LACE stay, GeoSphere Austria

24. 6. – 19. 7. 2024.

Verification of flow-dependent SPP and domain extension for C-LAEF

Author:

Endi Keresturi
endi.keresturi@cirus.dhz.hr

Supervisor:

Clemens Wastl

Preface

This work is a continuation of the previous two LACE stays - Keresturi (2022, 2023) on the topic of flow-dependent SPP (FD-SPP). FD-SPP is thoroughly explained in the previous two reports. Here, it will be assumed that the reader is already familiar with it.

Motivation

Keresturi (2022, 2023) implemented and assessed the general behavior of FD-SPP and its impact on the C-LAEF (ensemble system of GeoSphere Austria) behavior, which was found to be beneficial, especially for ensemble spread. However, no long-term verification was done which is necessary to make a final judgement about the method. In this work, we will analyze the results of two one-month long C-LAEF experiments. Also, FD-SPP code needed to be modified in order to be in accordance with the ARPEGE/IFS coding norms and optimizations.

Additionally, since Croatia aims to join the new CLAEF1k project (joint cooperation of several LACE countries on a common 1km ensemble system), the current C-LAEF domain needs to be extended to accommodate the interests of Croatia. Here, we will explore the possible domain extension further to the south.

Finally, as the new domain will increase the SBU consumption significantly, the lagging approach is being considered for C-LAEF. Such a lagging approach is used by several national weather services to increase their ensemble sizes but on the other side it is known to cause ensemble member clustering with respect to forecast quality (Buizza, 2008) which can have various negative impacts on ensemble from theoretical point of view. Here, we will take a look at one model run and try to visually observe clustering in C-LAEF lagged runs.

Part 1 – Changes in the code

Due to time constraints, the original FD-SPP implementation in CLAEF1k cy46t1 was done without paying attention to ARPEGE/IFS coding norms (link) or recommended Fortran optimizations (El Khatib, 2019). In the current LACE stay, the code has been revisited and appropriate corrections were applied. Most of the corrections concerned the changes in variable names, usage of array syntax (:), variable allocations, loops structure and removal of unnecessary commented lines.

Part 2 – Long-term verification

Verification methodology

The original FD-SPP was implemented to CLAEF1k cy46t1, but the operational C-LAEF is on 2.5 km and cy43t1. In order save billing units (SBUs) and to have a fair comparison with the standard SPP implementation in the operational C-LAEF system on 2.5km, FD-SPP was phased back to cy43t1 for the long-term verification of the new perturbation scheme. We expect that the impact FD-SPP has at 1 km, doesn't significantly change at 2.5 km.

Our goal is to assess the added value of FD-SPP perturbations in C-LAEF. For this reason, we define two experiments: a) CLAEF_oper – operational C-LAEF configuration on 2.5 km and cy43t1 using regular SPP and b) CLAEF_FD – configured the same as CLAEF_oper except that FD-SPP is used instead of standard SPP.

Verification will be performed separately for February 2024 and June 2024 and separately for surface and upper air variables. Variables used in verification are the following: Temperature (T), wind speed (WS), wind gusts (WG), total cloudiness (TC), relative humidity (RH), dewpoint temperature (Td), geopotential height (Z) and mean sea level pressure (MSLP). Standard scores for probabilistic forecasts are used – RMSE of ensemble mean, spread, bias, CRPS, spread/skill ratio, ROC and Brier score.

Different domains/verification packages were used for verification because of the following reasons. C-LAEF operational suite uses HARP as a verification package and hence all observations and CLAEF_oper files were already prepared. However, with the HARP package used at GeoSphere Austria, it is currently only possible to verify over Austria (252 stations) for both surface and upper air (4 stations). For this reason, I have, additionally, used my own verification package to verify against Croatian stations (44 stations; surface) and on the whole domain for upper air (25 stations). All of this complicated verification process took some time to complete because of the following reasons: First, as I have never used HARP, I needed to install it under my user and get myself familiar with it. Second, the HARP scripts were taken from HARP training (June 2024), and some of them contained bugs which had to be identified and corrected. Third, my own verification package can handle both surface and upper air verification (over the whole domain), but it couldn't yet fully handle ensembles. Because of this, I needed to finish the implementation of the ensemble support.

It is worth noting that the differences between the experiments are expected to be small because we are only slightly modifying the model perturbations which are known to have small impact on overall verification scores. The biggest impact is expected to be on the ensemble spread.

Results

a) Over Austria

Results for February are generally positive. Ensemble spread is slightly increased for all variables and all lead times (Figure 1). Impact on RMSE is more neutral and slightly positive for some variables and some lead times (Figure 1). This is what we had hoped for – increase spread without deteriorating RMSE. Impact on CRPS is also positive (Figure 2).

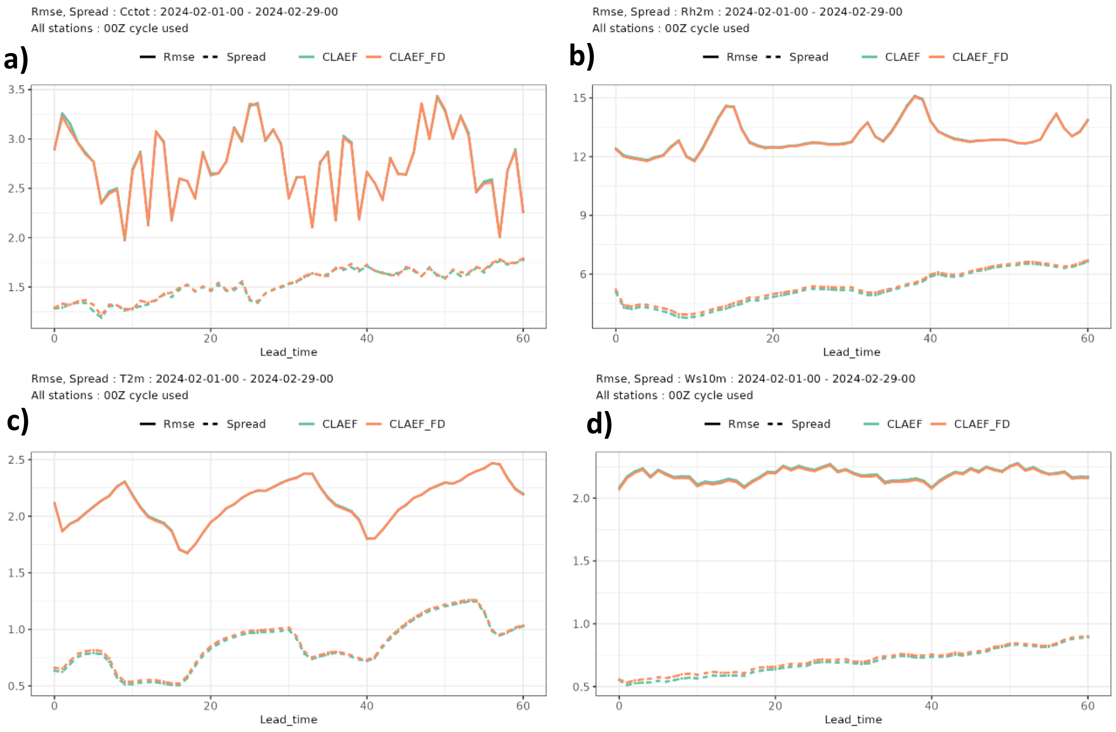

Figure 1. RMSE (solid) and spread (dashed) for CLAEF_oper (blue) and CLAEF_FD (orange) for February 2024 averaged over Austrian stations. TC is shown in a), RH in b), T in c) and WS in d).

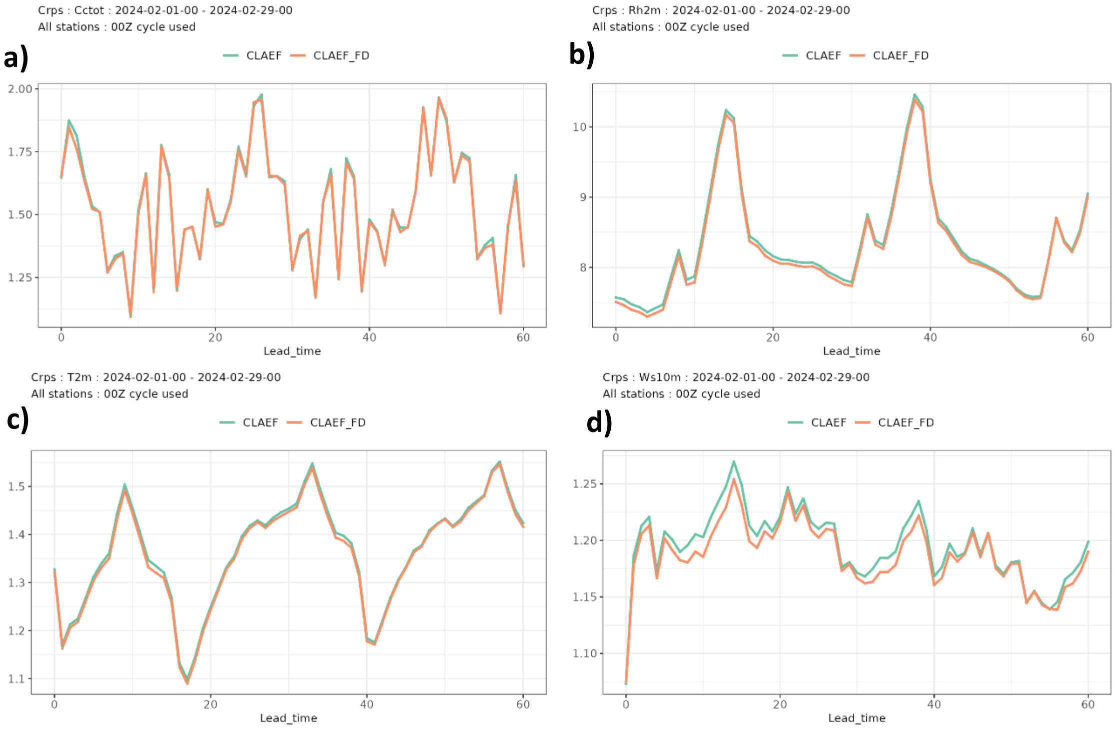

Figure 2. As in Figure 1, but for CRPS.

ROC area is also slightly increased for all variables and majority of thresholds (not shown). Bias is reduced for total TC and WS and neutral for T and RH (not shown).

Results for June are unexpectedly more neutral. Increase in spread is only visible for TC, while it is neutral for other variables (Figure 3). RMSE is decreased for WS and is neutral for other variables (Figure 3). CRPS is decreased for WS and slightly for TC (Figure 4). ROC is increased mostly only for WS (not shown). Bias is improved for WS and neutral for other variables (not shown).

The reason for this is unclear especially because, during summer, model physics is more active. In Frogner *et. al* (2022), it is shown that the parameters that they use in SPP are much more active during the summer. Our hypothesis is that SPP in C-LAEF is not properly tuned for the convective season, and we have less impact during the summer. In order to confirm this, an additional experiment has been run – without using SPP at all. This will allow us to assess the impact of standard SPP during the summer compared to winter, similar to Keresturi (2023). Unfortunately, it was possible to run only one day in the winter and summer period, respectively, because the ECMWF coupling files have not been stored. Figures A1-A2 in Appendix confirm our hypothesis, SPP in C-LAEF is significantly less (2-3 times) active during the summer for surface variables. Comparing Figures 3-4 with Figures 11-12 in Keresturi (2023) confirms that conclusion. For vertical levels, differences were much smaller.

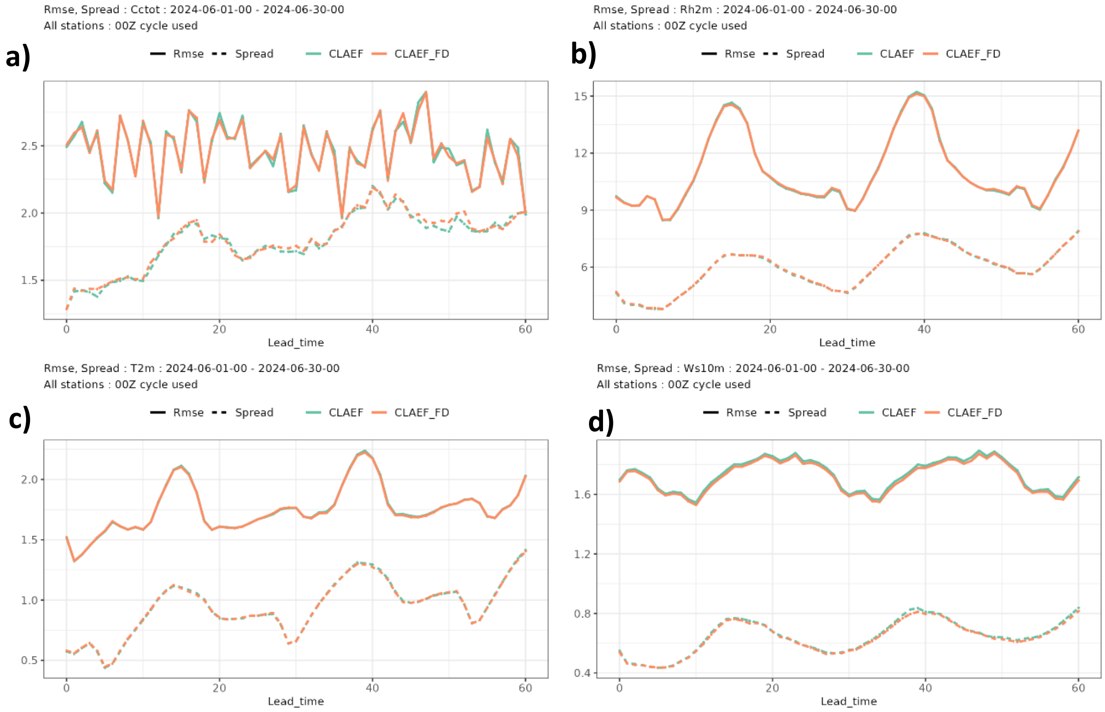

Figure 3. As Figure 1, but for June 2024.

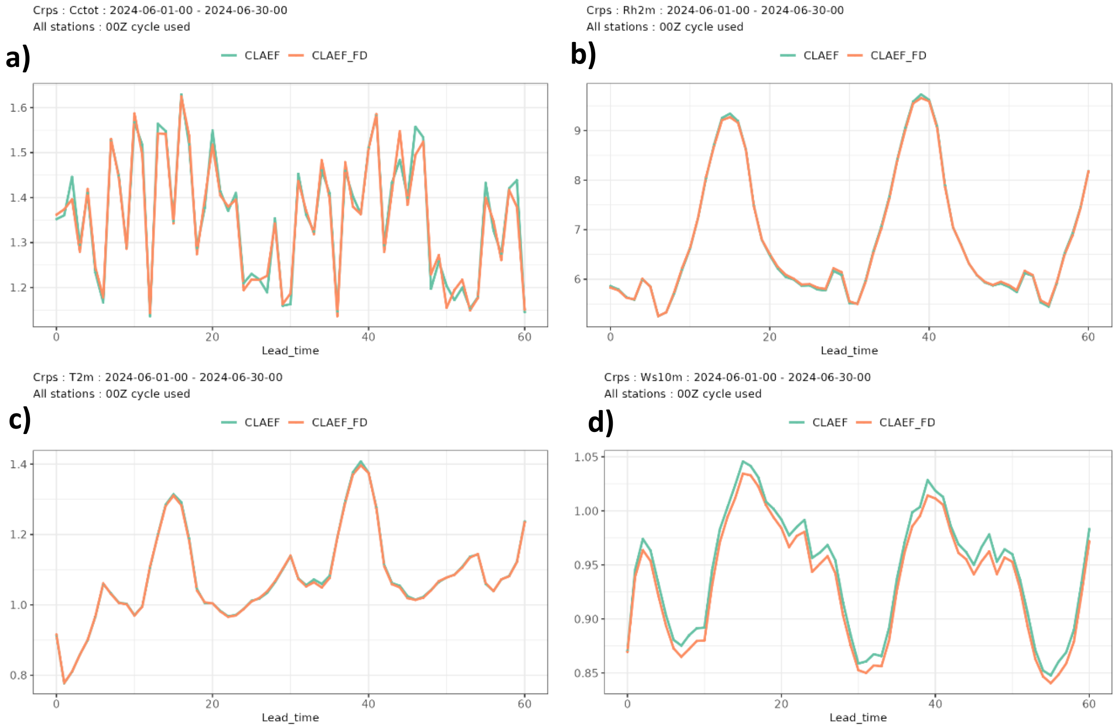

Figure 4. As Figure 2, but for June 2024.

b) Over Croatia

Results for February are more neutral but agree with the ones from Austria. Additionally, verification over Croatia was extended with MSLP and WG. Spread was increased for WG, while the results are neutral for MSLP (not shown). Results for June also agree with the ones from Austria, and results for WG and MSLP are neutral (not shown).

c) Upper air over the whole domain

Differences in upper air variables should be even smaller since the model physics is more active near the ground. Variables used here include T, Td, WS, Z and RH and vertical levels used are (in hPa): 925, 850, 700, 500 and 300. As there is a high number of possible combinations of variable/levels, we will only show plots for which the impact is the greatest. Results indicate that there are practically no differences for levels above 850 hPa and we will leave them out of the rest of this analysis. At 925 and 850 hPa impact is neutral or slightly positive for CLAEF_FD. The largest impact is on spread for WS as is for 10-m WS (Figure 5). For the rest of the variables the impact on RMSE/spread is similar to T. Regarding CRPS, which is slightly reduced (T and RH) or neutral (Figure 6).

The results for June are very similar, but slightly more neutral (not shown).

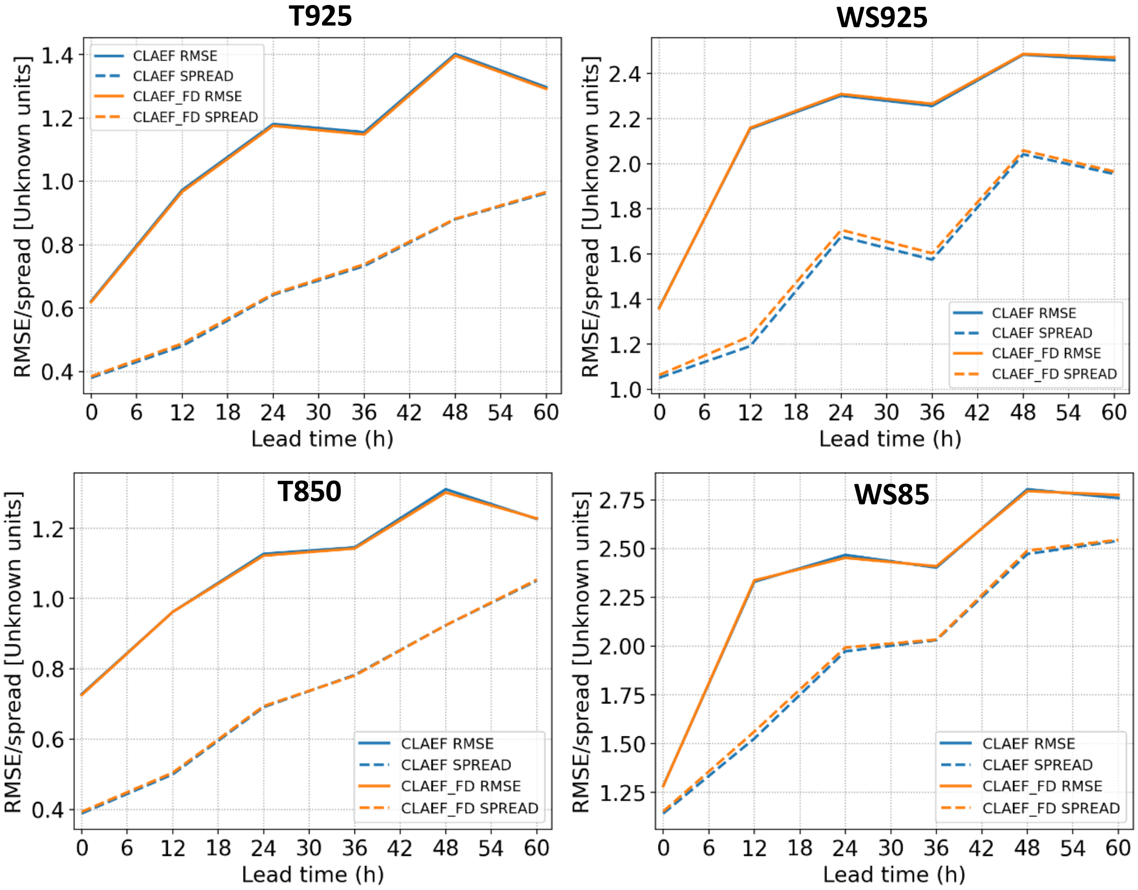

Figure 5. RMSE (solid) and spread (dashed) for CLAEF_oper (blue) and CLAEF_FD (orange) for February 2024. Variables and levels are indicated on each plot.

Conclusions

Overall, the results indicate that by using FD-SPP, ensemble performance is improved. The improvement is not big, but it is consistent. It is the strongest on the surface and weakens as we vertically move away from it. Ensemble spread is almost always increased (see monthly domain averaged spread in Figures A3-A6 in Appendix) and even the RMSE, bias and ROC are positively affected. No negative impact on RMSE and bias is a very good result since stochastic perturbations are known to potentially have negative impact on those scores (Wastl *et. al*, 2019). However, results for June are more neutral due to lack of SPP activity during summer in C-LAEF. Reasons for this remain unknown.

Next steps should include optimization of standard SPP configuration and revalidation of FD-SPP. Based on results presented here, it can easily be concluded that FD-SPP is a promising technique, and it is worth investing in.

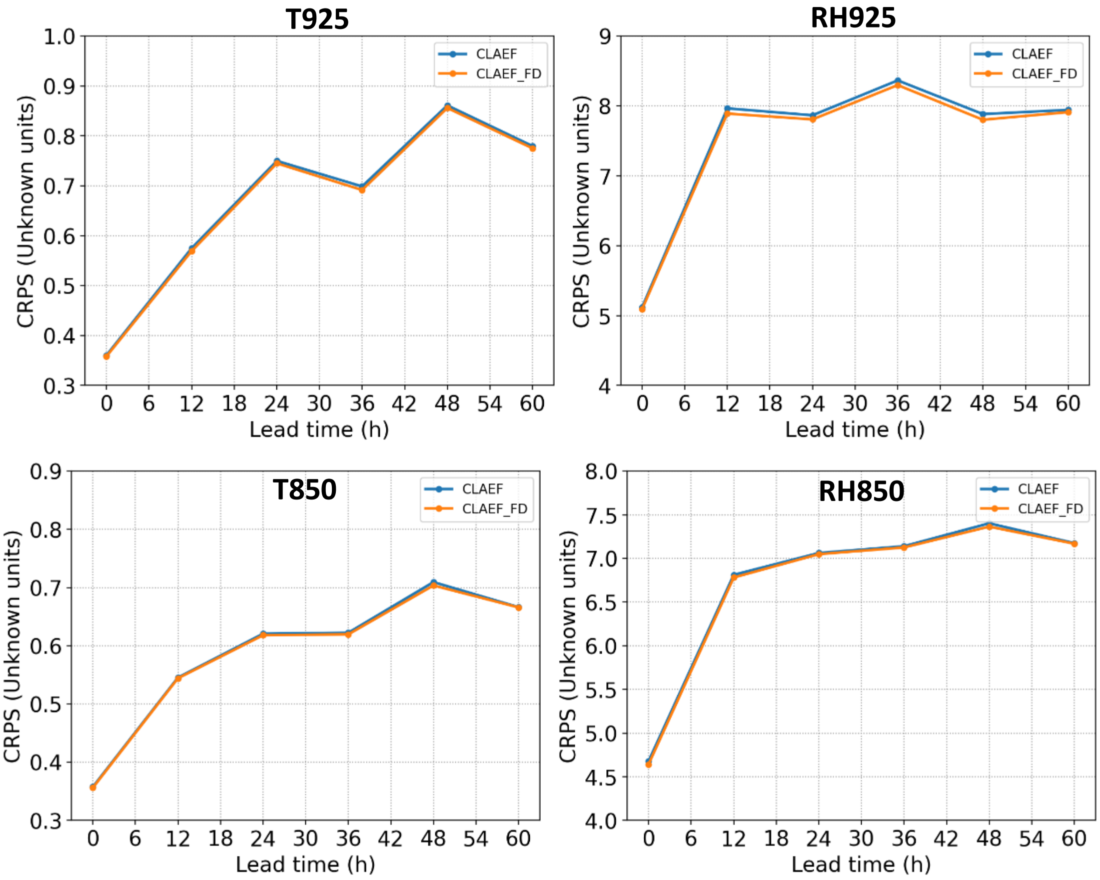

Figure 6. As for Figure 5, but for CRPS and with WS replaced by RH.

Part 3 – CLAEF1k domain extension

The current CLAEF1k e-suite domain doesn't include the whole Croatia (Figure 7). For this reason, it needs to be extended to the south before Croatia enters the C-LAEF1k project. Croatian forecasters expressed the need for the whole Adriatic Sea to be included which means that southern edge of domain needs to be, at least, at the Strait of Otranto. A new domain that fulfills this condition has been created (Figure 7).

The total size of domain has been increased from 1080 grid points to 1350 grid points in Y direction giving a total domain size of 1500x1350 grid points. As linear truncation is used, the highest wavenumber in Y direction was increased from 539 to 674. Configuration 001 consumes around 45800 SBUs on the old domain, which increases to 52300 on the new domain. This is an increase of about 15 %. The total runtime goes from 62 min to 71 min, an increase of 15 %. It is valid to note that those numbers were obtained from a single model run.

It is worth noting that the climate files used on both domains were produced with the following settings: Quadratic truncation for orography and linear truncation for all other fields with the use of additional filtering (ZS_FILTER=1). However, it seems that the CLAEF1k e-suite has climate files that were produced without this additional filtering. However, this doesn't affect the validity of results presented here because we have used the same settings for both domains.

Outputs of the 60-h 001 integration on the new domain were visually inspected and compared to the 001 integration on the old domain. No unusual behavior or unexpected differences were observed (not shown). The only thing to note is that the positions of lakes and cities (some differences were also observed in the orography field) in the climate files for the new domain were different than in the ones for the old domain (Figure 8). They look shifted by 1-2 grid points to the south. At first, this is very unexpected because the position of such features should not depend on the size of the domain. However, due to the spectral nature of our models, this is probably the consequence of the spectral transforms. With the new domain, we have also changed the number of waves used in spectral space and these waves do not have spatial locality like grid points.

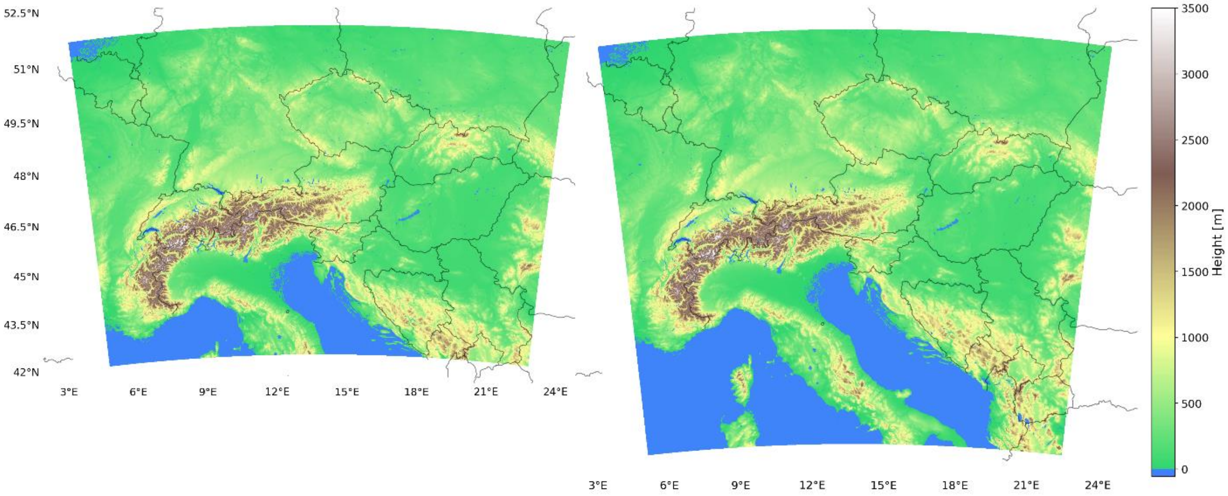

Figure 7. C-LAEF domain (C+I zone) for current configuration (left) and future configuration including Croatia (right).

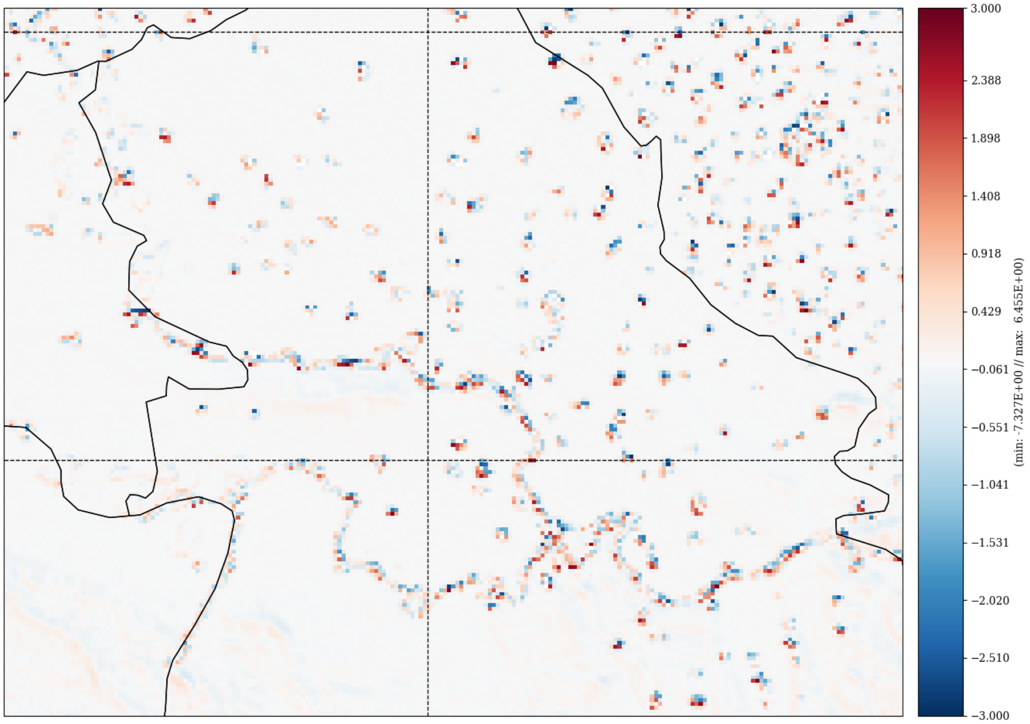

Figure 8. 2-m temperature difference between new and old domains.

Part 4 – C-LAEF lagged setup

Lagging approach is often utilized to cheaply increase the ensemble size and/or to reduce the computational requirements of the ensemble prediction system. Keresturi (2019) explored the possibility of increasing the ensemble size by using lagged deterministic forecasts. Long-term verification has shown benefits (i.e., Increased ensemble spread) of doing so but out of the "wrong" reasons. The main cause of the increased spread was the heavy clustering of the newly added deterministic members. Clustering is, of course, not desirable (see Keresturi, 2019). Here, C-LAEF can be configured so that every 3-h, 4 members produce long-range forecasts which are then combined with 3-, 6-, and 9-hours old forecast form different 4 members. When all of them are combined, a 16-member ensemble is formed, providing a long-range forecast every 3 hours. This is about 2 times cheaper than running the whole ensemble every 3 h. The question is, however, what is the degradation in quality of such a system and is the clustering of members present?

In order to answer those questions, a comprehensive assessment would be required which includes long-term verification and several case studies. Due to the lack of time, this kind of analysis is outside the scope of this work. Here, we'll have a look at one model run (11 July 2024, 00 UTC) and try to find the evidence of clustering. The experiment is done with the C-LAEF operational configuration (2.5 km) – CLAEF_oper and its lagged version - CLAEF_lag at 00 UTC. In CLAEF_lag for 00 UTC run, members 1-4 are current, members 5-8 are 9 h old, members 9-12 are 6 hours old and members 13-16 are 3 hours old.

The results indicate that the clustering is mostly not present, and when it is, it is at the beginning of the forecast. The situations where clustering is present will now be described. (1) The 2 m temperature over the open sea (Figure 9). In this case, the clustering is obvious, and it comes from different SST used by older members (we are not assimilating SST, it is taken from ECMWF). Coastal locations and islands can then be affected but, luckily, the intensity of clustering is reduced (Figure 10 shows an example of a coastal location). (2) Precipitation spatial distribution. Figure 11 shows spatial precipitation plots for all members of the ensemble for CLAEF_oper and CLAEF_lag at the beginning of the forecast (+4 h). It can clearly be seen that members do cluster, but with larger lead time, this clustering disappears (not shown). To some extent, this can also be seen in total cloudiness fields.

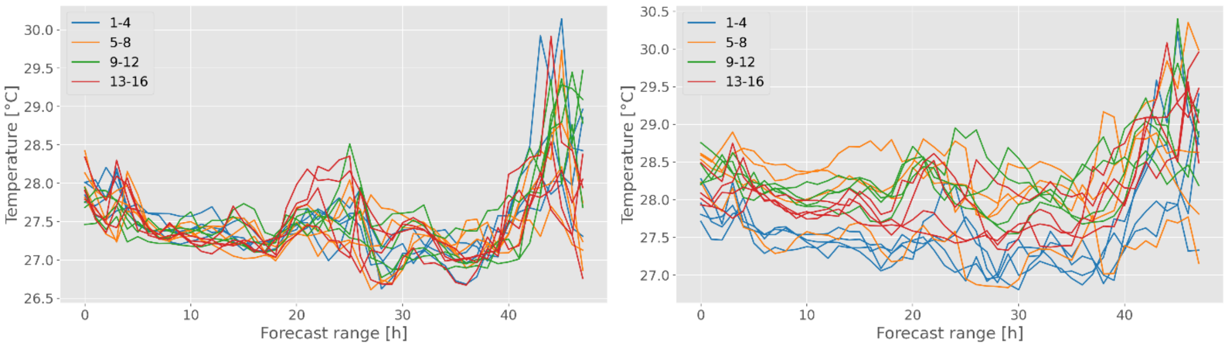

Figure 9. 2 m temperature at open sea for CLAEF_oper (left) and CLAEF_lag (right).

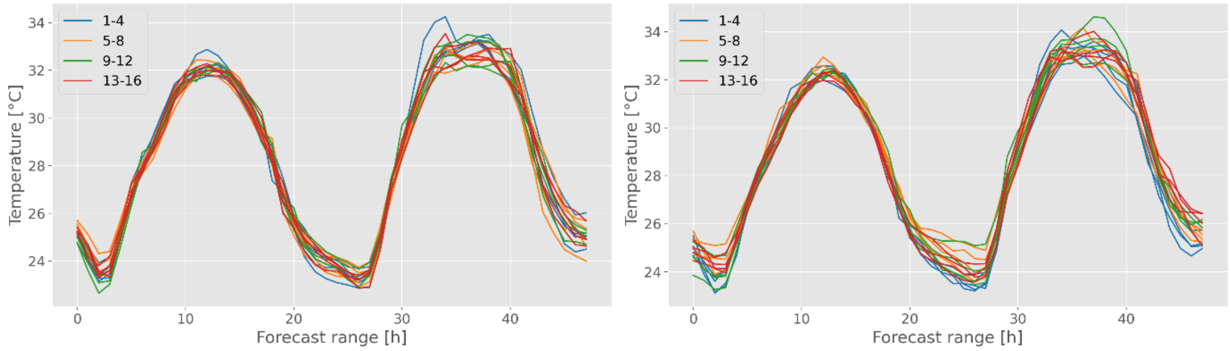

Figure 10. 2 m temperature at Zadar for CLAEF_oper (left) and CLAEF_lag (right).

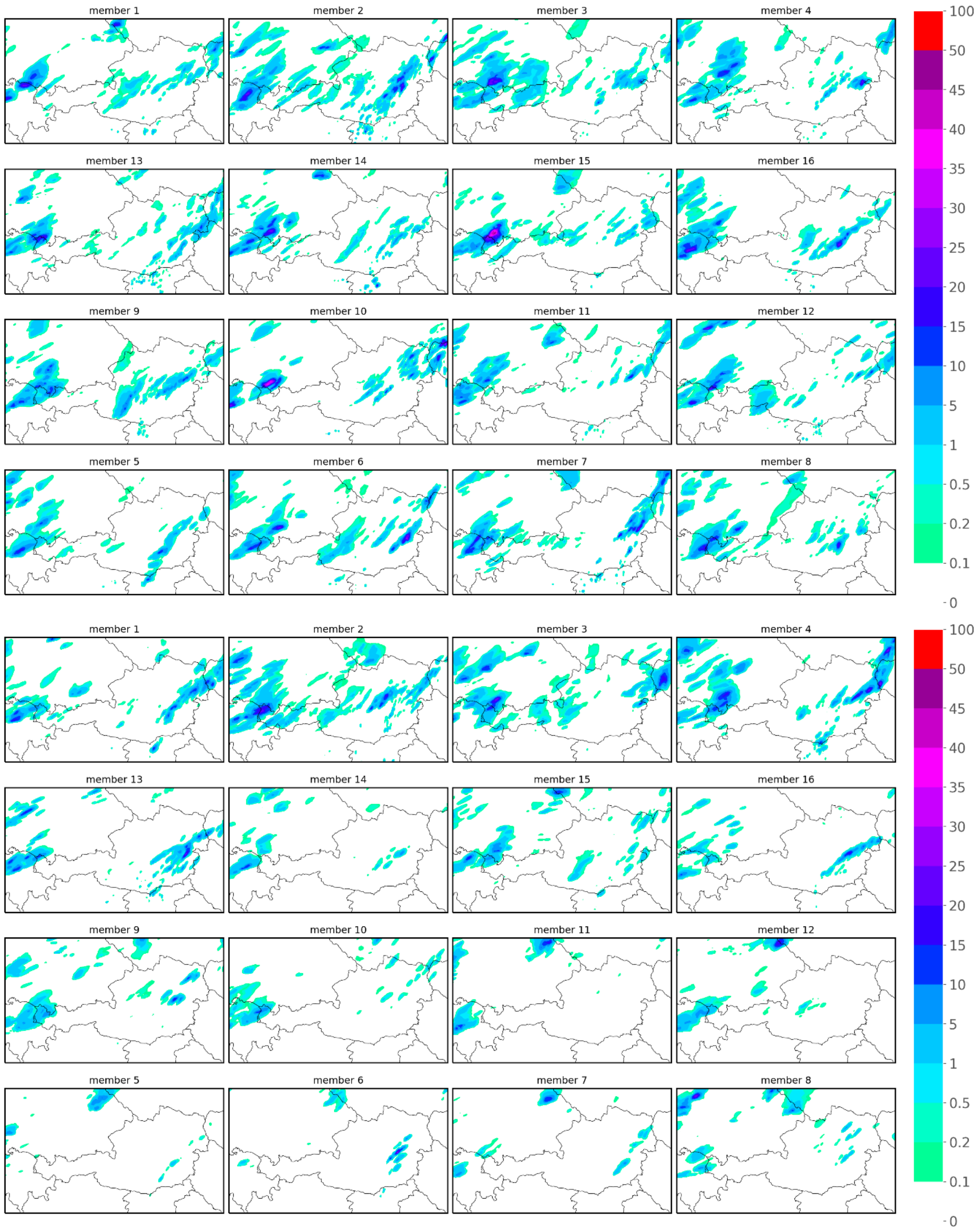

Figure 11. All ensemble members for CLAEF_oper (up) and CLAEF_lag (down) at forecast lead time of 4 h. Members are sorted by age from top (newest) to bottom (oldest).

Appendix

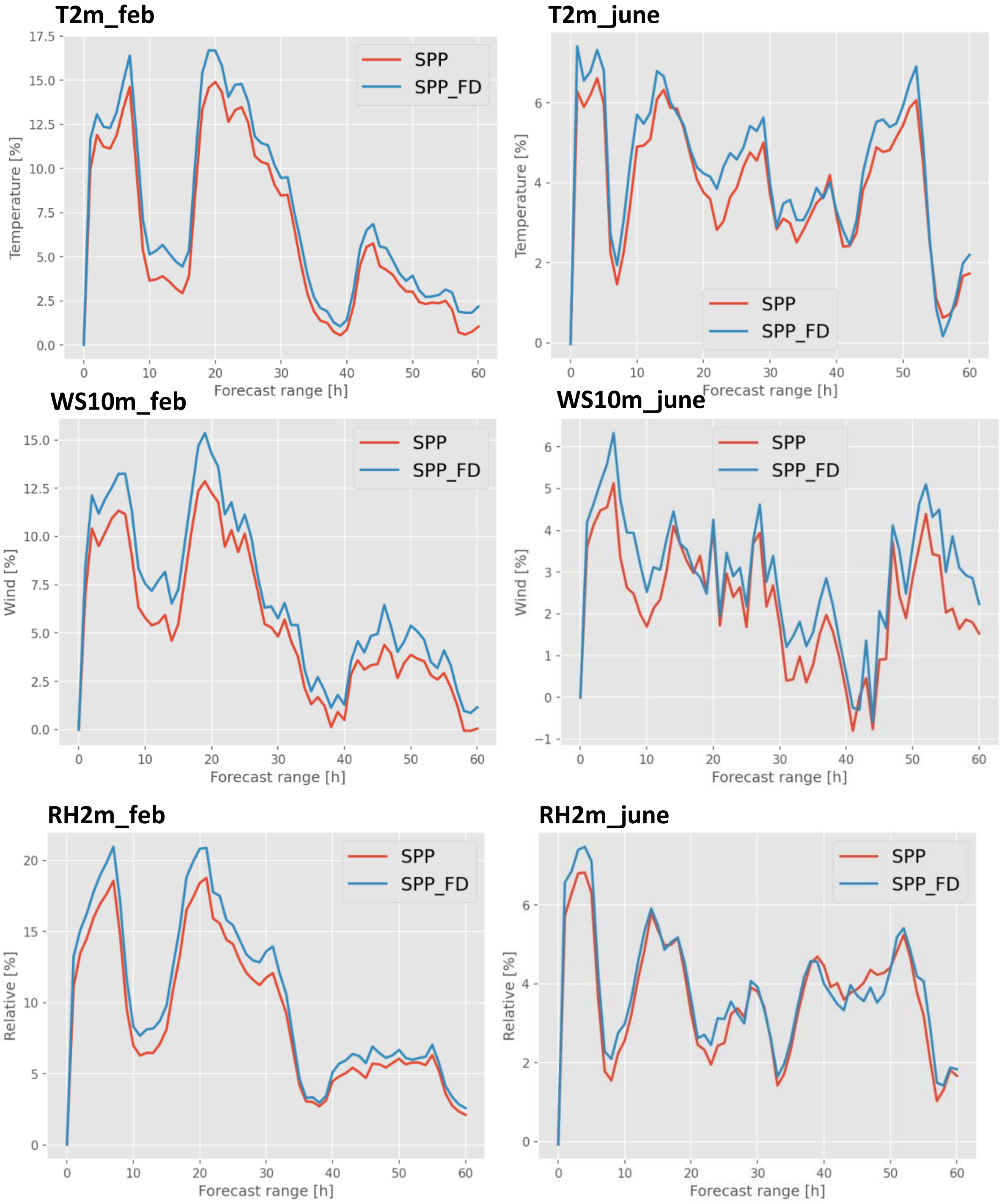

Figure A1. The percentage of domain averaged spread difference for CLAEF_oper (red) and CLAEF_FD (blue) for surface variables indicated above each plot. Plotted against reference - an experiment without SPP. *feb* denotes 29 February 2024 and *june* denotes 28 June 2024.

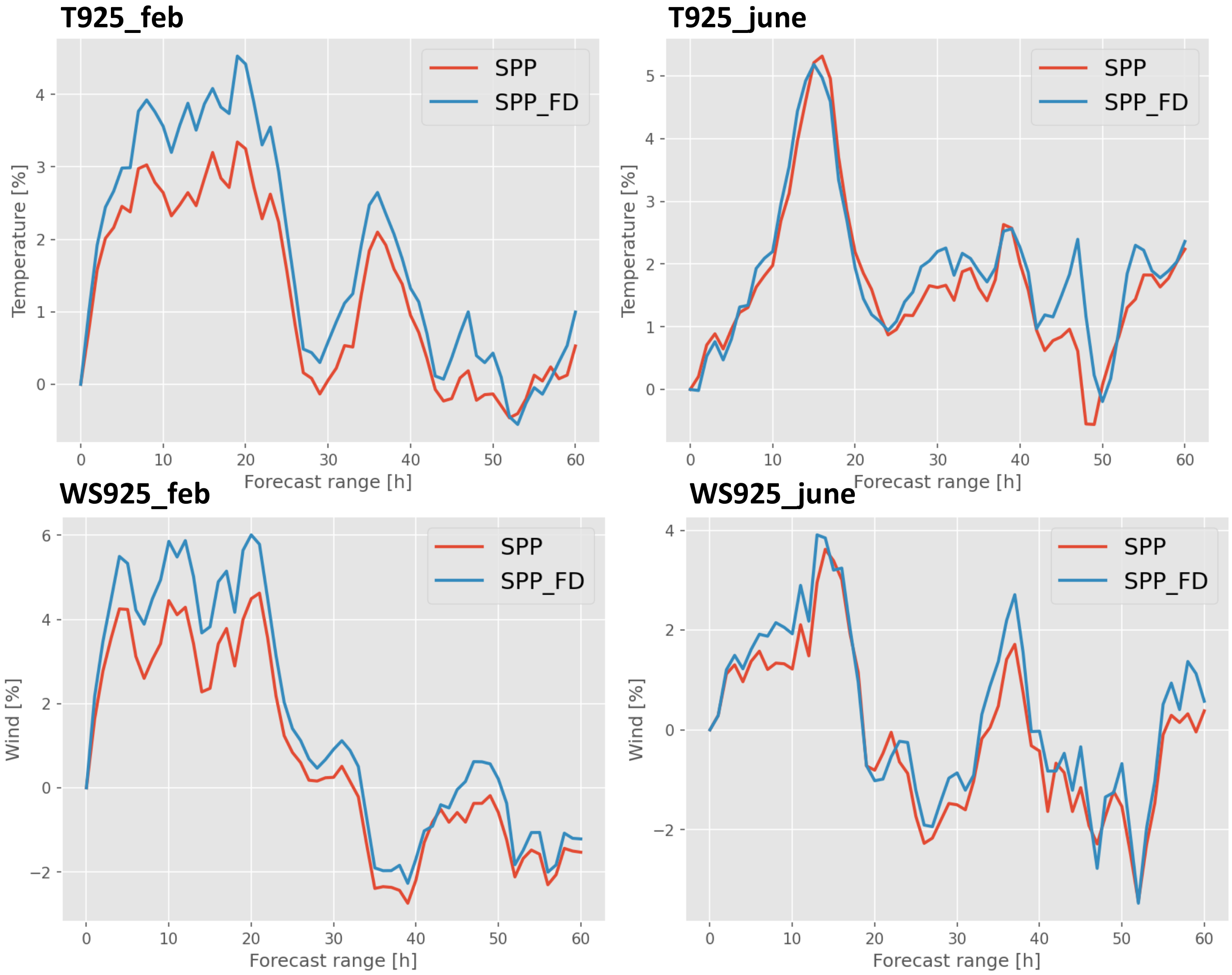

Figure A2. As Figure A1 but for temperature and wind speed at 925 hPa.

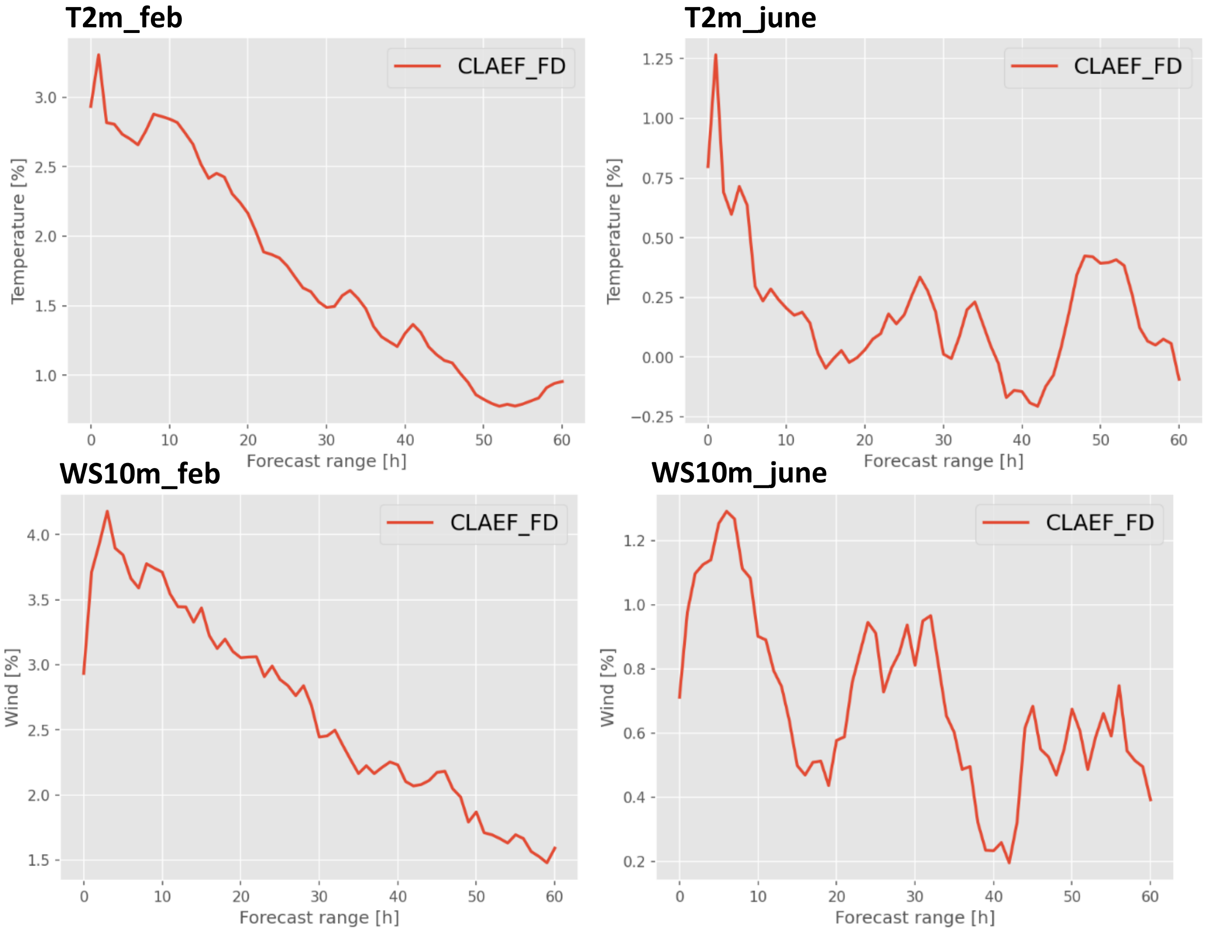

Figure A3. The percentage of domain averaged spread difference between CLAEF_oper and CLAEF_FD for variables, level and period indicated above each plot. Here *feb* and *june* denote entire months.

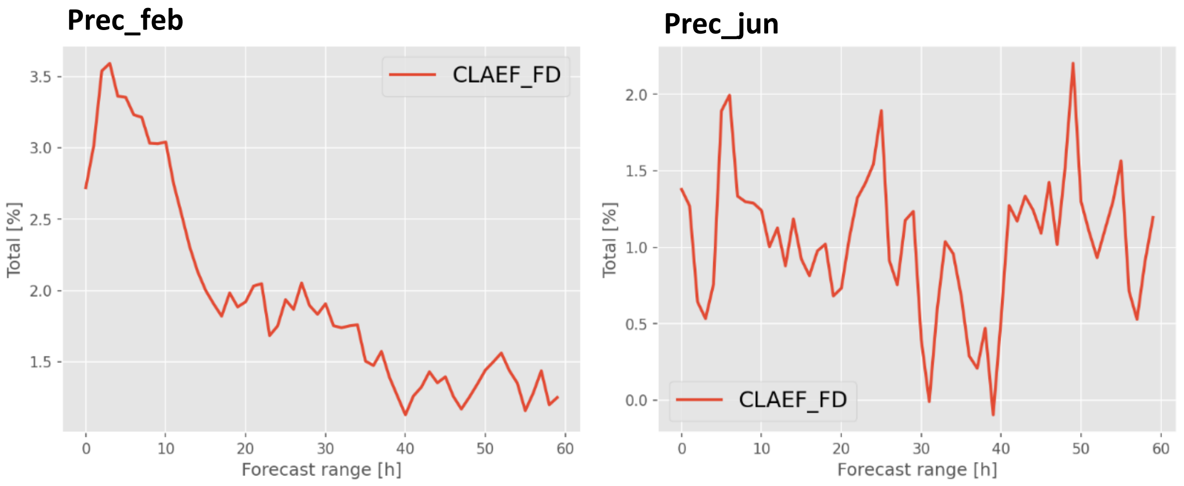

Figure A4. As for Figure A3.

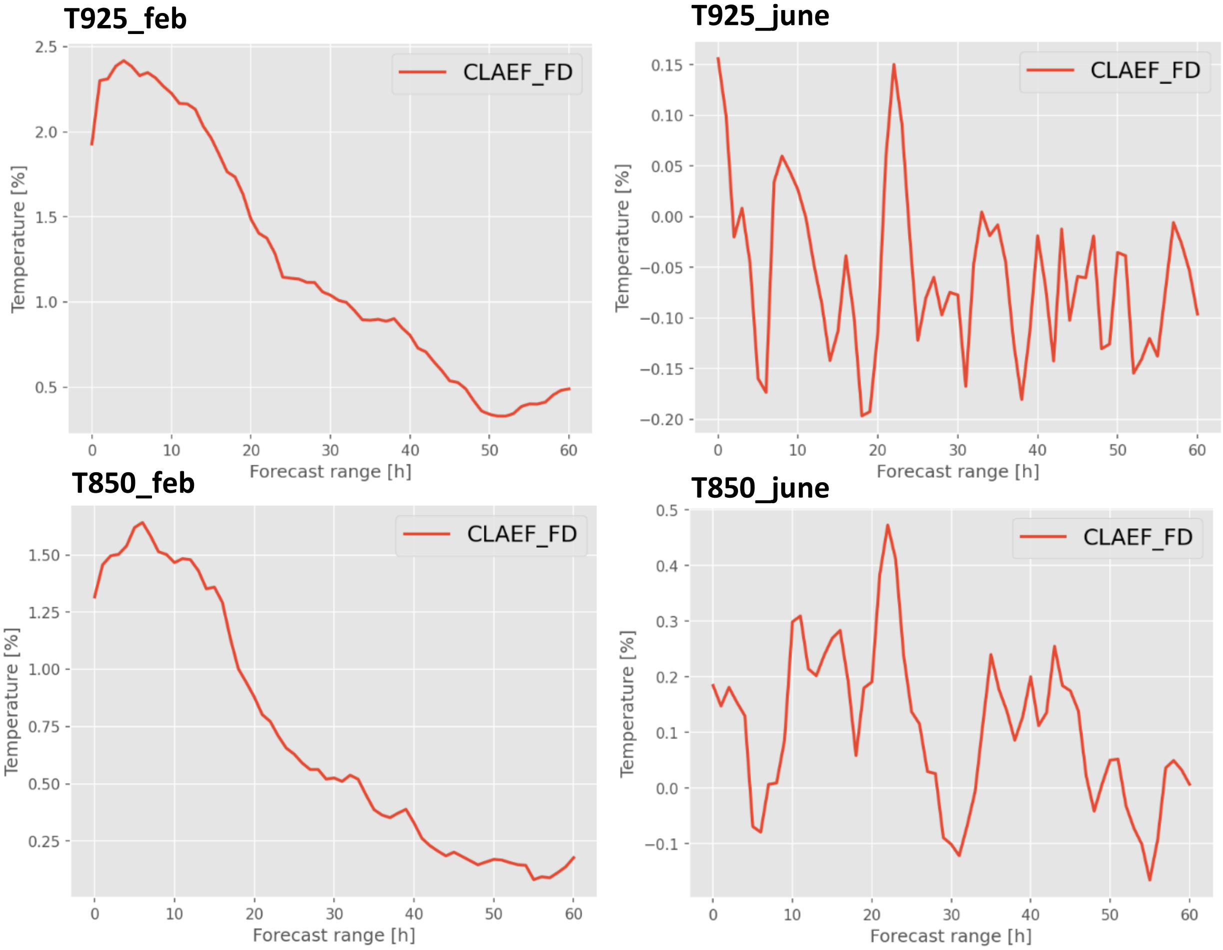

Figure A5. As for Figure A3.

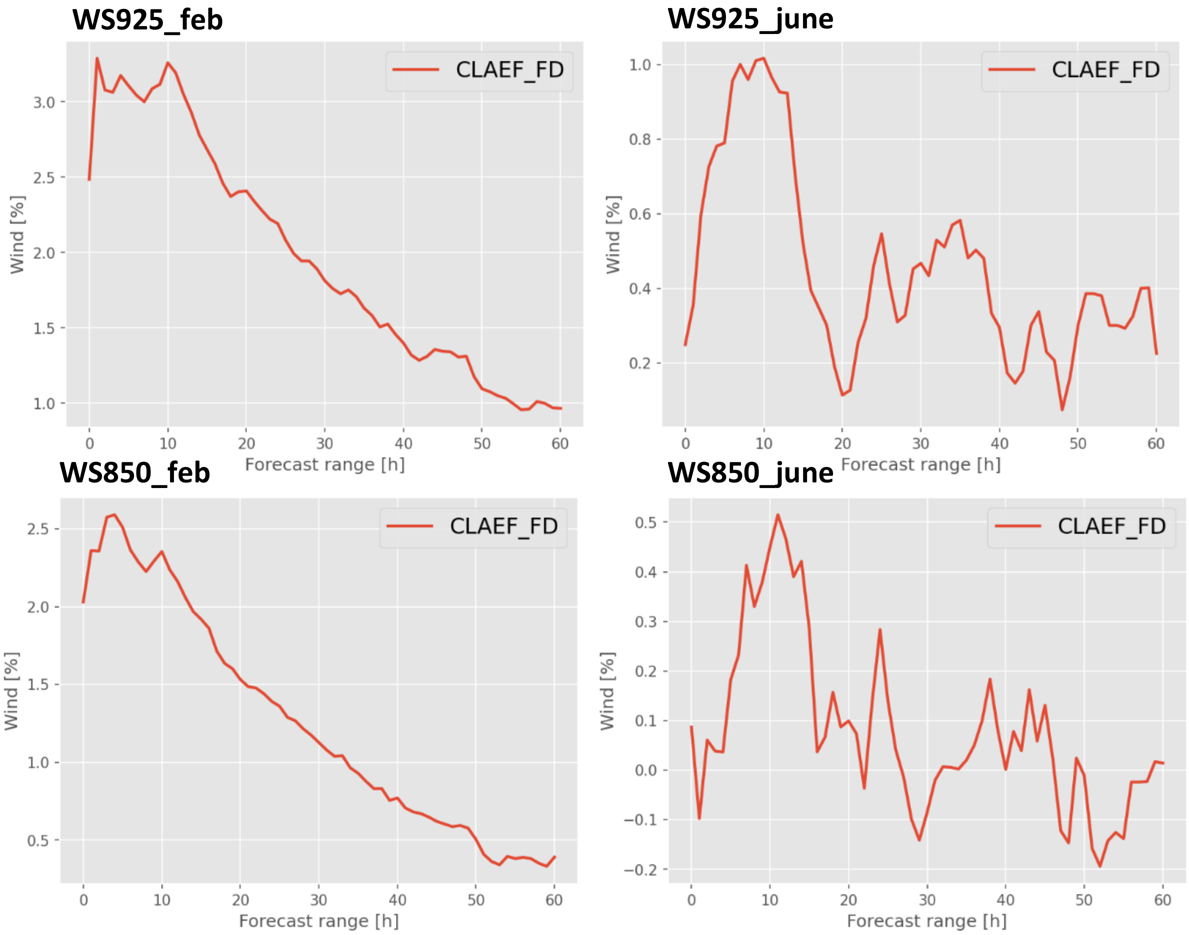

Figure A6. As for Figure A3.

References

Buizza, R., 2008: Comparison of a 51-Member Low-Resolution (TL399L62) Ensemble with a 6-Member High-Resolution (TL799L91) Lagged-Forecast Ensemble. *Mon. Wea. Rev.*, **136**, 3343-3362.

El Khatib, R., 2019: General Fortran optimizations guide. Available online: https://www.umr-cnrm.fr/gmapdoc/IMG/pdf/general_fortran_optimization_guide_manual.pdf

Frogner, I., U. Andrae, P. Ollinaho, A. Hally, K. Hämäläinen, J. Kauhanen, K. Ivarsson, and D. Yazgi, 2022: Model Uncertainty Representation in a Convection-Permitting Ensemble—SPP and SPPT in HarmonEPS. *Mon. Wea. Rev.*, **150**, 775–795.

Keresturi, E., 2019: Adding lagged deterministic forecasts to a convection permitting EPS. Available online: https://www.rclace.eu/media/files/Predictability/2019/Report_Endi_2019.pdf

Keresturi, E., 2022: Flow dependent SPP perturbations for C-LAEF. Available online: https://www.rclace.eu/media/files/Predictability/2022/Report_endy_2022.pdf

Keresturi, E., 2023: Flow dependent SPP perturbations for C-LAEF. Available online: https://www.rclace.eu/media/files/Predictability/2023/LACEstay2023_Endi_Keresturi_report.pdf

Wastl, C., Y. Wang, A. Atencia and C. Wittmann, 2019: Independent perturbations for physics parametrization tendencies in a convection-permitting ensemble (pSPPT). *Geosci. Model Dev.*, **12**, 261–273.