of the declarations block being fully analysed and not on the logical flow of executable statements. This appears to be something of a grey area where compilers may differ in their practice - and clearly they do! E.g., in coupling_surf_atmn.F90, the following declarations are made:

```
REAL, DIMENSION(KI), INTENT(IN) :: PTA    ! air temperature forcing (K)
...
REAL, DIMENSION(SIZE(PTA))    :: ZPEW_A_COEF ! implicit coefficients
```

Changing the declaration of ZPEW_A_COEF (and other variables declared similarly) from SIZE(PTA) to simply KI allowed the run to at least proceed past this point of failure.

Other source-code changes were also necessary, not because they were wrong, but simply to satisfy the PGI compiler – e.g., in some cases PGI was not able to “USE” a complete module, so the list of “USED” variables had to be spelled out explicitly.

With these and some other source-code changes, a HARMONIE executable built with PGI was eventually able to run to completion. However, this would only work with a single MPI task and no DR_HOOK, so some fundamental problems remain unsolved and this is not a satisfactory situation.

5.2 Build with OpenACC directives

Once a working HARMONIE was built, it was almost trivially simple to insert some OpenACC directives and to demonstrate that the concept worked. A code section, or “kernel”, suitable for offload to a GPU can be identified with a “!\$acc region” directive at the beginning of the section and a closing “!\$acc end region” directive at the end. Further clauses can be added to the basic directives to identify variables that need to be transferred in one direction or another between the host and the device, or that don’t need to be transferred at all (i.e., are purely local to the device).

Simple directives like this were added to the main loop in laitri.F90, and also to some loops in rtm_rtrn1a_140gp.F90.

Table 1 DR_HOOK output showing performance of the main HARMONIE routines on CPU & GPU

#	% Time (self)	Cumulated (sec)	Self (sec)	Self (1-CPU, CPU+GPU)	Total	# calls	Routine
1	15.24	108.87	108.87	(33.5, 10.1)	108.89	59475	LAITRI
2	5.17	145.80	36.94		371.81	3965	APL_AROME
3	4.73	179.59	33.78		33.80	63440	TRIDIAG_MASSFLUX
4	3.80	206.77	27.18	(12.2, 3.7)	27.18	325	RRTM_RTRN1A_140GP

Table 1 shows performance of the main HARMONIE routines with very small 50 x 50 x 60 point problem size on a single CPU-core and a single GPU. The values in red (10.1s for LAITRI and 3.7s for RRTM_RTRN1A_140GP) are run-times for the sections of those routines offloaded to the GPU. The GPU achieved a speedup factor of about 3 in both cases. That speedup includes time taken for data transfer, and is quite encouraging, given that this problem is so small.

6 Experiments with the Intel Xeon Phi coprocessor

Some interesting recent results were obtained on a small test system consisting of a 4-core Intel SandyBridge server with an Intel 5110P Xeon Phi coprocessor attached. The Phi coprocessor is effectively an alternative to GP-GPUs for the purposes of code “acceleration”. From a user perspective, the Phi had 240 “logical” cores (consisting of 60 physical cores, each of which appeared as 4 logical cores or “hyper-threads”), and 8 GB of memory. Each Phi core is (or at least emulates) an x86 processor, and all Intel’s main software development tools (compilers, MPI and MKL libraries) support Phi, and applications can run on it much as on any other x86-based platform.

Applications can use the Phi either in “native” mode (i.e., running entirely and directly on the Phi), or in “offload” mode (where jobs are based on the host server and work is offloaded to the Phi as determined by user-programmed directives, much as OpenACC directives offload work to GPUs).

6.1 Kernel test results on Phi

Figure 2 shows performance of a simple “kernel” test code, which does repeat smoothing of an arbitrarily-sized 3-dimensional array using a 7-point stencil, and is parallelised with OpenMP. It is similar to HARMONIE’s LAITRI subroutine. The left panel shows results from the 1GB problem size; the right panel shows results from the 6GB problem size. The blue diamonds on the left of both charts show the performance of the OpenMP parallel version on the host node, and serve as a reference for the other results. Note that scalability on the host is quite poor, even for the 6GB case.

The green symbols on the 1GB chart shows results from the system when a “naïve” offload method is used, in which the input 3d array is transferred from host to device for each smoothing iteration, and the output array is moved back again afterwards. Performance is very slow, and largely independent of the number of coprocessor threads used, because total run-time is dominated by the data transfer time, which itself is independent of the number of coprocessor threads. There is no analogous result for the 6GB case: one test was enough to demonstrate that transferring 6GB of data is even slower than transferring 1GB, so there was no point in repeating that for different coprocessor thread-counts.

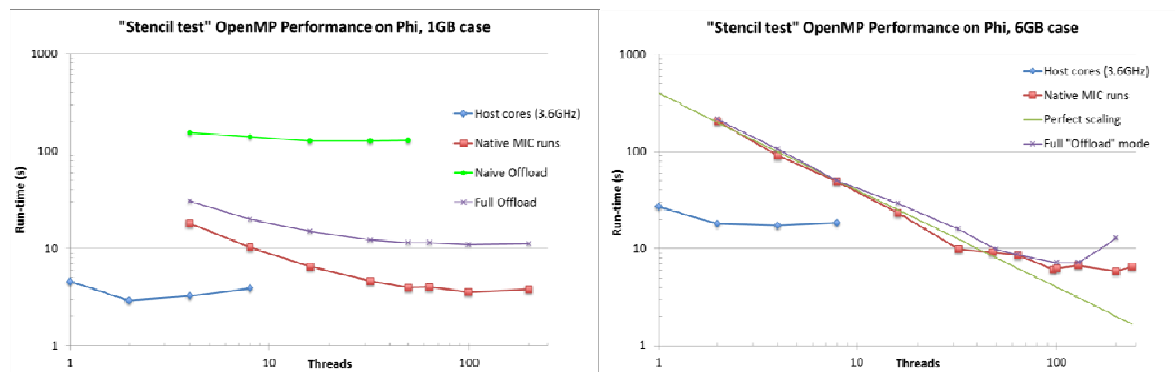


Figure 2 Performance of Stencil kernel test on Phi coprocessor and on host node.

The purple symbols in Fig. 2 show performance of the test in “full offload” mode, where the output from each smoothing iteration was left on the coprocessor, and used as input for the next iteration. In other words, no data transfer was required between host and coprocessor except before the first iteration and after the last. Not surprisingly, performance is much better than from the “naïve offload” case, but still not as good as on the host node for the smaller 1GB test case, at least. For the 6GB problem size, the “full offload” tests are faster than on the host node when 32 or more coprocessor threads are used.

The red symbols in Fig. 2 show performance when the entire application is run “natively” on the coprocessor. No directives are used in this case at all; the entire code is simply compiled and linked with the “-mmic” flag, and launched directly on the coprocessor. This is the fastest way to run the test

on the coprocessor, since no data transfer is required at all, and no coordination is needed with the host node.

Taken as a whole, Fig. 2 serves as a useful representative illustration of the strengths and weaknesses of the Phi coprocessor. Its main strength is that even jobs which scale poorly with OpenMP on a host node, can scale up to 200 or more threads on the coprocessor – subject to the proviso that the problem size is large enough (or does enough work), but not so large that it exceeds the 8GB memory available on the device. Indeed, jobs must scale to at least 20 or so threads on the Phi for any gain to be possible at all. These results also reinforce yet again the main lesson from GPUs, that data movement must be minimized in order to preserve any performance gain achieved by the “accelerator” device.

As a loose analogy, the Phi coprocessor is like the 8 pawns each player has in a game of chess, while the host node is like one knight. That knight is more powerful than any pawn, and more valuable than even 3 or 4 of them. But if deployed well, 8 pawns together can be more valuable than a knight.

6.2 HARMONIE results on the Phi coprocessor

As a first cut, the HARMONIE cycle 38 “benchmark” version was built with the “-mmic” compiler option and run “natively” on the Phi coprocessor (equivalent to the red symbols for the kernel test in Fig. 2). The auxiliary zlib, hdf5, and netcdf libraries also had to be built with “-mmic”, since absolutely every piece of software that runs on the Phi coprocessor must be compiled for that architecture.

While HARMONIE built for the host node ran smoothly (as it does all around Europe...), the first executables built for the Phi (i.e., with “-mmic” option) failed with error messages about arrays “out of bounds”. Two source code changes were needed to remove these: one was to change NUNDEFLD from -99999999 to 1 in sudim1.F90; the other was change lines 339-340 in suspstd.F90 to:

```
ALLOCATE ( IGRIB ( 1 ) )
```

```
CALL ALLOCATE_GRID( YGPSDT, 0, 1, igrib )
```

With those two changes, HARMONIE was able to run smoothly to completion on the Phi.

Table 2 shows results obtained so far from running HARMONIE on the host node and natively on the Phi coprocessor. Only the “extra small” test case supplied with the benchmark kit was used, since that was the only one that would fit in Phi’s 8GB memory. Results from the host node are in blue, and from the Phi in red, for different combinations of MPI tasks and OpenMP threads. The host results (in blue) scale reasonably well and are all plausibly self-consistent as the MPI task count and OpenMP thread count vary among the 8 logical cores on the host.

The results on the Phi are entirely plausible too, but they are very preliminary and are at least some of them are certainly not optimal. Firstly, they are all much slower than the equivalent run on the host – but as is clear from the right chart in Fig.2, that is most likely a function of the very low thread count used so far on the Phi. Moreover, this “extra-small” problem size, with just 50x50 horizontal points, is unlikely to run on more than a dozen or so threads simply because it has too few grid-points.

Notice that the 2x2 by 2-threads run with Phi reports two numbers (3403s and 545s). The first of these is the total run-time, while the second is the time to complete the 360-steps of the 6-hr forecast. In other words, the run was almost complete but took over 45 minutes just to finalize and stop. A similar phenomenon happened with the 3x4 by 1-thread run (2235s for the complete run, but only 379s to reach the end of the 360 time-steps). I suspect that this extreme slowness at the end is related to the KMP_STACKSIZE setting used for these runs. The stencil kernel tests exhibited the same phenomenon when KMP_STACKSIZE was “too big” (~4GB). Of course, the jobs fail to run at all if KMP_STACKSIZE is too small. So another lesson from running on the Phi coprocessor is that there is a “Goldilocks” range for KMP_STACKSIZE (and possibly other settings too); not too small, not too large, but “just right”.

Table 2: HARMONIE run-times (s) on host and (natively on Phi)

	Threads			
MPI-Tasks (NX x NY)	1	2	4	8
1	342(1634)	186(850)		98
1x2			104	
2x2		103.5 (3403; 545)		
2x4	102			
3x4	(2235; 379)			

7 Conclusions

GPUs and the Xeon Phi coprocessor are powerful and appealing performance tools, but releasing their performance potential with a complex (and especially a Fortran) code like HARMONIE will be a long and difficult project.

At least our work so far demonstrates that HARMONIE runs in “native” mode on the Phi coprocessor, and that some of the busiest routines can be offloaded to a GPU using relatively simple directives.

HARMONIE uses many Fortran features that are not currently supported by any “offload” compiler (e.g., derived data types containing pointers), and given the nature of accelerators with their separate memories, HARMONIE itself will probably need to be modified in order to work around those restrictions. Indeed, as we have experienced, HARMONIE contains some features that are difficult for even standard “host” compilers to handle!

Meanwhile, compilers and compiler directives are likely to evolve quickly themselves too, and the various flavours of offload directives may well converge to an expanded “OpenMP” standard.

We can be reasonably certain on one point, however, which is that acceleration of HARMONIE is unlikely to be accomplished by any “automatic” tool. Some human contribution will always be needed!

Reference

[1] Herb Sutter: “A Fundamental Turn Toward Concurrency in Software”. Dr.DobbsJournal, 30(3), March 2005. (*Available online*).

FROST-2014 – First experiences of deterministic Harmonie in Sochi region

Sami Niemelä, Evgeny Atlaskin, Sigbritt Näsman and Pertti Nurmi,
Finnish Meteorological Institute

1 Introduction

The next winter Olympic games will be held in Sochi, Russia, on 8-23 February 2014. The World Meteorological Organization (WMO) has organized both Research Development Projects (RDP) and Forecast Demonstration Projects (FDP) during the previous Olympics. The purpose of these projects has been to advance and demonstrate the status and benefits of the state-of-the-art forecasting and nowcasting systems in a highly demanding operational environment. The FROST-2014 (Forecast and Research: the Olympic Sochi Testbed) project will focus on the development/demonstration of modern short-term Numerical Weather Prediction (NWP) and nowcasting systems over mountainous region in winter-time conditions and on the assessment of practical use of this information.

FMI will contribute to the FROST-2014 FDP by running the high resolution mesoscale NWP-model Harmonie over Sochi region, on daily basis, before and during the Olympic games. The purpose of this study is to build and test the first experimental version of the Harmonie covering the highly mountainous areas of the eastern coast of the Black Sea. The numerical experiments will focus on three aspects: (i) the comparison between the mesoscale baseline model and operational global/regional models, (ii) the effect of an alternative orography dataset and (iii) the effect of the model grid size over Sochi region.

2 Characteristics of Sochi area

Figure 1 shows the venues and main characteristics of Sochi Olympic area. The Sochi area is located on the eastern coast of the Black Sea (43.4° N, 40.0° E). The events will take place in two specific areas, in the coastal (Adler) and mountain clusters. The coastal cluster hosts indoor sports, whereas all the weather dependent events, such as ski jumping, cross-country and alpine skiing, will take place in mountain cluster. The distance between the mountain cluster and the coast of Black sea is about 40 km. There is only one road going from the Adler to the mountain cluster along the valley. The valley itself is about 3-6 km wide and is surrounded by high mountains peaking approximately 2000-2500 m above sea surface level.

The main village in the mountain cluster is the called Krasnaya Polyana, which is located about 550 m above the sea surface. Krasnaya Polyana has an official WMO Synop weather station providing basic near surface weather observations. The sport events are organized in close vicinity in the mountains. The ski stadium, alpine skiing and ski jumping venues are at the heights of 1400 m, 2200 m and 600 m above sea surface, respectively.

The southern location, the vicinity of the sea and the high mountains pose significant challenges to weather prediction in the region of the Olympic games. Both numerical weather forecasts in FROST-2014 FDP and RDP and actions of duty forecasters will be under investigation during the games. These difficult conditions will certainly deepen our understanding of the performance of NWP-models, which in turn will help in the actual model development process.



Figure 1: Characteristics of the Sochi Olympic area

3 Model configuration

Harmonie is a non-hydrostatic model based on fully compressible Euler equations. The baseline mesoscale system over Sochi is based on Harmonie cy37h1.2 with 2.5 km horizontal grid size and with 65 levels in vertical (soc37h12). Figure 2a shows the domain of the baseline setup (640x500 points). This configuration uses Arome-physics package (Seity et al. 2011). The time step is 60s. Upper air data assimilation is handled by 3D-Var, whereas optimum interpolation is used for surface variables. Background error statistics were created for this domain by using an ensemble method. One day (+24h) forecast was made twice a day with 6-hourly assimilation interval including conventional observations only. Hourly boundary conditions are taken from the ECMWF global model. The only difference between baseline experiment and the reference Harmonie configuration is the use of orography dataset from the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007).

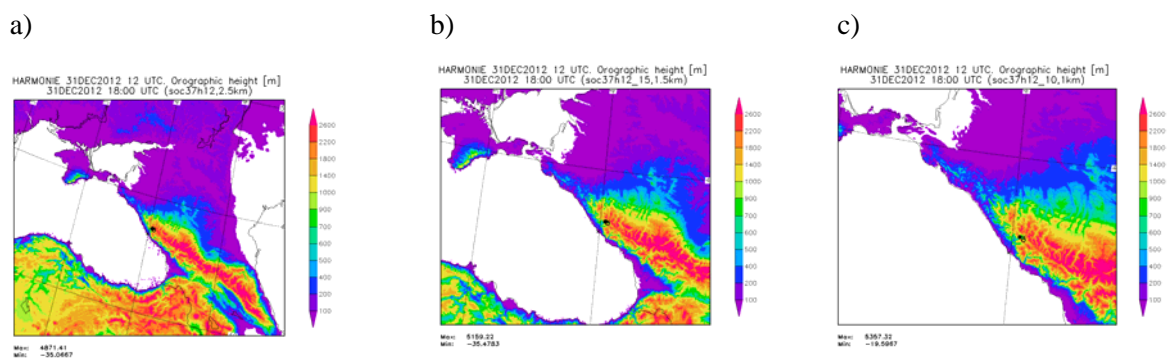


Figure 2: Domains of the different Harmonie configurations with a) 2.5km, b) 1.5km and c) 1.0 km horizontal grid size. All domains have 640x500 points. The Olympic venue is marked as a dot.

The second experiment (soc37h12old) is exactly like the baseline with one exception. The orographic dataset is the GTOPO30, which is used by default in Harmonie. Furthermore, the third (soc37h12_15) and fourth (soc37h12_10) experiments differ from the baseline by horizontal grid size (1.5 km and 1.0 km) and domain size (Figs. 2b and 2c). The time step for these experiments were 40s and 30s, respectively. Data assimilation was used in both high resolution experiments, however, the background error statistics were determined by interpolating the baseline statistics.

In addition, the mesoscale baseline model is compared to operational regional and global NWP-systems. The regional model is the FMI's operational HIRLAM-model (version 7.4) with a 7.5 km grid size, whereas the global model is the ECMWF-IFS with a 16 km grid size.

4 Results

The Harmonie configurations were run over different periods during three months between January and March 2013. However, the data from all model experiments exist only for January 2013. Therefore, the comparison presented in this paper covers that period only. From geographical point of view, the comparison focuses within about 300 km radius around Sochi region. The observations are taken from the WMO Synop weather station network, which includes 12 stations from the area of the interest.

Baseline vs. operational models

Figure 3 shows the bias and RMS-errors of 2m-temperature and 10m wind speed between baseline Harmonie (soc37h12), HIRLAM and IFS models. Harmonie with 2.5 km outperforms the operational models in terms of RMSE (temperature) and bias (both temperature and wind speed). Both operational models tend to overestimate the wind speed much more than Harmonie. This is a rather expected result indicating the benefit of high resolution, since mountain peaks and valleys are much better described with 2.5 km grid size than 7.5 or 16 km. Furthermore, Fig. 3 shows that the baseline mesoscale experiment works in reasonable manner to be used as a basis for further studies.

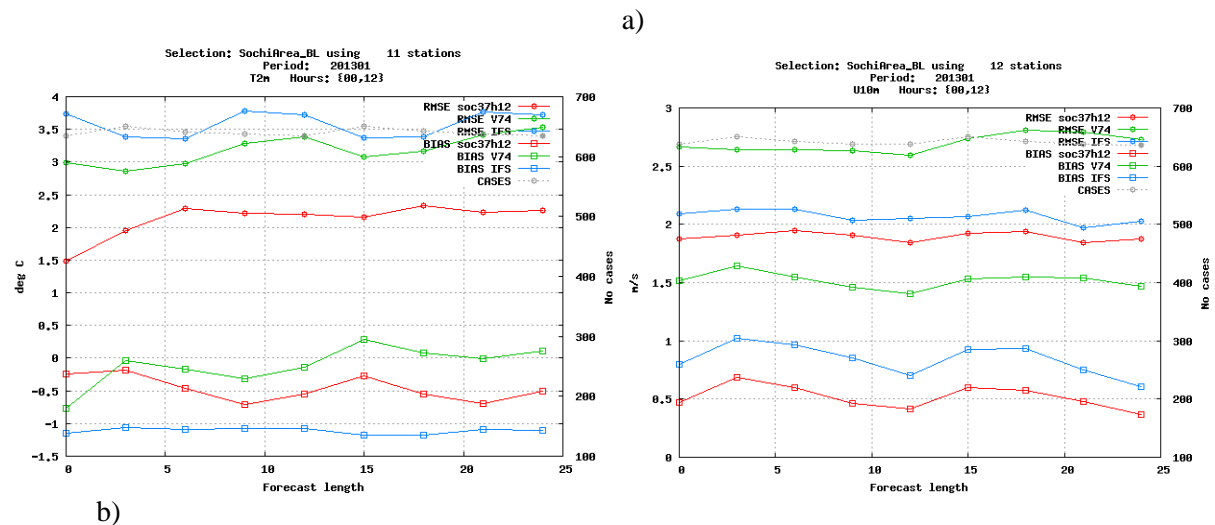


Figure 3: Bias (square) and RMSE (circle) of a) 2m-temperature and b) 10m wind speed as a function of forecast length for January 2013. Red is Harmonie baseline experiment (2.5 km), green is HIRLAM (7.5 km) and blue is ECMWF-IFS (16 km).

The effect of orography dataset

Figure 4 shows the bias and RMS-errors of 2m-temperature and 10m wind speed of Harmonie experiments with SRTM (baseline, soc37h12) and GTOPO30 (soc37h12_old) orographic datasets. The resolution of the SRTM and GTOPO30 datasets are about 100 m and 1 km, respectively. The impact of new SRTM data set is more or less neutral for both temperature and wind speed. The same result is valid for all other parameters as well (not shown). Since the new dataset is not degrading the results, it is encouraged and justified to use SRTM data with smaller grid size than 2.5km.

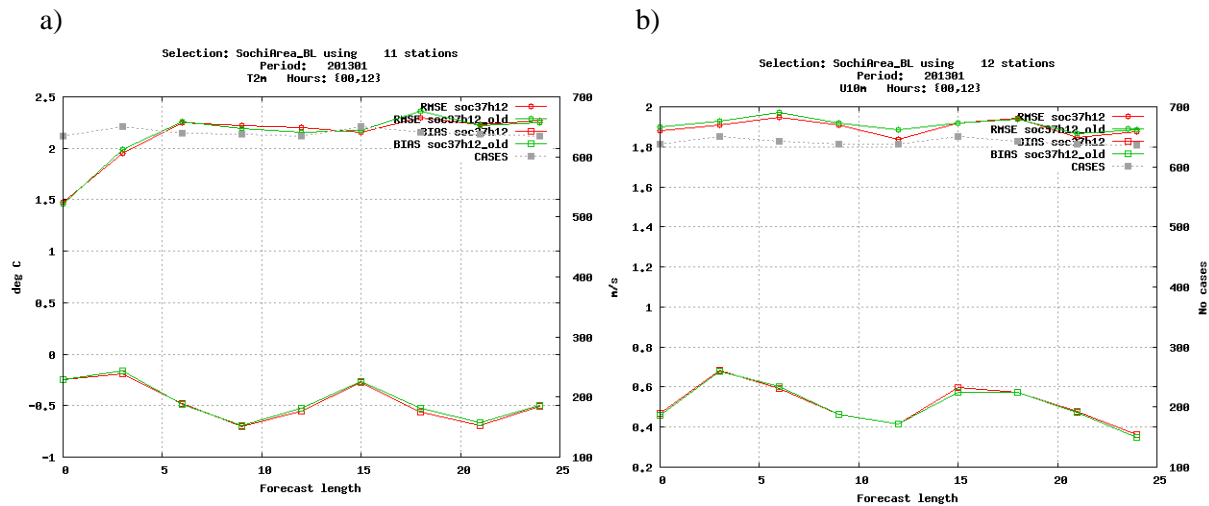


Figure 4: Bias (square) and RMSE (circle) of a) 2m-temperature and b) 10m wind speed as a function of forecast length for January 2013. Red is Harmonie baseline experiment (with SRTM and 2.5 km), green is Harmonie (with GTOPO30 and 2.5 km).

The effect of the grid size

Figure 5 shows the bias and RMS-errors of 2m-temperature for three different Harmonie experiments with 2.5, 1.5 and 1.0 km grid size. The overall results (Fig. 5a) over the Sochi region reveal some differences. The RMS-error is slightly smaller with 1.5 and 1.0 km than with 2.5 km grid size. Moreover, the bias of the 1.0 km experiment is smallest, whereas the 2.5 km model has the largest underestimation of the temperature. When focusing on single station statistics located in the valley of the Olympic mountain cluster (Karsnaya Polyana, 550 m above sea level, Fig. 5b) the same result is even more amplified. The model experiment with 1.0 km grid size has almost zero bias and the two highest resolution experiments outperform the baseline experiment with 2.5 km grid size. With the 2.5 km grid size the characteristics of the narrow valley cannot be represented accurately enough, whereas in 1.0 km model the height difference between the model surface and the Krasnaya Polyana station is only about 50 m.

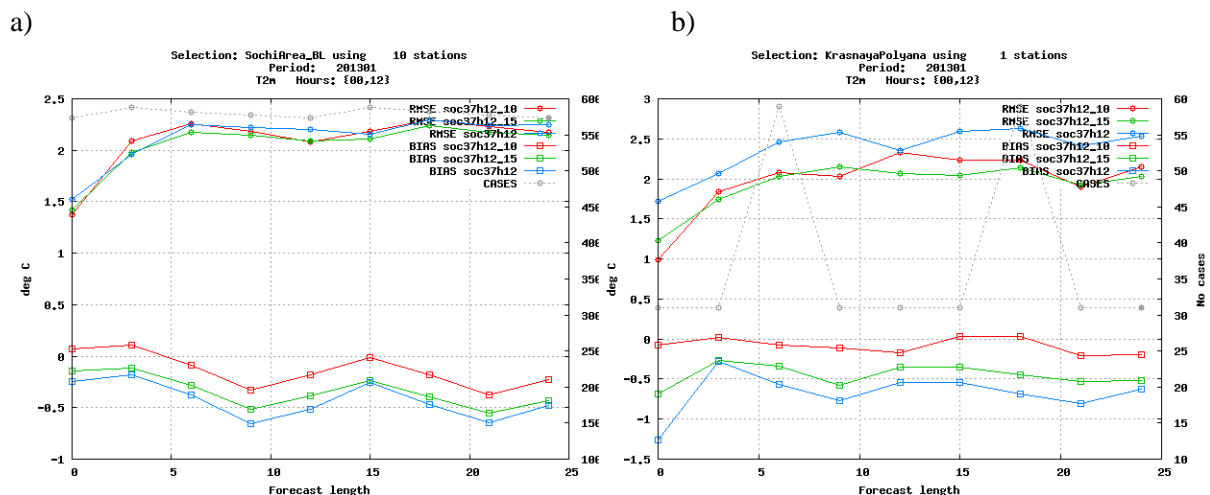


Figure 5: Bias (square) and RMSE (circle) of 2m-temperature at a) Sochi region (10 stations) and b) Krasnaya Polyana (single station) as a function of forecast length for January 2013. Harmonie experiments with 2.5 km (blue), 1.5 km (green) and 1.0 km (red) grid size.

Problems with high resolution

During the 3 month test period, the high resolution model experiments (1.5 and 1.0 km) were suffering frequent numerical instability problems. Figure 6a shows an example of the wavy pressure patterns over Black Sea that were typical for these unstable situations. Moreover, both high-resolution models crashed more than 5 times during the test period. These situations were solved simply by reducing the time step from 40s to 30s (1.5 km) and from 30s to 15-20s (1.0 km). However, reducing the time step did not remove the noisy patterns from the pressure field.

An additional model experiment based on cy38h1.alpha2 (soc38h12_10 with 1.0 km grid size) was prepared to study the most unstable cases. The namelist options, which are in line with Vivoda et al. (2008), were used as follows (Yann Seity, personal communication).

1. For time stepping, use of Predictor-Corrector (PC) scheme with one iteration instead of Stable Extrapolating Two Time Level Scheme (SETTLS):

```
&NAMCTO
  LPC_FULL=.TRUE.,
  LPC_NESC=.TRUE.,
  LPC_NESCT=.FALSE.,
  LPC_CHEAP=.TRUE.,
/
&NAMDYN
  LSETTLS=.FALSE.,
  LSETTLST=.TRUE.,
  NSITER=1,
  LRHDI_LASTITERPC=.TRUE.,
/
```

2. Advection of variable gw (in grid point space) instead of vertical divergence d_4 and switching off the SL diagnostic of bottom boundary condition:

```
&NAMDYNA
  LGWADV=.TRUE.,
  LRDBBC=.FALSE.,
/
```

3. Setting the subsequent GFL flags:

```
&NAMGFL
  YQ_NL%LPT=.TRUE.,
  YL_NL%LPT=.TRUE.,
  YI_NL%LPT=.TRUE.,
  YS_NL%LPT=.TRUE.,
  YR_NL%LPT=.TRUE.,
  YG_NL%LPT=.TRUE.,
  YTKE_NL%LPT=.TRUE.,
  YLRAD_NL%LPT=.FALSE.,
  YIRAD_NL%LPT=.FALSE.,
/
```

Figure 6b shows the positive effect of the proposed settings on the pressure field of soc38h1a2_10. The noisy features are not present anymore and the model run is stable even with 40s time step. In this case, 20s time step was required for default configuration in order to avoid the crash. The PC-scheme with one iteration was about 20% more expensive than SETTLS, however, due to the larger time step the PC-scheme was much more efficient overall.

These settings were tried also in version cy37h1.2, however, the test failed due to technical reasons. During the first time step TKE went down to 0 and consequently the run aborted in the APL_AROME routine. The source of the problem was not investigated further.

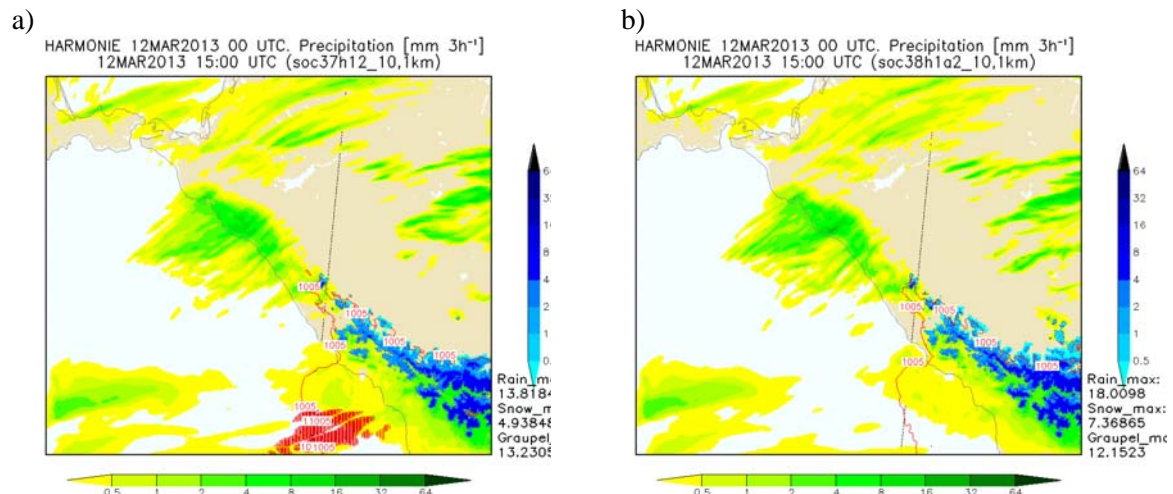


Figure 6: Accumulated precipitation [mm/3h] and mean sea level pressure [hPa]. a) soc37h12_10 with 1.0 km grid size, 20s time step, LSETTLS and LRDBCC. b) soc38h1a2_10 with 1.0 km grid size, 40 s time step, PC-scheme and LGWADV. Analysis: 12 Mar 2013, 00 UTC. Valid at: 12 Mar 2013, 15 UTC.

5 Conclusions

This paper presents the first results of the high resolution (1 – 2.5 km) Harmonie experiments over the region of Sochi Olympics. The study is related to the WMO FROST-2014 project, where one FMI contribution is to run deterministic Harmonie before and during the Olympics as part of the forecast demonstration project. In addition, some problems related to high resolution modelling are reported and subsequent corrections to solve these problems are suggested. The main results are the following:

- Baseline Harmonie experiment with a 2.5 km grid size outperformed the available regional (HIRLAM) and global (ECMWF) forecasts.
- The alternative orography dataset (SRTM) showed neutral impact in comparison with GTOPO30.
- Harmonie experiments with 1.5 and 1.0 km grid resolution showed overall improvement in comparison with the 2.5 km experiment, especially over mountainous Sochi region.
- Both highest resolution Harmonie experiments with default time stepping and advection settings suffered significant numerical instability problems. Most of the noisy patterns were removed by using the PC-scheme and by advecting the vertical velocity based variable in grid point space.

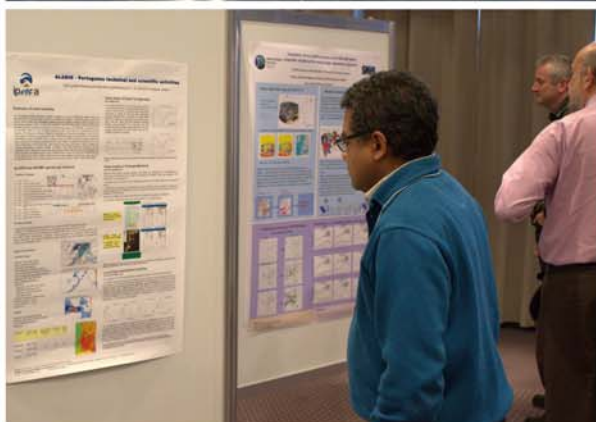
References

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf, 2007, The Shuttle Radar Topography Mission, *Rev. Geophys.*, **45**, RG2004, doi:[10.1029/2005RG000183](https://doi.org/10.1029/2005RG000183).

FROST-2014: Available online, <http://frost2014.meteoinfo.ru/doc>

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, V. Masson, 2011: The AROME-France Convective-Scale Operational Model. *Mon. Wea. Rev.*, **139**, 976–991.
doi: [http://dx.doi.org/10.1175/2010MWR3425.1](https://doi.org/10.1175/2010MWR3425.1)

Vivoda, J., J. Masek, F. Vana, M. Vörös, Y. Seity, 2008: How to run NH kernel. *AROME training course, 4-10 March 2008, Lisbon, Portugal*, 49 pp. Available online: <http://www.cnrm.meteo.fr/gmapdoc/spip.php?article189>



New model cycle 38t1

Joël Stein and Marc Tardy. Météo-France, DPREVI/COMPAS

1 Description of the e-suite

On 2 July 2013, a new model cycle of the operational models was implemented. The new cycle 38t1 included the following modifications for the global model ARPEGE:

- Active assimilation of the data of IASI, AMSU-A, MHS, GPS-RO and ASCAT provided by the satellite METOP B (Figure 1)
- Active assimilation of CRIS and ATMS provided by SUOMI-NPP (Figure 1)
- Active assimilation of OSCAT provided by OceanSat-2
- Increase of the numbers of assimilated channels for IASI, AIRS and of data for GPS-RO
- Use of wavelets in the assimilation cycle of the EDA and ARPEGE to provide daily estimations of the covariances for the guess
- Use of TKE in the shallow convection scheme KFB
- Faster damping of the SST analysis toward the OSTIA values

Specific modifications were added for the LAM hydrostatic (ALADIN) or non-hydrostatic (AROME):

- Use of a new orography and land-sea mask coming from GMTED2010 and new physiographic data from HWSD
- Switch off the canopy parameterization over sea
- Finer sampling of satellite data (IASI, SSMI-S, AMSU-A)
- Active assimilation of SEVIRI channels 8.7, 12 et 13.4 microns over land in AROME

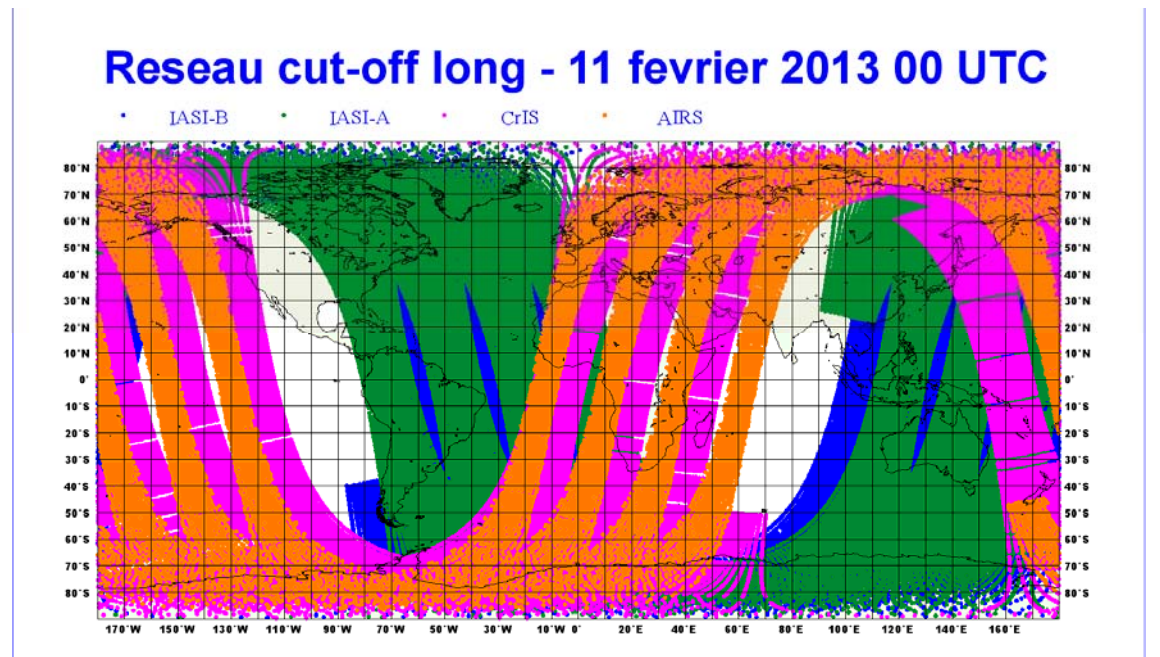


Fig1: Available data provided by the 4 infrared hyperspectral sounding assimilated by ARPEGE: IASI-A (green), IASI-B(blue), Cris (pink) and AIRS (orange) during 6 hours the 11 February 2013 .

2 Results of the comparison

2.1 Objective scores and subjective comparison

The operational suite and the e-suite were compared from 5 February to 2 July 2013, i.e. 139 days. The objective scores improved for the e-suite (Fig 2) only during the first day of the simulation in the upper troposphere and during the whole range in the stratosphere. This improvement is mainly due to the changes in the data assimilation (new observations and methods), the negative signal over Europe is of moderate amplitude and not representative of extratropical domains, while NORD20 and SUD20 show improvements. Results for others 3D fields are mainly neutral.

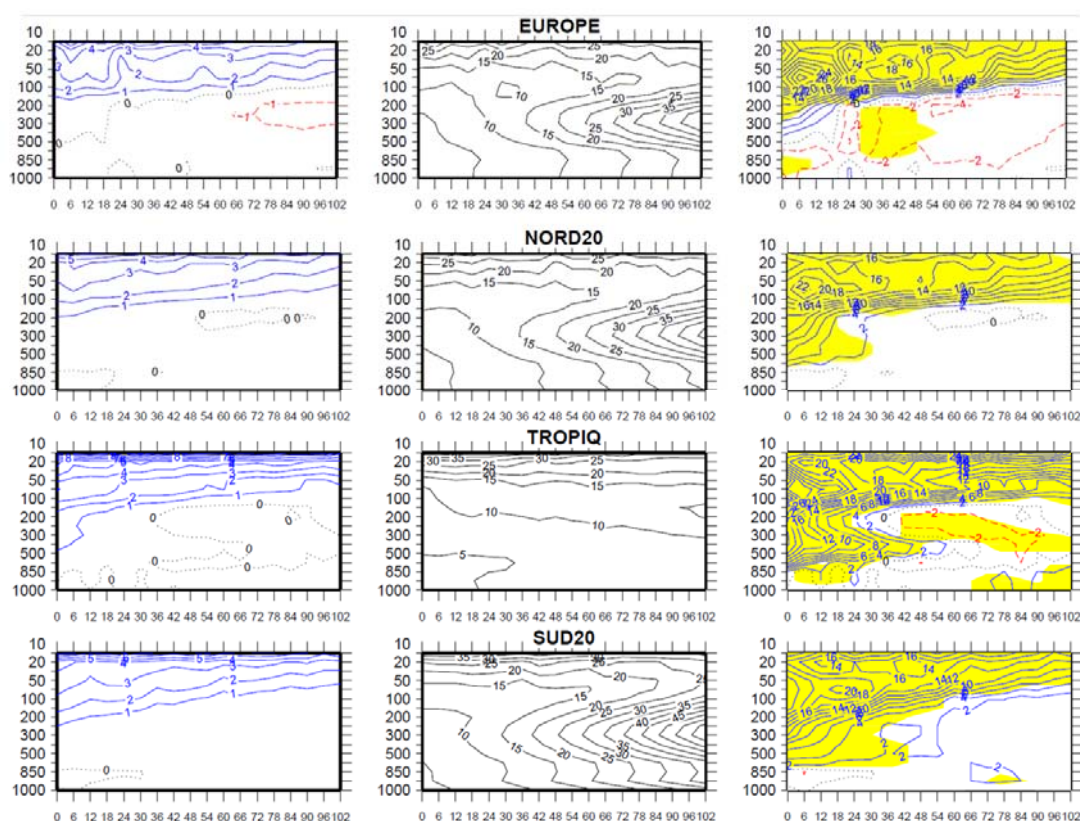


Fig 2 : Mean difference of the RMSE of the géopotential forecasted by the operational suite and the e-suite. The reference is provided by the ECMWF analysis for the domains: EUROPE (first line) NORD20 i.e. where the latitude is greater than 20° (second line), TROPICS i.e. where latitude is between -20° and $+20^\circ$, SUD20 i.e. where latitude is lower than -20° . The central row corresponds to the rmse (isolines every 5 hPa) of the operational model in function of the lead time and the pressure. The left panel to the difference of both rmse (blue isoline corresponds to an improvement drawn every hPa and red isoline to a degradation). The right panel displays the same information as the left panel but measured in % of the operational rms. Regions where a significativity test based on bootstrap method is positive are coloured in yellow.

A comparison has been performed by a forecaster to subjectively evaluate both suites. The result is in favour of the e-suite with 11 % of forecast improvements against 9% of degradations when the forecasts were different. The other cases were too far from analysis to allow the forecaster to decide which suite performs better in this case. A comparison of precipitation fields shows a light improvement for the e-suite mainly due to the modification of the shallow convection scheme.

ALADIN / HIRLAM

23rd Workshop / All-Staff Meeting 2013

Hotel Hilton Reykjavik Nordica, Reykjavik, 15-19 April, 2013

Sessions	
Monday 15 <i>Main room</i> <i>~85p</i>	Opening session S1: Data assimilation ALADIN LTM meeting Poster session Reception/ice breaker
Tuesday 16 <i>Main room ~85p</i>	S1: Data assimilation S2: Probabilistic forecasting and LAMEPS S3: Operational experiences Poster session
Wednesday 17 <i>Main room ~85p</i>	S4: Dynamics S5: Model physics Dinner
Thursday 18 <i>Main room ~85p</i> <i>+ 2nd room ~20p</i>	S6: System aspects and verification Closing session Working groups (system aspects, predictability/DA aspects)
Friday 19 <i>~22p</i>	HMG-CSSI: 9h-18h

Talks (including questions) : normally 20' each (10' for PM opening)
20' for discussions

Monday, 15 April (8:30-18:00)

08:30-09:00 Registration

09:00-10:10 Plenary opening session. Chair: Jan Barkmeijer

Opening of the Meeting

Welcome

Practical arrangements for the meeting

09:10-09:20 Piet Termonia: ALADIN status overview

09:20-09:30 Jeanette Onvlee : HIRLAM activities and highlights

09:30-09:50 Yong Wang : LACE news

09:50-10:10 Gergely Boloni : Status of the EUMETNET C-SRNWP project

10:10-10:40 Coffee break

10:40-12:20 Plenary session 1: Data assimilation and use of observations. Chair: Claude Fischer

10:40-11:00 : Jean-Francois Mahfouf : Recent usage of observations in Meteo-France data assimilation systems

11:00-11:20 : Magnus Lindskog : Harmonie ATOVS data assimilation and coordinated impact study

11:20-11:40 : Roger Randriamampianina : The radar data assimilation at OMSZ and met.no

11:40-12:00 : Siebren de Haan : On Mode-S EHS observations: quality issues, impact and the future

12:00-12:20 : Mate Mile : Arome data assimilation activities in Hungary

12:20-13:50 Lunch

13:50–15:30 : Plenary session 1 (cont.): Data assimilation and use of observations. Chair: Ole Vignes

13:50-14:10 : Tuuli Perttula : Towards assimilating IASI satellite observations over cold surfaces – the cloud detection aspect

14:10-14:30 : Carlos Geijo : Progress on radar data assimilation in Harmonie using the field alignment technique : Application to Doppler Wind Data

14:30-14:50 : Antonin Bucanek : Evolution of dispersion spectra during successive steps of the assimilation cycle

14:50-15:10 : Xiaohua Yang : Harmonie 3DVAR-RUC with asynoptic base time

15:10-15:30 : Jan Barkmeijer : Harmonie at KNMI and future plans

15:30-15:50 Coffee break

15:50-18:00 ALADIN LTM meeting

15:50-18:00 Poster session 1

18:00- Reception/ice breaker

Tuesday, 16 April (9:00-18:30)

09:00–10:20 Plenary session 1 (cont.): Data assimilation and use of observations. Chair: Gergely Boloni

- 09:00-09:20 : Nils Gustafsson : HIRLAM 4DEnsVar – discussion and first results
- 09:20-09:40 : Claude Fischer : Status of assimilation methods at Meteo-France
- 09:40-10:00 : Jelena Bojarova: Data assimilation and ensemble prediction: an enriching partnership
- 10:00-10:20 : Piet Termonia : A Short-Time Augmented Extended Kalman Filter for the surface

10:20–10:50 *Coffee break*

10:50–12:10 Plenary session 1 (cont.): Data assimilation and use of observations. Chair: Nils Gustafsson

- 10:50-11:10 : Annelies Duerinckx : Preliminary results on combining Surfex EKF with 3D-Var atmospheric assimilation for ALARO
- 11:10-11:30 : Mariken Homleid : Harmonie snow analysis
- 11:30-11:50 : Homa Kheyrollapour : improvement of objective analysis of lake surface state in Hirlam using satellite observations
- 11:50-12:10 : General discussion

12:10-13:40 *Lunch*

13:40-15:40 Plenary session 2: Probabilistic forecasting and LAMEPS. Chair: Roger Randriamampianina

- 13:40-14:00 : Theresa Gorgas : News on research and development in ALADIN-LAEF
- 14:00-14:20 : Inger-Lise Frogner : GLAMEPS and HarmonEPS experimentation and plans
- 14:20-14:40 : Geert Smet : Surface perturbations by cycling surface breeding
- 14:40-15:00 : Åke Johansson : Impact of a Hybrid Variational-Ensemble Data Assimilation System on the Performance of an ETKF-based EPS System
- 15:00-15:20 : Sibbo van der Veen : The impact of using CAPE singular vector perturbations in GLAMEPS
- 15:20-15:40 : General Discussion

15:40-16:10 *Coffee break*

16:10- 17:30 Plenary session 3: Operational experiences. Chair: Jean-Francois Mahfouf

- 16:10-16:30 : Sander Tijm : Systematic Harmonie problems as seen from a user perspective
- 16:30-16:50 : Javier Calvo : On Harmonie/Arome progress at AEMET
- 16:50-17:10 : Trygve Aspelien : Arome-Norway – from experiments to official public forecasts available for the whole wide world
- 17:10-17:30 : General discussion

17:30–18h30 : Poster session 2

Wednesday, 17 April (9:00-16:00)

09:00-10:40 Plenary session 4: Dynamics. Chair: Daan Degrauwe

09:00-09:20 : Jozef Vivoda : Vertical finite elements in the NH dynamical core of ALADIN

09:20-09:40 : Petra Smolikova : LACE – dynamics and coupling

09:40-10:00 : Daan Degrauwe : Localized horizontal discretizations with appropriate adjustment properties for the ALADIN dynamics

10:00-10:20 : Marco Kupiainen : Energy estimates and weak boundary procedures for LAM

10:20-10:40 : General discussion

10:40-11:10

Coffee break

11:10-12:30 Plenary session 5: Model physics. Chair: Sander Tijm

11:10-11:30 : Yves Bouteloup : Status of Arpege and Arome physics

11:30-11:50 : Laura Rontu : Summary and update of the stable boundary layer workshop

11:50-12:10 : Neva Pristov : ALARO-1: an overview

12:10-12:30 : Radmila Brozkova : ALARO-0 baseline: status and latest results on convective diurnal cycle

12:30-14:00

Lunch

14:00-15:40 Plenary session 5 (cont.) : Model physics. Chair: Radmila Brozkova

14:00-14:20 : Lisa Bengtsson : A closer look at fog over sea, clouds in cold conditions and deep convection in Harmonie

14:20-14:40 : Kristian Pagh Nielsen : Aspects of the radiation scheme in Harmonie

14:40-15:00 : Sigurdur Thorsteinsson : Impact of sub-grid orographic drag on weather forecasts in Iceland

15:00-15:20 : Ekatherina Kourzeneva : Recent HIRLAM lake-related activities

15:20-15:40 : General discussion

15:40-19:00 : free time

19:00-? : dinner

Thursday 18 April (9:00-13:00)

09:00-10:40 Plenary session 6: System aspects and verification. Chair: Sami Niemela

09:00-09:20 : Claude Fischer : Status of cycles; OOPS

09:20-09:40 : Ryad El-Khatib : Fullpos-2

09:40-10:00 : Ulf Andrae : The status of the Harmonie system

10:00-10:20 : Ulf Andrae : computational challenges and aspects of Harmonie

10:20-10:40 : Enda O'Brien : Accelerating Harmonie with GPU's

10:40-11:00 Coffee break

11:00-12:40 Plenary session 6: System aspects and verification. Chair: Alex Deckmyn

11:00-11:20 : Filip Vana : The OpenIFS project at ECMWF

11:20-11:40 : Christoph Zingerle : HARP – a common framework for verification

11:40-12:00 : Xiaohua Yang : Validation of Harmonie Cy38h1

12:00-12:20 : Sami Niemela : FROST-14 – first experiences of deterministic Harmonie in the Sochi region

12:20-12:40 : General discussion

12:40-13:00 : Plenary closing session. Chairs: Jeanette Onvlee and Piet Termonia

13:00- 14:30 Lunch

14:30-16:00 Working group on system aspects

Working group on transversal predictability – data assimilation issues

Friday 11 May (9:00-18:00) : HMG-CSSI meeting

Posters

Alica Bajic	The NWP activities in Croatia
Patrik Benacek	First test of water vapour regimes verification using SAL
Dag Bjorge	Norwegian Meteorological Institute- operational use of Harmonie AROME
Jelena Bojarova	Correction of non-additive errors in the HIRLAM hybrid ensemble variational data assimilation using image registration approach
Gergely Boloni	NWP at the Hungarian Meteorological Service ; Improving wintertime low cloud forecasts over Hungary: First experiments (on behalf of Balázs Szintai, Eric Bazile, Yann Seity, François Bouysse)
Javier Calvo	Long-term verification of HIRLAM and ECMWF models
Jure Cedilnik	ALADIN related activities in Slovenia
Maria Derkova	ALADIN related activities at SHMU
Emily Gleeson	Short-wave radiation schemes in Harmonie
Hassan Haddouch	Numerical Weather Prediction in Morocco 2013
Marek Jerczynski	ALADIN related activities in Poland
Ersin Kucukkaraca	ALADIN in Turkey
Jelena Bojarova	Correction of non-additive errors in the HIRLAM hybrid ensemble variational data assimilation using image registration approach
Andres Luhamaa	Harmonie activities in Estonia
Maria Monteiro	ALADIN - Scientific and Technical activities at IPMA
Bolli Palmason	Harmonie @ IMO – Visualisation and application
Tuuli Perttula	Impact of IASI satellite observations in harmonie during convective storm events in Finland
Emiel van der Plas	Assessing quality of high resolution forecasts
Patricia Pottier	The NWP systems at Meteo-France
Petra Smolikova	New developments and operational changes at CHMI
Simona Tascu	ALADIN related activities in Romania
Alena Trojakova	Latest status of BlendVar developments at CHMI
Boryana Tsenova	Status of operational applications in Bulgaria
Ole Vignes	MetCoOp – operational cooperation between SMHI and met.no
Eoin Whelan	NWP in MetEireann
Christoph Wittmann	NWP related activities in Austria

Participants

Dz Algeria			Dk Denmark	Xiaohua Yang (HMG) Kristian Pagh Nielsen	2
At Austria	Theresa Gorgas Yong Wang (LACE) Christoph Wittmann (LTM) Christoph Zingerle	4	Es Estonia	Andres Luhamaa Velle Toll	2
Be Belgium	Alex Deckmyn (LTM, CSSI) Annelies Duerinckx Daan Degrauwe (CSSI) Geert Smet Piet Termonia (PM) Michiel Vanginderachter	6	Fi Finland	Ekatherina Kurzeneva Sami Niemelä Tuuli Perttula Laura Rontu (HMG)	4
Bg Bulgaria	Boryana Tsenova (dep. LTM)	1	Is Iceland	Bolli Pálmason Guðrún Nina Petersen Hróbjartur Þorsteinsson Sigurður Thorsteinsson Sigurður Jónsson	5
Hr Croatia	Dijana Klaric Tomislav Kovacic (dep. LTM) Antonio Stanesic	3	Ir Ireland	Enda O'Brien Klara Finkle Emily Gleeson Eoin Whelan	4
Cz Czech Republic	Patrik Benáček Radmila Brozkova (LTM) Antonín Bucanek Petra Smolikova Alena Trojakova (CSSI)	5	Li Lithuania		
Fr France	Yves Bouteloup Ryad El Khatib (CSSI) Claude Fischer (LTM, CSSI) Jean-Antoine Maziejewski (ST) Jean-Francois Mahfouf (CSSI) Patricia Pottier (ST)	6	Nl Netherlands	Jan Barkmeijer Tilly Driesenaar (HMG) Siebren de Haan Jeanette Onvlee (PM) Emiel van der Plas Sander Tijn Sibbo van der Veen	7
Hu Hungary	Gergely Boloni (LTM) Mate Mile	2	No Norway	Solfrid Agersten Trygve Aspelien Jelena Bojarova (HMG) Dag Björge Inger-Lise Frogner (HMG) Nils Gustafsson Mariken Homleid Jörn Kristiansen Roger Randriamampianina Andrew Singleton Ole Vignes	11
Ma Morocco	Hassan Haddouch (LTM)	1	Es Spain	Javier Calvo Carlos Geijo Mariano Hortal (HMG)	3
Pl Poland	Marek Jerczynski (LTM)	1	Se Sweden	Ulf Andrae (HMG) Lisa Bengtsson Åke Johansson	5

				Magnus Lindskog Marco Kupiainen	
Pt Portugal	Maria Monteiro (LTM) Manuel Lopes	2	Canada	Homa Kheyrollahpour	1
Ro Romania	Simona Tascu (dep.LTM)	1			
Sk Slovakia	Maria Derkova (CSSI) Oldrich Španiel Jozef Vivoda (LTM)	3			
Si Slovenia	Neva Pristov (LTM) Jure Cedilnik	2			
Tu Tunisia	Zied Sassy (dep. LTM)	1			
Tr Turkey	Ersin Küçükcaraca (LTM)	1			
Total		39			44
ECMWF : Filip Vana					