



ALADIN-HIRLAM Newsletter

No 5, August 2015



Joint 25th ALADIN Workshop & HIRLAM All Staff Meeting 2015
13-16/04/2015, Elsinore, Denmark



DSCF2364_just.jpg



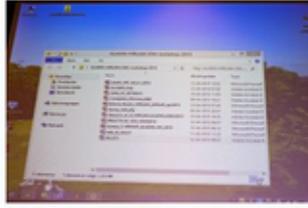
DSCF2365_just.jpg



DSCF2366_just.jpg



DSCF2367_just.jpg



DSCF2368_just.jpg



DSCF2369_just.jpg



DSCF2401_just.jpg



DSCF2404_just.jpg



DSCF2411_just.jpg



DSCF2391_just.jpg



DSCF2392_just.jpg



DSCF2393_just.jpg

CONTENTS

Introduction, Patricia Pottier	3
List of planned events	5
Articles	
Neighbourhood post-processing of AROME-EPS precipitation, François Bouttier, Tarik Kriat	6
The GLAMEPS real-time production system, Kai Sattler, Alex Deckmyn, Xiaohua Yang	15
A comparison between the default HarmonEPS perturbations and random field perturbations, Henrik Feddersen	26
PROFORCE – Bridging of Probabilistic Forecasts and Civil Protection, Clemens Wastl Martin Suklitsch, Yong Wang, Andre Simon	34
The new Numerical Weather Forecast Verification System in Tunisia, Wafa Khalfaoui.....	36
Can the compiler make the difference ?, Ryad El Khatib	45
Experiments with HARMONIE on Xeon Phi Coprocessors, Enda O'Brien	53
Impact experiments of HARMONIE radar data assimilation, Shiyu Zhuang, Mats Dahlbom, Xiaohua Yang, Bent Hansen Sass	61
Impact experiments of HARMONIE radar data assimilation, Jelena Borayova.....	67
Progress and plans in the ARPEGE and AROME models physics, Yann Seity, Jean-Marcel Piriou, Yves Bouteloup, Alexandre Mary, Sébastien Riette, Benoit Vié, Rachel Honnert, Clemens Wastl, Laura Rontu, Christoph Wittmann	79
Parameterization of orographic effects on surface radiation in AROME-SURFEX, Clemens Wastl, Alexandre Mary, Yann Seity, Laura Rontu, Christoph Wittmann	81
Description of the OCND2-option in the ICE3 clouds- and stratiform condensation scheme in AROME, Karl-Ivar Ivarsson	83
KNMI parameterization testbed as an evaluation tool: Status and plans, Wim de Rooy, Roel Neggers, Pier Siebesma, Bert van Ulf, Willem Vlot, Cisco de Bruijn, Emiel van der Plas, Eric Bazile, Henk Klein Baltink, Fred Bosveld, Toon Moene	88
Shortwave Radiation Experiments in HARMONIE : Tests of the cloud inhomogeneity factor and a new cloud liquid optical property scheme compared to observations, Emily Gleeson, Kristian Pagh Nielsen, Velle Toll, Laura Rontu, Eoin Whelan	92
Study of urban climatology over Budapest with SURFEX/TEB model at the Hungarian Meteorological Service, Gabriella Zsebeházi, Ilona Krüzselyi, Gabriella Szépszó	107

ALARO-1 first operational application, Radmila Brozkova,	114
Deep convection and downdraught in Alaro-1, Luc Gérard	123
On the necessity of modularity regarding the dynamical core, Daan Degrauwe	125
Testing accuracy of finite element scheme used in vertical discretization of ALADIN-NH, Petra Smolikova	133
A new Solver for the Non-Hydrostatic Semi-Implicit Scheme in HARMONIE using Green Functions, Carlos Geijo	138

Publications

Recently or soon published	163
Comment on: “Impact of changes in the formulation of cloud-related processes on model biases and climate feedbacks” by Carlo Lacagnina, Frank Selten and A. Pier Siebesma (J. Adv. Mod. Earth Syst.2014; 6(4): 1224-1243), Kristian Pagh Nielsen & Emily Gleeson	165

Introduction

Patricia Pottier

I am glad to provide you with the fifth edition of the combined Newsletter of the HIRLAM and ALADIN consortia. This edition is mainly dedicated to the “[25th ALADIN Workshop & HIRLAM All Staff Meeting 2015](#)” that took place on 13-16 April 2015 in Elsinore (Denmark), with additional several articles based on other work.

90 people from 23 countries gave more than 50 presentations and presented about 35 posters. Some working groups and many side meetings also took place during the week. The ALADIN Local Team Managers held their [18th meeting](#) (the first one ever where all ALADIN partners were represented). The ALADIN CSSI and the HIRLAM Management Group had their [annual coordination meeting](#) on Friday 17, at the Danish Met Institute premises in Copenhagen.

Please consult the [dedicated webpage](#) on the aladin website for the agenda, the list of participants, the pdf files of the presentations and posters, the minutes of the discussions and working groups, some photos.

The videos of the presentations are also available, on a protected part of the aladin website : please use the "log in" button, at the top left under the horizontal menu on the aladin home page (<http://www.cnrm.meteo.fr/aladin/>). Once identified, you have access to the "Partners only" part of the aladin website, including the videos. *If you need the user/password to log in, just ask me !*

A new part about “**Publications**” has tentatively been added in this newsletter, with the aim to announce or comment papers published in NWP specialised international journals. As this is the first time such a list is included in the Newsletter, the information used for it is likely still incomplete; hopefully, it will be better the next time.

I thank the authors for their contributions and hope you enjoy reading the fifth ALADIN-HIRLAM Newsletter.

Patricia

For additional information, please visit the [ALADIN](#) and [HIRLAM](#) websites, or just ask the authors of the articles.

List of planned events

1 Meetings

- [37th EWGLAM & 22 SRNWP](#) (& 19th ALADIN LTM meeting), **5-8 October 2015, Belgrade (Rs)**
- HAC Meeting, **27-28 October, De Bilt (NI)**
- [25th ALADIN GA & 2nd joint ALADIN GA & HIRLAM Council](#), **24-25 November 2015, Algiers (Dz)**, with new MoUs signature
- [21st ALADIN Workshop & HIRLAM All Staff Meeting 2016](#) (& 20th ALADIN LTM meeting), **4-7 April 2016, Tunisia**
- Coordination meeting between ALADIN CSSI and HIRLAM MG, **8 April 2016, Tunisia**
- [13th ALADIN PAC meeting](#), **23 May 2016 morning, Ankara (Tk)**,
- [4th joint HAC/PAC meeting](#), **23 May 2016 afternoon, Ankara (Tk)**,
- HAC meeting, **24 May 2016, Ankara (Tk)**,

2 Working Weeks / Working Days

- HARMONIE Data Assimilation & Use of Observations working week, 7-10 September 2015, Dublin (Ie)
- Training week on the use of the Harmonie system for climate modelling, 14-17 September 2015, Norrköping (Se)
- HARMONIE working week on EPS and predictability, 19-23 October 2015, Canary Islands (*place t.b.c.*)
- Forecasters meeting, 21-23 October 2015, Lisbon (Pt)
- LACE DA Working Days, 30 September – 2 October 2015, Bratislava (Sk)
- SURFEX validation WW, *date t.b.d.*, Brussels (Be)
- HARMONIE system working week, *date (prob. November) and place t.b.d.*
- [more information on-line](#)

Earlier this year ...



The [3rd joint HAC/PAC meeting](#) took place in Helsinki (21 May) with a major topic, the ALADIN-HIRLAM convergence : the outcomes of the common declaration adopted by the 1st joint ALADIN GA & HIRLAM Council, the outcomes of the data policy task force guidance about the ALADIN/HIRLAM common agreement for the next MoUs and the organization of the 2nd joint ALADIN General Assembly & HIRLAM Council and the MoUs signatures in Algiers.

Neighbourhood post-processing of AROME-EPS precipitation

François Bouttier, Tarik Kriat (CNRM/GMME, Météo-France&CNRS)

1 Introduction

Theis et al (2005) proposed to improve precipitation forecasts by applying a spatial tolerance on NWP model output, in recognition of that fact that smaller scales (<30km, typically) are less predictable than larger ones. We revisit this idea in the context of convection-permitting ensemble prediction, using the Arome-France-EPS system (nicknamed 'Arome-EPS' in this paper).

Our Arome-EPS has a very small ensemble size (12 members, once per day, with 36-hour range in this preoperational configuration). The model is the 2014 Arome-France one with a 2.5km mesh. The ensemble uses the global PEARP ensemble for its initial and lateral boundary conditions, and a representation of surface and model errors (Bouttier et al 2012, Bouttier et al 2015).

This study takes place during June/July 2014, during which most heavy precipitation events are caused by convective thunderstorm outbreaks in a weak southwesterly synoptic flow. They mostly occur over plains, so that predictability is quite low, as evidenced by member maps: taking 3-hourly accumulations (rr3) on 13 June 2014 as an example in Figure 1, one notices that, although the predicted intensities and patterns are consistent with the observations (Figure 2), the exact location of precipitation is inconsistent between members at the regional scale (Figures 1 and 2 both cover a 100x100km square in central France). The ensemble spread is correct, because the inconsistencies are similar between ensemble members and observations. Nevertheless, as we shall see, the small ensemble size and the large rr3 variability mean that point scores are rather poor and do not reflect the useful information embedded in the ensemble output.

This kind of situation is usually called the 'double penalty problem'. In the verification community, it is often solved by upscaling both model and observations in order to blur out small-scale detail (see a review in Ebert 2008, and a practical application in Amodei and Stein 2009). Upscaling often removes some important small-scale details, such as precipitation maxima in convective cells. Also, it is not helpful to users who require point forecasts: they are typically not interested to know whether precipitation was observed in a neighbourhood, they want the best possible (deterministic or probabilistic) forecast at a single point. In this study we investigate ensemble post-processing methods that address this issue; the post-processed precipitation products will be evaluated using objective scores at observation points.

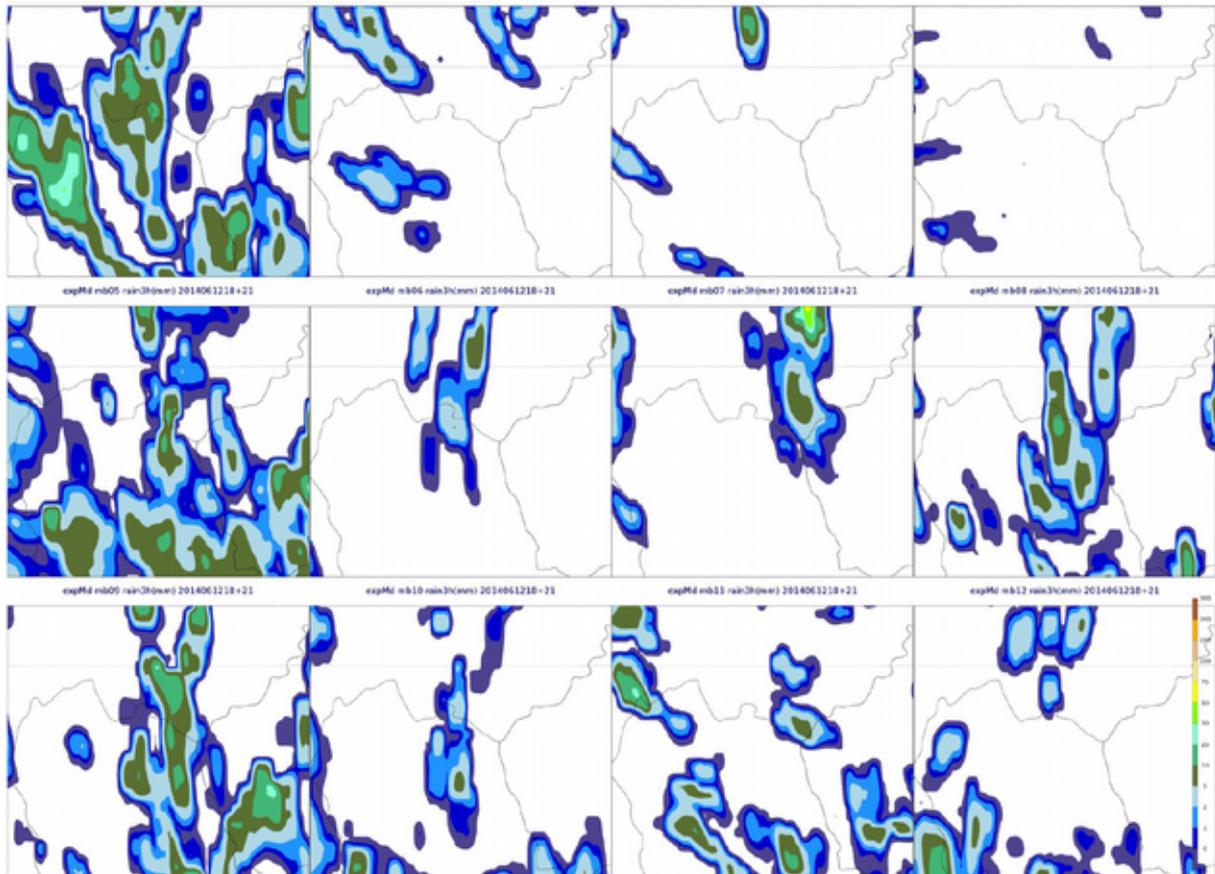


Figure 1: rr3 precipitation in each member of Arome-EPS in the forecast started on 12 June 2014, 18UTC, at range 18 hours.

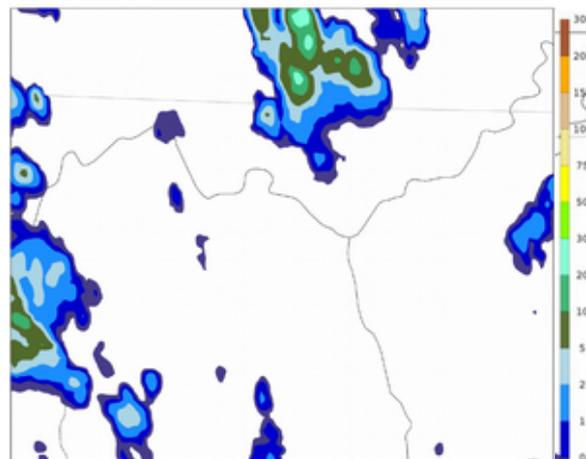


Figure 2: rr3 observations (radar + raingauge merged product) at the same time as Figure 1.

2 Methods for spatial tolerance

The basic idea of neighbourhood post-processing is to replace, at each output gridpoint x , the raw ensemble $A_n(x)$ of predicted values (i.e. the field values of the $n=12$ members at that gridpoint) by an ensemble $B_p(x)$ of p values drawn from the members in a neighbourhood in space and/or time. In this section we consider spatial tolerance only. Several methods are possible for implementation, in this section we restrict ourselves to disk-shaped spatial neighbourhoods with an empirically chosen radius r . If there are m gridpoints in the disk, then $n.m$ values are available to build the ensemble at point x ; we call $D_{nm}(x)$ this set of values.

In the so-called '**by-quantile sampling**' method, we assume that all $n.m$ values are equally likely, the implied PDF is sampled by setting the drawn values $B_p(x)$ to p equally spaced quantiles of the distribution $D_{nm}(x)$. For instance, if $nm=2$ and $p=1$, then $B_p(x)$ will be the median of $D_{nm}(x)$. If $p=nm$, then the quantiles are equal to the $D_{nm}(x)$ ensemble itself. In practice $p \ll nm$ because r is of the order of 10-50km, so that nm is a large number ($\sim 10^4$), too large for practical purposes. To convert D_{nm} into B_p , one must subsample distribution D_{nm} . Mathematically, the best way to sample a one-dimensional PDF is to use its quantiles, because most probabilistic products are integrals of the ensemble distribution: quantiles are evenly spaced values in probability space, and evenly spacing sample points generally is an optimal way of performing 1-D numerical integration of an unknown-shaped function (this is the textbook Riemann integration technique, with $O(1/p)$ convergence rate). In other words, using quantiles of D_{nm} to define B_p minimizes the sampling error.

In the '**uniform random sampling**' method, we replace the use of quantiles by a random draw of $B_p(x)$ in the set of $D_{nm}(x)$. This is equivalent to a Monte-Carlo integral; its theoretical convergence rate is in $O(1/\sqrt{p})$, which is slower than quantile sampling, but it has its own advantages as we shall see below. We have checked that, with reasonable values for m and p , it does not matter if the random sampling is done with or without replacement: scores from both methods are undistinguishable in our dataset.

Both uniform and by-quantile sampling methods assume that all points in the neighbourhood are equally likely. One could argue that points closer to the center are more likely, because they are less prone to location error, if the ensemble is indeed capable of predicting weather features smaller than radius r . For instance, orographically-forced precipitation often has a very predictable spatial distribution. To account for the fact that precipitation values are more likely near the disk centre than near the edge, we have tested a very simple non-uniform sampling, called '**centered random sampling**', in which the likelihood of a point being picked has an approximately Gaussian shape, and points at the edge of the disk have a 30% lower probability than in the uniform sampling. These choices are probably non-optimal, they are just devices to test the sensitivity to the sampling strategy.

In summary, the adjustable parameters of the spatial part of the neighbourhood post-processing are:

- the neighbourhood radius r , which implies the number m of points in the neighbourhood;
- the number p of values that are sampled in the neighbourhood, the end products are thus presented as an ensemble of size p , possibly different from the original ensemble size n .
- whether the sampling method is 'by-quantile', 'random uniform', or 'random centered'

3 Dataset and impact of spatial tolerance

We compare the impact of the above methods on 26 days of rr3 Arome-EPS forecasts in June 2014 (the results are the same on July, October or November data). Forecasts are verified every 3 hours from ranges 9 to 33 hours, against about 1000 quality-controlled stations at each time. The rr3 differences between Arome-EPS members and observations are used to compute a standard set of ensemble scores, including member and ensemble mean bias&rms, spread/skill consistency, rank diagrams, CRPS, Brier, ROC and economic value scores, taking observation error into account, and using a bootstrap technique to test for score variations. More detail on the score computation is given in Bouttier (2012).

The **dependency on the sample size p** is illustrated by the ROC diagram in Figure 3, which shows (as do the other scores) that the ensemble performance not sensitive to its size beyond about 100. This is consistent with results obtained with many other ensemble prediction systems, that it is not possible to prove a benefit of increasing the ensemble size beyond 100. In the rest of this study, we shall use $p=100$.

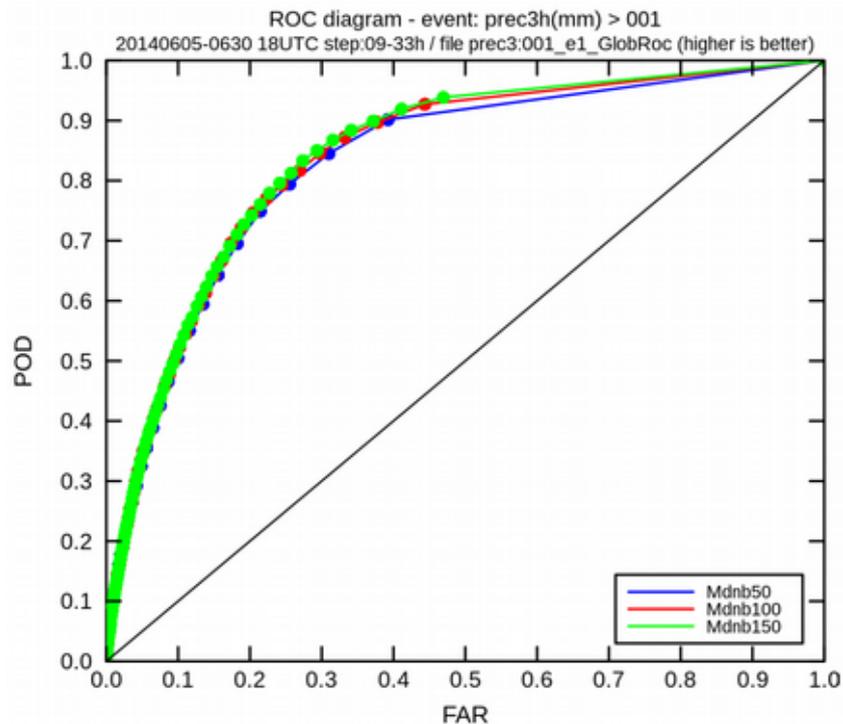


Figure 3: impact of the sample size p on the post-processed ensemble performance, illustrated by the ROC diagram for event 'rr3>1mm'. The blue, red, and green curves correspond to sizes 50, 100, 150, respectively.

The dependency on the sampling method is checked by comparing the three above mentioned methods. It was not possible to see any impact on the scores of the **centered vs uniform sampling**, so the study will focus on the uniform sampling. One possible explanation could be that the reduction of location error that the centered sampling brings, is compensated by an increase in inter-member error correlation in $B_p(x)$, such correlations are likely since $p > n$ and precipitation forecasts often have spatially correlated errors. Theoretically, inter-member error correlation reduces the information content and the spread of the post-processed ensemble, but it is difficult to confirm if this problem is really significant.

When **comparing by-quantile vs random sampling**, most scores were identical, but a few diagnostics showed undesirable features in the quantile sampling method. For instance, the Frequency Bias (Figure 4) becomes very poor for high precipitation events: the by-quantile sampling reduces the frequency of high precipitation events, because there is a physical link between the size and intensity of precipitation: high-precipitation cells tend to be more localized than light-precipitation. In any given disk, high precipitation values (when they exist) only cover a small part of the disk, so their frequency is nearly always too small to be picked up by even the highest quantile in the disk, in which case they will never show up in the $B_p(x)$. For instance, if $p=100$, precipitation values that do not occur more than 1% of the time any disk will never occur in the post-processed ensemble, although their frequency was non-zero in the original ensemble. By contrast, random sampling will preserve the frequency of even the rarest events.

In terms of threshold-based measures such as the ROC, Brier scores or reliability diagrams (not shown), the by-quantile and random sampling methods perform similarly for low precipitation thresholds ($rr3 < 3\text{mm}$), but the by-quantile method is the worst at higher thresholds (thresholds up to $\sim 10\text{mm}$ can be studied in this dataset). In summary, the by-quantile method will not be further used because it is inappropriate for practically important high precipitation events, while not showing any clear benefit over the random method for light precipitation.

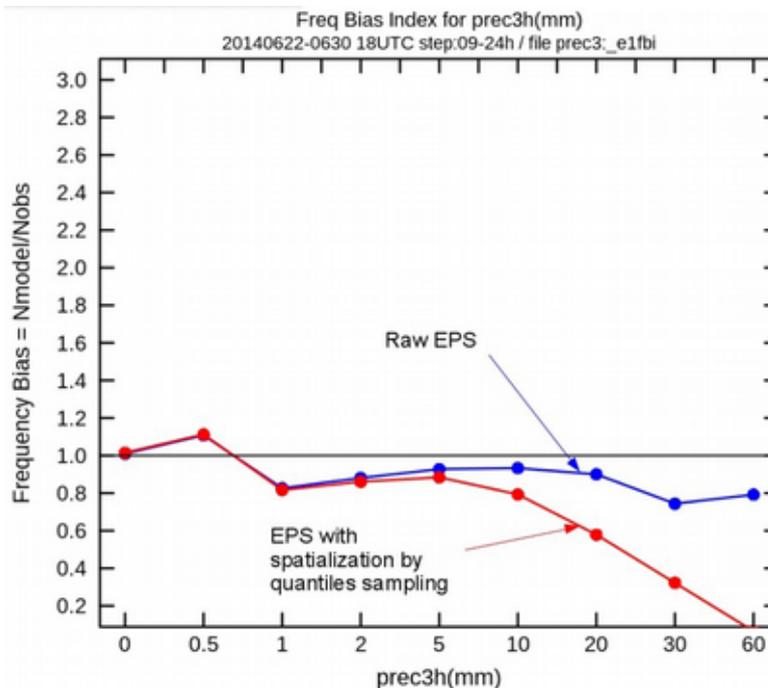


Figure 4: comparison of the Frequency Bias Index ($N_{\text{forecast}}/N_{\text{obs}}$) for various $rr3$ thresholds, between the by-quantile and uniform random sampling methods. The optimal value of the index is 1, lower values indicate an underprediction of the events.

Finally, the **dependency on tolerance radius r** is explored in the framework of the best method so far, the 100-member uniform sampling method. Figures 5 shows the ROC Area (or AUC, Area Under the Curve) and reliability diagrams for various values of r . The result is that the introduction of spatial tolerance significantly improves the forecasts in all aspects. For instance, at the 1mm threshold, the ROC area increases by about 5%, which is very significant (it amounts to fixing about a quarter of the wrong forecasts in the raw ensemble). At higher thresholds, the improvement is comparable, although it is more difficult to test statistically because the dataset is short. The spatial tolerance does not change the ensemble biases, since it is an equitable random resampling of the raw Arome-EPS output.

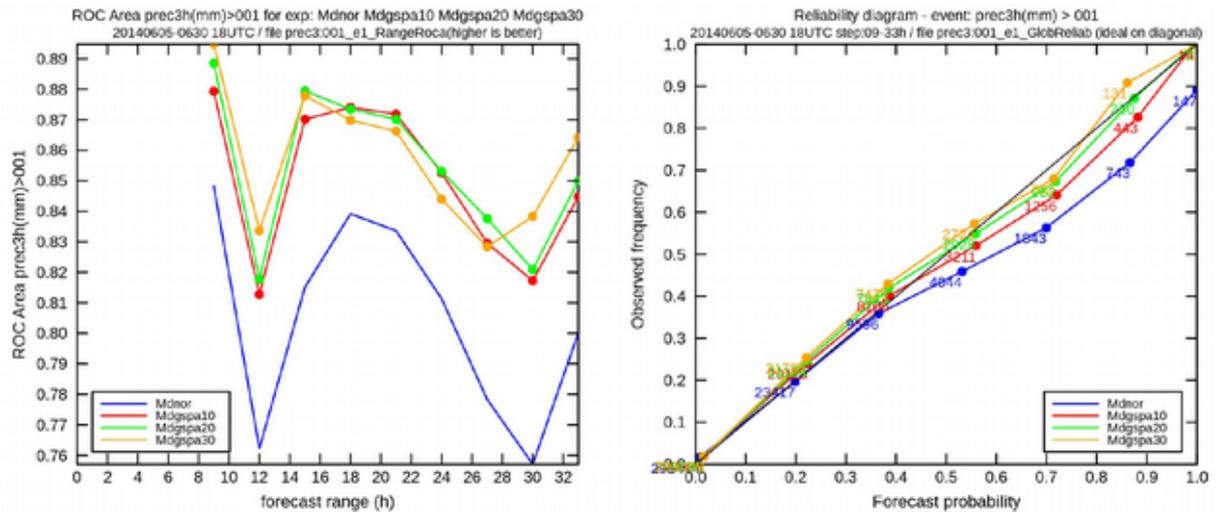


Figure 5: impact of the spatial tolerance with the uniform sampling method, with several tolerance radii of $r=0, 10, 20,$ and 30 (blue, red, green, and orange curves), which correspond to geographical disk radii of $0, 25, 50,$ and 75km , respectively.

The **mechanism of the improvement** brought by the post-processing is illustrated by figure 6, which shows a ROC diagram with and without spatial tolerance (using a 100-member, 30-km uniform random sampling). The left parts of the curves are very close to each other, which means that at most operating (i.e. probability) levels, the impact of spatialization is neutral. The main improvement to the area under the ROC curve is at the top of the figure, it comes from new points that add some 'roundness' to the angle of the Arome-EPS curve. This angle is a signature of the small size of the Arome-EPS ensemble: it corresponds to forecast probabilities of $1/12=8.3\%$, the lowest non-trivial probabilities that can be predicted in a 12-member ensemble. In other words, the main benefit of the post-processing is that it adds more members that have approximately the same information content as the 12 Arome-EPS original members. If Arome-EPS had 100 members, the post-processing would probably have no beneficial impact, as seen in our study on the sample size.

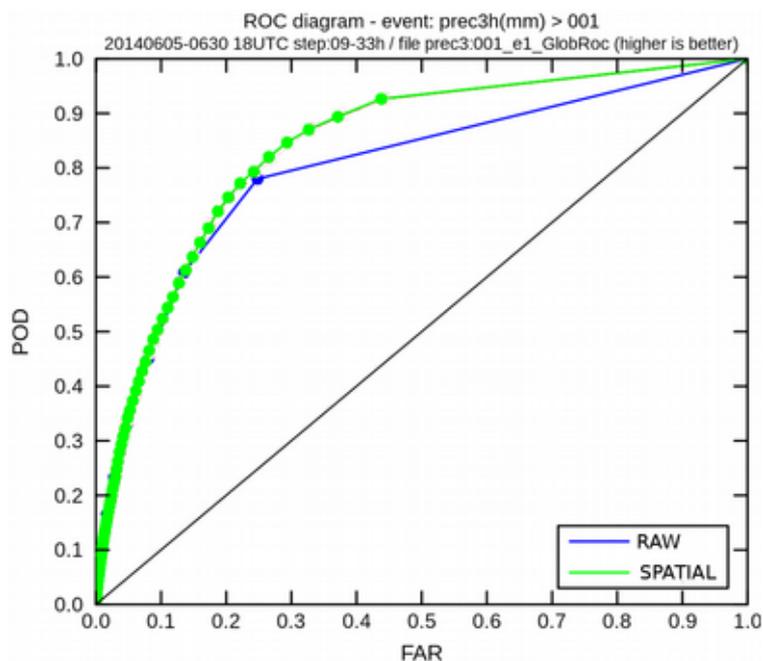


Figure 6: net impact of the spatial tolerance post-processed ensemble (green), compared with the ROC curve of the raw Arome-EPS ensemble (blue)

4 Test case with spatial tolerance

The 50-km uniform random sampling post-processing method is illustrated on a real case in Figure 7. Focusing on the 'rr3>1mm' event, the event was observed to occur in the green areas; the symbols show the Arome-EPS forecast probability of the event before and after the post-processing. As expected, one notices that the forecast is essentially unchanged for events with a large spatial extent. Isolated events, such as precipitation inside dry areas or blobs of high probabilities in areas of lower probabilities, are often smoothed out: the post-processing tends to erase forecast features that are likely to be the result of sampling errors. The post-processed field is less noisy than the raw one, and generally in better agreement with the observations. A key feature of the technique is that it does not smooth out probabilities of events that are consistently predicted in each member and neighbourhood, even if the physical size of these events is small (such as convective precipitation cells). If a simple spatial smoothing had been applied to the output of each member, then the information about local precipitation maxima would have been corrupted in small-sized precipitating events.

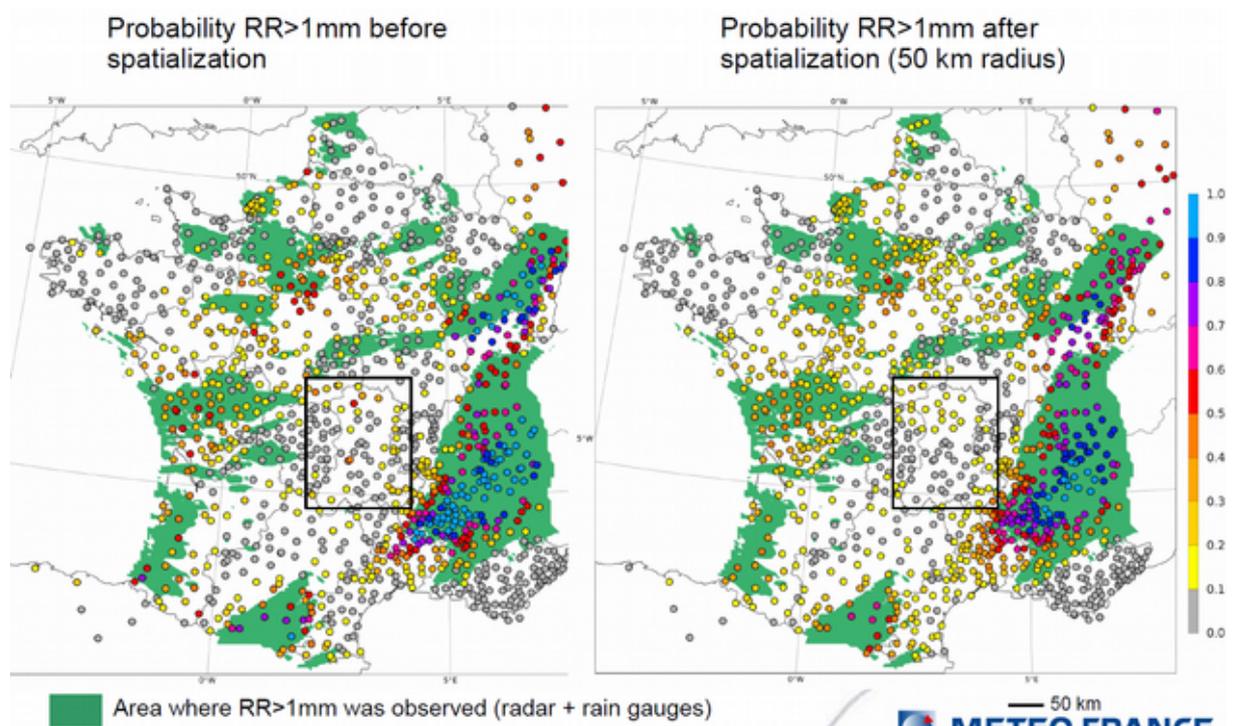


Figure 7: Observations of the 'rr3>1mm' event (green shading) on 29 June, 06UTC, and forecast probabilities (symbols, with colour key on right) of the event before (left), and after (right) application of spatial tolerance post-processing.

5 Impact of temporal tolerance

Most precipitating events have a strong time dependency. This may happen because they are linked to the diurnal cycle of convection, or because many precipitating systems are moving, so that in the course of a few hours a given system will have moved over a distance larger than its own size. Thus, prediction errors are likely to affect the timing as well as the location of the precipitation fields. Timing errors of fast-moving events will show up as location errors too, so they are already handled to some extent by the spatial tolerance methods outlined above. On the other hand, when predicting at a given point in space and time, there may be valuable information in the ensemble that can only be found by looking at the model output before or after the target time.

In order to test this idea, we extend the uniform random sampling method by expanding the input ensemble to forecast fields at 1 hour before and after the verification time, yielding a $3n=36$ -member equivalent Arome-EPS ensemble $\{ An(x,t-1h) , An(x,t) , An(x,t+1h) \}$ as input to the spatial neighbourhood post-processing. This is equivalent to considering that the sampling is done in a cylinder of $3n$ values that has the spatial shape of a disk, and a 3-hour temporal extent. The end result remains a $Bp(x,t)$ post-processed ensemble, with the same size $p=100$ as when only spatial tolerance is considered (in order to remove possible score biases linked to the ensemble size).

The result, shown in Figure 8, is that temporal tolerance brings a small but significant improvement on the quality of the ensemble output, on top of the spatial tolerance. It means that temporal tolerance brings information that cannot be obtained by purely spatial tolerance methods: in an ideal ensemble post-processing system, both post-processing methods should be combined.

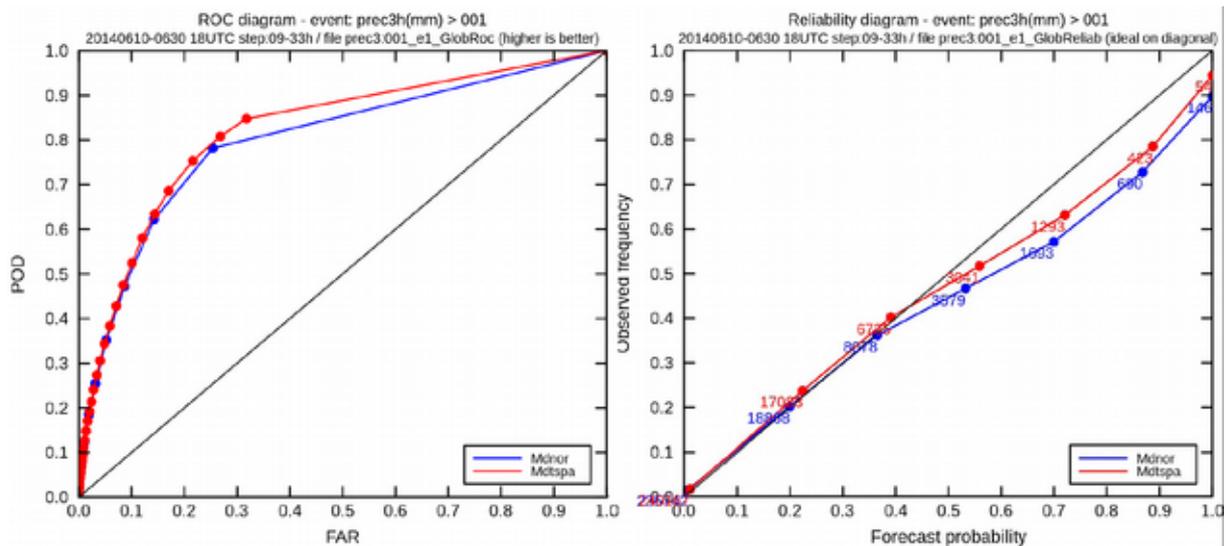


Figure 8: Impact of introducing temporal tolerance on top of spatial tolerance for predicting $rr3 > 1mm$ events, as measured using the ROC (left) and reliability (right) diagrams. The ROC area difference is statistically significant.

6 Summary and conclusions

Convective-scale ensemble prediction of precipitation can be significantly improved by introducing a simple tolerance to location and timing errors, according to this study carried out during June, July, October and November 2014 over France. The improvement is clear and significant for light precipitation. There seems to be similar benefits for heavy precipitation as well, although the sample size is too small to establish the statistical significance of this result.

The main conclusion is that double-penalty problems can probably be avoided, in convective-scale NWP systems, by an appropriate upscaling of the NWP system output. The ROC curves, in particular, clearly prove that there is valuable objective information in the Arome-EPS convective-scale ensemble prediction system. One must be careful not to corrupt the information about small-scale, high intensity events in the post-processing, otherwise one would defeat the purpose of running high-resolution NWP systems.

Another conclusion is that the 12-member Arome-EPS ensemble is too small: much of the ROC area improvement comes from mechanically extending, by random resampling, the ROC curve beyond the extremal decision thresholds that show up as angular points on the curves. The study of sensitivity to resampling size p suggests that Arome-EPS forecasts would be much improved by increasing the ensemble size to about 50, but not beyond. Future tests of larger Arome-EPS ensembles should be benchmarked against the post-processed 12-member ensemble, because additional Arome members will be much more expensive than the post-processing technique outlined here.

This work should be extended to larger datasets, and to other parameters for which timings and location errors are known to be a major forecasting problem. Perhaps the tolerance radius should be smaller in mountainous areas, where rain location may be more predictable than over plains. The sampling area could have a different shape, e.g. an ellipse or a terrain-dependent pattern. The post-processing should include some form of bias correction and calibration in order to alleviate systematic errors in the model output, such as biases in precipitation intensity. The value of the tolerance radius (in space and time) should be tuned adaptively in order to account for seasonal and regional variations of the predictability of precipitation.

7 References

- Amodei, M. and Stein, J. (2009), Deterministic and fuzzy verification methods for a hierarchy of numerical models. *Met. Apps*, 16: 191-203. doi: 10.1002/met.10
- Bouttier, F., B. Vié, O. Nuissier and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, 140: 3706-3721. doi:10.1175/MWR-D-12-00031.1
- Bouttier, F., L. Raynaud, O. Nuissier, B. Ménétrier, 2015: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Accepted for publication in *Quart. Jour. Roy. Meteor. Soc.* in June 2015. 16pp. DOI: 10.1002/qj.2622
- Ebert, E.E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appl.*, 15, 51-64
- Theis S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, 12, 257-268.

The GLAMEPS real-time production system

Kai Sattler, Alex Deckmyn, Xiaohua Yang

1 Introduction

GLAMEPS stands for Grand Limited Area Model Ensemble Prediction System, and is an initiative on a limited area models ensemble prediction system with participating countries from the ALADIN consortium (<http://www.cnrm.meteo.fr/aladin>), and from the HIRLAM consortium (<http://joomla.hirlam.org>).



GLAMEPS is a multi-model limited area ensemble prediction system including additional perturbation methods within the different models, like use of multiple parameterisation schemes or stochastic physics. GLAMEPS combines the advantages of the multi-model approach with the variability of configuration within a certain LAM model.

2 Development

2.1 GLAMEPS version 0

GLAMEPS was launched in summer 2006 in a cooperation between the ALADIN and the Hirlam consortium, with the idea to provide to the partners a LAM ensemble system for production of forecasts on a large common model domain covering the North Atlantic and the European Area, including as much as possible the areas of interest of each partner.

Originally GLAMEPS was thought as a decentralized production system, where several institutes contribute ensemble members to an overall ensemble, with possibility for central and individual post-processing.

With GLAMEPS development of the Hirlam EPS script system was started. Similarly ALADIN was also prepared for use within GLAMEPS. As a third model component, Euro-TEPS system was considered, but finally it didn't enter into production.

Version 0 was rather experimental. It included options to make use of boundary data from ECMWF-EPS or from the Euro-TEPS ensemble, and different perturbation approaches were experimented with in an effort to get to an ensemble system suitable for operational production.

2.2 GLAMEPS version 1

This version was established from late 2008 on, and it aimed to become the first GLAMEPS system to run in real-time. Until its final establishment as a time-critical (option 2) production system on the ECMWF computers in 2011, it was tested meticulously and further developed in order to be monitored by ECMWF operator staff.

Ensemble configuration:

- 12 + 1 Aladin members
- 12 + 1 Hirlam members using Kain-Fritsch convection parameterization
- 12 + 1 Hirlam members using Straco convection parameterization
- 14 + 1 ECMWF members from ECMWF-EPSC plus the high-resolution run

The "+1" represents the unperturbed control member, and all in all there are 54 ensemble members.

Cycling configuration:

- 12-hourly, twice a day
- 54h forecast length
- product generation twice a day

Model Data Domain (common GLAMEPS domain):

- rotated lat-lon grid
- 646 x 492 gridpoints, 11km mesh size
- domain area outlined in figure 1

The GLAMEPS version 1 system was used for real-time production until end of September 2014.

2.3 GLAMEPS version 2

Work on version 2 of GLAMEPS was started in mid 2013 with the aim to increase horizontal resolution, but especially also in order to configure the ensemble setup in a new way. Major elements of this new setup were:

- from 2 daily runs (12h-cycles) to 4 daily runs (6-h cycling)
- 6-hour lagged ensemble member configuration
- remove ECMWF-EPSC members from the ensemble member list, ECMWF-EPSC thus only providing boundary data, and no longer contributing to the GLAMEPS ensemble with direct ensemble member model output
- replace the Aladin model with two flavours of the Alaro model
- 6-hourly product generation
- extended product palette

The GLAMEPS version 2 system was introduced at the 24th Workshop/All-Staff Meeting 2014 (Alex Deckmyn, 2014), and it has become the real-time production system in October 2014. It is described in the next chapter.

3 GLAMEPS current production system

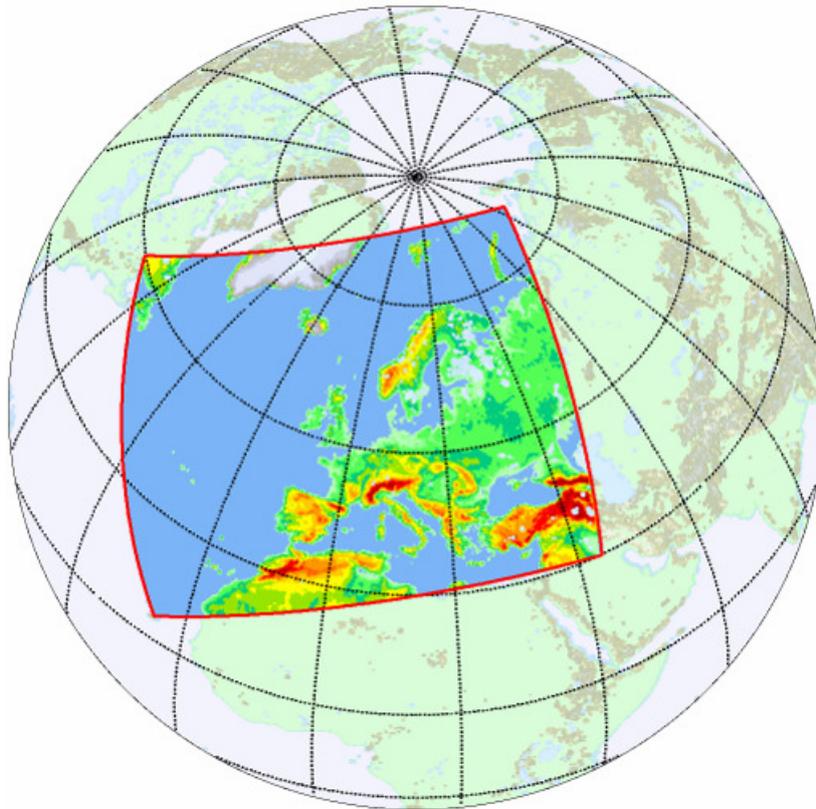


Figure 1: The GLAMEPS domain has been virtually the same for versions 0, 1 and 2.

The current GLAMEPS production system is based on GLAMEPS version 2. The common domain area for the system is shown in figure 1. The grid characteristics are as follows:

- rotated lat-lon grid
- 870 x 660 gridpoints
- 8km mesh size
- 40 vertical levels in the model runs

A detailed description of the GLAMEPS grid is given on the GLAMEPS production technical information page on <https://hirlam.org/trac/wiki/GlamepsProductionUserInfo>.

The configuration of the ensemble is as follows:

- 12 + 1 Alaro members using the ISBA model at the surface
- 12 + 1 Alaro members using the SURFEX model for surface
- 12 + 1 Hirlam members using Kain-Fritsch convection parameterization
- 12 + 1 Hirlam members using STRACO convection parameterization

The "+1" represents the unperturbed control member, and all in all there are 52 different ensemble members. Configuration of the daily cycling is as follows:

- 6-hourly, 4 times a day, with alternating perturbed members, i.e. 6+1 members per cycle for each model flavour
- 12-hourly boundary update (ECMWF-EPS)
- 54h effective forecast length
- product generation 4 times a day, based on latest two cycles
- 12h prolonged cycle forecasts for backup facility

Figure 2 outlines the data flow within the cycle configuration, from the boundary data to the products. Note that the data flow of observation data is not shown in the figure, which is mainly meant to illustrate the inter-connections concerning model data.

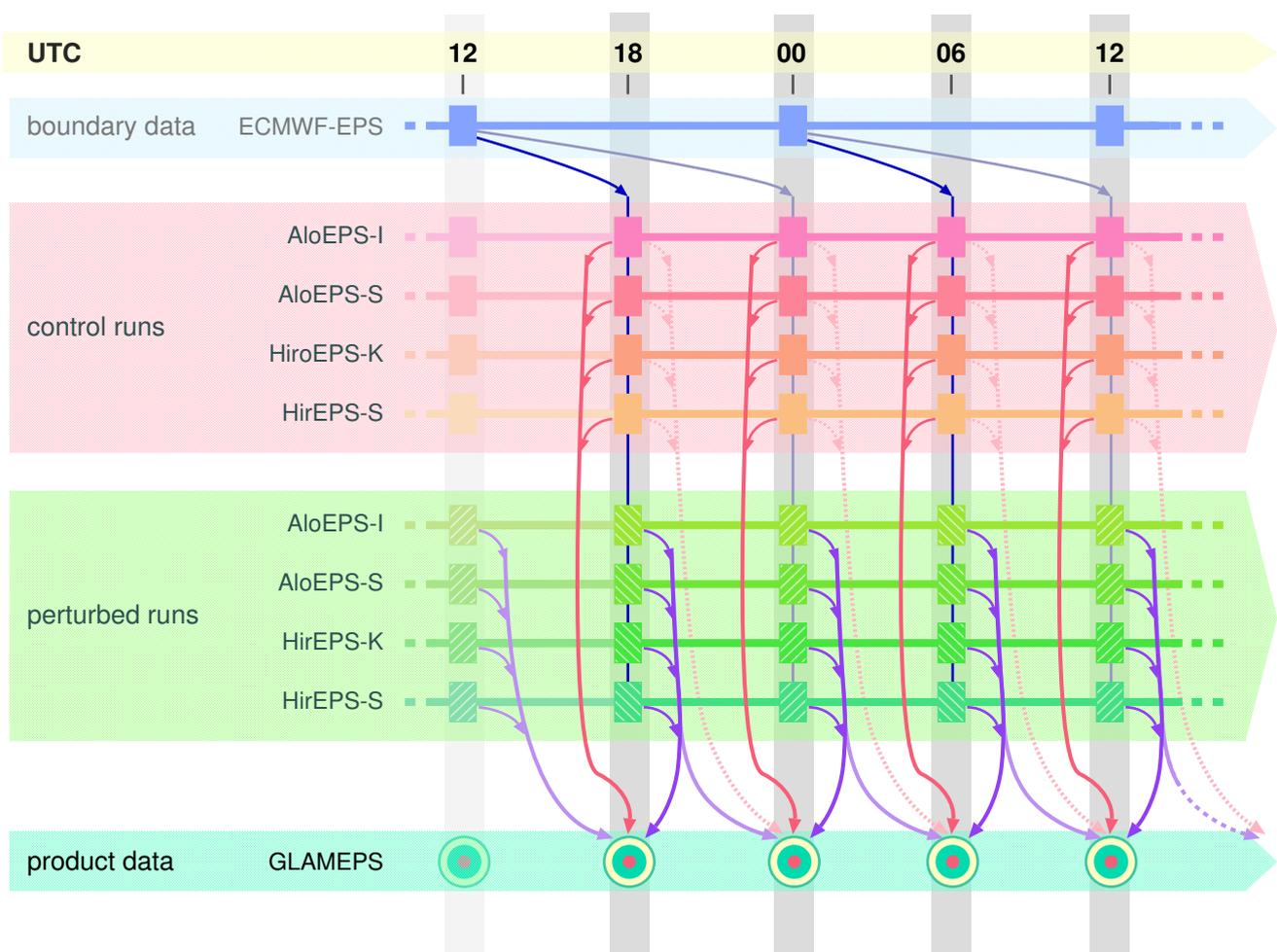


Figure 2: Data flow within the cycling configuration of the current GLAMEPS real-time production system.

The uppermost row (in yellow) shows the time of cycling, where 5 cycles are presented in 5 rows (grey). The boundary data from the ECMWF-EPS model (in blue) is available every 12 hours at cycle times 00 and 12 UTC

(blue blocks). As GLAMEPS cycling is every 6 hours (redish and greenish blocks), boundary data from one ECMWF-EPS cycle feeds two GLAMEPS cycles (indicated by the outward arrows from the blue blocks).

The four GLAMEPS control runs (in red) use the same boundary data twice, however, with a difference of 6 hours in boundary age (dark blue and grey arrows outward from the blue coloured ECMWF-EPS blocks). The boundary age is 6 hours for GLAMEPS cycles 06 and 18, and 12 hours for GLAMEPS cycles 00 and 12. The reason why the 00/12 GLAMEPS cycles don't use the respective ECMWF-EPS cycles is simply because EPMWF EPS boundary data is ready about 7-8 hours after the given cycle time, and GLAMEPS is running in real-time.

The perturbed GLAMEPS runs (in green) are indicated by one coloured block, even though there are several ensemble members for each of the four LAM models. Like the control runs the perturbed members are feeded by ECMWF-EPS boundary data (vertical blue lines between the redish and greenish coloured blocks), but now with perturbed data. Each GLAMEPS LAM member receives data from a different perturbed ECMWF-EPS member (not illustrated i the figure).

The horizontal lines between the coloured blocks within a row indicate the data flow within an ensemble member from cycle to cycle, like for example the first guess data.

The differently coloured curved arrows indicate the model data flow for product generation (petrol green row at the bottom). The butt at each cycle column within the lowermost row (product data) represents the bunch of GLAMEPS products, and the curved arrows reaching a butt indicate what data goes into the product generation process.

For example, cycle 00 product generation makes use of model data form the control runs of the current cycle 00 (dark red curved arrows), as well as from those perturbed members that were run at cycle 00 (dark violet curved arrows), plus those perturbed members that were run at the previous cyle 6 hours earlier, i.e. cycle 18 in the example, represented by the light violet curved arrows. The latter were already used in the former cycle 18, and now they are reused once more. Note that these members are not run at cycle 00, but first again at cycle 06 (alternating perturbed members). These three data sources make up the model data for the standard GLAMEPS product generation based on 52 ensemble members.

In addition to the three model data sources, the control members from the former cycle may also be utilized by those GLAMEPS users, who prefer to do their own postprocessing. Adding these data (indicated by the light red curved arrows) would add four more members and increase the ensemble size to 56.

4 Products

Since the first real-time production system of GLAMEPS, a standard product generation was established. It performs post-processing of the ensemble model output and its results are presented a/o on the GLAMEPS web page (<https://glameps.org>). But also the model output data is made available to users of the GLAMEPS system in order to facilitate individual post-processing by GLAMEPS users.

4.1 Model Output Data

This data is prepared for ready use in individual post-processing by GLAMEPS users. The output from the different LAM models has been converted onto the grid of the common GLAMEPS domain (figure 1).

The field data is made available in GRIB 1 format, and point data is available in an SQL table format. The field data is seperated into two *Output Data Streams*:

4.1.1 Main Data Stream

This represents the major GLAMEPS model output content:

- temperature at 2m, and pressure levels 925, 850 and 500
- winds at 10m, and pressure levels 925, 850 and 500
- MSLP, and pressure level heights for 925, 850 and 500hPa
- dew point temperature at 2m
- wind gust at 10m
- total accumulated precipitation, snowfall and cloud cover
- downward shortwave radiation

These data are stored into the archives at ECMWF, so they are available for historical reference.

4.1.2 Secondary Data Stream

This stream represents temporarily available GLAMEPS model output content available on the ECMWF computing system for a couple of days. It contains many more parameters like min-max temperatures at 2m, CAPE, humidity on pressure levels, cloud parameters, zero-degree level and many more. For details, please refer to the GLAMEPS product user informations on <https://hirlam.org/trac/wiki/GlamepsProductionUserInfo>.

Note that these data are available on the GLAMEPS data storage area at ECMWF only for up to about 5-7 days.

4.1.3 Table Extraction Data

These are point data extracted from the model output files, and provided as SQL table data for more than 10.000 locations from within the common GLAMEPS domain. Parameters comprise

- mean sea level pressure
- 2m temperature
- 10m wind
- total precipitation

4.2 Model Level Data

This is full model level data in the original models' grid and file format for the latest 4 cycles. These are available on ECMWFs super computer under a temporary storage area.

The data are provided *as is*, and users have to convert, process and administrate the data by themselves. It should also be noted that provision of these data is less stable than for example the Model Output Data, and it is very likely to fail in cases of production problems and in cases, where backup routines have to take effect in order to sustain operation.

4.3 Probability Maps

The probability maps are results from the GLAMEPS standard postprocessing, and following maps are provided (figure 3):

- 3h accumulated precipitation
- 10m wind speed
- 10m wind gust
- 925hPa wind speed
- 2m temperature

4.4 Maps for Ensemble Mean and Spread

These maps are part of the GLAMEPS standard postprocessing, and following parameters are made available (figure 4):

- 500hPa height
- 850hPa temperature
- mean sea level pressure
- 2m temperature

4.5 GLAMEPS-o-GRAMS

These are meteograms for several hundred locations from within the common GLAMEPS domain (1). The GLAMEPS-o-GRAMS show time series for following parameters:

- 2m temperature
- 10m wind
- 10m wind gust
- 3h precipitation

An example is shown in figure 5.

5 Further Development

Following items are under consideration for further developing the GLAMEPS production system:

- calibrated products (currently being tested)
- from 8km to 5km grid
- CAPE SVs in HirEPS (Sibbo van der Veen)
- analysis perturbation inflation for HirEPS and possibly also for AloEPS

- AloEPS model perturbation based on diffusion
- AloEPS surface perturbations

Not necessarily all of these items may enter into a GLAMEPS version 3 system. However, the calibrated products already have been developed (Thomas Nipen 2015), and are being tested within the real-time framework. Similarly, analysis perturbation inflation for HirEPS is being tested within a parallel test suite.

6 References

- Alex Deckmyn, 2014: Introducing GLAMEPS-v2, a presentaion at the ALADIN/HIRLAM 24th Workshop/All-Staff Meeting 2014, Bucharest
- GLAMEPS Production System: Operations Status, <https://glameps.org>
- GLAMEPS Production System: Technical Information for Users, <https://hirlam.org/trac/wiki/GlamepsProductionUser>
- Thomas Nipen, 2015: Calibrated probabilistic forecasts from GLAMEPS, 25th ALADIN Workshop & HIRLAM All Staff Meeting 2015, Helsingør

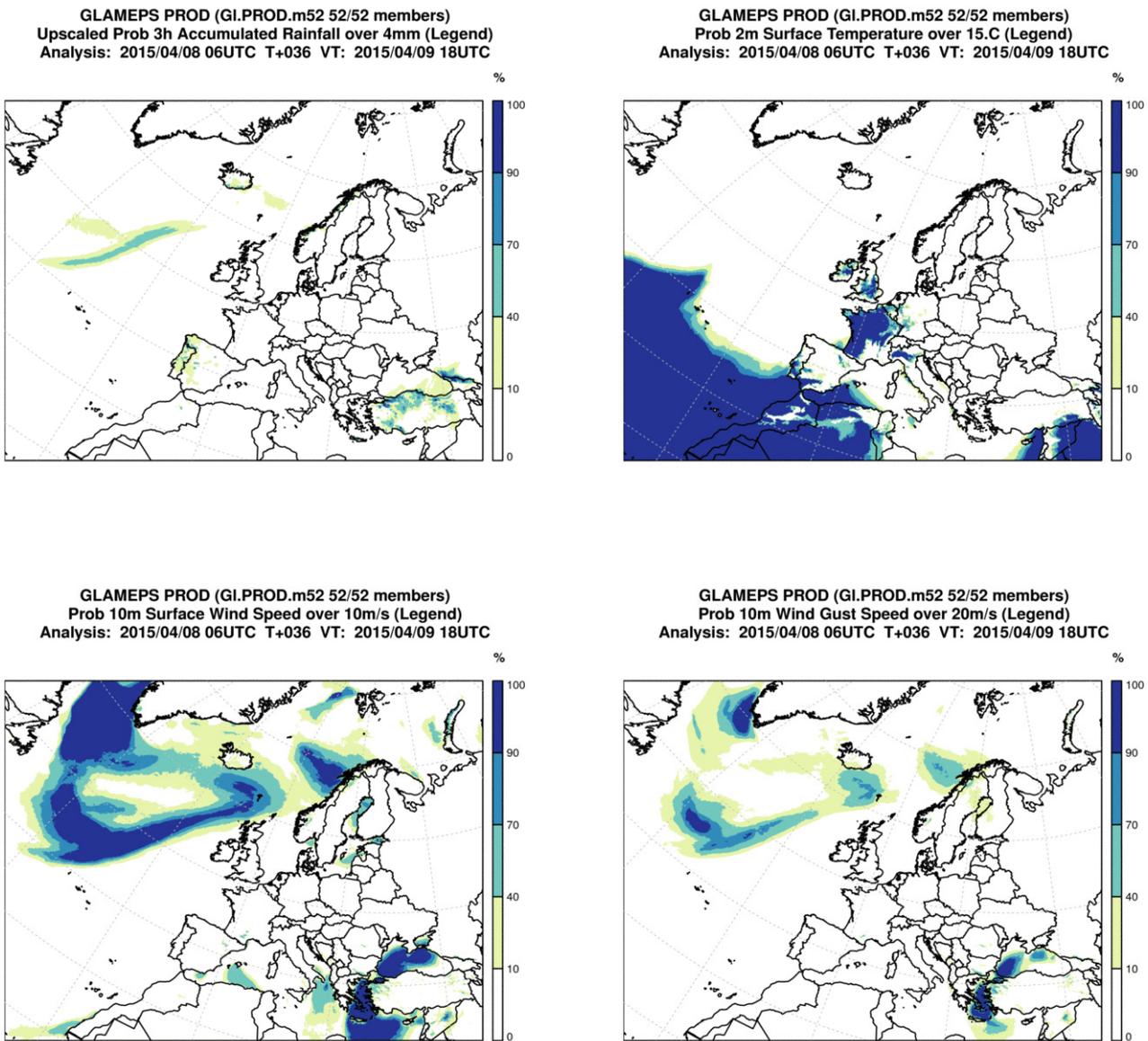
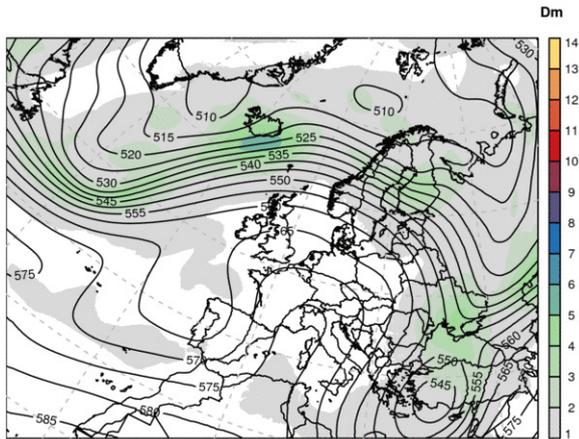
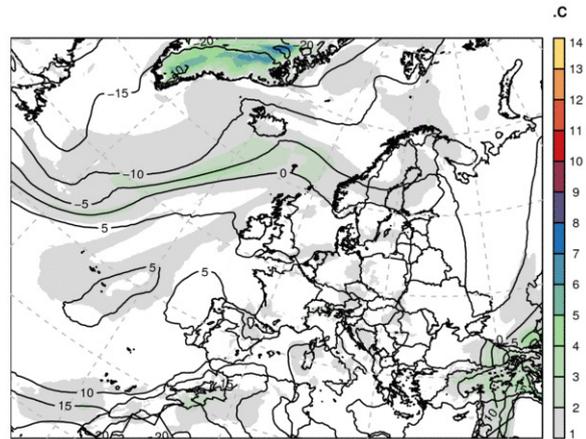


Figure 3: Example probability maps from GLAMEPS. These maps are available on <https://glameps.org>.

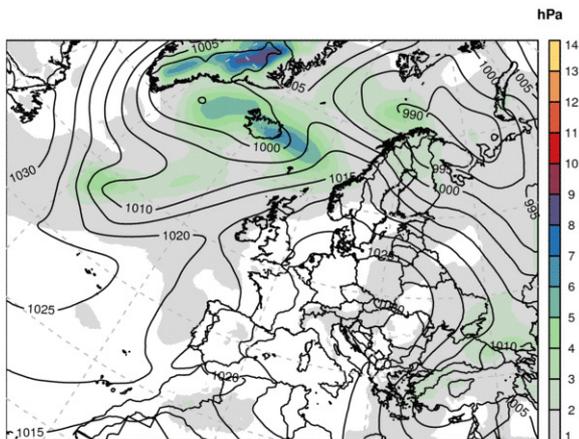
GLAMEPS PROD (GI.PROD.m52 52/52 members)
Spread & Emean 500hPa Height (Dm) (Legend)
Analysis: 2015/04/08 06UTC T+036 VT: 2015/04/09 18UTC



GLAMEPS PROD (GI.PROD.m52 52/52 members)
Spread & Emean 850hPa Temperature (.C) (Legend)
Analysis: 2015/04/08 06UTC T+036 VT: 2015/04/09 18UTC



GLAMEPS PROD (GI.PROD.m52 52/52 members)
Spread & Emean Mean Sea Level Pressure (hPa) (Legend)
Analysis: 2015/04/08 06UTC T+036 VT: 2015/04/09 18UTC



GLAMEPS PROD (GI.PROD.m52 52/52 members)
Spread & Emean 2m Surface Temperature (.C) (Legend)
Analysis: 2015/04/08 06UTC T+036 VT: 2015/04/09 18UTC

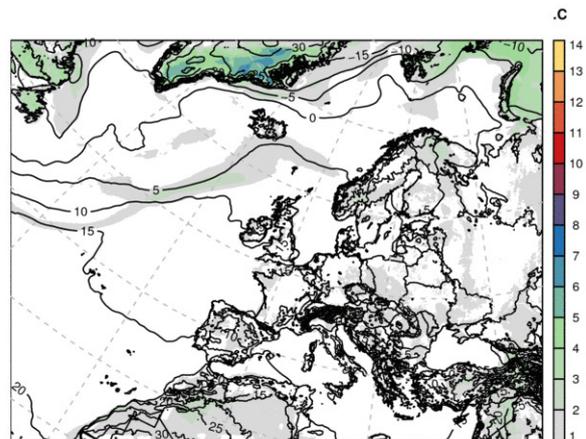


Figure 4: Example maps for ensemble mean and spread from GLAMEPS. These maps are available on <https://glameps.org>.

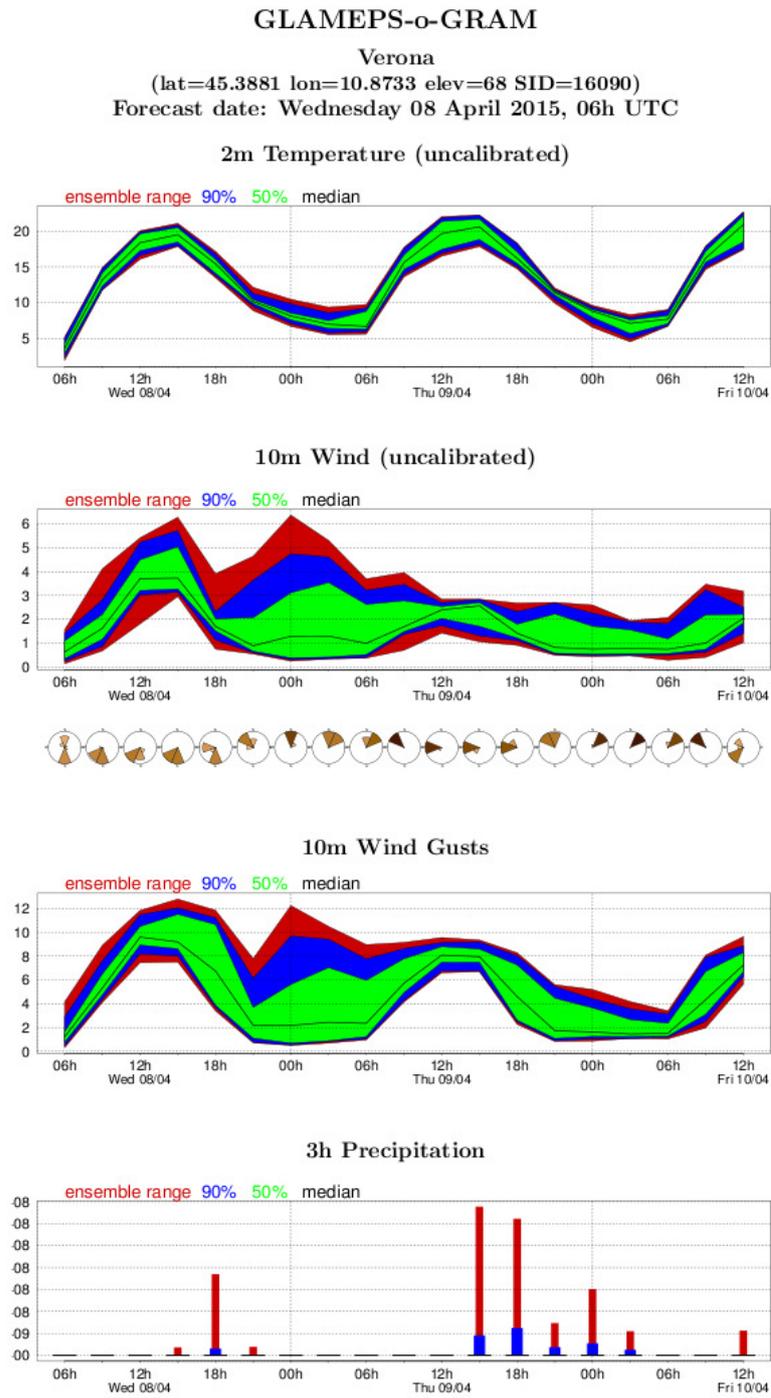


Figure 5: Example GLAMEPS-o-GRAM. The full range of locations is available on <https://glameps.org>.

A comparison between the default HarmonEPS perturbations and random field perturbations

Henrik Feddersen
Danish Meteorological Institute

1 Introduction

The default HarmonEPS perturbations of initial and lateral boundary conditions are based on members of the ECMWF global ensemble prediction system (EPS). An EC-perturbation, i.e. the difference between a member of the ECMWF-EPS and the ECMWF control run, is added to a Harmonie control run (after interpolation to the Harmonie grid). In particular, the interpolated EC-initial condition perturbation is added to a Harmonie analysis.

This approach is a natural extension of the way in which we run deterministic limited-area models. But the purpose of an ensemble prediction system is to describe the forecast uncertainty, roughly speaking given by the ensemble spread, and the ensemble spread of the global ECMWF-EPS does not necessarily give a correct description of the uncertainty in the short forecast range for a limited area.

In order to assess the value of using ECMWF-EPS perturbations for HarmonEPS, a baseline setup using “random field” perturbations is implemented, and two sets of ensemble forecasts for a 19-day period in June 2012 are verified and compared.

Random field perturbations where a perturbation is determined as a scaled difference between two randomly selected historical model states, were used by Magnusson et al. (2009) as a baseline method against which they could compare more advanced methods. They found in their examples that for a global ensemble the simple random field perturbation approach was performing just as well as both singular vector perturbations (Mureau et al., 1993) and a combination of bred vectors (Toth and Kalnay, 1997) and an ensemble transform (Bishop and Toth, 1999).

2 Perturbations of initial and lateral boundary conditions

A random field perturbation is determined as the scaled difference between two randomly selected model states $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$, from the operational ECMWF high-resolution runs, interpolated to the Harmonie grid, i.e.

$$\mathbf{p}_{\text{rf}}(t) = \alpha [\mathbf{r}_1(t) - \mathbf{r}_2(t)], \quad (1)$$

where t is the forecast length, and α is a scaling parameter that controls the size of the perturbation and is chosen so that all initial perturbations have the same amplitude in terms of the total energy norm.

The two random fields, \mathbf{r}_1 and \mathbf{r}_2 are selected from the same time of the day as the Harmonie control run to which the perturbation is added or subtracted; they are from the same season (in practice: the control date ± 15 days), they are at least five days apart and at least ten days away from the control date and year. The random fields are selected from the period 26 Jan 2010 – 24 Jun 2013 when the resolution of the operational ECMWF high-resolution model was fixed at T1279L91.

The random field perturbations are added (or subtracted) to the control run throughout the forecast, i.e. they provide both initial condition perturbations, $\mathbf{p}_{\text{rf}}(0)$, and lateral boundary conditions, $\mathbf{p}_{\text{rf}}(t)$, $t > 0$.

For comparison, the default HarmonEPS perturbations are given by

$$\mathbf{p}_{\text{def}}(t) = \beta [\mathbf{x}_m(t) - \mathbf{x}_0(t)], \quad (2)$$

where $\mathbf{x}_m(t)$ is the model state of member m of the ECMWF-EPS, interpolated to the Harmonie grid; member 0 is the ECMWF-EPS control run, and β is a scaling constant which is equal to 1. However, by increasing β the HarmonEPS ensemble can be inflated and the ensemble spread increased.

3 HarmonEPS experiments

For the period 10-28 June 2012 ECMWF ensemble forecasts have been rerun,¹ and model states of the members, including the control run, have been made available in the MARS archive. The control run and the first eight perturbed members are in the following used to provide initial and lateral boundary conditions as described above for a HarmonEPS experiment where all nine members use identical configuration (HARMONIE cy38h11, including AROME physics and SURFEX surface scheme), so that differences between the members are only due to differences between the initial and lateral boundary condition perturbations.

A parallel nine-member HarmonEPS experiment has also been made based on a control run plus four random field perturbations that are added to and subtracted from the control run.

The geographical domain that has been used for both experiments is the default HarmonEPS domain shown in Fig. 1. The horizontal resolution is 2.5km with 65 vertical levels, the ensembles are run twice per day (at 0 and 12UTC) with a forecast length of 36h, while the control forecasts, including analyses are run four times per day.

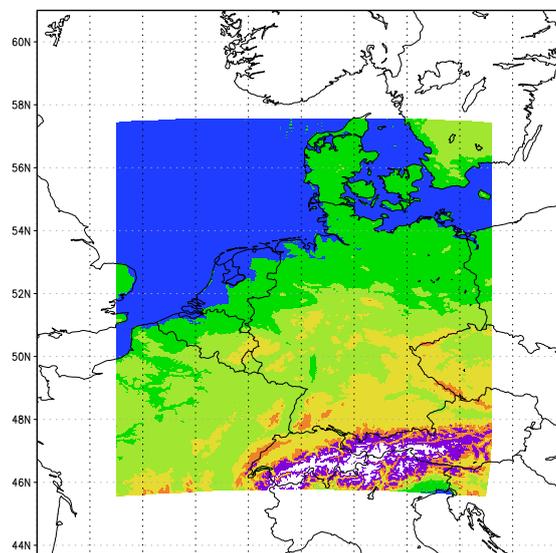


Figure 1: HarmonEPS domain used in experiments.

¹See <https://software.ecmwf.int/wiki/display/LAMEPS/Data+for+testing+resolution+sensitivity>

The ECMWF-EPS perturbations are already scaled such that their total energy over the HarmonEPS domain at the forecast initial time is approximately the same for all perturbations. The total energy norm (in units of Jm^{-2}) is a measure of the magnitude of the perturbation, i.e. the distance between the perturbed member and the control run; it is defined as (Errico et al., 2004)

$$E_{\text{tot}} = \frac{1}{2gN_{\text{hor}}} \sum_{i,j,k} (u'_{ijk}{}^2 + v'_{ijk}{}^2) \Delta p_{ijk} + \frac{c_p}{2gN_{\text{hor}}} \sum_{i,j,k} \frac{1}{T_k} T'_{ijk}{}^2 \Delta p_{ijk} + \frac{1}{gN_{\text{hor}} \bar{p}_s} \sum_{i,j} \Phi_{s_{ij}} p'_{s_{ij}}{}^2, \quad (3)$$

where u' , v' , T' and p'_s are wind, temperature and surface pressure perturbations; i , j and k are horizontal and vertical grid point indices, N_{hor} is the number of horizontal grid points, Δp_{ijk} is the pressure difference between vertical layers $k + \frac{1}{2}$ and $k - \frac{1}{2}$, Φ_s is the surface geopotential, and an overbar denotes a horizontal average.

The scaling parameter α in (1) is chosen such that the total energy of the initial random field perturbations becomes approximately the same as the total energy of the ECMWF-EPS perturbations.

4 Perturbation growth

Figure 2 shows a typical example of perturbation growth from the last HarmonEPS forecast in the experiment period. The total energy of the eight HarmonEPS members based on ECMWF-EPS perturbations is shown in red, and the total energy of the eight members based on random field perturbations in blue. During the first 12 hours of the forecast the perturbation growth is similar for the two ensembles; from around 12 hours into the forecast the perturbation growth levels off when the lateral boundary conditions start to dominate the ensemble divergence.

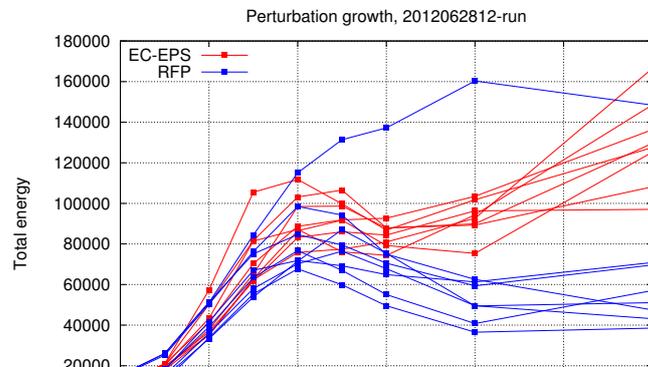


Figure 2: Total energy of HarmonEPS forecast perturbations. Red: Eight perturbed members based on ECMWF-EPS perturbations. Blue: Eight perturbed members based on random field perturbations.

The total energy of the ECMWF-EPS based perturbations is generally greater than the total energy of the random field perturbations in the last part of the forecast. This is due to the fact that the spread generally increases with forecast length between the ECMWF-EPS members, while it on average remains the same between two independent ECMWF deterministic forecasts (as used for the random field perturbations), see Fig. 3. If we let the scaling parameter α increase with forecast length the total energy of the random field perturbations will also grow with forecast length.

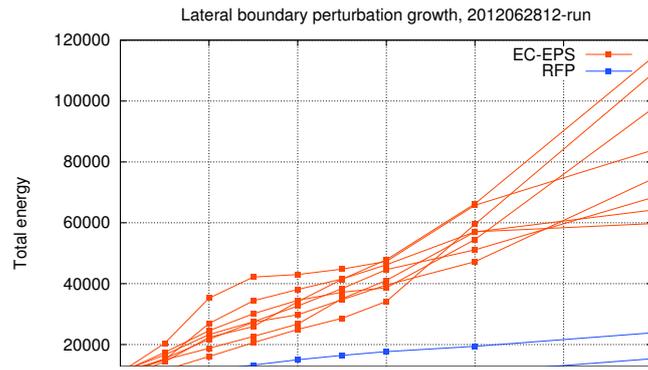


Figure 3: As Fig. 2, but for total energy of fields used to provide lateral boundary perturbations.

5 Forecast verification

The nine-member ensemble forecasts for the 10-28 June 2012 period are verified against available observations from approximately 200 German synop stations. In terms of ensemble spread and RMS error of the ensemble mean (Fig. 4) the differences between the two methods are marginal for 2m temperature, 10m wind speed and precipitation, while the spread in mean sea level pressure tend to follow that of the ECMWF ensemble (cf. Fig. 3).

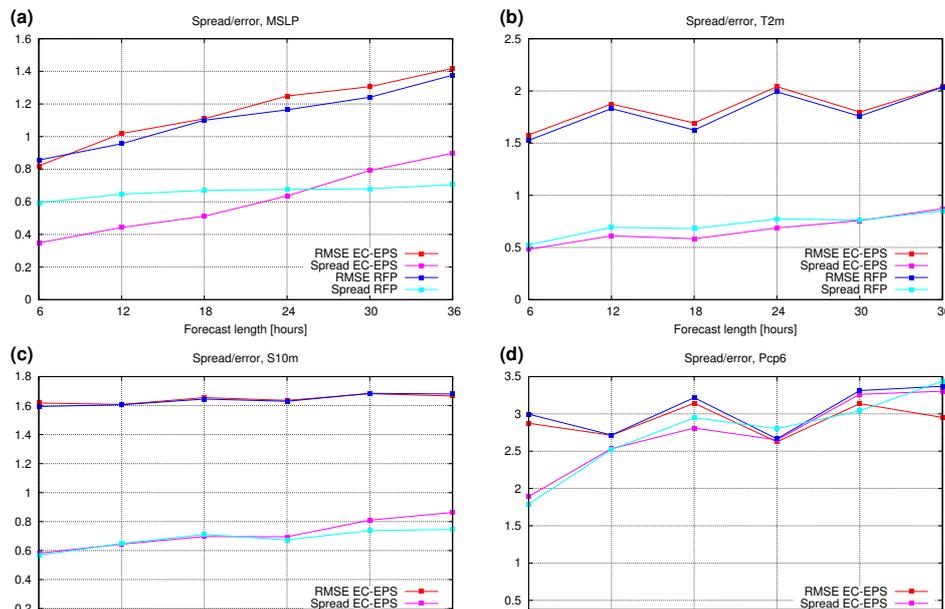


Figure 4: Ensemble spread and RMS error of ensemble mean for German synop observations for (a) mean sea level pressure, (b) 2m temperature, (c) 10m wind speed and (d) precipitation accumulated over 6h.

Rank histograms for 18h forecasts (Fig. 5) show practically no difference for 10m wind speed and 6h precipitation, while the rank histograms for MSLP and 2m temperature are slightly flatter (i.e. better) for the random field perturbations.

In terms of reliability we find no significant differences between the two methods as illustrated in Fig. 6.

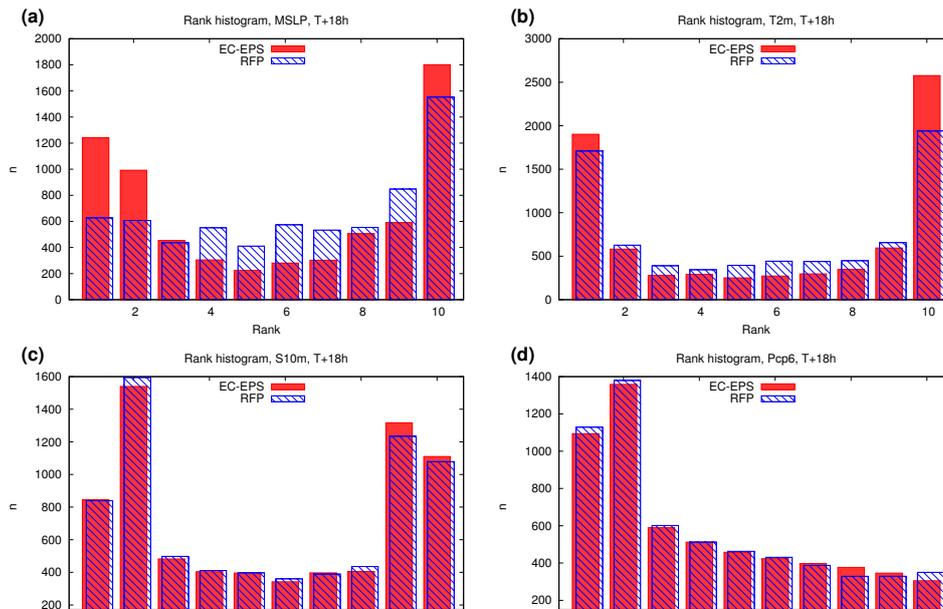


Figure 5: Rank histograms for 18h forecasts for German synop observations for (a) mean sea level pressure, (b) 2m temperature, (c) 10m wind speed and (d) precipitation accumulated over 6h. Red histograms for perturbations based on ECMWF-EPS; blue for random field perturbations.

6 Forecast example

On 12 June 2012 weakly forced convection in Northern Germany and Southern Denmark resulted in heavy showers with 15-30mm rainfall in a couple of hours as indicated in Fig. 7 that shows observations of 12h rainfall. The rainfall intensity was generally underpredicted by DMI's operational HIRLAM models, including both the 5km ensemble system and (especially) the 3km deterministic model.

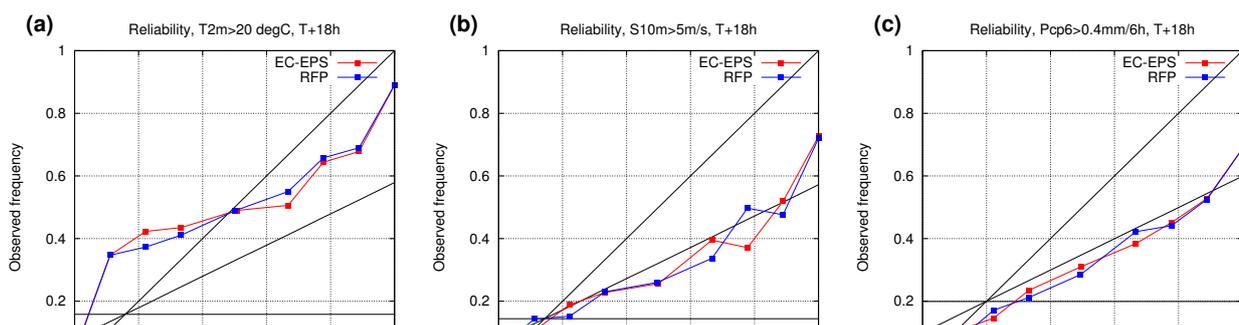


Figure 6: Reliability diagrams for 18h forecasts for German synop observations for (a) 2m temperature > 20°C, (b) 10m wind speed > 5ms⁻¹ and (c) precipitation > 0.4mm/6h. Red curves for perturbations based on ECMWF-EPS; blue for random field perturbations.

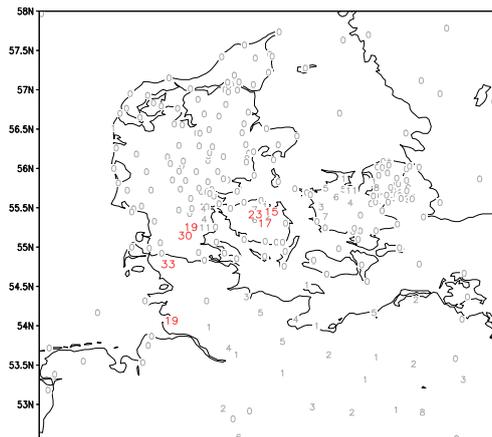


Figure 7: Observed 12h precipitation. Values of 15mm/12h and more are highlighted.

2012061200+15h, 3h accum. precip
Valid on 2012061215

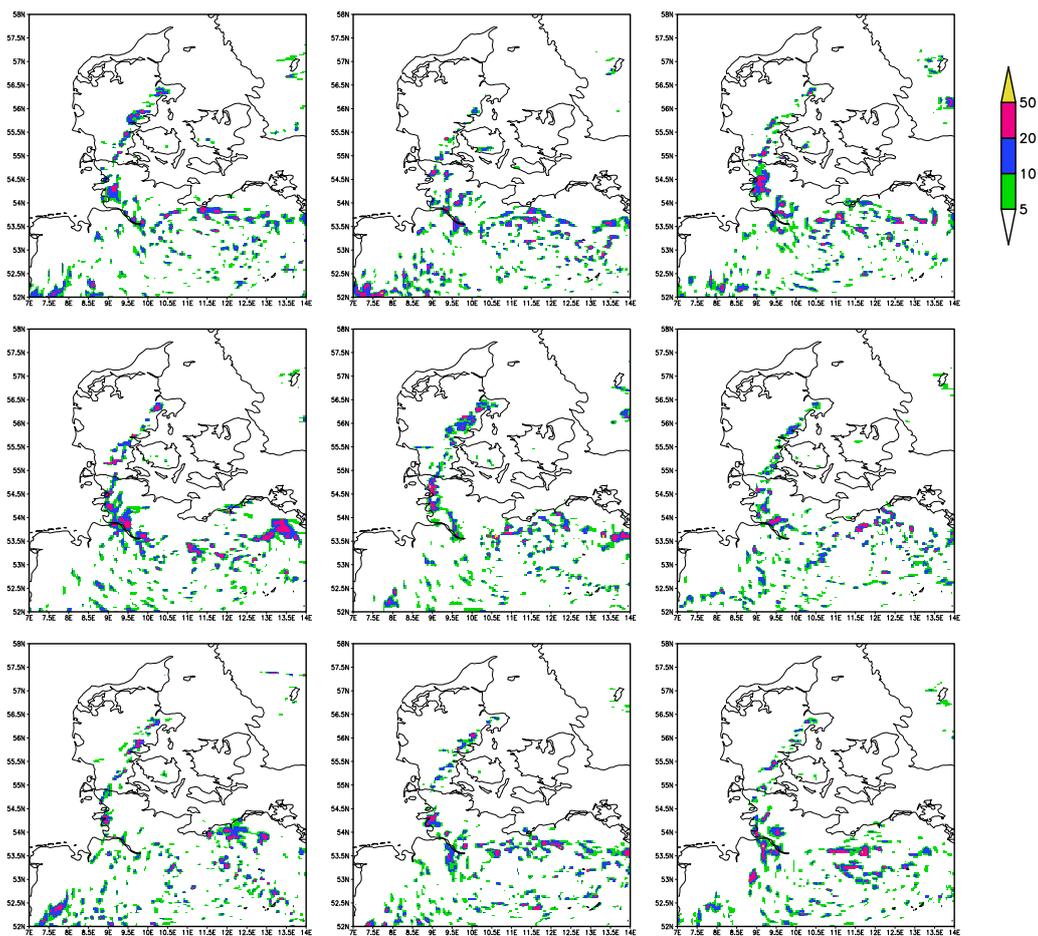
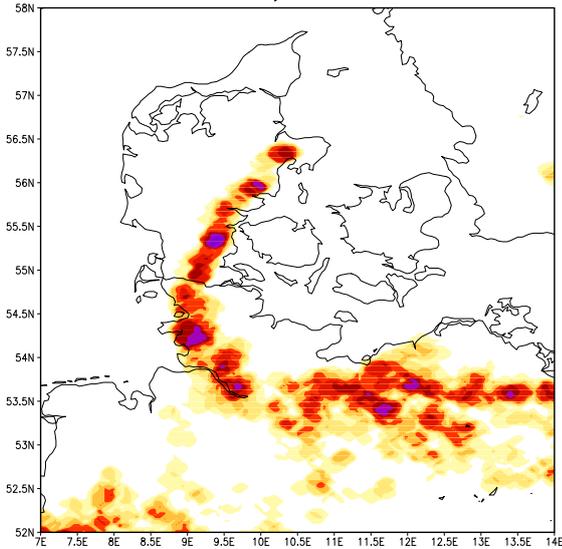


Figure 8: HarmonEPS 15h ensemble rainfall [mm/3h] forecast using ECMWF-EPS perturbations.

2012061200+15h ec perturbations: Prob($P_{cp}>20\text{mm}/3\text{h}$)
Valid on Tuesday 12 Jun 15:00 UTC



2012061200+15h rfp perturbations: Prob($P_{cp}>20\text{mm}/3\text{h}$)
Valid on Tuesday 12 Jun 15:00 UTC

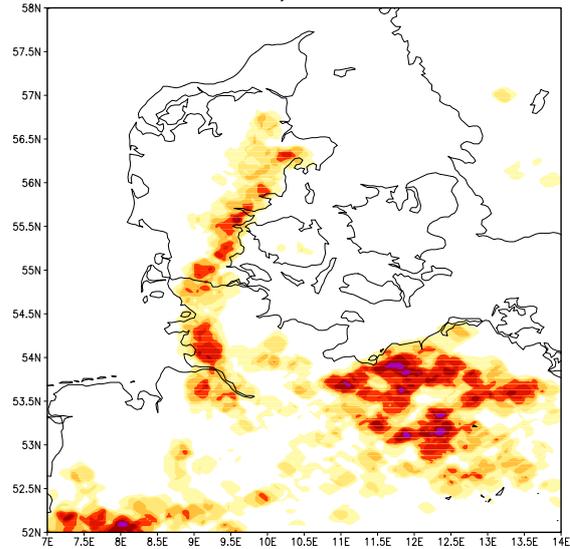


Figure 9: HarmonEPS 15h ensemble rainfall forecast probabilities of exceeding 20mm/3h somewhere within a 10km radius; left: ECMWF-EPS perturbations, right: random field perturbations.

A postage stamp plot of a HarmonEPS 15h ensemble forecast using the default ECMWF-EPS perturbations is shown in Fig. 8. It is evident that all nine ensemble members produce local showers in Northern Germany and Denmark. A similar postage stamp can be plotted from a HarmonEPS forecast based on random field perturbations (not shown).

Corresponding probability maps are shown in Fig. 9 where probabilities of exceeding 20mm/3h somewhere within a 10km radius are plotted (Feddersen, 2011). The two perturbation methods yield similar probability maps with the EC-perturbations giving slightly larger probabilities (possibly due to smaller ensemble spread), especially in Denmark. It is not possible to say whether one or the other probability forecast is better, and a period of only 19 days is too short to make a statistically significant verification of intense rainfall.

7 Concluding remarks

The differences between HarmonEPS ensemble forecasts based on ECMWF-EPS perturbations and random field perturbations are remarkably small for the test period. The fact that the simple random field approach yields ensemble forecasts of the same quality as the ensemble forecasts based on ECMWF-EPS perturbations suggests that using the latter “blindly” is not optimal.

So, should we use ECMWF-EPS perturbations at all for HarmonEPS? And if we do – is there a way to improve the perturbations?

An alternative perturbation method is the so-called scaled lagged average forecasting (SLAF; Hou et al., 2001) which uses previous forecasts that are available “for free” to compute perturbations. However, with access to the recently established ECMWF ensemble boundary condition programme the use of ECMWF-EPS perturbations has also become more attractive than it was previously.

In terms of improving the perturbations a very simple way to increase ensemble spread is to inflate the perturbations by increasing α in Eq. (1) or β in Eq. (2) which does indeed seem to improve the probabilistic verification scores (not shown). A more advanced approach that has been implemented successfully for global bred vectors, is an ensemble transform whereby the initial perturbations are orthogonalized in order to increase the initial ensemble spread (Wei et al., 2008; Keller et al., 2010). A similar ensemble transform might be applied to limited-area ensembles.

8 References

- Bishop, C. H. and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.
- Errico, R. M., K. Raeder and M. Ehrendorfer, 2004: Singular vectors for moisture-measuring norms. *Quart. J. Roy. Meteor. Soc.* **130**, 963–987.
- Feddersen, H., 2011: Predicting intense precipitation using upscaled, high-resolution ensemble forecasts. *HIRLAM Newsletter* **58**, 56–62.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.* **129**, 73–91.
- Keller, J. D., A. Hense, L. Kornblueh and A. Rhodin, 2010: On the orthogonalization of bred vectors. *Wea. Forecasting* **25**, 1219–1234.
- Magnusson, L., J. Nycander and E. Källén, 2009: Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus* **61A**, 194–209.
- Mureau, R., F. Molteni and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.* **119**, 299–323.
- Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.
- Wei, M., Z. Toth, R. Wobus and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast systems. *Tellus* **60A**, 62–79.

PROFORCE – Bridging of Probabilistic Forecasts and Civil Protection

Clemens Wastl, Martin Suklitsch, Yong Wang, Andre Simon



1 Introduction

Severe weather often causes natural disasters which lead to significant economic damages and sometimes also loss of lives. In order to increase the safety of the general public and infrastructure accurate and reliable weather forecasts are essential. Only then it is possible to facilitate the allocation of aid to those in need in time. However, weather forecasts are subject to inherent uncertainties: The atmospheric system is highly chaotic, and numerical weather prediction models face certain limitations and contain assumptions about the system they describe. In order to quantify such uncertainties ensemble prediction systems (EPS) have been established several years ago and are now widely used within the meteorological community. But only recently uncertainties started being communicated to the end users as well. For decision makers and civil protection agencies alike the access to information about uncertainties in the weather forecast has great potential to significantly improve their decision making process. In the framework of the EU funded project PROFORCE (Bridging of Probabilistic Forecasts and Civil Protection, 12/2013 – 11/2015) an easy to use seamless probabilistic forecasting system has been developed. It is aimed at improving preparedness and decision making procedures within civil protection agencies. The PROFORCE project is a co-operation between the national weather services of Austria and Hungary and civil protection partners from both countries.

2 Seamless forecasting system

The main goal of PROFORCE is a seamless probabilistic forecasting system including nowcasting, short-range and medium-range forecasting up to several days ahead which is especially customized to the needs of civil protection applications. The main characteristic of this seamless system is the probabilistic feature containing information about the uncertainty and predictability of severe weather events. This information should support civil protection agencies to optimize their decision making procedure in terms of preparedness and awareness, and therefore help them to better protect the society and environment from impacts of severe weather. The seamless probabilistic forecasting system developed within PROFORCE merges four types of ensemble systems (Table 1). The individual EPSs are based on different models and run with different horizontal resolutions and forecast ranges. All of them have their own role in the final seamless product depending on the nature of the predicted weather event (convective, large scale) and the forecast range.

Table 1: Short description of the models used within the seamless system of PROFORCE.

Model	Members	Hor. Resolution	Forecasting Range	Runs/day
ECMWF EPS	50 + 1	32 km	15 days	2
LAEF ALADIN-HUEPS	16 + 1 (HU 10 + 1)	11 km (HU 8 km)	3 days	2
AROME EPS	16 + 1 (HU 10 + 1)	2.5 km	30 hours (HU 36 hours)	8
Ensemble INCA	16 + 1 (HU 10 + 1)	1 km	12 hours	24

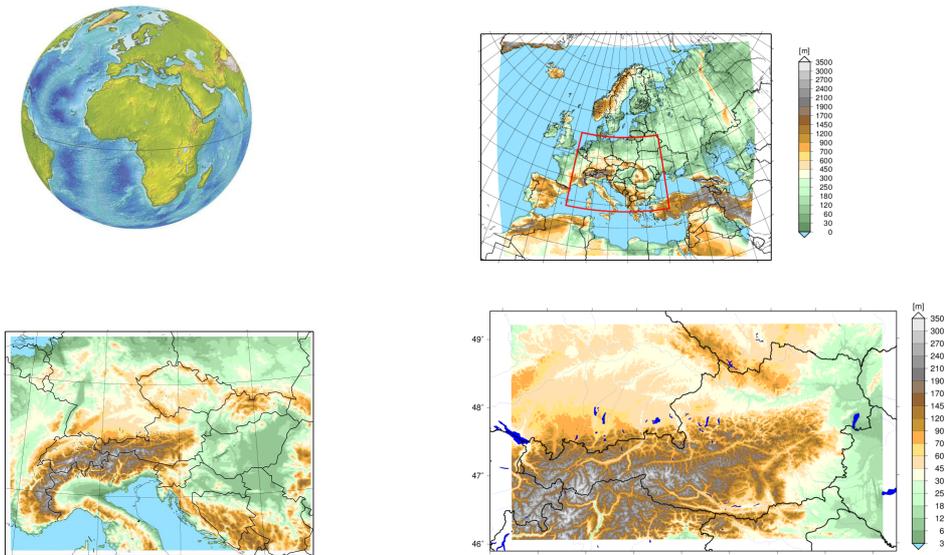


Figure 1: Domains of the models used within the seamless system of PROFORCE. ECMWF-EPS (upper left), LAEF (upper right), AROME-EPS (lower left), Ensemble-INCA (lower right). The images refer to the Austrian implementation, Hungarian domains are different for AROME-EPS and Ensemble INCA.

In the seamless system the output of the individual EPSs is visualized as probability maps (showing the probability of exceedance of certain thresholds) as well as meteograms/plumes and put on a purpose-built web portal that can be accessed by civil protection authorities and disaster management. The uncertainty information is illustrated in a very clear and simple way to foster a fast decision making process. Another main topic of PROFORCE are user tailored training sessions for civil protection to use the new system in an optimal and efficient way.

The seamless system has been tested during several case studies (large storm events, convective events), and its applicability for an operational use in civil protection agencies has been approved. The feedback of end-users and beneficiaries was wholly positive. By using the new and innovative system prevention and preparedness actions could be launched much earlier and more specific compared to the classical deterministic weather forecast. E.g. due to the combination of different models and scales, first prewarnings of the storm Niklas (March 30 – April 2, 2015) were issued already several days before the beginning of the event. Furthermore, the additional information about reliability and uncertainty facilitates an improved assessment of the potentially dangerous weather situation.

In the end the whole warning chain from official authorities down to local decision makers and finally also the general public can take benefit out of this project.

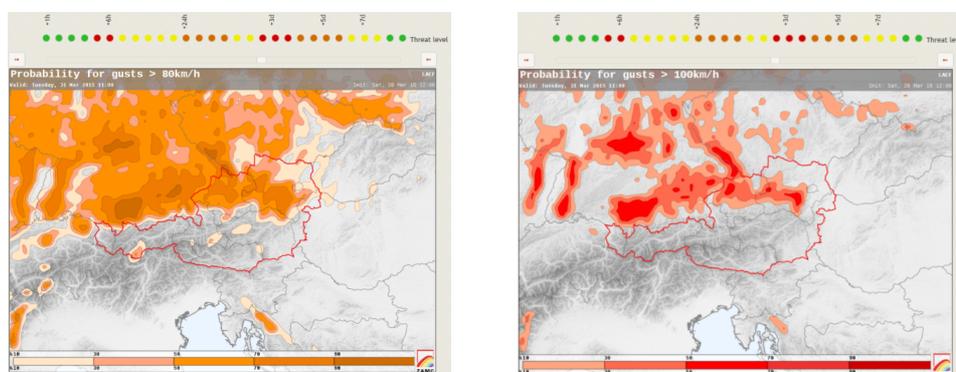


Figure 2: Example of a LAEF forecast map from the PROFORCE portal: Probability of wind gusts > 80km/h (left) and > 100km/h (right) for the test case of storm Niklas on 31 March 2005, 18 UTC.

The new Numerical Weather Forecast Verification System in Tunisia

Wafa Khalfaoui

1 Introduction

A new Numerical Weather Forecast Verification System (NWFVS) was established in order to investigate the performance and the quality of the numerical prediction models products in the National Institute of Meteorology of Tunisia. The present document aims to:

- give the state of the progress of the Numerical Weather Forecast Verification System (NWFVS)
- present the performance of ALADIN-Tunisia operational suite in the 1st trimester of 2015

2 NWFVS General Framework

The code is scripted in Shell / Fortran / Python / R. The verification concerned ALADIN model which is running operationally in the National Institute of Meteorology. The forecast data are compared and verified against GFS analysis and stations data. Figure 1 schematizes the flow diagram of the Verification process with the different steps and Inputs/Outputs.

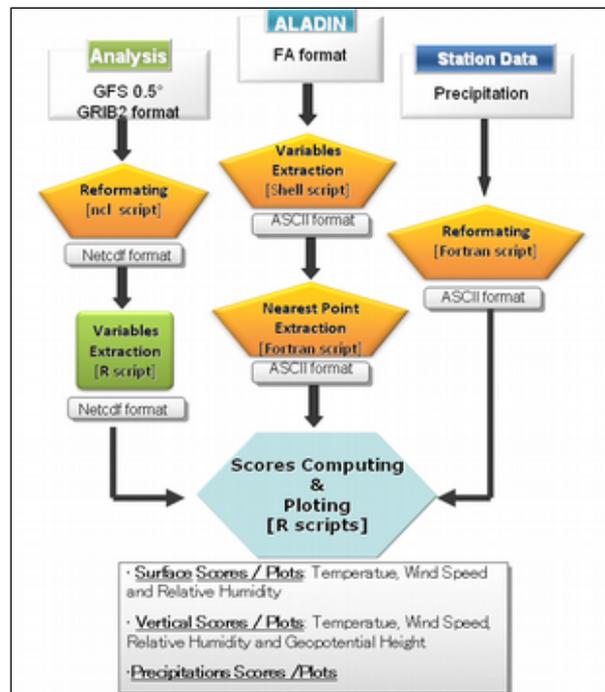


Figure 1: Numerical Weather Forecast Verification System Process flow Diagram

Verification domain lies between 7° and 12° west-east and 30° and 38° south-north to cover all the Tunisian territory. Figure 2 presents the verification domain and the projection of GFS analysis grid on red points. A zonal classification based on the geographic distribution of the stations was performed to identify different verification areas. The current classification relies on the criteria of spatial homogeneity (Figure 2). Different control areas based on homogeneous fixed verification grid at different scales (0.1°, 0.5°, 1°) are used on the verification procedure.

Forecast outputs are verified against GFS analysis and synoptic stations data. Temperature, wind speed, relative humidity and geopotential height are processed with a time step equal to 6 hours at 00H, 06H, 12H and 18H in the run of 00H every day for the surface and the vertical verification. The verification of the precipitation is performed for the 24 hour accumulated amount from 06 UTC to 06 UTC. Giving the limited number of the synoptic stations and their spatial sparseness (Figure 2), combined with the frequent unavailability of the 6 hour and 12h accumulated precipitation data, we are constrained for the moment to deal with 24 hour accumulated precipitation collected from nearly 300 pluviometric stations (synoptic, climatologic,...) (Figure 2).

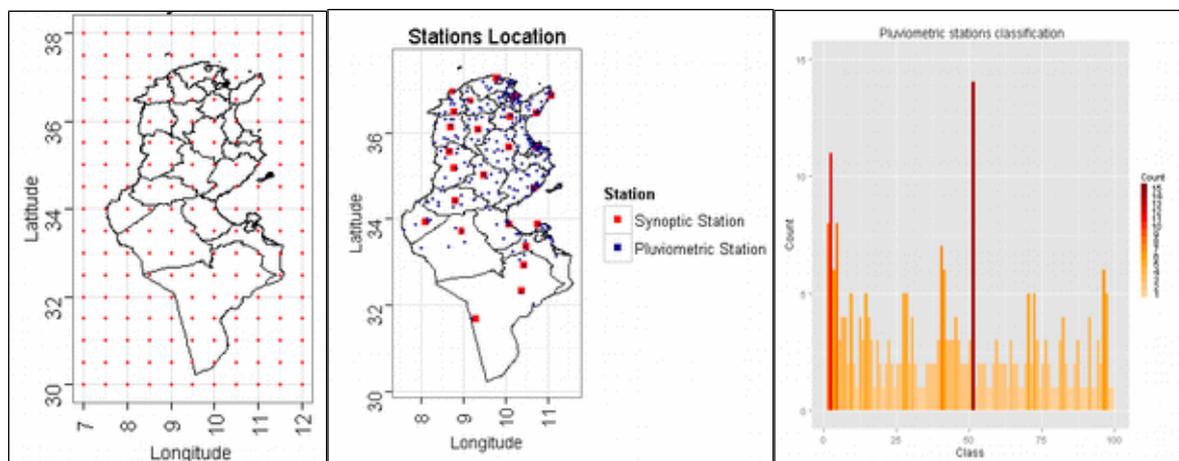


Figure 2: Synoptic / pluviometric stations (right box), Analysis grid point on the verification domain (middle box) and Pluviometric Station classification

As the ALADIN-Tunisia assimilation procedure is under implementation, and due to the operational constraint, GFS analysis seems to be suitable for the NWFVS. The model and analysis characteristics are summarized on the table 1 below.

Table 1: Model and analysis characteristics

Model	Analysis
Limited Area Model ALADIN Operational suite: cycle 29 Coupled to global model ARPEGE 12.5 km spatial resolution 3H time step	GFS model 0.5 ° spatial resolution 0H, 6H, 12H and 18H Grib2 file

3 Verification Results of the 1st trimester of 2015

3.1 Temperature (2 m)

A negative bias around 1.25 degree (Figure 3.a), especially for the 00H and 06H, indicates an over all tendency for under-estimating the temperature mainly on January and February (Figure 4.a). Temperature on 12H and 18H presents a bias almost equal to zero. The RMSE (Figure 3.d) and the standard deviation (Figure 3.c) are presenting the same signal as their amplitudes are smaller on 12H and 18H than on 00H and 06H reflecting smaller error amplitude and spread after 12H. This difficulty of the model to estimate the temperature on the first hours of the run is due to the influence of the initial and boundary conditions taken from the global model ARPEGE especially that we don't have yet an assimilation suite. After 12 hours, ALADIN seems to be able to move away from the initial conditions and calibrate the temperature. As January and February were characterized by the temperature fluctuation (compared to Mars where the temperature was more stable), the error amplitude (Figure 4.b) and spread (Figure 4.c) are more pronounced during January and February than on Mars.

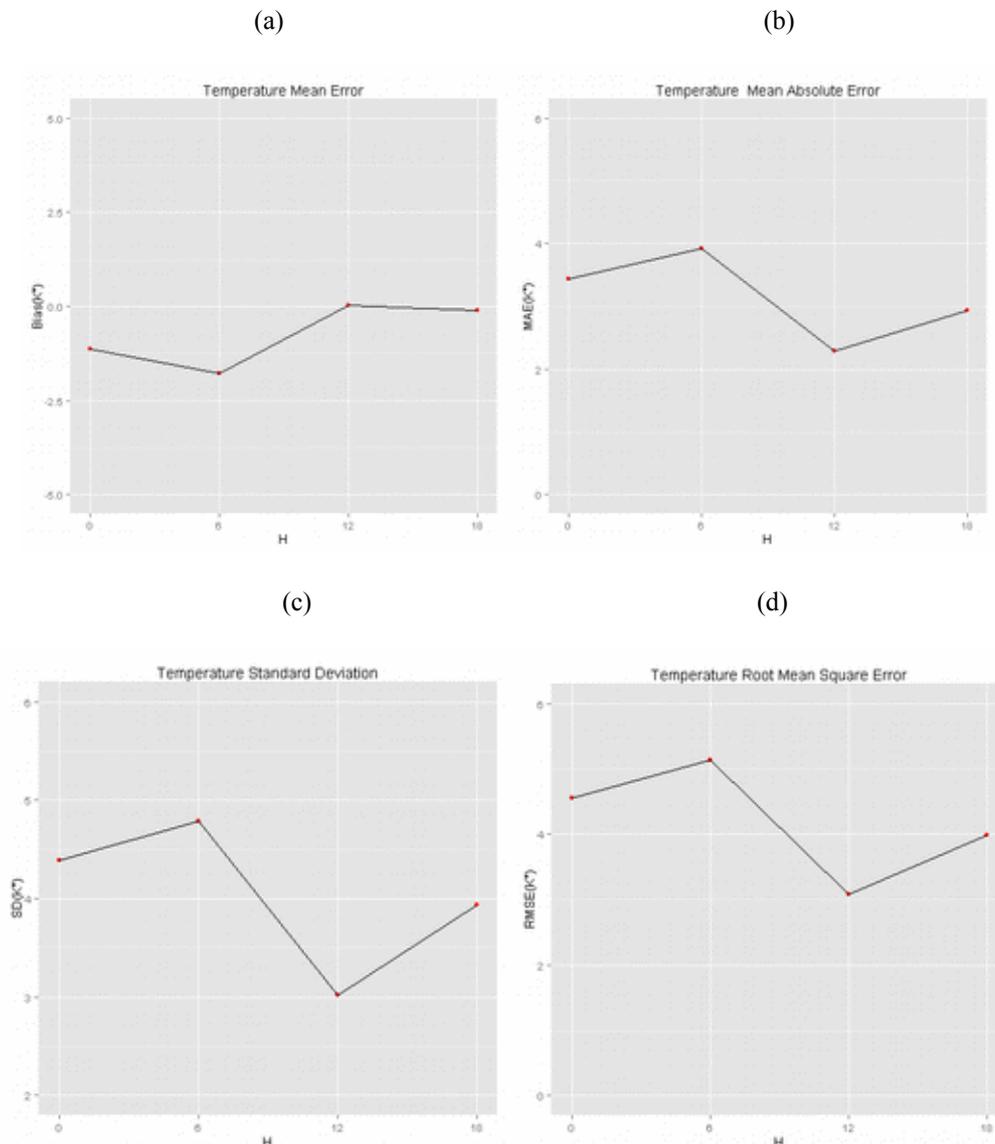
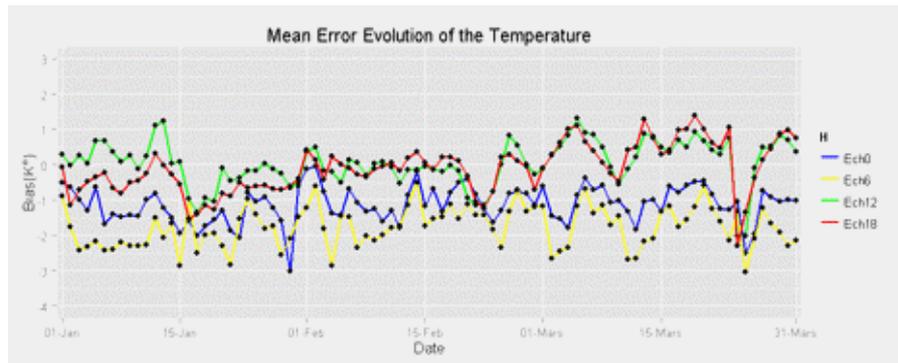
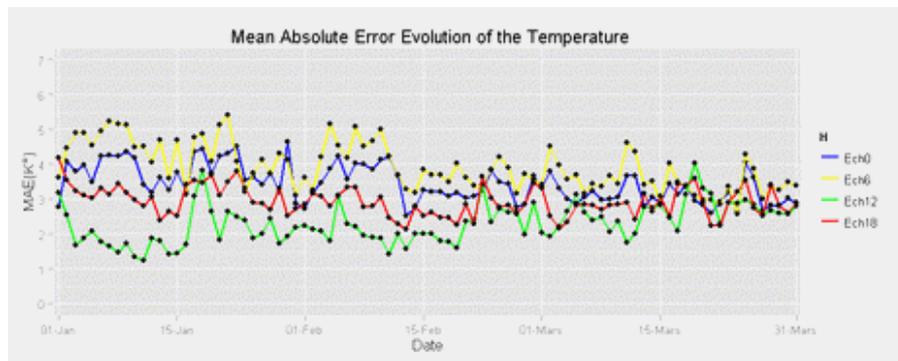


Figure 3: Temperature scores; bias (a), mean absolute error (b), standard deviation (c) and root mean square error (d); of ALADIN vs Analysis at 0H, 6H, 12H and 18H during the 1st trimester of 2015

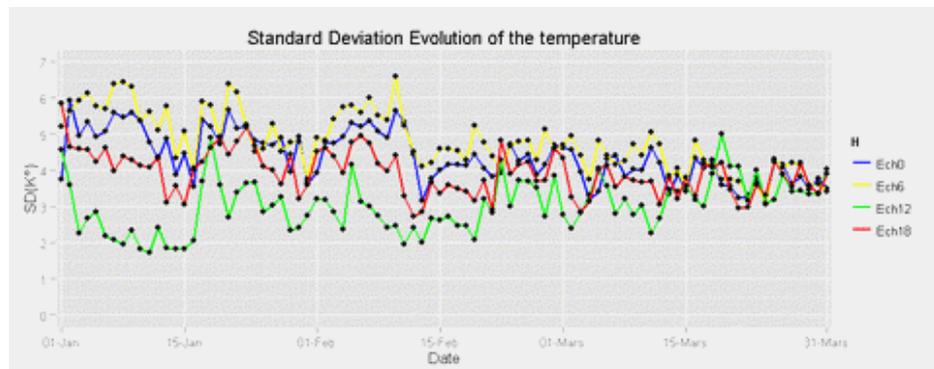
(a)



(b)



(c)



(d)

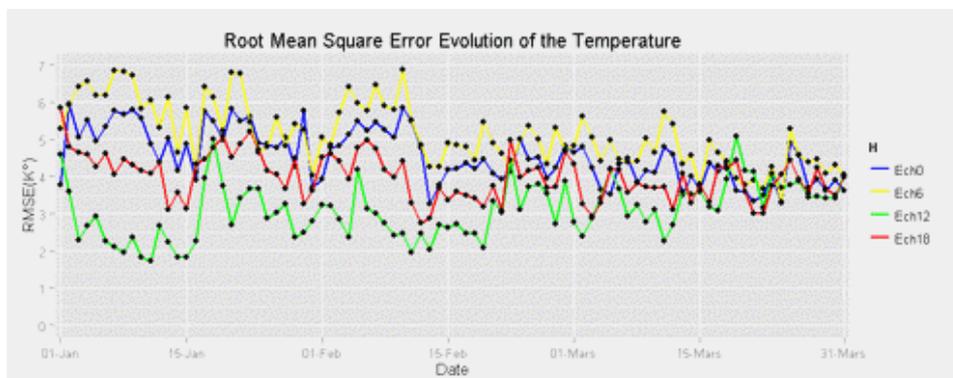


Figure 4: Evolution of the temperature scores; bias (a), mean absolute error (b), standard deviation (c) and root mean square error (d) ; of ALADIN vs analysis at 0H, 6H, 12H and 18H during the 1st trimester of 2015

3. 2 Vertical Profile

The vertical profiles of the wind speed, the temperature, the relative humidity and the geopotential height RMSE (Figure 5) are presenting the same general pattern: the RMSE maximum is recorded near the tropopause with a general tendency for ALADIN to underestimate (Figure 6) the variables (except for the temperature that was overestimated in some pressure levels).

The wind speed RMSE maximum lies between 300 hPa and 200 hPa (Figure 5.d) which corresponds to the current jet zone. The relative humidity RMSE maximum lies mostly between 950 hPa and 850 hPa (Figure 5.b) which corresponds to the low clouds development zone. The geopotential height presents a large negative error (Figure 6.c) especially near the tropopause indicating a general tendency to underestimate the geopotential height. This underestimation of the geopotential leads to the amplification of the low pressure centers which is common feature of the meteorological models during the winter season.

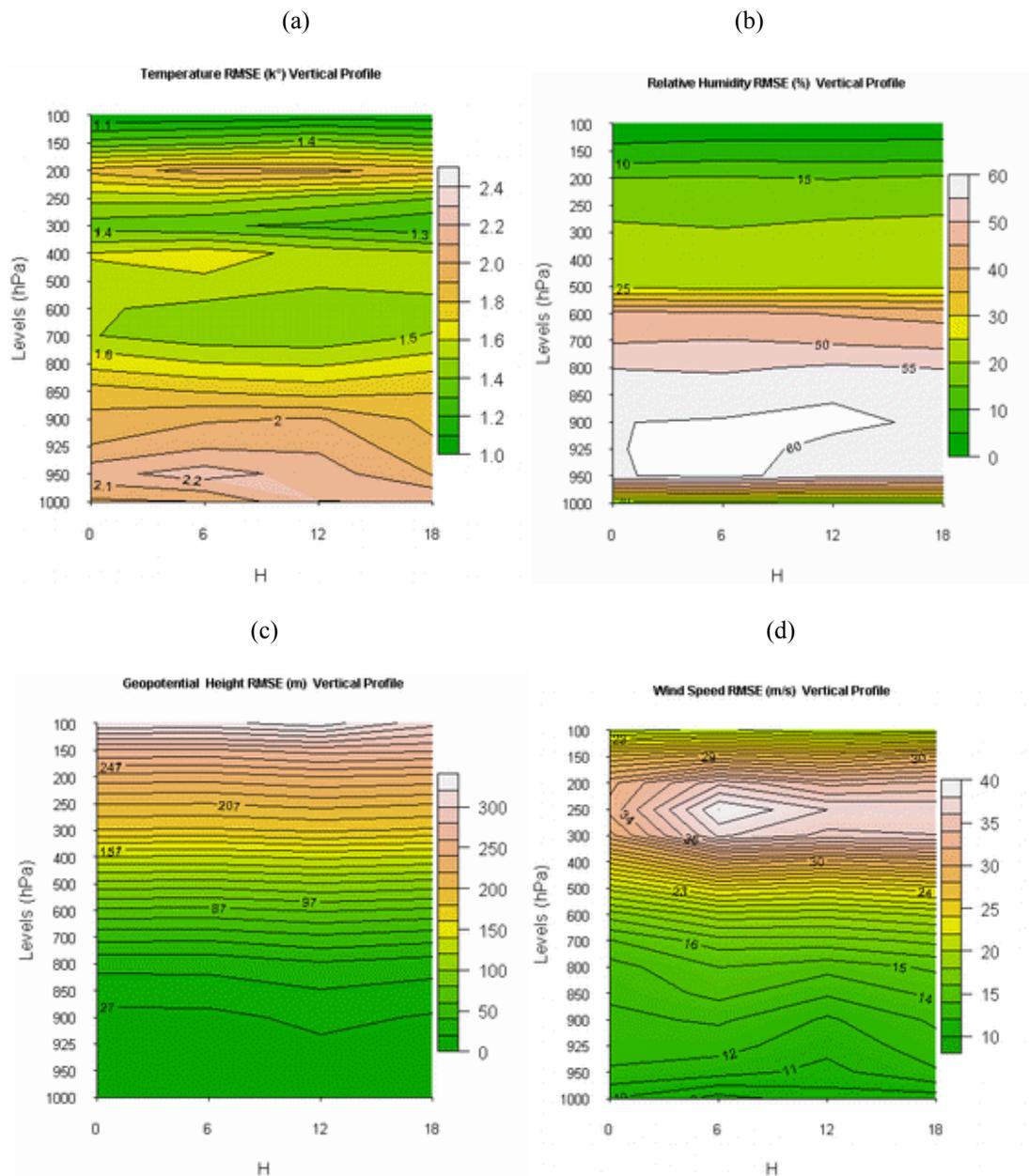


Figure 5: Vertical RMSE profile of the temperature (a), the relative humidity (b), the geopotential height (c) and the wind speed (d) ; of ALADIN vs analysis at 0H, 6H, 12H and 18H during the 1st trimester of 2015

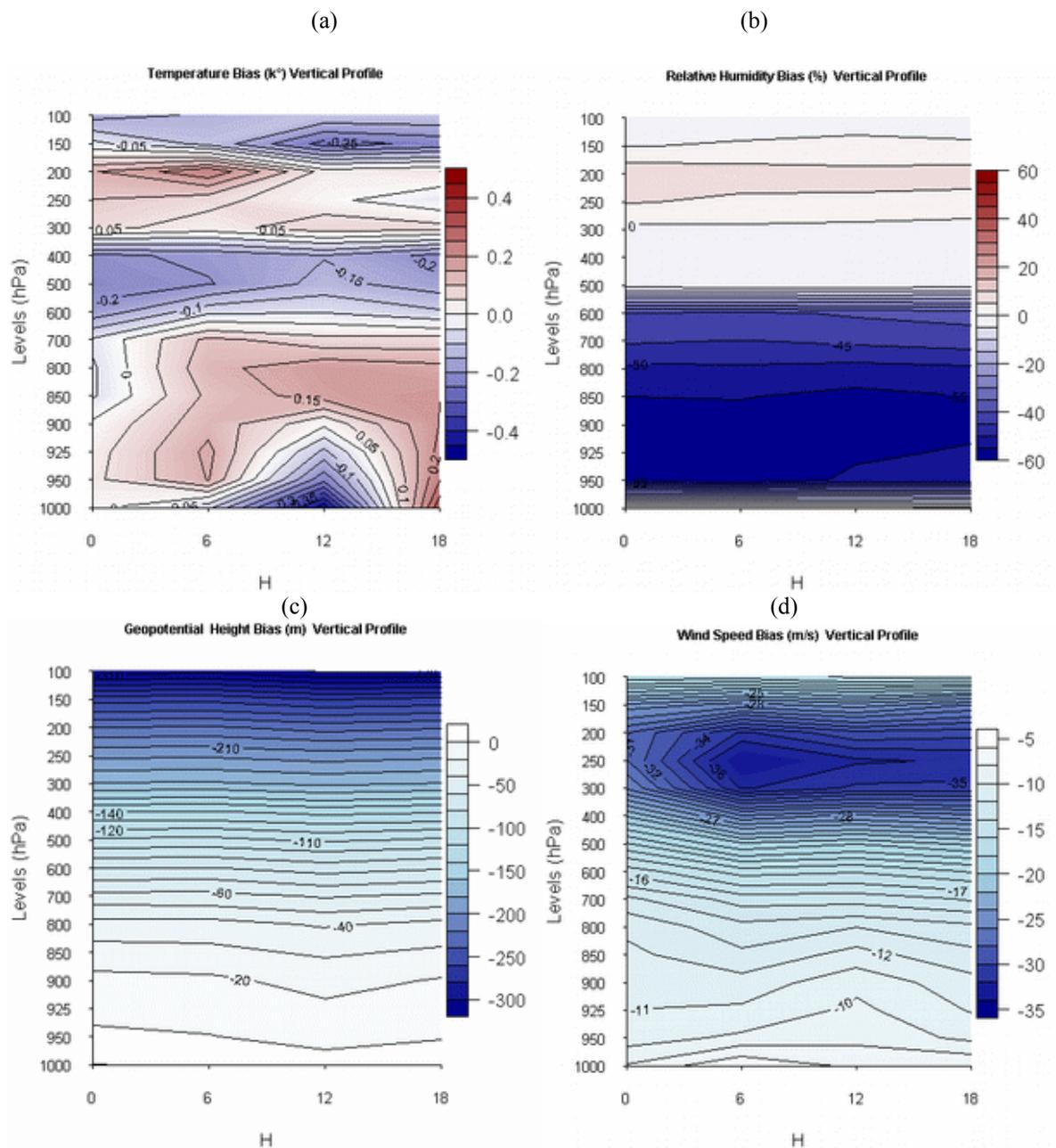


Figure 6: Vertical Bias profile of the temperature (a), the relative humidity (b), the geopotential height (c) and the wind speed (d); of ALADIN vs analysis at 0H, 6H, 12H and 18H during the 1st trimester of 2015

3.3 Precipitation

January was marked by a deficit of 30% compared to the normal. In February, several severe precipitation events were recorded and generated important flood events especially in the North West region. Actually, in February, precipitation increased by 120% compared to the normal. Mars was characterized by a rise of 27% compared to the normal mainly on the north and the Center East region. ALADIN underestimated the precipitation in the North West region that was marked by heavy precipitation in February and overestimated slightly the precipitation in the South and Center regions (Figure 7). When examining the different scores for the different thresholds (0.1mm, 1mm, 5mm, 10mm, 15mm and 20mm), an overall tendency of the model to deteriorate the scores as the threshold increases is noticeable. This deterioration reflects the difficulty of ALADIN to detect and predict amount of severe precipitation events (Figures 7, 8 and 9).

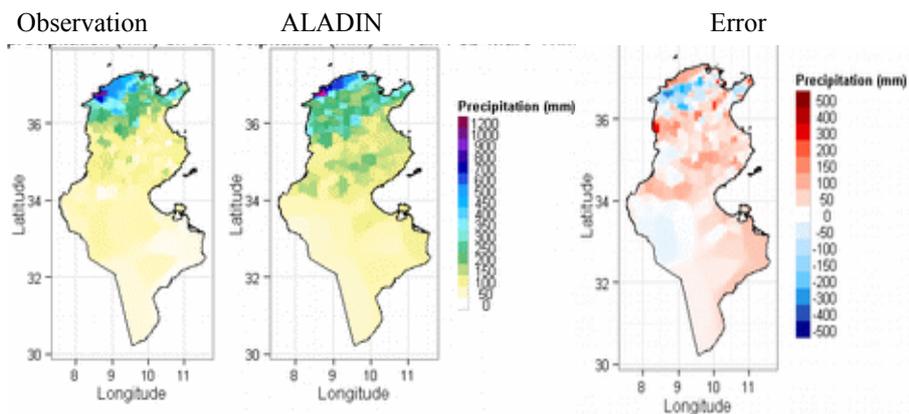


Figure 7: Accumulated precipitation in the 1st trimester 2015 of, respectively from up left to down right, Observation, ALADIN and their Error (Forecast-Observation)

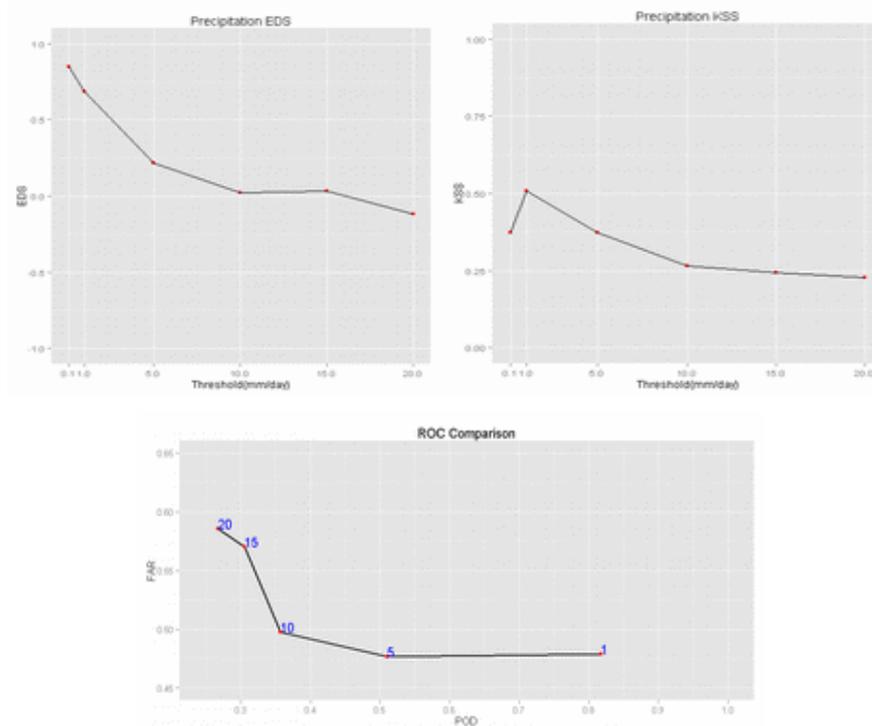


Figure 8: ALADIN 24H accumulated precipitation scores; respectively from top left to down right; EDS, KSS and experimental ROC for 0.1mm, 1mm, 5mm, 10mm, 15mm and 20mm thresholds during 1st trimester 2015

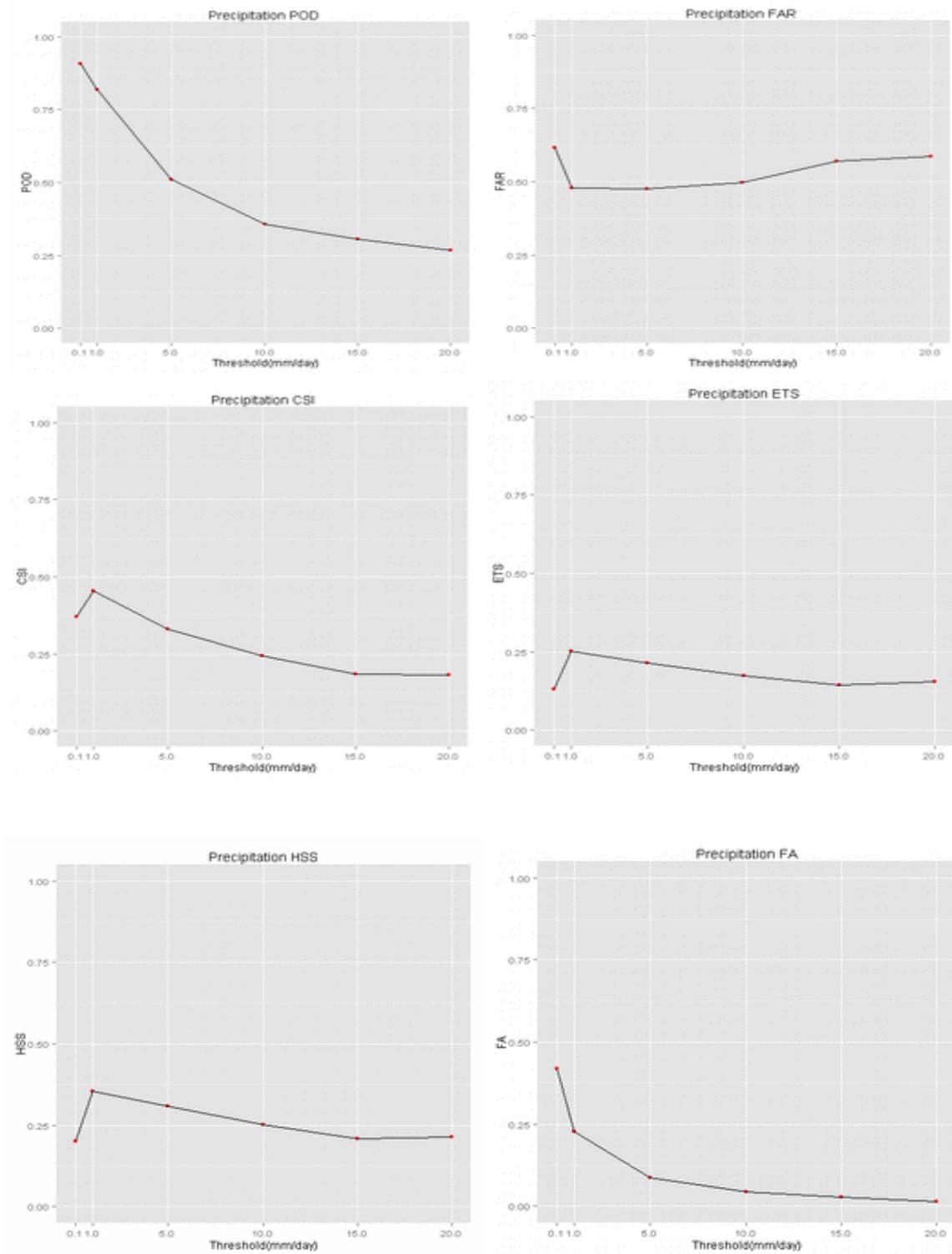


Figure 9: ALADIN 24H accumulated precipitation scores; respectively from top left to down right; POD, FAR, Bias, CSI, ETS, HSS and FA for 0.1mm, 1mm, 5mm, 10mm, 15mm and 20mm thresholds during 1st trimester 2015

4 Summary & State of progress of the NWFVS

Step	What is done	On going Work
Verification Events & O/F Data Base construction	Series of events (AROME 2.5 km cycle 38 / ALADIN 12.5 km operational cycle 29 / ALADIN 7.5 km cycle 38 / HARMONIE 2.5 km cycle 38) is already set.	Models (AROME/ALADIN) Comparison Study for Mars 2014
Observation/Forecast (O/F) Data base Matching	<p>Nearest point method :</p> <ul style="list-style-type: none"> ▪ from models grid point to station ▪ from station to models grid point ▪ Interpolation (p-nearest points average) <p>Control the observation data quality (screening)</p>	<ul style="list-style-type: none"> ▪ Downscaling / Upscaling
Verification Area	<ul style="list-style-type: none"> ▪ Classification based on the geographic / climatological distribution ▪ A homogeneous fixed verification grid (per variable) 	Higher resolution
Scoring	<ul style="list-style-type: none"> ▪ explorative methods ▪ numerical descriptive measures ▪ Categorical verification scores from contingency tables ▪ Spatial Verification Method : Fuzzy Neighborhood Method (FSS) 	<ul style="list-style-type: none"> ▪ Spatial verification methods (MODE, SAL) ▪ Distributions-oriented Approach ▪ Signal Detection Theory and The ROC
Verification Newsletter	First edition ready	Adding further scores

Can the compiler make the difference ?

Ryad El Khatib

1 Introduction

Each computer vendor provides a machine with a preferred compiler, chosen for its performance after application benchmarking. Unfortunately we have to trust the vendors, and eventually it is hardly possible for us end-users to compare the different compilers on a more “operationnal” basis. Indeed, the compilers may not be all available on a given machine for portability reasons, or just because an alternative compiler would not have been sold.

However the recent acquisition of a Cray computer at ECMWF is giving to our community a unique opportunity to compare simultaneously three compilers : Cray, Intel and Gfortran, from the point of view of operational forecasting, not from the point of view of vendors benchmarks.

2 Conditions of comparisons

Source code

The source code was the regular cycle 41. It was chosen because it is the first cycle which has been ported to work with the Cray compiler by ECMWF. In other words, it is the first cycle expected to be stable with the three compilers : Cray, Intel and Gfortran.

Note that earlier attempts to port the source code to work with the Cray compiler at FMI had been partly unsuccessful, because of the instability of the compiler itself.

Constraints on scientific results

Of course the scientific results have to be valid, but different compilers will return different results. The scientific validity was checked with respect to a reference set of spectral norms, exactly like for regular vendors benchmarks.

Besides, the scientific had to be bitwise-repeatable and bitwise-reproducible ; bitwise-repeatable means that the same experiment rerun several times should return exactly the same results. Bitwise-repeatable means that the scientific results must not be sensitive to any change of computing resources : the same experiment rerun several times with different numbers of MPI tasks, OPEN-MP threads, or “array slicing “ (better known as the NPROMA parameter, or the vector length for vector computers, or the cache-blocking size for the cache-based computers) should return exactly the same results. That last constraint needs first that the code be bug-free on this aspect, and then it could reveal a problem of threads or tasks synchronization, or memory alignment in the compiler. This property is particularly important for the source-code developers.

Background libraries

Linear algebra libraries (BLAS, LAPACK) are important for an optimal performance of the applications. Therefore the custom for each compiler vendor is to provide its own optimized scientific library. Cray works with its Cray SCI library. Intel can provide its MKL library but its usage revealed at a time a problem of bitwise-reproducibility for what we need to do. However the open-source library OpenBLAS and the library LAPACK from Netlib provide together an equivalent performance ; therefore they have been used for the tests with the Intel compiler and with Gfortran. It would have been more proper to use these open-source libraries for the tests with the Cray compiler too, but eventually the tests showed that they were not performing significantly less than the Cray library.

For all the compilers, the MPI library used was the Cray one (MPI Cray 6.3.1), compiled for each compiler. The mpi driver used was *mpiauto* : a wrapper on top of the mpi library driver developed at Météo-France. *mpiauto* proved to perform as well, or better (with hyperthreading, on Intel) than the Cray native driver *aprun*.

Unfortunately with Gfortran, there have been sporadic issues with MPI hanging up in the model I/O server at the very end of the runs. Hopefully a more recent version of the MPI library could fix it.

Compiler versions and options

For each compiler, the most recent version (at the time of the study) was used :

- Cray 8.2.7
- Intel 15.0
- GCC 4.9

The options used had to be equivalent from one compiler to another, and as optimal as possible with respect to the bitwise repeatability/reproducibility constraint described above :

- Cray :
-hflex_mp=conservative -hadd_paren -hfp1
- Intel :
-align array64byte -fp-model source -ftz -O2 -xAVX -finline-functions -finline-limit=500
- Gfortran :
-fstack-arrays -O2 -fvectorize *with* -ffast-math *at link time*

Remarks :

-O3 was not used for Intel, after it showed detrimental effects of the scientific forecasts at Météo-France. Anyway, it had no significant impact on the computational performance.

-O3 might be used for Gfortran ; however it has not been evaluated yet and it would need like for Intel, a careful scientific validation. As an alternative, one could evaluate the opportunity to activate sub-options of -O3 that would not modify the results.

Miscellaneous

Hyperthreading was disabled, and no paging module was used for the Cray compiler. The LUSTRE file system was using the same striping (4) for all the experiments.

Applications tested

The forecasting model AROME was tested, in a configuration close to what it is today in the operations at Météo-France :

- Resolution : 1.3 km L90 ; forecast range : H12
- I/O server activated (3 nodes)
- post-processing on-line
- Coupling frequency reduced, mostly for commodity
- Computational resources used in the model part : 200 nodes, as 400 MPI tasks x 12 OPEN-MP threads.

Remark : The global model ARPEGE was also tested, in a configuration close to what it is today in the operations at Météo-France ; but its results, close to the ones of AROME, will not be discussed in this paper.

3 File system troubleshooting

The file system LUSTRE installed at ECMWF appeared to be insufficiently reliable with small files. As a matter of fact, it makes non-repeatable elapse times from one job to another, with a variability of 5% to 10%. Therefore several developments have been conducted to alleviate this issue, the underlying idea being that any file should be read from, or written out to a user-defined directory :

- No output listing
- statistic file “ifs.stat”, DDH files, IO-server listings written out to specific directories on NFS
- Real time coupling script “atcp”, post-processing server control file “cnt3_wait” read from a specific directory on NFS

Though there have been still sporadic issues with the file system afterwards, and whatever the sizes of the files involved were, these modifications enabled to stabilize the elapse time with a reliability of +/- 2%.

4 Compilers performance on top of cycle 41

The picture below (Figure 1) shows for each compiler its relative performance with respect to a reference which is the mean performance of all three compilers.

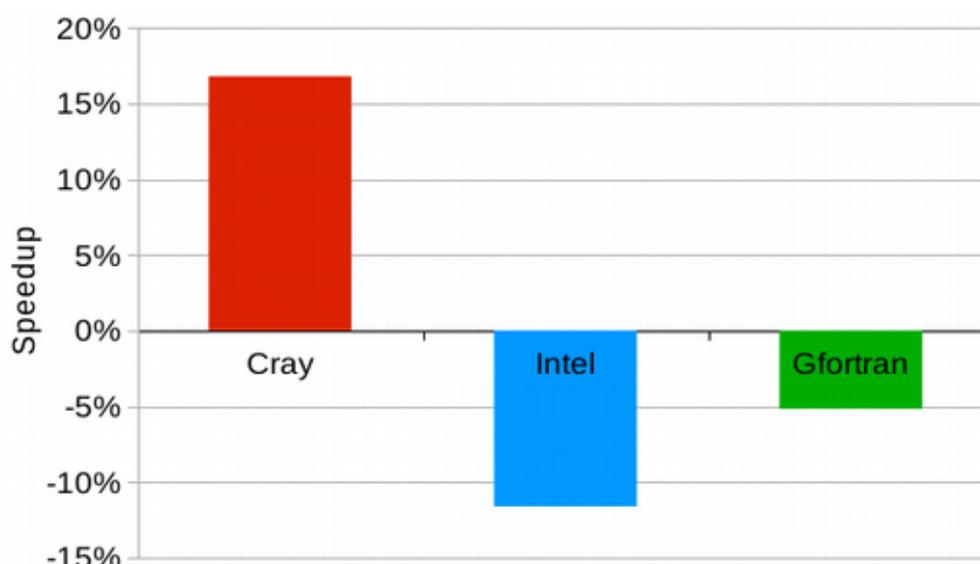


Figure 1 : Performance of each compilers (relatively to the mean performance of them) for AROME (cycle 41)

It is astonishing to observe that the Intel compiler is even slower (- 7%) than the open-source Gfortran compiler. One should feel also much concerned by the difference of performance between the Cray compiler and the Intel compiler (30 %), in the scope of future computers replacements.

Therefore a detailed analysis of the differences of performance has been conducted, at the level of the source code subroutines, and even the source code loops, with the help of the compilers reports.

5 Analysis of the differences of performance

Array syntax

The array syntax is a coding style such that loops indices are implicit. Example :

```
A(:,:)=B(:,:)*C(:,:)-D(:,:)/Z(:,:)
```

The equivalent “traditional” Fortran 77 style is :

```
DO JK=1,NK
  DO JI=1, NJ
    A(JI,JK)=B( JI,JK)*C( JI,JK) -D(JI,JK)/Z(JI,JK)
  ENDDO
ENDDO
```

The array syntax looks like a seducing coding style for the scientists because it make apparently simpler source code, and one could even claim that array syntax makes vector loops. This is true ; however this style is detrimental to the performance because it prevents memory cache re-use. Memory cache is a very fast but small memory where data can be stored temporary, and then re-used without the need to re-access the “normal” memory (exactly called : Random Access Memory). For instance, consider the consecutive two loops below :

```
A1(:)=B(:)*C1(:)
A2(:)=B(:)*C2(:)
```

Because of the array syntax, each line is a separate DO loop for the compiler, and then each element of the array B will be fetched twice from memory to be put in the cache for computation. If the same code is written in the traditional Fortran 77 style :

```
DO J=1,N
  A1(J)=B(J)*C1(J)
  A2(J)=B(J)*C2(J)
ENDDO
```

Then each element of the array B is fetched only once from memory, and can remain in the cache between the two lines of computation : the code runs faster.

The Cray compiler has an advantage over the two other compilers : it can analyse the array syntax and operate such a “loop fusion” if possible. According to the documentation, the Intel compiler would perform also the loop fusion with the optimization level -O3 ; however in the facts this is not the case, or there has been a misinterpretation of the Intel documentation.

The solution is to remove the array syntax formulation and recode the loop following the traditional Fortran 77 style. Note that it will also benefits a bit to the Cray compiler, as the re-writing in Fortran 77 style can raise opportunities to re-organize the loops in ways to avoid useless re-computations.

This technique has been applied in particular to the subroutine `th_r_from_thl_rt_1d.F90` as a first test, because it is fully written in array syntax but it is a short subroutine (170 lines, including comments). After being re-written in F77 style, the local speedup was 10 % to 15 %. On another, longer subroutine, it was clear that the more array syntax was eliminated, the bigger the speedup was.

Memory allocation

There are two kinds of arrays : the automatic arrays and the dynamic arrays.

The automatic arrays are those arrays one allocates at the moment they are declared. Example :

```
REAL :: ZAUTOMATIC(N)
```

Automatic arrays are allocated very fast, in a memory region called the *stack*. Unfortunately they make higher water mark of the total memory usage.

The dynamic arrays (as known as allocatable arrays) are those arrays one allocates explicitly at a time in the program. Example :

```
REAL :: ZDYNAMIC(:)
ALLOCATE(ZDYNAMIC(N))
```

Dynamic arrays are allocated in a memory region called the *heap*. They are more flexible than the automatic arrays since one doesn't need to know much in advance what the size of the array should be. They can also help to save memory. However, such allocations are slower than allocations of automatic arrays.

Here the Cray compiler has another advantage over the two other compilers : it can use a special allocator called *tmalloc* (thread-caching malloc). Though more expensive in memory than the standard allocator *malloc*, *tmalloc* is known to be faster for allocations/deallocations cycles inside parallel regions.

Gfortran and the Intel compiler use the standard allocator *malloc*. As a matter of fact, the allocation of an array does not take place at the exact line of code where it is allocated, but at the moment the array starts to be filled for the first time. Example :

```
REAL :: ZDYNAMIC(:)
ALLOCATE(ZDYNAMIC(N)) ! No, the array ZDYNAMIC is not really allocated yet.
...
ZDYNAMIC(:)=0. ! ZDYNAMIC is allocated now, and the filled.
```

Consequently a problem of performance due to an allocation can be very difficult to diagnose because the first time a dynamic array is filled can be very far away in the code from the place it is allocated.

Against this issue there are three possible solutions ; the choice of the best one depends on the piece of code itself :

- Use automatic arrays
- Allocate only once ; do not deallocate until the array becomes completely useless
- Deallocate and reallocate only if the size of the array has increased, or has changed

These techniques have been applied in the gridpoint calculations (semi-lagrangian-related arrays) and the FFTs transforms. Locally the speedup was bigger than 200 %.

Vectorization skill and efficiency

“Skill” is defined here as the ability for the compiler to vectorize a given loop. “Efficiency” is defined here as the ability for the compiler to generate the fastest binary code from a vector loop.

Each compiler has its own strategy for vectorization. Each one may decide to vectorize a loop or not, based on different criteria. Though the following observation is subjective, it seems from experience that :

- Cray is skilled and efficient.
- Gfortran is skilled but less efficient than Cray.
- Intel can be more efficient than Cray, but it is less skilled.

The vectorization or not of a loop, from one compiler to the other, seems to depend on the occurrence of array syntax or deferred-shape dummy arrays. There is no strict order among the compilers, but the Intel compiler seems to be the most penalized one, because of its limited skill in vectorization.

For the long term, one should try and report vectorization failures to the compiler vendor (if any ...). In the case of Intel, the upgrade of compiler from the version 13 to the version 15 has reduced the number of vectorization failures, and led to an overall improvement of the performance of nearly 5%.

Besides, one can “help” the compiler with compiler directives. Unfortunately there are not many of them, and there are no directives for vectorization with Gfortran.

Vectorization can be forced by the following directive :

- Cray compiler : `!DIR$ PREFER VECTOR`
- Intel compiler : `!DEC$ VECTOR ALWAYS`

For both of them, dependencies can be ignored by the directive : `!DEC$ IVDEP`

That last directive can be particularly helpful for the Intel compiler which can easily get confused by components of a Fortran structure. Also it often happens that the Intel compiler would report a vector dependency though the real reason for the compiler not to vectorize a loop would be another one : subroutine too long (!), indice of loop too complex, potential use of non-contiguous array, ... This directive can force the compiler to reveal the true reason for not vectorizing a loop.

However, it is interesting to check if one (or two) of the compilers managed to vectorize a given loop, and not the other(s) by observing the speed of the loop. In that case, a good - but not easy - trick is to find a way to re-write the loop in a manner it will vectorize on all compilers.

These techniques have been applied to several loops in the code. They led to a speedup varying from 20 % to 300 %. Best performances were achieved after array syntax or deferred-shaped arrays have been eliminated.

Transcendental functions

Transcendental functions are essentially the trigonometric functions, the logarithm and the exponential (including the power function). These functions are expensive (in cpu) and regularly used inside large computational loops ; therefore it is important that they vectorize. If they don't, they can break down the performance of a large computational loop.

Unfortunately these functions can't vectorize with Gfortran. With the Cray compiler they can vectorize, except `ASIN` and `ACOS`. With the Intel compiler they can't vectorize, or the bitwise reproducibility is lost.

Waiting for the compilers to support full vectorization of the transcendental functions, one can try to isolate them in specific loops (and preferably contiguous loops to enable loop fusion later on).

Applied on the subroutine `advprcs.F90`, this technique has improved the local speedup by 26 % for Intel, and 6 % for Gfortran.

Copies of arrays

This area has not been fully investigated yet. The first experiments are showing again that pure array syntax is not the best for performance : other aspects, especially the associativity property of the memory cache, should be considered.

6 Compiler performance after optimizations

The picture below (Figure 2) shows for each compiler its performance after optimizations as described above, with respect to its performance before optimizations. In other words, each compiler is compared to itself, after code optimization.

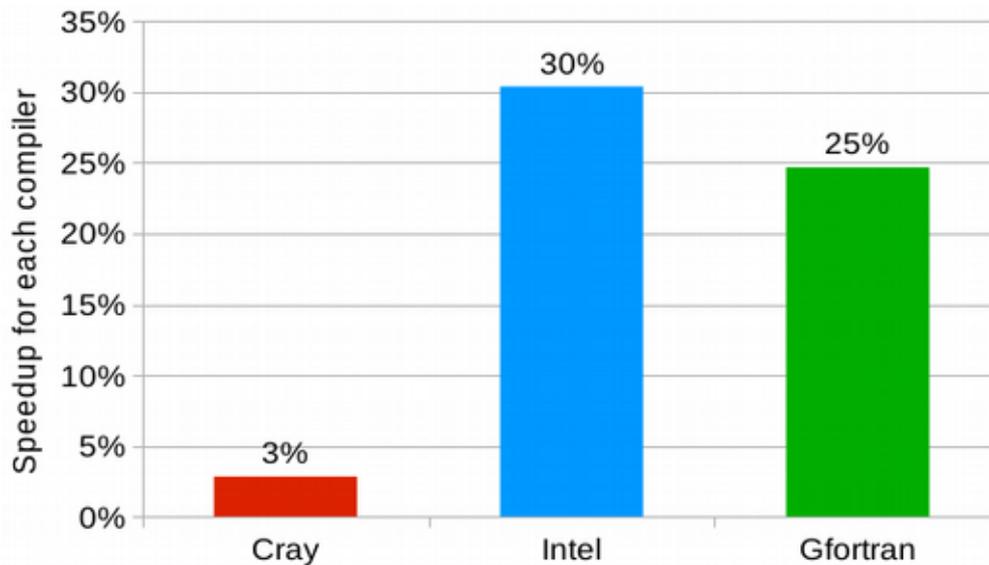


Figure 2: Performance of each compilers after optimizations (relatively to their performance before optimizations) for AROME (cycle 41)

As one can see, the Cray compiler gets only little benefits of the optimizations, as it was already very fast.

Similarly to the Figure 1, the picture below (Figure 3) shows for each compiler its relative performance with respect to a new reference which is the mean performance of all three compilers, after optimizations as described above have been applied to each of them.

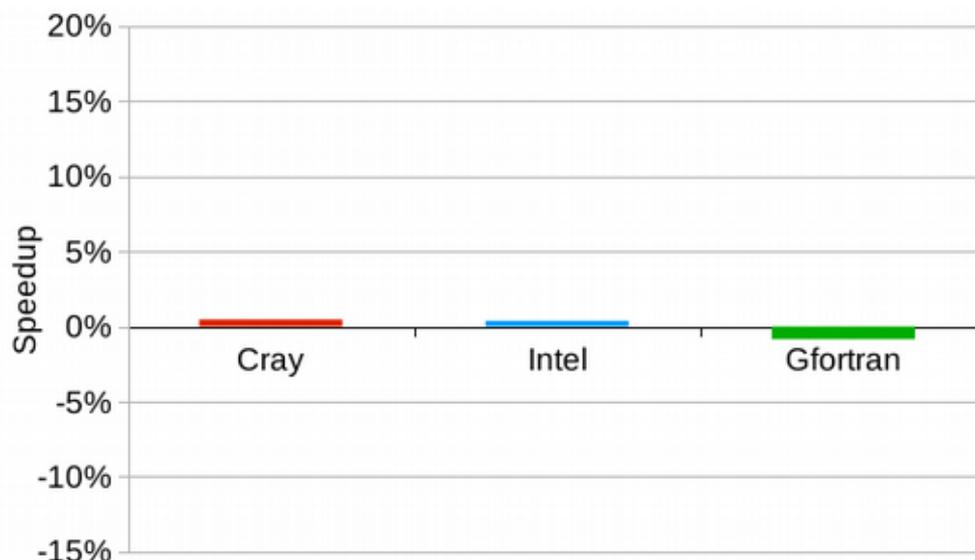


Figure 3 : Performance of each compilers (relatively to the mean performance of them) for AROME (optimizations on top of cycle 41)

There is not much to say : all three compilers are now *globally* equivalent. Nevertheless, looking to the details of performances, there are still differences of performances at the level of subroutines or loops ; which means that there is still room for optimizations, but one cannot say to which compiler(s) it will profit.

7 Conclusions

Indeed, the compiler can make the difference ! The Cray compiler has clearly advantages over the other two, and using it can help the developers focus on more scientific topics than optimizations. However, the competition is stimulating, therefore it would not be good to have one vendor far ahead from the others.

Surprisingly, the comparison of the compilers performances have revealed that there were still room for improvement at unexpected places. It told us how to code in a more “competitive” or “portable” way. Effectively as the programming languages become more and more sophisticated, we tend to use the new facilities of the language and forget the basis of High Performance Computing, that the compilers still won't compensate for us. An old notion like “vectorization” is well-known and still relevant, but we miss that the supercomputers of today are *cache-based* machines. A better understanding of the mechanism of the processors of today should help us adopt more efficient coding practices. Any developer should learn it !

Experiments with HARMONIE on Xeon Phi Coprocessors

Enda O'Brien

1 Introduction

Novel computing architectures like Xeon Phi Coprocessors and GPUs promise enhanced performance based on massively increased parallelism. However, they also have relatively limited memory and transferring data to them and back again is very slow. Currently these performance limitations, at least in the case of Xeon Phi, exceed the benefits for applications like HARMONIE. Nevertheless, this article reports on experiments with HARMONIE and a “standalone” version of the LAITRI subroutine on Xeon Phi coprocessors which provide some insight as to how array structures and other features of HARMONIE might be changed in future versions in order to improve performance.

2 OpenMP Multi-Threading in HARMONIE

As reported in an earlier Newsletter [1], MPI parallelism alone is not enough to obtain good performance from Xeon Phi coprocessors; OpenMP multi-threading is needed as well. Thanks to what was evidently a lot of hard work and clever programming, HARMONIE currently benefits from OpenMP directives at a very high-level in the source. Thus, the OpenMP directives encompass a few large parallel sections (i.e., they are “coarse-grained”). In general, this is much more efficient than having many smaller “fine-grained” parallel sections, since “forking” and “joining” each multithreaded section incurs a lot of overhead cost.

One consequence of coarse-grained OpenMP parallel sections is that all variables declared locally in subroutines called from within such parallel regions (i.e. not passed in the argument list or covered by a USE statement) are “private” to each OpenMP thread. In other words, each thread creates its own separate copy of such variables, using its own separate memory.

HARMONIE is designed so that much of the grid-space computational work is done in “blocks” of horizontal grid-points, of size determined by the NPROMA parameter. The value of NPROMA is typically chosen so that all (or almost all) the data required by each “block” can fit in cache, and so can be accessed quickly. “Cache-blocking” like this minimizes the need to access data from (the much slower) main memory. From a computational point of view, it is a very efficient and valuable feature of HARMONIE.

When HARMONIE is run without any multi-threading, the single thread of each MPI process works on each NPROMA-sized block in sequence, until all such blocks have been processed. In runs that use OpenMP multi-threading, the full set of NPROMA-sized blocks are distributed among the threads, so each available thread can work on its own complete block, in parallel with all other threads.

This block-distribution strategy is highly parallel and very efficient computationally, but since each thread uses its own full block-sized set of private variables in parallel, it is also quite profligate with memory – especially when many threads are running. In fact, the amount of memory consumed is almost directly proportional to thread-count. This is usually not a problem when OMP_NUM_THREADS is set to 2 (e.g., to use the “hyperthreads” or extra logical cores on a standard Xeon processor) – or even 4 or 8 – but can cause failure due to lack of memory when OMP_NUM_THREADS is set to values of 60 or more, as would normally be the case on a Xeon Phi coprocessor. When trying to configure HARMONIE to use all the computing resources of a Xeon Phi (typically with 8 or 16 GB memory and 240 logical cores), the system invariably runs out of memory long before it runs out of threads. Empirical evidence for this was reported in an earlier newsletter [1].

Such behaviour can also be inferred from the structure of the “working” arrays in the top-level grid-point routine CPG. In cycle 38, these arrays have typical dimensions (NPROMA, NFLEVG, KBLKS), or some variation of these, where NFLEVG is the number of vertical levels and KBLKS resolves to OMP_NUM_THREADS (see

gp_model.F90). Note that these array sizes are independent of the total number of horizontal grid-points: what really matters is the “block” size NPROMA, and the number of such blocks to be processed in parallel (KBLKS – i.e., the total number of OpenMP threads).

The main parallel loop then spans the full horizontal grid (IGPCOMP) in steps of size NPROMA:

```
!$OMP PARALLEL DO PRIVATE(J,L,...)
  DO J = 1, IGPCOMP, NPROMA
    ! (calculate offsets, start and end points for each block, etc.)
    L = OML_MY_THREAD
    CALL CPG_GP(J, L, ARRAY(1,1,L),...)
    ...
  ENDDO
```

The working arrays in subroutine CPG_GP are then just 2-dimensional (spanning NPROMA and NFLEVG). However, the local arrays in CPG_GP – and in all the other routines below it – are “private” to each thread, which means that each thread has its own separate copy. There are quite a lot of such arrays, and the memory required to hold them increases very fast as a function of the total thread-count. Such memory usage precludes the use of more than a handful of threads per MPI process. This is certainly unsustainable on a platform like Xeon Phi, but is not ideal even on standard host nodes.

The multi-threading of grid-point operations in HARMONIE cycle 41 is slightly different, in that the last dimension of working arrays in the new “CPG driver” CPG_DRV is not KBLKS (or OMP_NUM_THREADS), as above, but instead corresponds to the number of NPROMA blocks that fit into the total number of grid-points, i.e., the “number of grid-point blocks” NGPBLKS. For a typical NPROMA value of 50, and typical MPI subdomain sizes of 1,000 – 10,000 grid-points, NGPBLKS would have a value somewhere in the range between 20 and 200 – but possibly much larger as the number of MPI processes is reduced and the number of OpenMP threads is increased proportionately. So in cycle 41, the memory required by the higher-level 3-dimensional arrays will not increase in proportion to OMP_NUM_THREADS, but will be proportional to the fixed value of NGPBLKS. This is relatively large but at least is independent of thread-count, and is not a problem.

However, the multi-threaded loop fragment shown above has exactly the same characteristics in cycle 41 as in cycle 38: work is distributed among all available threads in block sizes of NPROMA. The block sizes are fixed, and independent of thread-count. Similarly, each thread acquires its own private copy of all “local” variables in subroutines below CPG, and memory consumption still increases very fast as more threads are used.

As in most applications, MPI parallelism in HARMONIE works by sub-dividing the full domain into ever-smaller sub-domains as the MPI process count increases. For a given problem size, the amount of computational work done by each MPI process goes down as the number of processes goes up. This is the so-called “strong-scaling” approach. A “weak scaling” approach is also possible, in which ever-larger problems are solved by using more MPI processes, while keeping the amount of work per process about the same.

OpenMP parallelism in HARMONIE is different: there is currently no domain decomposition as for MPI. Instead, the fixed NPROMA-sized blocks that are chosen for optimal cache-use are simply distributed among all available threads. The work-rate (and memory use) increases almost in direct proportion to the thread-count. While this design is highly parallel, there is a scalability limit to how far the thread-count can be increased, that is set by both the amount of available memory and by the number of shared-memory cores. On standard Xeon nodes, with fewer than 50 logical cores but 64 GB or more memory, the limiting factor is the core count. On Xeon Phi coprocessors, on the other hand, with 240 logical cores but only 8 or 16 GB memory, the limiting factor is invariably memory.

For OpenMP multi-threading to work properly on Xeon Phi coprocessors, a true “domain decomposition” or other more scalable approach is really needed to replace the fixed-size NPROMA block-distribution method described above. Given the size and complexity of HARMONIE, can any practical strategy accomplish this?

One option, at least on the Xeon Phi, is to reduce NPROMA. While the Xeon Phis do have small level-1 and level-2 caches, they have no analogue to the multi-megabyte level-3 caches that are present on standard Xeon processors, so the operation of “cache blocking” is quite different on Xeon Phi. Reducing the size of NPROMA allows the thread count to be increased by the same proportion without changing the total amount of memory

consumed. Results from such an exercise are shown in Fig. 1. This shows run-time performance as a function of NPROMA for a medium-sized (300 x 300 point) HARMONIE configuration run “natively” over 4 Xeon Phi coprocessors (i.e., 4 MPI processes), using 40 OpenMP threads per process for all points on the red curve, and 100 threads per process for the result represented by the blue square. The main result is that performance improves slowly as NPROMA increases; the larger values allocate larger blocks of work to each thread and not surprisingly this is more efficient than working on smaller blocks. However, for NPROMA=60, it was impossible to run this job using more than 40 threads per process, since not enough memory was available. As expected, reducing NPROMA to 10 allowed the job to run with 100 threads per process – but the small performance gain due to the extra threads was not enough to offset the performance loss due to the smaller NPROMA. So, back to the drawing board.

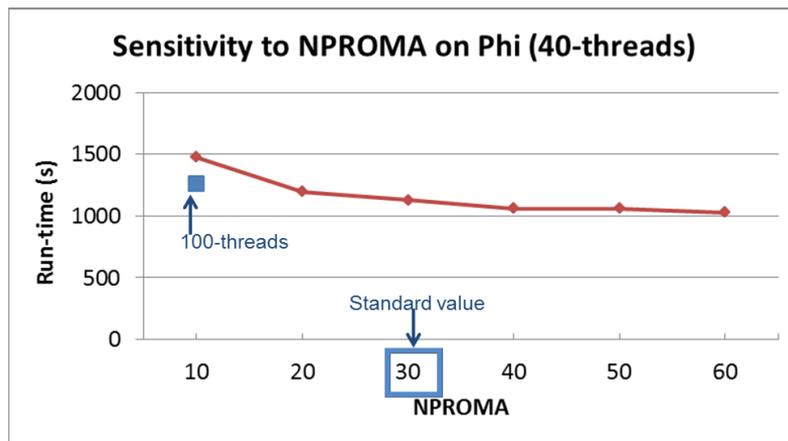


Figure 1. Run-times for the IRELAND55 case over 4 Xeon Phi cards as a function of NPROMA.

Perhaps a more promising option, but more difficult to implement, would be to modify the concept of “cache blocking” to one of “thread blocking” – at least for the Xeon Phi. Then the working arrays would have just the physical dimensions (NFLEVG, NPROMA_T) instead of (NPROMA, NFLEVG, KBLKS), where NPROMA_T could be an order of magnitude (or more) larger than the current NPROMA – indeed as large as memory permits. Note that it is necessary to have NFLEVG in the first dimension in the modified array structure, since OpenMP parallelism is performed over the last (outer) dimension, and vertical dependencies and a limited number of vertical levels would make it difficult to parallelise over NFLEVG. The working arrays could still be “blocked” into a new “thread block-size” by dividing NPROMA_T by the number of threads:

$$\text{THRED_BLK_SZ} = \text{NPROMA_T} / \text{OMP_NUM_THREADS}$$

The number of “outer” blocks (NBLKS) is then obtained from the total number of grid-points (IGPCOMP):

$$\text{NBLKS} = \text{IGPCOMP} / \text{NPROMA_T}$$

(For current purposes we can assume that the integers above divide evenly into each other, though of course adjustments would need to be made for situations where they did not).

The original code snippet shown above would then become something like:

```

DO I=1,NBLKS
  ... ! (calculate offsets, start and end points for each block, etc.)
!$OMP PARALLEL DO PRIVATE(J,L,...)
  DO J = 1, NPROMA_T, THRED_BLK_SZ
    L = OML_MY_THREAD * THRED_BLK_SZ
    CALL CPG_GP(I, J, L, ARRAY(1,L),...)
    ...
  ENDDO
ENDDO

```

As shown in this new construction, the size of the 2-dimensional working arrays in CPG_GP now goes down as the thread-count goes up, since THRED_BLK_SZ is inversely proportional to OMP_NUM_THREADS. This means that total memory consumption should not change as the number of threads increases. It is the key difference from the original construction, where each thread's block-size was fixed and memory consumption increased with the thread-count.

In the ideal case where NPROMA_T = IGPCOMP (and tests with cycle 41 suggest that this is also practical), the number of "outer" blocks (NBLKS) reduces to one, and the short outer loop above is effectively removed.

That, at least, is a schematic overview of the theory. For the time being, however, it is completely impractical to try this with the full HARMONIE code. The next step should be to experiment with the different structures described above using a simple yet representative standalone program, and see whether the modified array structures provide any performance benefit at all on Xeon Phi coprocessors. Perhaps such restructuring will prove to be futile, and increasing the thread-count while reducing the thread block-size will simply lead to a result such as that shown in Fig. 1, where the performance using 100 threads and NPROMA=10 is worse than using just 50 threads and NPROMA=40. Empirical performance results are still needed to clarify this.

3 Performance Tests with a Standalone LAITRI Program

While no "standalone" version of CPG has been written yet, a standalone version of the semi-Lagrangian dynamics routine LAITRI has. LAITRI is an attractive routine to work with since it is both self-contained (i.e., calls no other subroutines), and it accounts for a relatively large fraction of total run-time. The experiments with LAITRI described below were conducted mainly to explore the potential of Xeon Phi, and not from any attempt to reconfigure HARMONIE along the lines described in the previous section. They are all relevant nevertheless.

A "standalone" version of LAITRI from cycle 41t1 was produced, by writing a small "driver" program to repeatedly call LAITRI exactly as in the full HARMONIE code. A complete set of the input variables needed was saved to a file during a run of the full HARMONIE model with just one MPI process. Thus the array dimensions correspond to a relatively large MPI sub-domain (300 x 300 grid-points in this case). This file can then be read by the driver program, and so provides physically reasonable values for all the arguments used when calling LAITRI. The main "input" array is PXSL, the output is array PXF, while the main integer dimensions are KSLB1 ("horizontal dimension for grid-point quantities", equal to 63,517 in this case), and KPROMA ("horizontal dimension for interpolation point quantities", equal to 53 in this case, and evidently related to the HARMONIE input value of NPROMA=50).

In order to obtain representative run-times, calls to LAITRI were made from within a do-loop, with some PXF output written after every call, and the PXSL input array changed slightly too in order to track progress over each loop iteration. At least 100 loop iterations were performed, and run-time for the entire loop was measured. Within the loop, the call to LAITRI accounted for almost the entire run-time; writing some PXF output and updating PSXL input took only negligible time (as was verified by removing those steps from some test runs).

In order to run LAITRI on a Xeon Phi coprocessor, an "offload" directive (similar to an OpenMP directive) was added to the driver program just before the call to LAITRI. Separate "offload_transfer" directives were used to control (and to minimize) the movement of data between the host node and the coprocessor. In order to mimic how this might work in the context of HARMONIE itself, the input array PXSL was transferred to the coprocessor before each call to LAITRI, and the output PXF was transferred back to the host afterwards. Only one line had to be added to the LAITRI source: a directive to give the subroutine the "offload" attribute.

In the HARMONIE context, LAITRI is called from within an OpenMP parallel region, much as CPG_GP is called from within the code snippets shown above. In the standalone version, a simple OpenMP parallel directive was added above the outermost loop to provide basic multithreaded parallelism. The problem size could be increased by multiplying KPROMA by some factor, and by replicating all input arrays that had KPROMA as a dimension by the same multiple.

For each version of the driver program, an "offload" executable was compiled to run LAITRI on Xeon Phi, and a "host" executable was compiled (treating the offload directives as comments) to run entirely on a standard Xeon host node. Tests were also done with the entire package built to run entirely, or "natively", on a Xeon Phi.

The following four sets of benchmark tests were run in order to evaluate how LAITRI performance depended on certain features of its source code.

3.1 Scalar vs. Vector Algorithm

All versions of LAITRI contain two separate versions of the same algorithm, one for scalar and one for vector processors. The one used at run-time is selected by the LOPT_SCALAR variable (cycle 38) or LOPT_RS6K (cycle 41t1). The two schemes probably have their origins deep in the mists of time when big vector systems (such as the Fujitsu VPP5000, or NEC SX-8) were available along with scalar RISC platforms (such as the IBM RS6000). Those vector systems had no cache, while the scalar platforms did, and still do (as do standard Xeon processors, in particular). So the “scalar” algorithm in LAITRI uses several short internal loops over KPROMA in order to make best use of the cache, while the “vector” version just ploughs through one long loop, and makes no use of the KPROMA-sized local variables at all.

Results from a reference LAITRI test using both “scalar” and “vector” versions on both Xeon Phi (offload mode) and a host Xeon node are shown in Table 1. All available cores or threads were used in all cases.

Table 1 Run-times (seconds) of standalone LAITRI test using different platforms and solution algorithms

LAITRI algorithm	Xeon Host (20 core E5-2660 2.2 GHz)	Xeon Phi 5110P (240 threads)
Scalar	57.5	236.4
Vector	57.2	119.4

The simple conclusion from Table 1 is that while it really makes no difference to a standard host node which particular LAITRI algorithm is chosen, for Xeon Phi the vector scheme runs almost twice as fast as the scalar one. While the difference is dramatic, it is perhaps not too surprising given that Xeon Phi has no L3 cache to benefit from the “cache-blocking” provided by the scalar scheme. Even using the vector scheme, however, Xeon Phi is still about twice as slow as a full host node.

In all further results shown below, the vector algorithm was used on both host and coprocessor platforms.

3.2 Dimension Re-ordering

Seven of the eight arrays in the LAITRI subroutine argument list are dimensioned (KPROMA, KFLEV), followed in some cases by a third or even fourth dimension as well. OpenMP parallelism operates on the outermost dimension (the number of vertical levels KFLEV), but it is much easier to increase the horizontal dimension KPROMA than KFLEV. Therefore, the order of these dimensions was exchanged, as was the loop ordering in LAITRI, to use KPROMA as the “outer” dimension and so parallelise over KPROMA instead. In a further set of experiments, the array ordering was changed again to use (KFLEV, KPROMA) as the outermost dimensions in all LAITRI arrays. Thus, the original PCLO(KPROMA, KFLEV, 3, 2) was first changed to PCLO(KFLEV, KPROMA, 3, 2), and then to PCLO(3, 2, KFLEV, KPROMA).

The consequence of making these changes can be seen in the first three rows of Fig. 2, which shows run-times for a test case with KPROMA increased by a factor of 1000 on both Xeon Phi and Xeon “Host” platforms under different source code modifications. The array dimension re-ordering (and associated loop re-ordering) certainly improved performance on the Xeon Phi, while making essentially no difference on the host.

3.3 Hoisting “IF” blocks

There are several “IF” statements within the main LAITRI loops to control what should be done depending on the “monotonicity” parameter KQM. Since this input parameter does not change within LAITRI, it is possible to “hoist” these if blocks from inside the loops, and essentially put the “IF” blocks outside the “DO” loops. This was to test whether the “IF” statements interrupted the flow of computation enough to impair performance.

Comparing the 4th row in Fig. 2 with the row above shows how performance responded to this change: apparently it made little or no difference on either platform. In other tests, however, hoisting the “IF” statements produced a performance gain of about 20% on Xeon Phi (but no change on the host), so it seems like a worthwhile change to keep, especially since it is relatively simple to implement.

3.4 Replacing Repeated KLO Indirect Indices with Temporary Variables

One of the key features of LAITRI is its use of “indirect indexing” (via the KLO array) to select different elements of the PXSL field for use in the interpolation. In the original code, KLO is dimensioned (KPROMA, KFLEV, 0:3). Following the changes described above, the final KLO dimensioning is (0:3, KFLEV, KPROMA). The entries KLO (I, JLEV, JROF), where I=0, 1, 2, or 3, are repeated several times in the course of the main LAITRI vector loop. It is relatively easy to store each of these 4 values in temporary scalar variables at the top of each loop iteration, and use these temporary variables instead of the full KLO elements to locate the required PXSL entries in the body of the loop. The effect of this change is shown in the final row of Fig. 2. It produced a significant performance improvement on Xeon Phi, while having no discernable impact on the Xeon host.

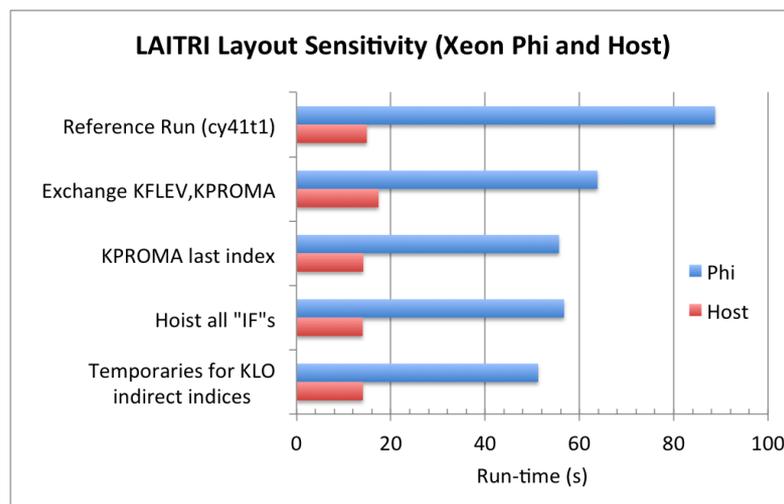


Figure 1 Response of LAITRI performance to various source-code changes, on both Xeon Phi and a standard Xeon host node. Changes are cumulative: configuration in any row includes changes in the rows above as well.

3.5 Xeon Phi Performance Limits

The real motivation behind all these experiments with a standalone version of LAITRI was to see how Xeon Phi could perform on a representative and amenable piece of HARMONIE without the handicap of running out of memory while using only a fraction of its compute resources, as it invariably does when running the complete model. In making all the changes described above and as shown in Fig. 2, Xeon Phi was given every possible advantage: the “vector” algorithm, the loop re-ordering, lifting of “IF” blocks outside the “DO” loops, and use of temporary variables instead of repeated array elements. Despite all this, run-times on the Xeon Phi (with 240 threads) remain over three times longer than on a host node (with 20 threads), as shown by the last row on Fig. 2.

One possible reason for the relative slowness of Xeon Phi could be that it just wasn't given enough work to do. To test this, the problem size (as a multiple of KPROMA) was increased as far as 10,000, at which point almost all the 8 GB memory on the Xeon Phi was being used. Results are shown in Fig. 3 for both Xeon Phi and host node, along with the performance ratio between the two. Fig. 3 suggests that run-times on the Xeon Phi saturate at just over double those of the full host node. The results on the last row of Fig. 2 correspond to the value of 1,000 on the x-axis of Fig. 3.

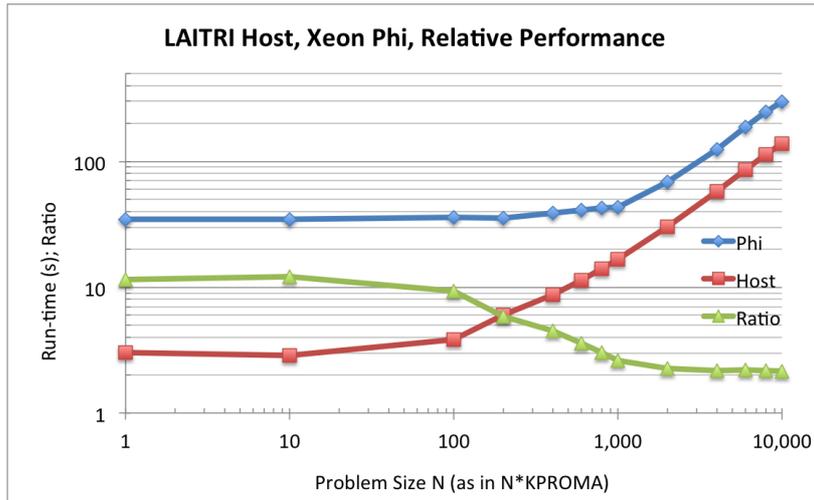


Figure 3 Performance of standalone LAITRI on Xeon Phi and host, as function of problem size.

All Xeon Phi results shown so far are from running LAITRI in “offload” mode, where the input array PXSL is copied to the coprocessor before each LAITRI call, and the output PXF is copied back to the host afterwards. Those transfers consume about 60% of total run-time, as may be inferred from Fig. 4, which shows performance of the standalone LAITRI as a function of thread-count, when run on the Xeon Phi in both “native” mode (i.e., entirely on the Xeon Phi) and “offload” mode, as well as on the host node. The very best times on the Xeon Phi were 51s in “offload” mode, 21s in “native” mode, while the same job ran in just under 14s on the host.

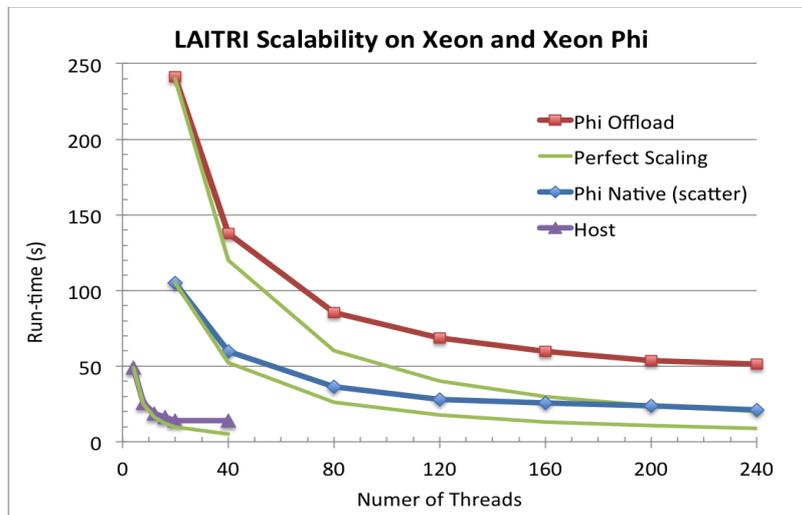


Figure 4: Scalability of standalone LAITRI on both Xeon Phi and Xeon host node, as function of thread-count.

4 Conclusions

It is clear that performance problems with HARMONIE on Xeon Phi coprocessors go beyond the primary limitation of running out of memory before running out of available threads. Despite having more cores (60 x 4-threaded cores) and more overall memory bandwidth (approx. 140 GB/s) than a standard 20-core host node, an interpolation algorithm such as LAITRI still runs slower on Xeon Phi than on the host. While it is easy to write interpolation or stencil applications that perform better on Xeon Phi than on a host node (see [2]), the indirect indexing in LAITRI prevents the assumption of a reliable “unit stride” through the main working array, and evidently this is enough of a complication to slow Xeon Phi performance significantly. Compiler reports show that the inner loops in LAITRI do indeed vectorize, but the Intel Vtune profiler shows that the “vectorization factor” is only about half what it could be in loops with a straightforward unit stride.

Even so, several relatively simple changes can be made to the code structure in LAITRI that significantly improve performance on Xeon Phi coprocessors (without changing the answer). Similar code changes are worth keeping in mind for the future as coprocessor architecture improves and coprocessor performance becomes competitive with that of standard hosts.

Acknowledgement

This work was supported by Intel Corp. through the Intel Parallel Compute Centre (IPCC) at ICHEC.

5 References

- [1] E. O'Brien, 2014. “Running HARMONIE on Xeon Phi Coprocessors”. ALADIN-HIRLAM Newsletter No. 3, Sept. 2014, pp. 69-78.
- [2] E. O'Brien and A. Ralph, 2013. “Can HARMONIE be accelerated with GPUs or coprocessors?”. ALADIN-HIRLAM Newsletter No. 1, Sept. 2013, pp. 64-71.

Impact experiments of HARMONIE radar data assimilation

Shiyu Zhuang, Mats Dahlbom, Xiaohua Yang, Bent Hansen Sass

1 Introduction

The technique of HARMONIE radar data assimilation has been developing for years in HIRLAM community and technically is available now for real time running. Several NWP centers in HIRLAM community include in DMI aims at implementing radar data assimilation operationally in the coming years. Before the operationally implementation, there is a requirement of carrying out impact experiments to evaluate how sensitive to and how much improvements in the short range forecasts will be when radar data are assimilated in the system. The additional benefit of performing this kind of impact studies is, because of being covered by radars from several countries, the experiments in Danish area could help us to get more information on the impact of different radar data of different quality. These information should be useful to handle radar data of different quality.

In the following, several impact experiments of HARMONIE radar data assimilation in DMI were described.

2 Model setup and sensitivity tested

The experiments were performed over the Danish operational DKA domain (Fig.1) utilizing HARMONIE version 38h1.2. The model grid mesh consisted of 65 vertical levels and 800×600 horizontal grid points at roughly 2.5 km grid spacing. The ECMWF forecasts were used as lateral boundary and the 3DVAR data assimilation method was applied in 3-hourly assimilation cycling with data window of 3 hours.



Figure 1: The model domain is illustrated by the red frame

Totally more than 10 sensitivity experiments in 2 main groups have been carried out. Among them one group is one month running starting from 02UTC 1st August 2014 (2014080102) and being with different model configurations, such as structure functions, scaling factor for background error, whether or not to use radar data (reflectivity and/or radial wind). Among them 5 experiments are selected as follows to be presented in this paper:

Experiment 1: no radar data+ reference structure functions (SFS)

Experiment 2: reflectivity + diurnal SFS

Experiment 3: reflectivity + EDA-based SFS

Experiment 4: reflectivity + reference SFS (available in DMI operational HARMONIE)

Experiment 5: reflectivity + EDA SFS + REDNMC=0.6 (rescaling factor of σ_b) (default REDNMC = 0.9)

Experiment 6: both reflectivity and radial wind + EDA SFS

Experiment 7 and Experiment 8 are not listed here.

Experiment 9: only Danish radar reflectivity + diurnal SFS

Another is a group of experiments for case study of a severe flash flood event affecting Copenhagen-Malmø region occurred in the early morning of 31 August 2014. With the same configurations as group 1, all of the model running in this group are just started with a same initial value of 3h forecasts at 2014083017 produced in Experiment 1.

We only illustrate part of the results of experiments (EXP1, EXP2, EXP3, EXP5, EXP9) in the following text due to some of the cases need to be further investigated.

3 Radar data usage

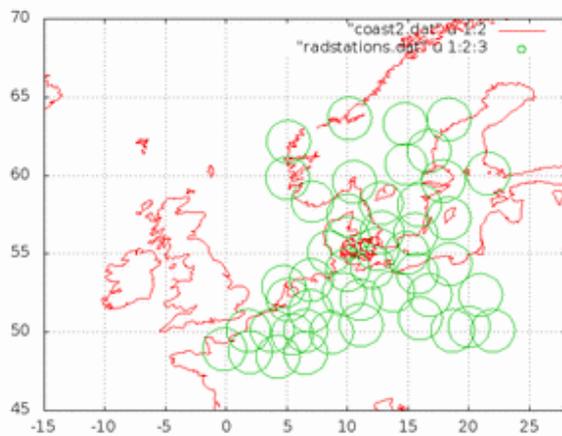


Figure 2 measurements from about 60 radars in Denmark and 9 neighboring European countries have been used for the experiments

In addition to the conventional measurements (TEMP, SYNOP, AIRCRAFT and PILOT), the radar measurements from about 60 radars from 10 European countries in OPERA data sets have been used for the above mentioned experiments (fig.2). With super-obbing preprocessing, the observed raw radar data is executed by QC and data reduction, then the reflectivity is converted to a pseudo-observation of humidity profile through 1D Bayesian retrieval. The humidity profile can be then assimilated into the Harmonie system through 3DVAR. The radial wind, i.e. the velocity of echoes moving towards or away from the radar, can be assimilated in the 3DVAR system directly after QC and dealiasing. For our inter-comparison, only radar reflectivity are applied in most of the experiments. Just one experiment have tested the case with both reflectivity and radial wind data.

4 Verifications

1h output of 24h forecasts at asynoptic time (02, 05, 08, 11, 14, 17, 20, 23UTC) were used to verify against SYNOP and TEMP observations. Figure 3 and Figure 4 shows the surface and upper air of root mean square error(RMS) and bias verification scores respectively for EXP1, EXP2, EXP3, EXP5 and EXP9.

Intercomparison of both synoptic and upper air forecast skills among the experiments indicate:

- Generally comparable forecast skills between the runs with and without radar data.
- Main impact of radar data is seen on modification of cloud cover in the first 12 h of forecast (larger negative bias). However, the impact of Danish radar on modification of cloud cover is smaller.
- Relatively significant skill differences between the radar data assimilation results with different configuration.
- For upper air profile, the standard deviation of humidity (dew point temperature) forecast error from 700hPa to 500hPa looks improved compared to other experiments.

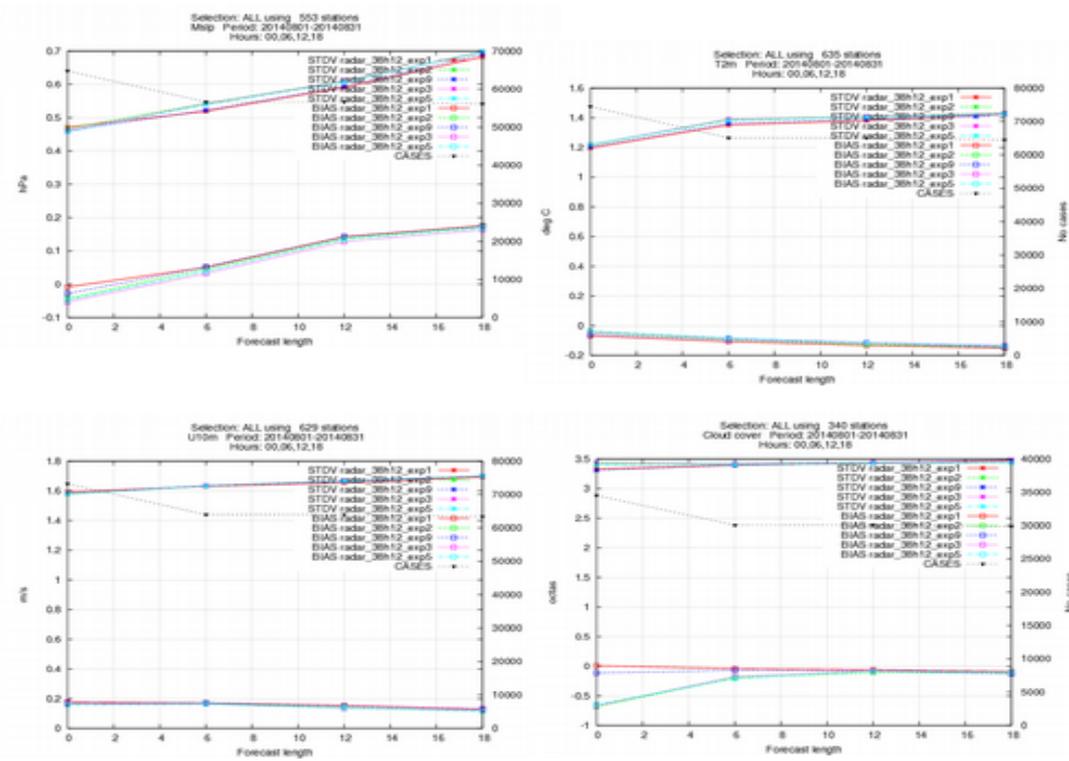


Figure 3: Surface verification against SYNOP observations in the domain and averaged over the one-month period for experiments with different configuration. RMSE and Bias of forecasts for: surface pressure (upper left, unit: hPa), two meter temperature (upper right, unit: K), ten meter wind speed (lower left) and cloud cover (lower right, unit: %) scores as function of forecast range.

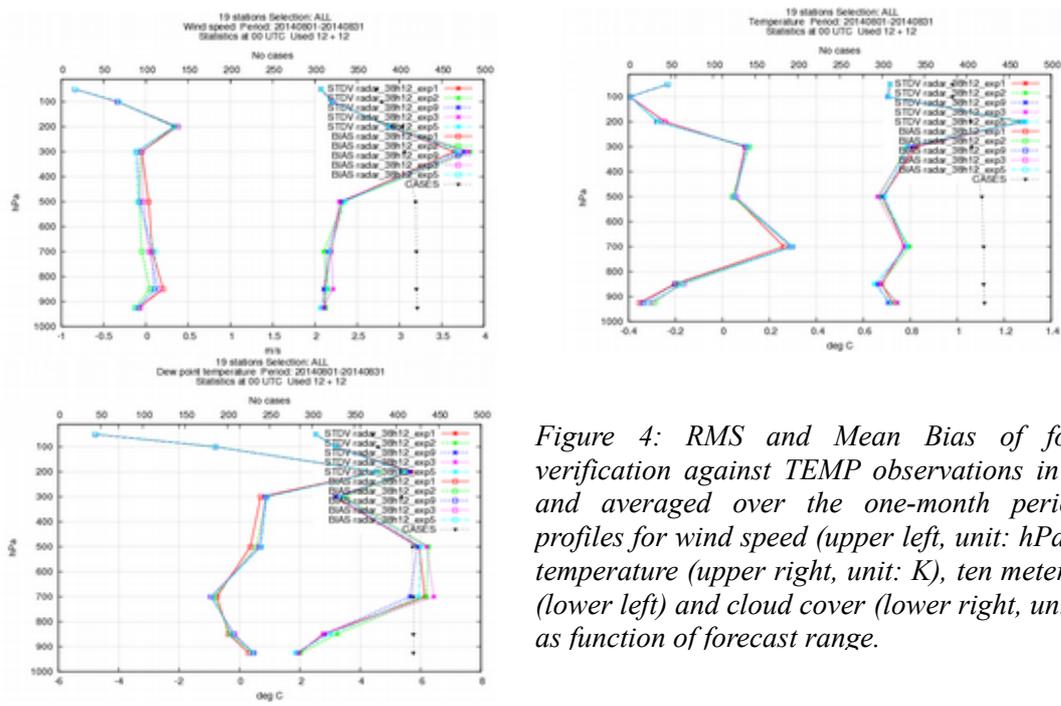


Figure 4: RMS and Mean Bias of forecasts for verification against TEMP observations in the domain and averaged over the one-month period. Vertical profiles for wind speed (upper left, unit: hPa), two meter temperature (upper right, unit: K), ten meter wind speed (lower left) and cloud cover (lower right, unit: %) scores as function of forecast range.

5 Impact on spin-up due to radar data assimilation

From our experiments, some basic features of moisture spin up is exposed by variation of rainfall rate with the forecast length (fig.5).

Moisture spin up in Harmonie are generally limited. For instance, for data assimilation cycling using little moisture observation data, the domain averaged rain rate appear to be very limited. When radar observation used, modification of humidity fields results in more pronounced spin up. In the rainy episode, such spin up in rain rate last up to 3 hours after initialization.

Currently in Harmonie, cloud fraction time series show typically a jump from zero to a stable level shortly after initialization. Initiating cloud fraction from a non-zero value may be helpful for a reduced moisture spin-up.

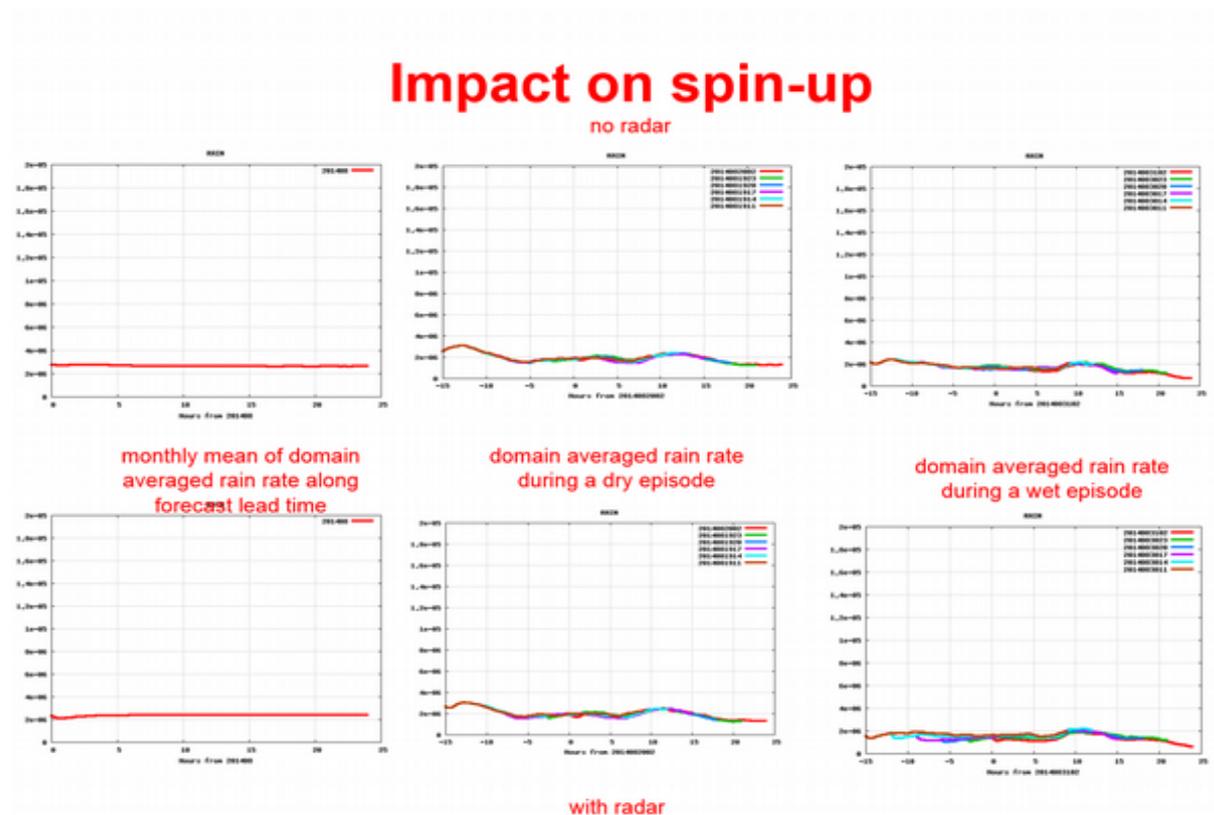


Figure 5: Rainfall rate as a function of forecast length (unit: $\text{kg/m}^2/\text{s}$). Upper: no radar; Lower: with radar

6 Case study of heavy rainfall

A severe flash flood event affecting Copenhagen-Malmö regions occurred in the early morning of 31 August, 2014. The maximum of 24h accumulated precipitation surpassed 130mm (fig.6).

A large amount of precipitation distribution analysis on 1h, 6h, 12h and 24h accumulated rain were carried out In our experiments, But unfortunately no obvious improvement on short range precipitation forecast with radar data was found. Thus, we don't want to present many results here, only some of 6h accumulated rain distribution were provided in this paper (fig.7).

Even though the month-long assimilation with radar data has produced a generally favorable verification results in average, prediction of extreme precipitation cases has shown to be a challenging

task. For the Aug 31 precipitation event over Copenhagen-Malmö region, precipitation forecasts by the continuous data assimilation cycling showed strong sensitivity to the varying configurations and most of them have insufficient forecast skill for predicting the magnitude and location as compared to the observed flash flood episode. It shall be cautioned that the configuration for radar data assimilation and hence these simulation experiments results are still very preliminary.

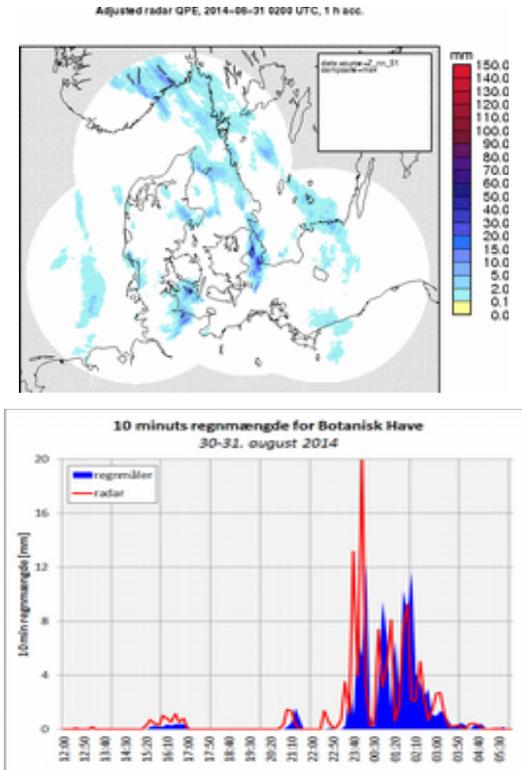


Figure 6: Radar quantitative precipitation estimation (QPE) for a heavy precipitation case happened in Copenhagen-Malmö regions in the early morning of 31 August 2014. 1h accumulated QPE (upper left, unit: mm), 24h accumulated QPE (upper left, unit: mm), 10 minutes measurement of rain gauge against QPE at Botanisk Have (lower left, Courtesy Flemming Vejen, DMI)

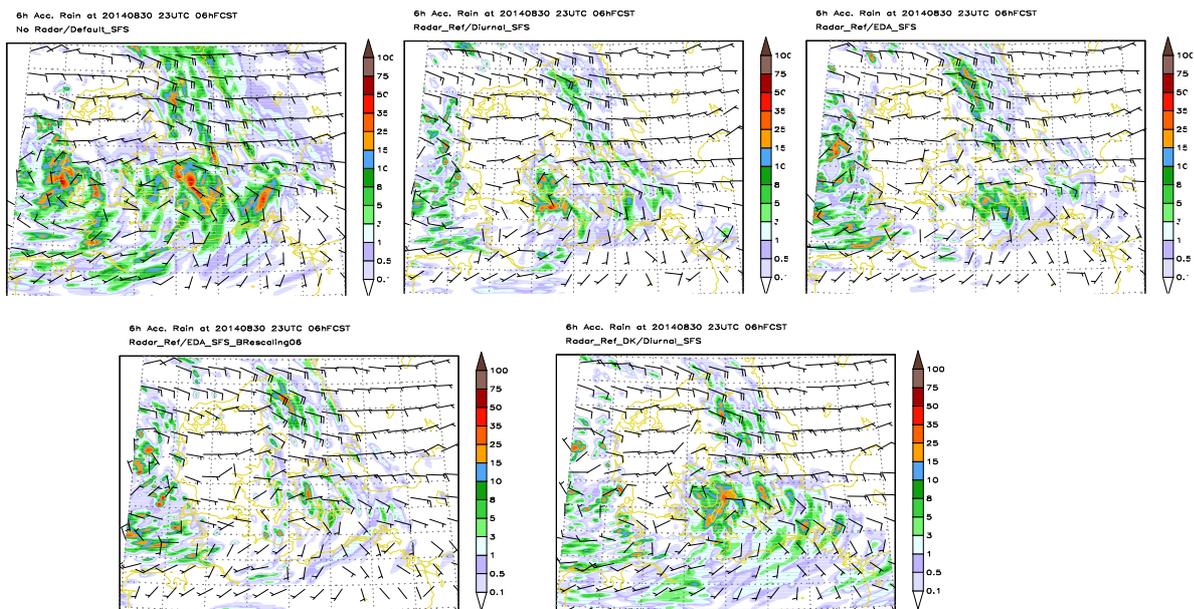


Figure 7: 6 hour accumulated precipitation valid at 05UTC 31, August, 2014 for different experiments. Upper from left to right are EXP1, EXP2, EXP3; Lower from left to right are EXP5 and EXP9.

7 Summary

A series of impact experiments for Harmonie 38h12 is carried out at DMI to investigate sensitivities of the short range forecast to assimilation of radar data with different configurations. These include tests of structure functions, scaling factor for background error, whether or not to use radar reflectivity and/or radial wind. Observation from about 60 radars from 10 European countries have been used with super-obbing preprocessing. By the nature the current attempt with radar data assimilation is still quite preliminary. While verification of radar data assimilation tests for a month-long summer episode have so far shown an overall positive results, an optimal configuration with radar data assimilation is yet to establish. We select in this text to show some case studies for the August 31 heavy precipitation in Copenhagen to illustrate the challenge in this endeavor. Examination of the moisture spin-up in the experiments indicate additional challenge in the use of observation data with moisture information.

8 References

- Caumont, O., et. al, 1D + 3Dvar assimilation of radar reflectivity data: A proof of concept, *Tellus A*, 62, 173-187, 2010
- Martin Ridal etc., Assimilation of radar data in HARMONIE: Activity and Plans, HIRLAM Newsletter No. 58, 8-11, 2011.
- Martin Ridal etc., [Assimilation of OPERA radar data in HARMONIE: A new preprocessing and data reduction](http://www.cnrm.meteo.fr/aladin/IMG/pdf/ridal_asm_apr2015.pdf), On-line presentations in http://www.cnrm.meteo.fr/aladin/IMG/pdf/ridal_asm_apr2015.pdf.
- Salonen, K., Towards the operational use of radar radial wind observations in HIRLAM, HIRLAM Newsletter 55, 23-28, 2009.

Relevance of climatological background error statistics for mesoscale data assimilation

Nils Gustafsson¹, Jelena Bojarova²

¹ Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

² The Norwegian Meteorological Institute, Oslo, Norway

1 Introduction

The number of observations for initialisation of NWP models is generally several orders of magnitude smaller than the number of model state variables. For this reason, the forecast error and observation error statistics are used to spread the observed information in space and time, taking into account the information on the representativity and accuracy of the observations and the balances and relationships valid between model state variables.

Although it was early recognized that ideal background error statistics need to be flow-dependent (Gandin 1963), most of operational NWP system are still using climatological background error statistics for their data assimilation schemes, derived by averaging in time. Besides this implicit assumption of stationarity, other simplifying assumptions like horizontal homogeneity and isotropy are also generally applied.

The model for the background error statistics of the Harmonie mesoscale forecasting system was formulated by Berre (2000). This background error model is based on a linear regression approach. The regression parameters are estimated from a time series of forecast differences which are considered as a proxy for background errors. This approach has been successfully applied for synoptic scale data assimilation, while it was more difficult to show a positive impact of data assimilation for cloud and convection permitting models with a grid resolution of a few km. The main objective of this study is to understand reasons for these problems and to outline the limitations of climatological background error statistics for meso-scale data assimilation

2 Atmospheric spectra and grid point structure function for Limited Area Model

Atmospheric energy spectra can be sampled from observations and aircraft observations of winds at flight levels (8 - 12 km above the ground) have proven to be particularly useful for this purpose (Nastrom and Gage, 1985). With regard to kinetic energy, spectra based on high-resolution model simulations and spectra based on observed data agree reasonably well. They indicate (1) a spectral energy slope proportional to k^{-3} , where k is the horizontal wave number, for horizontal scales ranging from the largest global scales down to approximately a few hundred km; (2) a more flat spectral energy slope proportional to $k^{-5/3}$ for smaller scales down to the smallest horizontal scales resolved by the models, generally up to wave-lengths above $6 \Delta s$ - $10 \Delta s$ where Δs is the grid separation. The spectral energy densities fall off quickly for the shortest waves.

The Harmonie forecasting model is a limited area spectral model (Bubnova et al., 1995) utilizing an area extension (Haugen and Machenhauer, 1993) for its spectral transforms. The application of the extension zone for the calculation of energy spectra is associated with horizontal aliasing problems producing distorted spectra

(Blazika et al., 2015). In order to obtain a result not influenced by aliasing effects, a physical space calculation of structure functions (Lindborg, 1999, Frehlich and Sharman, 2004) is used in this study.

We will thus calculate the second order structure function $f(s)$ for the longitudinal wind u component in the east-west direction:

$$f(s) = \frac{\overline{(u(x+s, y, t) - u(x, y, t))^2}^{xyt}}{\overline{u(x+s, y, t)^2}^{xyt} + \overline{u(x, y, t)^2}^{xyt} - 2\overline{u(x+s, y, t) \times u(x, y, t)}^{xyt}} \quad (1)$$

where s is distance separation in the x-direction and $\overline{(\)}^{xyt}$ denotes an average in the horizontal and in time. One can notice that the structure function $f(s)$ is proportional to the difference between the sum of the variances and the covariance multiplied by 2. A covariance in physical space corresponds to a variance spectrum in spectral space. It can be proven that the slope of kinetic energy spectrum $\sim k^{-5/3}$ in spectral space corresponds to slope of a structure function $f(s) \sim s^{2/3}$ in grid-point space (Frehlich and Sharman, 2004).

3 Two different ensemble sets for generation of background error statistics

Two different sets of time series of forecast differences were tried for estimation of background error statistics for HARMONIE data assimilation (Fischer et al, 2005). HARMONIE model (Seity et al, 2010) was run for the month of August 2011 at a horizontal resolution of 2.5 km with 65 vertical levels over Scandinavian domain. Both sets of experiments contain 6 ensemble members and use operational ECMWF EDA (Ensemble of Data Assimilations) ensemble 12h forecasts from 06 UTC and 18 UTC as lateral boundary conditions. The ECMWF EDA uses a T399 horizontal resolution and 91 vertical levels, and is based on 12h 4DVAR. The initial data for the first set of HARMONIE ensemble forecasts (denoted as *DS*) was obtained by a direct downscaling of the ECMWF EDA ensemble fields from 06 UTC and 18 UTC, namely by the horizontal and vertical interpolation of ECMWF fields to HARMONIE geometry. The initial data for the second set of the HARMONIE Ensemble forecasts (denotes as *EDA*) is obtained from a 3D-VAR assimilation of perturbed observations. The Harmonie 3D-VAR scheme uses 6h cycle in this study and assimilates conventional observations only.

To illustrate the properties of these two ensemble sets, in Figure 1 we show structure functions for the longitudinal component of the wind in the east-west direction computed for the *DS* and *EDA* HARMONIE ensemble sets. This Figure shows horizontally averaged structure functions at model level 47 (≈ 900 hPa) computed from *DS* (top) and *EDA* (bottom) ensemble sets valid on the 5th of August 2011 06 UTC with forecast length ranging from +0h to +12h. We may notice first that the structure functions from the downscaling experiment are associated with a strong spinup during first 6-12 h of the forecast integration. This is in an agreement with what was reported earlier by Skamarock et al. (2014). This is mainly because the high-resolution HARMONIE forecast model in the downscaling experiments is initialized with ECMWF coarse resolution data. In the EDA experiments the own high-resolution analysis is build and as a result the spinup of the structure functions in the EDA experiment is much weaker. It is interesting to notice that HARMONIE structure functions reflect the daily cycle of boundary layer processes (summer day time turbulence and convection) with spinup of the structures with the forecast length during the day and the spindown during the night. This behaviour is mostly pronounced for the shortest horizontal scales.

The structure functions from the EDA experiment in the lower troposphere (Figure 1) are in good agreement with the observed spectra on the scales in the range from 15-20 km to 500-1000 km, indicating that the real resolution of the HARMONIE model with the grid resolution $\Delta s = 2.5$ km is in the range $6\Delta s - 8\Delta s$ (Frehlich and Sharman, 2004). The decay of the energy on the longest waves is due to the size of the domain. It takes longer time to spinup meso-scale structures higher up. On the jet-stream level (≈ 300 hPa) *DS* and *EDA* ensemble sets were not able to develop meso-scale structures even though the presence of meso-scale structures

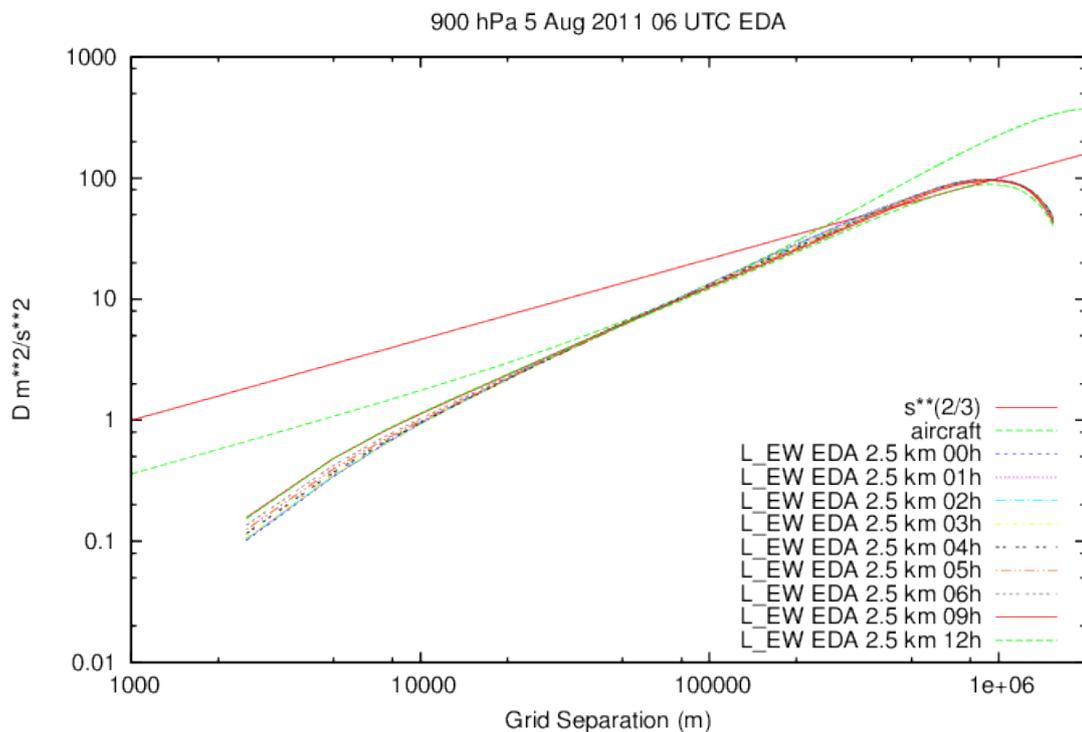
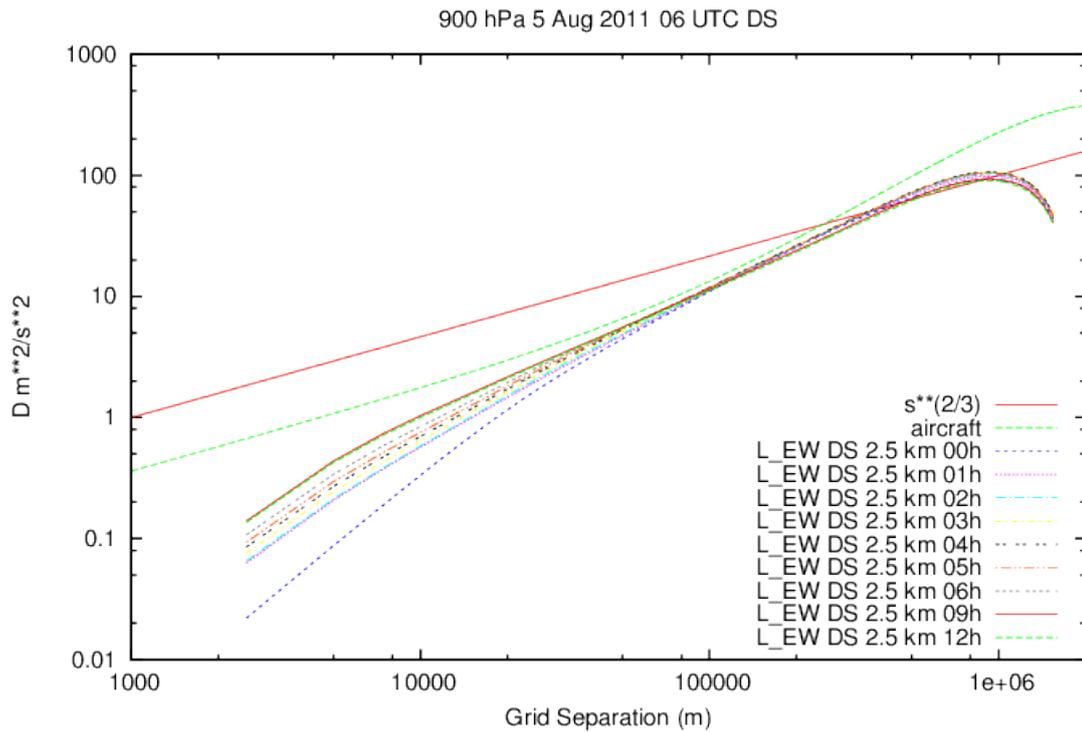


Figure 1: Structure functions for the longitudinal component of the wind in the east-west direction. Horizontal domain average at model level 47 ($\approx 900 \text{ hPa}$) for 5 August 2011 06 UTC. Separate curves for forecast lengths +0h, +1h, +2h, +3h, +4h, +5h, +6h, +9h and +12h. The theoretical slope of the structure function $s^{2/3}$ as well as the structure functions derived from aircraft observations are also included. Downscaling (top) and EDA (bottom).

at this level is confirmed by observations. We may speculate that a too weak turbulence in free atmosphere in the HARMONIE forecasting model in combination with a too small model domain prohibits meso-scale structures to develop.

4 The linear regression model for the background error statistics

The model for the background error statistics in the Harmonie data assimilation was formulated by Berre (2000).

$$\mathbf{B}^{-1/2} = VD^{-1}F \quad (2)$$

where $\mathbf{B}^{-1/2}$ is the inverse of the square-root of the background error covariance, F is a horizontal 2-dimensional Fourier transform from physical grid-point space to spectral space, D^{-1} is a de-correlation operator and V is a vertical transform utilizing the eigenvectors of vertical covariance matrices.

2-dimensional Fourier transforms (F) together with the assumption that spectral components for different wavelengths are statistically independent implies horizontally homogeneous background error statistics in physical space. The horizontal isotropy is assumed in addition, what allows to represent horizontal covariances via a 1-dimensional covariance spectra for the control variable.

The balance operator D is derived in spectral space with a step-wise multiple regression technique for each wave number component separately.

$$\begin{aligned} \zeta &= \zeta \\ \eta &= MH\zeta + \eta_u \\ (T, P_s) &= NH\zeta + P\eta_u + (T, P_s)_u \\ q &= QH\zeta + R\eta_u + S(T, P_s)_u \end{aligned} \quad (3)$$

where $(\zeta, \eta, (T, ps), q)^T$ are forecast errors of vorticity, divergence, temperature, surface pressure and humidity on the model levels, which are splitted into the balanced and unbalanced (statistically uncorrelated) parts. $\eta_u, (T, ps)_u, q_u$ are the unbalanced components of divergence, temperature, surface pressure and humidity, H is a horizontal balance operator, which relates spectral coefficients of vorticity to those of generalized geopotential, and M, N, P, Q, R and S are vertical balance operators, which relate vertical profiles of predictors to those of predictands.

4.1 Horizontal covariance spectra for the control variables

Control variables in the HARMONIE data assimilation are vorticity and unbalanced components of divergence, temperature, surface pressure and humidity. In this study we have compared spectral densities of the control variables estimated from DA and EDA ensemble sets. Horizontal spectral densities for all control variables estimated from the DS ensemble set are obviously affected by spin-up processes for wavelengths shorter than 500 km. The reason, as discussed above, is that the Harmonie downscaling forecasts were initialized from much coarser resolution initial ECMWF EDA data. Certain spectral spinup processes are still ongoing at +12h in the down-scaling case.

In Figure 2 we show the horizontal spectral density of unbalanced humidity errors, at model level 35 (≈ 500 hPa), valid at 06UTC. Spectral densities corresponding to different forecast lengths are shown on the top (estimated from DS ensemble set) and the middle (estimated from EDA ensemble set) panels of Figure 2.

The bottom panel shows comparison of the +12h spectra of the unbalanced humidity errors estimated from the *DS* and *EDA* ensemble sets. The horizontal spectral densities of vorticity, unbalanced divergence, unbalanced temperature and surface pressure show similar behaviour. It is evident that non-physical and unrealistic features may be injected into the data assimilation system if forecast error statistics are estimated from a data set severely infected by spin-up.

We may notice as well that the spectral densities of unbalanced humidity estimated from the *EDA* ensemble set are generally larger than those estimated from the *DS* set. In fact the same is valid for the balanced part of the humidity variable what allows us to conclude that the HARMONIE forecast model generate a more vivid moisture climate than the downscaling and that the time needed to spin-up "moist" structures are quite long, more than 12 hours.

We may also observe a relatively weak error growth of the unbalanced horizontal humidity spectra, clearly superimposed on the daily cycle variations (more energy during the day time and less energy during the night). This weak error growth is present for other control variables and can be seen as a rather weak increase in the HARMONIE ensemble spread between +3h and +12h. We may speculate that this problem is associated with the LBC conditions which have a strong impact over the small Harmonie domain. The LBCs are 3-12h forecasts which are efficiently from 12h 4D-Var assimilation window, and should be considered as the analyses rather than forecast errors.

4.2 Standard deviations and balance relations

The background error variance for the different variables is another useful diagnostic tool which can be obtained by summing the spectral densities over horizontal wave numbers. Figure 3 shows what part of the temperature (top) and surface pressure (bottom) error variances is explained by vorticity and by unbalanced divergence, as a function of horizontal wave number. The statistics are estimated from the *EDA* ensemble set.

One may notice that for the largest horizontal scales (wave length around 1000 km) almost 50% of the temperature error variance, and even more of the surface pressure error variance, is explained by a linear relation with vorticity via a linear geostrophic balance. For horizontal scales shorter than 100 km this balance almost disappear. The balance between mass field and unbalanced divergence is very weak, being the largest at small horizontal scales. We may observe that the degree of balance between the surface pressure field and vorticity increases with the model integration. We may speculate that the Harmonie model establishes a large scale geostrophic balance which is different from the geostrophic balance in the real atmosphere realised via the observations. Each time observations are inserted the model is destructed from her own climate but returns to it during the model integration. Verification scores from numerous HARMONIE experiment confirm presence of large scale surface pressure biases.

There seems to be a strong interaction between mass field, vorticity, divergence and humidity on very small horizontal scales, corresponding to 2-5 grid length, that is in the range below the efficient model resolution. One may notice a peak in the explained temperature variance at exactly $2\Delta s$ which most likely points towards the numerical noise in a form of grid scale interactions between wind and temperature. Note that we have not applied the de-alising of vorticity in these experiments. This option need be investigated.

The structures of the small scale interactions between surface pressure, vorticity and unbalanced divergence are more complicated. The bottom panel in Figure 3 indicates a strong increase in the explained variance of surface pressure by vorticity and unbalanced divergence on scales below 12km, $2\Delta s - 5\Delta s$. In fact, this strong variability on the smallest scales has been introduced on purpose by a reduction of the horizontal diffusion coefficient in the Harmonie model with AROME physics. The AROME physics does not include a deep convection parameterization scheme, and relies on the model dynamics to establish convective cells that carries out the needed vertical transports associated with convection. The Harmonie AROME model has a tendency to create very unrealistic "super-cells" of convection, that could only be cured in the operational setting by a reduction of horizontal diffusion.

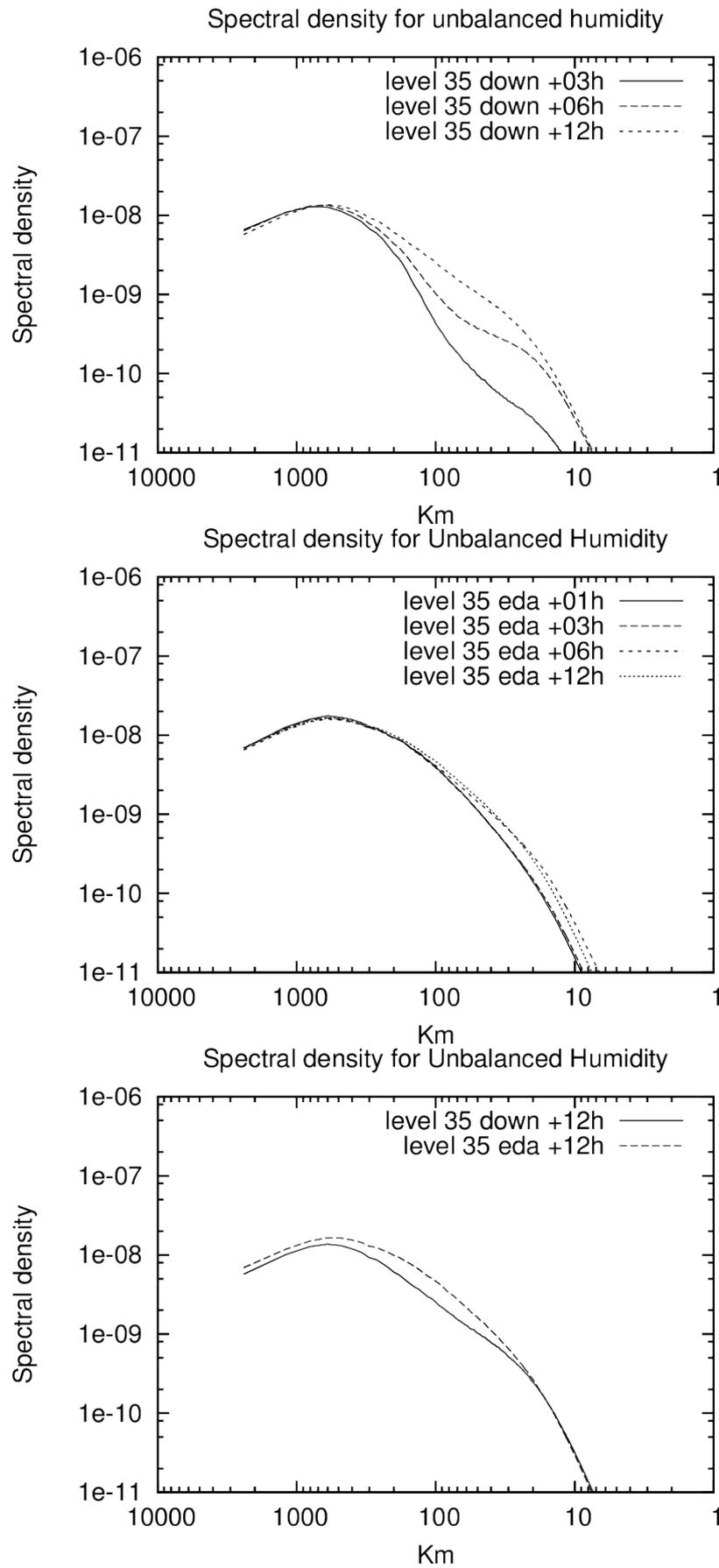


Figure 2: Spectral densities of +1h, +3h, +6h and +12h unbalanced humidity background error as estimated by the ensemble downscaling (top) and the Ensemble Data Assimilation (EDA, middle) techniques. Model level 35. Spectral densities of +12h unbalanced humidity background error as estimated by the downscaling and EDA techniques (bottom).

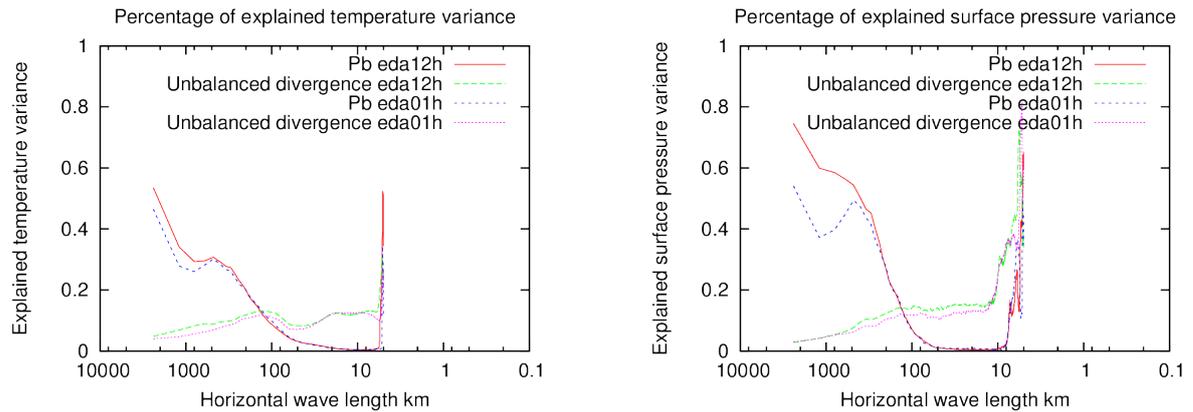


Figure 3: Part of temperature background error variance (left) and (logarithm of) surface pressure error variance (right) explained by vorticity and unbalanced divergence as a function of horizontal wave number. EDA +3h and EDA +12h.

The problem of small scale noise indicated in Figure 3 raises many questions:

- The Harmonie AROME model uses a linear grid (the shortest wave in the spectral model representation is 2 grid lengths). However the model dynamics, and in particular the model physics, include nonlinearities of higher order than the quadratic advection. The small scale noise might be due to aliasing effects. Possibly a cubic grid could be the solution.
- An important process is the forcing by orography. In order to avoid orography "Gibbs waves" over sea, the model orography is given full representation on the linear grid. Maybe it is over-optimistic to describe the forcing by orography with a 2 grid length resolution, when the model can accurately describe the time evolution of atmospheric structures of scales of 5 to 6 grid lengths or even longer.
- It seems over-optimistic to explicitly describe the effects of convective cells on the km-scale with a model using a 2.5 km grid resolution. Taking our consideration that efficient model resolution is 5-8 grid lengths, a model grid resolution below 500 meters would be needed.

5 Relevance of the statistical balances for meso-scale data assimilation

We have performed a number of simulation experiments in order to get a better insight into a relevance of the Harmonie statistical balances discussed above. Random realisations of the model state with the structures of the background error covariance were added to the Harmonie control run and the ensemble has been propagated forward in time with HARMONIE AROME model (2.5km horizontal resolution; 65 vertical levels). The internal HARMONIE perturbations were relaxed towards the ECMWF EDA perturbations on the lateral boundaries in order to assure a realistic variability on large scales. The experiment was carried out over Central European domain for a 2 weeks period in August 2012 when strong convective activities were observed.

Despite different time period and different domain, the findings in this experiment have confirmed those discussed above. Due to very weak balances on scales below 100 km and the lack of initialisation in HARMONIE data assimilation surface pressure perturbations provoke disturbances and large part of them escape from the domain with a speed of gravity waves. What happens to the remaining part of the surface perturbation depends on the weather regime (and the time of the day). The same is true for other model variables. The initial adjust-

ment process affects severely the inserted signal. The challenge of model initialisation at meso-scales should probably be addressed again.

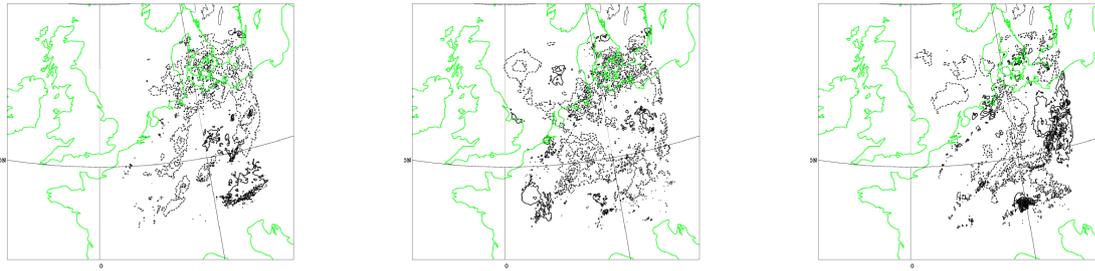


Figure 4: The initial random surface perturbation with the structure of the B-matrix covariance inserted at 16 of August 12 UTC (left) and its evolution after 1h (middle) and 3h (right) of HARMONIE AROME model integration. A strong convective activity is observed along the coast line.

The remaining part of the perturbations is advected with the flow under a relatively weak error growth under stable atmospheric conditions. After initial adjustment process has happened, a close to linear regime of the error growth lasts for for 2-4 hours for temperature and humidity. It may take a substantial time for the signal to propagate from one model state variable to another. To illustrate the mis-balance problem we show in Figure 4 the random surface pressure perturbation, with the structure of the B-matrix covariance, inserted into Harmonie control run on the 16 August 2012 at 12 UTC (left) and its evolution after 1h (middle) and 3h (right) of model integration. In order to generate the random perturbation with the structure of B-matrix covariance, the full-dimensional control vector was sampled and was projected to the model space (vorticity, divergence, temperature, surface pressure and humidity) via the square-root of the background error covariance, described above. The evolution of the surface pressure part of perturbation the "random" model state perturbation is shown in Figure 4.

On contrary in the areas with strong convective activities (vertical moisture divergence is taken here as proxy) a severe error growth may occur on the smallest scales through interactions between wind and moist air mass. Under such conditions even a very tiny humidity perturbation might result in significant changes to the cloud cover provoking feedback on the temperature, which in its turn feedbacks on cloud formation and humidity. Such process is physically consistent. One may question however how realistic it is if it happens on the wavelengths strongly contaminated by a noise due to aliasing of non-linear interactions. In Figure Figure 5 we show the inserted humidity perturbation(top left) at model level 45 and its evolution after 3 hours of model integration (top right). The 3h forecast of vertical divergence at model level 47 and the impact on the cloud cover (middle level clouds) is added to illustrate the convective development. The simulation are done in middle of the day, the 16 of August 2012 12 UTC. We may observe that small scale features in areas with the active convection along the coast line have even been developed for the surface perturbation after 3 hours of model integration (Figure 4, right). A deeper investigation is needed in order to understand to what extend the small scale process represent a realistic phenomena and what impact the numerical noise has provoking it.

6 Meso-scale structures and the assumption behind the background error covariance model

Weak balances on the scales of 10 -100 km derived from both *DS* and *EDA* ensemble sets indicate that the background error covariance model is not supported by the data. The model assumes stationarity, homogeneity and isotropy of horizontal forecast error covariance what obviously does not hold on meso-scales.

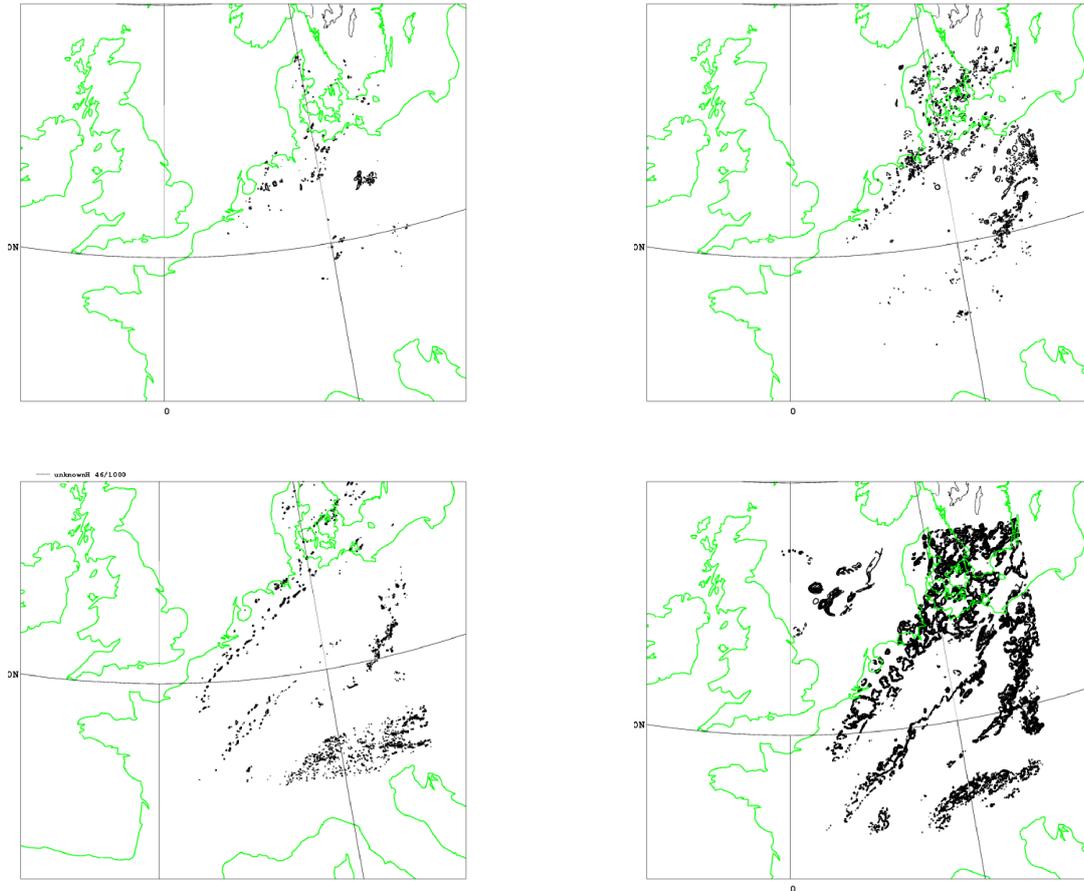


Figure 5: Humidity perturbation at initial time (16 August 2012 12 UTC) at model level 45 (top left) ; Humidity perturbation after 3 hours of model integration (top right); Vertical divergence at model level 47 (3h forecast) (bottom left) ; Change in the cloud cover (middle level clouds) after 3 hours of model integration (bottom right)

We illustrate what assumptions on homogeneity, isotropy and stationarity imply in Figure 6. We show here time-averaged horizontal correlations for 12h forecast error temperature at model level 35 with reference to a grid point with x- and y-coordinates = (235,595): the time-averaged "raw" correlations (upper, right), the correlations assuming horizontal homogeneity (lower, left) and the correlations assuming horizontal homogeneity and isotropy (lower, right). A map of orography has been added for illustration (upper, left).

Figure 6 clearly shows that time-averaging over 25 days and 6 ensemble members only is not enough to establish a stable model for the background error statistics. The effects of sampling errors are significant (upper, right). We may also notice that horizontal averaging, through the assumption of horizontal homogeneity, is a very efficient remedy to obtain stable and smooth background error statistics. The assumption of horizontal isotropy performs final polishing.

Although the smooth statistics obtained through the assumptions of homogeneity and isotropy may be attractive from a computational point of view, one may question how representative they are for real forecast error correlations with a strong case-to-case variability. To illustrate this, we have selected a single case, 12 August 2011 06 UTC + 12h, and we show in Figure 7 the same correlations maps as shown in 6 but estimated from the single case only, that is averaged over 6 ensemble members from 12 August 2011 06 UTC + 12h (Figure 7, middle). For comparison we show the horizontal correlations derived from the same data set when the homogeneity of

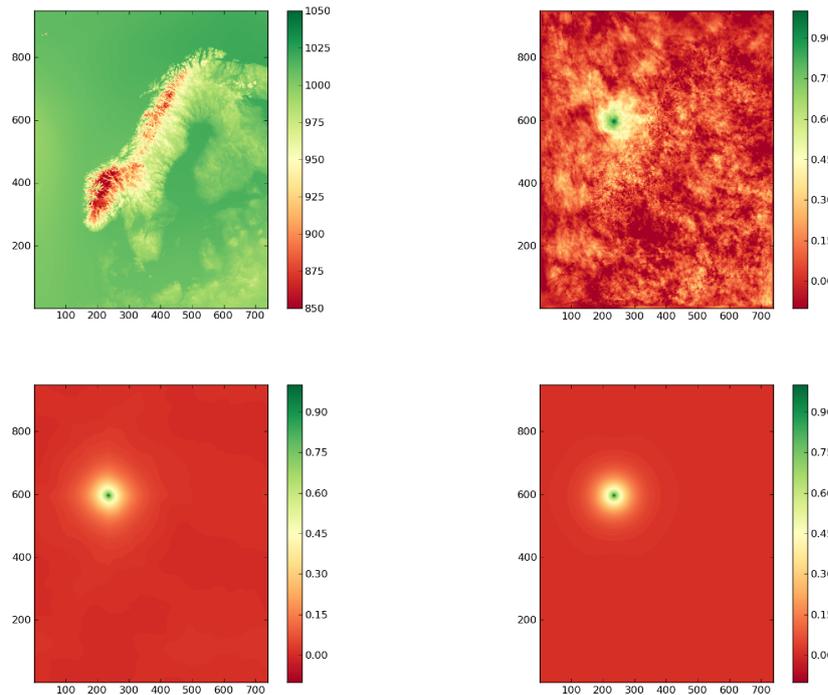


Figure 6: Surface pressure to illustrate orography (upper left). Horizontal correlations for temperature at model level 35 (≈ 500 hPa) with reference to the grid point $(x,y)=(235,595)$, averages over ensemble members and time: no further assumptions (upper right), homogeneity assumption (lower left), homogeneity and isotropy assumptions (lower right).

the forecast error covariance is assumed (Figure 7, right). We have also introduced the background state from one of the ensemble members (member 1) (Figure 7, left). The reference point for the horizontal correlations is the same as in Figure 6 and was selected based on this case. It located to be in the area of an upper air frontal zone (model level 35 ≈ 500 hPa) in Norwegian Sea close to Norwegian coast. It may appear too naive to try to estimate horizontal background forecast error correlations from 6 ensemble members only, but the message of the "raw" horizontal correlations of Figure 7 is clear and physically relevant. The correlation patterns in the vicinity of the reference point are large-scaled, elongated along the frontal zone and are indicating an uncertainty in the horizontal position of the upper air frontal zone. Note much stronger correlations in the areas with the same air mass and the drop down of the correlation in the direct vicinity of the front where different air masses meet. One can see that the assumption of homogeneity smooths away all interesting features, averaging over all grid points. Assumption on isotropy brings even more symmetry in the estimate of the background error covariance, removing remaining large scale irregularities (not shown here).

The time averaging (the assumption about stationarity) may create small scale almost isotropic horizontal correlation structures that differ significantly from correlation structures obtained for individual situations. In Figure 8 we show the horizontal correlation maps for two individual situations, 12 August (middle) and 18 August (right) 2011, all 06 UTC +12h, from the 25 cases used to create the time-averaged horizontal correlation map shown in Figure 6 (upper, left). These individual correlation maps have elongated structures in the vicinity of the reference point (235,595) but the elongation is in different directions. When stationarity is assumed, a feature typical for most of the individual cases, that the horizontal correlation decreases rapidly with distance becomes pronounced. Although the horizontal correlation maps are different for individual situation, one common property might be noticed. The correlation structures over Norwegian Sea area are generally larger than the correlation structures over the land, where convective phenomena happens. We include here again the surface pressure map to illustrate the orography (Figure 8, left). Even at 500hPa we may notice a clear response of

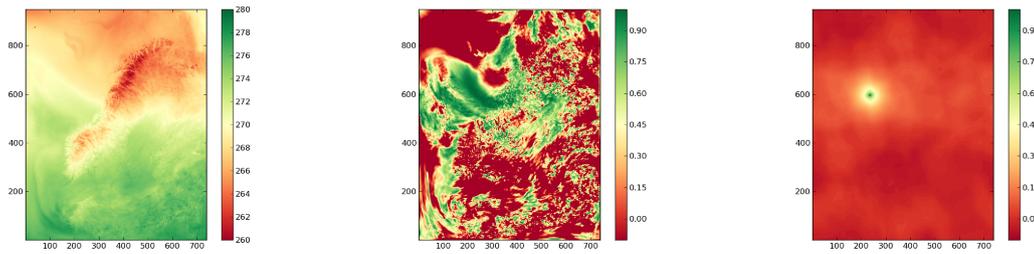


Figure 7: Temperature at model level 35 (≈ 500 hPa) 12 August 2011 + 12h, ensemble member 1 (left). Horizontal correlations for 12h forecast errors temperature with reference in the grid point $(x,y)=(235,595)$, averaged over 6 ensemble members from 12 August 2011 + 12h: (middle), the horizontal correlations assuming forecast error homogeneity (right), derived from 6 ensemble members from 12 August 2011 + 12h.

horizontal correlation structures on high orography over Norwegian mountains, resulting in the band with small scale features. The non-isotropic and inhomogeneous horizontal correlations, depending on orography and on land/sea differences, have been demonstrated in studies with observations and model simulations (Hägemark et al., 2000), (Sattler and Huang, 2002).

Examining ensemble based horizontal correlation maps for different individual cases we have noticed that a large scale signal is typically more clear than the small scale features which are stronger affected by sampling noise. Of course, six members ensemble is too small to make too general conclusions from. However, we may state that it is difficult to derive a correlation model by assuming stationarity, homogeneity and isotropy that at the same time could be useful to describe forecast error in an individual case for meso-scale phenomena. It is likely to be favourable to apply data assimilation for convection permitting models with ensemble background error statistics obtained from ensemble averaging only. One should be aware that advanced techniques might be needed to extract efficiently information on wide range of scales from "raw" ensemble covariances, because sampling noise affect different scale to different extend. When such techniques are developed, even relatively small size ensemble might be useful to describe flow-dependent covariances at meso-scales.

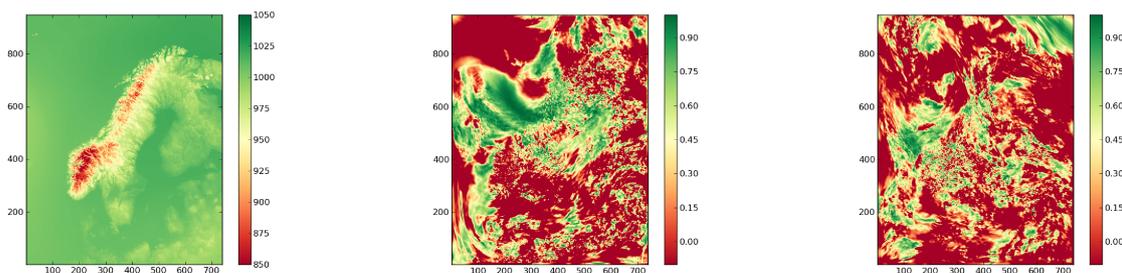


Figure 8: Surface pressure map to illustrate orography (left). Horizontal correlations for 12h forecast errors temperature with reference in the grid point $(x,y)=(235,595)$, averaged over 6 ensemble members from 12 August 2011 + 12h (middle) and from 18 August 2011 + 12h (right).

7 References

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Mon. Weather Rev.*, **128**, 644-667.

- Blazika, V., Gustafsson, N. and Zagar, N., 2015: The impact of periodization methods on the kinetic energy spectra for limited-area numerical weather prediction models. *Geosci. Model Dev.*, **8**, 87-97.
- Bubnova, R., Hello, G., Benard, P., 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. Weather Rev.*, **123**, 515-535.
- Gandin, L.S., 1963: Objective analysis of meteorological fields. *Leningrad Hydromet. Press*
- Fischer, C., Montmerle, T., Berre, L., Auger, L. and Stefanescu, S.E., 2005: An overview of the variational assimilation in the Aladin/France NWP system. *Q.J.R.Meteorol.Soc.*, **131**, 3477-3492.
- Frehlich, R. and Sharman, R., 2004: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Weather Rev.*, **132**, 2308-2324.
- Haugen, J.-E. and Machenhauer, B., 1993: A spectral limited-area model formulation with time-dependent boundary conditions applied to the shallow-water equations. *Mon. Weather Rev.*, **121**, 2618-2630.
- Hägmark, L., Ivarsson, K.-I., Gollvik, S. and Olofsson, P.-O., 2000: Mesan, an operational mesoscale analysis system, *Tellus*, **52A**, 2-20.
- Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *J. Fluid Mech.*, **388**, 259-288.
- Nastrom, G., and Gage, K., 1985: *J. Atmos. Sci.*, **42**, 950-960.
- Sattler, K. and Huang, X.-Y. 2002. Structure function characteristics for 2 meter temperature and relative humidity in different horizontal resolution. *Tellus*, **54A**, 14-33.
- Seity, Y., Brousseau, P., Malardel, S., Hello, G., Benard, P., Bouttier, F., Lac, C. and Masson, V., 2010: The AROME-France convective scale operational model. *Mon. Weather Rev.*, **139**, 976-991.
- Skamarock, W.C., Park, S.-H., Klemp, J.B., and Snyder, C., 2014: Atmospheric Kinetic Energy Spectra from Global High-Resolution Nonhydrostatic Simulations. *J. Atmos. Sci.*, **71**, 4369-4381.

Progress and plans in the ARPEGE and AROME models physics

Yann Seity, Jean-Marcel Piriou, Yves Bouteloup, Alexandre Mary, Sébastien Riette, Benoit Vié, Rachel Honnert, Clemens Wastl, Laura Rontu, Christoph Wittmann

1. Introduction

Operational configurations, especially ARPEGE T1200 and AROME 1.3 km with hourly assimilation cycle which are put into operation on 13 April 2015, are detailed in our national poster. What is described below is not operational, but in preparation for future e-suites.

2. ARPEGE

Concerning ARPEGE, PCMT convection scheme (Piriou et al,2007; Guérémy 2011) succeeds in creating more intermittent convective precipitations (Figure 1), with a delayed diurnal cycle compared to the maximum of incoming solar radiation. Inside the scheme, microphysics and transport are separated and 5 new prognostic variables are required. The microphysics is the same as the one used for stratiform part and several formulations are available for closure, entrainment and detrainment.

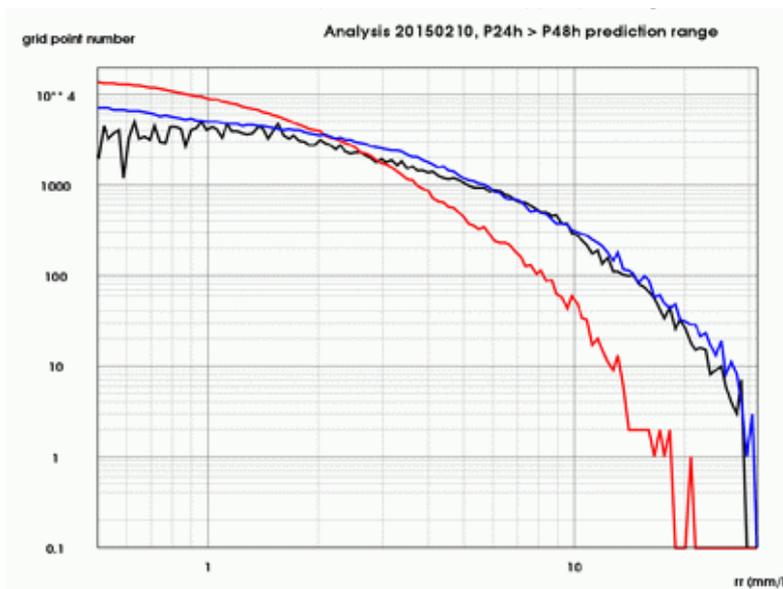


Figure 1: PDF of 3-hourly precipitations for P24 to P48 20150210 forecasts, global 0,25°x0,25° grid (red: oper, blue: PCMT, black : TRMM observations)

Numerical stabilization has been reached thanks to implicit resolution of the full EDMF equation (mass flux + turbulence).

3. AROME

Thanks to a collaboration between Météo-France, ZAMG and FMI, the HIRLAM parametrisation of orography/radiation interactions (Senkova et al,2007) has been implemented in the SURFEX version used by AROME (7.3) (Wastl et al. in this newsletter). It concerns 3 effects : orographic shadowing and slopes concern solar direct radiative fluxes, whereas reduced sky view factor concerns diffuse-solar and thermal radiative fluxes. Impact studies over the French Alps show a significant impact, even

at 2,5km grid-mesh. It increases at higher resolution (1,25km and 500m). It is stronger in winter than in summer, even if radiative fluxes are lower. With the parametrisation, the sunshine duration is much more realistic, as illustrated in Figure 2 for a winter case over the northern French Alps. Indeed, in the current operational configuration, as the orographic shadowing effects are not parameterized, the sunshine duration is the same on flat areas as over the mountains (there is only a North-South small variation due to the earth curvature). On the reverse, with the orographic shadowing effect (Figure 2, right), the sunshine duration is much more contrasted over the mountains. Maxima are observed on the tops, and minima in valleys. It can be noticed that now, as in the reality, there are some points in winter, at the northern bottom of the highest peaks, which never see the sun (sunshine duration = 0). The impact on T2m and V10m scores for 3 case studies (not shown here) is positive. Nevertheless, more systematic evaluation have shown that the reduced sky view effect leads to a significant warming in the valleys during night, which seems to be overestimated. Complementary work is needed on this part and for that reason, only shadowing and slope effects will be activated in the summer 2015 AROME e-suite.

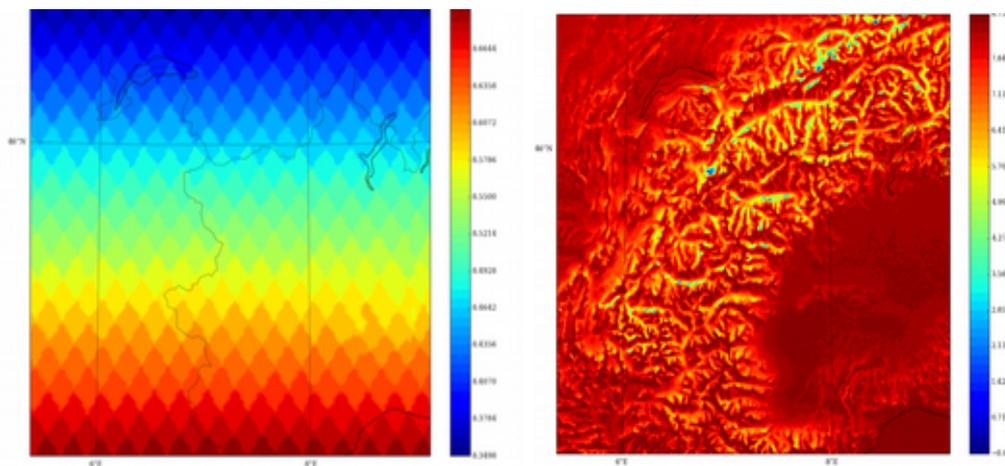


Figure 2: Sunshine time duration over the French Alps on 20131212 (left : oper with values ranging from 8h21 to 8h42, right E-suite with values from 0 to 8h42)

Concerning microphysics, thanks to a detailed study, the ICE3 sensitivity to the time step length has been reduced (there is still some on-going work on it). Concerning hail, as a few bugs have been fixed in ICE4 in MesoNH, the use of ICE4 in AROME will be re-evaluated. In parallel, there is an ongoing PhD work concerning the inclusion of microphysics in the Surface Boundary Layer scheme of SURFEX, in order to improve fog forecasts. Finally, LIMA 2- moments microphysics scheme is under validation process in Meso-NH, using for instance aircraft datas from the HYMEX dataset. A research version should be available in AROME by the end of 2015.

Concerning the turbulence scheme, an on-going work quantifies the theoretical horizontal and vertical eddy-diffusivities and the related mixing lengths at resolutions of the gray zone of turbulence where the turbulence is non-isotropic. Concerning surface, some tests of more advanced SURFEX schemes (DIFF for instance) would be performed in AROME after the implementation of SURFEX v8.

References

- Guérémy, J. F. ,2011, A continuous buoyancy based convection scheme: one- and three-dimensional validation. *Tellus A*, 63: 687–706. doi: 10.1111/j.1600-0870.2011.00521.x
- Senkova A. V., Rontu L. and Savijäri H., 2007, Parameterization of orographic effects on surface radiation in HIRLAM, *Tellus* 59A, 279-291
- Piriou J.-M., Redelsperger J.-L., Geleyn J.-F., Lafore J.P., and Guichard F., 2007, An Approach for Convective Parameterization with Memory: Separating Microphysics and Transport in Grid-Scale Equations. *J. Atmos. Sci.*, **64**, 4127–4139.

Parameterization of orographic effects on surface radiation in AROME-SURFEX

Clemens Wastl, Alexandre Mary, Yann Seity, Laura Rontu, Christoph Wittmann

1 Introduction

Topographic features like slope angle, slope aspect, sky view factor or shadowing by surrounding obstacles have a significant influence on radiation fluxes at the earth's surface. As a consequence, such characteristics can also alter temperature fields, the local circulation, the formation of clouds or the triggering of convection in such areas. With increasing spatial resolution of weather forecast models these effects gain also importance for numerical modelling. For a better representation of these topographic influences and with the aim to improve the model performance in areas with complex topography, a respective radiation parameterization scheme has been introduced to the High Resolution Limited Area Model AROME. The scheme called *ororad* is based on developments made in the HIRLAM model and has been developed in a collaboration between ZAMG, Météo France and FMI. Additional information can also be found in the contribution of Yann Seity about AROME physics in this newsletter.

2 Ororad scheme

The basic idea behind the *ororad* scheme has been developed by Müller and Scherer (2005) who designed a unified radiation parameterization scheme based on high resolution topography data. For the present *ororad* scheme in AROME high resolution orography data from the Shuttle Radar Topography Mission (SRTM, Rodriguez et al., 2005) with a horizontal resolution of three arc-seconds (about 90m in Central Europe) have been used. Slope and shadow factors have been applied to account for the different slope angles and directions with respect to the current solar position and estimate the relief shadows due to neighbouring obstacles. These factors are influencing the downwelling direct shortwave radiation flux at the surface (SW). The restricted visibility of sky is described by the sky-view factor and modifies the diffuse SW and downwelling longwave radiation (LW, Figure 1).

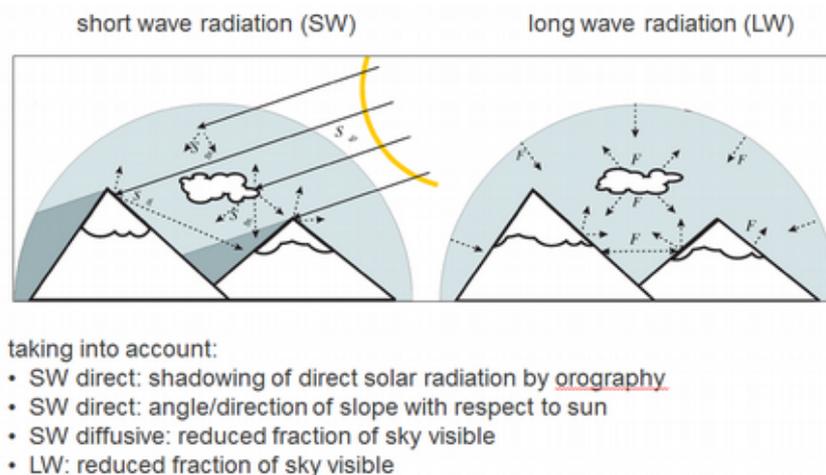


Figure 1: Ororad scheme taking into account orographic effects on short wave and long wave radiation fluxes at the surface.

Some scripts/programs for preprocessing of the SRTM data and to calculate the necessary orographic parameters were applied from HIRLAM, but finally these will be replaced by processing within the SURFEX PGD system currently developed at Météo France. The respective orographic fields have to be calculated only once and are finally included in the PGD output file. During the model integration, when the solar elevation and azimuth angles are known, the direction-dependent coefficients accounting for the shadowing by neighbouring obstacles are calculated. The three factors slope, shadow and sky-view can easily be switched on/off by SURFEX namelist settings.

Figure 2 (left) shows a comparison of measured downwelling SWR flux and the modelled one for an Alpine station in a narrow north-south oriented valley in Austria for a clear day in March 2014. Model simulations were done with AROME-SURFEX cy40t1 and a horizontal resolution of 1km. Slope effect is not considered here, because the station observations are made on a plane area. The shadow factor (DSH) causes a significant improvement of the modelled SWR during the morning and evening hours (reduction of about 100 – 150 W/m²). Sky view has a comparatively low influence on SWR, but a big influence on the LWR fluxes during the night (not shown, because no LWR measurements available at this station). Looking at the slope effect of the *ororad* scheme on the temperature at the same day 12 UTC in Figure 2 (right) shows a temperature increase on the sunny slopes up to 3°C and a decrease on the shady slopes up to -4°C compared to the reference run without *ororad*.

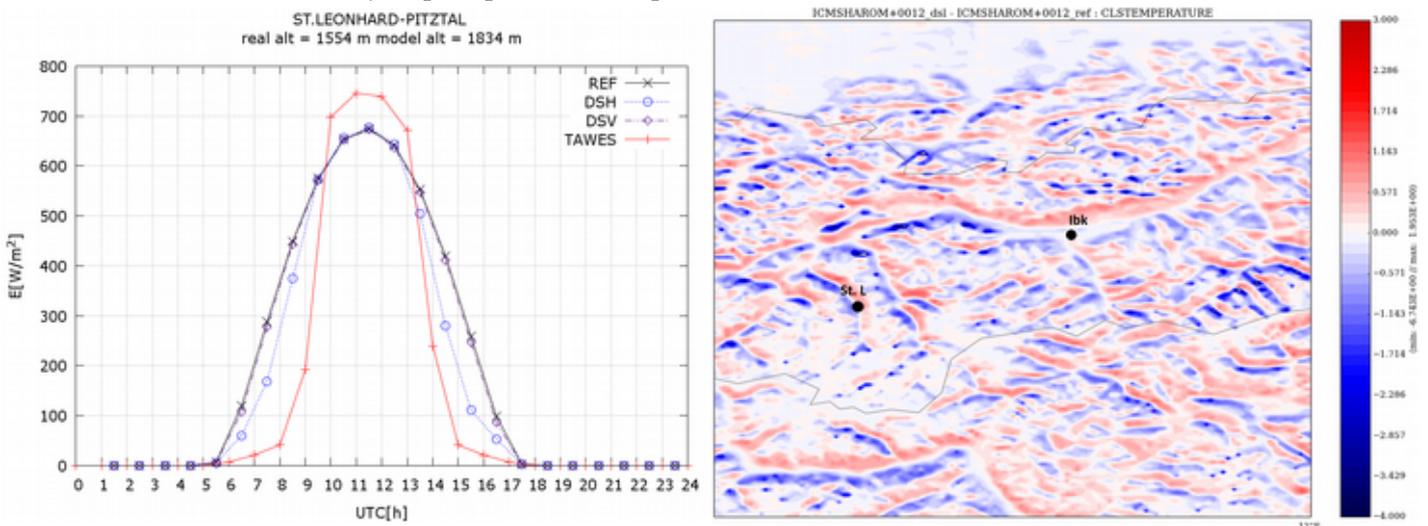


Figure 2: Left: Comparison of measured and modelled SWR flux (1 hourly mean values) at the valley station of St. Leonhard –Pitztal in Tyrol. Red curve shows the station observation (TAWES), black curve the reference AROME run without *ororad* (REF), blue one is for shadow effect only (DSH), magenta one for sky view effect (DSV). Right: 2m temperature difference *ororad* – REF for 12 UTC at March 12, 2014 for the area of Tyrol in Austria. Black dots show the location of the capital city Innsbruck (Ibk) and of the station St. Leonhard-Pitztal (St. L).

First tests and verification of case studies have shown that slope and shadow effect help to reduce temperature BIAS in narrow valleys. But, the restricted sky-view during the night results in a general warming of the valleys of more than 3 °C (not shown) in the model which is far too high compared to observations. Hence further tests and improvements of the scheme have to be done in near future, e.g. an E-suite of CY41T1 with *ororad* will be activated end of August at Météo France, and in autumn 2016 at ZAMG (based on CY40T1).

3 References

- Müller, M. D. and D. Scherer, 2005: A grid- and subgrid-scale radiation parameterization of topographic effects for mesoscale weather forecast models. *Mon. Wea. Rev.* 133, 1431-1442.
- Rodriguez, E., C. Morris, J. Belz, E. Chapin, J. Martin, W. Daffer, and S. Hensley, 2005: An assessment of the srtm topographic products. Technical Report JPL D-31639, Jet Propulsion Laboratory, Pasadena, California, 143. Available at <http://www2.jpl.nasa.gov/srtm>.

Description of the OCND2-option in the ICE3 clouds- and stratiform condensation scheme in AROME

Karl-Ivar Ivarsson, Swedish Meteorological and Hydrological Institute, Sweden

1 Introduction

Although the cloud physics in AROME is advanced and generally works well, there are two weaknesses that are addressed here.

The first one is that low clouds disappear too quickly in 'moderate' cold conditions which means typical 2m- temperatures about 0 to -10 degrees Celsius. This also leads to too much outgoing long-wave radiation and too low temperatures near the surface.

The second one is that in case of more severe cold weather, below -25 C, there is too much low clouds instead, including clouds at the very lowest level, 'ice-fog'. There also seems to be a moderate overprediction of cirrus clouds.

The reason for that low clouds disappear too quickly in 'moderate' cold conditions is a too quick decay of mixed-phase clouds. This decay is caused by a too fast growth of cloud ice water and solid precipitation, which too quickly removes moisture.

In case of more severe cold weather, there is too much of ice clouds. The reason for this is that those clouds appear as soon as the relative humidity with respect to ice reaches 100% in any part of a grid-box. In reality, supersaturation with respect to ice is common, and in such cases, the amount of cloud ice water is often too low for generating a cloud.

On the other hand, solid precipitation may have cloud-like properties, so occasionally there is 'ice-clouds' also in case of a relative humidity with respect to ice that is less than 100%. Another reason for this is that cloud ice water takes some time to evaporate.

In section 2, there is a short description of the modifications of the ICE3-physics within the OCND2 option. In section 3, the experiment with this new option is described, in section 4 the result of this experiment is presented, followed by a short summary in section 5.

2 A short description of the modifications of the ICE3-physics within the OCND2 option.

One important modification is to have a more rigorous separation of the fast liquid water processes from the slower ice phase processes. This is achieved by the following modifications:

- The statistical cloud scheme only handles water clouds and mixed phase clouds. Only the amount of cloud liquid is calculated from this scheme.
- The Bergeron-Findeisen process is derived as a conversion from vapour to ice.
- Tunings and modifications of the conversion from ice to vapour and vice versa, since those processes seem to be too fast in the original version of ICE3.
- Tuning of the ice nucleus (IN) concentration.

- A separate ice cloud fraction is derived using the relative humidity with respect to ice and the content of cloud ice water and of solid precipitation. The ice cloud fraction is also dependent of model thickness, since ice-clouds are generally optical thinner than water clouds.
- Total cloud cover is the sum of the liquid (mixed-phase) fraction and the ice fraction.

The modifications are partly based on earlier work with the HIRLAM cloud physics, for details see Ivarsson (2010) and influenced by Liu X et al. (2007) and Xu and Randall, (1996) .

3 The experiment set up used for the comparison between the modified scheme (OCND2=T) and original one (OCND2=F)

One cold winter period (November 15 to December 10 2010) and one wet summer period (August 10-23 2011) are used. Some details of the experiment is described in table 1.

Table 1: Experiment description

Model version	Cy 38h1.1
Resolution	2.5 km
Levels	65
Area	MetCoOp domain, 750x960 grid points. Covers north-western Europe.
Initial conditions	3DVAR and surface analysis
Boundaries	ECMWF boundaries every 3 hours
Surface scheme	SURFEX
Starting times used for verification	00 UTC and 12 UTC
Cycle interval	3 hours
Verification times	Every 3 or 6 hours (every 12 hours for precipitation)

The first 5 days are omitted for winter period and the first 2 days for the summer period. This is done in order to avoid too much spin-up effects.

4 Verification results

Some general comments

The OCND2-option is derived in order to improve the model performance in winter. As expected, the effects seen for the summer period is limited and the only difference of particular interest is that the over-prediction of large precipitation amount is reduced in summer. (An unexpected improvement.) Therefore, only some results from the winter period are described in some more detail.

2-metre temperature and low clouds for the winter period

The result for the winter period is seen in figure 1

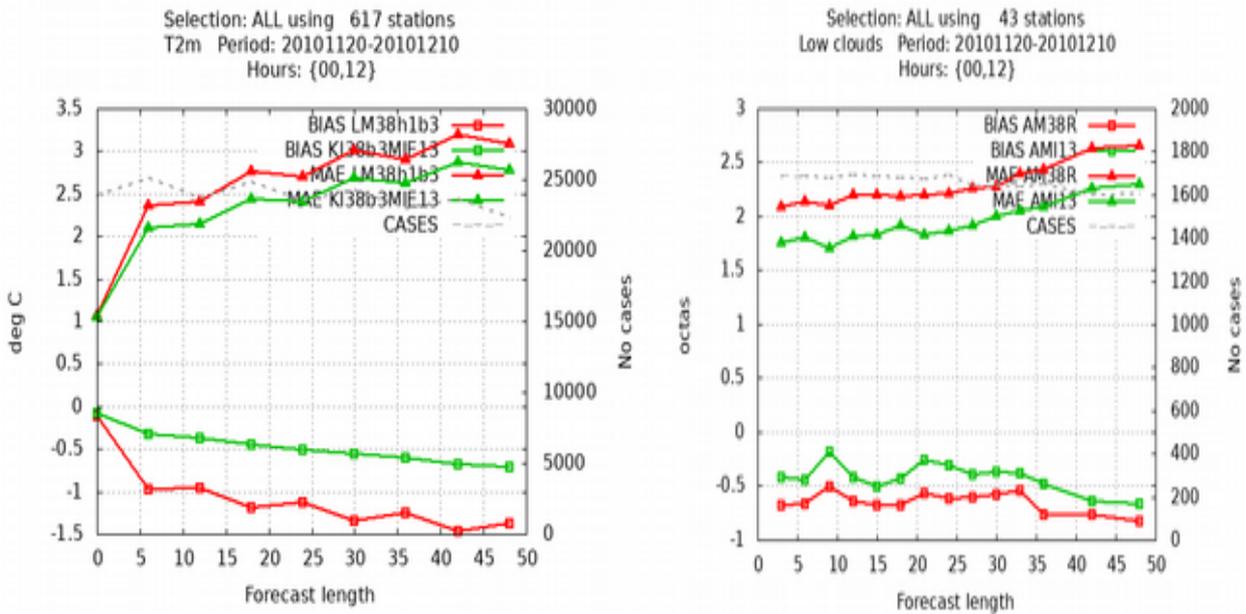


Figure 1: To the left: Mean error (BIAS) and mean absolute error (MAE) for 2-metre temperature. OCND2=F in red and OCND2=T in green. To the right : The same for low clouds compared to Swedish automatic stations.

With OCND2=T, the low clouds are increased (and contains more of liquid water, not shown) which reduced the outgoing long wave radiation. The reduces the negative bias of the 2-m temperature.

Cloud base

The result for cloud base verified against Swedish automatic stations is seen in figure 2. The frequency bias (FB, which should be near unity) and equitable treat score (ETS, higher value is better) are used.

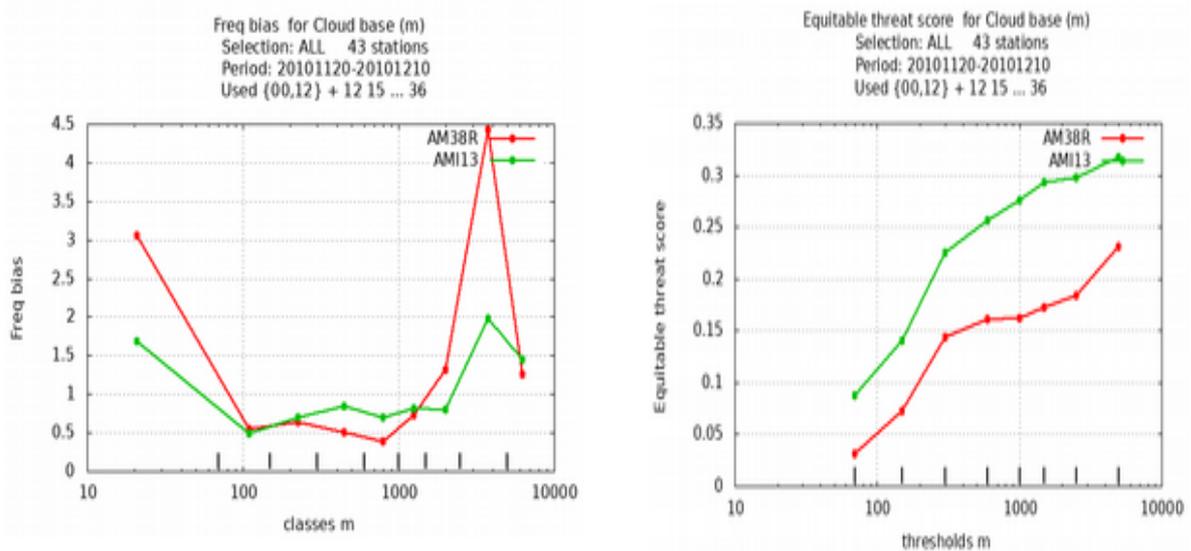


Figure 2: To the left: Frequency bias for different cloud bases for cloudiness of at least 3 octas. OCND2=F in red and OCND2=T in green. To the right: The same, but ETS for cloud base exceeding different thresholds of cloud base.

The over-forecasting of cloudiness for the lowest level (fog) is reduced with OCND2=T . There are too often cloud bases around 3000 metres with OCND2=F. This is probably due too lack of lower clouds. This error is reduced with OCND2=T. Probably the reduction of those errors contribute to a large part of the increase of ETS seen when OCND2 is set to true.

Other results of interest.

Those (both improvements and deficiencies with OCND2=T) are summarized in the following table :

Table 2: Important differences seen between OCND2=FALSE and -TRUE:

Parameter	Difference using OCND2=T relative to using OCND2=F
2m dew point temperature winter	Better (less bias and absolute error)
Total cloud cover winter	Better (less bias and absolute error, higher ETS)
2m specific humidity winter	Better (less bias and absolute error)
10m wind speed winter	Little worse (somewhat more positive bias and increased absolute error)
Mean sea level pressure winter	Little worse (somewhat more positive bias and increased absolute error)
Upper air temperature winter	Little worse since there is a cooling in the lower troposphere.
Upper air relative humidity winter and summer	Increased absolute error, but more realistic distribution. (Both better and worse)
Upper air specific humidity winter	Little better (somewhat less bias, and absolute error in lower troposphere, a little more realistic distribution)
Spread of the 12 hours precipitation amounts winter and summer	Little better (somewhat less spread, which is closer to observations)
12 hour precipitation frequency bias winter	Increased for lower amounts (too much of lower amounts which is worse)
Cloud liquid water	Increased for supercooled clouds (probably better in most cases, but hard to confirm by observations)
Cloud ice water	Decreased for supercooled clouds (probably both better and worse dependent on the situation, but hard to confirm by observations)
Graupel	Increased, probably a deterioration since it seems to be already too much graupel with OCND2=F

5 Summary

A modification of the ICE3-physics which can but used by setting OCND2 to true, has been tested for a cold winter period and for a wet late summer period. This modification has only small effects on the forecasts in summer but reduces the negative bias of the 2m – temperature in winter, since the amount of low clouds (mostly supercooled liquid water) increases. The prediction of cloud base is improved in winter.

Important drawbacks noticed is that there is a some cooling of lower stratosphere, leading to a small increased positive bias of the mean sea level pressure. Small amounts of precipitation occurs too often in winter with this modification.

6 References

- Ivarsson K.I 2010 : The impact of cloud condensation nuclei (CCN) concentration and ice nucleus (IN) concentration on clouds and precipitation in HIRLAM.HIRLAM newsletter No 55 part B, 48-54
- Liu X et al. 2007: Inclusion of ice microphysics in the NCAR Community Atmospheric Model Version 3. (CAM3) Journal of Climate Vol 20 4526-4547
- Xu and Randall, 1996 : A semiempirical cloudiness parameterization for use in climate models. J.A.S Vol 53 3084-3101

KNMI parameterization testbed as an evaluation tool: Status and plans

Wim de Rooy, Roel Neggers, Pier Siebesma, Bert van Uft, Willem Vlot, Cisco de Bruijn, Emiel van der Plas, Eric Bazile, Henk Klein Baltink, Fred Bosveld, Toon Moene

1 Introduction

As described in Neggers et al. 2012, the KNMI Parameterization Testbed (KPT) consists of two main components: i) an archive of data streams and ii) an interactive graphical user interface (GUI) for the visualization and intercomparison of these data streams.

The data streams include a wealth of observations, mainly for Cabauw, as well as model output. Among the available Cabauw observations are measurements from a 213m high tower, all kind of surface and radiation flux measurements, lidar data, cloudnet data etc. etc. Model data can be divided in extractions for the Cabauw location from 3D model runs, LES runs (limited periods), as well as 1D runs. The 1D runs mainly concern different flavours and experimental versions of the Harmonie model and are driven by RACMO (Regional Climate Model) which provides the large scale forcing.

The graphical interface allows quick visualization of all data. At the externally accessible website, testbed.knmi.nl, only prefabricated plots are available. Once some security issues are solved the full interactive graphical user interface will become available also outside KNMI.

This paper mainly describes the presentation at the ASM in Helsingor 2015. More results from the testbed will be presented in forthcoming papers.

2 The objectives of the Testbed

Why do we need a testbed?

Typical problems related to validation and verification studies are:

- Different verifications concerning the same process can not be compared because they use different set-ups, e.g. concerning the cycle version or verification period.
- If a modification has positive impact on the scores, it does not tell us if this change is a real improvement or just a compensating error.
- Longer-term verification is needed to produce statistically significant results.

Some of the above mentioned problems can be tackled in the testbed:

- Modifications are compared within the same set-up (i.e. one cycle version and for the same time period)
- In depth investigation of all processes involved (chain) thanks to the availability of many (advanced) observations and sometimes LES. This enables “real” model improvements. A nice example of this approach is described in Neggers & Siebesma 2013.
- Models run daily, so not only on “golden” days. Also, reforecasting for long periods can be done relatively rapidly in 1D mode.

However, there are also some disadvantages of the testbed approach.

- How representative are 1D results, driven by RACMO, for 3D results? As we also have extractions for the Cabauw location available from 3D model runs with the same version as the 1D runs, this can and will be investigated.
- How representative are the results for other locations as only one (or a few) locations are studied? Additional verification over different (large) areas is necessary.

3 Testbed results and plans

The testbed turned out to be a useful tool to perform model sanity checks. Runs with cycle version 38h1.1 revealed a large imbalance of the energy at the surface. This imbalance was probably related to a bug in the radiation routines and was removed in cycle 38h1.2 (see Fig. 1).

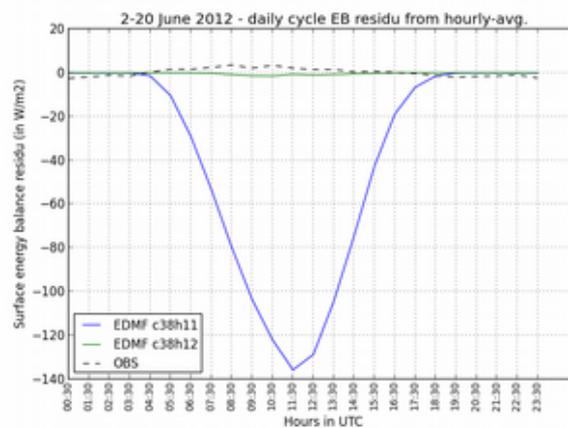


Figure 1: Mean daily cycle of the surface energy balance (W/m^2) from 2 to 20 June 2012 for cycle 38h1.1 and 38h1.2 1D runs and observations (Vlot 2015)

Apart from removing the surface energy imbalance, running with cycle 38h1.2 instead of 38h1.1 also resulted in a substantial improvement of the results. This is illustrated in Fig. 2 showing the impact on the latent heat flux and the T2m.

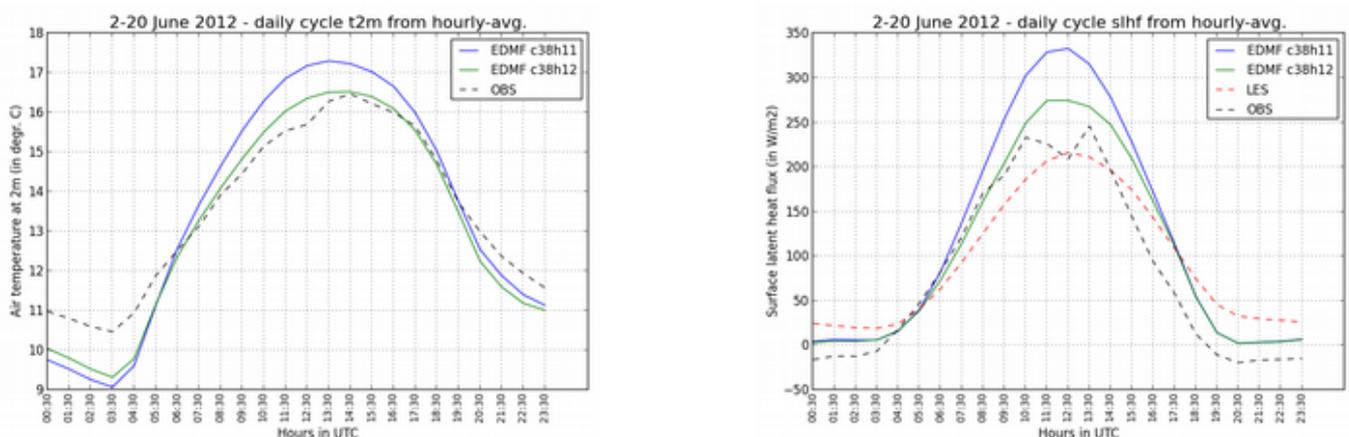


Figure 2a, b: Mean daily cycle of a) the 2m temperature [$^{\circ}C$] and b) the latent heat flux (W/m^2) from 2 to 20 June 2012 as observed and for cycle 38h1.1 and 38h1.2. 1D runs (Vlot 2015). The right panel also shows latent heat flux of LES runs.

The large influence of the updates and bugfix in cycle 38h1.2 meant that the large dataset of 1D runs with cycle 38h1.1 could not be trusted, at least not in an absolute sense. However, the impact of many modifications can be understood and are clearly recognized in testbed plots. For example, we can see the large (positive) impact of the extra variance term (de Rooy et al. 2010), the effect of extra top entrainment when the RACMO turbulence scheme is used (Fig.3) and the effect of putting the inhomogeneity factor from 0.7. to 1 results in less SW radiation and consequently lower H, LE, T2m, and q2m.

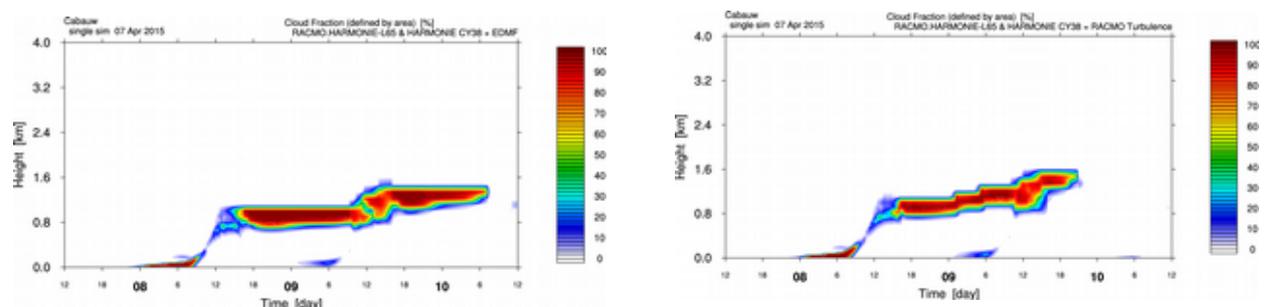


Figure 3a, b: Time versus height contour plot of cloud cover in Cabauw starting at 7 April 2015 at 12 UTC. On the left a) 1D run (cycle 38h1.2) with the default Harmonie options and at the right panel b) a 1D run with Racmo turbulence. The effect of more top entrainment with Racmo turbulence is reflected in increased rising of the cloud cloud layer and the acceleration of the breaking up of the stratus.

Plans

One of the objectives of the Testbed is to study the validity of the 1D approach for validation purposes. This is done by comparing output for the Cabauw location of 1D and 3D runs in the same configuration. Monthly mean averages of several parameters of such comparisons reveal that significant differences can occur. Note that if we look at monthly means, cases with e.g. the passage of frontal systems are included. The description of such large scale systems will be mainly determined by the host model (for 1D RACMO) so differences between 1D and 3D runs can be expected. A more detailed discussion on the validity of the 1D approach can be found in a internship report of Willem Vlot (to be published in August, 2015). Based on the conclusions we will start to develop driver files from Harmonie 3D runs i.o. RACMO. Additional advantage of using a Harmonie driver file is the initialisation and characterization of the surface which can simply be copied from the host model.

Within the Testbed we will continue to perform daily evaluations of

- 1D runs with many options; EDMF, EDKF, HARATU (HARmonie with RACmo TURbulence), LOCND2=FALSE, runs with ARPEGE driver files, radiation updates, and possibly your modifications!
- extractions for Cabauw from 3D runs: Harmonie cycle 36h1.4 (operational version at KNMI), default cycle 38h1.2, cy38h1.2 with HARATU. Results of the extractions are also send to cloudnet (see www.cloud-net.org) for an accurate validation of vertical profiles of cloud cover, liquid water and ice.

Additionally, reforecasts will be done, e.g. for 2012 when LES results for the full year are available (Schalkwijk et al. 2015) and August 2006 (high precipitation amounts, see Vlot 2015)

4 Conclusions

Much technical work including model sanity checks are done to revitalize the KNMI parameterization testbed. Some of the strengths of the testbed are:

- Combining many observations (and LES) to investigate a chain of processes to gain insight in the model.
- Long-term validation of many model versions becomes possible.
- Validation is based on one set-up to facilitate a fair comparison of different model versions
- Link the Hirlam/Aladin community
- Potential for real model improvements

On the other hand a possible weakness is the representativeness of 1D runs for 3D performance. This depends on the process studied. For example effects of a modification on deep convective precipitation can not be studied due to the lack of meso-scale circulations in a 1D environment. Therefore, focus should be on the fast (mainly) boundary layer processes. Some of the concerns about the representativeness will be reduced by embedding 1D runs in a Harmonie host model.

5 References

R. A. J. Neggers, A. P. Siebesma, and T. Heus, 2012: Continuous Single-Column Model Evaluation at a Permanent Meteorological Supersite. *Bull. Amer. Meteor. Soc.*, **93**, 1389–1400.

R. A. J. Neggers and A. P. Siebesma, 2013: Constraining a System of Interacting Parameterizations through Multiple-Parameter Evaluation: Tracing a Compensating Error between Cloud Vertical Structure and Cloud Overlap. *J. Climate*, **26**, 6698–6715.

de Rooy, W.C., de Bruijn E.I.F., Tijn A.B.C., Neggers, R.A.J., Siebesma A.P., and Barkmeijer J., 2010: Experiences with Harmonie at KNMI, HIRLAM newsletters, **56**, 21-29.

Vlot, Willem, to be published in August 2015: Validation of parameterizations in Harmonie using SCM & LES output at the Cabauw meteorological supersite. MSc Internship report Wageningen University, KNMI. (contact rooyde@knmi.nl for a copy).

Schalkwijk, J., Jonker H.J.J., Siebesma A.P., Bosveld, F.C., 2015: A year-long large-eddy simulation of the weather over Cabauw: An overview. *Monthly Weather Review*, **143**, 828-844

Shortwave Radiation Experiments in HARMONIE

Tests of the cloud inhomogeneity factor and a new cloud liquid optical property scheme compared to observations

Emily Gleeson, Kristian Pagh Nielsen, Velle Toll, Laura Rontu, Eoin Whelan

1 Introduction

Within the European Union, at least 20% of each country's total energy consumption must come from renewable sources by the year 2020, increasing to at least 27% by 2030. Solar energy is currently one of the least expensive forms of clean energy. With the growing interest in and use of solar power comes the need for reliable solar or shortwave (SW) radiation forecasts.

SW radiation strongly impacts on weather at the Earth's surface and the development of the atmospheric boundary layer. Thus, improvements in the representation of SW radiation are important for the ongoing improvement of numerical weather prediction (NWP) forecasts. Accurate SW radiation output from NWP models relies on the accuracy of **1.** cloud cover, **2.** the physical properties of clouds (liquid water load, ice water load, effective water and ice radii), **3.** the optical properties of the clouds, **4.** the radiative transfer approximations and **5.** aerosols and atmospheric gases. In this study, we mainly focus on 2 and 3, the influence of the physical and optical properties of clouds.

Regarding aerosols (**5**), HARMONIE uses monthly climatologies of vertically integrated aerosol optical depth (AOD) at 550nm (Tegen et al., 1997). The aerosol optical properties (single scattering albedo (SSA), asymmetry factor (g) and AOD scaling for each SW radiation band) are parametrized following Hess et al. (1998). Toll et al. (2015) found a noticeable improvement in the HARMONIE NWP forecast for a heavily polluted Russian wildfire case study when the direct radiative effect of the real time aerosol distribution was used instead of the climatological distribution. For situations where the aerosol distributions are close to average, updating the aerosol climatology or using real time aerosol distributions only results in small improvements (Toll et al., in preparation). The indirect radiative effect of aerosols in HARMONIE has not yet been extensively studied.

The radiative transfer approximations (**4**) have been extensively tested by Nielsen et al. (2014) using MUSC, the single column version of HARMONIE. In that study, the SW radiation schemes in the model were compared to the accurate DISORT model run within the libRadtran framework (Stamnes et al., 1988, 2000; Mayer and Kylling, 2005) for a range of clear sky, cloud liquid and cloud ice experiments. The benchmark tests included a study of the cloud inhomogeneity factor, the current SW cloud liquid optical property parametrizations available in HARMONIE (Fouquart, 1987 and Slingo 1989) and a new parametrization proposed by Nielsen.

The present study using 3D HARMONIE expands on the 1D benchmark tests done using MUSC. Here we focus on the influence of the cloud inhomogeneity factor, which effectively modifies the cloud water and ice loads used in the radiation calculations, and the new Nielsen cloud liquid optical property scheme on forecasts for the Irish operational domain.

Regarding the cloud inhomogeneity factor, there are several studies that try to quantify the radiative response of the inhomogeneity of different cloud types using observations (e.g. Pomroy and Illingworth, 2000; Hogan and Illingworth, 2003). Oreopoulos and Cahalan (2005) have investigated the climatology of cloud inhomogeneity using MODIS data.

Subgrid scale variability in cloud properties induces errors in simulated longwave (LW) and SW radiative fluxes in global climate model simulations when this variability is not accounted for (Scheirer and Macke, 2003; Fu et al., 2000). The relationships between both SW albedo and LW emissivity and cloud optical thickness are nonlinear so that inhomogeneous clouds have a lower mean SW albedo and a lower mean LW emissivity than homogeneous clouds with the same mean optical thickness. The inhomogeneity parameter χ is defined as the ratio of the exponential of the logarithmic average of the cloud optical thickness to the linear average of the cloud optical thickness (Equation 1). This parameter can be used as a good approximation of the effective optical thickness in the case of inhomogeneous clouds (Oreopoulos et al. 2005; Cahalan et al. 1994).

$$\chi = \frac{e^{\overline{\ln \tau}}}{\overline{\tau}} \quad (1)$$

In the current SW and LW radiation schemes used by default in the HARMONIE model (Seity et al., 2011, Brousseau et al., 2011) an inhomogeneity correction factor has been used to scale the cloud optical thickness (cloud water and ice loads) to account for subgrid cloud variability following Tiedtke (1996). A value of 0.7 was chosen following an observational study by Cahalan et al. (1994). This inhomogeneity factor and the radiation parametrizations originate from cycle 25R1 of the ECMWF global model IFS (see ECMWF, 2012 and Mascart and Bougeault, 2011). Tiedtke (1996) showed that this cloud inhomogeneity factor led to a 9 W/m² increase in the average net downward SW radiation flux over tropical oceans.

In HARMONIE deep convection is treated explicitly at the default horizontal grid spacing of 2.5km. In addition, cloud structure is better resolved compared to the global IFS model which has coarser grid spacing (T511 or ~40km grid spacing was used in the IFS model at the time when the cloud inhomogeneity factor was introduced). Consequently, it is not physically correct to use a correction factor of 0.7 for cloud optical thickness in HARMONIE at high horizontal resolution (pers. comm. Hogan 2014 unpublished Townsend dissertation 2015) and in this study we investigate the effect of this inhomogeneity factor on HARMONIE NWP forecasts for Ireland and the UK.

In Nielsen et al. (2014) the Nielsen scheme was shown to be better than the Fouquart and Slingo schemes for a suite of 1D cloud liquid tests. The new scheme was developed because initial tests of the Fouquart and Slingo cloud liquid optical property schemes in HARMONIE showed significant deviations from the Mie calculations. The new Nielsen scheme is based on empirical fits to the Mie calculations (see Nielsen et al., 2014 Supplement 1). To complete the validation of the scheme, 3D experiments were carried out and are presented here.

The final part of this study involved using downward global SW radiation as a proxy for cloudiness by computing the clear sky index (downward global SW radiation normalised by the clear sky downward SW radiation) and the evaluation of HARMONIE cloud cover and cloud condensate compared with MSGCPP satellite data (Roebeling et al., 2006). Perez et al. (2014) performed a detailed analysis of the HARMONIE cloud cover forecasts. They also discussed the diagnosis of cloud cover variability in terms of the three-dimensional fractional cloud cover defined in each gridbox. Three-dimensional fractional cloud cover is also used by the radiation parametrizations to derive the grid-scale cloud condensate load from the cloud liquid and ice content given by the microphysics parametrizations. The analysis that we show here is complementary to the analysis of Perez et al. (2014) in that we also analyse the cloud condensate load, which can vary independently from the cloud cover.

2 HARMONIE Set-up and Experiments

HARMONIE cycle 38h1.2 on a 2.5km horizontal grid, with 540x500 grid-points in the x- and y-directions and 65 hybrid model levels (Simmons and Burridge 1981; Laprise 1992), was used for the experiments described in this article. The model domain was centred over the island of Ireland and includes the UK and part of northern France.

Lateral boundary conditions from the ERA-Interim re-analysis project (Dee et al., 2011) at a 3-hour frequency were used. Observations, retrieved from ECMWF's MARS archive and Met Éireann's observation database, were assimilated using the model's three-dimensional variational (3DVAR) data assimilation system (with no digital filtering). Observations from land surface stations, ships and drifting buoys, radiosonde ascents and aircraft were assimilated. Surface observations were also assimilated in the model's surface analysis system using optimal interpolation. The experiments were configured to use a 3 hourly cycling strategy with a 1.5 hour observation window.

This body of work includes 3 radiation experiments. The default SW radiation scheme in HARMONIE is the IFS scheme (ECMWF cycle 25R1) with six SW spectral bands, three in the ultraviolet/visible spectral range and three in the solar infrared range (Mascart and Bougeault, 2011). The delta-Eddington approximation (Joseph et al., 1976; Fouquart and Bonnel, 1980) is used for the radiative transfer calculations with the cloud liquid and cloud ice optical properties calculated with the Fouquart (1987) and the Ebert and Curry (1992) parametrizations respectively. Before being used for radiation calculations the cloud water load is modified by so-called cloud SW and cloud LW inhomogeneity factors, each of which is set to 0.7 in the default set-up.



Figure 1. Map of Ireland showing the locations of the shortwave radiation measurement sites.

In addition to the default IFS SW radiation scheme (REFEXP), we ran 2 comparison experiments (INHOMEXP, COPEXP). In the first of these (INHOMEXP), we set both the cloud SW and cloud LW inhomogeneity factors to 1.0 as the current default values of 0.7 were set for IFS model versions with horizontal grid spacings of ~10-100km. HARMONIE is run on a 2.5km grid. At this grid spacing sub-grid cloud inhomogeneity has a smaller effect on the average cloud optical thickness. Thus, we suggest inhomogeneity factors of 1.0 instead (Nielsen et al., 2014). In the second experiment (COPEXP), we used a new cloud liquid optical property scheme (the Nielsen scheme, Nielsen et al., 2014). The default cloud ice optical property scheme and the delta-Eddington approximation for radiative transfer

were still used. Both the cloud SW and LW inhomogeneity factors were set to 1 as in the previous experiment. Therefore, COPEXP is identical to REFEXP except for changes to the cloud liquid optical property scheme and the inhomogeneity factors in the radiation parametrization.

Two month-long simulations were carried out: June and December 2013, with each having a 10-day spin-up period during the previous month. 48-hour forecasts were run at 00 and 12 UTC with 3-hour forecasts run at each of the remaining cycles for data assimilation continuity.

In HARMONIE's Webgraf verification package observations from synoptic stations and upper-air soundings covering the domain were used. SW radiation observations from 7 synoptic stations in Ireland (Belmullet, Clones, Dublin Airport, Gurteen, Johnstown Castle, Malin Head, Valentia – see Figure 1) and MSGCPP cloud cover and cloud condensate data were also used to verify the output from HARMONIE.

3 Results and Discussion

The results are presented in four sections, focusing on June 2013 because during this period the investigated effects (difference in cloud inhomogeneity, cloud liquid optical property scheme) induced changes. This was expected as the changes are predominantly SW radiation-related and during December in Ireland, SW radiation is significantly lower than in June.

Forecast verification using observations recorded at synoptic stations and from upper-air ascents (for stations in the experiment domain) is presented in Section 3.1. Section 3.2 shows areal comparisons of relevant output from the experiments. More detailed analysis of the global SW radiation output from HARMONIE compared to measurements over Ireland is presented in Section 3.3. Finally a comparison of HARMONIE cloud cover and cloud condensate (cloud water and ice) versus MSGCPP satellite data is detailed in Section 3.4.

3.1 Verification versus synoptic station data and radiosondes

The results presented in this section are mainly temperature focused where available observations from synoptic stations and upper-air ascents over the experiment domain are compared to the output from each HARMONIE experiment. Figure 2(a) shows a clear increase of ~ 0.1 degrees in the negative bias in 2m temperature when the cloud inhomogeneity factor is removed (i.e. set to 1) with a further ~ 0.1 degree reduction using the Nielsen cloud liquid optical property scheme. INHOMEXP and COPEXP have more negative biases in specific humidity (Figure 2b) and precipitation (Figure 2c) and neutral effects on wind speeds, cloud and mean sea level pressure. Lower temperatures cause less evaporation, lower humidity and less precipitation. However, the statistical significance of these results has not been tested. Figure 3(a,b) shows the decreases in temperature at the 925hPa and 850hPa pressure levels as a function of forecast length, with the decreases more pronounced at the 850hPa level. This is also indicated in the temperature profiles shown in Figure 3(c) which shows the average temperature bias at each pressure level valid at 12 UTC.

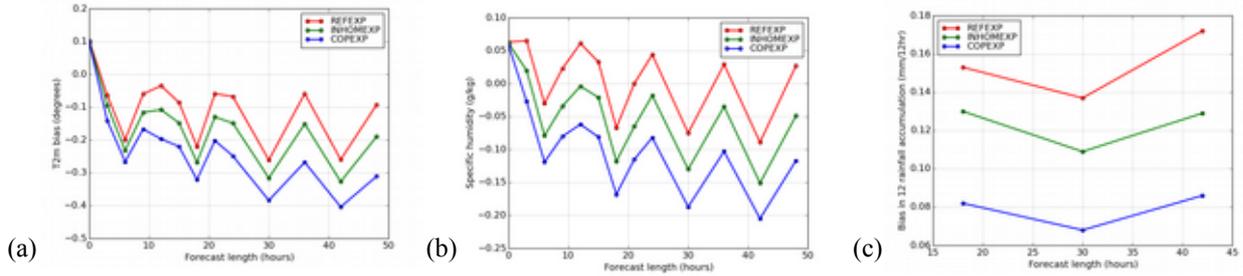


Figure 2. Bias in (a) 2m temperature, (b) specific humidity and (c) precipitation as a function of forecast length for the REFEXP, INHOMEXP and COPEXP experiments.

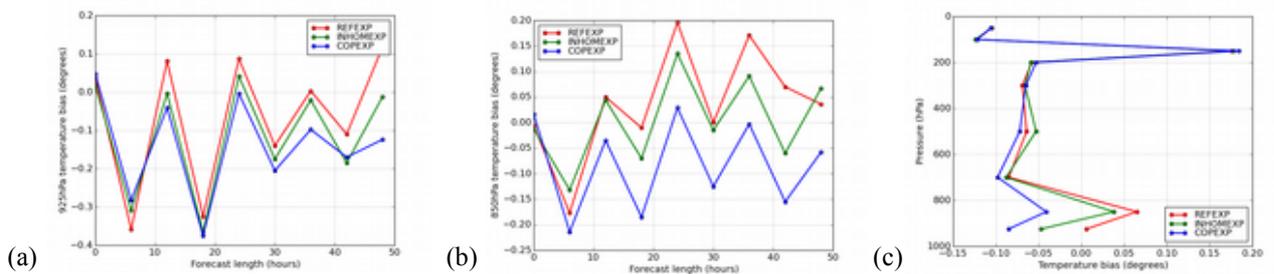


Figure 3. Bias in (a) 925 hPa and (b) 850 hPa temperature as a function of forecast length (c) bias in temperature at various pressure levels (valid at 12UTC) for the REFEXP, INHOMEXP and COPEXP experiments.

3.2 Areal comparisons of the INHOMEXP and COPEXP versus REFEXP temperature and energy fluxes

Mean monthly surface temperature over the domain for June 2013 is shown in Figure 4(a) for REFEXP, the default HARMONIE set-up, where hourly forecasts up to 24 hours from each of the 00UTC runs were used in the calculation. The INHOMEXP and COPEXP relative biases compared to REFEXP are shown in Figure 4(b) and 4(c). The biases are mostly negative and more pronounced in COPEXP which includes the effect of having no cloud inhomogeneity as well as the new cloud liquid optical property scheme. As expected, these biases are only over land as sea surface temperatures (SSTs) in the model are derived from ERA-Interim reanalysis data. The lateral boundary conditions are updated at 3-hour intervals, during which time the SSTs vary insignificantly compared to surface temperatures over land. Figure 5(a) shows the mean surface temperature over all land points in the domain for each experiment by time of day at 3-hour intervals, where again forecasts from the 00UTC runs were used in the calculation. Figure 5(b) is similar but shows the biases in surface temperature for INHOMEXP and COPEXP relative to REFEXP. Both figures clearly illustrate the diurnal dependence of the effect of the cloud inhomogeneity and liquid optical property scheme on surface temperatures.

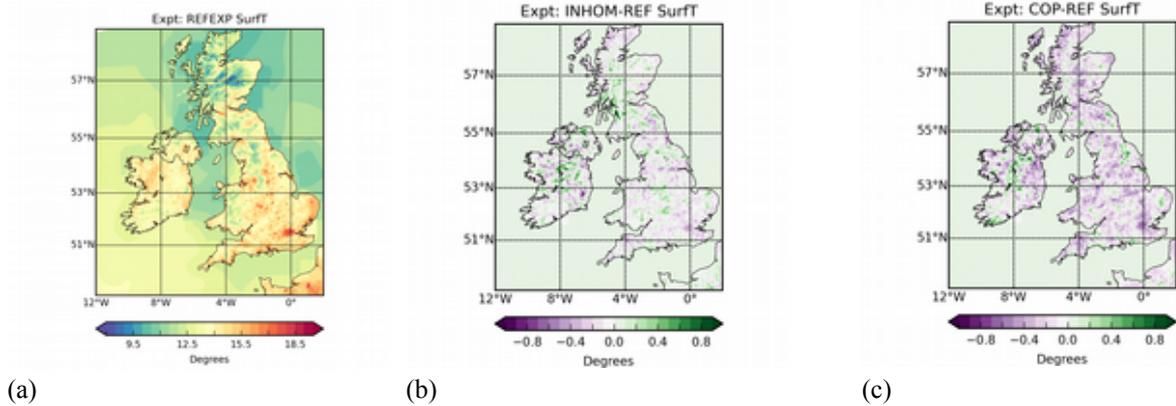


Figure 4. (a) REFEXP (b) INHOMEXP minus REFEXP and (c) COPEXP minus REFEXP monthly mean surface temperature. In each case the hourly forecasts (up to 24 hours) from the 00UTC cycles were used.

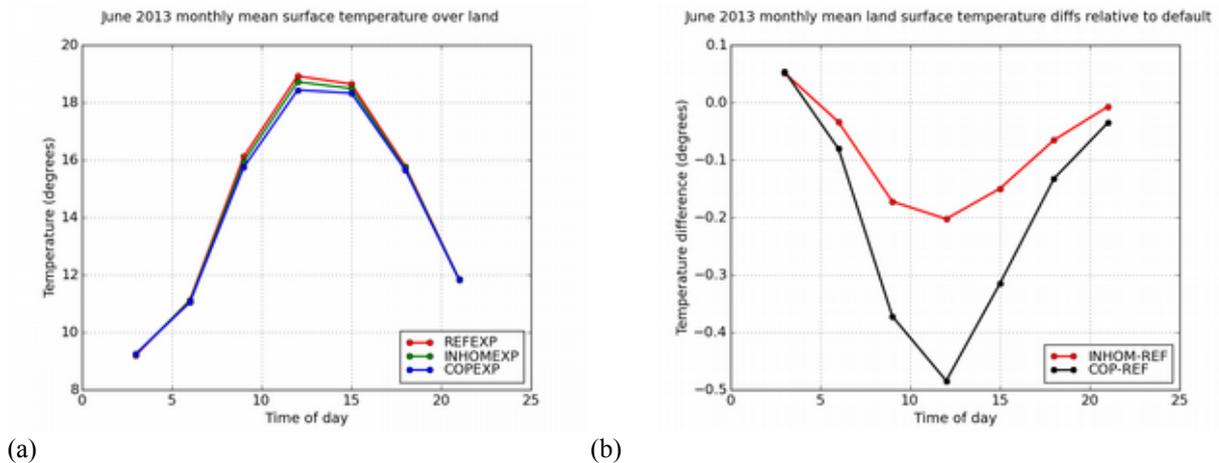


Figure 5. (a) Monthly mean surface temperature over all land grid points and at 3-hour intervals during the day, using forecasts from the 00UTC forecast cycles. (b) Similar to (a) but differences relative to the default set-up are plotted.

Most of these changes in temperature (surface, 2m, lower troposphere below 850hPa) can be explained by the decrease in the downwelling SW radiation flux (SWD) shown in Figure 6. The diurnal cycle of the surface temperature differences and also the differences in this cycle for INHOMEXP and COPEXP can also be attributed to the differences in SWD. The average decrease in SWD over land grid points is 11 W/m^2 for INHOMEXP and 22 W/m^2 for COPEXP. On the other hand, the changes in downwelling LW radiation at the surface (LWD, Figure 7) are an order of magnitude smaller ($< 1 \text{ W/m}^2$ average change over land grid points for both experiments compared to REFEXP).

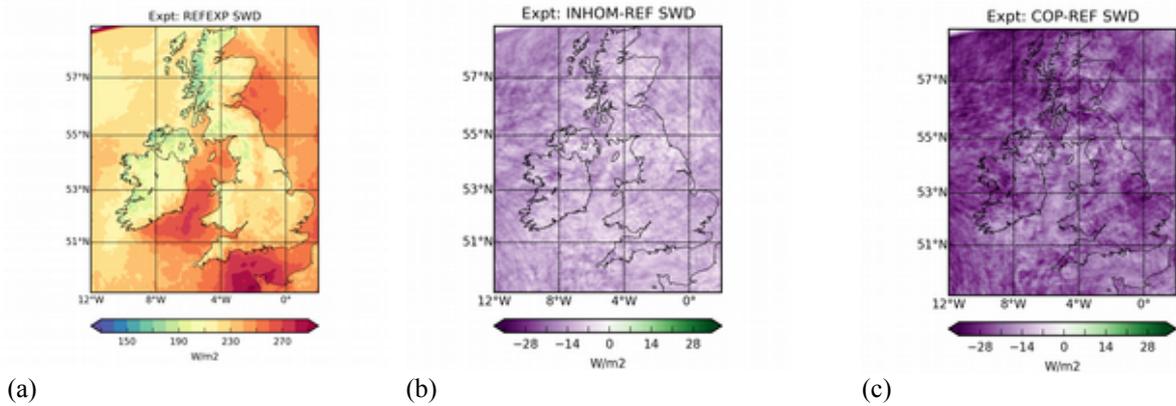


Figure 6. (a) REFEXP (b) INHOMEXP minus REFEXP and (c) COPEXP minus REFEXP monthly mean of daily mean global SW radiation. In each case the hourly forecasts (up to 24 hours) from the 00UTC cycles were used.

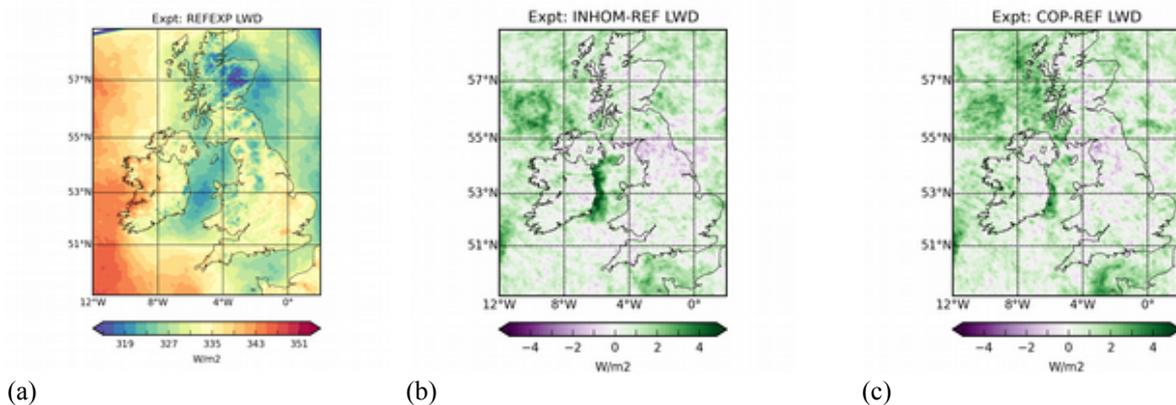


Figure 7. (a) REFEXP (b) INHOMEXP minus REFEXP and (c) COPEXP minus REFEXP monthly mean of daily mean downwelling LW radiation at the surface. In each case the hourly forecasts (up to 24 hours) from the 00UTC cycles were used.

3.3 Downwelling SW Radiation Verification and the Clear Sky Index (CSI)

In this section we focus on downwelling global SW radiation in more detail comparing hourly accumulations of SWD at 7 synoptic stations (Figure 1) in Ireland to HARMONIE output for the 3 experiments for June 2013. The clear sky index is a useful parameter for comparing SW radiation because it also acts as a proxy for cloud cover and cloud condensate amounts. The index is the ratio of global SWD divided by the maximum possible global SWD and is dependent on the location, date and time. In this case, the index was computed using observations from the 7 Irish stations and data from the 3 HARMONIE experiments, bi-linearly interpolated to the station locations. At each location, and for the observations and HARMONIE experiments separately, the CSI was computed using the average solar zenith angle (SZA) over the previous hour. The average SZA over the previous hour was used because the accumulations of SWD are available at hourly intervals in HARMONIE. For computational reasons it was not possible to output and store data at a time resolution of one minute. Hence, since hourly averages of SWD are used in the CSI calculations, the hourly average SZA for the same period was also used.

It is important to note that the clear sky irradiances from HARMONIE were not used; instead the maximum solar irradiances were computed separately using locations, dates and times. These were computed using the `hlsolar.F90` solar astronomy subroutine from HiRLAM and our newly tuned version of the Savijärvi et al., 1990 clear sky equation (Equation 2) for global SW radiation at the surface using a 2.5 g/cm^2 water vapour load. We tuned constants $C1$, $C2$ and $C3$ using the clear sky experiments discussed in Nielsen et al., 2014 where $C1$ and $C2$ are associated with integrated water vapour u and $C3$ is associated with backscattering from reflected beams. $C1$, $C2$ and $C3$ are 0.11, 0.25 and 0.07 respectively. S_0 is the solar irradiance at the top of atmosphere, which varies with the Sun-Earth distance, and h is the angular solar height.

$$S(\text{sfc}) = S_0 \sin(h) \left(1 - 0.024 \sin(h)^{-0.5} - C1 \left(\frac{u}{\sin(h)} \right)^{C2} - \left(\frac{0.28}{1 + 6.43 \sin(h)} - C3 \alpha \right) \right) \quad (2)$$

This approach was taken because only instantaneous clear sky SW fluxes at the surface are available in HARMONIE but accumulated fluxes are required. A water vapour load of 2.5 g/cm^2 was used as it is a typical value at midlatitudes. For low water vapour loads the clear sky irradiances will be approximately 10% higher, as was shown in Nielsen et al. (2014). A 10% difference does not affect the results strongly. For greater accuracy, modelled water vapour loads and the water vapour loads from MSGCPP should be used in the HARMONIE and observed CSI calculations respectively.

To aid analysis the data were then binned by CSI and hourly mean cosine of the SZA. Data from the 7 Irish stations were amalgamated to produce Figure 8, where SWD biases relative to observations are plotted. The CSI (observation data) and $\cos(\text{SZA})$ bins are each in steps of 0.1. $\cos(\text{SZA})$ is denoted negative when the solar azimuthal angle is negative (i.e. the Sun is in the eastern sky) to investigate whether the positioning of the Sun has an obvious effect on the biases. CSI values greater than 1 mainly occur at low SZAs and can also be caused by optically enhancing sun cloud geometries. The reduction in global SW radiation can be clearly seen when the inhomogeneity factor is increased from its default value of 0.7 to 1.0 (i.e. less transparent clouds) and further still when the Nielsen SW cloud liquid optical property scheme is used. These biases are explored in more detail in Figures 9-11. Subsets of the data in Figure 8 are plotted in Figure 9 where typical CSI ranges (0.4 to 0.5 and also 0.5 to 0.6) were selected. In each case, the decrease in SWD relative to REFEXP is greater for higher SZAs. This effect is more pronounced in the CSI=0.5 to 0.6 case.

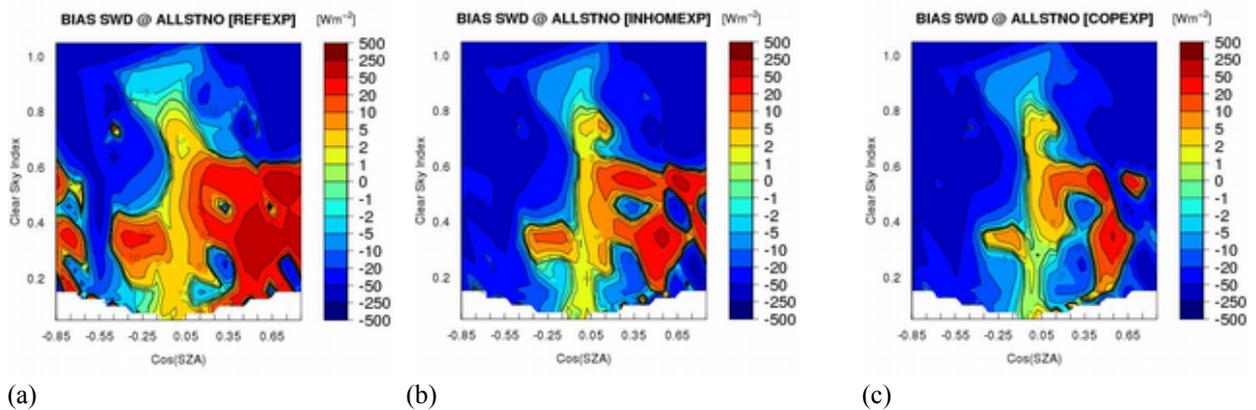


Figure 8. (a) REFEXP (b) INHOMEXP and (c) COPEXP mean bias in global SW radiation compared to measurements at 7 Irish stations where the data are binned by CSI and hourly mean $\cos(\text{SZA})$. $\cos(\text{SZA})$ is denoted negative when the solar azimuthal angle is negative (i.e. the sun is in the eastern sky). Only data where $N > 10$ (i.e. number of data points per bin > 10) are shown and CSI refers to the CSI of the observation data.

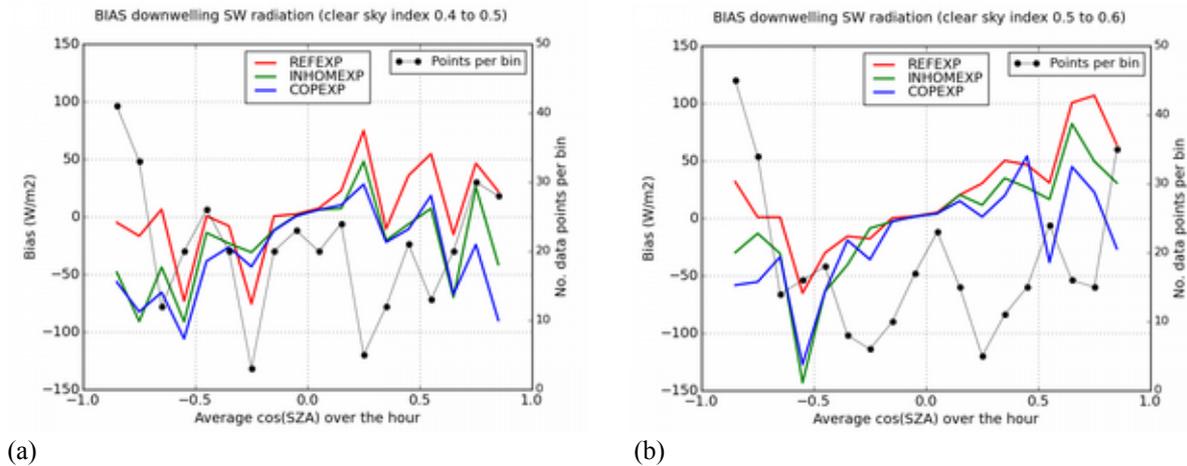


Figure 9. Mean bias in global SW radiation compared to measurements at 7 Irish stations where the data are binned by hourly mean $\cos(SZA)$ for (a) CSI between 0.4-0.5 and (b) CSI between 0.5-0.6.

In Figures 8 and 9 the CSI refers to the clear sky index calculated using the observation data. In Figure 10, the CSI for the observations and also the CSIs for each of the 3 radiation experiments are used to generate probability density functions of CSI. We can use these to draw conclusions about cloudiness in the HARMONIE model.

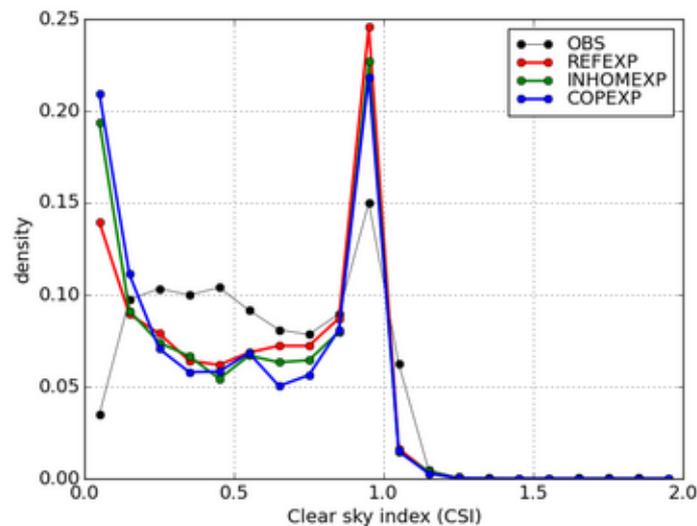


Figure 10. Probability density functions of CSI using hourly SW observations and forecast data for June 2013 valid at 7 Irish observation locations. Only data where $\cos(SZA) > 0$ are included. Observations (black), REFEXP (red), INHOMEXP (green), COPEXP (blue).

The binary distribution of the CSI (peaks at 0 and 1) gives complementary information to the binary cloud cover shown by Perez et al. (2014). In general the CSI is close to 1 under clear sky conditions. However, the CSI may also be high for cases with 100% cover when the clouds are thin. The main explanation for the results in Figure 10 is that the cloud water loads are too high in the HARMONIE model – this is discussed and illustrated in more detail in Section 3.4. We do not think that the binary cloud issue is due to 3D cloud cover or the overlap assumptions (see Section 3.4). We think that there is too much cloud water in the thickest clouds, and for thick clouds cloud overlap is not really an issue.

3.4 Cloud Water Path and Cloud Cover compared to MSGCPP Satellite Data

As described in the introduction the accuracy of SW radiation forecasts depends on a chain of parameters that include cloud cover, cloud water load, cloud liquid water and ice effective radii and cloud optical properties. In order to determine which of these explain the model issues shown in Figure 10, we performed separate tests of the modelled cloud cover and cloud water loads against MSGCPP SAF products from EUMETSAT. Here the cloud water load is the vertically integrated cloud liquid water and cloud ice.

The mean cloud cover for July 2013 for each of the HARMONIE experiments is over-predicted compared to the MSG cloud cover (not shown). However, the diagnostic cloud cover in HARMONIE is different to the radiative cloud cover as the former uses a random overlap algorithm as opposed to maximum overlap for clouds used in the radiation schemes. The diagnostic cloud cover is higher than the maximum overlap cloud cover (per. comm. Lisa Bengtsson, June 2015). Due to this, we do not know whether the cloud cover in HARMONIE is biased relative to MSG data. In the tests of cloud cover we do not see significant differences between the results of the three HARMONIE experiments.

The second important physical cloud parameter is cloud water load or path (CWP). The monthly mean CWP is shown in Figure 11 (a)-(d) for the three HARMONIE experiments and the MSGCPP satellite derived product. Only data between 07 and 17 UTC (06.45 to 16.45 for MSG) were used. In addition to this, only times where both MSGCPP and HARMONIE have full cloud cover were included, so that the result applies to cloudy cases. It is clear from Figure 11 that HARMONIE, regardless of cloud liquid optics and inhomogeneity, over-predicts cloud condensate, with Figure 11 (e)-(g) showing the HARMONIE biases relative to MSGCPP which are mostly in excess of 0.1g/m^2 . It is also clear that there are no significant differences in CWP between the three HARMONIE experiments.

Finally Figure 12 focuses on SW radiation biases relative to Irish station data for cases where the observed MSG cloud cover is 1 and the HARMONIE cloud cover exceeds 0.9 with these cloud cover criteria applied to each of the 7 stations separately. Figure 12(a) is a density plot of SWD biases for cloudy cases where data for each of the 7 stations were included to generate the figure; hence in this plot the data are not binned by CWP bias relative to MSG. The SWD biases are clearly skewed towards negative biases, consistent with positive biases in CWP. This is also illustrated in the Figure 12(b) scatter plot of the HARMONIE SW biases relative to station observations versus CWP biases relative to MSG data where again data for each of the 7 stations are amalgamated. The results for the REFEXP are shown here. The red curve with error bars shows the mean \pm the standard deviation of the CWP bias for SWD bias bins at 50 W/m^2 intervals and the cyan curve shows the percentage of data points in each of the SW bias bins. Most of the SW biases lie in the 0 to -50 W/m^2 range and within this sub-range the SW bias is correlated to the positive bias in CWP.

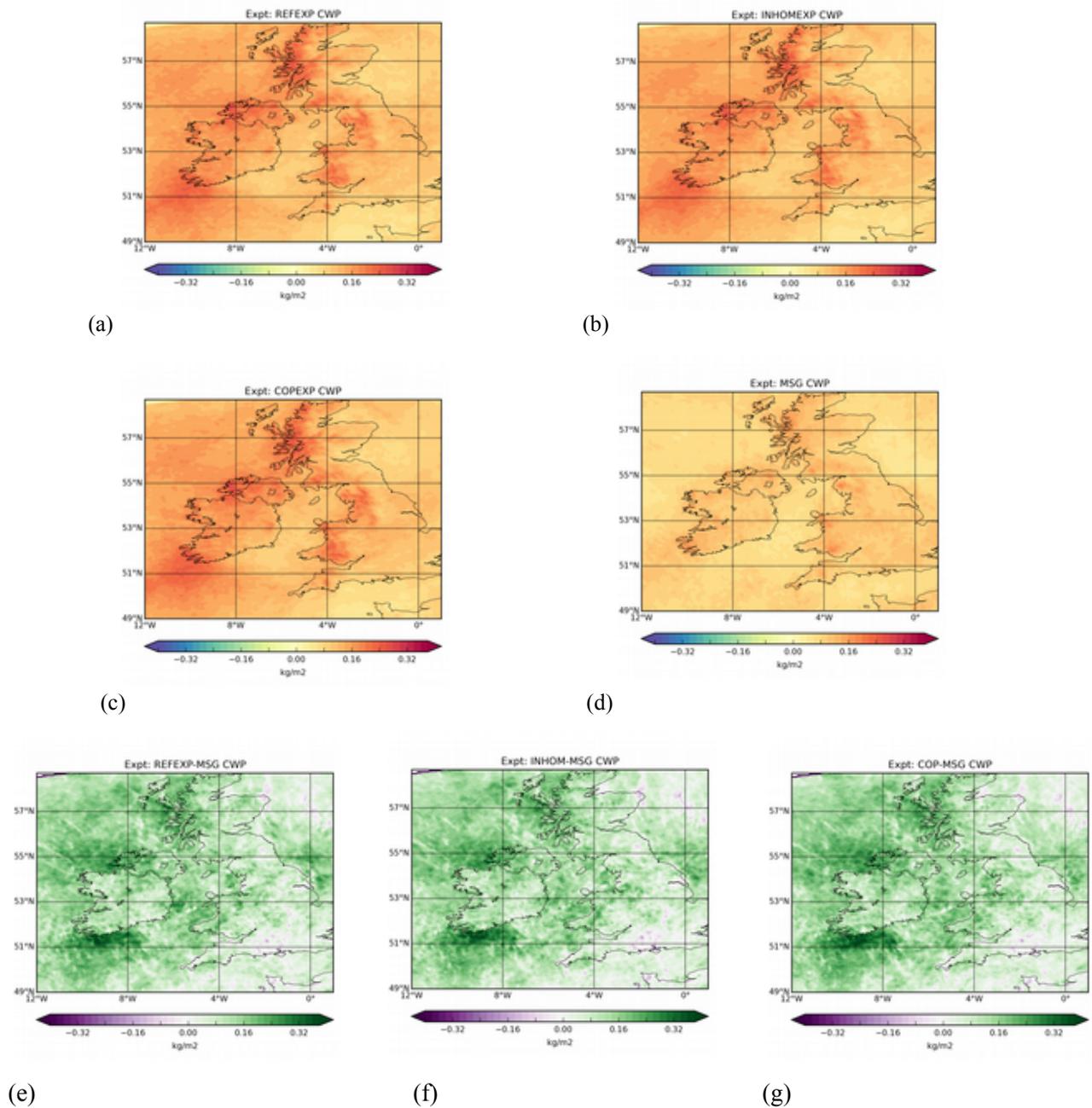


Figure 11. Monthly mean instantaneous integrated cloud condensate (water + ice) (kg/m^2) for (a) REFEXP (b) INHOMEXP (c) COPEXP and (d) MSG where data between 07 and 17 UTC were used (i.e. the times for which MSG data were available over the domain). (e), (f), (g) show the HARMONIE minus MSG biases for the REFEXP, INHOMEXP and COPEXP experiments where only grid points where both HARMONIE and MSG have full cloud cover are included.

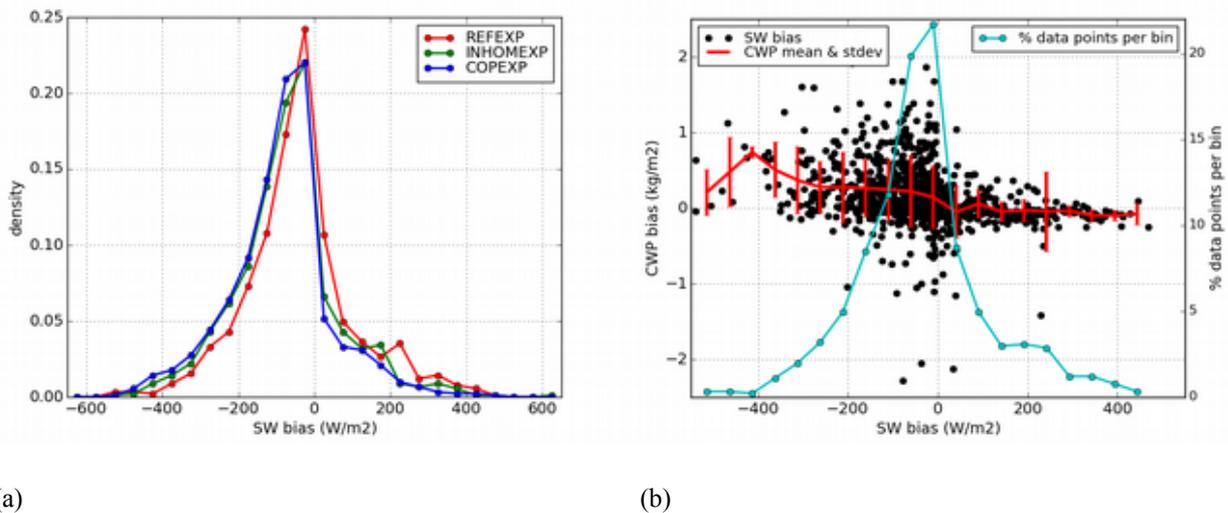


Figure 12. (a) SWD bias relative to observations for 7 stations in Ireland for REFEXP, INHOMEXP and COPEXP for cases where MSG has full cloud cover and the cover in HARMONIE is >0.9 . (b) Scatter plot of the bias in SWD relative to synoptic station data versus the bias in cloud water path relative to MSG data for the REFEXP. The red bar depicts the mean and standard deviation of the CWP bias for SW bias bins at 50W/m^2 intervals.

4 Conclusions and Next Steps

In this study we have used a novel method for testing clouds and radiation output from the HARMONIE NWP model by testing the cloud cover and cloud physical parameters independently. In addition our use of measured SW radiation to verify the modelled clouds is an improved method compared to traditional verification using synoptic surface observations, where only the cloud cover is verified. When verifying the model only by means of cloud cover, all correctly forecast overcast cases are considered true, despite the fact that they can have wrong cloud water loads or effective radii. This is why the verification using the clear sky index and cloud water loads gives complementary verification information that previously has often been ignored. Using this approach, our results indicate that cloud water loads in HARMONIE are too high.

Using the clear sky index as a proxy for cloudiness has also highlighted the binary (on/off) cloud cover in HARMONIE where the cloud cover tends to be 0/8 (zero octa) or 8/8 (8 octa); a similar result to that by Perez et al. (2014). Such behaviour is unlike observed cloud cover.

Our experiments were designed to test the influence of cloud inhomogeneity factor and the Nielsen cloud optical scheme on the NWP forecasts. Both have the effect of reducing surface, near surface and lower atmosphere temperatures, which results in a slightly larger negative bias over the Irish domain. In light of the verification of the cloud water path, the cloud inhomogeneity factor offset the effect of the positive bias in cloud water path by effectively reducing the cloud water and ice loads used by the radiation scheme by 30%.

Further work will be carried out using MUSC to comprehensively test the influence of cloud condensate on radiation output compared to observations. The modelled cloud effective radius data

could also be compared to the corresponding MSGCPP product. The results presented here demonstrate the need for a collaborative effort between the radiation and the cloud working groups to resolve some of the issues outlined.

5 References

Brousseau P., Berre L., Bouttier F., and Desroziers G. Background-error covariances for a convective-scale data-assimilation system: AROME-France 3D-Var. *Q.J.R. Meteorol. Soc.*, 137:409–422, 2011.

Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Bell, T. L., & Snider, J. B. (1994). The albedo of fractal stratocumulus clouds. *Journal of the Atmospheric Sciences*, 51(16), 2434-2455.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, 137: 553–597. doi:10.1002/qj.828

Ebert, E.E. and Curry, J.A.: A parameterization of ice cloud optical properties for climate models, *J. Geophys. Res.*, 97, 3834–3836, 1992.

ECMWF: IFS documentation – Cy38r1, Part IV: Physical processes, available at: <http://old.ecmwf.int/research/ifsdocs/CY38r1/IFSPart4.pdf>, 2012.

Fouquart, Y.: Radiative transfer in climate modeling, in: NATO Advanced Study Institute on Physically-Based Modeling and Simulation of Climate and Climatic Changes, edited by: Schlesinger, M. E., 223–283, 1987.

Fouquart, Y. and Bonnel, B.: Computations of solar heating of the earth’s atmosphere – a new parameterization, *Beitr. Phys. Atmos.*, 53, 35–62, 1980.

Fu, Q., Carlin, B., & Mace, G. (2000). Cirrus horizontal inhomogeneity and OLR bias. *Geophysical research letters*, 27(20), 3341-3344.

Hess, M., Köpke, P., and Schult, I.: Optical properties of aerosols and clouds. The software package OPAC, *B. Am. Meteorol. Soc.*, 79, 831–844, 1998.

Hogan, R.J., Personal communication regarding “The Effect of Cloud Structure on Radiation”, by Anna Mary Townsend, B.Sc. Dissertation, University of Reading.

Hogan, R. J., & Illingworth, A. J. (2003). Parameterizing ice cloud inhomogeneity and the overlap of inhomogeneities using cloud radar data. *Journal of the atmospheric sciences*, 60(5), 756-767.

Joseph, J.H., Wiscombe, W.J., and Weinman, J.A.: The Delta-Eddington approximation for radiative flux transfer, *J. Atmos. Sci.*, 33, 2452–2459, 1976.

Laprise, R., 1992: The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Wea. Rev.*, 120, 197–207.

Mascart, P.J. and Bougeault, P.: The Meso-NH Atmospheric Simulation System: Scientific Documentation, Tech. rep., Meteo France, Toulouse, France, 2011.

- Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations – description and examples of use, *Atmos. Chem. Phys.*, 5, 1855–1877, doi:10.5194/acp-5-1855-2005, 2005.
- Nielsen, K.P., Gleeson, E., and Rontu, L.: Radiation sensitivity tests of the HARMONIE 37h1 NWP model, *Geosci. Model Dev.*, 7, 1433-1449, doi:10.5194/gmd-7-1433-2014, 2014.
- Oreopoulos, L., & Cahalan, R. F. (2005). Cloud inhomogeneity from MODIS. *Journal of climate*, 18(23), 5110-5124.
- Daniel Martin Perez, Gema Morales Martin and Javier Calvo Sanchez, 2014: Statistical condensation scheme: Vertical variation of the saturation and convolution, ALADIN-HIRLAM Newsletter 3: 36.
- Pomroy, H. R., & Illingworth, A. J. (2000). Ice cloud inhomogeneity: Quantifying bias in emissivity from radar observations. *Geophysical research letters*, 27(14), 2101-2104.
- Roebeling, R.A., A.J. Feijt, and P. Stammes (2006). Cloud property retrievals for climate monitoring: implications of differences between SEVIRI on METEOSAT-8 and AVHRR on NOAA-17, *J. Geophys. Res.*, 111, 20210, doi:10.1029/2005JD006990.
- Savijärvi, H., 1990: Fast Radiation Parameterization Schemes for Mesoscale and Short-Range Forecast Models. *J. Appl. Meteor.*, 29, 437–447.
doi: [http://dx.doi.org/10.1175/1520-0450\(1990\)029<0437:FRPSFM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1990)029<0437:FRPSFM>2.0.CO;2)
- Scheirer, R., & Macke, A. (2003). Cloud inhomogeneity and broadband solar fluxes. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 108(D19).
- Seity Y., Brousseau P., Malardel S., Hello G., Benard P., Bouttier F., Lac C. and Masson V. The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, 139:976–991, 2011.
- Simmons, A., and D. M. Burridge, 1981: An energy and angular-momentum conserving finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, 109, 758–766.
- Slingo, A.: A GCM parameterization for the shortwave radiative properties of water clouds, *J. Atmos. Sci.*, 46, 1419–1427, 1989.
- Stammes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K.: Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, *Appl. Optics*, 27, 2502–2509, 1988.
- Stammes, K., Tsay, S.-C., and Laszlo, I.: DISORT, a GeneralPurpose Fortran Program for Discrete-Ordinate-Method Radiative Transfer in Scattering and Emitting Layered Media: Documentation and Methodology, Tech. rep., Stevens Institute of Technology, Hoboken, NJ, USA, 2000.
- Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob, and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J. Geophys. Res.*, 102, 23895-23915
- Tiedtke, M. (1996). An extension of of cloud-radiation parameterization in the ecmwf model: The representation of sub-grid scale variations of optical depth. *Mon. Wea. Rev.*, 124, 745-750.
- Toll, V., Reis, K., Ots, R., Kaasik, M., Männik, A., Prank, M., & Sofiev, M. (2015). SILAM and MACC reanalysis aerosol data used for simulating the aerosol direct radiative effect with the NWP

model HARMONIE for summer 2010 wildfire case in Russia. *Atmospheric Environment*. <http://dx.doi.org/10.1016/j.atmosenv.2015.06.007>

Toll, V., Gleeson, E., Nielsen, K.P., Männik, A., Mašek, J., Rontu, L., Post, P., Importance of the direct radiative effect of aerosols in numerical weather prediction for the European region, In preparation.

Wyser, K., Rontu, L., and Savijärvi, H.: Introducing the effective radius into a fast radiation scheme of a mesoscale model, *Contr. Atmos. Phys.*, 72, 205–218, 1999.

6 Acknowledgements

The authors would like to thank Noelle Gillespie, Chief Scientist, Valentia Observatory for providing shortwave radiation measurement data.

Study of urban climatology over Budapest with SURFEX/TEB model at the Hungarian Meteorological Service

Gabriella Zsebeházi, Ilona Krüzselyi, Gabriella Szépszó

1 Introduction

At the Hungarian Meteorological Service (HMS) two regional climate models (RCMs; ALADIN-Climate and REMO) provide climate change estimations for the 21st century on 10, 25, 50 km horizontal resolutions, particularly over the Carpathian Basin. RCM outputs serve input data for climate impact studies, nevertheless, users have constant demand to receive more detailed data than available from RCMs. In Hungary special attention is dedicated to investigate the effects of climate change in cities (especially of Budapest) considering their vulnerability to amplified temperature rise. However, RCMs alone are not appropriate for such aims due to their resolution and their simplified parameterizations for atmospheric interactions with artificial surfaces. One of the most state-of-the-art parameterization methods for describing the specific energy budget of urban areas is the Town Energy Balance (TEB) urban canopy model (Masson, 2000). TEB is included in the SURFEX (SURface EXternalisée; Le Moigne, 2009) externalised surface scheme, which describes surface-atmosphere interactions also over nature, lake and sea.

At HMS, SURFEX/TEB (hereafter SURFEX) is coupled offline to ALADIN-Climate (hereafter ALADIN; Csima and Horányi, 2008) for urban climate investigations. In the first steps, a detailed validation has been conducted over some Hungarian cities, in order to reveal the capabilities of our model set-up. Vértesi (2011) performed simulations over Budapest for a 10-year period (1961–1970) and compared the 2-m temperature and 10-m wind speed results with station measurements. The validation was limited on two observational stations operating in Budapest and its neighbourhood in this period: one in the city centre and the other one in the suburban area. Promising results were concluded with some shortcomings: SURFEX adds extra heat to the ALADIN fields principally over the city, catching the urban heat island phenomenon (especially in spring and autumn), but ALADIN results were mostly identified with too low temperature values, while SURFEX overestimated the temperature in comparison with measurements.

A possible reason for these outcomes supposed to be the land cover classification and physiographical information derived from the ECOCLIMAP database (Masson et al., 2003). First version of ECOCLIMAP was created in 2003 using existing land cover maps, climate maps, satellite data, and lookup tables. It refers to recent conditions, meaning that in Vértesi's study building density of 2000s was combined with climate conditions of the 1960s. However the built-up areas in the suburb of Budapest have increased notably since the 60s and its impact can largely influence the validation results. To scrutinise this problem, we performed a SURFEX integration on a later 10-year period, 1991–2000.

The objective of this paper is to present our findings focusing on the 2-m temperature. The paper is organised as follows: the overview of the experimental design are presented in Section 2, in Section 3 the temperature results and the added value of SURFEX to ALADIN are investigated. Finally a short summary, conclusions and future plans are given in Section 4.

2 Models and methods

The SURFEX land surface model describes surface-atmosphere interactions by simulating turbulent fluxes. In the model each grid box is divided into four tiles: nature (additional 12 patches can be distinguished), sea, inland water and urban surface. The surface energy and soil moisture budgets are computed separately for different tiles. The fluxes are then weighted with fraction of the tiles and aggregated over each grid cell. For urban surfaces the Town Energy Balance model is applied. In TEB complex city structures are simplified with canyon approach, i.e., streets are represented by roads with buildings on their two sides. The scheme computes energy budget separately for roof, wall and road, and moisture budget for roof and road using 3 or 4 layers in each surface. The anthropogenic heat and moisture fluxes derived from traffic, industry and domestic heating are also taken into account. In the model there is no interaction between the grid points, therefore, advection can be treated only in forcing fields.

SURFEX needs downward long- and shortwave radiation, surface pressure, precipitation, temperature, and horizontal wind components as inputs, and computes latent and sensible heat fluxes, upward radiative fluxes and momentum fluxes. Atmospheric forcings in our study were obtained from the ERA-40 re-analysis-driven (Uppala et al., 2005) simulation of ALADIN-Climate v5.2 regional climate model (developed at Météo France). The 10 km horizontal resolution RCM results on 30 m above surface were interpolated into a 1 km resolution domain over Budapest using the special (EE927) configuration of ALADIN (Fig. 1) that takes into account the topography as well. We applied the 5.1 version of SURFEX in stand-alone mode on 1 km resolution. Over natural surfaces the ISBA (Noilhan and Planton, 1989) scheme operated and the boundary layer was treated explicitly with the Surface Boundary Layer Scheme (also known as Canopy model; Hamdi and Masson, 2008). The surface coverage was derived from the 1 km resolution ECOCLIMAP database.

In the evaluation both spatial and temporal temperature characteristics were analysed. Quantitative validation was conducted for the same points as in the case of the earlier period (Vértesi, 2011; Fig. 2) due to the availability of data. However it turned out that these points are not the most appropriate for analysing the annual and daily cycle of urban heat island (UHI, see in Section 3). Hence we performed further investigations over two additional grid points considered as pure urban and rural points (Fig. 2).

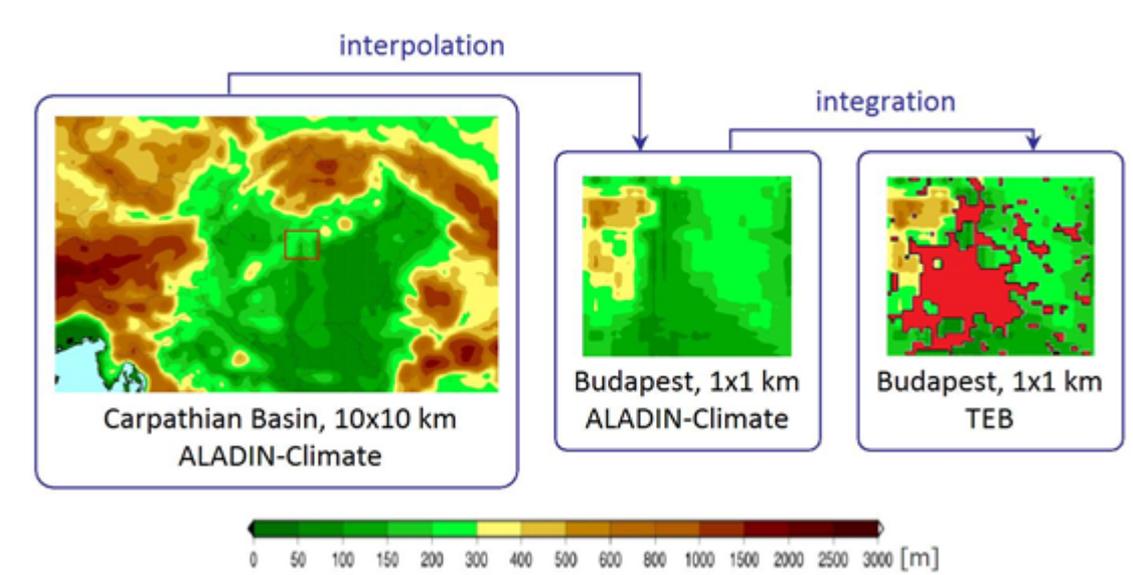


Figure 1: Flow chart of the use of SURFEX with topography [m] of the domains, and the grid cells that include urban surface (last panel; red).

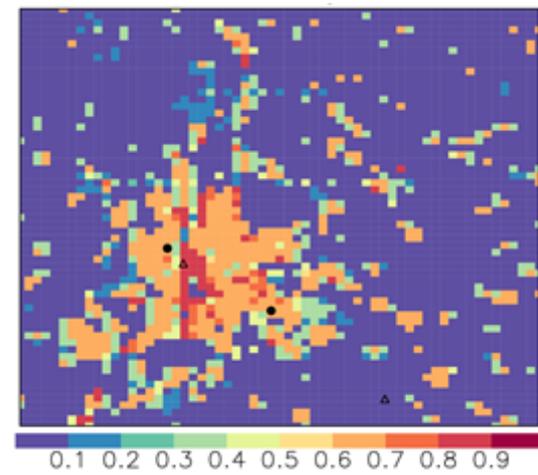


Figure 2: Fraction of urban land cover in grid cells according to ECOCLIMAP. Gridpoints closest to the observational stations are marked with solid circles (suburban station locates on the South-Eastern Budapest). Further selected urban and suburban gridpoints used in our study are marked with empty triangles.

3 Results

One of the most visible effects of energy budget differences between artificial and natural surfaces is the urban heat island phenomenon, i.e. the emergence of the urban areas (as well as smaller settlements) from their surroundings by their warmer air temperature, especially during night times. Its spatial structure and intensity show daily and annual variability. Figure 3 presents the maps of seasonal mean UHI intensity (reference is the average of four neighbouring grid points at the South-Eastern part of the domain) performed by SURFEX at a nocturnal (03 UTC) and diurnal time (12 UTC), when the highest and lowest UHI are detected, respectively. The simulated spatial characteristics of night time UHI coincide the general structure described by Oke (1987), namely the horizontal gradient of the relative temperature surplus is the largest near the border of the city (this area is called “cliff”), the inner parts appear as a “plateau”, and the city core peaks with the highest intensity.

The seasonally averaged daily cycles of UHI are analysed between a densely built-in urban point and a pure rural point (Fig. 4) which were selected based on their urban land cover fractions. Since measurements are not available in these two locations, we can only determine the general skill of SURFEX by comparing the model results against the findings of Probáld (1974), who investigated the climate characteristics of Budapest for 1965–1967 based on a pair of station data. The simulated UHI characteristics follow our expectations. Namely, maximum temperature difference occurs at night, before sunrise, and during the morning it quickly declines reaching a minimum at 12 UTC. In spring, summer and autumn it even turns into negative (i.e. the outer point is warmer than the city centre). This latter attribute is on the one hand due to the more efficient heat conduction and larger heat capacity of buildings compared to natural surfaces that lead to a delay in heat uptake and also release. On the other hand the low morning sun-rays cannot reach the bottom of the narrow canyons.

Focusing on the seasonal variability of the daily evolution of UHI intensity, our results shows that the strongest intensity occurs at spring night (3.5 °C in the core, with reference to a rural point, see Fig. 3) and not in summer, as Probáld (1974) presented. These discrepancies may be explained by the different time period and reference points. On the other hand it is well known that ALADIN heavily overestimates the precipitation in summer (Csima and Horányi, 2008), that may lead to a lower UHI

intensity. In winter the daily cycle of UHI is more flat (Fig. 4) and horizontal gradient of temperature difference remains lower as well.

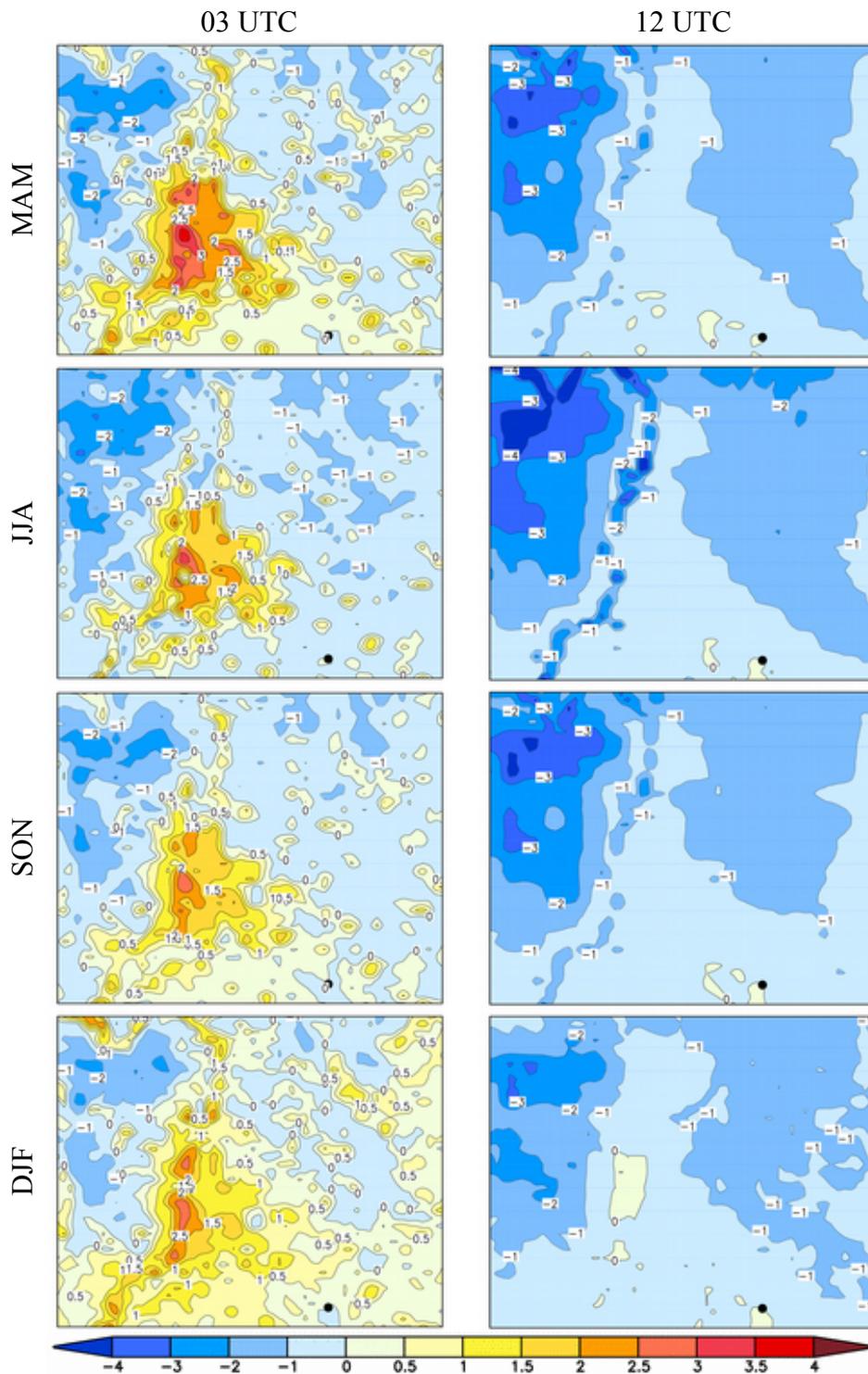


Figure 3: Seasonal mean urban heat island intensity (in °C; 2-m temperature difference in each gridpoint with reference to the average of four rural points indicated with a black dot) at 03 UTC and 12 UTC over Budapest according to SURFEX simulation in 1991–2000.

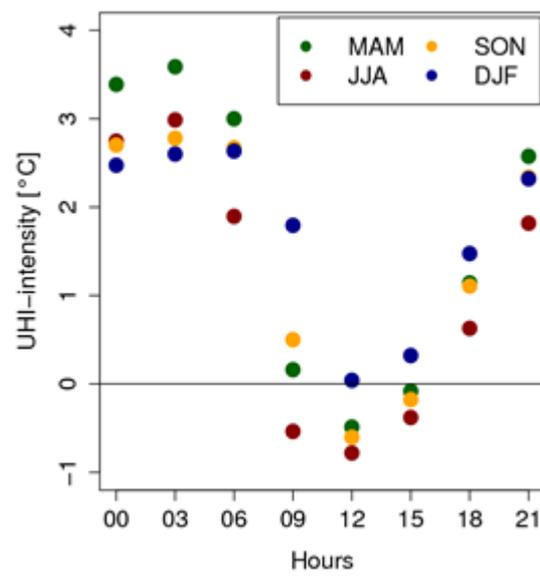


Figure 4: Daily cycle of the seasonal mean urban heat island intensity (in °C; 2-m temperature difference between the urban and rural points marked with triangles in Fig. 2) according to SURFEX simulation in 1991–2000.

Quantitatively validating the UHI results, the simulated monthly means of daily average temperature difference between gridpoints over Kitaibel Pál Street and Pestszentlőrinc is compared against station measurements (Fig. 5). Contrary to the measurements, positive urban heat island cannot be detected in SURFEX in any months. To find the reason of this spurious behaviour, the monthly mean temperature biases of SURFEX and ALADIN are also investigated in these two locations (Fig. 6). ALADIN is featured with negative bias throughout the whole year (but especially in January, April and October) over both points, and the underestimation is larger with 0.7–1.7 °C in the urban point (Kitaibel Pál Street) than in the rural point (Pestszentlőrinc). SURFEX adds extra heating to the ALADIN results in every month, however, this warming is almost the same in the two reference points meaning that SURFEX is unable to compensate the errors originated from ALADIN. All this lead to the conclusion that main error characteristics of SURFEX remained unchanged choosing the 1991–2000 period. Investigating the geographical location of the reference points, the inner site locates close to the Buda hills, on higher altitude (with roughly 60 m in the model) compared to Pestszentlőrinc, which can have a cooling effect on the forcings coming from ALADIN. Moreover in ECOCLIMAP the cover-types of both sites are defined as “temperate suburban” (i.e. the gridpoint cover is composed of 60% town and 40% nature), therefore, calculations of SURFEX led necessarily to same results in the two points.

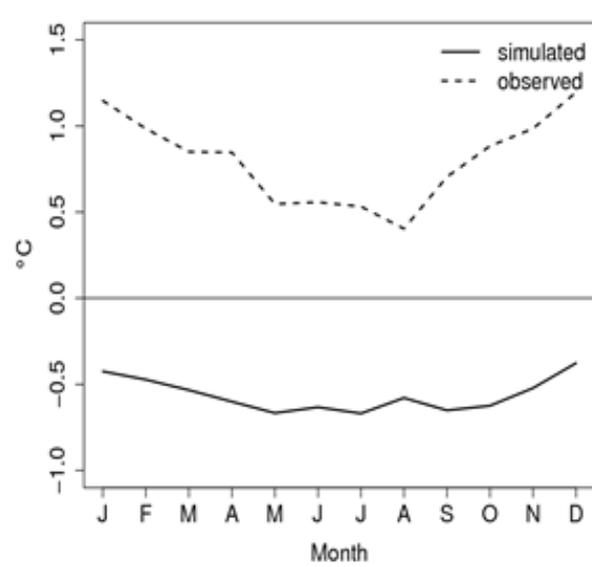


Figure 5: Monthly mean urban heat island intensity (in °C; 2-m temperature difference between Kitaibel Pál Street and Pestszentlőrinc) according to homogenised station measurements (dashed line) and SURFEX simulation (solid line) in 1991–2000.

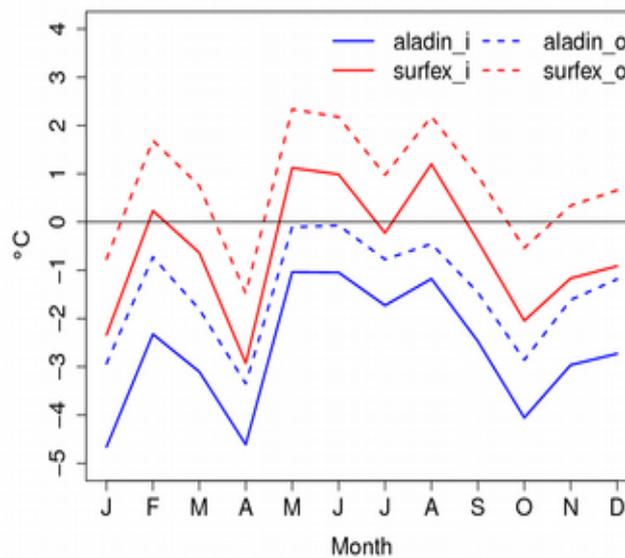


Figure 6: Monthly mean temperature difference (in °C) between the ALADIN (blue), SURFEX results (red) and observations in Kitaibel Pál Street (solid line) and Pestszentlőrinc (dashed line) in 1991–2000.

4 Summary and conclusions

At the Hungarian Meteorological Service dynamical urban climate research is conducted with the state-of-the-art SURFEX/TEB externalised surface model, whose Town Energy Balance (TEB) module describes the surface-atmosphere interactions over artificial surfaces in detail. Previously the model set-up was validated for 1961–1970 and it turned out that SURFEX cannot capture the UHI

feature that was detected based on two station measurements located in an urban and a suburban sites. In this paper our aim was to find an explanation for the abovementioned shortcoming. A new 10-year simulation was conducted for 1991–2000, because we supposed that in this case more realistic land surface information is provided to SURFEX, since ECOCLIMAP data was collected from around 2000. We applied SURFEX v5.1 coupled in stand-alone mode to ALADIN-Climate v5.2 over Budapest and investigated the 2-meter temperature results. It is found that SURFEX is able to detect the general spatial characteristics and diurnal and annual cycle of the urban heat island effect over Budapest. However, the choice of more recent period does not improve the UHI results over the two reference gridpoints where measurements are available. The reason of this behaviour is the inaccurate classification of land cover for these two reference points in the ECOCLIMAP database. Thus SURFEX could not eliminate the bias coming from ALADIN-Climate at the two sites. It is noted that the inner point locates in a complex area, where it may be exposed to the cooling effect of the Buda Hills through the RCM.

In the next step we intend to continue and deepen the validation. We will carry on some sensitivity analyses with urban surface parameters (e.g. albedo, building height) and model set-up in order to better understand the behaviour of SURFEX and to find its optimal settings. Our far goal is to implement future climate change scenarios for Hungarian cities on very high resolution that serves quantified information for impact studies and decision making.

5 References

- Csima, G. and Horányi, A.*, 2008: Validation of the ALADIN-Climate regional climate model at the Hungarian Meteorological Service. *Időjárás*, **112**, 155–177.
- Hamdi, R. and V. Masson*, 2008: Inclusion of a Drag Approach in the Town Energy Balance (TEB) Scheme: Offline 1D Evaluation in a Street Canyon. *J. Appl. Meteor. Climatol.*, **47**, 2627–2644.
- Le Moigne, P.*, 2009: *SURFEX Scientific Documentation*. Note de centre (CNRM/GMME), Météo-France, Toulouse, France. 211p.
- Masson, V.*, 2000: A Physically-based Scheme for the Urban Energy Budget in Atmospheric Models. *Bound.-Layer Meteor.*, **94**, 357–397.
- Masson V., J.-L. Champeaux, F. Chauvin, C. Meriguet and R. Lacaze*, 2003: A global database of land surface parameters at 1km resolution in meteorological and climate models. *J. Climate*, **16**, 1261–1282.
- Noilhan, J., and Planton, S., 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.
- Uppala, S.M., Kallberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.*, 2005: The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.*, **131**, 2961–3012.
- Vértési, Á. É.*, 2011: *Modelling possibilities of the urban heat island effect in Budapest* (in Hungarian). Master Thesis, ELTE, Budapest, Hungary. 78p.

ALARO-1 First Operational Application

Radmila Brožková

Introduction

The multi-scale physics, namely the moist deep convection parameterization, was a flagship of the ALARO-0 baseline. Indeed the 3MT scheme, together with other important ingredients such as prognostic microphysics handling jointly convective and resolved inputs, offer the possibility to exploit the model at wide range of horizontal resolutions. At the same time, a good methodology, adopted in the ALARO concept, starting with governing equations for moist physics, allows further scientific sophistications while keeping the seamless features of the ALARO-0. In this context, the ALARO-1 represents an enhancement of the ALARO-0, based on the need to improve radiative gaseous transmission functions and on the opportunities appearing in the parameterization of turbulence and convective drafts. The ALARO-1 thus represents a lot of scientific innovations and it aims to be used at all scales.

Owing to the strategy of the step-by-step improvements in NWP, we present here the ingredients of the first operational version of the ALARO-1, running at CHMI since 22 January 2015.

Radiation scheme ACRANEB2

Alike its predecessor ACRANEB in ALARO-0, the new scheme is a broadband one, having one solar and one thermal band. The choice to stay with the broadband approach is important for the strategy to keep the cloud-radiation interaction at every model time-step for a reasonable price, in contrast to k -distribution methods. Where beneficial, ACRANEB2 inherited the methodology from its predecessor, such as the delta-two stream formulation, idealized optical paths, and the use of the Net Exchange Rate method to treat the long wave radiation. Here below we mention important points of improvements without going to details. Full description of the new scheme will be published; the article on the solar part (Mašek et al. 2015) will appear soon.

Gaseous transmissions

Absorbing gases are water vapour, ozone and CO_2^+ , for which optical depths are fitted using recent spectroscopic database and laboratory measurements of ozone absorption cross-sections. Additional gases can be added at no extra cost for the scheme, except that the broadband fits would have to be recomputed. The important improvement is reached by taking into account the secondary saturation of gaseous absorption and also by applying the secondary corrective fits, which are temperature and pressure dependent. The effect of non-random gaseous spectral overlaps is parameterized for each pair of gases, where fits are done using a set of non-homogeneous optical paths.

Rayleigh scattering, cloud optical properties and intermittency

In contrast to other broadband schemes, a correction due to Rayleigh scattering is applied a posteriori, neglecting interactions with other processes, since these were accounted for via the delta-two stream and adding system. Cloud optical properties are based on the fits of the cloud mass extinction coefficient, the single scattering albedo and the asymmetry factor with respect to the effective radius of particles, determined for liquid and ice clouds. Recent fitting references are used. In the broadband case the cloud optical saturation must be evaluated with the multiple scattering included; here again the delta-two stream system makes the computation simpler. The non-random gas-cloud spectral overlaps are usually neglected in broadband schemes, however in ACRANE2 there is a parameterization of this effect, which may become important in case of the water vapour and cloud overlap in near-infrared wavelengths. Altogether, the accuracy of ACRANE2 compared to a narrow band references is very good for the NWP purposes, see a long wave example on Figure 1.

In general, the impact of cloud-radiation interaction should be accounted for as much precisely as possible. Given the rapidly evolving cloud scene, the best option is to compute this interaction at every time-step of the model, while an appropriate intermittency strategy can be adopted for slower evolving gaseous transmissions. Here the broadband scheme is well adapted for this goal. In ACRANE2, the intermittency is introduced in both short wave and long wave parts. While the short wave intermittency is a simple one with typically one hour interval to re-compute the transmission coefficients, there is a two-layer intermittency applied in the long wave band. Here one part of the Net Exchange Rate terms ought to be refreshed every hour about, while less important terms may be refreshed once every three hours about. This strategy significantly reduces the overall cost of the scheme. The cost/accuracy ratio is competitive compared to other state-of-the-art schemes used in NWP; however the every model time-step cloud-radiation interaction is the advantage on the ACRANE2 side. The above mentioned intermittency strategy is implemented in the first operational version of ALARO-1.

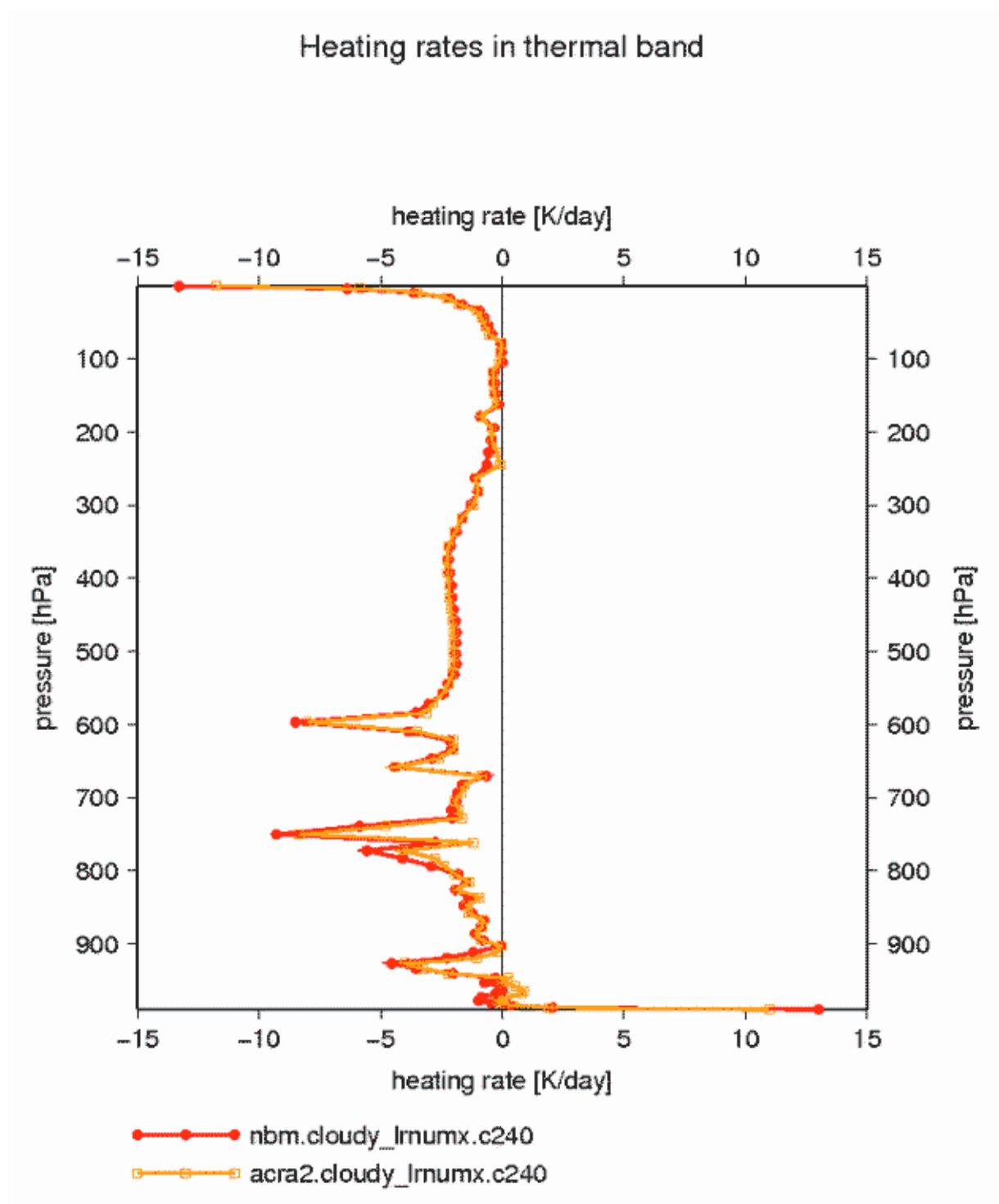


Figure 1: Thermal band heating rates for a non-isothermal cloudy atmosphere, with the maximum-random overlap assumption for clouds. Red – narrow band reference, yellow – ACRANE2.

TOUCANS – the first operational setup

TOUCANS (Third Order moments Unified Condensation Accounting and N-dependent Solver) is more the framework than a concrete scheme of turbulence. A key stone of this framework is the unified concept of stability functions valid for Richardson number going from $-\infty$ to $+\infty$, hence there is no critical Richardson number needed. Several models of turbulence can be emulated, like revisited Mellor-Yamada basic system, QNSE (Quasi Normal Scale Elimination) system and also EFB (Energy and Flux Budget) system. In addition, we may go progressively from the pseudo-prognostic TKE formulation (choice in the ALARO-0 baseline) to the fully prognostic TKE and TTE (moist Total Turbulence Energy) and also we can make choices of the mixing length parameterizations and ways to compute the shallow convection.

The choice of the first operational setup is a result of a progressive step-by-step validation of the TOUCANS parts. While all the above cited turbulence models yield reasonable results, we retained the so-called modified model II of the stability functions (revisited Mellor-Yamada Level 3 system, see Bašták-Žurán et al., 2014) as the best one for the future operational use. The first validation of the TOUCANS structure has been made by emulating the previous operational pseudo-prognostic TKE scheme, still using Louis stability functions. Then, we switched to the stability functions of the model II and new ingredients were added, such as the parameterization of the third order moments and turbulent diffusion of cloud condensates. While the first part of the shallow convection parameterization still relies on the Ri^* based algorithm, the influence of skewness in moist Brunt-Vaisalla Frequency computation is introduced, based on the fit of LES data, as illustrated on Figure 2. However, the most important improvement was achieved by the activation of the moist Total Turbulent Energy option.

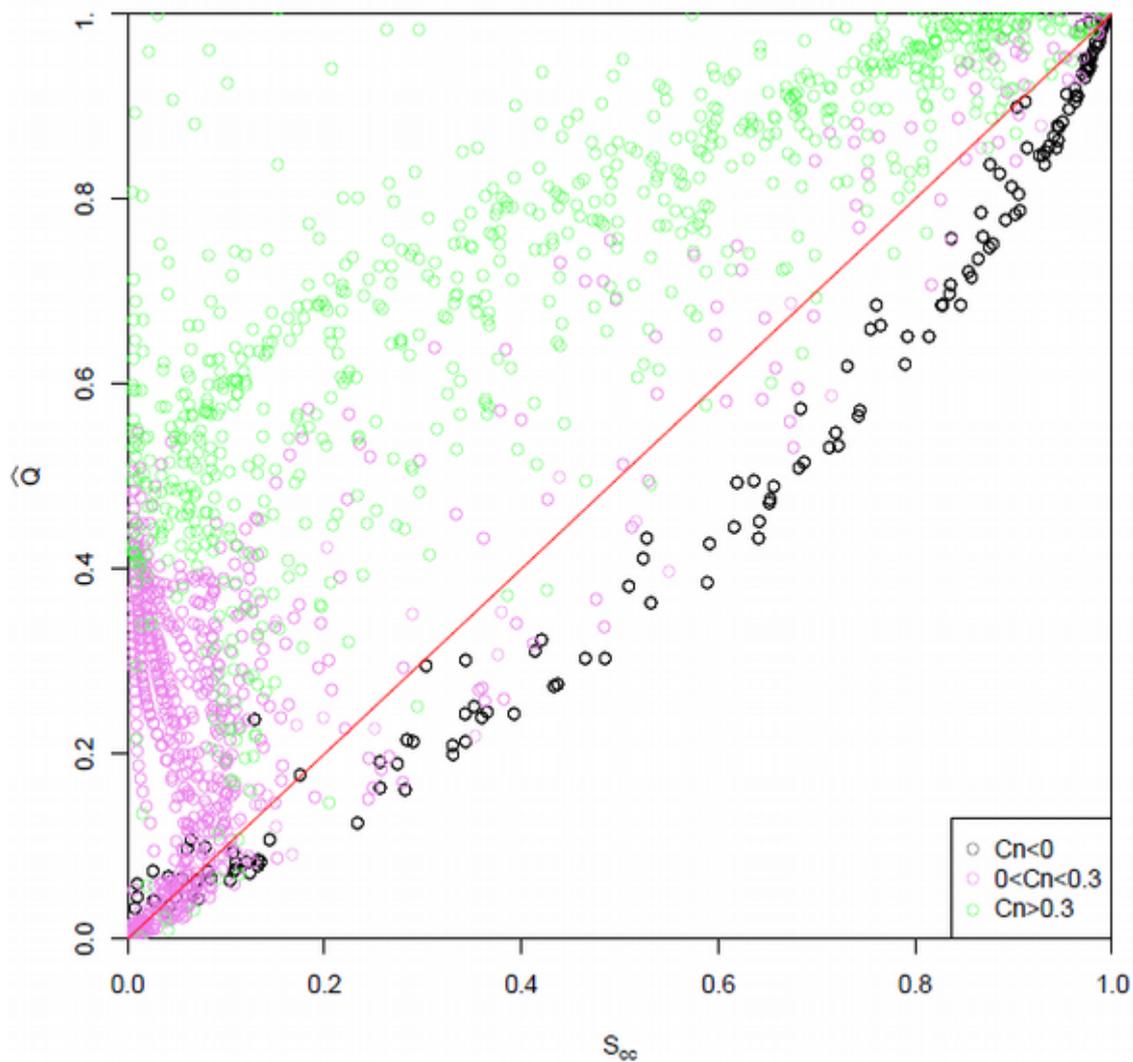


Figure 2: Relationship between a parameter \hat{Q} and Shallow convection cloudiness S_{cc} based on LES data, where \hat{Q} is equal to the skewness parameter C_n under the condition of neutrality.

Enhancements in Microphysics

Although these are improvements to the existing microphysics scheme, we count them already as being the part of ALARO-1, fitting into the ALARO-1 spirit of ‘putting in more physics’, and capitalizing on the structure given by ALARO-0.

The first modification concerns the geometry of cloud and rain, which is an important ingredient to handle jointly the convective (sub-grid) and resolved processes. Previously, two options were available for the vertical geometry – maximum-random or random overlaps, where the maximum-random option was used operationally. However, a simulation of real cloud scenes would correspond to a mixed situation (Shonk et al., 2010). Therefore we introduced a bit of randomness to the maximum-random overlap choice; the “randomness” coefficient is computed for each level, taking into account its thickness.

The second modification consists in the improvement of rain drop size distribution, following the work of Abel and Boutle (2012). According to observations, there are higher numbers of small rain drops than given by the Marshall-Palmer distribution. Modified distribution leads to more evaporation of smaller droplets and helps to diminish drizzle, which is often exaggerated by NWP models.

Results

New operational configuration based on the first version of ALARO-1 improves the scores for most of the parameters. For summer convective conditions it was tested over the period June-July 2009, while for the winter it was tested right in the parallel suite. The scores were computed with respect to radiosounds for the altitude fields and with respect to ground measurements for the screen level values. The impact in the upper air fields is more important in summer than in winter. It shows the improvement of the simulation of the ‘atmospheric machine’, as demonstrated on Figure 3.

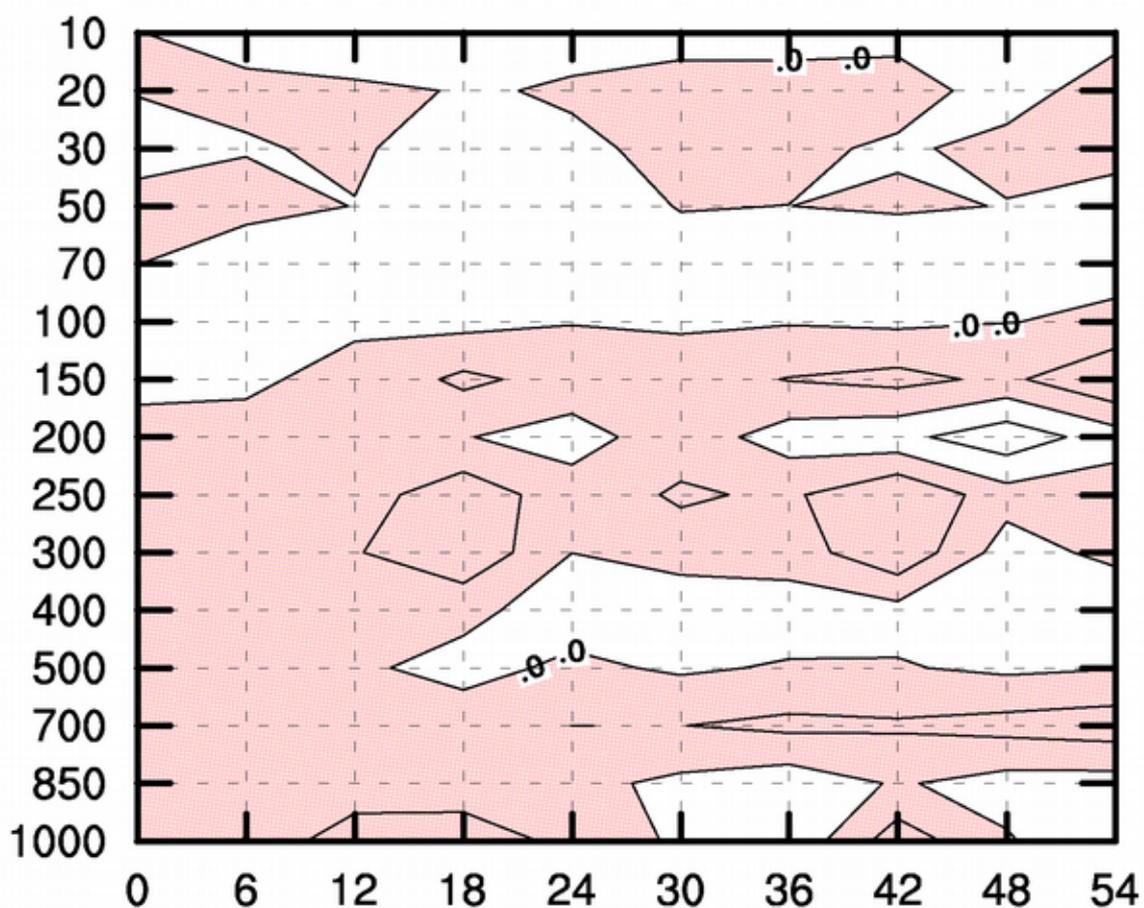


Figure 3a: Root Mean Square Error difference for temperature between ALARO-1 and the reference for June-July 2009. Red colour means the improvement, i.e. ALARO-1 has less error.

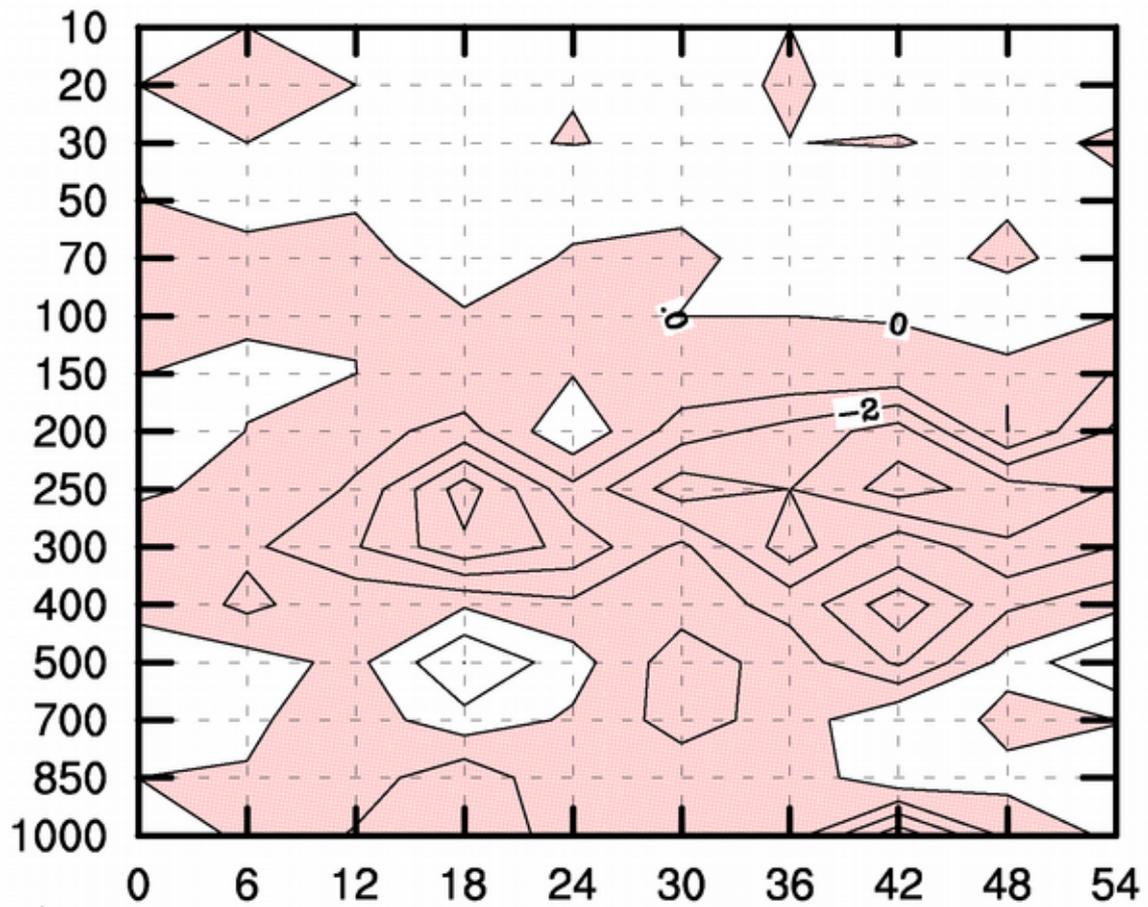


Figure 3b: Root Mean Square Error difference for relative humidity between ALARO-1 and the reference for June-July 2009. Red colour means the improvement, i.e. ALARO-1 has less error.

In winter the impact is concentrated into the boundary layer and screen level, as demonstrated on Figure 4.

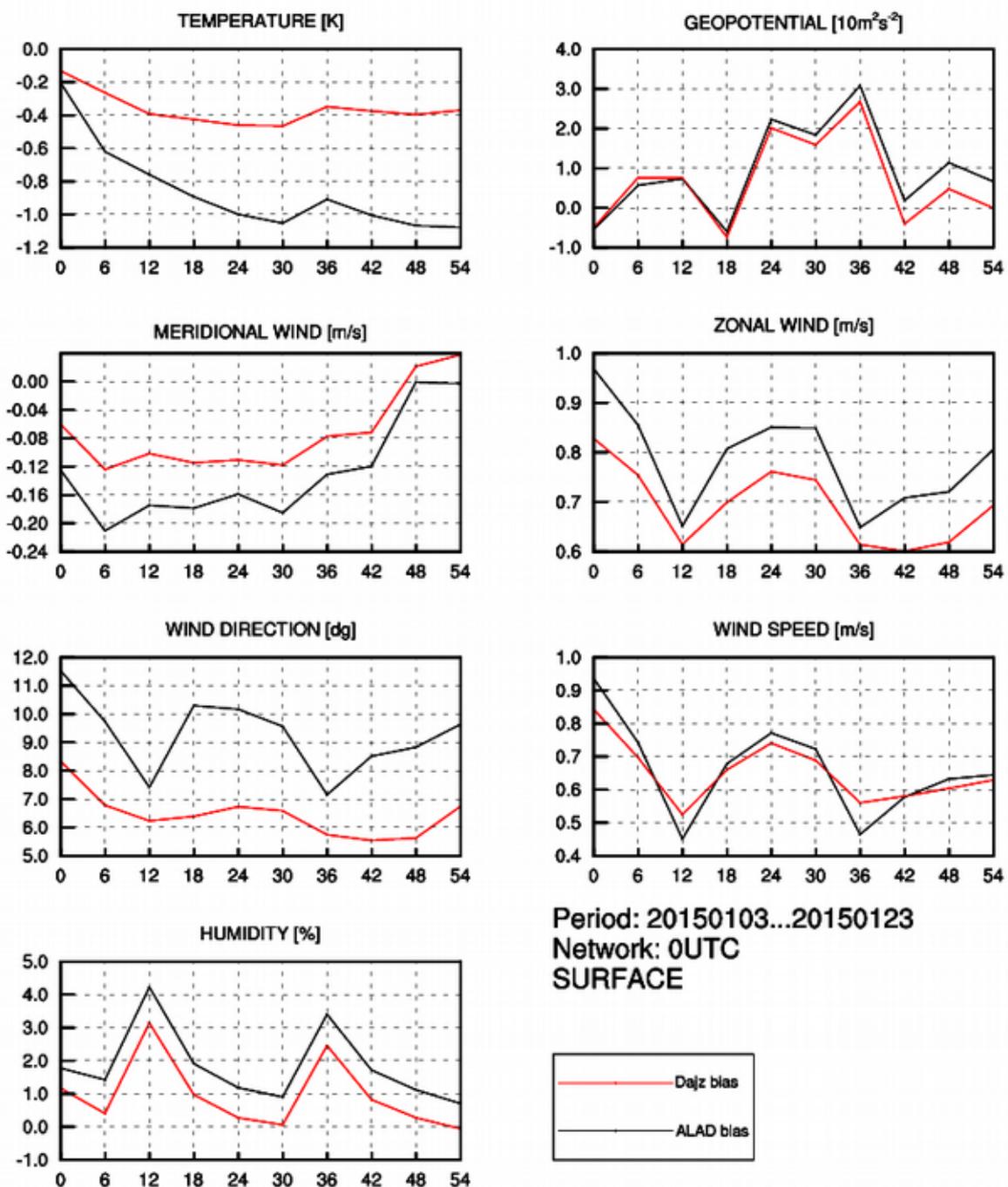


Figure 4: Bias of temperature, geopotential, wind and relative humidity at screen level for a winter period of January 2015. Black - operational reference (ALAD), red – parallel test (Dajz) of ALARO-1.

Conclusion

Although the first operational version of ALARO-1 does not yet fully profit from the scientific investment put to the development, namely on the side of the turbulence scheme, the model results are nicely improved. Return is also expected from the refinements of the parameterization of the convective drafts. The most important however is the multi-scale character of the ALARO-1, yielding seamless solutions across a wide range of horizontal resolutions, including the grey zone of moist deep convection, down to 1km.

References

Abel, S. J. and I. A. Boutle, 2012: An improved representation of the rain drop size distribution for single-moment microphysics schemes. *Q. J. R. Meteorol. Soc.*, **138**, 2151-2162.

Bašták-Đurán, I., J.-F. Geleyn and F. Váňa, 2014: A Compact Model for the Stability dependency of TKE production–destruction–conversion terms valid for the whole range of Richardson numbers. *J. Atmos. Sci.*, **71**, 3004–3026.

Mašek, J., J.-F. Geleyn, R. Brožková, O. Giot, H. O. Achom and P. Kuma, 2015: Single interval shortwave radiation scheme with parameterized optical saturation and spectral overlaps. Accepted in *Q. J. R. Meteorol. Soc.*

Shonk, J. K. P., R. J. Hogan, J. M. Edwards and G. G. Mace, 2010: Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part I: Review and parameterization. *Q. J. R. Meteorol. Soc.*, **136**, 1191-1204.

Deep convection and downdraught in Alaro-1

Luc Gerard

1 Non saturated downdraught

A precipitation-driven downdraught implies the following negative feedback loop: when the downdraught velocity increases, the adiabatic heating increases, while the water transfer, hence the evaporative cooling, is reduced; the downdraught becomes warmer and less saturated, the buoyancy decreases, reducing the velocity. Seen the need to transfer condensed water to evaporate, saturation would require a very low downdraught velocity, and most of the time the downdraught is unsaturated (Fig.1).

Our parameterisation includes a prognostic downdraught velocity, the estimation of the water transfer including effects of detrainment, melting and precipitation-area inhomogeneity. A prognostic downdraught mesh fraction is used.

We have shown that the downdraught tunings can help to tune the diurnal cycle, but Alaro-1 has the uncommon feature to produce a quite delayed cycle, especially for the descending slope. Further study of the diurnal cycle in Alaro-1 seems advisable.

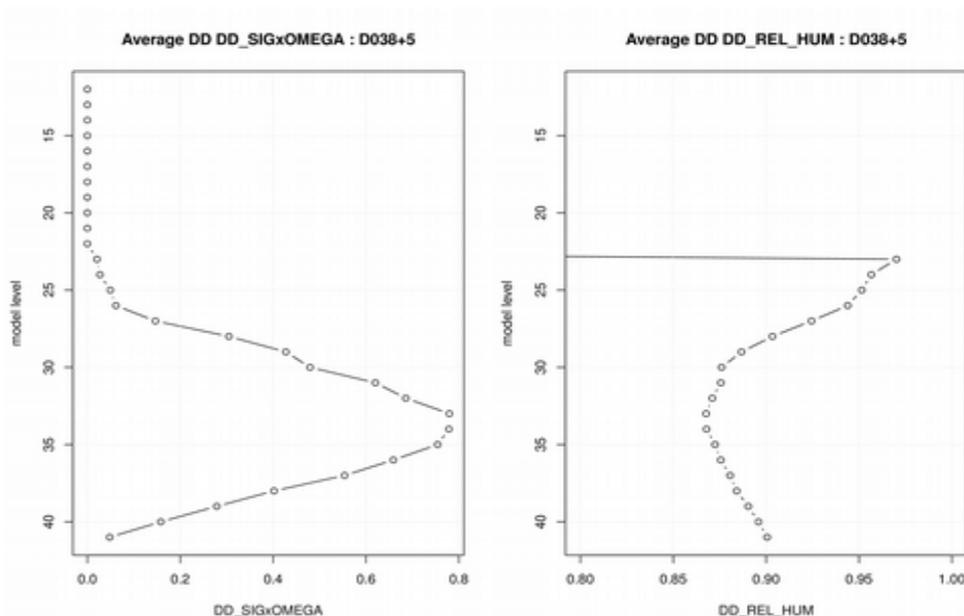


Figure 1: Mean profiles of downdraught mass flux (left) and relative humidity (right).

2 Complementary subgrid updraught

The complementary subgrid draught (CSD) scheme was developed to fix the weaknesses of 3MT at fine resolutions; it allows the subgrid signal to decrease gradually when the updraught mesh fraction increases, hence when the deep convection is increasingly resolved by the model grid (more details in Gerard, 2015).

The distinct behaviour of the CSD wrt the 3MT was illustrated by comparing the subgrid vs resolved precipitation shares on the Cold-Air Outbreak case of the WGNE Grey-Zone Model-Intercomparison Project (Fig.2).

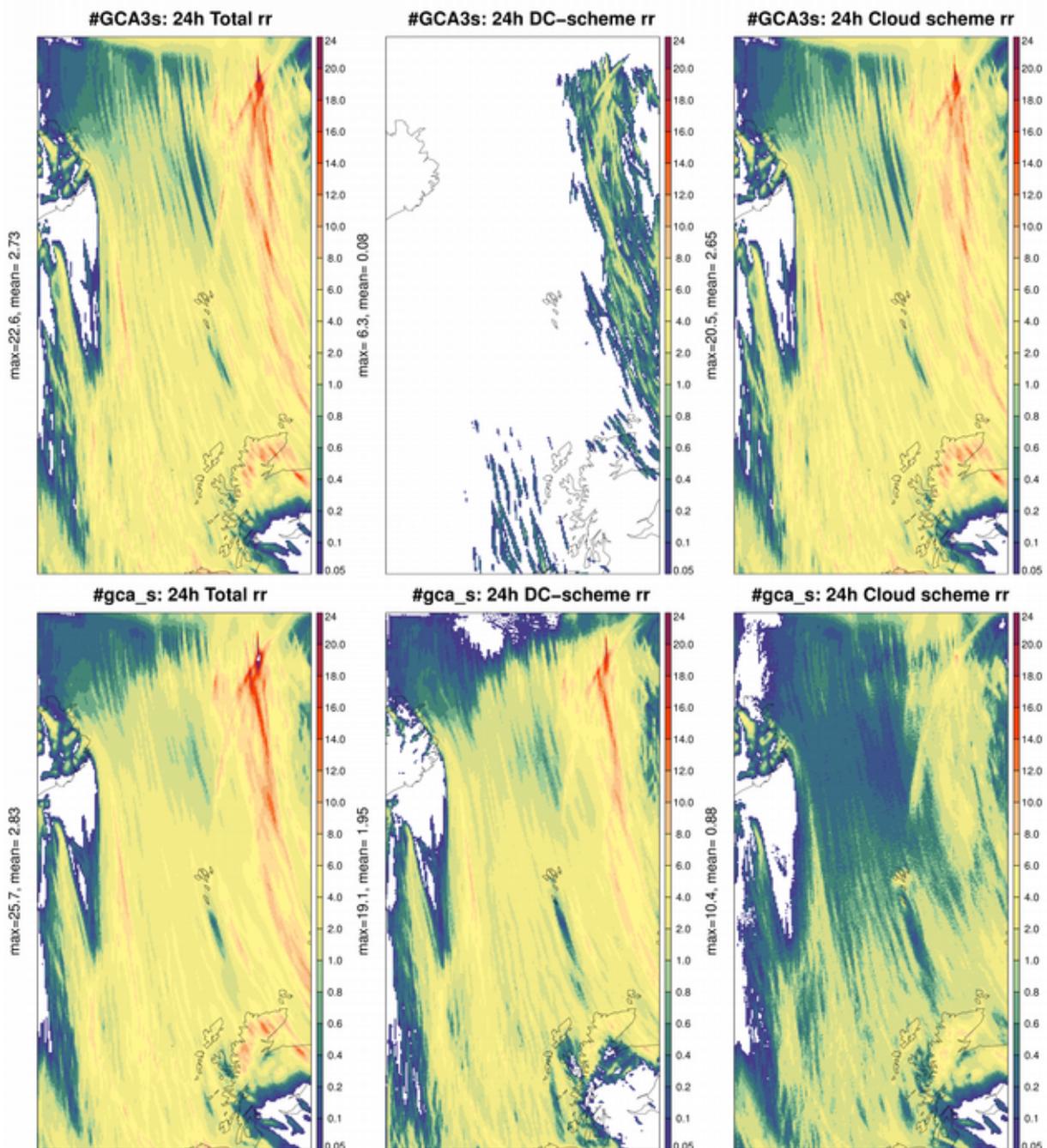


Figure 2:24-hours surface accumulated precipitation shares: total, subgrid convection scheme and cloud-scheme. Top: CSD, Bottom: 3MT, resolution 2km.

3 References

Gerard, L., 2015: Bulk mass-flux perturbation formulation for a unified approach of deep convection at high resolution. *Mon. Wea. Rev.*, in press.

On the necessity of modularity regarding the dynamical core

Daan Degrauwe, Steven Caluwaerts, Fabrice Voitus and Piet Termonia

1 Introduction

The current dynamical core of the ALADIN/ARPEGE/IFS models is semi-implicit semi-Lagrangian (SISL), with a spectral horizontal discretization. However, due to the fact that spectral methods are nonlocal (i.e. they require intensive communications between *all* gridpoints), the scalability of this spectral discretization on massively parallel machines is uncertain (Caluwaerts et al., In Press). Therefore, we should consider alternative, more localized, discretizations. While doing so, it is important to keep in mind that the accuracy of the spectral method is unsurpassable. This means that moving to a different discretization unavoidably means that some sacrifice will be made in the accuracy. Moreover, since the SISL scheme allows for large timesteps, our current dynamical core is also quite efficient (in terms of number of operations). It is an open question whether (or when in the future) the gain in scalability of a localized scheme will compensate the loss in accuracy and efficiency.

A new challenge in NWP is the energy-efficiency of running forecasts. In fact, this topic is considered an even more pressing one than the scalability of the model (Kallen, 2015). Especially in view of the development of new types of hardware accelerators (Xeon Phi, GPU, optical processors), it is important to consider the suitability of the dynamical core to run on such devices.

In this contribution, we do not intend to answer the question which dynamical core is the best for the foreseeable future. Instead, we will argue that there is no definitive answer to this question. Therefore, we should not put all of our eggs in one basket, and aim for modularity regarding the dynamical core. This means the development of a system where the most suitable alternative can be chosen in a specific situation.

2 The Z-grid scheme: the best of different worlds?

In this section, different localized grids are compared. More information on the methodology can be found in Caluwaerts et al. (In Press).

2.1 The test bed: SWE with SISL

It is easier to test alternatives for the spectral discretization on a manageable toy model than on the full atmospheric model. The shallow water equations (SWE) are often used for this purpose, since they support slow and fast waves that are representative for the atmospheric Rossby and gravity waves.

We consider the linearized 2D SWE that are given by:

$$\frac{Du}{Dt} = fv - \frac{\partial\phi}{\partial x} \quad (1)$$

$$\frac{Dv}{Dt} = -fu - \frac{\partial\phi}{\partial y} \quad (2)$$

$$\frac{D\phi}{Dt} = -\Phi \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \quad (3)$$

where u and v are velocity perturbations, ϕ is the geopotential perturbation, D/Dt is the total derivative, f is the Coriolis number and Φ is the background water geopotential.

For the time being, we will focus on the spatial discretization, and stick to the semi-implicit semi-Lagrangian time integration scheme that is used by the ALADIN/ARPEGE/IFS model family. Since advection is handled in a semi-Lagrangian fashion, it is not relevant at this moment for our tests. Therefore, we can simplify our set of equations to

$$\frac{\partial u}{\partial t} = fv - \frac{\partial\phi}{\partial x} \quad (4)$$

$$\frac{\partial v}{\partial t} = -fu - \frac{\partial\phi}{\partial y} \quad (5)$$

$$\frac{\partial\phi}{\partial t} = -\Phi \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \quad (6)$$

2.2 Discretizations under consideration

Three localized grids are considered: A-grid, C-grid and Z-grid (Mesinger and Arakawa, 1976). Figure 1 shows the location of the variables on these grids.

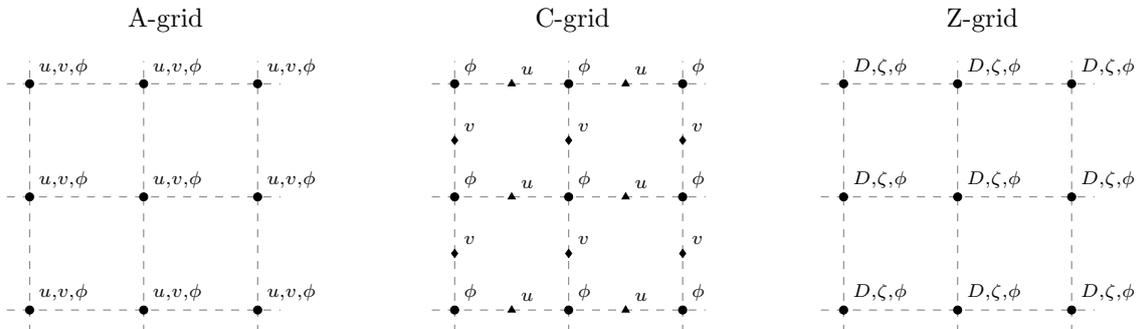


Figure 1: Different grids for the shallow water equations

The Z-grid uses the same gridpoints as the A-grid, but it reformulates the SWE in terms of vorticity and divergence, instead of in terms of velocities (Caluwaerts et al., 2015a):

$$\frac{\partial D}{\partial t} = f\zeta - \nabla^2 \phi \quad (7)$$

$$\frac{\partial \zeta}{\partial t} = -fD \quad (8)$$

$$\frac{\partial \phi}{\partial t} = -\Phi D \quad (9)$$

where divergence $D = \partial u/\partial x + \partial v/\partial y$ and vorticity $\zeta = \partial v/\partial x - \partial u/\partial y$.

We only consider second-order finite difference schemes on these grids. However, the qualitative results presented here can be shown to be also valid for higher-order discretizations.

2.3 Behaviour of SWE: dispersion relations

The behaviour a numerical scheme to solve the SWE is characterized by its dispersion relation, i.e. the way how waves (Rossby and gravity) are propagated by the scheme. Figure 2 shows the dispersion relation of the gravity waves on the different grids presented in the previous section.

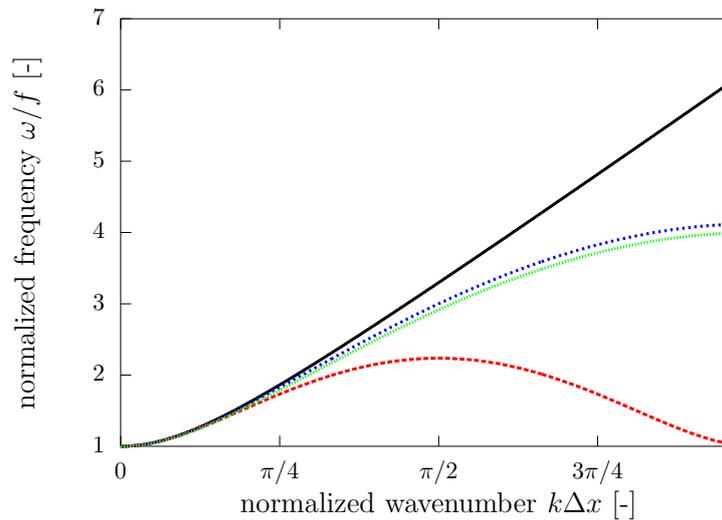


Figure 2: Dispersion relation for the gravity waves in the SWE on different grids: spectral (black), A-grid (red), C-grid (green) and Z-grid (blue). (from Caluwaerts et al. (In Press))

The most important conclusion from this figure is the following: the group velocity (which is given by the slope of the dispersion relation) for the shortest waves ($k\Delta x > \pi/2$) becomes *negative* for the A-grid discretization. This means that the A-grid scheme will propagate the energy of the shortest gravity waves in the opposite direction of where it is supposed to go. This is a long-known issue (Mesinger and Arakawa, 1976) which has stimulated the development of staggered grids such as the C-grid.

Although its dispersion relation is better, the C-grid also has disadvantages. Running a semi-Lagrangian advection scheme on a C-grid is roughly three times more expensive than on a non-staggered grid. Also, the staggering has consequences on the physics, on the data formats, etc. which makes the switch to a C-grid quite cumbersome.

The Z-grid scheme seems to offer the best of both worlds (Caluwaerts et al., 2015): it combines the organizational simplicity of the A-grid (all variables are defined at the same gridpoints) with the relatively good dispersion relation of the C-grid. It should be noted however, that this comes at the price of having to solve a Poisson equation to retrieve the velocities from the divergence and the vorticity.

3 Dispersion relations are not all that matter ...

3.1 Results of a standard adjustment test

We now will subject the different schemes to a standard adjustment test. The initial state is given by $u(x, t = 0) = 0$, $v(x, t = 0) = 0$ and

$$\phi(x, t = 0) = \phi_0 \text{sgn}(x - x_0). \quad (10)$$

This discontinuity in the geopotential obviously causes the initial state to be out of geostrophic balance. Gravity waves will form and propagate to restore this balance. Figure 3 shows the evolution of the geopotential on the different grids. The exact settings of the numerical parameters can be found in Caluwaerts et al. (In Press). This figure confirms the conclusions from the dispersion relation. The solutions with the Z-grid and C-grid are very close together and slightly decelerated w.r.t. the spectral solution. The solution with the A-grid is clearly inferior. The deceleration for the long waves is larger, while some small-scale noise indicates the faulty propagation of the shortest waves.

However, when investigating the other components of the solution, e.g. the velocity u as shown in figure 4, some defective behaviour of the Z-grid scheme becomes apparent: this field is very noisy. In fact, it turns out that this noise already appears after a single timestep. The other grids do not show this behaviour, which cannot be explained when only looking at the dispersion relation.

3.2 Impact of the eigenvectors of the amplification matrix

In order to explain the observed noisy u -field, we should consider the SWE from a wave-based perspective, instead of reasoning in terms of the variables u , v and ϕ .

To further simplify the analysis, we'll consider a linearized 1D SWE system without the Coriolis term:

$$\frac{\partial}{\partial t} \begin{pmatrix} u \\ \phi \end{pmatrix} = \begin{pmatrix} 0 & -\frac{\partial}{\partial x} \\ -\Phi \frac{\partial}{\partial x} & 0 \end{pmatrix} \begin{pmatrix} u \\ \phi \end{pmatrix}, \quad (11)$$

or, in matrix form:

$$\frac{\partial}{\partial t} \mathbf{X} = \mathbf{L} \mathbf{X}. \quad (12)$$

Now let's transform this equation, which is formulated in terms of the variables (u, ϕ) , into an equation in terms of (decoupled) waves. Let $\mathbf{\Lambda}$ be a diagonal matrix containing the eigenvalues of \mathbf{L} , and \mathbf{V} a matrix containing the respective eigenvectors, then we can define a transform

$$\mathbf{W} = \mathbf{V}^{-1} \mathbf{X}, \quad (13)$$

which reduces Eq. (12) to a decoupled system:

$$\frac{\partial}{\partial t} \mathbf{W} = \mathbf{\Lambda} \mathbf{W}. \quad (14)$$

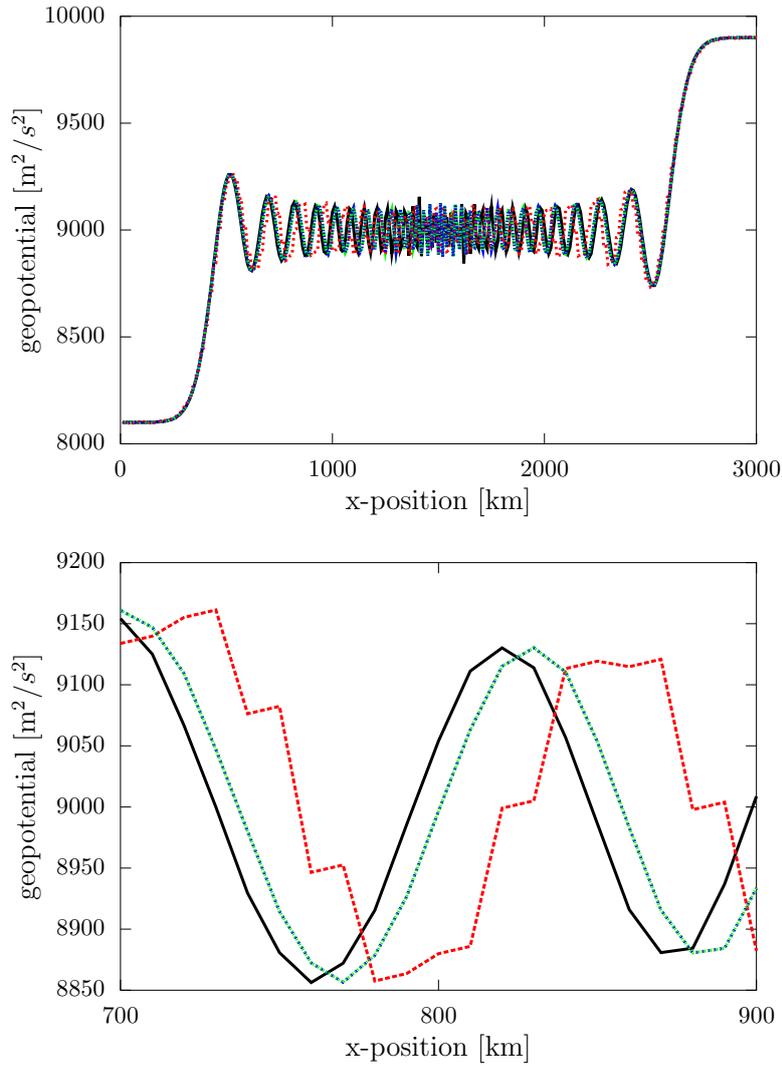


Figure 3: Result of the SWE adjustment test after 40 timesteps: geopotential $\Phi + \phi$. Spectral (black), A-grid (red), C-grid (green), Z-grid (blue). (from Caluwaerts et al. (In Press))

This shows how the propagation of the waves is determined by the *eigenvalues* of the matrix \mathbf{L} , while the relation (projection) of the physical variables on these waves is determined by the *eigenvectors*. For a time- and space-discretized equation, the amplification matrix takes the role of \mathbf{L} , but the fact remains that eigenvalues determine propagation, while eigenvectors determine projection.

All this is well-known, but how is it relevant to the noisy u -field that is obtained on a Z-grid? If we apply the eigenvalue decomposition on the matrix \mathbf{L} of the 1D SWE system (11), we get

$$\lambda = \pm ikc \qquad \mathbf{V} = \begin{pmatrix} 1 & 1 \\ c & -c \end{pmatrix}, \qquad (15)$$

where $c = \sqrt{\Phi}$ is the speed of the gravity waves.

Two things should be noted in this equation: (i) the eigenvectors (and hence the relation between u , ϕ and the wave amplitudes) are independent of the wavenumber k , and (ii) the wave amplitudes are mostly determined by the value of ϕ , since a typical value of the speed of gravity waves is $c \sim 100 \text{ m/s} \gg 1$.

Now let's move to the space-discretized schemes on the different grids. One can show that on an A-grid and on

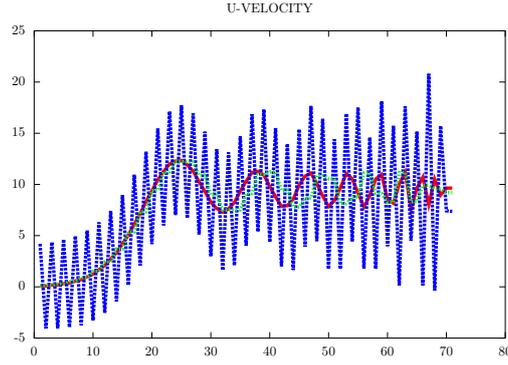


Figure 4: Result of the SWE adjustment test: velocity u after a 20 timesteps. Spectral (green), A-grid (red), Z-grid (blue).

a C-grid, the eigenvalues and eigenvectors become

$$\lambda^A = \pm \frac{ic}{\Delta x} \sin k\Delta x \quad \mathbf{V}^A = \begin{pmatrix} 1 & 1 \\ c & -c \end{pmatrix} \quad (16)$$

$$\lambda^C = \pm \frac{2ic}{\Delta x} \sin \frac{k\Delta x}{2} \quad \mathbf{V}^C = \begin{pmatrix} 1 & 1 \\ c & -c \end{pmatrix} \quad (17)$$

It is observed that on the A-grid, the eigenvalues go to zero for the shortest waves ($k \rightarrow \pi/\Delta x$), which explains the bad propagation of short waves on this grid. More importantly for the current analysis, the eigenvectors remain equal to the exact eigenvectors of eq. (15), hence independent of the wavenumber k . This means that although the propagation of the waves is modified by the spatial discretization, the decomposition of the initial state into the decoupled waves is the same as in the exact case.

Doing the same calculation for the Z-grid scheme, however, the eigenvalues and eigenvectors become

$$\lambda^Z = \pm \frac{2ic}{\Delta x} \sin(k\Delta x/2) \quad \mathbf{V}^Z = \begin{pmatrix} \sqrt{2 - 2 \cos(k\Delta x)} & \sqrt{2 - 2 \cos(k\Delta x)} \\ c \sin(k\Delta x) & -c \sin(k\Delta x) \end{pmatrix} \quad (18)$$

This reveals a fundamental difference between the Z-grid scheme and the other schemes: the eigenvectors become dependent on the wavenumber. For short waves, the decoupling of the initial state into the two waves will become dominated by u instead of by ϕ .

Figure 5 illustrates this effect. We consider an initial state with $u = 0$ and $\phi \neq 0$ (as was the case for the geostrophic adjustment test). In the exact case, as well as for the A-grid and C-grid schemes, this state is decoupled into two waves w_1 and w_2 , both of which have very limited u -components. For the Z-grid scheme, however, the orientation of the eigenvectors changes, and gets closer to the u -axis for short waves. As such, the initial state is decoupled into two waves, each of which has a substantial u -component.

Figure 6 shows what this means in physical space. For all grids, the initial state $u = 0$ is decomposed into two waves which are in exact counterphase so they cancel each other out at $t = 0$. The difference between the Z-grid and the other grids is that the amplitude of these waves is much larger (tending to infinity as $k \rightarrow \pi/\Delta x$). The two waves propagate in opposite direction, so after a single timestep, they are no longer in counterphase. For the A-grid and the C-grid, this will generate some small-scale fluctuations in u that are not even unphysical. For the Z-grid, however, this will result in very large short-wave oscillations in u .

To summarize, this test shows that the dispersion relation is not all that matters for appropriate behaviour of a scheme. It turns out that the eigenvectors of the amplification matrix may also play a role.

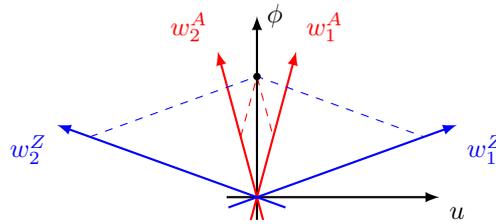


Figure 5: Projection of an initial state on two decoupled waves for the A-grid and for the Z-grid.

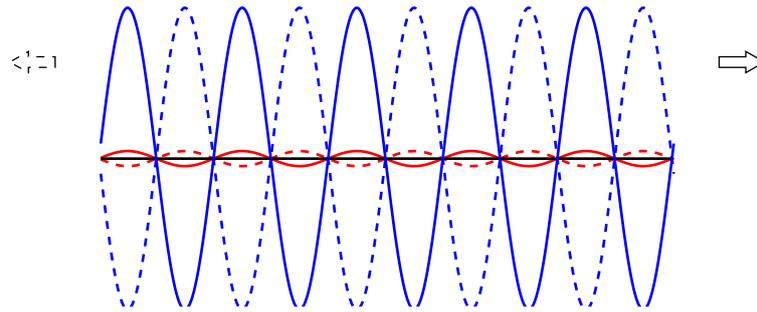


Figure 6: Decoupling of a zero initial state into two waves for the A-grid (red) and for the Z-grid (blue). Solid lines are waves travelling in the positive x -direction, dashed lines are waves travelling in the opposite direction.

4 Conclusions

The aim of this contribution is not to present new science or an elaborative overview of existing methods. The attentive reader has noticed that many simplifications were made in the presented analysis. We started from an already simple (though representative) system: the shallow water equations. Then this system was further simplified by linearization, and reduction to 1D. The discussion of spatial discretizations was limited to second order finite differences on three different grids. For the time discretization, only the semi-implicit semi-Lagrangian approach was considered.

Notwithstanding these huge simplifications, it is still impossible to draw any firm conclusions about which grid is the most suitable for localized methods. Each grid has its specific advantages and disadvantages; even the Z-grid, which for some time was considered to combine the benefits of the other grids (Caluwaerts et al., In Press), turns out to be defective for some specific tests. It is an open question how important the propagation of the shortest waves actually is in a 3D model, where physics can trigger these short waves, while horizontal diffusion can filter them.

The quality of the different schemes was only evaluated with a rather unphysical adjustment test. Other aspects of a scheme, like accuracy, scalability, or energy efficiency were not considered. Let alone the suitability of the schemes on exotic hardware that may be developed in the future. Given that there is no firm conclusion about the best grid for the very simplest model, means that we face an impossible choice about the most suitable grid for NWP. The only conclusion from the presented tests can be that we should try to be prepared for anything. This means that the dynamical core should be designed in a modular way, with different alternatives existing next to each other. Only such a system will allow to pick the most suitable scheme in a specific situation.

References

Caluwaerts S., Degrauwe D., Termonia P., Voitus F., Bénard P. and Geleyn J.-F., Importance of temporal symmetry in spatial discretization for geostrophic adjustment in semi-implicit Z-grid schemes, *Q. J. R. Meteorol. Soc.* 141, 128–138, 2015

Caluwaerts S., Degrauwe D., Voitus F. and Termonia P., Discretization in Numerical Weather Prediction: a modular approach to investigate spectral and local SISL methods. Springer Special Issue: Mathematical Problems in Meteorological Modelling, In Press.

Kallen, E., Research developments at ECMWF of relevance to HARMONIE. Presentation at the joint 25th ALADIN Wk & HIRLAM ASM, Denmark, 2015.

Mesinger F. and Arakawa A. Numerical methods used in atmospheric models. GARP Pub. Series No. 17 1. World Meteorological Organisation, Geneva, 1976.

Testing accuracy of finite element scheme used in vertical discretization of ALADIN-NH

Petra Smolíková, Jozef Vivoda and Juan Simarro

1 Introduction

The topic of the design of vertical finite elements scheme for ALADIN-NH is being solved since 2006 mainly in the frame of RC LACE. The main objective of this task is to have a stable and robust vertical finite elements (VFE) discretization to be used in high resolution real simulations with orography with the expected benefit being the enhanced accuracy for the same vertical resolution when comparing with vertical finite differences (VFD) method. We want to stick as much as possible to the existing choices in the design of dynamical kernel (SI time scheme, mass based vertical coordinate) and to stay close to the design of VFE in hydrostatic model version (according to Untch and Hortal, 2004).

As the most important task for the year 2014 it was identified the need to find the benefits of the currently implemented version of VFE in NH model (cy40t1). In Section 1 we discuss the accuracy of vertical operators used, in Section 2 we describe the results of idealized test with known analytical solution and we summarize results obtained in high resolution 3d simulations with ALADIN-NH.

2 Accuracy of vertical operators

The accuracy of all vertical operators has been studied carefully. It was shown by approximation of cubic splines by Taylor series expansion in terms of Fourier components (by Staniforth and Wood, 2005) that the truncation error of the FE first derivative and integral operators of order 4 (cubic splines) is 8 (this phenomenon being called superconvergence) for uniformly placed η levels with ignored top and bottom boundary effects. We showed that the truncation error of the second derivative operator calculated under the same conditions for uniformly placed η levels is only 6 which is still two orders better than for finite difference method used in ALADIN-NH. Truncation errors for several discretization method and derivative operators may be seen on Figure 1. Moreover, we have shown that these results are in good agreement with the real accuracy of vertical operators of ALADIN-NH being applied on smooth function $\sin(6\pi\eta)$ with uniformly placed η levels and ignored top and bottom boundary effects. See Table 1, where Calculated order comes from real application of vertical operators, while Approximated order is the result following Staniforth and Wood. We have to admit that the vertical derivative operator which transforms a function defined on half levels to a function being defined on full levels which is used for vertical velocity w has truncation error only 4, since there is no superconvergence effect. Details in the FE vertical operators definition have been described by Vivoda and Smolíková (2013).

Since an arbitrary choice of the order of splines used for FE discretization has been enabled recently in ALADIN-NH, we have investigated RMSE of distinct vertical operators applied on a smooth function depending on the order of splines used and number of vertical levels. We may observe strong "saturation" for higher orders and vertical resolutions, see Figure 1. We see the explanation in rounding error cumulation for needed high number of operations.

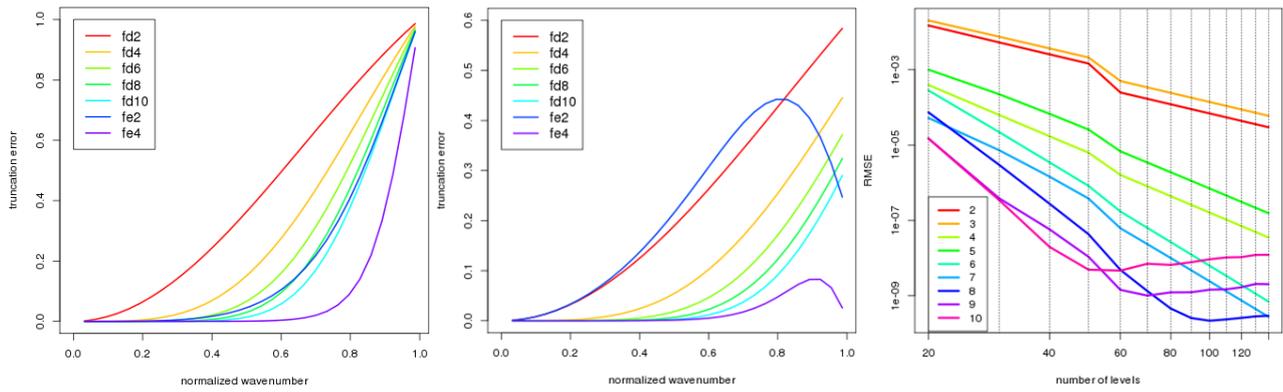


Figure 1: Left and middle: The absolute value of the truncation error of the first and the second derivatives determined by approximation of given functions by Taylor series expansion for different finite difference (fd) and finite element (fe) discretizations with uniform resolution (after Staniforth and Wood); in the legend the number indicates the order of scheme used. Right: RMSE of the first derivative vertical operator applied on a smooth function with the increasing number of regularly placed vertical η levels; colors indicate the order of splines used in the FE scheme.

Table 1: Accuracy of vertical operators. For details consult Vivoda and Smolřková (2013).

Operator	MAE	Calculated order	Approximated order
first derivative	8.62e-12	8.02	8
first derivative with BC	8.62e-12	11.92	8
first derivative with BC half levels -> full levels	1.04e-05	4.00	4
second derivative with BC	4.47e-08	6.06	6
integral with BC	4.22e-16	9.13	8

3 Idealized tests

Further, we have concentrated ourselves on the theoretical explanations and studies of the simple cases with known analytical solution to see the impact of vertical discretization method choice.

A slightly modified version of the idealized test setup used by Skamarock and Klemp was proposed by Baldauf and Brdar, 2012: the quasi linear 2-dimensional expansion of sound and gravity waves in a channel induced by a weak warm bubble. The modification allows derivation of an exact analytical solution for the compressible, nonhydrostatic Euler equations that are the basis for ALADIN-NH model. The derived analytical solution is supposed to be used as a benchmark to assess compressible dynamical cores. This test is designed for usual height based vertical coordinate models with vertical velocity imposed to be zero at bottom and top of the domain. These boundary conditions are natural for height based vertical coordinate models, but not for mass based vertical coordinate models as ALADIN-NH. In the original solution, the evolution of the perturbation is a set of waves that propagate horizontally.

However, for a mass based model, the vertical velocity is imposed to be zero at the bottom but at the top the model is open and there are not boundary conditions for the vertical velocity: atmosphere can evolve freely and can move up and down at the top. As a consequence, in a mass based vertical coordinate model, the evolution of the initial perturbation is a set of waves that propagate both in the horizontal and vertical directions. Trying to fix vertical velocity to zero at a given height is not an easy task, as the model itself is not prepared nor designed for such an imposition. Our simple proposition to solve this difficulty was to impose vertical velocity to be zero by the sponge instantaneously and directly on the upper boundary of 10km. However, the results show that

mass is being lost through the upper boundary and the evolution of wave is smooth down by this fact.

As a consequence, the difference between the vertical velocity field of the experiment with FE used in vertical discretization and the one with FD used is order of magnitude smaller than the overall error of both experiments compared to the analytical solution. See Figure 2 for vertical velocity fields and Figure 3 for comparison of vertical velocity value in the height of 0.5km, 5km (middle of the domain) and 9.5km. The curves for VFE and VFD in various levels are almost indistinguishable, except for the higher level, where VFE seem to work better in the middle of the domain. There is some noise generated with VFE close to lateral boundaries.

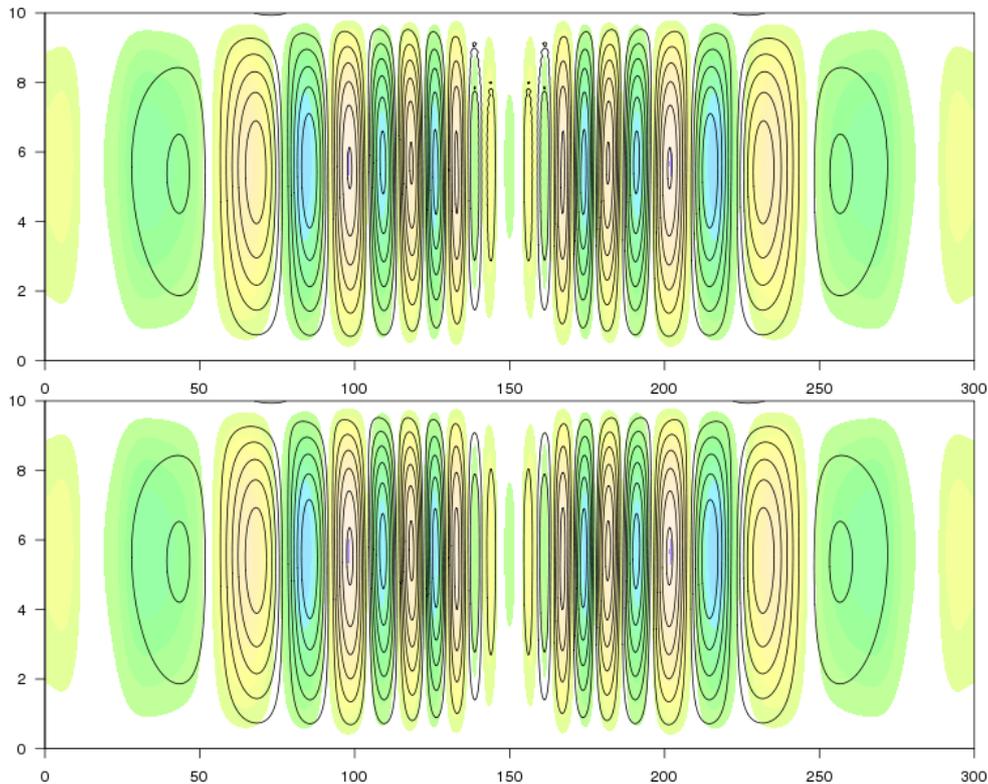


Figure 2: Potential temperature field in Baldauf-Brdar test (gravity waves in a channel); shading shows analytical solution, while contours represent solution with VFE (top) and VFD (bottom).

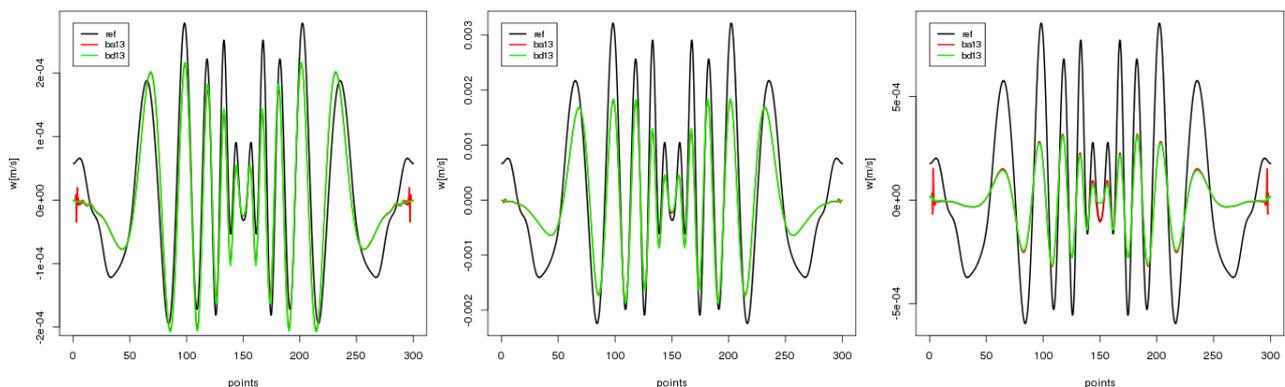


Figure 3: Potential temperature in Baldauf-Brdar test for distinct vertical levels in 500m vertical resolution; left: 4th level in 0.5km; middle: 40th level in 5km (middle of the domain in vertical); right: 76th level in 9.5km. Analytical value is in black, VFE in red and VFD in green.

Fortunately, we may modify the initial perturbation of temperature to localize its maximum in the lower atmosphere. This perturbation would evolve as a set of waves propagating horizontally and vertically, and, because the initial perturbation is located in the lower atmosphere, it will take some minutes to get the upper atmosphere. During this time, the vertical velocity at the top will be zero in the analytical solution and, therefore, also should be zero in the numerical solution for any model, open or not at the top. He has modified the initial condition in the analytical solution calculation and we have run another set of experiments for this new simple case.

In this experiment the maximum perturbation was 0.01K from the basic value 250K at the middle of the domain in horizontal and in the height of 2km. The horizontal resolution was 500m with 256 points in horizontal and successively, 500m, 1000m and 2000m in vertical with corresponding number of vertical levels to get 40km of vertical extent of the whole domain. The timestep used was 0.5s to avoid errors of time discretization to influence the results and the integration has continued until 600s has been reached.

Unfortunately, it is not true that the vertical velocity at the top of the model will remain zero. In the mass based vertical coordinate model as ALADIN-NH is, the sink of mass through the top is unavoidable and the solution is again distorted. On the other hand, we could notice that the accuracy of the experiment with FE was enhanced compared to the one with FD in vertical discretization. And this claim holds for all three resolutions used, 500m, 1000m and 2000m, see Figure 4 for an illustration of this fact.

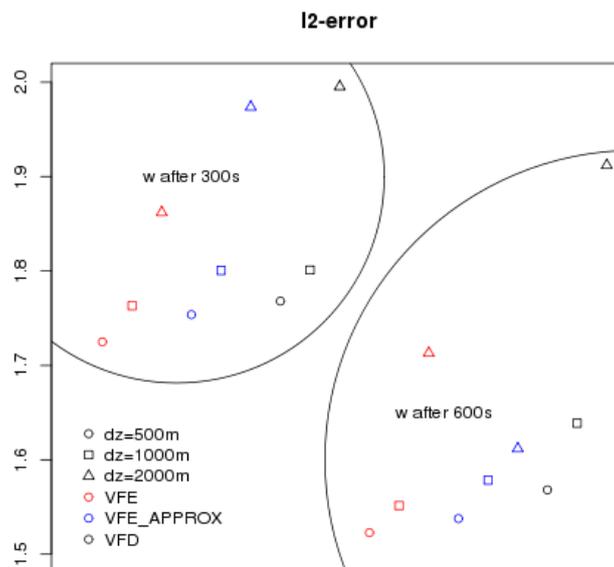


Figure 4: Relative l_2 -error for vertical velocity field in distinct experiments with modified Baldauf-Brdar test after 300s and 600s. VFE_APPROX covers choice LVFE_APPROX=TRUE.

As a conclusion we have to admit that even though the vertical operators used in FE discretization are more precise the effect of their usage is difficult to be seen in idealized tests. Nevertheless, it was shown that gravity waves are better represented in VFE than VFD. The reduction of the error (not big, but about 5% in the ideal case tests) could be appreciable in some situations, for instance in the simulation of strong downslope winds where gravity waves are an important ingredient.

4 High resolution 3D simulations

Another important task for VFE topic was to show that FE may be as stable as FD in 3D real simulations in high horizontal resolutions. We have chosen two months of experiments, January 2014 and July 2014, over the domain covering the Alpine region. The horizontal resolution was 1.25km and vertically we have used Czech operational setting of 87 levels. We have run one integration per day from 00UTC up to +24hours. Results obtained have confirmed previous results obtained with coarser resolution of 2.2km over Czech domain. We may conclude that:

- VFE scheme used in NH with proper setting of FE parameters and proper setting of vertical levels may be as stable as FD scheme; the time step used in our experiments was 50s.
- It is difficult to find any benefit from FE used in vertical discretization concerning objective scores.
- The precipitation field is modified by FE in such a way that there is bigger number of grid points without rain (cumulated precipitations for 1 hour < 0.1mm) and bigger number of grid points with highest values of cumulated precipitations (>30mm/hour). Consequently, there is smaller number of grid points with modest rain between 0.1 and 30mm. We consider this trend as beneficial since the field of precipitation is more sharpened and the rain locations are more restricted. We should admit that this phenomenon is clearly present but not as intensive as to be observed easily. Our conclusion comes from the statistical analysis of results. Concerning the intensity of observed trend, we may summarize: we have 24 events per one day (hourly cumulated precipitations), 31 days in a month, i.e. 744 events per grid point; we have 403x694 grid points in the area; thus we have $2 \cdot 10^9$ events in the area; if we prepare 17 histogram intervals, the two experiment series differ only in 0.3% of these events. Nevertheless, the phenomenon is observed in both, summer and winter series.

5 References

- Baldauf, M., and Brdar, S., An analytic solution for linear gravity waves in a channel as a test for numerical models using the non-hydrostatic, compressible Euler equations. *Q. J. R. Meteorol. Soc.*, 139(677), 1977-1989, 2012.
- Juan Simarro, Baldauf and Schär analytical and numerical solution for the Euler Equations. A test., HIRLAM internal report, 2014.
- Simarro, J., Homar, V., and Simarro, G., A non-hydrostatic global spectral dynamical core using a height-based vertical coordinate, *Tellus A*, 65, 1-20, 2013.
- Simarro J., Hortal M., A semi-implicit non-hydrostatic dynamical kernel using finite elements in the vertical discretization, *Quart. J. Roy. Meteor. Soc.*, 138,
- Staniforth A., Wood N., Comments on 'A finite-element scheme for the vertical discretization in the semi-Lagrangian version of the ECMWF forecast model' by A.Untch and M.Hortal (April B, 2004, 130, 1505–1530), *Q. J. R. Meteorol.Soc.*, 131,765–772, 2005.
- Untch A., Hortal M., 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, 2004.
- Vivoda J., Petra Smolřková, Finite elements used in the vertical discretization of the fully compressible forecast model ALADIN-NH, ALADIN/HIRLAM Newsletter no.60, 2013.

A new Solver for the Non-Hydrostatic Semi-Implicit Scheme in HARMONIE using Green Functions

Carlos Geijo, (cgeijog@aemet.es)

Spanish State Meteorological Agency, AEMET

This article is longer than usual but the topic requires this length (very specific issue of the vertical discretization of the semi-implicit scheme in the ALADIN non-hydrostatic dynamical kernel) .

1 Introduction

This work focuses on the very specific issue of vertical discretization of the semi-implicit (SI) scheme in the ALADIN non-hydrostatic dynamical kernel. As it is well known, a semi-implicit treatment of fast non-hydrostatic dynamics is absolutely necessary in operational weather forecasting to avoid stringent CFL limitations that if violated would cause instabilities. The quest for a stable and reliable scheme goes back as far as 1995 [1] when a first implementation was done following ideas by Laprise on how to extend to non-hydrostatic (NH) dynamics the adiabatic formulation of hydrostatic Primitive Equations (HPE) in mass-based terrain-following coordinates. Since then many important and successful steps forward have been achieved culminating in a “state-of-the-art” weather model (see [2], [3] for a detailed history of the development of this scheme). The first formulation of the vertical discretization was a finite differences scheme (FD) which is still used as default in the current version of the HARMONIE NWP package. Much later, around 2005, started the development of a different scheme based on finite elements (FE). Again the inspiration came from the formulation of this approach in HPE [4], but the first attempts met no success due to difficulties originated in the presence of derivative and integral operators mixed in the equations [2], a feature that is absent in HPE mass-based coordinates. Recently the efforts have started to give encouraging results and first versions of the FE scheme are now included in the latest HARMONIE releases [5].

This work was originally motivated by the interest in understanding better the dynamical kernel with the problem of initialization in mind. However, I think that it has evolved into something rather different. The difficulties found with the FE scheme above mentioned, in particular, the problems encountered when trying to reduce the SI system into a single Helmholtz equation as it is done in the FD scheme, have been central in the formulation of the ideas here presented. The outcome of this work is an alternative approach for the SI scheme based on the use of Green Functions and interpolating polynomials which permits to find a unique C^2 solution by quadratures. Numerical tests performed for a wide range of horizontal scales (100m to 10000Km) give very satisfactory results. Of course more tests are necessary. However, I believe that the ideas here presented are interesting in themselves and this paper aims at communicating them in as a complete and elaborated way as I can at this moment.

The ideas here presented have been codified in a set of FORTRAN95 and C programs and these are kindly available to anyone interested in studying in more detail this approach. This software is the one used in the numerical tests described in section 5 of this paper.

2 The SI System in the NH-PDVD Model

After time discretization and before vertical discretization, the system to solve can be written as ([6], page 17, map factor is omitted for clarity):

$$\begin{aligned}
 D + (kH)^2 (N\Delta t)^2 (-G^*[T] + G^*[P] - P - \pi_s) &= D^\bullet \quad ; \quad D = D' \Delta t; \quad N^2 = \frac{g}{H}; \quad H = \frac{RT^*}{g} \\
 T + \left(\frac{R}{c_v}\right)(D+d) &= T^\bullet \quad ; \quad T = \frac{T'}{T^*} \\
 P - \left(S^*[D] - \left(\frac{c_p}{c_v}\right)(D+d)\right) &= P^\bullet \quad ; \quad P = P' \quad Eq 1 \\
 \pi_s + N^*[D] &= \pi_s^\bullet \quad ; \quad \pi_s = \frac{\pi_s'}{\pi_s^*} \\
 d + (N\Delta t)^2 L^*[P] &= d^\bullet \quad ; \quad d = d' \Delta t
 \end{aligned}$$

The unknowns in this system correspond to the variables in the next time step and the “dot terms” on the r.h.s include the dynamical, physical and SI corrections [6]. Later, in section 4.4, these terms are considered in bigger detail, but for the time being let us regard them as given, and that we have some smooth approximation in the vertical coordinate for them (i.e. that we can take derivatives on them). We see that there are only a few parameters that define the system: T^* gives the scale height of the isothermal atmosphere and also (through N^2) the buoyancy frequency, π_s^* (values for top and bottom) gives the depth of the atmosphere. The space and time resolutions (k and Δt) complete the set. The unknowns have been converted into pure numbers by using these parameters. Also a few integral and differential operators appear: $G^*[\]$, $S^*[\]$, $N^*[\]$ and $L^*[\]$. These operators are:

$$\begin{aligned}
 G^*[X] &= \int_{\eta}^1 \frac{m^*}{\pi^*} X d\eta = \int_{\xi}^{\xi^-} X d\xi' \quad ; \quad \xi = \ln\left(\frac{\pi^*(\eta)}{\pi_{top}^*}\right); \quad \xi^- = \ln\left(\frac{\pi_s^*}{\pi_t^*}\right); \quad m^* = \frac{d\pi^*}{d\eta} \\
 S^*[X] &= \left(\frac{1}{\pi^*(\eta)}\right) \int_0^{\eta} m^* X d\eta = e^{-\xi} \int_0^{\xi} e^{\xi'} X d\xi' \\
 N^*[X] &= e^{-\xi} \int_0^{\xi} e^{\xi'} X d\xi' \quad ; \quad L^*[X] = (\partial^* \partial^* + \partial^*)[X] = (\partial^2 + \partial)[X] \quad ; \quad \partial^* = \left(\frac{\pi^*}{m^*}\right) \partial_{\eta} \quad \partial = \partial_{\xi}
 \end{aligned}$$

They are expressed in the hybrid η coordinate (used by the model) and in a height based coordinate ξ (0 at the top, ξ^- at the bottom) which we will find more convenient to work with. These operators satisfy a number of relations [2]. Simple but lengthy manipulations reduce the system to a single linear second order differential equation (DE) for the vertical divergence:

$$\begin{aligned}
 \left(\frac{c_p}{c_v}\right) (-\lambda + L^*)[d] &= \frac{-1}{(N\Delta t)^2} \left(1 + \frac{c_p}{c_v} k^2\right) [d^\bullet] + \left(\partial - \frac{c_p}{c_v} L^*\right) [D^\bullet] \quad Eq2 \\
 &+ k^2 \left(\frac{c_p}{c_v} (\partial + 1) - 1\right) [T^\bullet] + (L^* + k^2 \left(\left(1 - \frac{c_p}{c_v}\right) (\partial + 1)\right)) [P^\bullet]; \quad k^2 = (kH)^2 (N\Delta t)^2
 \end{aligned}$$

We see that these manipulations involve derivatives only up to second order in the dot terms. Therefore C^2 smoothness is sufficient to get to this differential equation (in fact C^1 and C^0 are

sufficient for $T \cdot d'$ respectively). This suggests that a discretization in the vertical coordinate by means of cubic splines may be of interest. How this discretization can be implemented is explained in section 3 of this paper. The next problem then is how to solve the differential equation in this basis. The right answer is to use Green functions (GF). The basic theory of GF is presented in annex A.

A GF gives the inverse operator of a differential operator. It allows express the solution of a linear differential equation:

$$L[u] = f \quad Eq3$$

and also the derivatives of this solution (up to order n-1, where n is the order of L) by means of quadratures (i.e. integrations). For instance, if L is second order,

$$u(z) = \int_a^b GF(z, \eta) f(\eta) \quad ; \quad u'(z) = \int_a^b GF'(z, \eta) f(\eta) \quad Eq4$$

and u'' is read off the defining equation Eq3. Higher order derivatives, if they exist, can be computed by deriving both terms of the equation. The specification of the boundary conditions (BC) for the problem is incorporated in the construction of the GF (see annex A). It is also worth mentioning here that according to Eq4 a solution can be obtain for Eq3 also in the case when f fails to be continuous in a set of measure zero. Eq4 is then to be interpreted in a “weak sense” or “distributional sense”. The solution u fails in this case to have continuous n^{th} order derivative, although it's got that derivative as a distribution [8].

A key point in this work is simply that the above integrals Eq4 already provide an algorithm to solve the problem. The problem is then cast into an interpolation problem. Interpolations are at the heart of numerical differentiation and integration. This connection is established through the use of the spline basis, and in this basis they are of course inverse one of each other.

Performing integrations on splines is an easy matter, but to obtain a complete solution it is necessary to undo all the substitutions that lead to Eq2. This was a major difficulty in the successful completion of the design of the algorithm. By repeat use of integration by parts, we will see that all the variables, their derivatives (up to second order) and also the values they take when submitted to the $G^*[]$, $S^*[]$ and $N^*[]$ operators can be computed also by quadratures in a convenient way.

The method outlined has some clear advantages. I can think of the following ones (not by order of importance):

It works with non-staggered vertical grids. This probably will help with problems like the X-term issue. In fact, it is possible to compute the solution and its derivatives (up to the 2nd) at any point.

The algorithm permits the introduction of BC in a natural way, opening in this manner a number of possibilities in the issue of nesting in the vertical direction within a host model.

The calculation by integrals of the solution is always stable (the GF is a compact operator) and as will be seen also accurate. Therefore we can use realistic values for scaling parameters T^* , π_s^* which should have a beneficial impact on the non-linear “remaining”, and hence on stability.

The method can equally be applied *mutatis mutandis* to the NH-GEOGW model (see annex C).

3 Vertical Discretization using Cubic Splines

The presentation given before considers the problem as one of infinite dimension. In order to implement a numerical scheme, we have to find a suitable finite dimension approximation. As already mentioned, the fact that C^2 smoothness puts out of the way the problems related with the operator constrains required to reduce the problem to a single differential equation suggests that using cubic splines can be an interesting approximation.

As is well known, cubic splines are piecewise polynomials (PP) of third order that interpolate a set of values $\{y_i\}$ given at N knots $\{a_i\}$. The $4N-4$ required parameters can be specified by imposing continuity at the interior knots for the 0th, 1st and 2nd derivatives ($3N-6$ conditions). The remaining $N+2$ unknowns are determined using the N values plus 2 additional conditions. Usually one can make one of two choices: a) natural splines, where the second derivative at the 1st and N^{th} knots are taken as zero, b) clamped splines, where the values of the first derivatives at the 1st and N^{th} knots are given. This last choice usually gives better results than the first, but of course it requires knowledge of these derivatives, which are not always available. Both methods establish an $N \leftrightarrow N+2$ correspondence which looks odd for a "change of coordinates".

There is however another alternative which is found to be more convenient [7]. It consists in not requiring the endpoints to be knots, but still impose $\text{Spl}(a_1)=y_1$, $\text{Spl}(a_N)=y_N$. This scheme, which does not need additional information (i.e. it is " $N \leftrightarrow N$ "), appears to be better than natural splines because the vanishing of the second derivative at the end points usually introduces weird variations near the boundaries. To be precise, the proposed transformation of coordinates reads (Spl'' stands for second derivative in the vertical):

$$\begin{pmatrix} 2 & \lambda_2 & 0 & \\ \mu_3 & 2 & \lambda_3 & 0 \\ 0 & & 0 & \lambda_{N-2} \\ \mathbf{M} & 0 & \mu_{N-1} & 2 \end{pmatrix} \begin{pmatrix} M_2 \\ \mathbf{M} \\ \mathbf{M} \\ M_{N-1} \end{pmatrix} = \begin{pmatrix} d_2 \\ \mathbf{M} \\ \mathbf{M} \\ d_{N-1} \end{pmatrix} ; \quad \text{Spl}''(a_i) = M_i \quad ; \quad i = 2, \dots, N-1$$

$$\lambda_i = \frac{h_{i+1}}{h_i + h_{i+1}} ; \mu_i = 1 - \lambda_i ; h_i = a_i - a_{i-1} ; \quad i = 3, \dots, N-2 \quad \text{Eq5}$$

$$d_i = \left(\frac{6}{h_i + h_{i+1}} \right) \left(\left(\frac{y_{i+1} - y_i}{h_{i+1}} \right) - \left(\frac{y_i - y_{i-1}}{h_i} \right) \right)$$

and the remaining $d_2, \lambda_2, d_{N-1}, \mu_{N-1}$ are obtained from the two conditions $\text{Spl}(a_1)=y_1, \text{Spl}(a_N)=y_N$:

$$\lambda_2 = 2 \left(\frac{h_2}{h_{32}} \right) \frac{(h_3^2 - h_2^2)}{(h_{32}^2 - h_3^2)} ; h_{32} = h_3 + h_2 ; \quad \mu_{N-1} = 2 \left(\frac{h_N}{h_{NN-1}} \right) \frac{(h_{N-1}^2 - h_N^2)}{(h_{NN-1}^2 - h_{N-1}^2)} ; h_{NN-1} = h_N + h_{N-1}$$

$$d_2 = \left(\frac{12}{h_{32}} \right) \left(\frac{h_3 y_1 + h_2 y_3 - h_{32} y_2}{h_{32}^2 - h_3^2} \right) ; \quad d_{N-1} = \left(\frac{12}{h_{NN-1}} \right) \left(\frac{h_{N-1} y_N + h_N y_{N-2} - h_{NN-1} y_{N-1}}{h_{NN-1}^2 - h_{N-1}^2} \right)$$

The trick behind this construction is to extrapolate to a_1 and a_N the polynomial pieces in the 2nd and $(N-2)^{\text{th}}$ intervals. Therefore, at a_2 and a_{N-1} the interpolating spline has now also continuous 3rd derivative. Once the set $\{M_i\}_{i=2, \dots, N-1}$ is calculated by inverting the tridiagonal (and diagonal dominant, [7]) matrix in Eq 5, the set of splines is uniquely determined. For example, for the generic i -th interval $[a_{i-1}, a_i]$, $i=3, \dots, N-1$ we have:

$$\begin{aligned}
Spl_{i-1}(x) &= M_{i-1} \frac{(a_i - x)^3}{6h_i} + M_i \frac{(x - a_{i-1})^3}{6h_i} + \left(y_{i-1} - \frac{M_{i-1}h_i^2}{6} \right) \left(\frac{a_i - x}{h_i} \right) + \left(y_i - \frac{M_i h_i^2}{6} \right) \left(\frac{x - a_{i-1}}{h_i} \right) = \\
&= \alpha(i-1,1)B^3 + \alpha(i-1,2)B^2 + \alpha(i-1,3)B + \alpha(i-1,4) \quad B = \frac{(x - a_{i-1})}{h_i}
\end{aligned}$$

where the α 's are (linear) functions of the M_i 's and y_i 's

$$\begin{aligned}
\alpha(i-1,1) &= (M_i - M_{i-1}) \left(\frac{h_i^2}{6} \right); \quad \alpha(i-1,2) = M_{i-1} \left(\frac{h_i^2}{6} \right); \\
\alpha(i-1,3) &= y_i - y_{i-1} - 2(M_{i-1} + M_i) \left(\frac{h_i^2}{6} \right); \quad \alpha(i-1,4) = y_{i-1}
\end{aligned}$$

For the intervals at the edges: $[a_1, a_2]$ and $[a_{N-1}, a_N]$, the formulas are more complex but still linear functions of M_2, M_3 and M_{N-2}, M_{N-1} respectively. Given that the M 's are also linear functions of the d 's, which in turn are also linear functions of the y 's, we find that the proposed change of coordinates (α 's \leftrightarrow y 's) is a linear one! . Considering that the dynamical kernel uses spectral representations of the fields to compute horizontal derivatives, which permits to solve the SI system for one k at a time, is interesting to note that we have the following commutative diagram:

$$\begin{array}{ccc}
y_i & \leftrightarrow & DFT[y_i] \\
\mathbf{b} & & \mathbf{b} \quad \quad \quad Eq6 \\
\alpha(i,:) & \leftrightarrow & DFT[\alpha(i,:)]
\end{array}$$

These splines have still other nice properties. For example, according to [7] it is true that

$$\max_{a \leq x \leq b} |f^k(x) - Spl^k(x)| = O(h^{4-k}) \quad k = 0, 1, 2$$

if $f(x)$ has a continuous fourth derivative, and if $\max_i (h/h_i) \leq \beta < \infty$ as $h \rightarrow 0$ for a fixed β . Therefore the calculation of the second derivative implicit in this transformation is "second order", and for the first derivative and the function value is better than that.

It is fair to raise also some objections. It is known that the splines interpolate with the least curvature, i.e. they minimize the integral:

$$E(g) = \int_a^b [g''(x)]^2 dx$$

therefore it is clear that this spline basis cannot be "complete" in C^2 , i.e., given a "point" in C^2 it is not possible to find an spline arbitrarily close to it. Although we have the following results [7], for $f \in C^2$ and $h = \max \{h_i\}$

$$\max_{a \leq x \leq b} |f(x) - Spl(x)| \leq h [hE(f)]^{1/2} \quad ; \quad \max_{a \leq x \leq b} |f'(x) - Spl'(x)| \leq [hE(f)]^{1/2}$$

which would allow us to conclude that, taking h sufficiently small we can make these norms as small as we want, a similar result for the second derivative cannot hold. The cubic splines interpolate linearly the second derivative between the knots, a C^2 function with second derivative which wanders in an arbitrary (but continuous) way between the knots will never satisfy similar conditions in either the L^2 norm nor much less the uniform norm.

One may be surprised by the use of splines as a basis in numerical calculations. Usually, splines are employed to interpolate data and the least curvature property is beneficial for filtering noise. However, one may also regard numerical calculations as not completely noise free.

The method gives just C^2 solutions. It is not clear at this moment whether working with interpolation schemes of higher degree of smoothness (e.g., Hermite Interpolation) will be practical and retain the appealing features of this method (see section 4.4).

4 The Algorithm

Let us present the computational scheme in detail

1. The Green Function for $L^* - \lambda$

This section will be illustrated with the problem with homogeneous BC on the function, i.e.: $u(0)=u(\bar{\xi})=0$. The kernel of the operator $L^* - \lambda$ is in this case spanned by the two following solutions

$$u_1(\xi) = e^{a^+\xi} - e^{a^-\xi} \quad ; \quad u_2(\xi) = e^{a^+\xi} - e^{\Delta\bar{\xi}} e^{a^-\xi} \quad ; \quad W(u_1, u_2) = \Delta(e^{\Delta\bar{\xi}} - 1) e^{-\xi}$$

These two solutions are linearly independent if the wronskian is different from zero. When this happens the construction of a GF is always possible and a unique solution can be found. The roots of the characteristic equation take the values:

$$a^\pm = \frac{-1 \pm \sqrt{1+4\lambda}}{2} = \frac{-1 \pm \Delta}{2} \quad ; \quad \Delta = \sqrt{1+4\lambda}$$

and λ is:

$$\lambda = \frac{\left(1 + k^2 \frac{c_p}{c_v} \left(1 + (N\Delta t)^2 \frac{R}{c_p}\right)\right)}{\left((N\Delta t)^2 \frac{c_p}{c_v}\right)}$$

This λ is always positive and it follows that for the problem at hand Δ is never zero. In passing we note that the λ values for which $\Delta=0$ give the spectrum of the full elastic model (without the Lamb wave) on which the ALADIN-NH dynamics is based. This is shown in annex B.

The value of $\bar{\xi}$ depends on the depth of the atmosphere. We will assume that the vertical coordinate stretches along a finite interval $[0$ (top), $\bar{\xi}$ (bottom)]. If we take $\pi_s^* = 950hPa$, $\pi_{top}^* = 7.5hPa$ then $\bar{\xi} = 4.84$. As explained in annex A, we build the GF for this problem, which happens to be:

$$g(z, \xi) = \begin{cases} z < \bar{\xi} & \Omega_R(z) \left(e^{\hat{a}^+ \xi} e^{-\frac{\Delta}{2} \bar{\xi}} - e^{\hat{a}^- \xi} e^{\frac{\Delta}{2} \bar{\xi}} \right) ; \quad \Omega_R(z) = \frac{\sinh\left(\frac{\Delta}{2} z\right) e^{-\frac{z}{2}}}{\Delta \sinh\left(\frac{\Delta}{2} \bar{\xi}\right)} \\ \xi < z & \Omega_L(z) \left(e^{\hat{a}^+ \xi} - e^{\hat{a}^- \xi} \right) ; \quad \Omega_L(z) = \frac{-\sinh\left(\frac{\Delta}{2} (\bar{\xi} - z)\right) e^{-\frac{z}{2}}}{\Delta \sinh\left(\frac{\Delta}{2} \bar{\xi}\right)} \end{cases}$$

where $\hat{a}^\pm = a^\pm + 1 = -a^m$;

and $g'(z, \xi)$ is given by replacing the Ω by Ω' . We see that, as it must be, the GF satisfies the homogeneous BC of the problem, i.e. $g(z=0, \xi) = g(z = \bar{\xi}, \xi) = 0$. The subscripts ‘‘R’’ and ‘‘L’’ remind us that in the formal solution:

$$u(z) = \Omega_L(z) \int_0^z (e^{\hat{a}^+ \xi} - e^{\hat{a}^- \xi}) f(\xi) \, d\xi + \Omega_R(z) \int_z^{\bar{\xi}} \left(e^{\hat{a}^+ \xi} e^{-\frac{\Delta}{2} \bar{\xi}} - e^{\hat{a}^- \xi} e^{\frac{\Delta}{2} \bar{\xi}} \right) f(\xi) \, d\xi \quad Eq7$$

the $\Omega_{L,R}$ weight the integral on the left interval and right interval respectively. Note that this notation may result a bit confusing because they are determined from the solutions which satisfy the right and left BC respectively.

The calculation of these functions entails the problem of hyperbolic functions with very big arguments, which in finite arithmetic translate into loss of accuracy or even produce overflows. The following table shows the values for the key parameters for different horizontal scales, calculated with 14-digits precision.

Table 1: Values of the key parameters for different spatial scales. Time step is always 60 seconds. The atmosphere depth is $\bar{\xi} = 4.84$ in this example

Scale	λ	.a+	.a-	Δ	$\sinh(\Delta/2 \bar{\xi})$
100 m	0.571E+06	0.755E+03	-0.756E+03	0.151E+04	$+\infty$
1 Km	0.571E+04	0.750E+02	-0.760E+02	0.151E+03	0.367E+159
10Km	0.572E+02	0.708E+01	-0.808E+01	0.151E+02	0.438E+16
100Km	0.732E+00	0.491E+00	-0.149E+01	0.198E+01	0.607E+02
1000Km	0.167E+00	0.146E+00	-0.114E+01	0.129E+01	0.112E+02
10000Km	0.161E+00	0.141E+00	-0.114E+01	0.128E+01	0.111E+02

In spite of the overflows, the GF is perfectly finite for all scales. In fact it is always, in absolute value, $\leq 1/\Delta$. Luckily it is not difficult to achieve good accuracy by judicious approximations of hyperbolic functions by exponentials. The figures 1 and 2 display an example: $g(z=\text{level}17, \xi)$ and

$g'(z=\text{level}17, \xi)$. The most difficult cases (small scales) show that the upper bound $1/\Delta$ is reached and also that the “unit jump” (see annex A) is satisfied with high accuracy (more about this in section 4.2).

The general features of these functions are apparent in the plots. The GF becomes more and more narrow the shorter the scale, indicating that for small waves only nearby zones contribute to the solution. The GF is always negative, as one expects from a negative operator (eigenvalues always < 0 , see annex B).

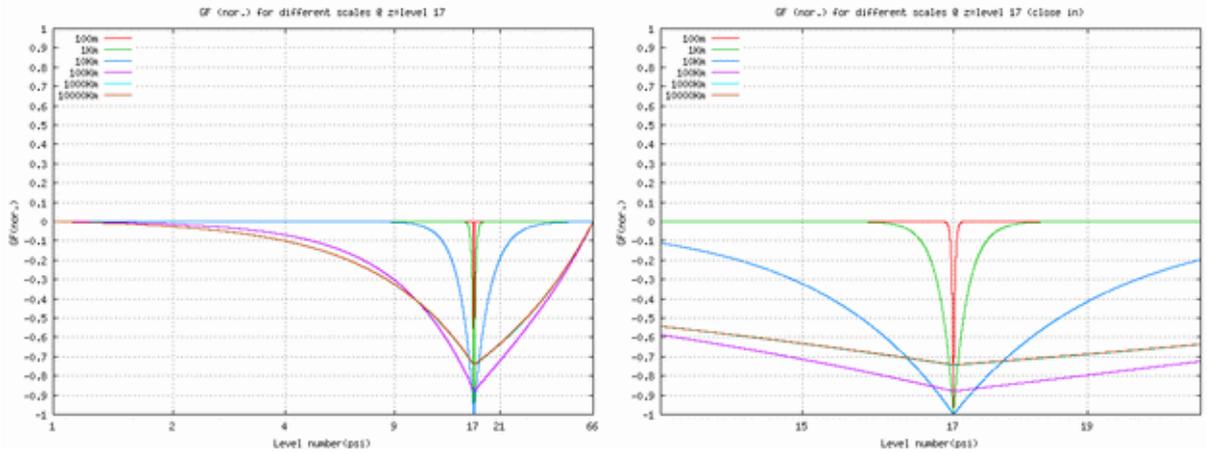


Figure 1. Green Functions (normalized) for the problem $u(z_0)=u(z_1)=0$ for different horizontal scales for the calculation of the solution at ξ corresponding to level 17 in the 65-levels vertical discretization. For the shortest scale (100m) the solution picks up contributions from only very near distances. The longest scale considered is 10000Km. In this case the GF has much more depth and, constrained by the boundaries, is no longer symmetric. The figure on the right is a zoom in around level 17

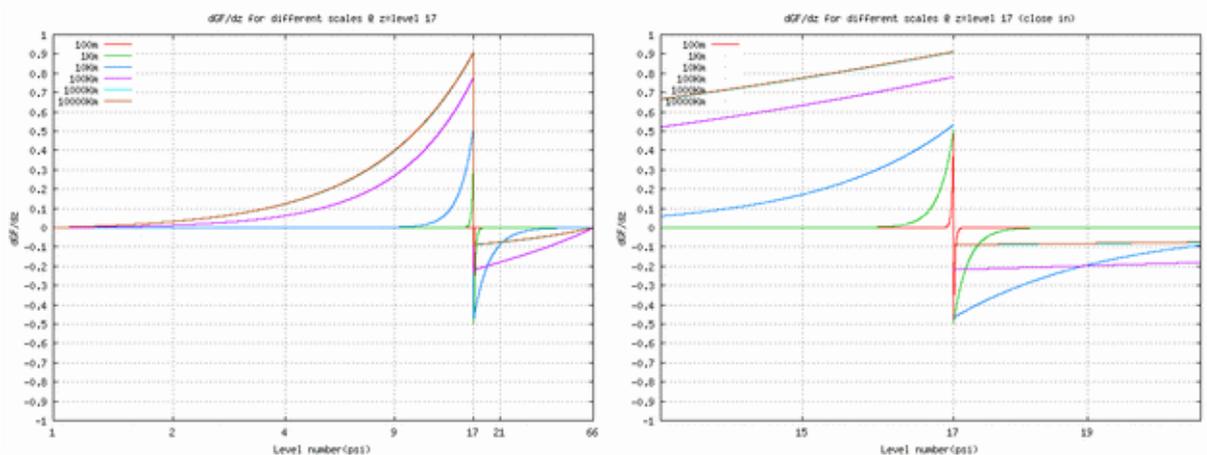


Figure 2. Green functions for the calculation of the derivative in the problem $u(z_0)=u(z_1)=0$. “The unit jump” is clearly visible in all cases. The dependence on horizontal scale follows the same patterns as in fig1

2. The Numerical Solution for d

Let us see now how the calculations can be carried out with satisfactory accuracy. We will make use many times of the following formula. If $P_n(x)$ is a polynomial of degree n:

$$\int e^{ax} P_n(x) dx = e^{ax} Q_n^a(x) = e^{ax} \sum_{j=0}^n \frac{(-1)^j}{a^{j+1}} \frac{d^j P_n(x)}{dx^j} \quad Eq 8$$

$Q_n(x)$ is another polynomial also of degree n and its dependency on the scale factor “ a ” is explicitly indicated. With this notation the solution for the problem when the “ P ” (the r.h.s of Eq3) is a whole polynomial in $[0, \bar{\xi}]$ (i.e. not piecewise) can be quickly written down:

$$u(z) = -\frac{1}{\Delta} \left(Q_n^+(z) - Q_n^-(z) \right) + (-\Omega_L(z)) \left(Q_n^+(0) - Q_n^-(0) \right) + \Omega_R(z) e^{\frac{z}{2}} \left(Q_n^+(\bar{\xi}) - Q_n^-(\bar{\xi}) \right) \quad Eq 9$$

where the +,- signs refer to which \hat{a}^+ or \hat{a}^- term in the GF has been used in the calculation (see Eq7). We have made also use of the following interesting identities, which are consequences of the continuity and “unit jump” conditions imposed on the GF, that is, they are always fulfilled by construction of the GF:

$$\Omega_L(z) e^{\hat{a}^+ z} - \Omega_R(z) e^{\hat{a}^+ z} e^{-\frac{\Delta}{2}\bar{\xi}} = \frac{-1}{\Delta} \quad ; \quad \Omega_L(z) e^{\hat{a}^- z} - \Omega_R(z) e^{\hat{a}^- z} e^{\frac{\Delta}{2}\bar{\xi}} = \frac{-1}{\Delta} \quad Eq 10$$

These identities play an important role in the implementation of the algorithm, allow a great simplification of the expression of the solution and also show immediately, when they are particularize for $z=0$ and $z=\bar{\xi}$, that the BC conditions are satisfied. The expression for the derivative $u'(z)$ can readily be obtained by repeating the operations with the corresponding GF. The generalization of this scheme to more general BC is also not difficult to work out.

Given that this notation is found very convenient, it will be used as a guide at the time of arranging the calculations when the r.h.s is given as a set of PP. We denote this set, the nodes and interval widths by:

$$\left\{ P_j(B) \right\}_{j=1, nlev-1} \quad ; \quad \left\{ \xi_j \right\}_{j=1, nlev} \quad ; \quad \xi = \xi_j + \Delta \xi_j B \quad ; \quad \Delta \xi_j = \xi_{j+1} - \xi_j \quad B \in [0, 1]$$

and the subscript now stands for interval number instead of polynomial degree (actually, the degree of the polynomial does not play any role here). The basic calculation on each interval is

$$\int_{\xi_j}^{\xi_{j+1}} e^{\hat{a}^+ \xi} P_j(B) d\xi = e^{\hat{a}^+ \xi_{j+1}} Q_j^{\pm}(1) - e^{\hat{a}^+ \xi_j} Q_j^{\pm}(0) \quad ;$$

with the definition of the Q 's now slightly different to that in Eq8 because of the chosen parameterization in B for each subinterval:

$$Q_j^{\pm} \rightarrow \Delta \xi_j Q_j^{\pm}$$

The value of the solution at a generic point z in $[0, \bar{\xi}]$ is

$$\begin{aligned}
& \Omega_L(z) \left[-Q_1^+(0) \right] - \Omega_L(z) \left[-Q_1^-(0) \right] + \\
& \Omega_L(z) e^{\hat{a}^+ \xi_2} \left[Q_1^+(1) - Q_2^+(0) \right] - \Omega_L(z) (+ \rightarrow -) + \quad \text{Eq 11} \\
& \dots \\
& \Omega_L(z) e^{\hat{a}^+ \xi_{j(z)}} \left[Q_{j(z)-1}^+(1) - Q_{j(z)}^+(0) \right] - \Omega_L(z) (+ \rightarrow -) + \\
& \left[\Omega_L(z) - e^{\frac{-\Delta}{2} \bar{\xi}} \Omega_R(z) \right] e^{\hat{a}^+ z} Q_{j(z)}^+(B) + \left[-\Omega_L(z) + e^{\frac{\Delta}{2} \bar{\xi}} \Omega_R(z) \right] e^{\hat{a}^- z} Q_{j(z)}^-(B) + \\
& \Omega_R(z) e^{\frac{-\Delta}{2} \bar{\xi}} e^{\hat{a}^+ \xi_{j(z)+1}} \left[Q_{j(z)}^+(1) - Q_{j(z)+1}^+(0) \right] - \Omega_R(z) (+ \leftrightarrow -) + \\
& \dots \\
& \Omega_R(z) e^{\frac{-\Delta}{2} \bar{\xi}} e^{\hat{a}^+ \xi_{j(\bar{\xi})}} \left[Q_{j(\bar{\xi})-1}^+(1) - Q_{j(\bar{\xi})}^+(0) \right] - \Omega_R(z) (+ \leftrightarrow -) + \\
& \Omega_R(z) e^{\frac{-\Delta}{2} \bar{\xi}} e^{\hat{a}^+ \bar{\xi}} Q_{j(\bar{\xi})}^+(1) - \Omega_R(z) e^{\frac{\Delta}{2} \bar{\xi}} e^{\hat{a}^- \bar{\xi}} Q_{j(\bar{\xi})}^-(1)
\end{aligned}$$

where the notation $(+ \rightarrow -)$ and $(+ \leftrightarrow -)$ means substitution of $-$ for $+$ and exchange between $+$ and $-$ respectively. The subindex $j(z)$ stands for the interval to which z belongs, i.e. $z \in [\xi_{j(z)}, \xi_{j(z)+1}]$. When z coincides with a node ($z = \xi_j$) we take $B=0$ unless $z = \bar{\xi}$ in which case $j(z)=nlev-1$ and $B=1$.

That this expansion reduces to the previous result for whole polynomials (Eq9) can be seen from the expression for the Q 's in terms of the integrand and its derivatives (Eq8). When the integrand is C^∞ the "jumps" at the nodes are zero. Also the central terms reduce to a difference proportional to $-1/\Delta$ by application of the identities mentioned above (Eq10). During the numerical tests, it was found also that in certain special cases this difference term is several orders of magnitude bigger than the others, rendering the solution identical to Eq9 (see section 5 "Numerical Tests").

In practice we will particularize the above calculation for a number $nlev$ of nodes. For this purpose let us define the following matrices:

$$\begin{aligned}
[\partial] W_L^+(k, j) &= \begin{cases} 0 & \xi_k < \xi_j \\ [\partial] \Omega_L(\xi_k) e^{\hat{a}^+ \xi_j} & \xi_k \geq \xi_j \end{cases} ; \quad [\partial] W_R^+(k, j) = \begin{cases} [\partial] \Omega_R(\xi_k) e^{\frac{-\Delta}{2} \bar{\xi}} e^{\hat{a}^+ \xi_j} & \xi_k \leq \xi_j \\ 0 & \xi_k > \xi_j \end{cases} \\
[\partial] W_L^-(k, j) &= \begin{cases} 0 & \xi_k < \xi_j \\ [\partial] \Omega_L(\xi_k) e^{\hat{a}^- \xi_j} & \xi_k \geq \xi_j \end{cases} ; \quad [\partial] W_R^-(k, j) = \begin{cases} [\partial] \Omega_R(\xi_k) e^{\frac{\Delta}{2} \bar{\xi}} e^{\hat{a}^- \xi_j} & \xi_k \leq \xi_j \\ 0 & \xi_k > \xi_j \end{cases} \quad \text{Eq12}
\end{aligned}$$

where the $[\partial]$ indicates the corresponding matrix element for the $g'(z, \xi)$. A convenient numerical test of the accuracy in the calculations is to check to what precision the diagonal elements of these matrices satisfy the following equalities (just a rephrasing of Eq10):

$$\begin{aligned}
W_L^+(j, j) - W_R^+(j, j) &= \frac{-1}{\Delta} \quad ; \quad -W_L^-(j, j) + W_R^-(j, j) = \frac{1}{\Delta} \\
\partial W_L^+(j, j) - \partial W_R^+(j, j) &= \frac{-a^-}{\Delta} \quad ; \quad -\partial W_L^-(j, j) + \partial W_R^-(j, j) = \frac{a^+}{\Delta} \quad Eq 13
\end{aligned}$$

Also other similar checks on the values that these quantities must take in order to satisfy the BC of the problem can be done. In the implementation carried out for the numerical tests presented below, all these equalities are fulfilled with nearly 14 digits precision, and this for all scales considered (from 100m to 10000Km).

3. The Back Substitution Process

Solving Eq2 by this method is, by no means, the end of the story. It is still necessary to undo all the substitutions to obtain the solutions for the other unknowns. This represents at first sight a very hard task because the solution we have found for d (Eq11) comes in a form which deters from the idea of doing calculations on it like $G^*[d]$ or $S^*[d]$, which are required to find the rest of the unknowns D , T , P and π_s . Fortunately, it turns out that integration by parts solve all these difficulties and permits the calculation of these variables in an efficient way and without making any approximation.

The D equation reads:

$$\left(1 + \frac{c_p}{c_v} k^2\right) D - k^2 \left(G^*[d] - \frac{c_p}{c_v} d \right) = D^* + k^2 \left(\pi_s^* - \left(G^*[P^*] - P^* \right) + G^*[T^*] \right)$$

And to calculate D the quantity $G^*[d]$ is required. The formal solution for d can be written (Eq7):

$$d(z) = \Omega_L(z) \int_0^z A(\xi) + \Omega_R(z) \int_z^{\xi} B(\xi)$$

Then, integrating by parts, we operate as follows

$$\begin{aligned}
G^*[d](\xi) &= \int_{\xi}^{\xi} d(\eta) = \bar{\Omega}_L(\xi) \int_0^{\xi} A(\eta) - \left[\bar{\Omega}_L(\xi) \int_0^{\xi} A(\eta) + \bar{\Omega}_R(\xi) \int_{\xi}^{\xi} B(\eta) \right] + \\
&\int_{\xi}^{\xi} \left(-\bar{\Omega}_L(\xi) A(\xi) + \bar{\Omega}_R(\xi) B(\xi) \right) \equiv Gd(\xi) - Gd(\xi) + \left(\frac{1}{a^+ a^-} \right) G^*[\{P_j(B)\}](\xi)
\end{aligned}$$

where Gd denotes the function obtained by an expression formally identical to that for d but with Ω replaced by $\bar{\Omega}$, a primitive of Ω , i.e. $\bar{\Omega}' = \Omega$. The identities (Eq10) allow the reduction of the third term, with $\{P_j(B)\}$ denoting the set of PP that represents in the spline basis the r.h.s of Eq2. The matrices associated to the $\bar{\Omega}$'s are defined in a similar way and their diagonal elements satisfy certain equalities similar to (Eq13) that are again very useful at the time of checking the accuracy of the implemented numerical code.

Similarly we have for $S^*[d]$:

$$e^{\xi} S^* [d](\xi) = \int_0^{\xi} e^{\eta} d(\eta) = -\bar{\bar{\Omega}}_R(0) \int_0^{\bar{\xi}} B(\eta) + \left[\bar{\bar{\Omega}}_L(\xi) \int_0^{\xi} A(\eta) + \bar{\bar{\Omega}}_R(\xi) \int_{\xi}^{\bar{\xi}} B(\eta) \right] +$$

$$\int_0^{\xi} \frac{e^{\eta} \{P_j(B)\}}{a^+ a^-} \equiv Sd(\xi) - Sd(0) + \left(\frac{1}{a^+ a^-} \right) e^{\xi} S^* [\{P_j(B)\}](\xi)$$

and $\bar{\bar{\Omega}}' = e^z \Omega$. We eventually will also need the calculation of $G^*[D]$, which in turn requires the calculation of $G^*[\xi d(\xi)]$, that again can be worked out following this method.

$$G^*[\eta d](\xi) = \int_{\xi}^{\bar{\xi}} \eta d(\eta) = \bar{\bar{\bar{\Omega}}}_L(\bar{\xi}) \int_0^{\bar{\xi}} A(\eta) - \left[\bar{\bar{\bar{\Omega}}}_L(\xi) \int_0^{\xi} A(\eta) + \bar{\bar{\bar{\Omega}}}_R(\xi) \int_{\xi}^{\bar{\xi}} B(\eta) \right] +$$

$$\int_{\xi}^{\bar{\xi}} (-\bar{\bar{\bar{\Omega}}}_L(\xi) A(\xi) + \bar{\bar{\bar{\Omega}}}_R(\xi) B(\xi)) \equiv G\eta d(\bar{\xi}) - G\eta d(\xi) + \int_{\xi}^{\bar{\xi}} \left(\frac{\eta}{a^+ a^-} + \frac{1}{(a^- a^+)^2} \right) \{P_j(B)\}$$

and $\bar{\bar{\bar{\Omega}}}' = z\Omega$.

So it is seen that the GF algorithm does not only allow calculate derivatives of the solution by deriving the “ Ω piece”, in a similar way it allows calculate integrals of the solution by means of primitives of the same piece.

To compute D' and D'' derive the D equation one and two times respectively. The resulting equations involve now only known quantities. For example, for the 1st derivative we have:

$$\left(1 + \frac{c_p}{c_v} k^2\right) \partial D + k^2 \left(d + \frac{c_p}{c_v} \partial d \right) = \partial D^{\bullet} + k^2 \left((\partial P^{\bullet} + P^{\bullet}) - \partial T^{\bullet} \right)$$

We see that free-slip solutions in the D can be obtained by solving Eq2 for d with “mixed BC”

$$k^2 \left(d + \frac{c_p}{c_v} \partial d \right) = \partial D^{\bullet} + k^2 \left((\partial P^{\bullet} + P^{\bullet}) - \partial T^{\bullet} \right) \quad Eq 14$$

To find the pressure departure P one can think of two ways:

a) One is direct back substitution in the P equation (third equation of the original system). This method requires $S^*[D]$ which by means of the identity

$$S^* \left[G^*[X] \right](\xi) = G^*[X](\xi) - e^{-\xi} G^*[X](\xi=0) + S^*[X](\xi)$$

is given in terms of $S^*[d]$ and $G^*[d]$, both previously computed.

b) The second possibility is to make use of the last equation in the original system, which gives P as the solution of another second order DE, with the operator $-\lambda + L^*[\]$ now replaced by $L^*[\]$ (i.e. $\lambda=0$). Now one of the roots is null ($a^+ = 0$). It can be shown that the integrations necessary in this case can also be expressed in terms of $S^*[d]$ and $G^*[d]$. To avoid inconsistencies, the BC for this DE must be extracted from the P equation. For instance, we could specify a “mixed type” BC

$$(\partial + 1)P - D + \frac{c_p}{c_v}(\partial + 1)(D + d) = (\partial + 1)P^\bullet \quad \text{at } \xi = 0 \quad \text{and} \quad \xi = \bar{\xi}$$

or BC on P alone:

$$P = P^\bullet - \left(\frac{c_p}{c_v}\right)(D + d) \quad \text{at } \xi = 0; \quad P = P^\bullet + S^*[D] - \left(\frac{c_p}{c_v}\right)(D + d) \quad \text{at } \xi = \bar{\xi}$$

or BC on P at the top and mixed at the bottom, etc... Note however that the operator now involved is not invertible for BC specified only on P' at *both* ends.

The solution obtained from different BCs will be different. For big λ values, the difference will be noticeable only in the proximity of the borders, but for large horizontal scales it will also be important far away from them. Therefore it is clear that the original system of equations can not be completely satisfied if we choose this second alternative for computing P. The second possibility is only apparent.

According to this analysis, the original system has only the degrees of freedom (DoF) contained in the specification of the BC for the solution of the first DE for d. It is worth mentioning here that in the treatment along these lines of the NH-GEOGW model, we get to two second order DEs, one for D and another one for w (see annex C). That is, the equivalent to the D equation in NH-PDVD model is a second order DE for D in the NH-GEOGW model. Therefore the NH-GEOGW model has more intrinsic DoF than the NH-PDVD model.

The rest of the unknowns at the different levels are obtained without difficulty by direct substitution in the original system and its derived equations. We have now computed the following set of variables at the nlev knots.

Table 2 Set of computed quantities at every knot by application of the method described in sections 4.1-4.3.

$.d, d', d'', G^*[d], S^*[d], G^*[z d(z)]$
$D, D', D'', S^*[D], G^*[D]$
$P, P', P'', G^*[P]$
$T, T', T'', G^*[T], \pi_s$

4. Calculation of the Dot Terms

So far we have not considered how the r.h.s of the original system (the “dot terms”) can be computed in the spline basis. In order to discuss this point, we need to examine in some detail how this r.h.s is calculated. As starting point we refer again to [6] (page 6) and write schematically the 2TL-SL scheme as (omitting for simplicity the tuning parameter β and the uncentering factor ε):

$$X^{t+\Delta t} - 0.5\Delta t B^{t+\Delta t} = X^t + \Delta t(A + F) - \Delta t B^{t+0.5\Delta t} + 0.5\Delta t B^t \quad \text{Eq 15}$$

Here B denotes the linear part that results after linearization of the adiabatic dynamics using an isothermal, hydrostatic, flat and resting base state (see section 1). These Bs can actually be read

straight off the NH-PDVD system (Eq1), which has the form of the l.h.s of Eq15. To fix ideas, let's consider the momentum equation (see [2], page 21). The pressure term is linearized as follows:

$$RT \frac{\mathbf{u} \cdot \nabla P}{p} + \left(1 + P + \frac{\pi}{m} \partial_\eta P\right) \mathbf{u} \cdot \nabla \left(\Phi_s + \int_\eta^1 \partial_\eta \Phi \right) \Rightarrow RT^* \mathbf{r} \cdot \nabla \left(\frac{\pi'_s}{\pi_s^*} + P - G^*[P] + G^*[T] \right) + NLT$$

where NLT stands for “non-linear terms”. After taking the divergence, the linear piece gives the “B” for the horizontal divergence. The (divergence of) NLT is moved to the r.h.s and becomes part of the “A” term in the 2TL-SL scheme (“dynamical contribution”). “A” also includes the non-inertial sources. “F” comprises the physical contribution.

NLT contains a great variety of terms. For example, the non-linear contribution to the pressure term due to inclination of the η surface is

$$\left(1 + P + \frac{\pi}{m} \partial_\eta P\right) \mathbf{u} \cdot \nabla \left(\Phi_s + \int_\eta^1 (\partial_\eta \Phi)_{NL} \right) + \left(P + \frac{\pi}{m} \partial_\eta P \right) \mathbf{r} \cdot \nabla \left(RT^* \left(\frac{\pi'_s}{\pi_s^*} - \frac{\pi'}{\pi^*} - G^*[P] + G^* \left[\frac{T'}{T^*} \right] \right) \right)$$

where $(\partial_\eta \Phi)_{NL}$ is an energy density due to non-linear departures from the base state, which not necessarily have to be small

$$(\partial_\eta \Phi)_{NL} = RT^* \left(\frac{m^*}{\pi^*} \pi' P - m' \frac{T'}{T^*} \right) / (P\pi + \pi')$$

The rest of the equations can equally be decomposed.

In the 2TL-SL scheme (Eq15) the superscripts indicate the time step at which each term belongs. The terms with superscript “t” are computed at the origin point of the SL trajectory (not a grid point in general). The terms with upper script “t+ Δt ” are computed at the final point of this trajectory (a grid point). The terms $B^{t+0.5\Delta t}$ and A+F are computed either at the medium point or as an average between the origin and final point of the trajectory [6]. This lagrangian aspect of the calculation should not interfere with the scheme here outlined. It is clear that, having at our disposal the variables given in table2 at all grid-points and levels, the r.h.s of Eq15 can be calculated at any other point by interpolations, and this can be done in several ways.

We may consider the following algorithm for computing the r.h.s of Eq15. The basic idea is that whenever we have to perform local and/or global operations along the vertical we first go to the spline basis, this avoids the need to consider a discrete formulation for these operators.

Interpolate to the point “t”, or any other in the SL trajectory, the solutions for D, T, P, d and π_s for the previous time step. Project on the spline basis as described in section 2. Calculate the B terms. For the NLT we need some kind of approximation. Quantities like $(\partial_\eta \Phi)_{NL}$ produce rational functions which break the simple scheme based on polynomials. The natural approach is to perform the products at the discrete set of levels and then transform the result into a spline. The situation is in some sense similar to that of dealing with the representation of grid-point products in spectral space. The convolution of spectral components generates a spectrum that may not be adequately resolved by the original reciprocal grid and some sort of “resizing” is required. To take the analogy a little bit further, one can note that polynomial multiplication can be seen as a convolution.

When we project the solutions of the previous step d, D, P, T and π_s on the spline basis, we do not use the solutions for first and second derivatives that the method also makes available. In fact, the calculation of the B terms with the spline solutions will not agree with those values read off table 2. This opens the possibility to improve the method by performing Hermite interpolations, that is, interpolations when not just the values of the function are specified, but also the values of derivatives of any order. If up to 2nd derivatives are given we will have piecewise C^2 fifth order polynomials instead of third order polynomials. As noticed before, the increase of the degree of the polynomials in the basis poses no problem in principle to the formulation of the algorithm, although it may have impact on its accuracy. How this extension with Hermite interpolation or even higher order smoothness solutions can be implemented is left for another paper.

Once the SL algorithm has converged, what we have is the SI forcing vector $(\mathbf{D}^*, \mathbf{T}^*, \mathbf{P}^*, \pi_s^*$ and $\mathbf{d}^*)$ at n_{lev} knots. We go to spectral space and then again project to splines, or according to the scheme Eq6, it is also possible to proceed in reverse order.

It may have gone unnoticed one detail. In the NLT above there is a ∂P term that, given that P is C^2 , will make \mathbf{D}^* just C^1 . The same problem manifests in the equation for pressure departure due to the ∂V term in the definition of D_3 . Although we have also ξ -derivatives in the vertical divergence equation, and in the temperature equation (D_3) these ones make no harm as the condition on \mathbf{d}^* is just C^0 and on \mathbf{T}^* is C^1 (see Eq2). Schematically we can represent the problem in the following way:

$$\begin{pmatrix} c_p \\ c_v \end{pmatrix} (-\lambda + L^*) [d] = LT(\partial^k, k \leq 2) + NLT(\partial^3 P, \partial^3 D + other\ terms)$$

where “other terms” stands for other terms which pose no problem. This is an inconsistency in the algorithm that must be treated by means of some sort of approximation. In the “computation receipt” sketched above, the problem is absorbed in the last projection of the SI forcing vector $(\mathbf{D}^*, \mathbf{T}^*, \mathbf{P}^*, \pi_s^*$ and $\mathbf{d}^*)$ to spline space which produces a C^2 approximation for this forcing vector. There are other ways of dealing with it.

5 The Numerical Tests

The algorithm presented in this communication has been tested in order to assess its accuracy. Two different tests have been conducted.

The first test consists in a “forward-backward” calculation: starting from a set of arbitrary splines for D , T , P , π_s and d , the r.h.s of Eq2 is computed (forward step), the system is afterwards solved (backward step) and the solutions so obtained compared with the known original splines and the degree of accuracy determined. When this test was carried out, it was found that out of the sum which gives the solution for d (Eq11), just one single term (the “difference term”) was enough to approach the solution with very good accuracy. In retrospective, this peculiar result looks natural: from all the terms in the sum, the only one that is a cubic PP is precisely this “difference term”:

$$u(z) \cong \frac{-1}{\Delta} \left[Q_{j(z)}^+(B) - Q_{j(z)}^-(B) \right] + (-\Omega_2(z)) \left[Q_1^+(0) - Q_1^-(0) \right] +$$

$$e^{\frac{\bar{\xi}}{2}} \Omega_1(z) \left[Q_{j(\bar{\xi})}^+(1) - Q_{j(\bar{\xi})}^-(1) \right]; \quad z = \xi_{j(z)} + \Delta \xi_{j(z)} B$$

$$u'(z) \cong \frac{-1}{\Delta} \left[a^- Q_{j(z)}^+(B) - a^+ Q_{j(z)}^-(B) \right] + (-\Omega_2'(z)) \left[Q_1^+(0) - Q_1^-(0) \right] +$$

$$e^{\frac{\bar{\xi}}{2}} \Omega_1'(z) \left[Q_{j(\bar{\xi})}^+(1) - Q_{j(\bar{\xi})}^-(1) \right]$$

This fact, surprising and puzzling at first, cast doubts about the generality of the test. Therefore a second kind of test was designed and carried out.

In this second test, “backward-forward”, a set of 14 arbitrary splines plus 1 arbitrary value for π_s is employed to compute the “dot terms” (fourteen is the number of terms of the l.h.s of Eq1). The system is then solved (backward step) and this solution utilized to calculate again the r.h.s (forward step). The accuracy is determined on the base of agreement between the original r.h.s and that calculated with the solutions, that is, we test to which accuracy the solutions satisfy the SI system. This second test is closer to the situation expected in an integrated test, where the forcing vector to the SI system (D^* , T^* , P^* , π_s^* and d^*) will not take values constrained by the use of a small set of splines as it happens in the first test.

For these tests we work with the set of levels known as “Prague 86-levels”. $\bar{\xi} = 4.95$ in these tests. The tests have been carried out for a different BC and their associated GF. The tests have also been conducted for a wide range of horizontal scales Dx , where Dx is 100m, 1Km, 10Km, 100Km, 1000Km or 10000Km.

To generate arbitrary splines we use 5-parameter analytical functions like:

$$p(1) e^{(p(2)\xi)} \left(\sin(2\pi(\xi + p(3))) \right)^{p(4)} \left(\cos(2\pi(\xi + p(5))) \right)^{p(6)}$$

and project them onto the spline basis.

In the first test, the values of $p(1)$ are selected in such a way that some reasonable dependency on the horizontal scale, as well as on the type of variable, is introduced. For instance, for the horizontal divergence $p(1)=tstep*Vo/Dx$, Vo is a typical wind speed (30m/s) and “tstep” is a typical time step (60 seconds). For the vertical divergence $p(1)=tstep*Wo/(H/n)$ with $Wo=5$ m/s and H is the height scale, and $n = 9 - \log_{10}(Dx)$ (it goes from 2 for the longest scale to 7 for the shortest scale). For the pressure departure P , $p(1)=0.05$ (that is, 5% of the reference hydrostatic pressure) and $p(2)=-2/\bar{\xi}$. For temperature T , $p(1)=0.2$ (that is 20% of the reference temperature) and $p(2)=-1/\bar{\xi}$. The parameters $p(4)$ and $p(6)$ take values close to “ $2n$ ” (in effect $2n \pm 1$), where $n = 9 - \log_{10}(Dx)$ putting in this way more fine vertical structure for the smaller horizontal scales. The phase lags $p(3)$, $p(5)$ are chosen at random. Fig. 3 shows the test functions for D for the longest scale (10000Km) and the shortest scale (100m). The results shown in this section do not depend substantially on variations of these choices.

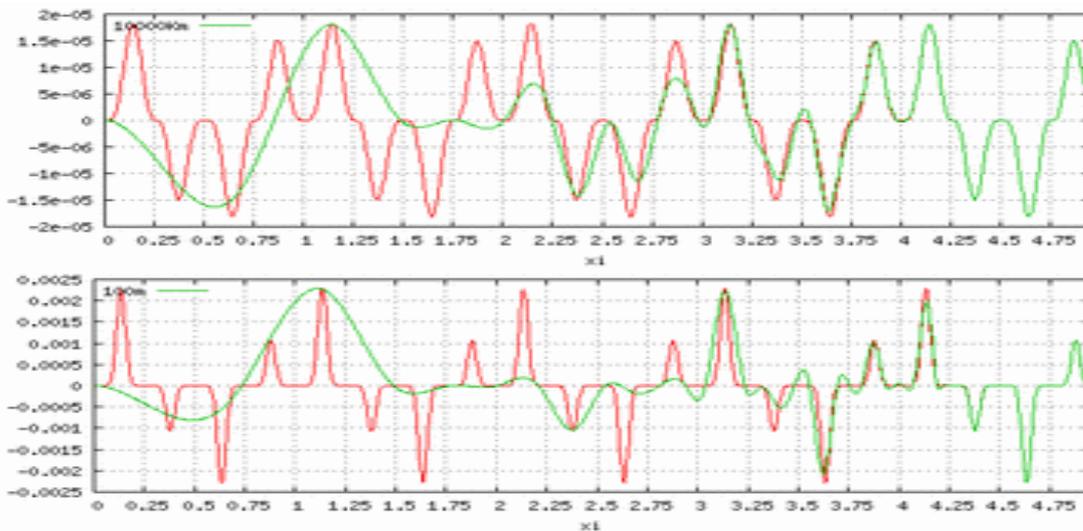


Figure 3. Test functions for D for two horizontal scales: top long scale, bottom short scale, built according to the model described in the text. Red lines are the “underlying” analytical function and the green lines are the splines. Note that at the bottom, where the vertical resolution is high, the splines closely match the smooth model. Note also that the higher powers of \sin and \cos functions produce sharper features. Actually, the spline approximation requires in this case higher density of levels to match the analytical model.

To measure the degree of accuracy attained we use the following expression:

$$\log_{10}\left(\frac{|a-b|}{|a|}\right) = \log_{10}(|a-b|) - \log_{10}|a|$$

It gives approximately the significant digit at which the quantities a and b differ. A value of -1 represents agreement up to the first significant digit, -2 up to the second and so on. The “accuracy tables” that will be shown, contain the *worst* case out of the 86 levels. When $|a|$ is below 10^{-10} times the scale of the set, the value does not enter in the evaluation of accuracy. By “scale of the set” it is meant the difference $|a|_{\max} - |a|_{\min}$. By doing this, those cases where the value $|a|$ is close to “epsilon” in 14-digits precision are disregarded. This makes the following tables represent a more faithful picture of the situation because it is frequent to find gross disparities when the calculation is done close to maximum machine precision. The “accuracy tables” presented below have the layout of that of table 2.

In order to keep within a reasonable size this paper, only the tables for the problem with “mixed BC” (Eq14) are shown. The rest of types of BC considered give similar results.

Table 3. Summary of the accuracy of the results in the first test "forward-backward" (i.e. known solutions for d, D, P, T and π_s) by direct comparison with the known solutions. Each box corresponds to the horizontal scale as indicated on top and each number to the quantity as indicated in table 2. The integer part of the numbers give the significant digit to which agreement between computed and reference values is achieved in the worst case of the 86 cases comprising a profile. See text for more explanations

100m	1Km	10Km
-4.6 -3.1 -2.4 -5.3 -5.7 -5.7	-5.3 -4.3 -3.6 -6.0 -6.4 -6.3	-5.7 -5.9 -7.3 -6.6 -6.4 -7.3
-4.7 -7.3 -6.8 -6.9 -7.3	-3.9 -6.8 -6.8 -7.0 -7.2	-4.2 -5.8 -7.3 -6.3 -7.3
-4.1 -3.5 -4.5 -3.9	-3.9 -3.4 -5.9 -5.3	-3.9 -4.4 -7.2 -6.4
-3.9 -5.7 -3.7 -5.7 -13.1	-4.2 -6.0 -5.8 -5.9 -13.6	-4.6 -5.7 -7.2 -6.0 -14.0
100Km	1000Km	10000Km
-4.9 -5.7 -7.3 -6.1 -7.1 -7.3	-2.3 -4.5 -7.3 -3.8 -4.9 -5.5	-2.0 -6.2 -7.3 -4.1 -6.1 -5.4
-4.1 -6.9 -7.3 -5.2 -7.2	-4.0 -6.7 -7.3 -3.9 -5.5	-5.2 -7.3 -7.2 -5.4 -6.8
-4.9 -7.1 -7.3 -6.9	-3.4 -6.1 -7.2 -4.7	-1.3 -6.4 -7.3 -4.3
-5.7 -7.3 -7.2 -5.8 -14.6	-3.0 -7.2 -7.3 -3.6 -15.7	-5.1 -7.3 -7.3 -4.1 -15.7

In this first "forward-backward" test it is also possible to calculate from the obtained solutions the "dot terms" and compare them with the truth. The result is shown in the next table. On each row the numbers correspond to D^* , T^* , P^* , π_s^* and d^* respectively.

Table 4. Summary of the accuracy of the results in the first test as indicated by the precision to which the solutions fulfil the SI system of equations

100m	1Km	10Km
-14.0 -5.8 -10.1 -14.0 -7.2	-14.0 -5.8 -10.1 -14.0 -8.7	-14.0 -5.8 -5.8 -14.0 -10.0
100Km	1000Km	10000Km
-14.0 -6.8 -7.5 -14.0 -10.3	-13.8 -7.8 -8.1 -14.0 -12.1	-11.7 -9.4 -9.1 -14.0 -13.9

In the light of these results it is possible to conclude that the method works well with a reasonable good accuracy.

We now move on to the second test, the "backward-forward" test. As mentioned at the beginning of this section, the fact that the solution for d in the first test takes a very specific form threw some suspicion about the generality of the test. For this second check we start off from a set of 14 arbitrary splines and one arbitrary value for π_s . These splines are generated also from the same analytical model as in the previous test. Now we have more freedom to pick up different combinations of parameters. In order to make the values presumably more realistic we can constrain a bit this freedom by distinguishing in the SI system terms of "type D", of "type T", of "type d" and of "type P" depending on which term they substitute. The splines are then scaled (i.e. parameter value for $p(1)$) in a similar way to the previous test. We then compute a r.h.s and apply the procedure to obtain our solutions. As the following table shows, these solutions satisfy with very good accuracy the system.

Table 5. As table 4 but for the second type of test.

100m	1Km	10Km
-14.0 -6.2 -12.8 -14.0 -0.6	-14.0 -5.8 -11.3 -14.0 -2.9	-14.0 -6.8 -10.9 -14.0 -5.3
100Km	1000Km	10000Km
-14.0 -7.0 -10.6 -14.0 -6.6	-13.5 -9.3 -10.9 -14.0 -11.3	-10.1 -11.6 -12.4 -14.0 -13.7

Although the values for d^* for the short scales are quite low, a closer look at the results shows that this low accuracy happens only for a few levels, and always with small $|a|$ values. For instance, the number -0.6 in the table arises from the difference between $0.24e-09$ and $0.18e-09$ when the scale of the set is $0.1e-01$. The accuracy tests are therefore considered as satisfied. These results were obtained for the mixed BC (Eq14). The last figure shows the solution for D' which depicts the expected behaviour of going to zero at the boundaries

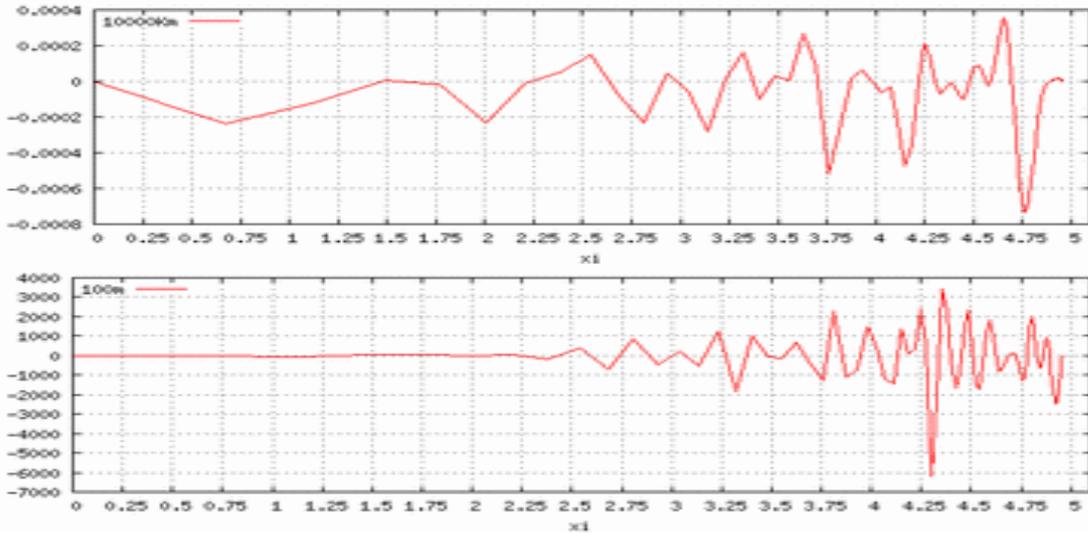


Figure 4 . Solution for D' (derivative of D) obtained in the second test with BC on d such that D' must become zero at the end points (see section 4.3). The condition is met with good accuracy (values below 10^{-11}). The top figure is for large horizontal scale, while the bottom figure displays the solution for the shortest scale considered in this work, 100m.

ANNEX-A Basic Green Functions Theory

This brief presentation follows [8]. The inverse of a linear differential operator L is an integral operator whose kernel $g(\eta, z)$ satisfies the differential equation

$$L_{\eta} g(\eta, z) = \delta(\eta - z) \quad \text{Eq A1}$$

where $\delta(\eta-z)$ is a distribution known as “delta”. That is, we seek a function $g(\eta, z)$ for which the “generalized operation” Lg satisfies:

$$\langle Lg, \phi \rangle = \phi(z) \quad \text{Eq A2}$$

for any Φ in the space of functions under consideration S and $\langle \cdot, \cdot \rangle$ denotes its interior product. The differential equation must be completed with homogeneous BC that define the domain of L , $D(L) = \{ f \mid L[f] \text{ is in } S \text{ and } BC[f]=0 \}$. That is g , as a function of η , belongs to $D(L)$. The form of BC depends of the form of L . For a n^{th} order operator on a compact domain $[a, b]$, BC can be a linear combination of the $0, 1, \dots, n-1$ derivatives at each end of the interval. BC can also impose periodicity.

For the adjoint of L , L^+ , we have also a $g^+(\eta, z)$ such that

$$L_\eta^+ g^+(\eta, z) = \delta(\eta - z) \quad \text{Eq A3}$$

and g^+ , as function of η (i.e. its first argument), is in $D(L^+) = \{ f \mid L^+[f] \text{ in } S \text{ and } BC^+[f]=0 \}$ where BC^+ denotes the adjoint BC (i.e. the BC that makes the boundary term after integration by parts vanish). It is not difficult to see that

$$g^+(\eta, z) = g(z, \eta) \quad \text{Eq A4}$$

from which we infer that g , as a function of z , is in $D(L^+)$ and g^+ , as a function of z , is in $D(L)$ and, of course, a self-adjoint operator $L=L^+$ and $BC=BC^+$ has symmetric GF. With these results, it is not difficult to see that the differential equation with homogeneous BC (i.e. $BC[u]=0$)

$$L_\eta[u(\eta)] = f(\eta) \quad \text{Eq A5}$$

has the following solution

$$u(z) = \langle \delta(\eta - z), u(\eta) \rangle = \langle L^+ g^+, u \rangle = \langle g^+, Lu \rangle = \int_a^b g^+(\eta, z) f(\eta) d\eta = \int_a^b g(z, \eta) f(\eta) d\eta \quad \text{Eq A6}$$

Construction of the GF

The following rules provide a formal way to find the GF for an n th order linear differential operator

$$L_x = a_n(x) \frac{d^n}{dx^n} + \dots + a_0(x)$$

- 1) $L_\eta g(\eta, z) = 0$ for $\eta \neq z$
- 2) $\frac{d^k}{d\eta^k} g(\eta, z)$ continuous at $\eta = z$ for $k = 0, \dots, n-2$
- 3) $\frac{d^{n-1}}{d\eta^{n-1}} g(\eta, z) \Big|_{\eta=z^-}^{\eta=z^+} = \frac{1}{a_n(z)}$ "jump condition"

In addition we require $g(\eta, z)$ to satisfy all appropriate homogeneous BC.

For a second order differential operator: $L = a_2(x) \frac{d^2}{dx^2} + a_1(x) \frac{d}{dx} + a_0(x)$

$$g(z, \eta) = \begin{cases} z < \eta & u_1(z) u_2(\eta) A; \quad Lu_1 = 0; \quad \alpha_{left} u_1(a) + \beta_{left} u_1'(a) \equiv B_L u_1 = 0 \\ \eta < z & u_1(\eta) u_2(z) A; \quad Lu_2 = 0; \quad \alpha_{right} u_2(b) + \beta_{right} u_2'(b) \equiv B_R u_2 = 0 \end{cases}$$

and "A" is determined from the "jump condition":

$$A = \frac{1}{a_2(\eta) [u_1(\eta) u_2'(\eta) - u_1'(\eta) u_2(\eta)]} = \frac{1}{a_2(\eta) W(\eta)} \quad \text{Eq A7}$$

Extension to problems with non-homogeneous BC

In order to include problems where the solution satisfies $BC[u]=Bu \neq 0$ we have to add into the analysis the boundary terms that result in an integration by parts whenever we swap from L to the adjoint or viceversa. To fix ideas let us consider only the operators of interest in this application: L ($a_2=a_1=1$, $a_0=-\lambda$), and L^+ ($a_2=1$, $a_1=-1$, $a_0=-\lambda$). Then, we have:

$$\langle Lu, v \rangle = \left| u'v - uv' + uv \right|_a^b + \langle u, L^+v \rangle = J_L(u, v) + \langle u, L^+v \rangle \quad Eq A8$$

Eq A6 now becomes

$$u(z) = \langle g(z, \eta), Lu(\eta) \rangle - J_L(u(\eta), g(z, \eta)) = \int_a^b g(z, \eta) f(\eta) + \left| -u'(\eta)g(z, \eta) + u(\eta)g'_\eta(z, \eta) - u(\eta)g(z, \eta) \right|_{\eta=a}^{\eta=b}$$

and we notice that the non-continuous character of g' does not cause trouble because only “one-side” derivatives enter in the J term. To simplify this term we note that $A'(\eta)=A(\eta)$ (this is a direct consequence of the equation for the Wronskian $W' = -W$) and then we obtain for $g'_\eta(z, \eta)$

$$g'_\eta(z, \eta) = \begin{cases} z < \eta & u_1(z) u'_2(\eta) A(\eta) + g(z, \eta) \\ \eta < z & u'_1(\eta) u_2(z) A(\eta) + g(z, \eta) \end{cases}$$

and the J term becomes:

$$\left(u(b) u'_2(b) - u'(b) u_2(b) \right) A(b) u_1(z) - \left(u(a) u'_1(a) - u'(a) u_1(a) \right) A(a) u_2(z)$$

which after some manipulations can be put in the form of our final result:

$$u(z) = \int_a^b g(z, \eta) f(\eta) d\eta + \frac{B_R u}{B_R u_1} u_1(z) + \frac{B_L u}{B_L u_2} u_2(z)$$

$$u'(z) = \int_a^b g'_\eta(z, \eta) f(\eta) d\eta + \frac{B_R u}{B_R u_1} u'_1(z) + \frac{B_L u}{B_L u_2} u'_2(z)$$

These expressions show that the solution to the problem depends linearly on the BC: B_{Ru} and $B_L u$.

ANNEX-B Spectrum of the L Operator

As mentioned in section 4.1, the zeros of $W(u_1, u_2)$ give the conditions when two independent solutions of $(L-\lambda)u=0$ cannot be found and the construction of the GF fails. We have just found one eigenfunction of the operator L and its corresponding eigenvalue λ . For this λ , $(L-\lambda)^{-1}$ does not exist.

For the problem chosen to illustrate the method, ($u(0)=u(\bar{\xi})=0$) we readily see that these values are given by:

$$e^{\Delta \bar{\xi}} = 1 \quad \Delta_n = \sqrt{1+4\lambda_n} = \frac{2\pi n}{\bar{\xi}} i \quad ; \quad \lambda_n = \frac{-1}{4} \left(\left(\frac{2\pi n}{\bar{\xi}} \right)^2 + 1 \right)$$

the spectrum is discrete and consists of negative values. The eigenfunctions are (up to normalization)

$$u_n(\xi) = e^{\frac{-\xi}{2}} \sin\left(\frac{n\pi}{\bar{\xi}} \xi\right)$$

which are not orthogonal, i.e. $\langle u_n, u_m \rangle \neq 0$. However, we see that they are orthogonal with the inner product

$$\int_0^{\bar{\xi}} d\xi e^{\xi}$$

or, alternatively, we can make L self-adjoint with the change of variable $u \rightarrow u \exp(\xi/2)$. However, it will be an error to think that orthogonality can always be achieved. For example, for the problem $u(0)=u'(\bar{\xi})=0$ the spectrum is given by the roots of the transcendental equation:

$$\tan\left(\hat{\Delta}_n \frac{\bar{\xi}}{2}\right) = \hat{\Delta}_n \quad \hat{\Delta}_n = \sqrt{4|\hat{\lambda}_n| - 1} \quad ; \quad \hat{\lambda}_n < 0 \quad ; \quad |\hat{\lambda}_n| \geq |\lambda_n| \geq \frac{1}{4} \text{ for all } n$$

with eigenfunctions

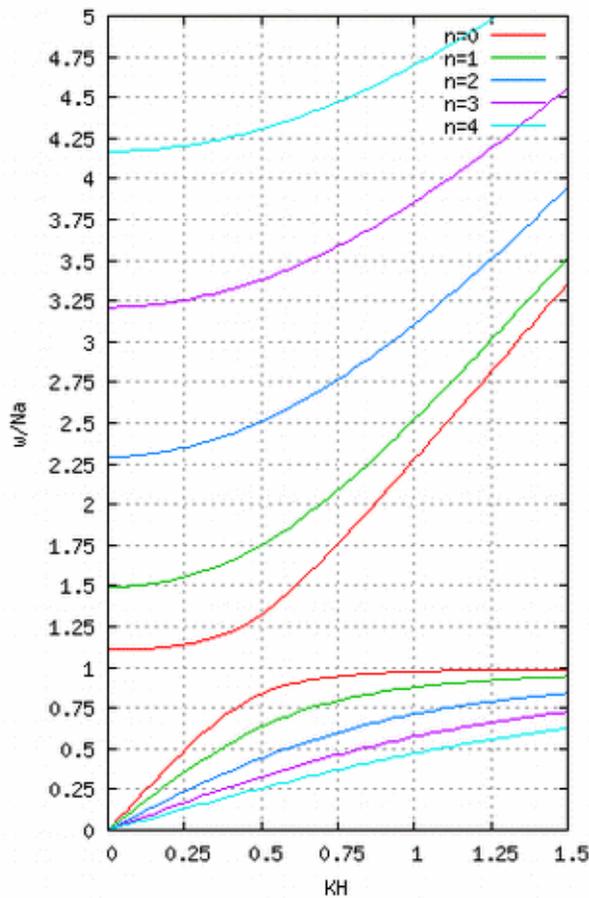
$$u_n(\xi) = e^{\frac{-\xi}{2}} \sin\left(\frac{\hat{\Delta}_n}{2} \xi\right)$$

and these functions are not orthogonal for any inner product. The reason for this can be seen from the fact that the substitution does not work for the BC on the derivative.

The dispersion relations for the simple case $u(0)=u(\bar{\xi})=0$ can be obtained by going back to the continuous in time structure equation [2, eq 3.33] and replacing L by their λ_n , that is:

$$\frac{1}{4} \left[\left(\frac{2\pi n}{\bar{\xi}} \right)^2 + 1 \right] = -\lambda_n = (kH)^2 \left[\frac{\left(\frac{N}{N_a} \right)^2}{\left(\frac{\omega}{N_a} \right)^2} + \frac{\left(\frac{\omega}{N_a} \right)^2}{(kH)^2} - 1 \right]$$

where $N_a^2 = N^2 c_p/c_v$ is the ‘‘acoustic cut off frequency’’. This is a quadratic equation in ω^2 which displays two branches, the acoustic and the gravitational branch. The dispersion relations for an isothermal atmosphere $T=273.16$ and depth from 1hPa to 1013hPa ($\bar{\xi}=6.92$, $H=8\text{Km}$) are shown below. The number n indicates the vertical mode number ‘‘n’’.



ANNEX C The SI System in the NH-GEOGW Model

The SI system for this choice of variables can be written as follows ([6], page 20)

$$\begin{aligned}
 D - (kH)^2 (N\Delta t)^2 (\pi_s + T + (\partial+1)\Phi) &= D^\bullet & ; & \quad D = D' \Delta t \\
 T + \left(\frac{R}{c_v}\right) \left(D - \left(\frac{T^*}{T_a}\right) \partial(gw) \right) &= T^\bullet & ; & \quad T = \frac{T'}{T^*} \\
 \Phi - gw - (N^* [D] - S^* [D]) &= \Theta^\bullet & ; & \quad \Phi = \frac{\Phi'}{RT^*} \quad C1 \\
 \pi_s + N^* [D] &= \pi_s^\bullet & ; & \quad \pi_s = \frac{\pi_s'}{\pi_s^*} \\
 gw + (N\Delta t)^2 \left(\frac{T^*}{T_a}\right) (-(\partial+1)T - (\partial+1)\partial\Phi) &= (gw)^\bullet & ; & \quad gw = \frac{gw'}{RT^*} \Delta t
 \end{aligned}$$

and now there are two scale temperatures T^* and T_a^* . Let us denote the ratio between these two temperatures “z”, and let us define the three operators:

$$\hat{O}_1 = \left(\frac{R}{c_v} z + 1 \right) \partial + 1 \quad ; \quad O_1 = \frac{c_p}{c_v} \partial + \frac{R}{c_v} \quad ; \quad O_2 = \left(\frac{R}{c_v} z + 1 \right) (\partial^2 + \partial) \equiv A(z) L^*$$

Eliminating variables in favour of D and gw, we get to two second order differential equations, with the same differential operator for both equations:

$$(N\Delta t)^2 z O_1 \hat{O}_1 + \left(1 - (N\Delta t)^2 z O_2 \right) \left(\frac{1}{k^2} + \frac{c_p}{c_v} \right) = A(z) z \left[\lambda - (\partial^2 + (1 - (\frac{c_p}{A(z)c_v} - 1) k^2) \partial) \right]$$

$$\text{where } k^2 = (kH)^2 (N\Delta t)^2 \quad ; \quad \lambda = \frac{\left(1 + k^2 \frac{c_p}{c_v} \left(1 + (N\Delta t)^2 \frac{R}{c_p} z \right) \right)}{\left((N\Delta t)^2 A(z) z \right)}$$

The consideration of values $z \neq 1$ would have a deleterious effect on the numerical scheme presented in this paper. Note that for big values of k^2 , the differential operator becomes nearly 1st order. The two roots are very different, one very close to zero and the other one very big in absolute value. This operator reduces to the same operator as in the PDVD case if we set $z=1$, i.e. $T^* = T^*_a$. For this value of z the two second order DE read:

$$\left(\frac{c_p}{c_v} \right) (-\lambda + L^*) [D] = \frac{-k^2}{(N\Delta t)^2} \left(1 + \frac{c_p}{c_v} \partial \right) [gw^\bullet] + \left(\frac{c_p}{c_v} L^* - \frac{1}{(N\Delta t)^2} \right) [D^\bullet] \quad C2$$

$$- k^2 (\partial + 2) [T^\bullet] + k^2 \left(\frac{R}{c_v} L^* - \frac{1}{(N\Delta t)^2} (\partial + 1) \right) [\Theta^\bullet] - \frac{k^2}{(N\Delta t)^2} \pi_s^\bullet$$

and for the vertical velocity

$$\left(\frac{c_p}{c_v} \right) (-\lambda + L^*) [gw] = \frac{-(1 + k^2 \frac{c_p}{c_v})}{(N\Delta t)^2} [gw^\bullet] + \left(\frac{c_p}{c_v} \right) (\partial + 1) - 1 [D^\bullet] \quad C3$$

$$- (\partial + 1 + k^2) [T^\bullet] - \left(L^* - \frac{R}{c_v} k^2 (\partial + 1) \right) [\Theta^\bullet] + k^2 \left(\frac{R}{c_v} \right) \pi_s^\bullet$$

We notice also that the “consistency problem” mentioned at the end of section 4.4, manifests here also in EqC2 already at the level of linear terms, due to the presence of $\partial^3 \Phi$ terms. As it stands, the scheme would allow the calculation of the following table of values at each knot:

gw, gw', gw''
D, D', S[D]
Φ, Φ', Φ''
T, T', π _s

and some sort of approximation on $\partial^3 \Phi$ is required to include D'' and T'' in the set.

References

- [1] Bubnová R., Hello G, Bénard P, Geleyn JF. 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. Weather Rev.* **123**, 515:535
- [2] Bénard P, Vivoda J, Mašek J, Smolíková P, Yessad K, Smith Ch, Brožková R and 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 Part A
- [3] Bénard P and Mašek J, 2011. "Scientific Documentation for ALADIN –NH Dynamical Kernel". Available from <http://www.cnrm.meteo.fr/gmapdoc>
- [4] 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
- [5] Smolíková P et al, 2015 "LACE-news in Dynamics" Presentation at the 25th ALADIN Workshop & HIRLAM ASM, available from <http://www.cnrm.meteo.fr/aladin/>
Smolíková P et al, 2014 "LACE Dynamics" Presentation at the 24th ALADIN Workshop & HIRLAM ASM, available from <http://www.cnrm.meteo.fr/aladin/>
Vivoda J et al, 2013 "Vertical FE in the NH Dynamical Core of ALADIN" Presentation at the 23rd ALADIN Workshop & HIRLAM ASM, available from <http://www.cnrm.meteo.fr/aladin/>
- [6] Yessad K, 2014 "Semi-Implicit Spectral Computations and Predictor-Corrector Schemes in the Cycle 41 of ARPEGE/IFS". Available from <http://www.cnrm.meteo.fr/gmapdoc>
- [7] Ralston A. and Rabinowitz P. "A First Course in Numerical Analysis. Second Edition". Dover Publications, 1978
- [8] Keener J.P "Principles of Applied Mathematics. Transformation and Approximation" Addison-Wesley, 1995

Recently or soon published

Patricia Pottier

In the “references” chapter of the previous articles, you may find links to interesting publications. Below are a few additional references of publications either recently published, or on early view or in press :

- Auger, L., Dupont, O., Hagelin, S., Brousseau, P. and Brovelli, P., *AROME–NWC: a new nowcasting tool based on an operational mesoscale forecasting system*. Q.J.R. Meteorol. Soc.. doi: 10.1002/qj.2463
- Berre L., H. Varella et G. Desroziers: *Modelling of flow-dependent ensemble-based background error correlations using a wavelet formulation in 4D-Var at Meteo-France*. Quart. Jour. Roy. Meteor. Soc., DOI: 10.1002/qj.2565
- Besson F., E. Bazile, C. Soci, J.-M. Soubeyroux, G. Ouzeau and M. Perrin, 2015: *Diurnal temperature cycle deduced from extreme daily temperatures and impact over a surface reanalysis system*, Adv. Sci. Res., 12, 137-140. doi:10.5194/asr-12-137-2015.
- Bouttier F., L. Raynaud, O. Nuissier and B. Ménétrier: *Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX*. Quart. Jour. Roy. Meteor. Soc., DOI: 10.1002/qj.2622
- Björkqvist, J.-V., Tuomi, L., Tikka, K., Fortelius, C., Pettersson, H., Kahma, K.K., 2015: *Modelling waves in complex fetch and bathymetric conditions in the northern Gulf of Finland*. Submitted to Coastal Engineering.
- Campins, J., and Navascues, B., 2015: *Impact of targeted observations on HIRLAM forecasts during HyMeX-SOP1*, submitted to QJRMS.
- Caluwaerts, D. Degrauwe, P. Termonia, F. Voitus, P. B´enard and J-F. Geleyn, 2015: *Importance of temporal symmetry in spatial discretization for geostrophic adjustment in semi-implicit Zgrid schemes*. Q. J. R. Meteorol. Soc., 141 128138. doi: 10.1002/qj.2344.
- Caluwaerts, S., D. Degrauwe, F. Voitus, P. Termonia, 2015: *Discretization in Numerical Weather Prediction: a modular scheme to investigate spectral and local SISL methods*. To appear in Springer Special Issue “Mathematical Problems in Meteorological Modelling”.
- De Meutter, P., L. Gerard, G. Smet, K. Hamid, R. Hamdi, D. Degrauwe, P. Termonia, 2015: *Predicting small-scale, short-lived downbursts: case study with the NWP limited-area ALARO model for the Pukkelpop thunderstorm*, Mon. Wea. Rev., 143, 742-756. doi:10.1175/MWR-D-14-00290.1.
- Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P. and Cébron, P., *PEARP, the Météo-France short-range ensemble prediction system*. Q.J.R. Meteorol. Soc.. doi: 10.1002/qj.2469.
- Duerinckx, A., R. Hamdi, J.-F. Mahfouf, P. Termonia, 2015: *Study of the Jacobian of an Extended Kalman Filter for soil analysis in SURFEXv5*, Geosci. Model Dev., in press.
- Fourrié N., E. Bresson, M. Nuret, C. Jany, P. Brousseau, A. Doerenbecher, M. Kreitz, O. Nuissier, E. Sevault, H. Bénichou, M. Amodei and F. Pouponneau, 2015 : *AROME-WMED, a real-time mesoscale model designed for the HyMeX Special Observation Periods*. Geoscientific Model Development Discussion, 8, 1801-1856, doi:10.5194/gmdd-8-1801-2015.
- Gerard, L., 2015: *Bulk mass-flux perturbation formulation for a unified approach of deep convection at high resolution*. Mon. Wea. Rev., in press.

- Giot, O., P. Termonia, D. Degrauwe, R. De Troch, S. Caluwaerts, G. Smet, J. Berckmans, A. Deckmyn, L. De Cruz, P. De Meutter, A. Duerinkcx, L. Gerard, R. Hamdi, J. Van den Bergh, M. Van Ginderachter, and B. Van Schaeybroeck: *Validation of the ALARO-0 model within the EURO-CORDEX framework*, submitted to Geoscientific Model Development.
- Karsisto, P., Fortelius, C., Demuzere, M., Grimmond, C., Oleson, K., Kouznetsov, R., Masson, V., Järvi, L., 2015: *Surface urban energy balance and wintertime stability simulated using three land-surface models in the high-latitude city Helsinki*. Submitted to QJRMS.
- Legrand R., Y. Michel, and T. Montmerle, 2015: *Diagnosing non-Gaussianity of forecast and analysis errors in a convective scale model*. Processes Geophys. Discuss., 2, 1061-1090, 2015 doi:10.5194/npgd-2-1061-2015
- Lindstedt, D., Lind, P., Kjellstrom, E., Jones, C., 2015: *A new regional climate model operating at the meso-gamma scale; performance over Europe*, Tellus-A Vol.67 24138, doi:10.3402/tellusa.v67.24138
- Mahfouf, J.-F., F. Ahmed, P. Moll, and F. N. Teferle, 2015: *Assimilation of zenith total delays in the AROME France convective scale model: a recent assessment*. Tellus A, 67, 26106, <http://dx.doi.org/10.3402/tellusa.v67.26106>
- Marquet, P., 2015: *On the computation of moist-air specific thermal enthalpy*. Q.J.R. Meteorol. Soc., 141: 67–84. doi: 10.1002/qj.2335
- Masek J., Geleyn J.-F., Brozkova, R., Giot O., Achom H. O. and Kuma P., 2015: *Single interval shortwave radiation scheme with parameterized optical saturation and spectral overlaps*. QJRMS, in press.
- Ménérier B., T. Montmerle, Y. Michel, and L. Berre, 2015: *Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part II : Application to a Convective-Scale NWP Model*. Monthly Weather Review, 143, No. 5, 1644-1664.
- Rontu, L., Niemela, S., and Wastl, C., 2015: *Influence of the details of topography on weather forecast – evaluation of HARMONIE experiments in the Sochi Olympics domain over the Caucasian mountains*, subm. to Frontiers.
- Sanchez-Ariola, J., Lindskog, M., Thorsteinsson, S., Bojarova, J., 2015: *Variational bias correction of GNSS ZTD in the HARMONIE modeling system*, subm. to JAMC.
- Simarro, J.P., 2015: *Un modelo no hidrostático global con coordenada vertical basada en altura (A global non-hydrostatic model with vertical coordinate based on height)*, PhD Thesis at Valencia University, obtained the highest degree 'cum laude', June 2015

Comment on:

“Impact of changes in the formulation of cloud-related processes on model biases and climate feedbacks”
by Carlo Lacagnina, Frank Selten and A. Pier Siebesma
(J. Adv. Mod. Earth Syst. 2014; 6(4): 1224-1243)

Kristian Pagh Nielsen & Emily Gleeson

1 Introduction

The following comment has previously been submitted to the AGU journal: “*Journal of Advanced in Modeling Earth Systems*.” The chief editor Robert Pincus and an anonymous associate editor rejected it for publication and replied:

“Your central point, that the CIF used in the Lacagnina et al. paper is unrealistically low, is entirely valid. But the unphysical nature of the CIF has been known since the 1990s and pointing out one dubious choice in one model doesn't do much to advance the field or to interpret the previous work. You may suspect that this "bias correction" impacts the results reported by Lacagnina et al.; if so I would encourage you to submit a manuscript in which you make this point and support it with evidence.”

This is all reasonable, but we have no plans for running tests with the EC-EARTH model used by Lacagnina et al. and nevertheless think that our comment is relevant for the scientific community. Therefore, we publish it here.

2 Comment

Recently, we published a paper (Nielsen et al., 2014) on detailed testing of the shortwave (SW) radiation physics in the HARMONIE cycle 37h numerical weather prediction (NWP) model that included testing of the cloud optical property and radiative transfer parametrizations. The default radiation scheme in this version of HARMONIE is based on the Integrated Forecast System (IFS) cycle 25r1 (ECMWF, 2002; Mascart and Bougeault, 2011). This is the same radiation scheme that was used in the version of the EC-Earth model runs included in the recent study by Lacagnina et al. (2014).

Using HARMONIE we showed that the default SW cloud inhomogeneity factor (CIF) of 0.7 has a large impact on the cloud radiative forcing. The impact as a function of the cloud water load is illustrated in Fig. 1 for a solar zenith angle (SZA) of 56° and standard atmospheric conditions as described by Nielsen et al. (2014). Liquid cloud droplets with effective radii of 10 μm are assumed.

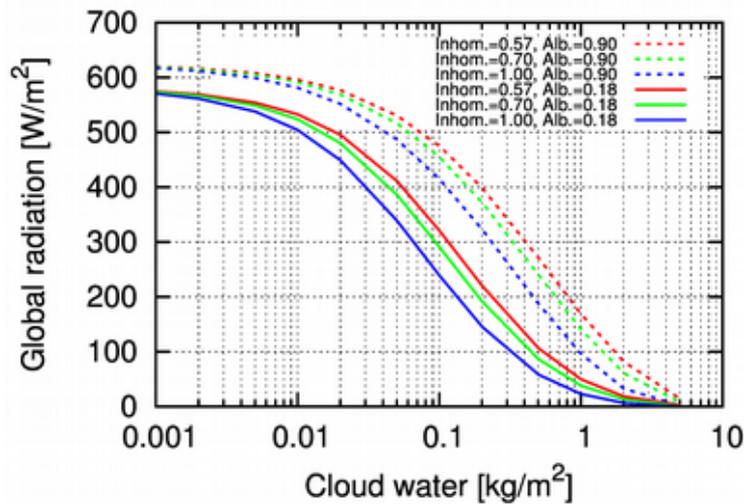


Figure 1. Global shortwave radiation at the surface as a function of cloud water load for three different cloud inhomogeneity factor settings: 0.57 (red lines), 0.7 (green lines) and 1.0 (blue lines), and for two different surface albedos: 0.18 (solid lines) and 0.9 (dashed lines).

Lacagnina et al. (2014) state that they reduced the CIF further in order to achieve radiative balance at the top of the atmosphere. They do not give the actual value of the factor. From the EC-Earth source code, we find this to be 0.57. The effect of having a SW CIF that is further reduced from 0.7 to 0.57 is also illustrated in Fig. 1. The effect is almost a 100 W/m^2 increase in the SW irradiance transmitted through clouds with $\sim 0.1 \text{ kg/m}^2$ cloud water compared to the transmittance through homogeneous clouds. This large effect is not surprising since the physical implication of using the CIF is to factorize the entire cloud water load before performing the radiation calculations. In the simulations of Lacagnina et al. (2014) 43% of all cloud water has thus been removed before the radiation calculations. Certainly this is important for the cloud feedbacks studied in the paper, and it is important for future investigators to be aware of this when they read the paper. The values of CIF for both SW and LW irradiances ought to have been specified in the paper, which is why we write this comment. Assumptions about CIF in general are important for understanding differences in irradiance fluxes from different atmospheric models, as described for instance by Oreopoulos et al. (2012).

Cloud inhomogeneity is an issue that can be physically quantified (Oreopoulos and Davies, 1998; Oreopoulos and Cahalan, 2005; Shonk et al., 2010, 2012) and thus ought not be used as a tuning factor to achieve radiative balance. As noted by Lacagnina et al. (2014) the next versions of EC-EARTH will not use CIF. In stead the McICA parametrization (Pincus et al., 2003) will be used.

3 References

- References ECMWF (2002), IFS documentation CY25r1, <http://old.ecmwf.int/research/ifsdocs/-CY25r1/index.html>, Accessed: 2015-03-16.
- Lacagnina, C., F. Selten, and A. P. Siebesma (2014), Impact of changes in the formulation of cloud-related processes on model biases and climate feedbacks, *J. Adv. Mod. Earth Syst.*, 6 (4), 1224–1243, doi:10.1002/2014MS000341.
- Mascart, P. J., and P. Bougeault (2011), The Meso-NH atmospheric simulation system: Scientific documentation, Tech. rep., Météo France, Toulouse, France.
- Nielsen, K. P., E. Gleeson, and L. Rontu (2014), Radiation sensitivity tests of the harmonie 37h1 nwp model, *Geosci. Model Dev.*, 7, 1433–1449, doi:10.5194/gmd-7-1433-2014.
- Oreopoulos, L., and R. F. Cahalan (2005), Cloud inhomogeneity from MODIS, *J. Climate*, 18, 51105124, doi:10.1175/JCLI3591.1.
- Oreopoulos, L., and R. Davies (1998), Plane parallel albedo biases from satellite observations. Part II: Parameterizations for bias removal, *J. Climate*, 11, 933–944.
- Oreopoulos, L., E. Mlawer, J. Delamere, T. Shippert, J. Cole, B. Fomin, M. Iacono, Z. Jin, J. Li, J. Manners, P. Räsänen, F. Rose, Y. Zhang, M. J. Wilson, and W. B. Rossow (2012), The continual intercomparison of radiation codes: Results from phase I, *J. Geophys. Res.*, 117, D06,118, doi:10.1029/2011JD016821.
- Pincus, R., H. W. Barker, and J.-J. Morcrette (2003), A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, *J. Geophys. Res.*, 108 (D13), 4376, doi:10.1029/2002JD003322.
- Shonk, J. K. P., R. J. Hogan, J. M. Edwards, and G. G. Mace (2010), Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part I: Review and parametrization, *Q. J. R. Meteorol. Soc.*, 136, 1191–1204, doi:10.1002/qj.647.
- Shonk, J. K. P., R. J. Hogan, and J. Manners (2012), Impact of improved representation of horizontal and vertical cloud structure in a climate model, *Clim. Dyn.*, 38, 2365–2376, doi:10.1007/s00382-011-1174-2.