SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Evaluating the impact of single precision forecasts in HarmonEPS for various perturbation strategies
Computer Project Account:	spiefann
Start Year - End Year :	2022 - 2023
Principal Investigator(s)	James Fannon
Affiliation/Address:	Met Éireann Glasnevin Hill Dublin 9 D09 Y921 Ireland
Other Researchers (Name/Affiliation):	Alan Hally, Met Éireann

The following should cover the entire project duration.

Summary of project objectives

The objective of this project was to evaluate the behaviour of HarmonEPS, the ensemble realisation of the HARMONIE-AROME NWP model, in single precision for different perturbation configurations. In particular, this project focussed on assessing the stability and performance of the Stochastically Perturbed Parameterizations scheme in single precision. The ultimate aim of this work was to help identify single precision model stability issues and expedite the transition from double to single precision forecasts within the HARMONIE-AROME community.

Summary of problems encountered

No significant technical problems relating to this Special Project or the HPC facilities at Reading and Bologna were encountered.

Experience with the Special Project framework

This was my first Special Project as a Principal Investigator and I found that the administrative aspects were straightforward, clearly signposted, and reasonable. As such, I was very satisfied with the Special Project framework overall.

Summary of results

The computational resources provided by this Special Project allowed for an extensive set of double and single precision HarmonEPS cycle 43h2.2 experiments to be carried out with both the default suite of perturbation methods in HARMONIE-AROME and with the Stochastically Perturbed Parameterizations (SPP) scheme. The main conclusions from these experiments, which included both debugging and longer cycling runs over Ireland and the UK, are given below:

- Single precision HarmonEPS forecasts with the default set of perturbations are generally stable and perform well relative to double precision, apart from a small positive PMSL bias and negative 2 m temperature bias. Runtime savings of close to 40% are achieved when using single precision forecasts on the Atos machine.
- The SPP scheme in HarmonEPS is found to perform well in double precision, with significant improvements in ensemble spread, CRPS, and spread-skill ratio for the majority of surface parameters over all test periods considered.
- In HarmonEPS cycle 43h2.2, SPP perturbation patterns differ in single and double precision in general. While various source changes were made to address this issue in the case of a pattern update frequency of every hour, differences are still evident when using a pattern update frequency of every timestep.
- In the case of a pattern update frequency of every timestep, forecast crashes can occur when using single precision and SPP together. Erroneous rainfall forecasts can also be observed for some successfully completed single precision SPP perturbed members.
- Single precision SPP stability and performance appears to be improved when using a pattern update frequency of every hour as the single and double precision SPP patterns are almost identical in this case.

For a complete analysis and discussion of the results, we have appended a detailed internal technical report to this document.

List of publications/reports from the project with complete references

Fannon J and Hally A (2023). Single Precision and SPP performance in HarmonEPS Cycle 43h2.2. Met Éireann NWP Note 2023/01. [Internal technical report, appended to this report]

Future plans

The results and experiences from this project have resulted in various source changes to the common HARMONIE-AROME NWP code and are being used to inform the operational configuration of the United Weather Centres-West common NWP model. Research is ongoing into the performance of single precision SPP in cycle 46 and applications for additional Special Projects related to this work may be submitted in the near future.

Met Éireann
NWP Note 2023/01

Single Precision and SPP performance in HarmonEPS Cycle 43h2.2

James Fannon and Alan Hally

Met Éireann
2023

Abstract

This NWP note provides an overview of a series of ensemble experiments conducted using the harmonEPS-43h2.2 branch of HARMONIE-AROME to analyse the performance of:

- single precision forecasts using the default suite of perturbations in HARMONIE-AROME,
- the recently proposed Stochastically Perturbed Parameterizations (SPP) scheme over Ireland and the United Kingdom using two SPP configurations, and
- single precision in combination with the SPP scheme.

The purpose of this testing was to help identify single precision model stability and performance issues, particularly when SPP is activated, and to assess the benefits of SPP. Computational resources for these experiments were provided by an ECMWF Special Project, SPIEFANN, in 2022.

The initial phase of testing focused on degugging tests with SPP and an extensive analysis of SPP perturbation pattern behaviour in single and double precision. Longer cycling experiments of four two week periods, one in each season, were then carried out to assess model stability and performance in different meteorological conditions. The results presented herein suggest that:

- No major stability or performance issues are observed for single precision HARMONIE-AROME forecasts when using the default set of perturbations, apart from a small positive PMSL bias and negative 2 m temperature bias relative to double precision. Runtime savings of close to 40% are achieved when using single precision forecasts on Atos.
- The SPP scheme is found to perform well in double precision, with significant improvements in ensemble spread, CRPS, and spread-skill ratio for the majority of surface parameters over all test periods considered. However, SPP also appears to introduce a negative wind speed bias in the perturbed members relative to the control.
- SPP perturbation patterns differ in single and double precision in general. These differences can be avoided when using a pattern update frequency of every hour, but are evident when using an update frequency of every timestep.
- Forecast crashes can occur when using single precision and SPP together. These failures appear to be linked to significant differences in the behaviour of the single precision SPP perturbation patterns compared to double precision.
- Erroneous rainfall forecasts can be observed for some successfully completed single precision SPP perturbed members. These erroneous forecasts are associated with extremely large PSIGQSAT perturbation values.
- Single precision SPP stability and performance appears to be improved when using a pattern update frequency of every hour as the single and double precision SPP patterns are almost identical in this case.

Contents

1 Introduction

The double precision (DP) floating-point format has traditionally been used to represent real numbers in all elements of an NWP model. A DP floating-point number, which utilises 64 bits of memory, is given by:

$$X_{DP} = (-1)^S \left(1 + \sum_{i=1}^{52} \frac{b_i}{2^i}\right) \left(2^{E-1023}\right), \tag{1}$$

where 1 bit is used for the sign (S) of the number, 11 bits for the exponent (E), and 52 bits for the significant (b_i). The DP format allows for both a wide range (the largest, $X_{\max}$, and smallest, $X_{\min}$, numbers representable in DP are $X_{\max} \sim 10^{308}$ and $X_{\min} \sim 10^{-308}$) and high accuracy (machine epsilon $\epsilon \sim 10^{-16}$). DP is the natural extension of the single precision (SP) floating-point standard, in which only 32 bits of memory are used. This results in a reduction to both the range of numbers which can be represented in SP (i.e. $X_{\max} \sim 10^{38}$ and $X_{\min} \sim 10^{-38}$) and their accuracy ($\epsilon \sim 10^{-7}$).

Over the last decade or so, there has been significant interest in the viability of switching from DP to SP arithmetic in NWP models (e.g. Palmer (2014) and Düben and Palmer (2014)). This has been motivated by several factors, including:

1. The increased interest in forecasting at sub-kilometre resolutions has emphasised the need to improve NWP model efficiency so as to compensate for and minimise the increased computational costs at such high resolutions.

2. NWP models are now primarily memory-bound, i.e. a large portion of the computational time is spent in communication between many individual cores. As such, any way to effectively increase the memory bandwidth would decrease the overall computational time significantly.

3. From a physical perspective, given the various sources of model and initialisation error, is it necessary/semsible to have 16 significant digits of accuracy for (the vast majority of) model variables?

This topic has been investigated extensively at ECMWF using the IFS model, where a variety of code adaptions, including a reformulation of poorly conditioned code and isolating some precision-sensitive routines/variables to run exclusively in DP, were required to allow the forecast component of IFS to run in SP (Váňa et al., 2016). Utilising SP forecasts within IFS was found to give a $\sim 40\%$ runtime saving (relative to DP) without significantly degrading forecast quality (Váňa et al., 2017). This research culminated in the operationalisation of SP forecasts in IFS-HRES and IFSENS as part of ECMWF's upgrade to Cycle 47r2 on May 11$^{\text{th}}$ 2021 (Lang et al., 2021a). It is important to note that only the forecast component of the model is run in SP, as other model aspects e.g. data assimilation, can be highly precision-sensitive. However, research into topics such as SP data assimilation and the use of half precision (8 bits) as part of a mixed precision model is ongoing at ECMWF (Hatfield et al., 2019, 2020).

In light of the successful use of SP in IFS, significant efforts have been made within the HIRLAM and ACCORD communities to enable the HARMONIE-AROME forecast component to run in SP. The option to run SP forecasts was first made available as part of HARMONIE-AROME Cycle 43h2.1 (HIRLAM, 2020), and has been tested by a number of ACCORD members (e.g. Vignes (2019), Feddersen (2021), and Suárez-Molina and Calvo (2021)). These tests have typically observed a $\sim 30\%$ forecast runtime reduction in SP relative to DP. As such, given the potential for substantial savings, there is an increased emphasis within ACCORD over the last number of years on the perspective operational use of SP.

A successful migration from operational DP to SP forecasts requires robust and methodical testing of HARMONIE-AROME in a variety of contexts in order to identify and rectify model performance and stability issues in SP. In Met Éireann, initial exploratory work in SP at Met Éireann was carried out during pre-operational deterministic testing of Cycle 43h2.1 (Bessardon et al., 2021). This was subsequently followed by SP EPS testing using a similar configuration to our operational system (Fannon and Hally, 2021), and SP runs at high horizontal resolutions as part of an ECMWF Special Project on hectometric scale HARMONIE-AROME (Clancy et al., 2022). The main conclusions arising from these studies are:

- With the exception of some isolated crashes caused by relatively trivial bugs, which are now addressed, deterministic and ensemble experiments, using the standard suite of initial, boundary, and surface perturbations, are typically stable in SP. This is true even in the case of 500m horizontal resolution forecasts over complex orography and storm conditions.
- SP has a relatively neutral impact overall on surface verification scores, with the exception of a small positive MSLP bias relative to DP.
- SP may slightly degrade upper-air humidity profiles relative to DP.
- Runtime savings of $\sim 30\%$ are observed at the standard 2.5km horizontal resolution. However, for hectometric-scale forecasts at 750/500m resolution, the saving is $\sim 40\%$ relative to DP.

These results are generally consistent with results obtained at other NHMs. However, while testing to date has suggested that SP forecasts generally remain stable and demonstrate good meteorological performance relative to DP, questions still remain to be addressed before SP can be considered ready for operational use.

One area of particular interest for SP EPS testing in HarmonEPS is that of the recently proposed Stochastically Perturbed Parameterizations (SPP) scheme (Frogner et al., 2022), which has been in operational use at ECMWF for several years (Ollinaho et al., 2017; Lang et al., 2021b). The SPP scheme introduces stochastic perturbations to selected closure parameters in the physical parameterizations of the model, fifteen of which are currently available in HarmonEPS. SPP has been shown to have a positive impact in ensemble spread, particularly for cloud variables, and was first operationalised by MetCoOp in August 2022 using a five parameter configuration.

However, compatibility of SPP and SP in HarmonEPS is a significant outstanding issue to be investigated. While all standard model perturbation methods (initial, LBCs, and surface) are carried out before entering the Forecast step of HARMONIE-AROME (and hence are run in DP), the SPP scheme occurs within the Forecast component itself. As such, if the forecast is run in SP, so too must the SPP scheme. Initial SP SPP testing was carried out at Met Éireann in 2021 (Fannon and Hally, 2021), which identified and resolved a number of model crashes and concluded that significant further testing was required.

This NWP note describes a series of extensive tests with SP HarmonEPS, with a particular focus on the performance and stability of the SPP scheme. This work was carried out as part of an ECMWF Special Project in 2022 (SPIEFANN), whose ultimate aim was to help identify SP model stability issues and expedite the transition from DP to SP forecasts within the HARMONIE-AROME community. The remainder of this note is structured as follows. In Section 2 we outline the code used, details regarding how to use SP and SPP in HARMONIE-AROME, and a number of exploratory technical tests with SPP. The performance of SP HarmonEPS using the standard suite initial, LBC, and surface perturbations is detailed in Section 3, along with associated runtime savings. Double precision SPP

experiments are detailed in Section 4, where we focus on the meteorological performance of several five parameter SPP configurations over a number of test periods. An extensive analysis of SP SPP stability and performance relative to DP is given in Section 5, including a discussion of SP model failures and SP SPP patterns. Finally, a summary of findings and recommendations for future work is given in Section 6.

2 Technical details and initial testing

This section provides details regarding the code used in the project, how to use SP in HARMONIE-AROME, the SPP scheme, and some initial technical tests with SP SPP. Note that, due to the migration of ECMWF services in 2022, the work carried out as part of this project utilised both the Cray HPC in Reading and the new Atos HPC in Bologna. However, for consistency purposes, the experiments described in this NWP note will generally refer to those run on Atos, unless otherwise stated.

2.1 Code and common configuration settings

The harmonEPS-43h2.2 branch of HARMONIE-AROME was utilised for all experiments discussed herein. This branch had been used extensively for SPP tuning experiments in 2021 and was the main repository for SPP code updates, and hence it was deemed most suitable for our purposes. Porting of this branch to the Atos HPC was required as it lagged significantly behind the default cycle 43 branch of HARMONIE-AROME. Details of the changes required to allow for the use of harmonEPS-43h2.2 on Atos are given in Appendix 7.1.2. A list of all binary versions used is also given in Appendix 7.1.1.

Some noteworthy configuration settings, which are common to all experiments discussed in this NWP note, are given in Table 1. The experimental IRELAND25 domain was used instead of the larger operational domain (IRELAND25_090) in order to reduce computation costs. Additional settings, such as the binaries used, precision, ensemble size, etc., will be detailed for each experiment as approriate. All other settings, if not explicitly stated, can be assumed set to be the default choice in the harmonEPS-43h2.2 branch.

Component	**Description**
Domain	IRELAND25, L65, linear grid
Data Assimilation	3DVAR, CANARI_OI_MAIN, conventional observations only
Cycles	3hr cycling
ECOCLIMAP	ECOCLIMAP SG
LSMIXBC	'yes' for the control member, 'no' otherwise
SLAF	SCALE_PERT=yes, default SLAFLAG and SLAFDIFF
Compiler	gnu

Table 1: Configuration settings common to all experiments.

2.2 Running HARMONIE-AROME in single precision

The namelist option "FP_PRECISION" in ecf/config_exp.h is used to control the floating-point precision in HARMONIE-AROME cycle 43h2.2. This option sets pre-processor flags in the makeup configuration file to control model precision. For example, from util/makeup/config.ECMWF.atos.gnu:

```
ifeq ($(FP_PRECISION),single)
      FDEFS += -DPARKIND1_SINGLE -DB2O_HAVE_IFSAUX -DHIRLAM_SP_HACKS
      CDEFS += -DPARKIND1_SINGLE -DB2O_HAVE_IFSAUX
      AUTODBL=
else
      AUTODBL=-fdefault-real-8 -fdefault-double-8 -DREAL_8
endif
```

where the "PARKIND1_SINGLE" flag is used by src/ifsaux/module/parkind1.F90:

```
!     Real Kinds
!     ----------
!
INTEGER, PARAMETER :: JPRT = SELECTED_REAL_KIND(2,1)
INTEGER, PARAMETER :: JPRS = SELECTED_REAL_KIND(4,2)
INTEGER, PARAMETER :: JPRM = SELECTED_REAL_KIND(6,37)
#ifdef PARKIND1_SINGLE
INTEGER, PARAMETER :: JPRB = SELECTED_REAL_KIND(6,37)
#else
INTEGER, PARAMETER :: JPRB = SELECTED_REAL_KIND(13,300)
#endif

! Double real for C code and special places requiring
!    higher precision.
INTEGER, PARAMETER :: JPRD = SELECTED_REAL_KIND(13,300)
```

Hence "FP_PRECISION" controls the real kind "JPRB", which is the default choice used throughout the model for real variables. As noted above, a variable of kind "JPRD" is always in DP, and these are used for various precision-sensitive calculations or variables.

There are three choices available for "FP_PRECISION":

- double: This is the default choice in HARMONIE-AROME and uses DP reals everywhere in the model.
- single: Switches variables of kind "JPRB" to SP everywhere in the model. As discussed in Section 1, this is generally not advisable as it will run all model elements, including boundary preparation, data assimilation, etc., in SP.
- dual: When building with FP_PRECISION=dual, the model will generate two sets of binaries; one using SP "JPRB", and the other using DP "JPRB". These binaries are stored under BINDIR/R32 and BINDIR/R64, respectively. For this "dual" option, the model will then use DP binaries (i.e. BINDIR/R64) for all model elements except the Forecast component, which uses the SP binaries (BINDIR/R32). This is evidenced in scr/Forecast:

  ```
  bindir=$BINDIR
  [ "$FP_PRECISION" = "dual" ] && bindir=$(readlink -f $BINDIR/../R32)
  $MPPEXEC $bindir/$MODEL || exit
  ```

  This methodology is also utilised at ECMWF. Note that "dual" will assume that the R32 and R64 binaries exist under BINDIR, and as such, BUILD=yes is required initially.

As such, the FP_PRECISION=dual option provides a convenient means of running SP forecasts in HARMONIE-AROME cycle 43h2.2. Throughout this note, the term "single precision experiment" or "single precision forecast" will refer to this FP_PRECISION=dual setting, and implies that only the Forecast component of the model is run in SP.

2.3 Using SPP in HARMONIE-AROME

SPP perturbations are simply switched on via the "SPP=yes" namelist option in ecf/config_exp.h. The perturbation patterns for a given parameter are generated using the Stochastic Pattern Generator (SPG) routine, with these patterns evolving in space and time according to specified spatial and temporal correlation scales. The pattern update frequency is controlled by "NPATFR_SPP" in ecf/config_exp.h; for example, SPP perturbation patterns can be updated every model timestep (NPATFR_SPP=1) or every X hours (NPATFR_SPP=-X), in which case the perturbation patterns are linearly interpolated in time within the pattern update interval. Hourly update intervals are typically used operationally to reduce computation cost (see Section 4.5).

The perturbed parameter values are themselves drawn from either lognormal or pseudo-uniform distributions whose mean value is approximately the default value used for the parameter in the model. Individual parameters to be perturbed using SPP are switched on/off under the "NAMSPP" namelist in nam/harmonie_namelists.pm, where other settings related to the parameter PDF, such as distribution type (lognormal/uniform) and standard deviation (known as the "CMPERT" value), are also set.

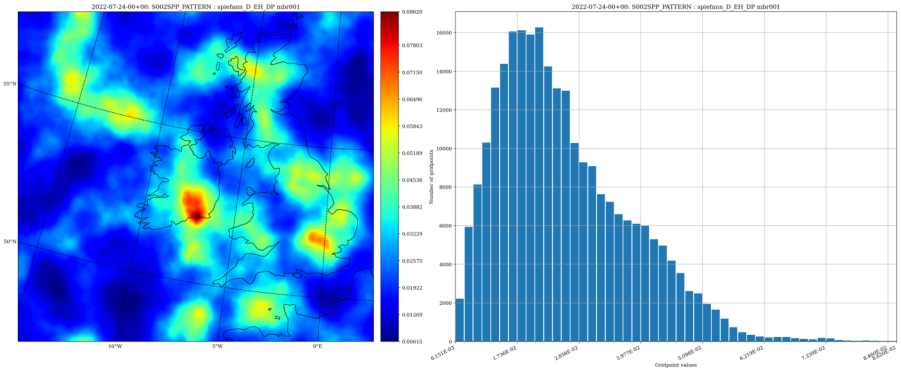
In this NWP note, we will explore the use of two proposed SPP configurations; a five parameter configuration implemented by the MetCoOp consortium in their operational EPS and a five parameter configuration proposed at KNMI (referred to here as the "5PMC" and "5PAT" configurations, respectively). Namelist settings for these setups are provided in Table 2. The two configurations are broadly similar and share four common parameters.

Config	Parameter	LUNIFORM	OFFSET	CMPERT	Additional details
5PAT	PSIGQSAT	FALSE	0.5	0.6	
	ICE_CLD_WGT	TRUE	0.5	1.5	
	RFAC_TWOC	FALSE	0.5	0.6	
	RZC_H	TRUE	0.475	1.35	LCORR_RZL_INF=TRUE
	RZL_INF	TRUE	0.25	2.25	LCORR_RZL_INF=TRUE
5PMC	PSIGQSAT	FALSE	0.5	0.6	
	CLDDPTHDP	FALSE	0.5	0.6	
	ICE_CLD_WGT	TRUE	0.5	1.2	
	RZC_H	TRUE	0.475	1.05	LCORR_RZL_INF=TRUE
	RZL_INF	FALSE	0.5	0.45	LCORR_RZL_INF=TRUE

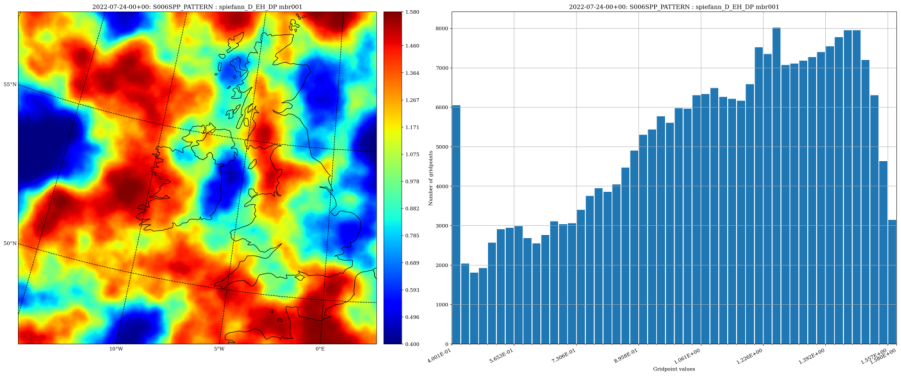
Table 2: NAMSPP namelist settings for the MetCoOp (5PMC) and KNMI (5PAT) SPP configurations. Note that the corresponding "LPERT" and "LLNN_MEAN1" flags are TRUE for each parameter listed. The column "OFFSET" refers to the "UNIFORM_OFFSET" flag.

Sample SPP perturbation pattern fields for parameters PSIGQSAT and ICE_CLD_WGT are illustrated in Figure 1. These fields are generated from the hour zero ICM file of a DP experiment utilising the 5PMC configuration and one perturbed member.[1] Note that "TEND_DIAG=yes" in ecf/config_exp.h is required in order to save the SPP patterns in the ICM files. Associated histograms of the pattern fields are also given in Figure 1. As indicated in Table 2, PSIGQSAT values are drawn from a lognormal distribution, while ICE_CLD_WGT uses a pseudo-uniform distribution.

[1] See configuration settings for "Exp A" in Table 3.



(a) PSIGQSAT



(b) ICE_CLD_WGT

Figure 1: Sample scaled SPP perturbation pattern fields (left) and corresponding histograms (right) at hour 0 for parameters (a) PSIGQSAT and (b) ICE_CLD_WGT. See text for additional details.

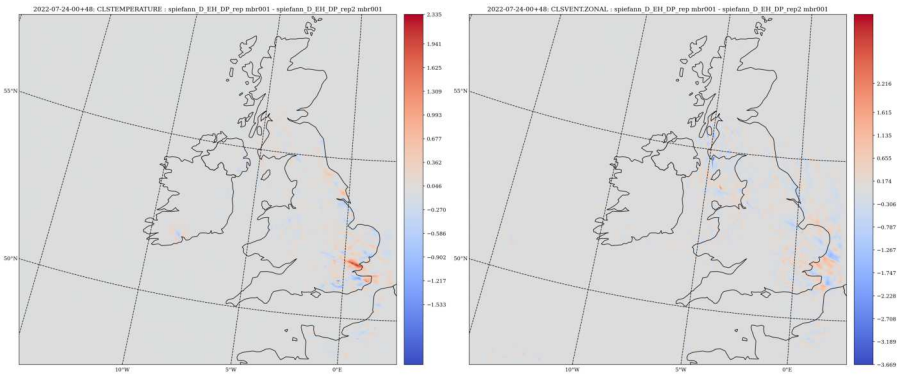
2.4 SPP forecast reproducibility

One feature to note when utilising SPP perturbations is the issue of forecast reproducibility. As an illustrative example, in Table 3 we describe two standard DP ensemble experiments with 0+1 members; one a reference experiment with the default set of EDA, surface, and SLAF perturbations only (RefA), and the other the same reference experiment but with SPP switched on (ExpA). A single 48 hour forecast was run for the 0000 UTC cycle on July 24th 2022, with each member started from the same first guess files and provided with identical boundary and observation files. Each experiment was then copied and repeated under a new name to test for model reproducibility, i.e. experiment RefB is an identical copy of RefA, while ExpB is a copy of ExpA.

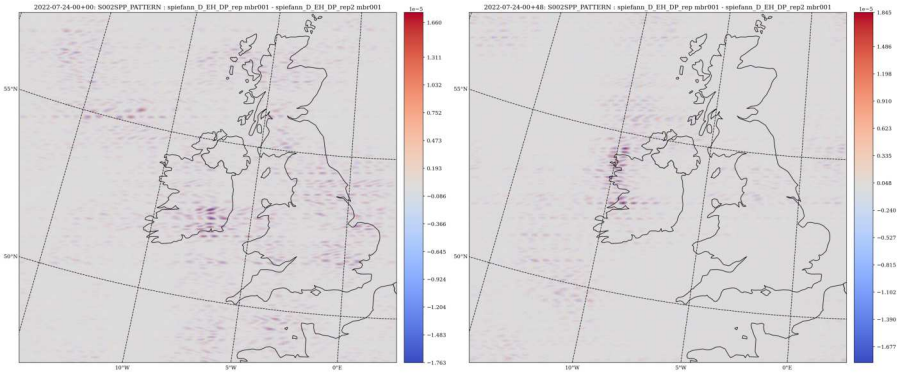
As expected, experiments RefA and RefB produce bit reproducible forecasts at hour 48 for both the control and perturbed member (not shown). However, when SPP is activated the pertubed members in ExpA and ExpB differ slightly at hour 48, as illustrated in Figure 2 for 2 m temperature and 10 m zonal wind speed. This lack of reproducibility appears to be linked to small differences in the underlying SPP perturbation patterns. These differences are present at hour zero and persist throughout the forecast, as illustrated for parameter PSIGQSAT in Figure 2. This is also the case for other perturbation parameters (not shown). While the forecast differences are evidently quite small, and could be deemed somewhat irrelevant, it is worth being aware of.

Exp.	Binaries	Precision	ENSMSEL	SPP	SPP config	NPATFR_SPP
RefA	A_1	double	0-1	no	-	-
RefB	A_1	double	0-1	no	-	-
ExpA	A_1	double	0-1	yes	5PMC	-1
ExpB	A_1	double	0-1	yes	5PMC	-1
ExpC	A_1	single	1	yes	5PMC	-1
ExpD	A_2	double	1	yes	5PMC	-1
ExpE	A_2	single	1	yes	5PMC	-1
ExpF	A_2	double	1	yes	5PMC	1
ExpG	A_2	single	1	yes	5PMC	1

Table 3: Experiment settings for various SPP forecast reproducibility and SP SPP tests discussed in Section 2. For each experiment a single 48 hour forecast was run for the 0000 UTC cycle on July 24th 2022. SLAF, EDA (PERTATMO=CCMA) and surface (PERTSURF=model) perturbations are switched on for each experiment.



(a) Hour 48 forecast differences for 2 m temperature (left) and 10 m zonal wind speed (right)



(b) Forecast differences at hour 0 (left) and hour 48 (right)

Figure 2: (a) Differences at hour 48 for ExpA and ExpB 2 m temperature and 10 m zonal wind speed forecasts. (b) Differences in the scaled PSIGQSAT perturbation pattern for ExpA and ExpB at hours 0 and 48. Experiment details in Table 3.

2.5 Technical testing of single precision SPP

As discussed in Section 1, exploratory SP SPP testing carried out at Met Éireann in 2021 identified a number of forecast model crashes which were subsequently resolved (Fannon and Hally, 2021). As such, initial technical testing with SP SPP in this project focused on assessing basic model stability and SP perturbation pattern behaviour relative to DP. Both single forecast and longer cycling tests (over a two week period) were carried out for debugging purposes.

In contrast to previous experience, no immediate forecast crashes were observed with SP SPP during these technical tests. However, these tests did reveal significant differences between the SPP perturbation patterns generated in SP and DP, despite the fact that the same pattern characteristics (e.g. correlation scales, pattern mean and standard deviation) were used for each. Indeed, the technical tests suggested a precision-dependence in the perturbation patterns. The underlying source of this precision-dependence was investigated extensively and highlighted several issues, which are detailed in the following sections.

2.5.1 Static initial seeds

During the initialisation of the SPP scheme in src/arpifs/phys_dmn/ini_spp.F90, an initial random number seed for the SPP pattern generator is set using the "KGET_SEED_SPP" routine (as defined in src/arpifs/module/spp_mod.F90). The default behaviour in HARMONIE-AROME is to set this initial seed using SETRAN (src/arpifs/var/setran.F90), where the seed is a function of the current forecast cycle, ensemble member number, and an offset unique to each perturbation parameter. This ensures that the patterns are "reproducible".[2]

Technical SPP testing[3] revealed that the initial seeds set by KGET_SEED_SPP for the SPP patterns differed in SP and DP. Moreover, it was found that the dependence of the seed on the forecast cycle was essentially lost when generating the patterns in SP. This is illustrated in Figure 3, where we plot the initial seed used in the SPP pattern generation for each 00 UTC cycle over a two week period in February 2020. Three indicative SPP parameters are considered for two ensemble members. The initial seed varies significantly with the forecast cycle in the reference DP experiment (green lines), and are different for each ensemble member and parameter. However, these seeds are "static" in time when the experiment is repeated in SP (orange lines).

This issue was traced back to the section of the SETRAN routine given in Listing 1. Precision-dependence is inherited from the "DIGITS" function, whose value is 53 and 24 for DP and SP reals, respectively. The static seed issue arises from the definition of "ZTIM", as follows:

- The "RJUDAT" function (arpifs/function/fcttim.func.h), which incorporates the year, month, and day variation, has an absolute value of $\sim 10^6$ but varies slowly over time, e.g. on the scale of 10^1 over the two week February 2020 period.
- The variable "KCONSEED" is a preset offset in the SPP settings, and is $\sim 10^4$. Hence the last component of ZTIM is on the order of 10^{10}.
- As such, given there are only ~ 7 digits of accuracy in SP, the $\sim 10^1$ variation in RJUDAT was lost when subtracting variables of scale 10^{10} and 10^6 in the definition of ZTIM. Thus ZTIM remained static upon varying the forecast cycle when computed in SP.

[2] This is not strictly true, however, as evidenced in Section 2.4.

[3] Using the C_1 binary version on cca, see Appendix 7.1.1

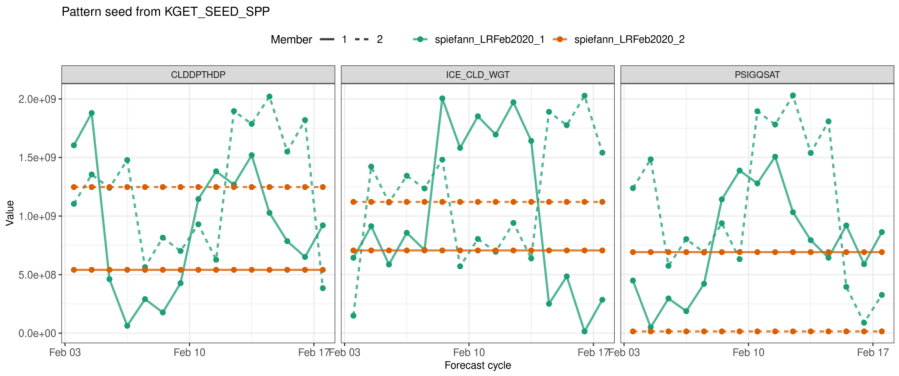


Figure 3: Initial pattern seed used for the SPP pattern generation for all 00 UTC cycles over February 3rd-17th 2020. The green and orange lines represent DP and SP experiments, respectively, using binary version C_1 in Table 8 (i.e. before the SETRAN modification discussed in the text). Ensemble members 1 and 2 are indicated by different line types.

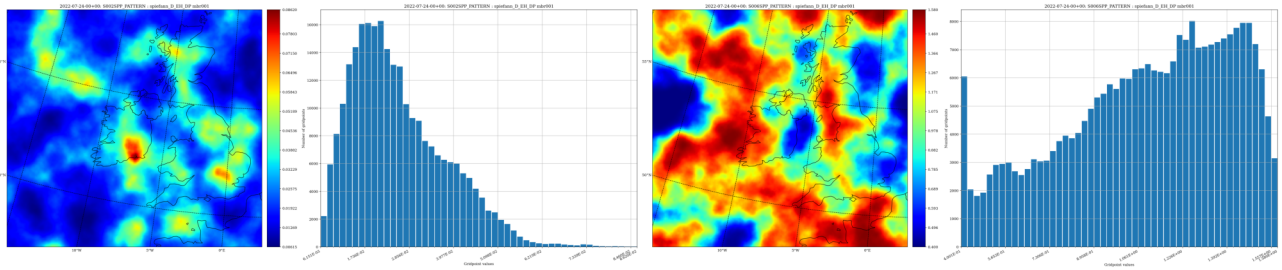
```
IRADIX  = RADIX(ZTIM)
IDIGITS = DIGITS(ZTIM)
ZIRR1 = 0.5_JPRB*(SQRT(5._JPRB)-1.0_JPRB)

!--- generate a unique number from the date and the input KCONSEED

ZTIM = RJUDAT(NCCAA(NINDAT),NMM(NINDAT),NDD(NINDAT))&
 & -1720994.5_JPRB + REAL(NSSSSS,JPRB)/RDAY &
 & -2581470.3_JPRB*KCONSEED

!--- multiply by an irrational number to randomize the bits and scale
!--- to between 0 and 1.

ZTIM = FRACTION(ZIRR1*ABS(ZTIM))

!--- reverse the bits

ZS = 0.0_JPRB
ZT = ZTIM
DO JDIGIT=1,IDIGITS
  ZT = ZT*IRADIX
  IS = INT(ZT)
  ZT = ZT-IS
  ZS = (ZS+IS)/IRADIX
ENDDO

!--- Scale to an odd number between 0 and HUGE-100000000
!--- (Allow some headroom because some routines use setran to set an initial seed,
!---  and then generate new seeds by incrementing.)

ISCALE=(HUGE(ISEED)-100000000)/2
ISEED = 1 + 2*INT( ISCALE*ZS )
```

Listing 1: Segment of the SETRAN routine from src/arpifs/var/setran.F90.

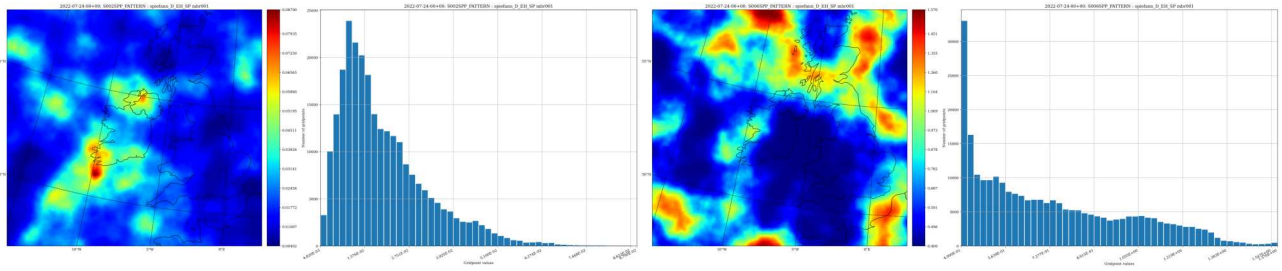
This issue was addressed by simply converting this SETRAN routine to exclusively use JPRD reals in all calculations, hence ensuring identical initial seeds for the SPP patterns regardless of the precision choice. Details of this change can be found in pull request #404 in the harmonEPS-43h2.2 branch. This change is assumed in all experiments discussed henceforth.

2.5.2 SPP dependence on precision and pattern update frequency

While the fix discussed in Section 2.5.1 ensured the same initial pattern seeds in SP and DP, differences in the SPP perturbation patterns were still observed during technical testing. A typical example of this is given in Figure 4 for member 1 of a DP (ExpA) and SP (ExpC) experiment, as described in Table 3. Both the parameter pattern maps and associated distributions are significantly different at hour 0 in this case. Such differences are present over all leadtimes, as illustrated for PSIGQSAT in Figure 5, where the maximum, minimum, mean, and standard deviation of the SPP perturbation pattern at each hour is plotted as a function of leadtime. A comparison of ExpA (green line) and ExpC (orange line) indicates that the DP and SP patterns are uncorrelated. The same behaviour is evidenced for all perturbation parameters (not shown).



(a) DP. Left: PSIGQSAT. Right: ICE_CLD_WGT



(b) SP. Left: PSIGQSAT. Right: ICE_CLD_WGT

Figure 4: Scaled SPP perturbation pattern fields and corresponding histograms at hour 0 for parameters (left) PSIGQSAT and (right) ICE_CLD_WGT from ExpA (top, DP) and ExpC (bottom, SP) in Table 3.

The cause of the DP and SP pattern differences were subsequently investigated in detail, and upon consultation with Ole Vignes (Met Norway), various source changes were made, particularly to the "RANDOM_NUMBERS_MIX" routine which is heavily used in the SPP pattern initialisation and evolution. The changes are discussed in more detail in Appendix 7.1.3. In Figure 6 we consider the difference in SP and DP perturbations patterns before and after these changes were implemented (see experiments ExpA, ExpC, ExpD, and ExpE in Table 3). The SPP perturbation patterns now appear to be essentially identical upon switching precision, modulo the small-scale "reproducibility noise" which was previously discussed in Section 2.4. This is also evidenced in Figure 5 for the bulk pattern statistics over all leadtimes, where the behaviour of the DP reference run (ExpA) is now recovered in SP (ExpE). Note from Figure 5 that the changes introduced in Appendix 7.1.3 have no impact when running in DP.

While these source changes clearly have a desirable impact on reducing the differences between SP and DP generated SPP patterns, they do not entirely alleviate the issue. All of the technical SPP experiments discussed above utilise a pattern update frequency of every hour, which is used operationally by MetCoOp to reduce computational cost (see Section 4.5). However this differs from

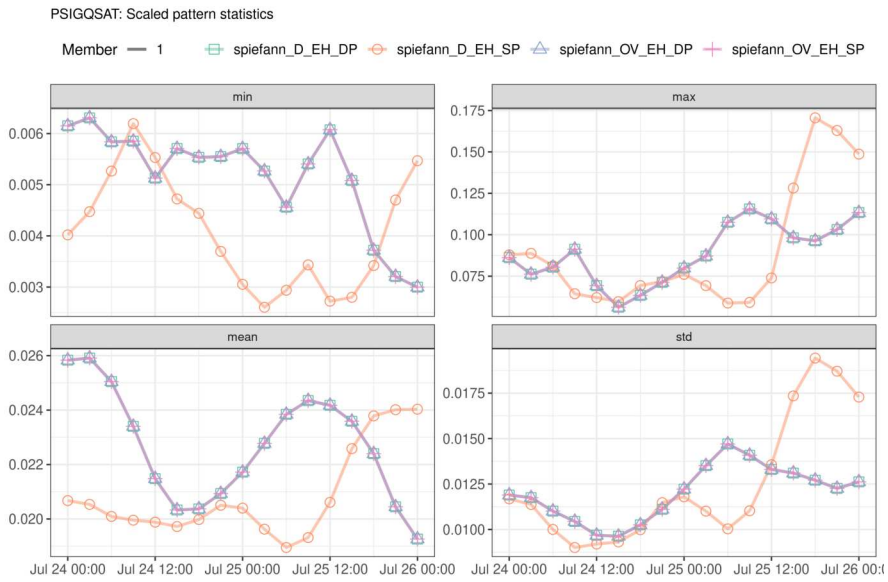
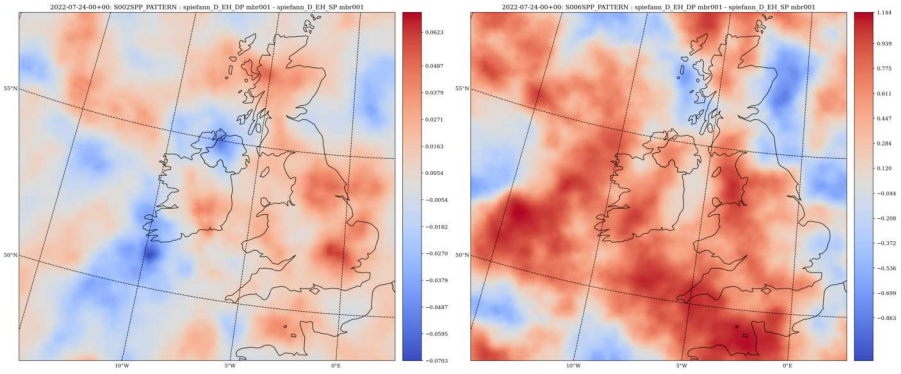
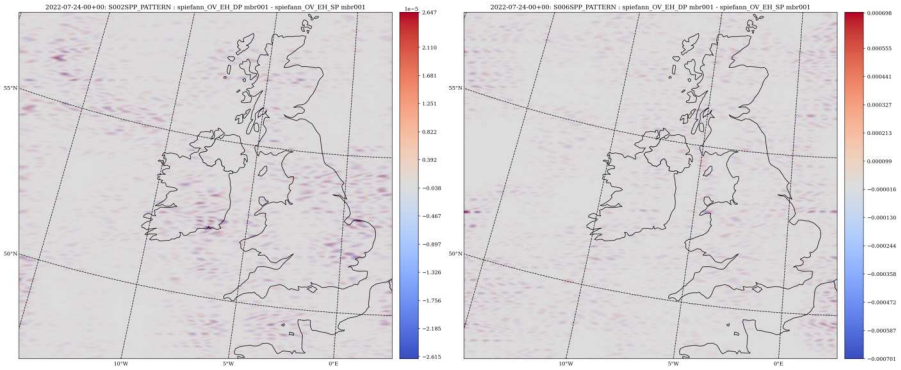


Figure 5: Minimum, maximum, mean, and standard deviation of the scaled PSIGQSAT SPP perturbation pattern as as function of leadtime for member 1 of ExpA (green), ExpC (orange), ExpD (blue), and ExpE (magenta) in Table 3. Symbols are included to aid visualisation of overlapping lines.



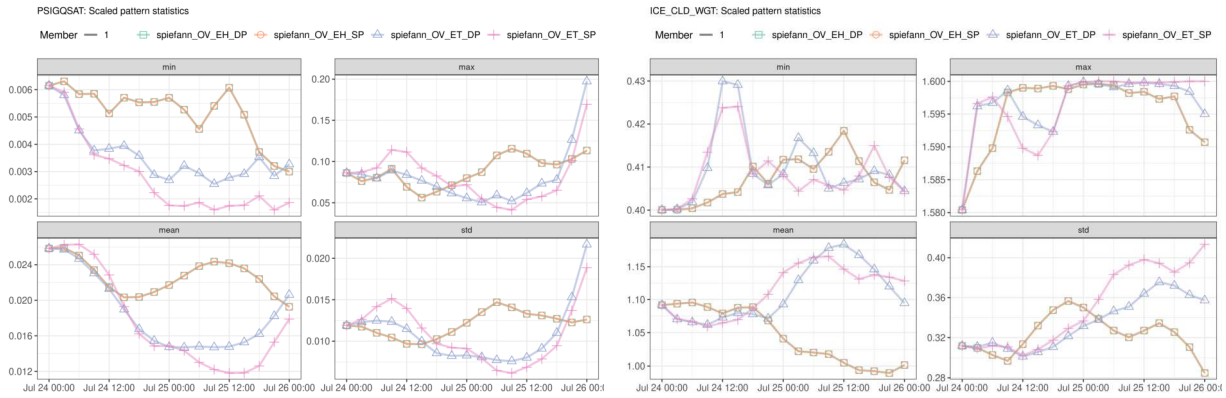
(a) ExpA - ExpC (DP - SP): Before changes in Appendix 7.1.3.



(b) ExpD - ExpE (DP - SP): After changes in Appendix 7.1.3.

Figure 6: Differences at hour 0 for the scaled (left) PSIGQSAT and (right) ICE_CLD_WGT perturbation patterns. The top and bottom rows indicate ExpA - ExpC and ExpD - ExpE, respectively, in Table 3. Experiments ExpD and ExpE include the additional changes discussed in Appendix 7.1.3.

the default choice of every timestep (NPATFR_SPP=1) in harmonEPS-43h2.2. If we now repeat experiments ExpD and ExpE, which include the RANDOM_NUMBERS_MIX routine changes, with a pattern update frequency of every timestep, we now observe a divergence in the SP and DP patterns. This is clearly illustrated in Figure 7, which is the same as Figure 5 but includes ExpF and ExpG in Table 3. While the SP and DP SPP patterns are initially almost identical, they quickly diverge over time when the SPP patterns are updated every timestep. In this particular case, the differences are relatively small for the PSIGQSAT parameter; however, this does not hold in general, and indeed the differences in the variance for parameter ICE_CLD_WGT, also included in Figure 7, are quite significant.



(a) PSIGQSAT (b) ICE_CLD_WGT

Figure 7: Minimum, maximum, mean, and standard deviation of the scaled PSIGQSAT (left) and ICE_CLD_WGT (right) SPP perturbation patterns as as function of leadtime for member 1 of ExpD (green), ExpE (orange), ExpF (blue), and ExpG (magenta) in Table 3. Symbols are included to aid visiualisation of overlapping lines.

Additional technical changes were tested in an attempt to address this issue, however none of these were ultimately successful during the course of the project. Therefore, in general the SPP patterns are precision-dependent, however this dependence can be masked when using a longer pattern update interval. As such, the source changes outlined in Appendix 7.1.3 represent a "partial fix" to the SP SPP perturbation patterns, and will be referred to as "partial pattern fixes" henceforth. This update interval dependence is quite curious, and highlights the need for additional changes to obtain pseudo-identical SPP patterns in SP and DP regardless of pattern update frequency (see the discussion in Section 6).

Finally, one may question the importance of having precision-independent SPP perturbation patterns and if the partial pattern fixes in Appendix 7.1.3 are really essential. This is particularly true in the context of an ensemble experiment, where differences in individual SP and DP SPP patterns and forecasts are tolerable provided that the overall pattern statistics and behaviour are similar. This question will be revisited in detail in Section 5.3.1.

3 Single precision performance with default perturbations

In this section we provide a summary of SP versus DP performance using the default suite of perturbations in HARMONIE-AROME; namely lateral boundary conditions via SLAF, surface perturbations, and EDA. As previously discussed in Section 1, these model perturbation methods occur outside of the Forecast step, and are hence always run in DP. This comparison allows one to assess both SP forecast stability and provide a benchmark for SP forecast quality relative to DP before SPP perturbations are switched on.

3.1 Experiment and verification details

Two experiment configurations will be considered; one a reference DP ensemble experiment (referred to as "DPert_DP" here), and the other a SP version of this (referred to as "DPert_SP"). For DPert_DP, we use the following settings:

- Binary versions C_2 or A_1 (cca and Atos, respectively) in Table 8 with FP_PRECISION=double.
- ENSMSEL=0-6, SLAF (see Table 1), PERTATMO=CCMA, PERTSURF=model, SPP=no.
- 48hr forecasts at 00 UTC, with 3hr cycling otherwise.

Experiment DPert_SP is identical to the above with the exception of FP_PRECISION=dual. The common settings described in Table 1 are also assumed and the same boundary and observation files are used in each experiment. A 0+6 ensemble size is used following standard practice within the ACCORD consortium, which is deemed "large enough" to provide meaningful ensemble statistics. Each experiment is carried out over a two-week period in each season, as detailed in Table 4. The use of multiple seasons allows us to gauge SP performance and stability in different meteorological conditions. The test periods chosen also contain various meteorological cases of interest, such as three named storms in February 2022 and a mini-heatwave over Ireland in July 2022. A five-day period is used to spin-up the DP control member of DPert_DP for each season, with each member in DPert_DP and DPert_SP initiated from this spun-up DP control at the start of each two-week period.

Season	Period	Conditions
Spring	Mar 28th - Apr 10th 2022	Several fog cases
Summer	Jul 11th-24th 2022	Mini-heatwave and thunderstorms
Autumn	Oct 16th-29th 2021	Wet spell
Winter	Feb 10th-23rd 2022	Storms Dudley, Eunice, and Franklin

Table 4: Two-week testing periods considered for the longer cycling experiments.

Point verification of all experiments discussed in this note was carried out using the harp package developed by the ACCORD consortium (`https://github.com/harphub/`). In this section, the primary purpose of point verification is to assess the performance of DPert_SP relative to DPert_DP, i.e. the gauge the impact of SP forecasts on overall ensemble performance, and not necessarily the overall meteorological performance of the experiments. As such, it is reasonable to use all synoptic stations available in the IRELAND25 domain for verification purposes.

Local Met Éireann vobs files (i.e. those used in the verification of our operational systems) are utilised in the verification process for all variables except accumulated precipitation, where instead we use the vobs files from the CAMP climate network of stations (covering ~ 80 stations over Ireland). Scorecards are used significantly throughout this note in order to conveniently summarise the performance of experiment X relative to experiment Y, e.g. DPert_SP relative to DPert_DP. The statistical significance of score differences between the two experiments is calculated using 1000 bootstrap replicates with observation/forecast data pooled by each forecast start date. Fair ensembles scores (such as the fair CRPS) are computed by scaling relative to an infinite number of ensemble members (i.e. num_ref_members = Inf in harpPoint::ens_verify).

3.2 Verification results

3.2.1 Surface

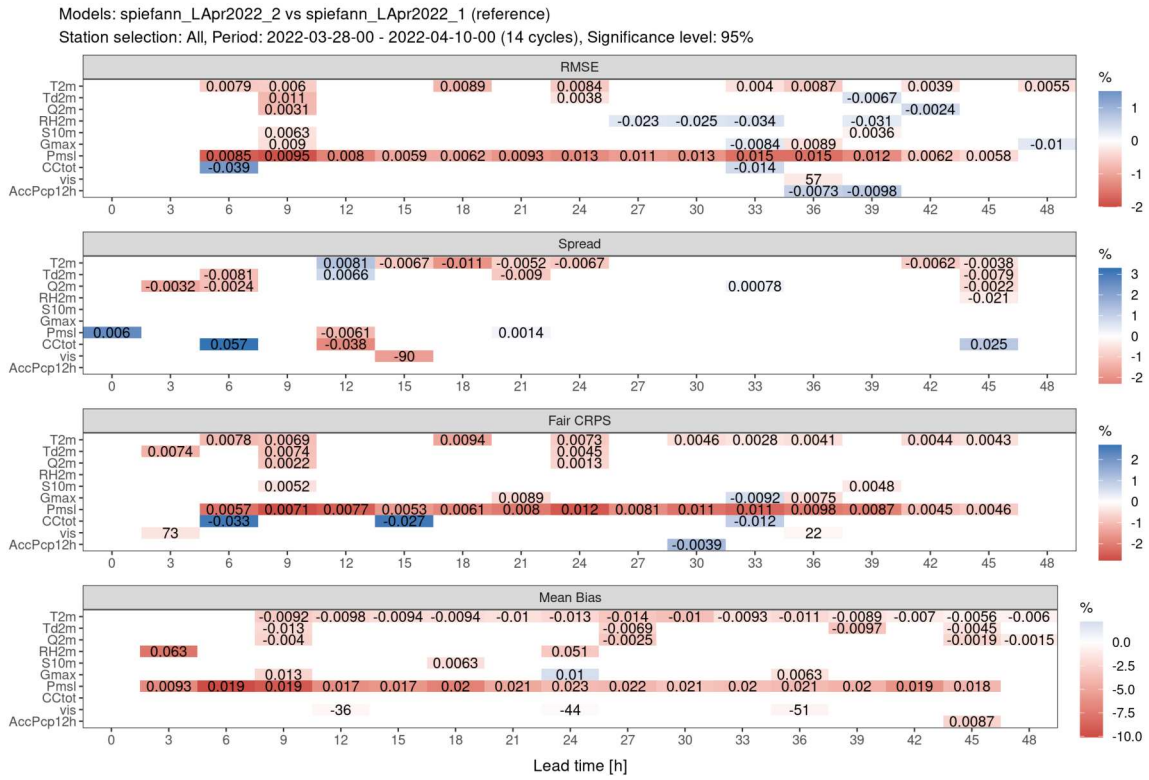
The statistical significance of surface parameter score differences between the SP and DP ensembles for the spring test period are illustrated in Figure 8. The scorecard should be read as follows:

- The surface parameters indicated in each row of the heatmap are 2 m temperature, 2 m dew point temperature, 2 m specific humidity, 2 m relative humidity, 10 m wind speed, 10 m max gust, PMSL, total cloud cover, visibility, and 12hr accumulated precipitation.
- Filled tiles indicated where score differences are significant at the 95% level. Blank squares can thus be interpreted as areas where the score differences are not statistically significant.
- The colour bar for each metric represents $100 \times (X - Y)/Y$, where X and Y are the mean ensemble scores for the SP and DP experiments, respectively i.e. the percentage difference of the mean SP score relative to DP. Blue indicates an improvement in the SP ensemble score, while red is a degradation. Note the separate color bars for each score.
- The number in each statistically significant tile represents $X - Y$, i.e. the actual difference in the mean scores for the SP ensemble relative to DP.

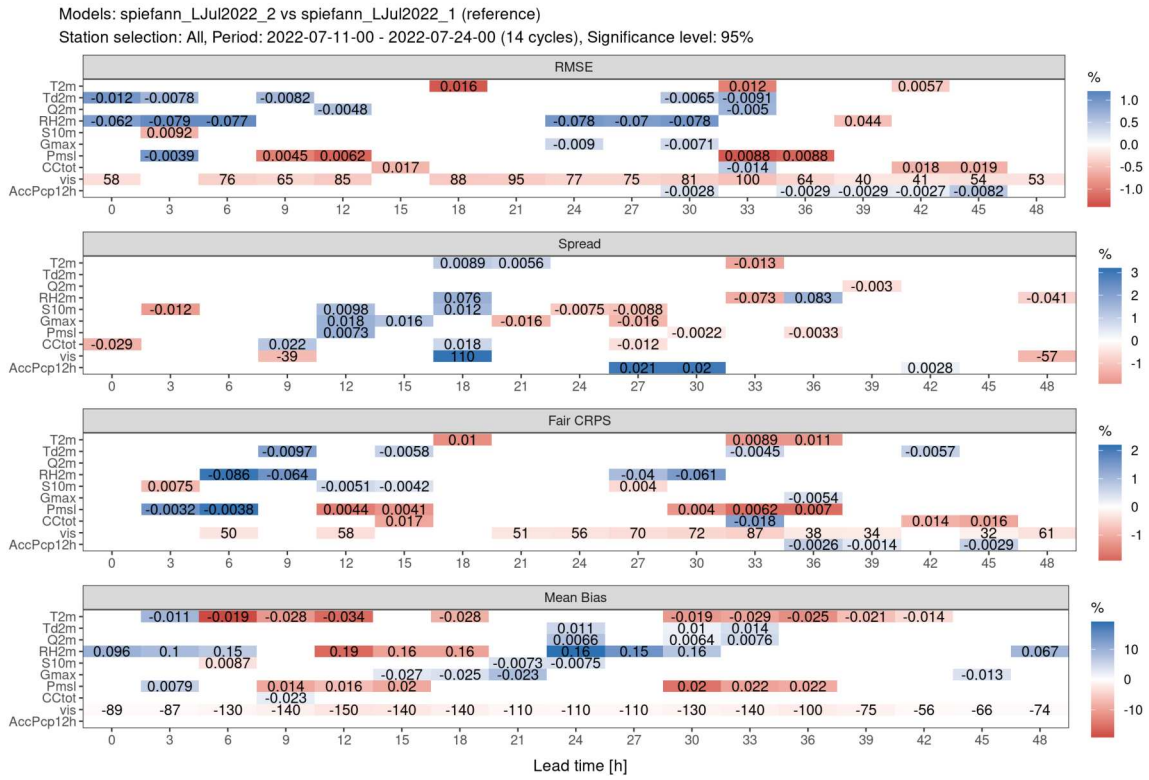
This methodology for representing the statistical significance has a number of advantages over the "standard" scorecard plotting tool from harp; namely, an indication of the actual score differences along with the relative importance of these differences. As such, it provides a more comprehensive data summary and reduces the need for visualisation of other surface scores.

As illustrated by the general sparsity in Figure 8(a), the DPert_DP and DPert_SP ensembles yield very similar results for the this spring period. The only significant signal present is a degradation in SP PMSL scores relative to DP. In particular, the SP ensemble has an additional positive PMSL bias of ~ 0.02 hPa, which commensurately degrades the RMSE and CRPS scores. Note, however, that the overall impact on PMSL is still relatively small (i.e. limited to $\sim 2 - 5\%$ generally). In addition, the SP ensemble also possess a very small cold bias relative to DP; however this is typically limited to ~ 0.03 K.

The same scorecards for the summer, autumn, and winter test periods are given in Figures 8 and 9. Across all periods we again see relatively little impact on the bulk ensemble scores when running in SP. The additional positive PMSL bias of ~ 0.02 hPa, and the very slight cold bias, in the SP ensemble are generally found to be the only consistent signals.

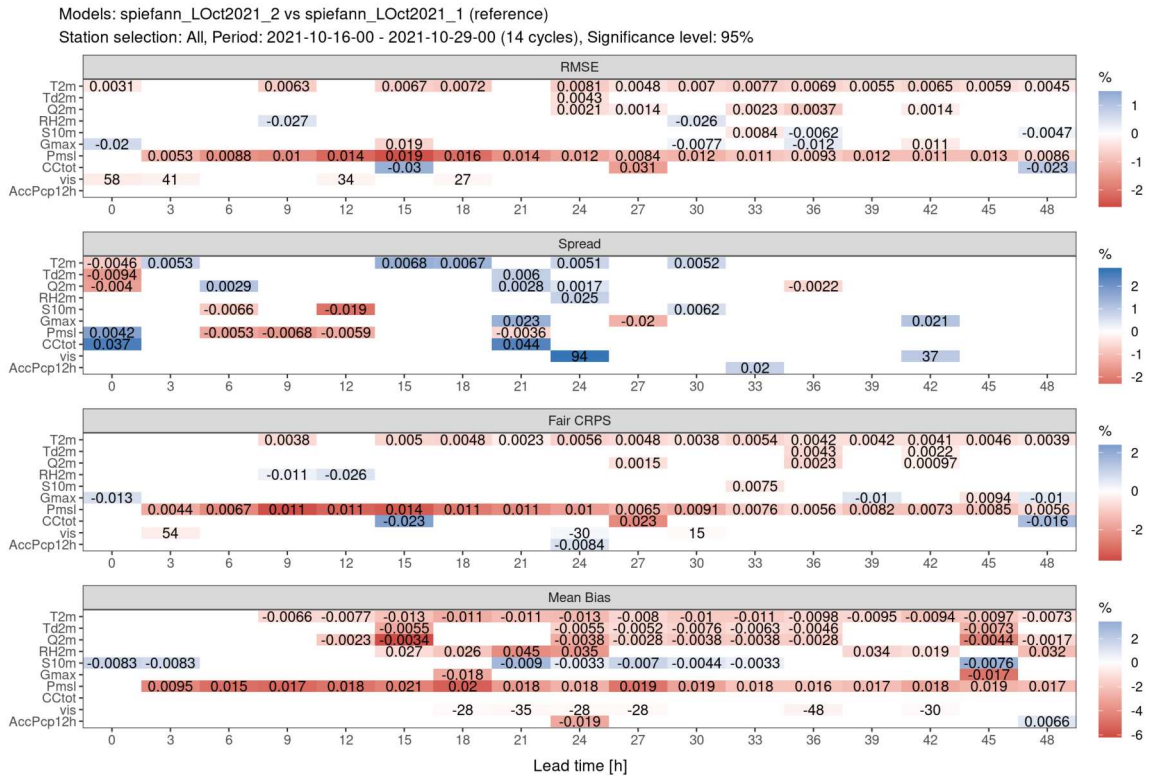


(a) Spring

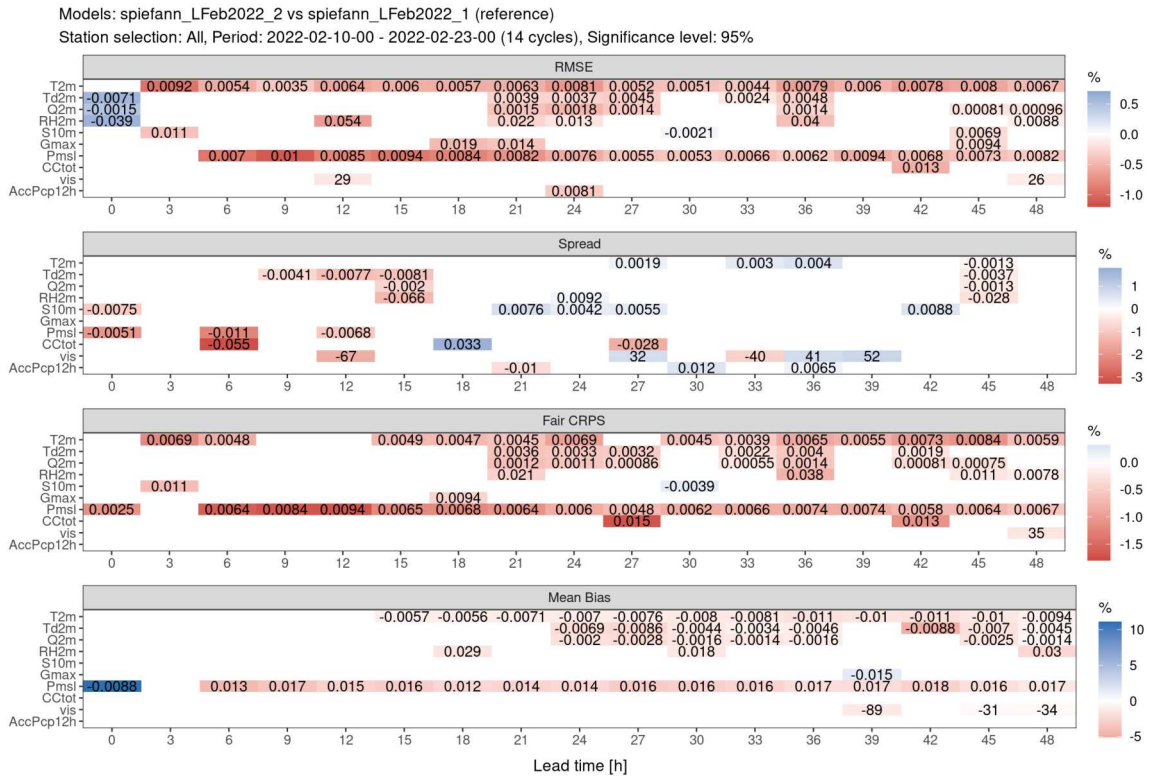


(b) Summer

Figure 8: Scorecard illustrating the differences in ensemble RMSE, spread, fair CRPS, and bias for DPert_SP versus DPert_DP (reference) over the spring and summer test periods. See text for additional details.



(a) Autumn

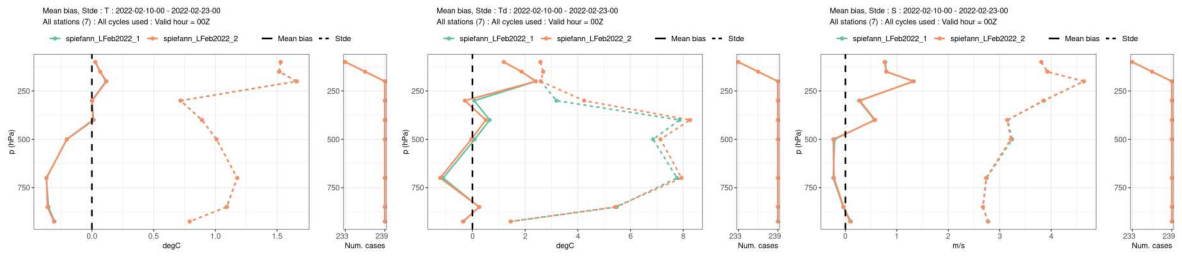


(b) Winter

Figure 9: Scorecard illustrating the differences in ensemble RMSE, spread, fair CRPS, and bias for DPert_SP versus DPert_DP (reference) over the autumn and winter test periods. See text for additional details.

3.2.2 Upper Air

Indicative ensemble mean bias and standard deviation upper-air profiles for temperature, dew point temperature, and wind speed over the stormy winter test period are given in Figure 10. It is first important to note that as the number of radiosonde launch sites in the IRELAND25 domain is quite small (i.e. only 7 stations), the number of observations available for verification over a two-week test period is quite small ($\sim$ 225 in Figure 10). This small number of cases clearly restricts what conclusions can be made regarding the statistical significance of score differences between the SP and DP ensembles, and this should be considered when interpreting the upper-air verification results.



(a) Temperature (b) Dew point (c) Wind speed

Figure 10: Upper-air profiles valid at 00 UTC for temperature, dew point, and wind speed for DPert_DP (green) and DPert_SP (orange) over the winter test period.

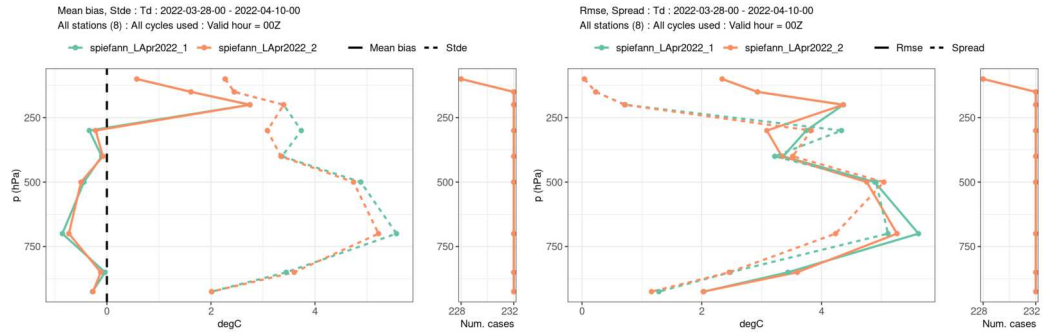
With that in mind, Figure 10 suggests that SP does not appreciably impact the upper-air temperature and wind speed profiles. A slight difference between the DP and SP experiments is evident for dew point, and this signal is also present for the other test periods, as illustrated in Figure 11. This possible impact on upper-air humidity is consistent with previous findings (e.g. Bessardon et al. (2021), Fannon and Hally (2021), and Feddersen (2021)). Thus a more rigorous analysis of upper-air humidity changes due to SP may be justified, however it falls beyond the scope of this work. Temperature and wind speed results for the other test periods are also very similar in SP and DP (not shown).

3.3 Runtime savings

For a given ensemble member x, the Forecast runtime saving achieved by using SP, denoted as RS_x, can simply be defined as

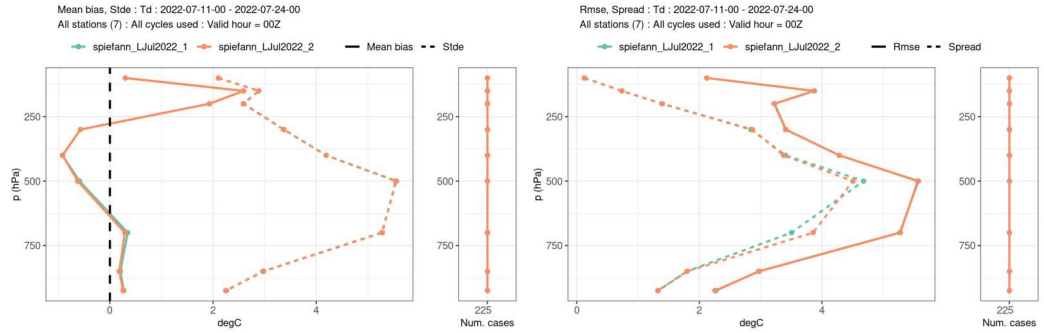
$$\mathrm{RS}_x = 100 \times \left(\frac{\text{Runtime for DP member } x - \text{Runtime for SP member } x}{\text{Runtime for DP member } x} \right), \tag{2}$$

Member runtimes are evaluated from the ECMWF job epilogue. In Figure 12 we illustrate RS_x for all ensemble members in experiments DPert_DP and DPert_SP for the summer and winter periods in Table 4. Only the 48 hour forecasts at 00 UTC are considered. Consistent runtime savings of between 30-40% are observed for all members, again in-line with previous findings. It is also interesting to note a larger saving for the summer experiment, which was run on Atos, compared to the winter experiment, which was run on cca.



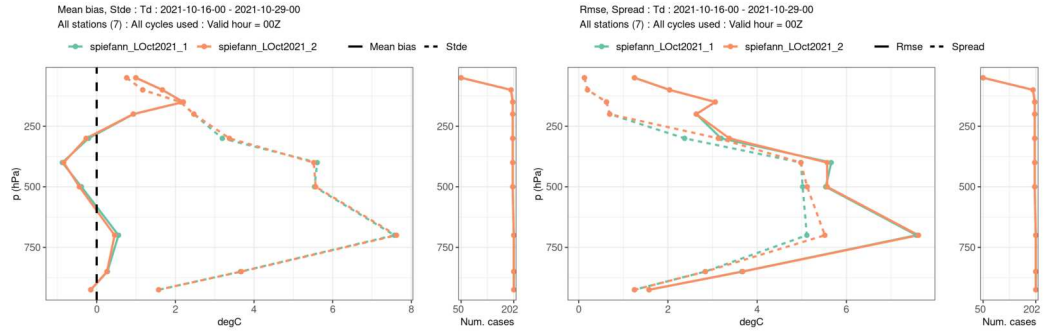
(a) Mean bias and STDE

(b) Spread Skill



(c) Mean bias and STDE

(d) Spread Skill



(e) Mean bias and STDE

(f) Spread Skill

Figure 11: Upper-air dew point profiles valid at 00 UTC for (left) mean bias and STDE and (right) spread skill for DPert_DP (green) and DPert_SP (orange) over the winter test period.

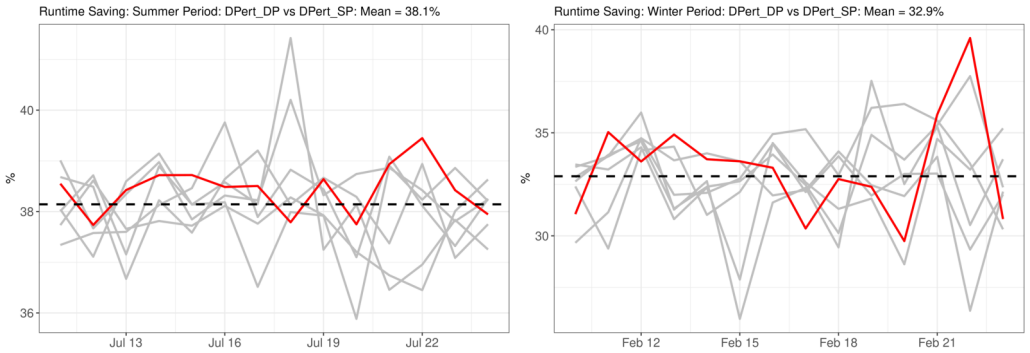
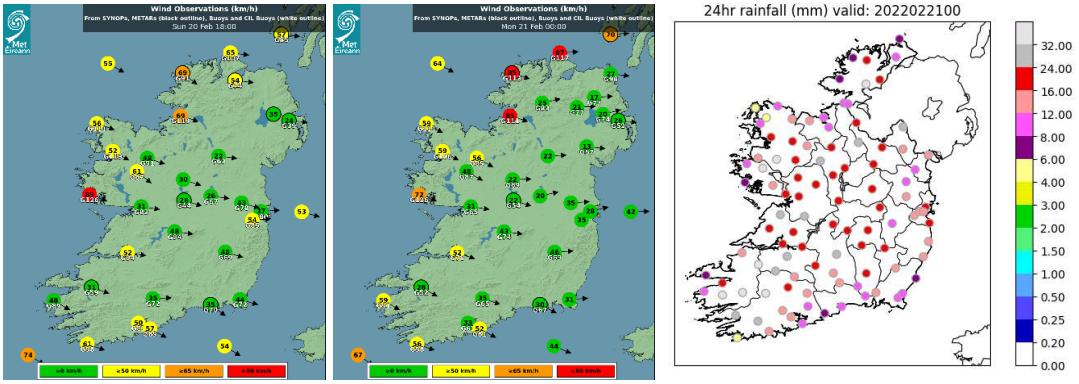


Figure 12: RS_x, as defined by Equation 2, comparing DPert_DP and DPert_SP for the summer (left) and winter (right) periods in Table 4. The control member is given in red, with perturbed members in gray. Note that the summer and winter periods were run on Atos and cca, respectively.

3.4 Sample forecasts for Storm Franklin

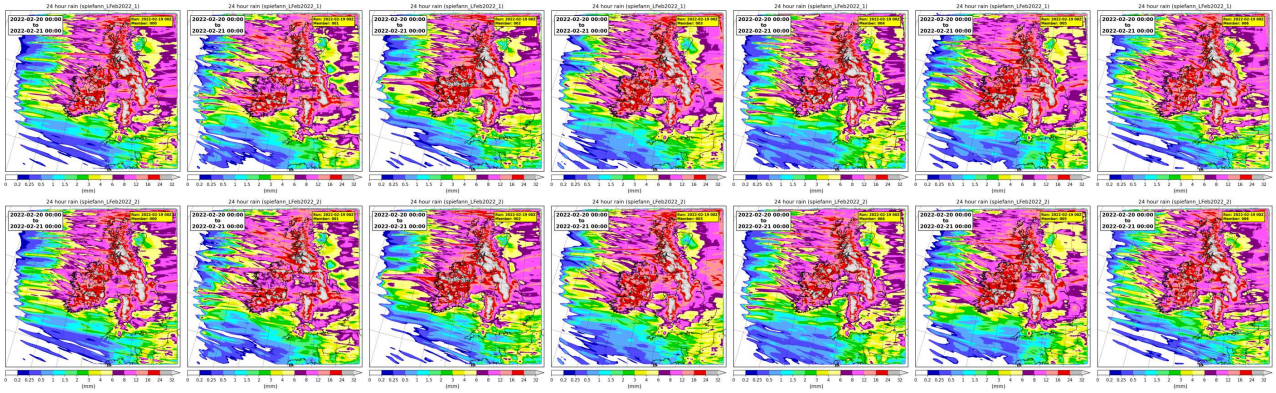
To conclude this section we briefly compare sample DPert_DP and DPert_SP forecasts for a severe weather event (Storm Franklin) which impacted Ireland on February 20th 2022, in which widespread yellow to red level winds were observed along with significant rainfall accumulations, as indicated in Figure 13. Once again, for this comparison we are not necessarily interested in the performance of HARMONIE-AROME for this event, but more-so a sanity check that there are no concerning discrepancies in the SP forecasts relative to DP.



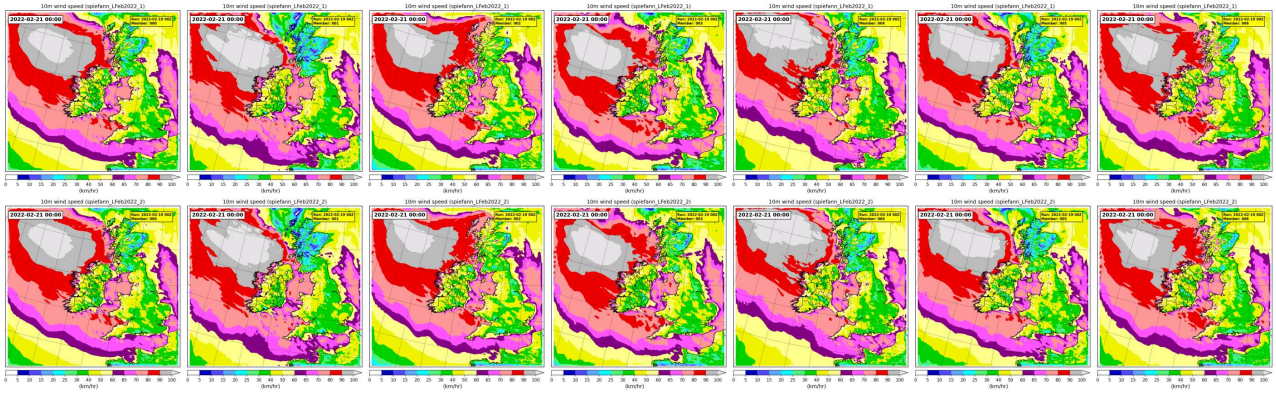
(a) Feb 20th 1800 UTC (b) Feb 21st 0000 UTC (c) Feb 21st 0000 UTC

Figure 13: (a)-(b) Observed wind speeds for Storm Franklin at the valid times indicated. (c) 24hr rainfall valid at February 21st 0000 UTC at all Synop and CAMP stations over Ireland.

Forecasts for 10 m wind speed and 24hr rainfall valid at February 21st using the 2022/02/19/00 cycle for experiments DPert_DP and DPert_SP are given in Figure 14. At a leadtime of 48 hours, the DP and SP forecasts are reassuringly similar for each member, even for a discontinuous variable such as precipitation.



(a) 24 hour rainfall



(b) 10 m wind speed

Figure 14: (a) 24 hour rainfall and (b) 10 m wind speed using the 2022/02/22/00+48 forecasts for experiments DPert_DP (top row) and DPert_SP (bottom row). Members 0-6 are given from left to right.

4 SPP experiments in double precision

In this section we extend the default reference ensemble experiment, with SLAF, EDA, and surface perturbations only, to include SPP perturbations. We first focus on the meteorological impact of SPP when running in DP to assess the potential benefits which this new perturbation scheme over the IRELAND25 domain. SP SPP experiments will be discussed in detail in Section 5.

4.1 Experiment details

We use DPert_DP, as detailed in Section 3.1, as the reference/control experiment throughout. All DP SPP experiments share the same settings as DPert_DP except for SPP=yes in config_exp.h. Experiments for both the MetCoOp and KNMI SPP configurations are carried out, as detailed in Table 5. Note that a pattern update frequency of every timestep is used for the majority of experiments discussed, which was done for consistency with the SP SPP experiments (see Section 5.1 for justification). However, a winter period run with a pattern update frequency of every hour was also carried out for comparison purposes. The four testing periods in Table 4 were again used and each ensemble member was warm-started from the corresponding spun-up DPert_DP control member (as discussed in Section 3.1).

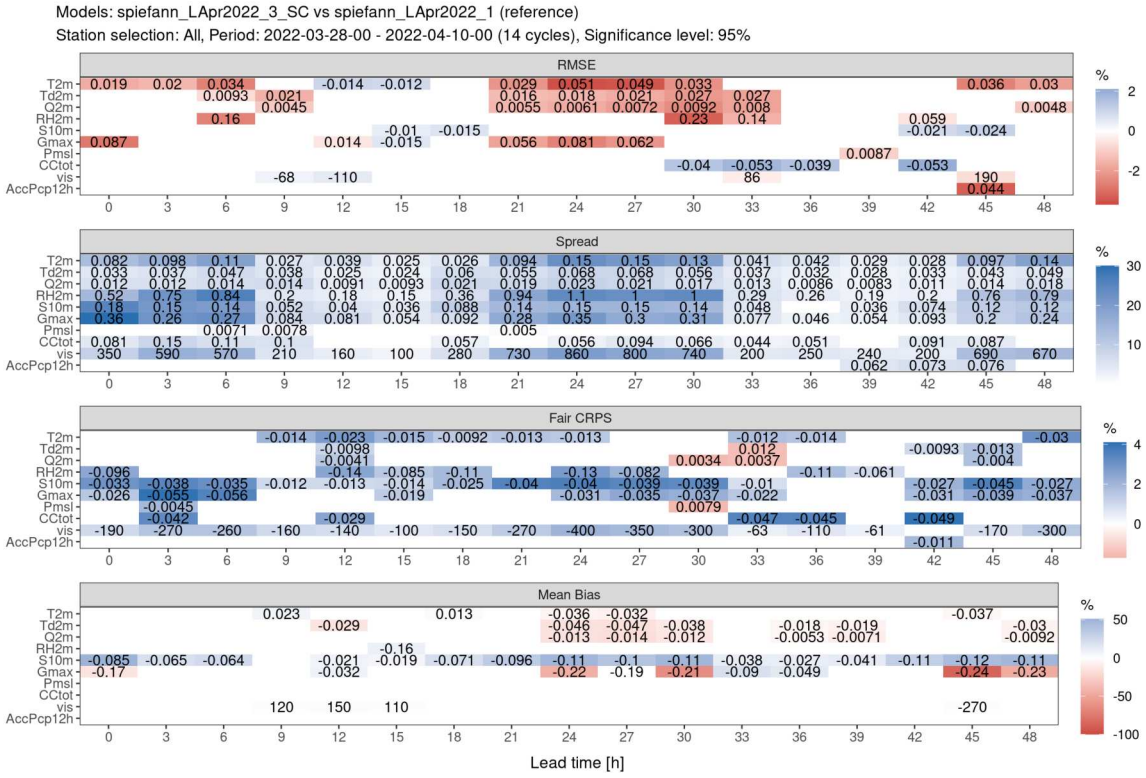
Name	SPP config.	NPATFR_SPP	Test period(s)
SPP_ATET_DP	5PAT	1	All seasons
SPP_MCET_DP	5PMC	1	All seasons
SPP_MCEH_DP	5PMC	-1	Winter

Table 5: DP SPP experiments considered in this note. All settings are identical to DPert_DP (Section 3.1) but with SPP=yes, and SPP configurations are those given in Table 2. Note that the test periods are those given in Table 4 and all experiments utilised either binary versions C_2 or A_1 (Table 8).

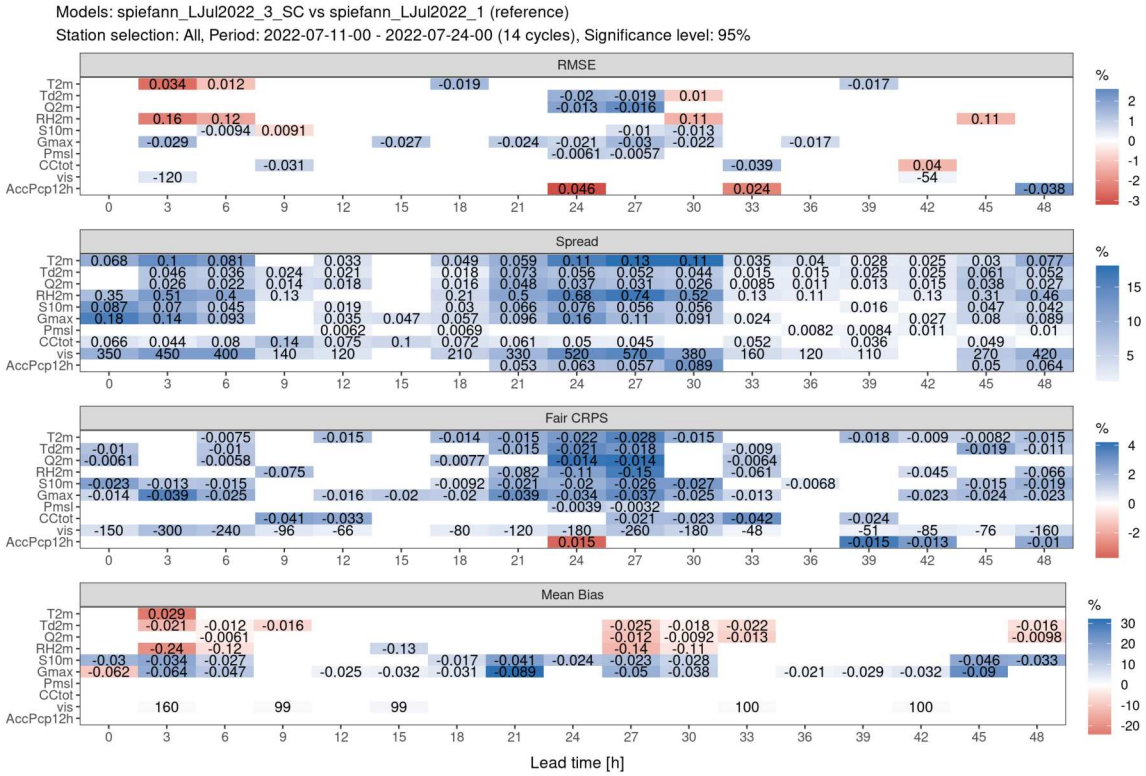
4.2 Results for KNMI SPP configuration

We first consider the surface verification results using the KNMI SPP configuration, i.e. experiment SPP_ATET_DP in Table 5. The surface parameter scorecard comparing SPP_ATET_DP to the DPert_DP reference experiment for the spring test period is given in Figure 15. As expected, the introduction of the SPP scheme has a positive impact on overall ensemble spread for the majority of variables, with a sizeable relative increase of $\sim 10-20\%$ in some cases. A small but positive impact on CRPS is also evident, while nightime mean RMSE for 2 m temperature, humidity, and wind gusts are slightly degraded. However, the increase in model spread more than compensates for this slight degradation, resulting in a significant improvement in the spread-skill ratio for the majority of variables, as illustrated in Figure 17 (comparing green and blue lines).

Initial tuning experiments with SPP in HarmonEPS demonstrated that SPP can have an undesirable impact on ensemble mean bias, as discussed in Frogner et al. (2022). Considering the mean bias in Figure 17, we find very similar scores for SPP_ATET_DP compared to the reference apart from 10 winds and gusts. Considering the individual member biases for these parameters (Figure 18) indicates a general shift to lower wind speeds in SPP_ATET_DP. Indeed, when SPP is activated, perturbed members tend to be become more negatively biased relative to the control.

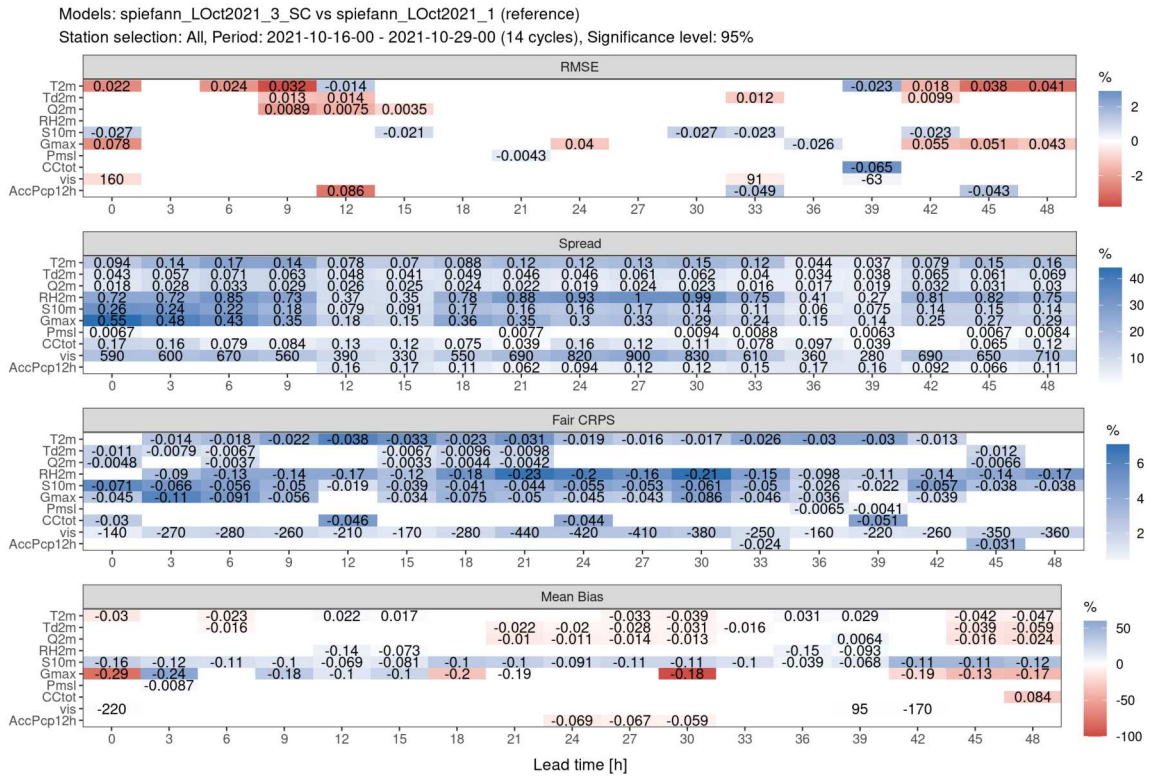


(a) Spring

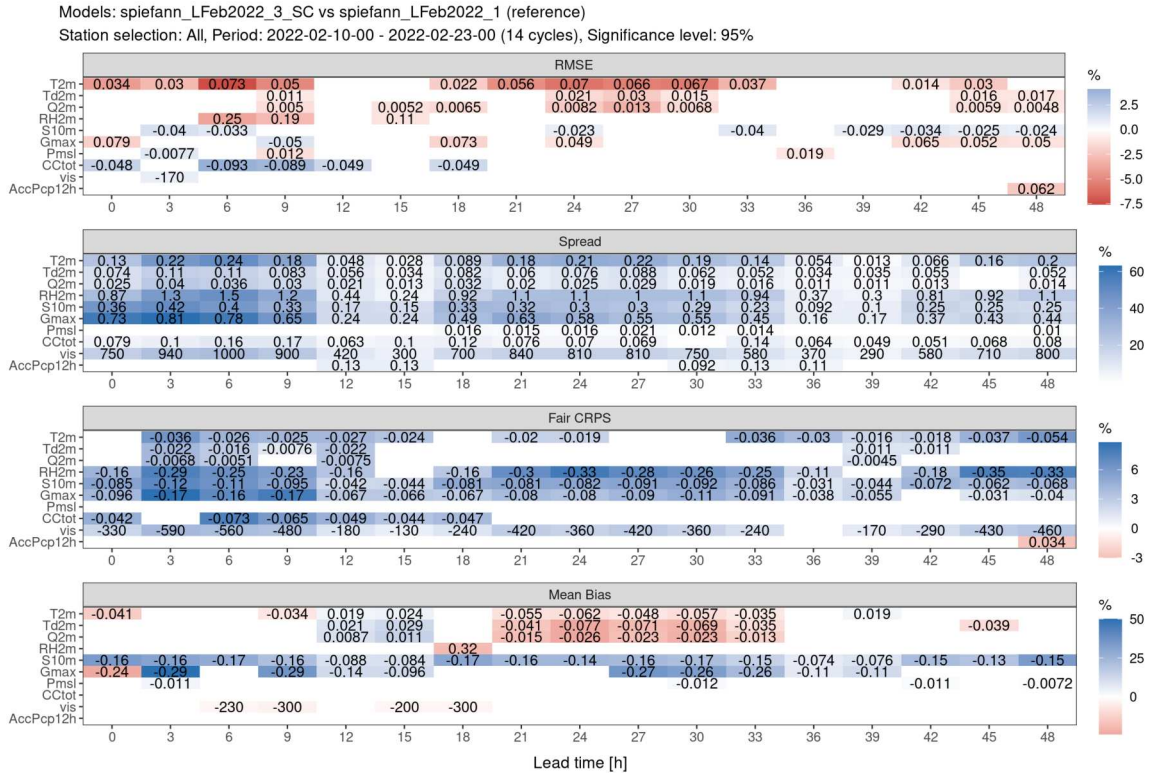


(b) Summer

Figure 15: Surface parameter scorecard comparing experiments SPP_ATET_DP to DPert_DP (reference) over the spring and summer test periods.

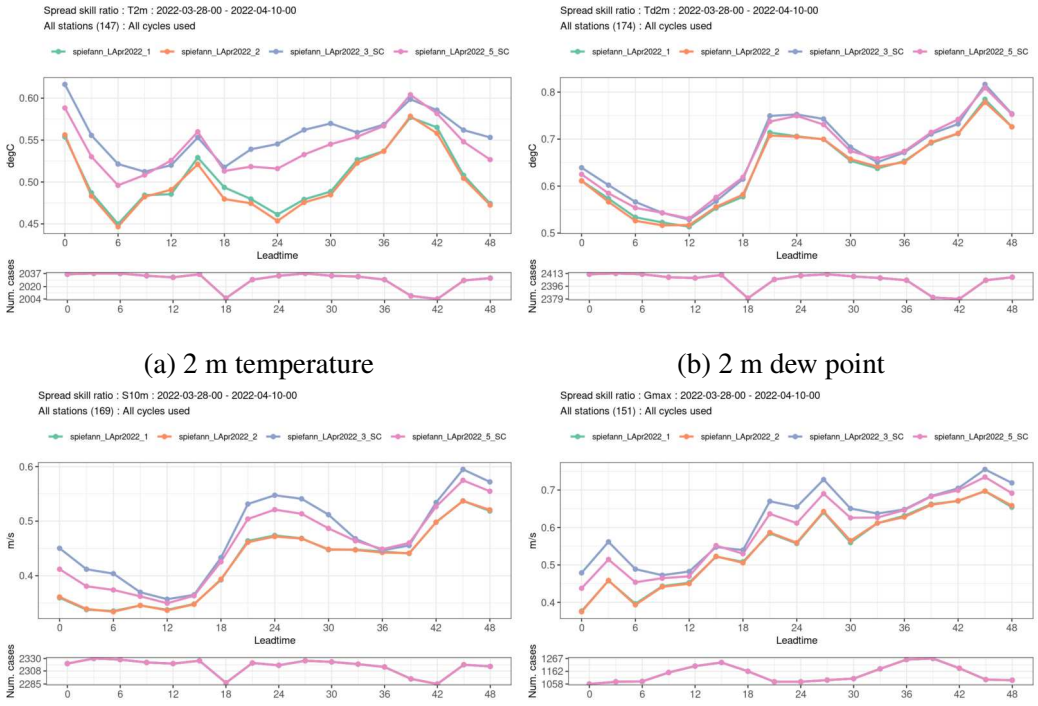


(a) Autumn



(b) Winter

Figure 16: Surface parameter scorecard comparing experiments SPP_ATET_DP to DPert_DP (reference) over the autumn and winter test periods.



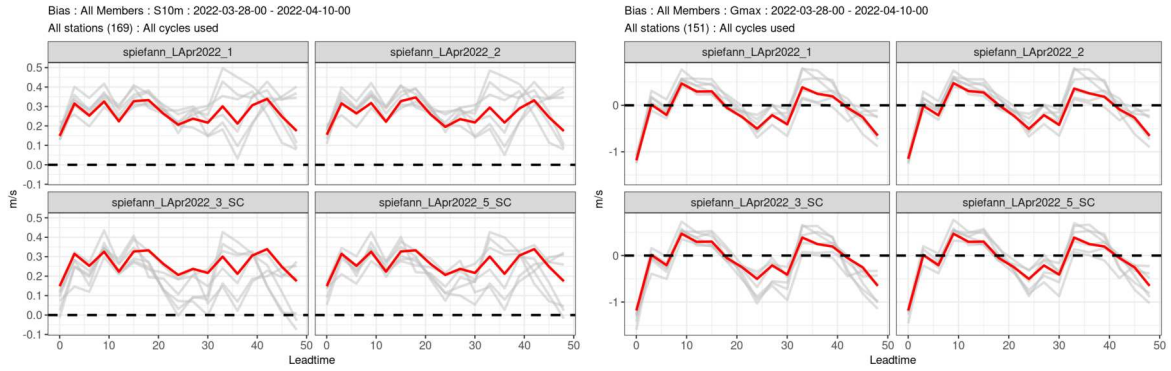
(a) 2 m temperature

(b) 2 m dew point

(c) 10 m wind speed

(d) 10 m wind gust

Figure 17: Spread-skill ratio for selected variables over the spring period for experiments DPert_DP (green), DPert_SP (orange), SPP_ATET_DP (blue), and SPP_MCET_DP (magenta).



(a) 10 m wind speed

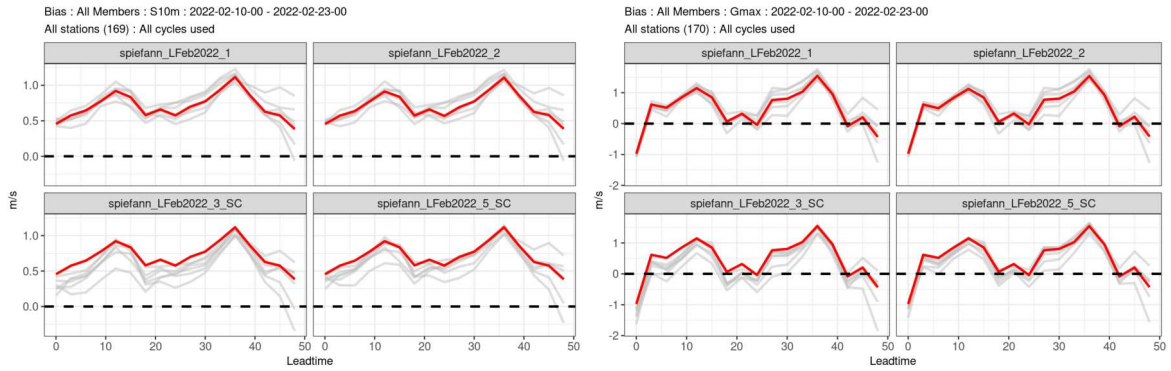
(b) 10 m wind gusts

Figure 18: Individual member biases over the spring test period for experiments DPert_DP (top left), DPert_SP (top right), SPP_ATET_DP (botoom left), and SPP_MCET_DP (bottom right).

Similar conclusions can be made from the scorecards over the summer, autumn, and winter test periods in Figures 15 and 16. A significant and consistent increase in ensemble spread for most surface variables is evident across each test period. This, coupled with a neutral to slightly negative impact on RMSE, leads to a significantly better spread-skill ratio in SPP_ATET_DP for each season. This is also clearly reflected in the improved CRPS scores.

Nighttime temperatures for the spring, autumn, and winter periods appear to be degraded slightly due to an additional small cold bias in SPP_ATET_DP. However the most consistent impact of SPP on the ensemble mean bias is clearly the introduction of a negative wind speed bias relative to the control experiment. This is particular evident in the results for the stormy winter period in Figure 16. While

the scorecard indicates an overall improvement in the mean bias, as HARMONIE-AROME typically overestimates wind over the IRELAND25 domain, this is due to the ensemble members becoming increasingly biased relative to the control. This is again illustrated for the winter period in Figure 19. As such, while SPP has the desired effect of decreasing the wind speed bias, it comes at the cost of "off-centering" the control member to a certain extent.



(a) 10 m wind speed

(b) 10 m wind gusts

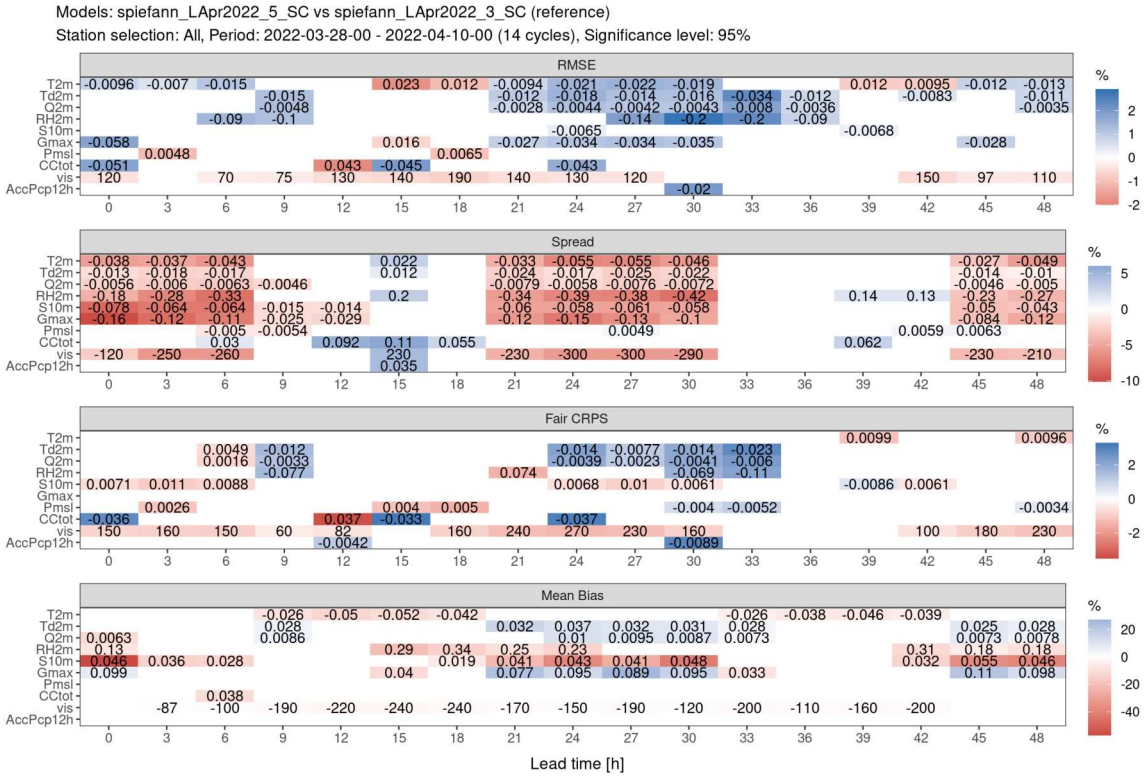
Figure 19: Individual member biases over the winter test period for experiments DPert_DP (top left), DPert_SP (top right), SPP_ATET_DP (botoom left), and SPP_MCET_DP (bottom right).

4.3 MetCoOp vs KNMI SPP configuration

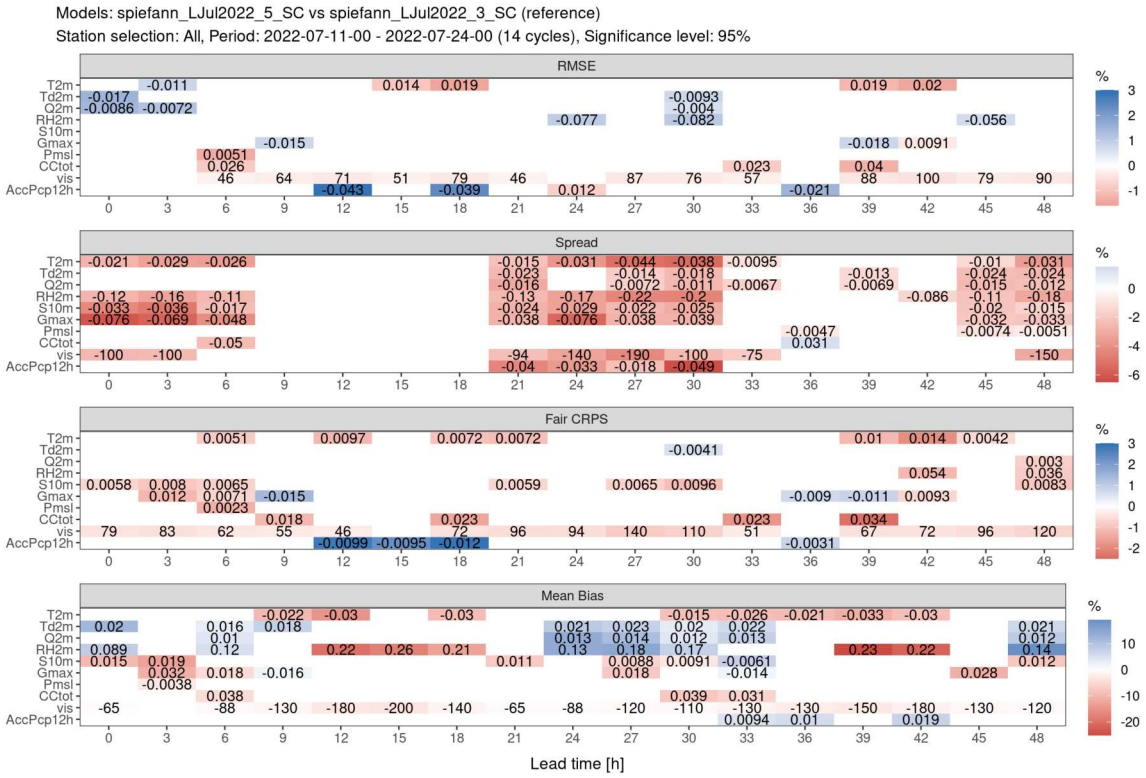
A similar analysis can also be performed for the MetCoOp SPP configuration ensemble, i.e. experiment SPP_MCET_DP in Table 5, with results presented in Appendix 7.2.1. On the whole one can draw similar conclusions to the SPP_ATET_DP experiment discussed in Section 4.2, i.e. a significant impact on overall ensemble spread, a small positive impact on CRPS, and a negative wind bias relative to the control experiment. This is of course to be expected given that the two SPP configurations have four common parameters. As such, it is instead useful to briefly contrast the performance of the two configurations over the IRELAND25 domain.

The surface scorecards comparing SPP_MCET_DP to SPP_ATET_DP for each test period are given in Figures 20 and 21. A reduction in ensemble spread for the MetCoOp configuration compared to KNMI is evident across each period, particularly at nighttime. While the RMSE for temperature and humidity is slightly improved for SPP_MCET_DP, the sizeable reduction in spread leads to a lower spread-skill relationship for the MetCoOp configuration in general. This is illustrated in the indicative plots over the spring test period in Figure 17.

CRPS scores for the two configurations are similar overall, with a tendency for slightly improved performance with the KNMI configuration. One can also note small differences in the mean biases for the two SPP configurations, i.e. the MetCoOp ensemble tends to produce slightly colder daytime temperatures, improved nighttime humidity, and higher wind speeds. The wind speed member biases for SPP_MCET_DP over the spring and winter periods are also indicated in Figures 18 and 19, respectively. An overall shift to lower wind speeds relative to the control is present, but to a slightly lesser extent than the KNMI configuration ensemble.

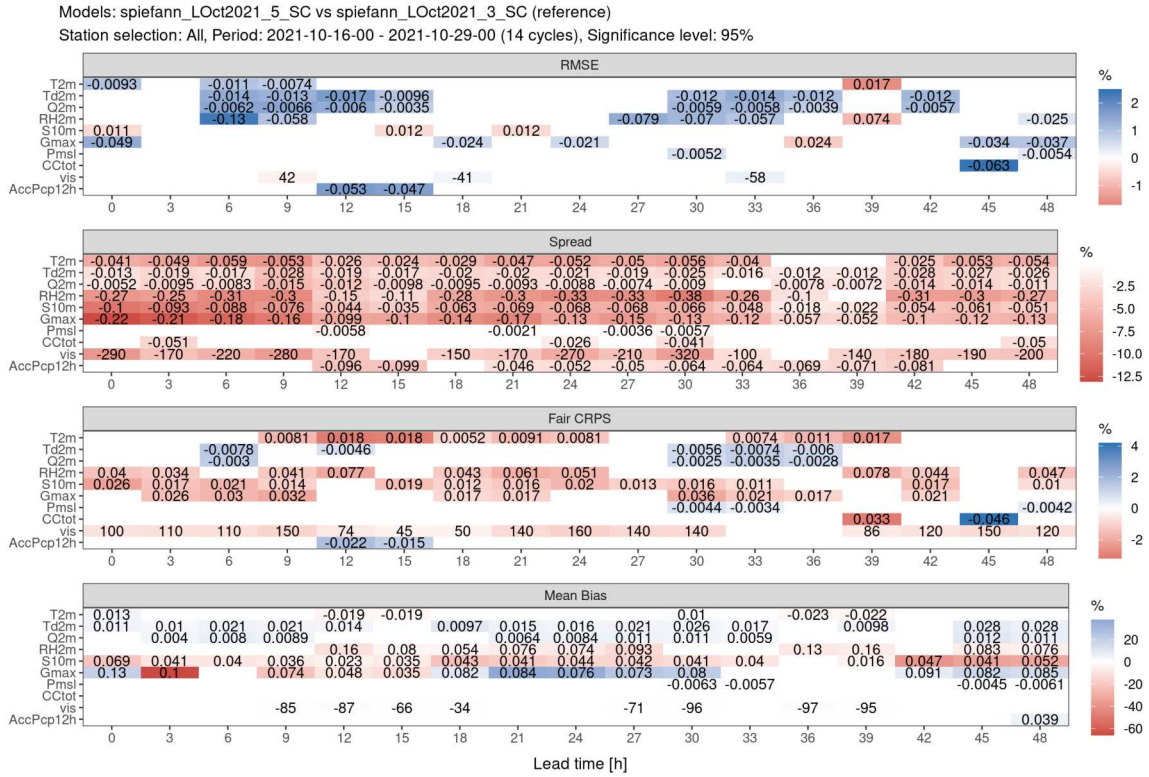


(a) Spring

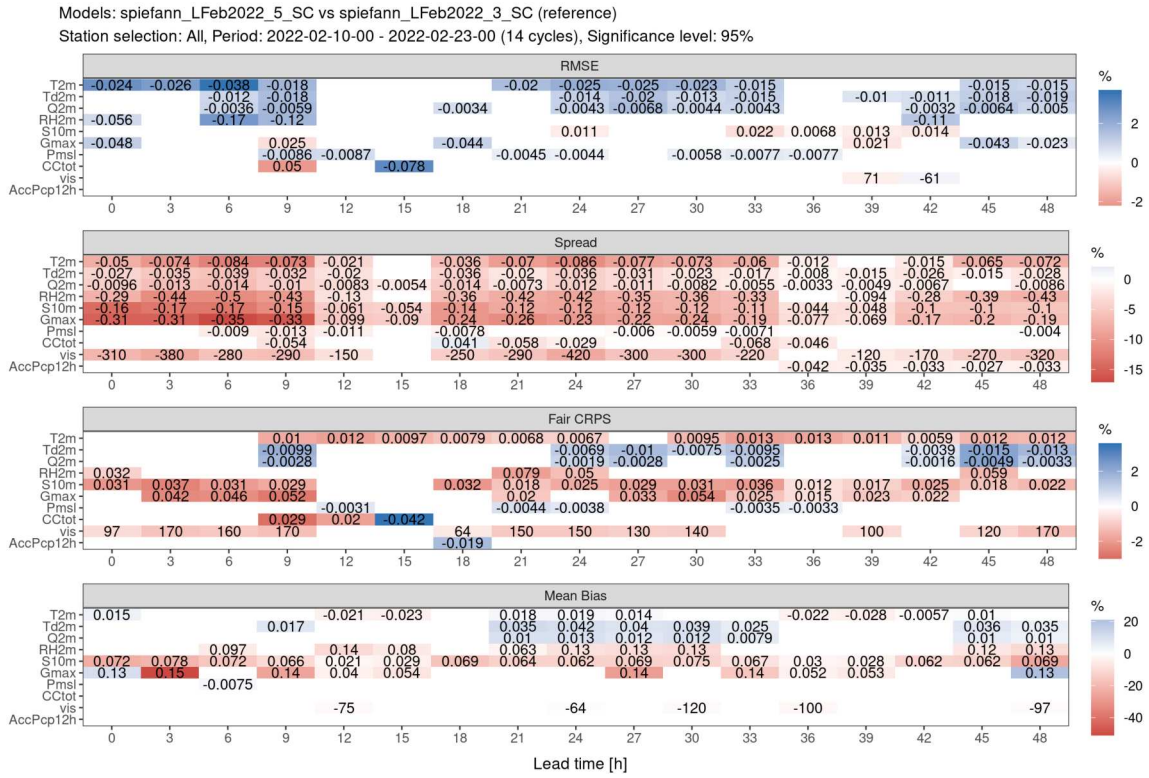


(b) Summer

Figure 20: Surface parameter scorecard comparing experiments SPP_MCET_DP to SPP_ATET_DP (reference) over the spring and summer test periods. Blue/red indicates an improvement/degradation for the MetCoOp configuration.



(a) Autumn



(b) Winter

Figure 21: Surface parameter scorecard comparing experiments SPP_MCET_DP to SPP_ATET_DP (reference) over the autumn and winter test periods. Blue/red indicates an improvement/degradation for the MetCoOp configuration.

4.4 Impact of hourly pattern updates

As discussed in Section 4.5, utilizing a pattern update frequency of every hour, as opposed to every timestep, reduces computational overhead associated with the SPP scheme. Indeed, this is the approach used in MetCoOp's operational implementation of SPP is used. To assess the meteorological impact of this switch, an additional MetCoOp SPP configuration run over the stormy winter test period was carried out using hourly pattern updates (experiment SPP_MCEH_DP in Table 5). The surface scorecard in Figure 22 provides a summary of model performance using pattern updates every hour versus every timestep. It is clear that changing the pattern update frequency has little to no impact on the overall surface verification scores, confirming that the update frequency can be changed without significantly impacting SPP performance.

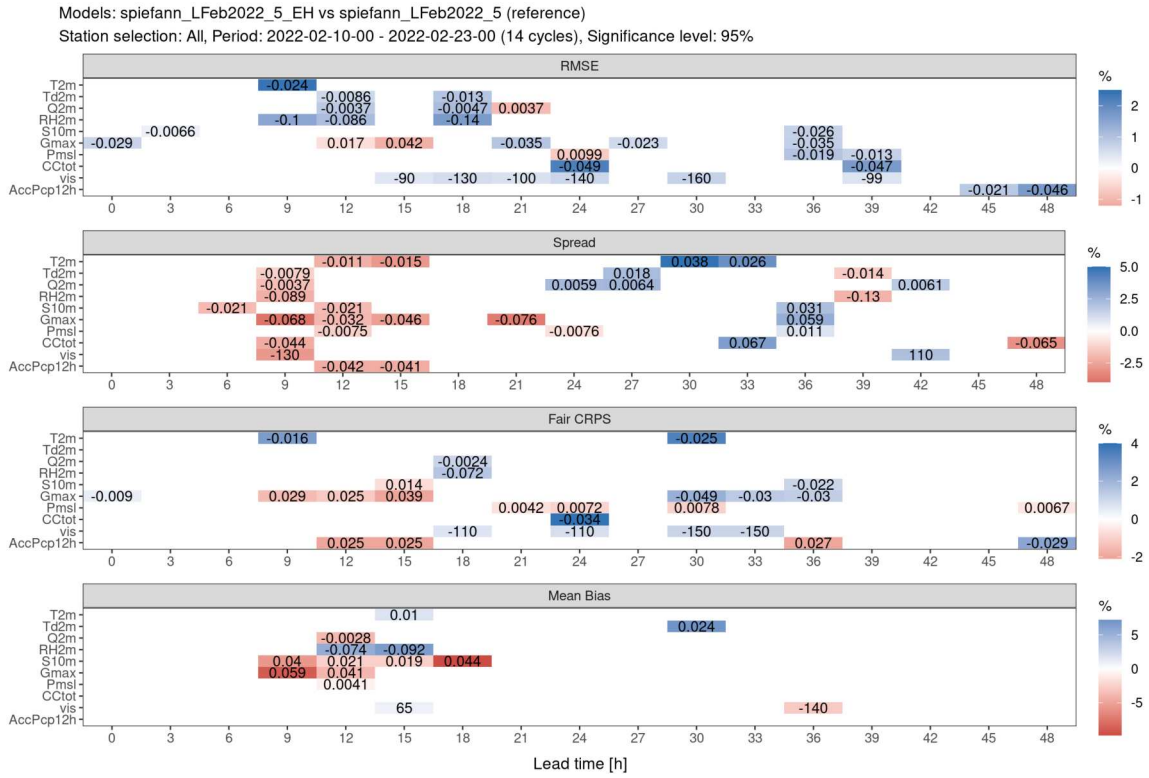


Figure 22: Surface parameter scorecard comparing SPP_MCEH_DP to SPP_MCET_DP (reference) for the winter test period. Experiments as indicated in Table 5. Blue/red indicates an improvement/degradation when using an hourly pattern update frequency compared to every timestep.

It is important to note that while the patterns generated by SPG are not reproducible upon changing the pattern update frequency, the underlying statistical properties are retained. For example, Figure 7 clearly demonstrates that the patterns produced using update frequencies of every hour and every timestep are clearly different for a single forecast cycle and ensemble member. However, when considering the bulk patterns statistics at each leadtime for every ensemble member and forecast cycle over a two week period, we find that the pattern statistics for SPP_MCEH_DP and SPP_MCET_DP are broadly similar. This is illustrated for the parameter PSIGQSAT in Figure 23, and is also the case for the other SPP parameters (not shown).

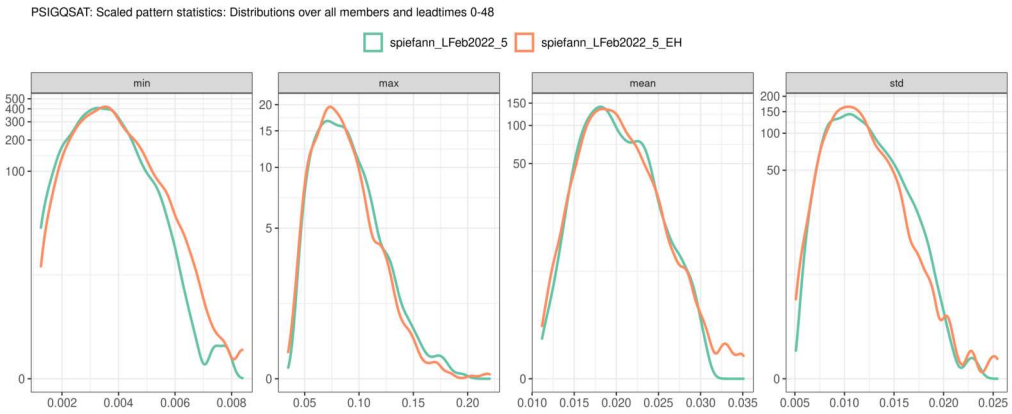


Figure 23: Bulk scaled PSIGQSAT pattern statistics for experiments SPP_MCET_DP (green) and SPP_MCEH_DP (orange) over the winter test period. The distributions are generated using the pattern statistics for each leadtime, ensemble member, and 48hr forecast cycle over the two weeks.

4.5 SPP cost

Finally, we briefly consider the additional cost associated with SPP for the experiments discussed in this section. Previous investigations suggest that calling the pattern update every timestep can add an additional cost of 10-15% to the forecast runtime, depending on the platform. However this cost can be significantly reduced by updating the pattern less frequently, e.g. approximately no additional cost is observed when updating the pattern every hour.

In Figure 24 we compare forecast runtimes over the summer and autumn test periods for experiments SPP_MCET_DP and SPP_ATET_DP to the reference DPert_DP experiment for all perturbed members. Only the 48hr forecasts are considered. An average additional cost of approximately 4-5% is observed for the IRELAND25 domain on Atos, although significant fluctuations from this can be observed for individual forecasts. This is somewhat lower than the 10-15% reported above, however these tests utilised a different domain and platform. One can also assess the impact of the pattern update interval by comparing experiments SPP_MCET_DP and SPP_MCEH_DP for the available winter period; in this case, using an update frequency of every timestep is approximately 6% more expensive than every hour (not shown). As such, SPP cost is essentially neutral when updating the patterns every hour, in-line with previous investigations.

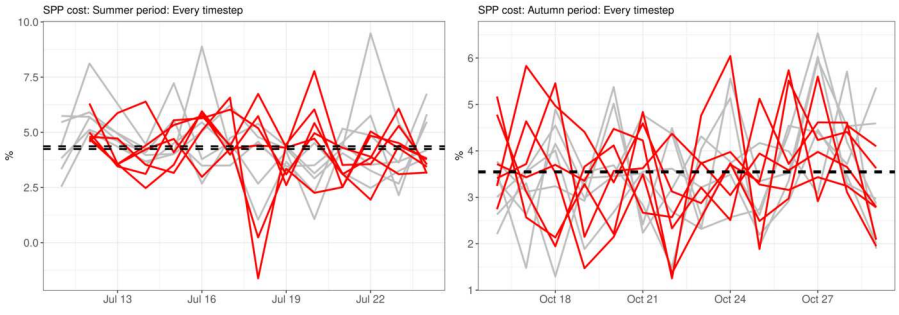


Figure 24: Additional runtime cost of using SPP with a pattern update frequency of every timestep. Gray/red lines indicate experiments SPP_ATET_DP/SPP_MCET_DP compared to the reference DPert_DP for the summer (left) and autumn (right) periods. All tests carried out on Atos.

5 SPP experiments in single precision

5.1 Experiment details

In this section we focus on the stability and performance of SPP in single precision. Analogous experiments to those discussed in Section 4 were repeated in SP, as described in Table 6. As the stability and robustness of the SPP perturbations in SP was of primary interest here, the following configuration details were used for the majority of experiments:

- The partial SP pattern fixes proposed in Appendix 7.1.3 were not implemented. As discussed in Section 2.5.2, these changes were insufficient to ensure reproducible patterns in SP and DP when updating every timestep.
- A pattern update frequency of every timestep was used.

These choices were made in order to "push" SP stability and thus help identify possible model crashes. However, additional tests utilising the partial SP pattern fixes, and a pattern update frequency of every hour, were also carried out. These additional experiments are detailed in Table 6. For each two week test period all SP ensemble members were again warm started from the corresponding spun-up DPert_DP control member.

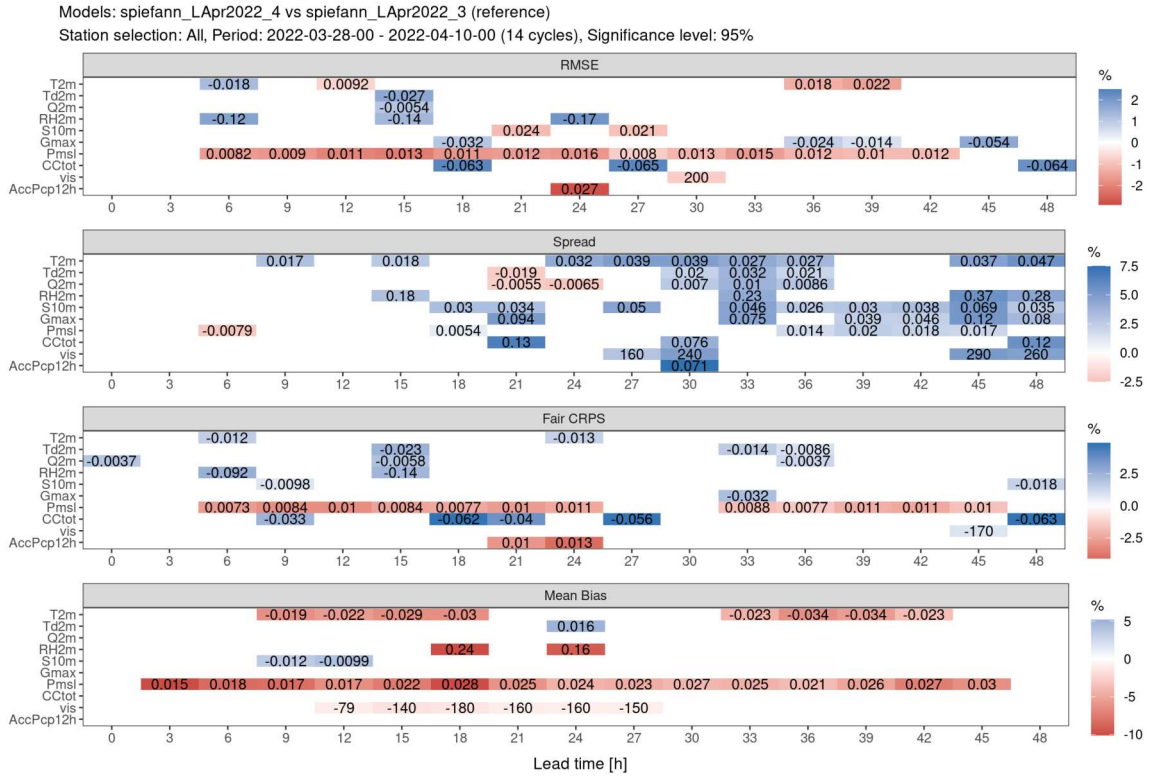
Name	Binary	SPP config.	NPATFR_SPP	Test period(s)
SPP_ATET_SP	C_2 or A_1	5PAT	1	All seasons
SPP_ATET_SP_RNM	A_2	5PAT	1	Autumn
SPP_MCET_SP	C_2 or A_1	5PMC	1	All seasons
SPP_MCET_SP_RNM	A_2	5PMC	1	Autumn, Winter
SPP_MCEH_SP_RNM	A_2	5PMC	-1	Winter

Table 6: Single precision SPP experiments considered in this section. All settings are identical to DPert_SP (Section 3.1) but with SPP=yes, and SPP configurations are those given in Table 2. Binary versions are those listed in Table 8, where suffix "RNM" indicates the use of the partial pattern fixes in Appendix 7.1.3. The test period(s) are those detailed in Table 4.

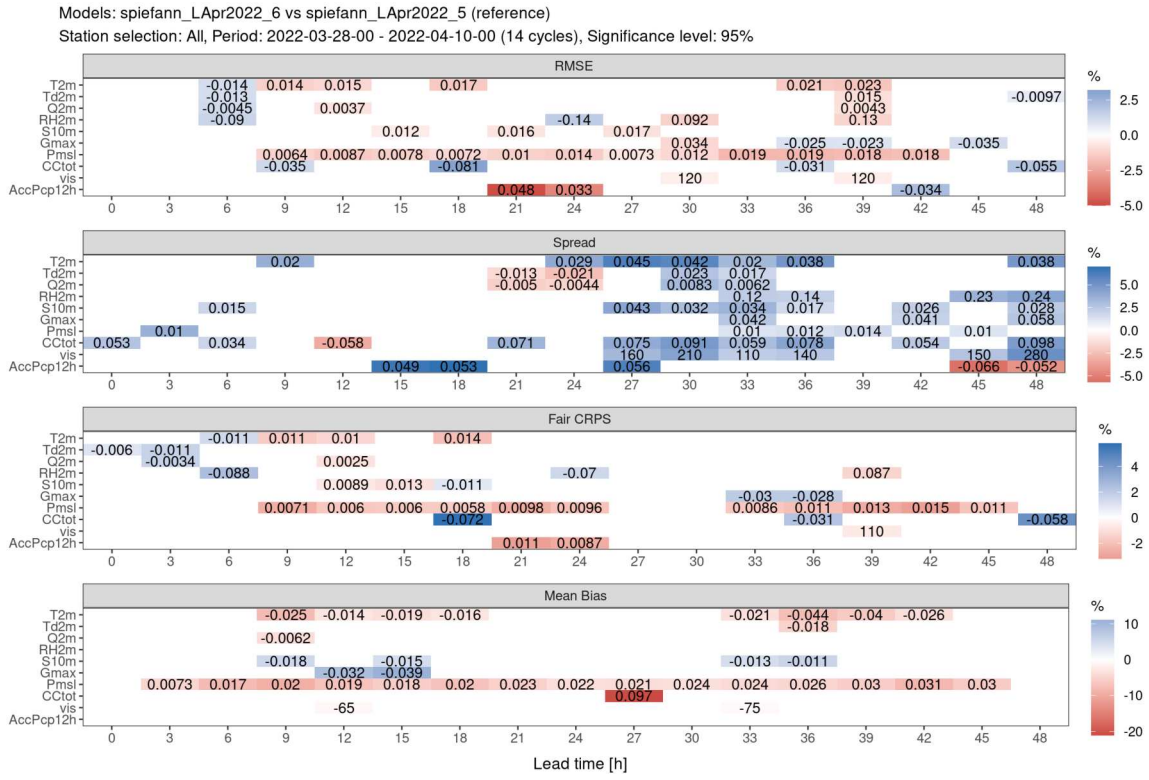
5.2 Point verification comparison

We first consider the standard surface verification scores for the KNMI and MetCoOp SP SPP experiments using a pattern update frequency of every timestep and no partial pattern fixes (i.e. experiments SPP_ATET_SP and SPP_MCET_SP, respectively). All scores are given relative to the corresponding DP experiment. Indicative results for the spring test period are given in Figure 25. A positive PMSL bias in the SP ensemble, which in turn degrades RMSE and CRPS scores, is clearly evident for both SPP configurations, along with a small cold bias of ~ 0.02 K. Both of these features are consistent with the findings of Section 3 when no SPP perturbations are active, and reflect a common feature of running HARMONIE-AROME in SP.

The surface scorecards also highlight a trend for slightly increased spread at longer leadtimes in the SP SPP ensemble relative to DP for the majority of variables. This is also illustrated in the corresponding spread-skill plots for 2 m temperature and 10 m wind speed given in Figure 26. Considering



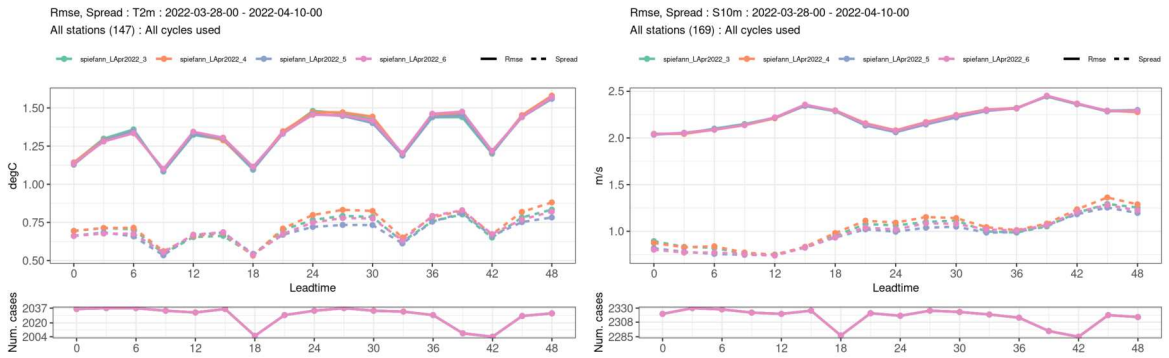
(a) SPP_ATET_SP



(b) SPP_MCET_SP

Figure 25: Surface parameter scorecard comparing experiments (a) SPP_ATET_SP to SPP_ATET_DP (reference) and (b) SPP_MCET_SP to SPP_MCET_DP over the spring test period. Blue/red indicates an improvement/degradation for SP ensemble.

the same scorecards for the other test periods,[4] as presented in Appendix 7.2.2, illustrates that a signal for increased spread over time in the SP SPP experiments, relative to DP, is a common feature across all runs. This signal is markedly different to the behaviour observed when comparing SP and DP ensembles with no SPP perturbations (see Figures 8 and 9 for comparison). As such, these results suggest that the combination of SP and SPP is having an undesirable impact on bulk ensemble behaviour relative to DP by artificially inflating ensemble spread.



(a) 2 m temperature

(b) 10 m wind speed

Figure 26: Ensemble RMSE and spread for experiments SPP_ATET_DP (green), SPP_ATET_SP (orange), SPP_MCET_DP (blue), SPP_MCET_SP (magenta) over the spring test period for 2 m temperature and 10 m wind speed.

5.3 Model failures

While the vast majority of the SP SPP tests over the four test periods remained stable, a number of model crashes were also encountered during testing, which are listed in Table 7. For experiments SPP_ATET_SP and SPP_MCET_SP, i.e. using the "default" code with pattern updates every timestep, only one crash was observed. This occurred during the autumn test period for cycle 2021/10/21/00; member 2 failed just after hour ~ 36 in the forecast, with all other members completing successfully. This in itself highlights one of the major difficulties with SP testing, as the crash could have been easily missed if, for example, experiments were only carried out over the summer and winter periods, or if the forecast length was restricted to 36 hours. As such, it emphasises the need the robust SP testing over multiple test periods with long forecasts. Moreover, given that only one of the six perturbed members failed for this cycle, it also suggests the need for a relatively large ensemble size. With that said, this methodology does not of course guarantee that all possible SP model crashes will be observed, and again emphasise the need for a more systemic methodology for SP testing within the ACCORD community.

The backtrace of this 2021/10/21/00 crash is included in Appendix 7.1.4 for reference. A division by zero floating point exception is observed in arpifs/phys_radi/swni.F90 at

```
ZRR2(JL,JAJ)=1.0_JPRB/PAKI(JL,JAJ,KNU)
```

where the variable "PAKI" is derived from arpifs/phys_radi/swu.F90 as "PAKI = -LOG(ZR)/ZUD", with "ZR" computed in arpifs/phys_radi/swtt1.F90. Additional investigations suggested that the prob-

[4]Note that for the autumn test period, the failed cycle 2021/10/21/00 (discussed in Section 5.3) is excluded when computing the verification scores.

Exp.	Member	Cycle	Cause	Additional details
SPP_ATET_SP	2	2021/10/21/00	SIGFPE in SWNI	Failed at hour 36
SPP_MCET_SP	2	2021/10/21/00	SIGFPE in SWNI	Failed at hour 36
SPP_ATET_SP_RNM	6	2021/10/29/00	SIGFPE in SWNI	Failed at hour 45
SPP_MCET_SP_RNM	6	2021/10/29/00	SIGFPE in SWNI	Failed at hour 45
SPP_MCET_SP_RNM	1	2022/02/22/00	SIGFPE in SWNI	Failed at hour 39

Table 7: List of all failed cycles for the SP SPP experiments detailed in Table 6. Note that in each case, all other members completed successfully for the cycle indicated.

lem was caused by ZR=1, possibly due to roundoff in swtt1.F90. However, conversion of this script to use JPRD was not sufficient to avoid the error, and it was not possible to identify the upstream cause of the issue. As such, a bugfix for this model crash remains outstanding.

Note that this autumn test period was repeated using the partial SP pattern fixes described in Section 2.5.2 (experiments SPP_ATET_SP_RNM and SPP_MCET_SP_RNM in Table 6), in order to assess if they could potentially stabilise this crash. However the same model failure was observed for these runs but for a different forecast cycle and member, i.e. cycle 2021/10/29/00 for member 6. Moreover, a run over the winter period for SPP_MCET_SP_RNM also encountered the same crash in swni.F90 (see Table 7). It is interesting to note that the winter period crash did not occur when the pattern partial fixes were not implemented (i.e. for experiment SPP_MCET_SP) and, as such, these partial fixes do not help to stabilise the SP SPP forecasts.

5.3.1 Analysis of single precision SPP patterns

In order to investigate these crashes further, it is natural to analyse and compare the behaviour of the SP and DP SPP perturbation patterns for the failed cycles. In Figure 28(a) we consider a timeseries of the PSIGQAT scaled pattern statistics for experiments SPP_ATET_SP and SPP_ATET_DP over cycle 2021/10/21/00. Only forecasts up to hour 33 are considered as SP mbr002 failed at hour ~ 36. Evidently one of the SP ensemble members, which corresponds to the crashed mbr002 in this case, has a sharp increase in the mean, max, and standard deviation of the PSIGQSAT perturbation pattern in the hours before the crash. A simple comparison with the equivalent behaviour of the DP members indicates that mbr002 diverges significantly from what is observed in DP. An analysis of the same pattern statistics for the other SPP parameters does not indicate similar "divergent" behaviour for mbr002 (not shown), suggesting that the crash may be linked to this spike in the PSIGQSAT pattern.

Given that PSIGQSAT represents the saturation limit sensitivity, it is interesting to consider the corresponding rainfall forecasts for this failed cycle. This is done so in Figure 27, where we consider the 2021/10/21/00+33 24 hour rainfall forecast for all members in experiments SPP_ATET_DP and SPP_ATET_SP. A quite significant area of rainfall is observed just south of Ireland in mbr002, which appears to be a somewhat anomalous given that all other ensemble members demonstrate little to no rainfall in the vicinity. As one might expect, this area of intense rainfall for mbr002 is also associated with extremely large values for the PSIGQSAT SPP perturbation (not shown). Note that experiment SPP_MCET_SP, which fails at the same hour for the same forecast cycle (as indicated in Table 7), demonstrates the same anomalous rainfall and spike in the PSIGQSAT perturbation pattern. This is due to the fact that the KNMI and MetCoOp SPP configurations share identical settings for the PSIGQSAT parameter.

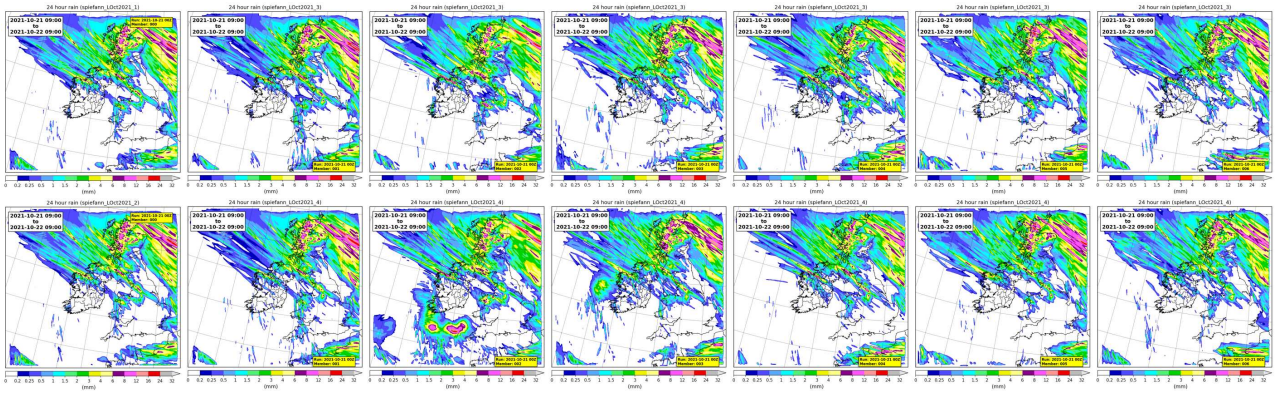
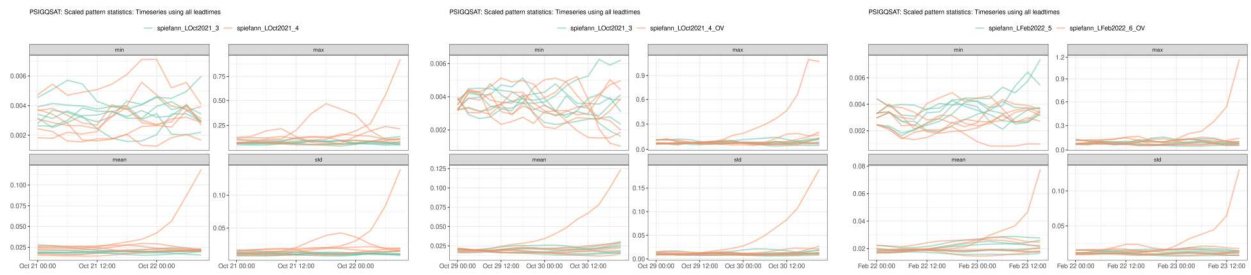


Figure 27: 24 hour rainfall using the 2021/10/21/00+33 forecasts for experiments SPP_ATET_DP (top row) and SPP_ATET_SP (bottom row). Members 0-6 are given from left to right.



(a) 2021/10/21/00 cycle (b) 2021/10/29/00 cycle (c) 2022/02/22/00 cycle

Figure 28: Timeseries of the PSIGQSAT scaled pattern statistics for experiments (a) SPP_ATET_DP (green) and SPP_ATET_SP (orange), (b) SPP_ATET_DP (green) and SPP_ATET_SP_RNM (orange), and (c) SPP_MCET_DP (green) and SPP_MCET_SP_RNM (orange) over the 2021/10/21/00, 2021/10/29/00, and 2022/02/22/00 cycles, respectively. Individual lines represent each ensemble member, and the divergent member in each case corresponds to the failed SP member from Table 7.

A similar analysis can be carried out for the other two crashed forecast cycles in Table 7, with timeseries of the PSIGQSAT pattern statistics for these cycles illustrated in Figure 28. For cycles 2021/10/29/00 and 2022/02/22/00, we compare the failed SP experiments to their DP equivalents (i.e. SPP_ATET_SP_RNM to SPP_ATET_DP, and SPP_MCET_SP_RNM to SPP_MCET_DP, respectively). In each case, the PSIGQSAT pattern for the failed SP member diverges significantly from the DP behaviour in the hours before the model crash. The corresponding 24 hour rainfall totals for the SP and DP ensembles, given in Appendix 7.2.3, confirm that these spikes are again associated with anomalous heavy showers for the failed members.

Therefore all SP crashes encountered during testing appear to be linked to a significant divergence in the PSIGQSAT pattern for one ensemble member from the behaviour observed in DP. To assess if this difference between the SPP patterns in SP and DP is a general feature, we can again construct distributions of pattern statistics over each two week test period using all ensemble members and forecasts (in a similar manner to Figure 23). These are given for PSIGQSAT in Figure 29 for the DP (SPP_MCET_DP) and SP (SPP_MCET_SP) MetCoOp SPP configuration experiments (with identical results for the KNMI configuration experiments).[5] We again consider the parameter PSIGQSAT given the role it plays in the observed model crashes. While the DP and SP ensemble distributions are similar overall, one can also observe a general tendency for larger pattern means, maxima, and standard deviation in the SP experiments. Comparing these distributions to Figure 28 indicates that these more "extreme" values in the SP patterns may be reflective of SP member divergences.

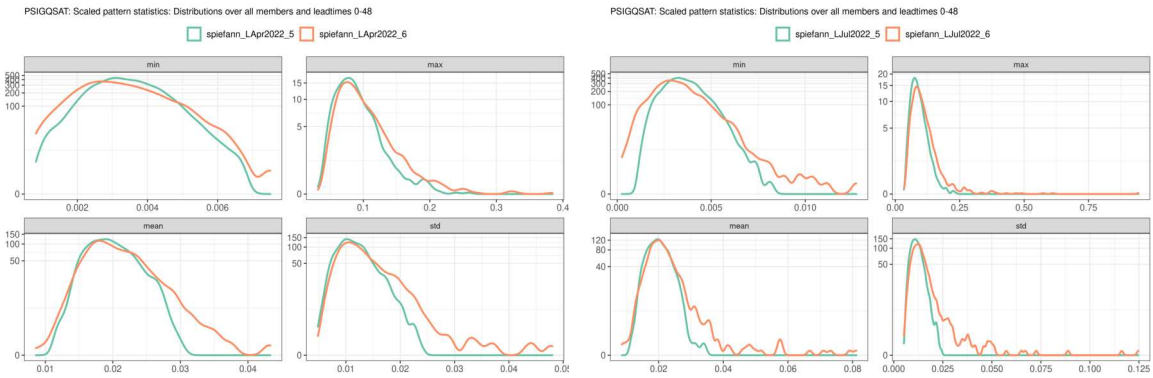
The results presented in Section 2.5 demonstrated that SPP perturbation patterns were not reproducible upon switching precision for a pattern update frequency of every timestep. However the analysis outlined above suggests a wider problem with the SP SPP patterns, in the sense that:

1. The overall bulk behaviour of the SP SPP patterns can differ significantly from that observed in DP, with anomalously large values for SPP perturbations present for "diverging" SP ensemble members.

2. Anomalously large SP SPP perturbations are not addressed when using the partial fixes to the SP patterns proposed in Appendix 7.1.3.

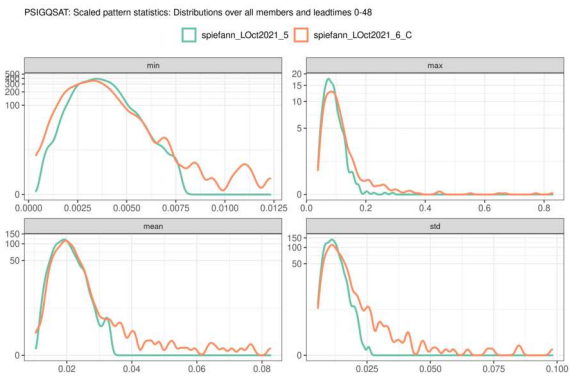
3. Crashes in the SP forecasts appear to be linked to divergent behaviour in the SP SPP patterns.

[5] Again, this is because the two SPP configurations share the same settings for PSIGQSAT.

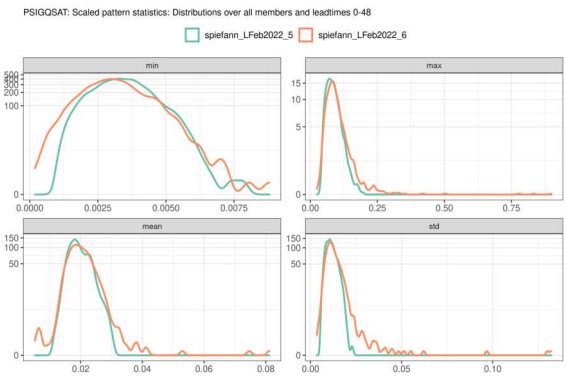


(a) Spring

(b) Summer



(c) Autumn



(d) Winter

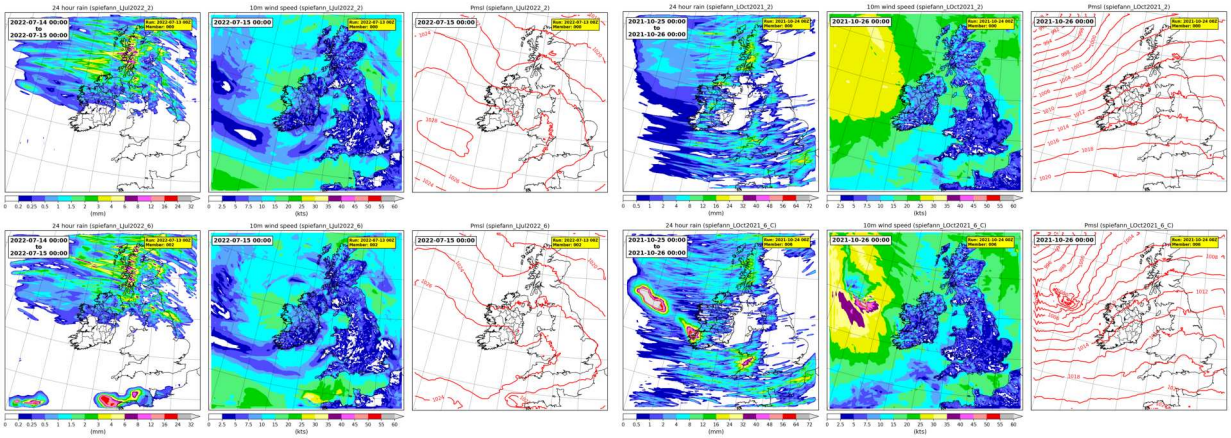
Figure 29: Bulk scaled PSIGQSAT pattern statistics for experiments SPP_MCET_DP (green) and SPP_MCET_SP (orange) over all test periods. These distributions are generated using the pattern max, min, mean, and standard deviation for each leadtime, ensemble member, and 48hr forecast cycle over each two week period. For the autumn period, the failed 2021/10/21/00 cycle is omitted.

5.3.2 Additional heavy rainfall cases

All model crashes discussed in Section 5.3 were connected with anomalously high rainfall in one of the SP members. However, these very high rainfall cases in SP can also be found for forecast cycles which completed successfully. Two examples of this are illustrated below for cycles 2022/07/13/00 and 2021/10/24/00, where we contrast control and perturbed member forecasts from experiment SPP_MCET_SP. In both cases the perturbed members considered in Figure 30 are found to have very large PSIGQSAT perturbation values at longer lead times (not shown).

For the July case (Figure 30(a)) a relatively heavy shower close to the southern boundary of the domain can be observed in the SP perturbed forecast, which is completely absent in the SP control member. Distorted wind and PMSL patterns relative to the control are also evident. In Figure 30(a) a clearly erroneous rainfall pattern to the west of Ireland is evidenced for SP mbr006 for the 2021/10/24/00 cycle. The corresponding 10 m wind speed and PMSL fields demonstrate alarming behaviour. Indeed, it is impressive that model stability is maintained in this case. Analysis of the corresponding DP forecasts (experiment SPP_MCET_DP) for these cycles indicates no such high rainfall in the perturbed members (not shown).

In order to fully demonstrate the role played by the PSIGQSAT perturbation in both the model crashes (Table 7) and the high rainfall cases in Figure 30 a number of experiments with modified SPP settings



(a) Cycle 2022/07/13/00 (b) Cycle 2021/10/24/00

Figure 30: (a) Left to right: 24hr rainfall, 10 m wind speed, and PMSL at hour 48 for cycle 2022/07/13/00. The top and bottom rows indicate the mbr000 and mbr002 forecasts for experiment SPP_MCET_SP. The same plots are repeated in (b) but for the 2021/10/24/00 cycle, where the mbr006 forecast is given on the bottom row.

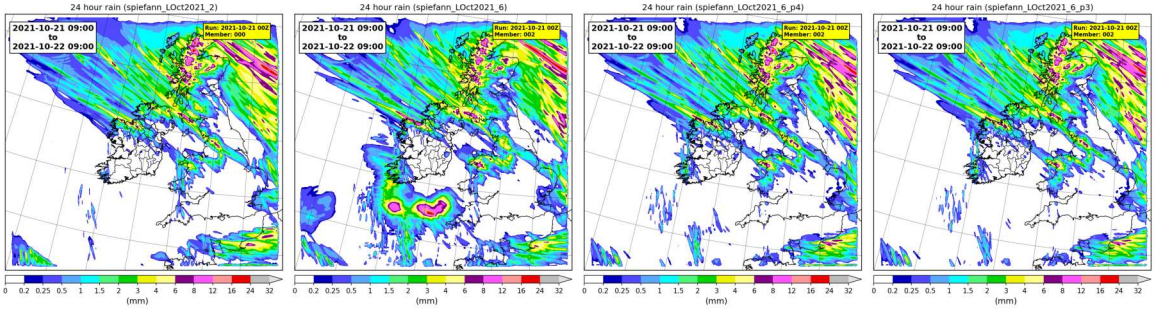


Figure 31: 24 hour rainfall at hour 33 for the 2021/10/21/00 cycle. From left to right: SPP_MCET_SP control member, failed SPP_MCET_SP mbr002 (Table 7), SPP_MCET_SP mbr002 with CMPERT=0.3 (change P1 in the text), and SPP_MCET_SP mbr002 with a uniform distribution (change P2 in the text).

were carried out. In particular, the following settings were considered:

- P1: Reduce the standard deviation of the PDF for PSIGQSAT from 0.6 to 0.3 (see Table 2).
- P2: Switch to using a pseudo-uniform PDF for PSIGQSAT instead of a lognormal distribution. This was done via "LUNIFORM_PSIGQSAT=TRUE" without changing the corresponding CMPERT (0.6) or default offset (0.5). Note that this configuration was used simply for testing purposes, and is not necessarily proposed as a suitable PDF for PSIGQSAT.

The failed mbr002 2021/10/21/00 forecast for experiment SPP_MCET_SP was re-run for settings P1 and P2 using the corresponding mbr002 2021/10/20/21 first guess files. The resulting 24 hour rainfall forecasts at hour 33 are illustrated in Figure 31. In this case changes P1 and P2 are found to both stabilise the forecast and eliminate the erroneous rainfall. P1 and P2 tests have also been carried out for the rainfall cases in Figure 30, with a similar removal of the erroneous showers (not shown).

These results therefore suggest that the SP SPP model stability and rainfall issues could be avoided by simply modifying the SPP settings for PSIGQSAT, with the reduction in CMPERT an obvious can-

didate. However, the SPP configurations considered in this note have been arrived at after significant tuning, and such a change would impact overall SPP performance. More importantly, this "solution" essentially just dodges instead of addressing the underlying problem with the SP SPP patterns when updating every timestep, and may not be sufficient to ensure SP stability in general.

5.4 Impact of hourly pattern updates

As discussed in Section 2.5.2, switching from a pattern update interval of every timestep to every hour appears to eliminate the differences between the SP and DP SPP patterns (assuming the use of the partial pattern fixes in Appendix 7.1.3). As such, an additional two week run with NPATFR_SPP=-1 was carried out over the winter test period (as detailed in Table 6). The distributions of the pattern statistics over this two week run were computed and found to be essentially identical to those observed in DP, confirming the results of Section 2.5.2. As such, experiment SPP_MCEH_SP_RNM therefore provides an assessment of SP SPP stability and performance when the perturbation patterns are independent of model precision.

The winter run with hourly pattern updates was found to be stable, in contrast to the equivalent experiment with updates every timestep which crashed at the 2022/02/22/00 cycle (see Table 7). A comparison of the surface verification scores for SPP_MCEH_SP_RNM to its DP equivalent is given in Figure 32. We again observe an additional PMSL bias and slight degradation in 2 m temperature scores when running in SP, which is entirely consistent with the results obtained when comparing the SP and DP ensembles without SPP (see Figure 9(b)). Importantly, however, there is no signal for increased spread with leadtime in the SP ensemble when updating every hour, which contrasts with the behaviour observed when updating every timestep (e.g. Figure 38(b)).

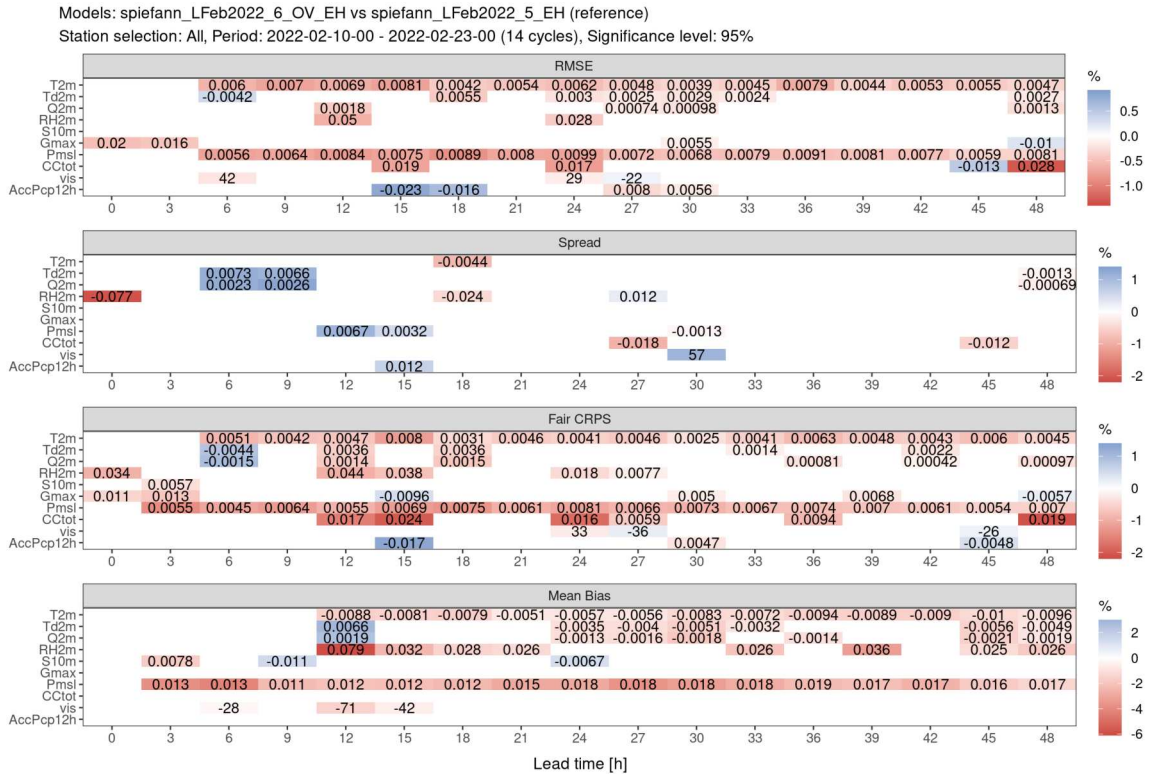
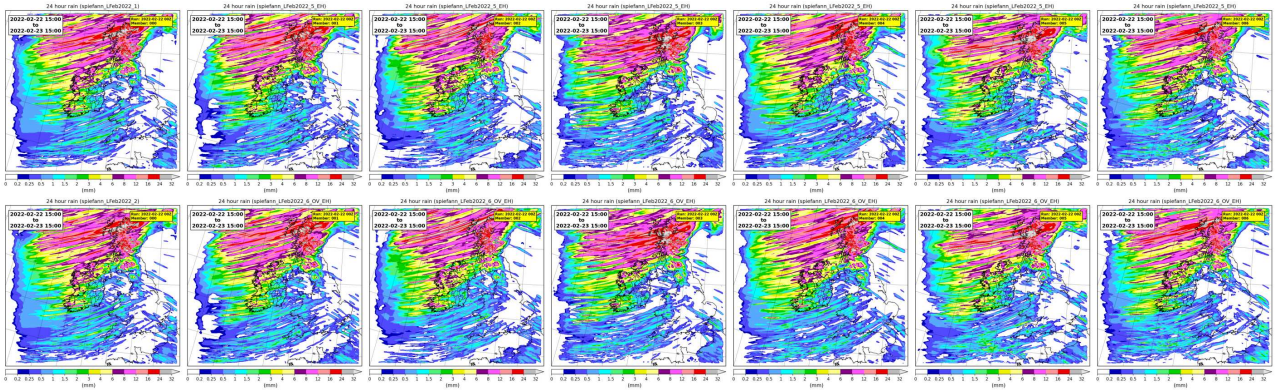
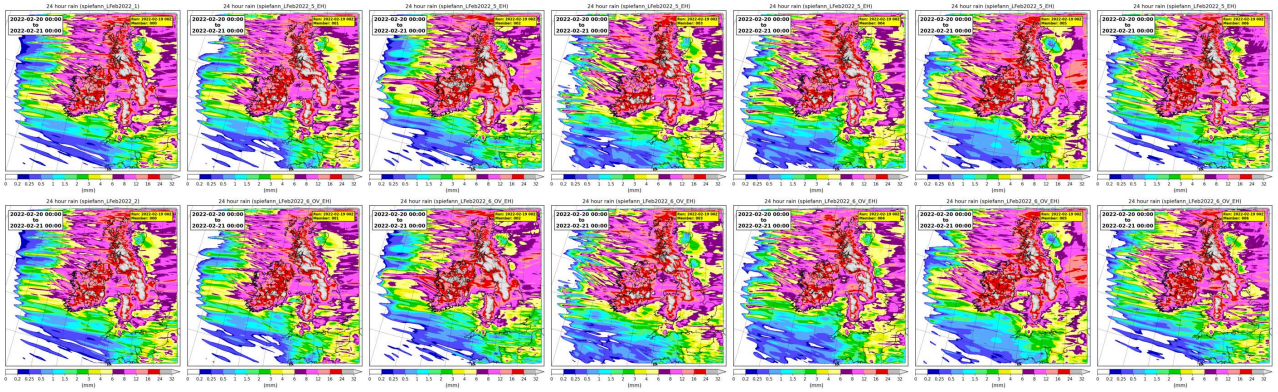


Figure 32: Surface parameter scorecard comparing SPP_MCEH_SP_RNM to SPP_MCEH_DP (reference) over the winter test period. Blue/red indicates an improvement/degradation for SP ensemble.

Finally, Figure 33 illustrates two sample 24 hour rainfall forecasts for the SP and DP ensembles with hourly pattern updates. In Figure 33(a), we consider the 2022/02/22/00 cycle which crashed in SP when updating the SPP patterns every timestep (Table 7). This failure was associated with an anomalous heavy shower in southeast Ireland for SP mbr001 (see Figure 39). In contrast, the SP and DP perturbed member forecasts with hourly updates are almost identical, with no indication of significant rainfall in the southeast. Also given in Figure 33(b) is the 2022/02/19/00+48 rainfall forecast for Storm Franklin (as previously considered in Section 3.4). Reassuringly similar behaviour in the SP and DP SPP ensembles are observed.



(a) 2022/02/22/00 cycle



(b) 2022/02/19/00 cycle

Figure 33: 24 hour rainfall for (a) 2022/02/22/00+39 and (b) 2022/02/22/00+48 forecasts for experiments SPP_MCEH_DP (top row) and SPP_MCEH_SP_RNM (bottom row). Members 0-6 are given from left to right.

6 Summary and Conclusions

The suite of ensemble experiments discussed in this note has highlighted a number of important aspects regarding both SPP and SP HARMONIE-AROME:

- With the default set of surface, LBC, and EDA perturbations, no major stability or performance issues are observed in SP. A small positive PMSL bias in SP is evident across all testing periods, as well as a slight cold bias relative to DP. Runtime savings of close to 40% are achieved on Atos.
- In DP, the SPP scheme using either the KNMI or MetCoOp configuration is found to perform well overall, with a significant improvement in ensemble spread and CRPS over the IRELAND25 domain for all test periods considered. The slight degradation in ensemble RMSE is more than compensated by the increased spread, leading to improved spread-skill ratios for the majority of surface variables. However, SPP also appears to introduce a negative wind speed bias in the perturbed members relative to the control and leads to some "off-centering" of the members.
- The SP and DP SPP perturbation patterns can differ substantially in general. A number of partial fixes were implemented which ensure almost identical SP and DP patterns when updating the patterns every hour, but differences still exist when updating the patterns every timestep. These differences reflect a tendency for larger pattern means, maxima, and standard deviation when running in SP.
- A number of SP crashes with the SPP scheme were encountered when updating the perturbation patterns every timestep. These crashes appear to be linked to a divergence in the behaviour of the SP SPP patterns relative to DP. Moreover, erroneous rainfall forecasts have also been observed for SP SPP members, which are caused by extremely large PSIGQSAT perturbation values.
- For the single test period considered, SP SPP appears to work well when switching to hourly pattern updates.

While the performance of SP HARMONIE-AROME without SPP perturbations provides additional evidence to support its possible operational use, the SP SPP crashes encountered in this note again emphasise the need for a more rigorous and coordinated approach to SP testing within ACCORD. Such testing could involve long-term cycling of SP e-suites (which is currently carried out at MetCoOp) and the use of the URANIE software for sensitivity analysis.

Of most pressing concern is of course the SP SPP crashes which appear to be caused by the SP pattern differences relative to DP. Personal correspondence with Ole Vignes (Met Norway) has indicated that additional modifications have recently been made to ensure that SP and DP patterns are almost identical for all pattern update intervals. These changes are to be committed to HARMONIE-AROME cycle 46 and were not available for testing in this special project.

Finally, there are a number of avenues for future work. In particular, a dedicated study of the positive PMSL bias in SP may be warranted, along with the possible degradation in UA humidity profiles. Extensive SP SPP testing in cycle 46, with the additional pattern fixes mentioned above, also needs to be carried out, while an exploratory evaluation of URANIE for SP HARMONIE-AROME testing should be considered.

7 Appendix

7.1 Code details

7.1.1 Binary versions

The following versions of the harmonEPS-43h2.2 branch, and slight modifications thereof, were used at various points during the ECMWF Special Project discussed in this note. The column "Host" indicates the HPC used.

Binary verison	Host	From commit	Local src changes
C_1	cca	8f9f5b2554	-
C_2	cca	0e03cb9996	-
A_1	Atos	0e03cb9996	Adaptions for running on Atos (see Table 9)
A_2	Atos	0e03cb9996	Adaptions for running on Atos (see Table 9) and OV changes for pattern reproducibility (see Table 10)

Table 8: List of harmonEPS-43h2.2 binaries used in this report.

7.1.2 Porting to Atos

As the harmonEPS-43h2.2 branch was not initially compatible for the new Atos HPC in Bologna, manual porting of the branch to Atos was required in order to:

- bring in new configuration and submission files,
- allow for the use of a virtual ecflow server,
- change various hardcoded paths, the location of job files, etc..

As these changes had already been made to the harmonie-43h2.2_bf branch, it was sufficient to "git cherry-pick" the relevant commits into the harmonEPS-43h2.2 branch (checked out at commit 0e03cb9996). The list of "cherry-picked" commits used are given in Table 9. Note that these changes were subsequently merged into the official harmonEPS-43h2.2 branch in early 2023 (see pull request #590 in the Harmonie repository).

7.1.3 Partial single precision pattern fixes

The main methodology to address the divergence between SPP perturbation patterns generated in single and double precision was proposed by Ole Vignes (Met Norway). Table 10 provides details of the src files modified, and the relevant changes can be viewed in pull request #595 in the Harmonie repository. The main change related to the "RANDOM_NUMBERS_MIX" routine, which is called by "SET_SEED_ARP" and "EVOLVE_ARP_SPG" in spectral_arp_mod.F90 (src/arpifs/module). These changes were then used to generate the A_2 binairies listed in Table 8.

#	Commit	Note
1	bbd12c194d	Add Atos config files
2	6aec80ffbb	Add intel Atos config files
3	7e81458883	-
4	92e0c62dbc	-
5	2469b464ce	-
6	14dc77c3c5	-
7	77904af556	-
8	765a327def	-
9	c4a254a7ff	For ecflow virual machine
10	acf3467f14	-
11	2fe0b6eae1	Required to avoid failure in Canari at "read_namelists_tebn.F90"

Table 9: List of commits used to port harmonEPS-43h2.2 to Atos.

Subsequently, various other attempts were also made to address outstanding discrepancies between the patterns when the pattern update frequency is set to every timestep. However, none of these were successful, and as such are not discussed here.

File	Description
src/algor/module/random_numbers_mix.F90	Introduce specific 64-bit uniform/Gaussian distribution functions
src/arpifs/module/spectral_arp_mod.F90	Force the use of a 64-bit random number stream and modify various computations to double precision

Table 10: List of files changed for the partial SP pattern fixes.

7.1.4 SWNI crash

Listing 2 gives an indicative backtrace for the SWNI crash encountered by all failed cycles listed in Table 7.

7.2 Additional SPP results

7.2.1 Double precision SPP results for MetCoOp configuration

Figures 34 and 35 illustrate the surface scorecards for the DP MetCoOp SPP ensemble experiment (SPP_MCET_DP) relative to the DP reference experiment (DPert_DP) for all test periods.

```
22:00:12 STEP 1750 H=  36:27 +CPU=  0.409
22:00:12 STEP 1751 H=  36:28 +CPU=  0.410
JSETSIG: sl->active = 0
[drhook.c-l1133] signal_harakiri(SIGALRM=14): New handler installed at 0xb21035; old preserved at (nil)
***Received signal = 8 and ActivatED SIGALRM=14 and calling alarm(10), time =1664316013.12
[myproc#166,tid#1,pid#102784,signal#8(SIGFPE)]: Received signal :: 2679MB (heap), 421MB (rss), 0MB (stack), 0 (paging), nsigs 1,
↪  time 1664316013.12
tid#1 starting drhook traceback, time =1664316013.12
[myproc#166,tid#1,pid#102784]: 2679 MB (maxheap), 421 MB (maxrss), 0 MB (maxstack), walltime = 1664316013.12s
[myproc#166,tid#1,pid#102784]:  MASTER
[myproc#166,tid#1,pid#102784]:   CNT0
[myproc#166,tid#1,pid#102784]:    CNT1
[myproc#166,tid#1,pid#102784]:     CNT2
[myproc#166,tid#1,pid#102784]:      CNT3
[myproc#166,tid#1,pid#102784]:       CNT4
[myproc#166,tid#1,pid#102784]:        STEPO
[myproc#166,tid#1,pid#102784]:         SCAN2M
[myproc#166,tid#1,pid#102784]:          GP_MODEL_STACK
[myproc#166,tid#1,pid#102784]:           GP_MODEL
[myproc#166,tid#1,pid#102784]:            CPG_DRV
[myproc#166,tid#1,pid#102784]:             CPG
[myproc#166,tid#1,pid#102784]:              MF_PHYS
[myproc#166,tid#1,pid#102784]:               APL_AROME
[myproc#166,tid#1,pid#102784]:                ACRADIN
[myproc#166,tid#1,pid#102784]:                 RECMWF
[myproc#166,tid#1,pid#102784]:                  RADLSW
[myproc#166,tid#1,pid#102784]:                   SW
[myproc#166,tid#1,pid#102784]:                    SWNI
JSETSIG: sl->active = 0
...
>>signal_drhook(at 0xb2105c): Calling previous signal handler in chain at 0x1486373d5e10 (if possible)
[aa2-3064:102769:0:102769] Caught signal 8 (Floating point exception: floating-point divide by zero)
==== backtrace (tid: 102769) ====
 0 0x0000000000b21650 signal_drhook()
 ↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/ifsaux/support/drhook.c:1259
 1 0x0000000000012b20 .annobin_sigaction.c()  sigaction.c:0
 2 0x00000000017f62e3 swni_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_radi/swni.F90:363
 3 0x00000000017f41d6 sw_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_radi/sw.F90:242
 4 0x00000000017af7a7 radlsw_()
 ↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_radi/radlsw.F90:1565
 5 0x0000000001d85926 recmwf_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_radi/recmwf.F90:437
 6 0x0000000001d2fb0b acradin_()
 ↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_radi/acradin.F90:203
 7 0x0000000001b86106 apl_arome_()
 ↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_dmn/apl_arome.F90:2018
 8 0x0000000001816199 mf_phys_()
 ↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/phys_dmn/mf_phys.F90:2179
 9 0x00000000017292a1 cpg_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/adiab/cpg.F90:801
10 0x0000000001723223 cpg_drv_._omp_fn.0()
↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/adiab/cpg_drv.F90:386
11 0x0000000000011706 GOMP_parallel()  ???:0
12 0x0000000001724ed8 cpg_drv_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/adiab/cpg_drv.F90:386
13 0x00000000015ea637 gp_model_()
↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/gp_model.F90:466
14 0x000000000010828fe gp_model_stack_()
↪  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/gp_model_stack.F90:71
15 0x000000000108cfc6 scan2m_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/scan2m.F90:503
16 0x0000000000b72d0f stepo_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/stepo.F90:374
17 0x0000000000b4a76a cnt4_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/cnt4.F90:1142
18 0x0000000000b457ca cnt3_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/cnt3.F90:152
19 0x0000000000b45459 cnt2_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/cnt2.F90:109
20 0x0000000000b450f3 cnt1_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/cnt1.F90:125
21 0x0000000000b408ac cnt0_()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/arpifs/control/cnt0.F90:185
22 0x000000000061c2a3 master()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/master.F90:148
23 0x000000000061c2a3 main()  /lus/h2resw01/scratch/dujf/hm_home/spiefann_TBR_080822/lib/R32/src/master.F90:3
24 0x0000000000023493 __libc_start_main()  ???:0
25 0x000000000061c41e _start()  ???:0
================================
```

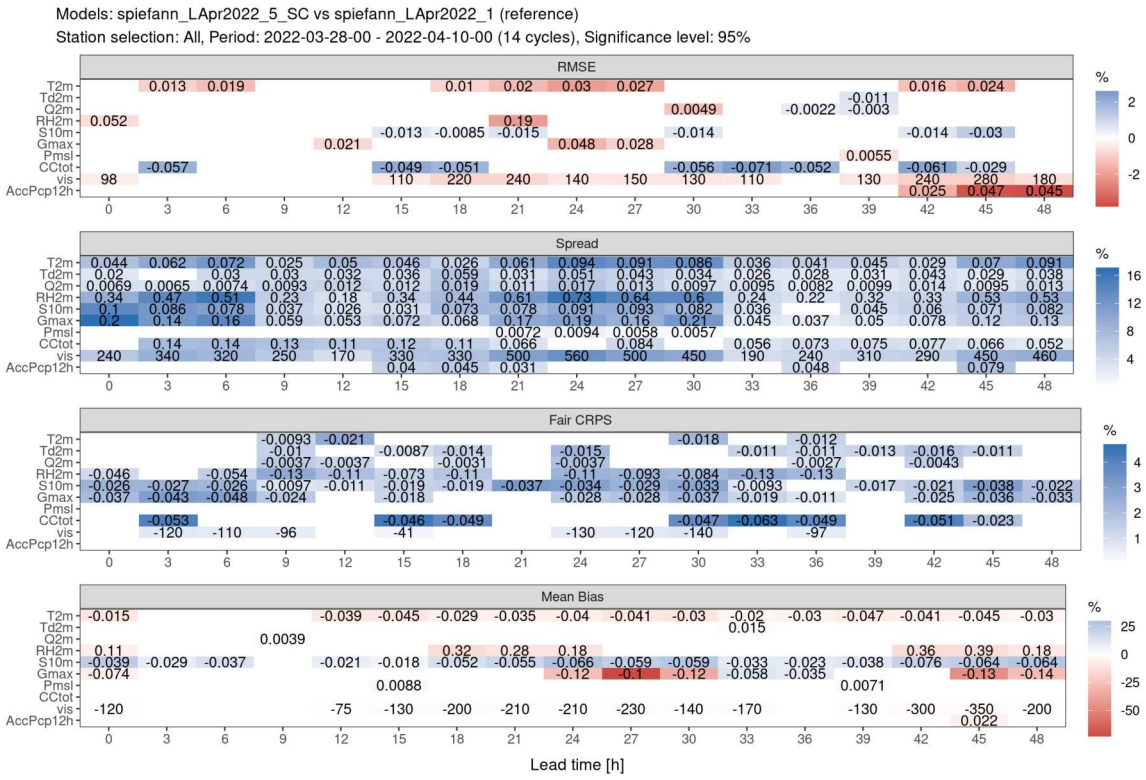
Listing 2: Backtrace for the SWNI crash indicated in Table 7.

7.2.2 Single precision SPP results for KNMI and MetCoOp configurations

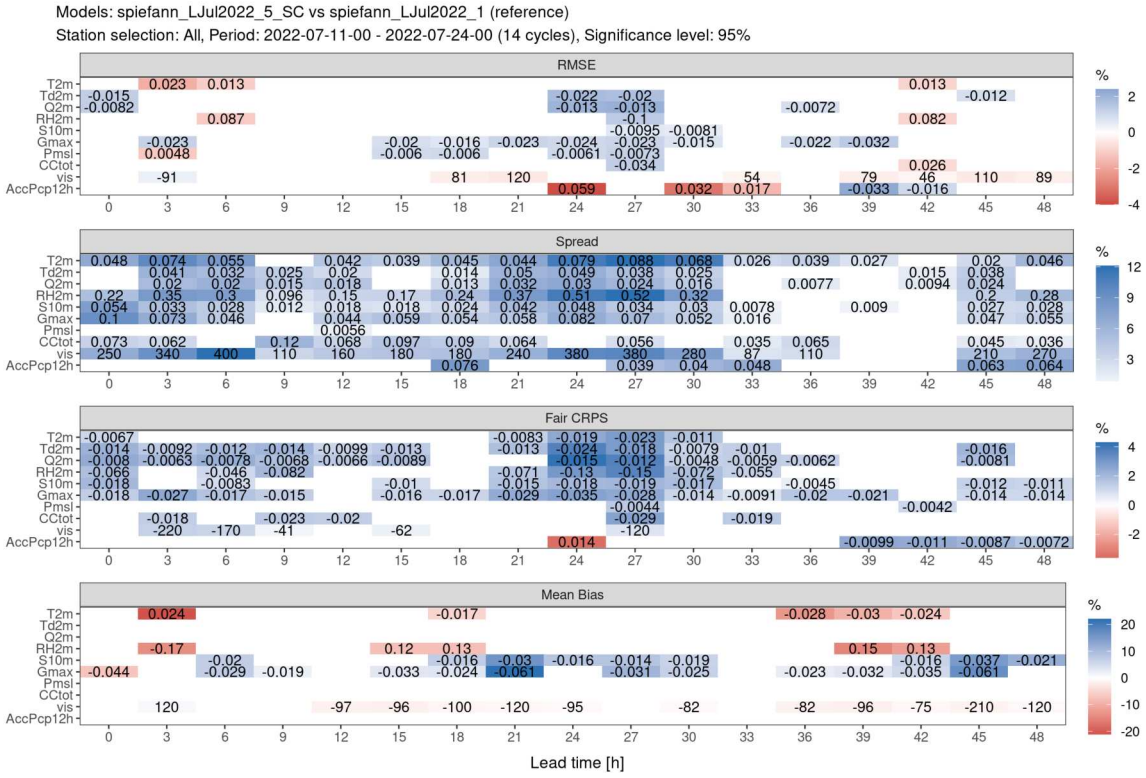
Figures 36, 37, and 38 illustrate the surface scorecards for SPP_ATET_SP and SPP_MCET_SP relative to their DP equivalents (i.e. the same as Figure 25) for the summer, autumn, and winter test periods.

7.2.3 Additional rainfall cases

In Figure 39(a) we consider all 24 hour rainfall forecasts at hour 45 for experiments SPP_ATET_DP and SPP_ATET_SP_RNM for the failed 2021/10/29/00 cycle. An anomalous shower for the SP mbr006 is clearly evident. The same plot at hour 39 for the 2022/02/22/00 cycle of experiments SPP_MCET_DP and SPP_MCET_SP_RNM is also given in Figure 39(b).

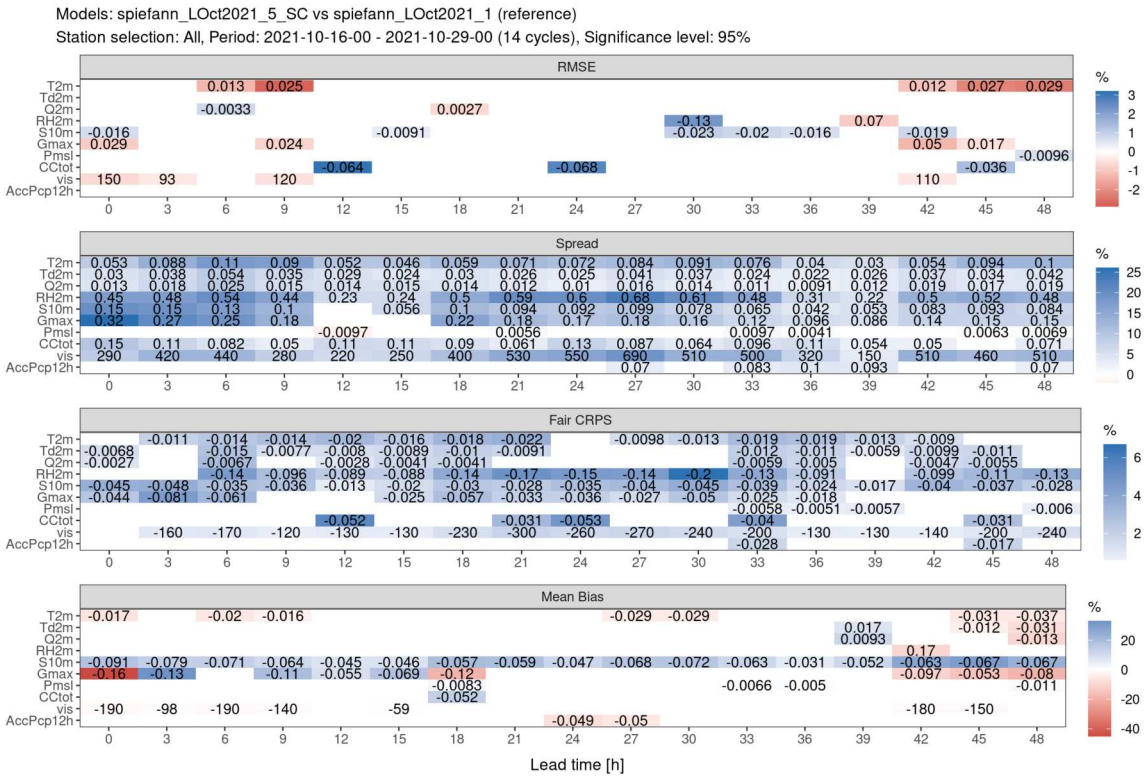


(a) Spring

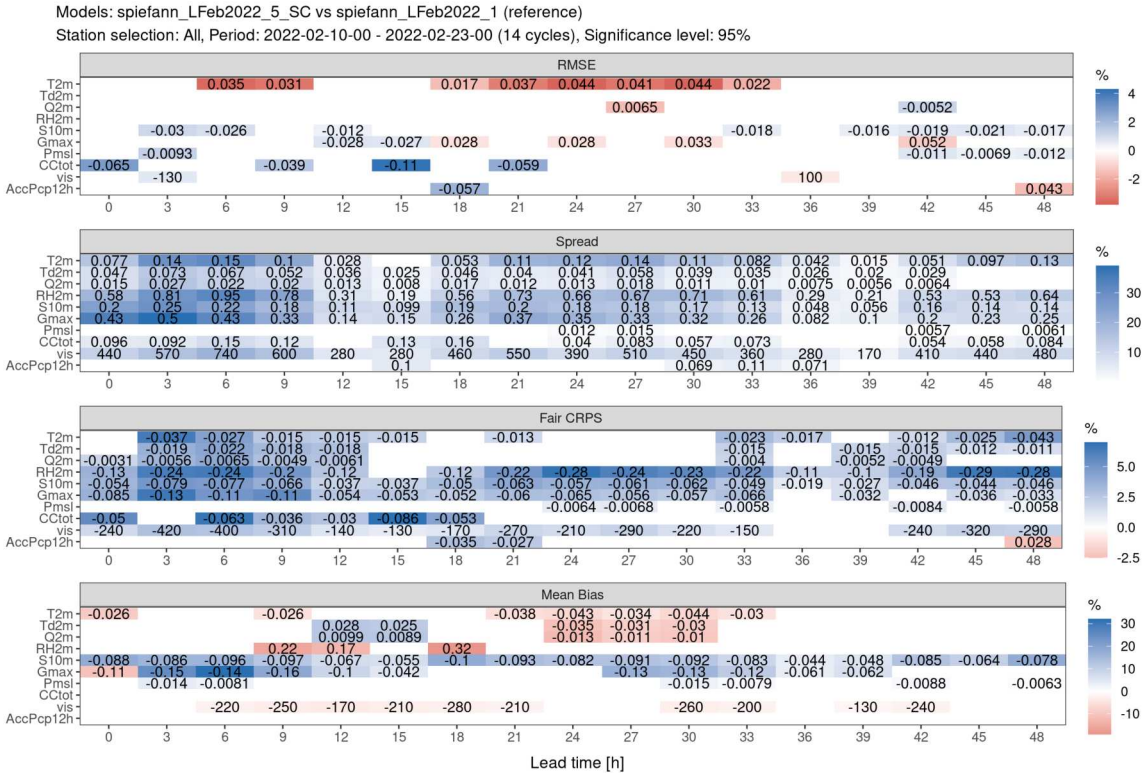


(b) Summer

Figure 34: Surface parameter scorecard comparing experiments SPP_MCET_DP to DPert_DP (reference) over the spring and summer test periods.

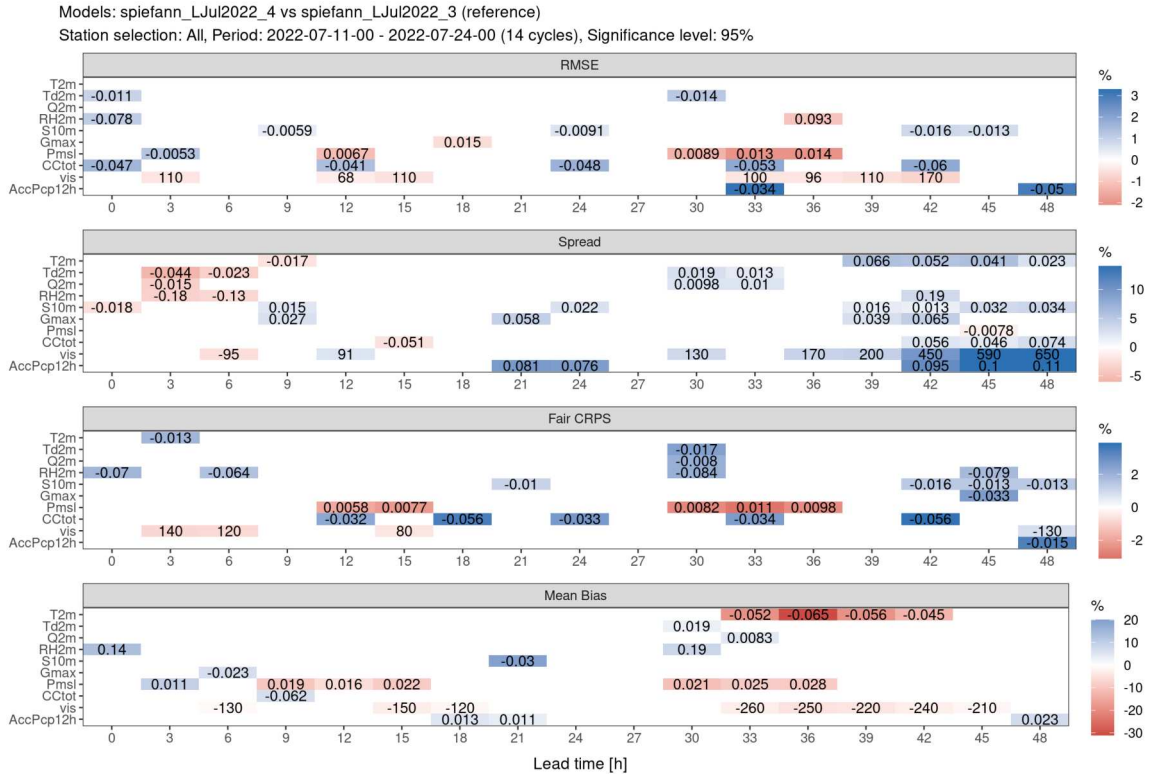


(a) Autumn

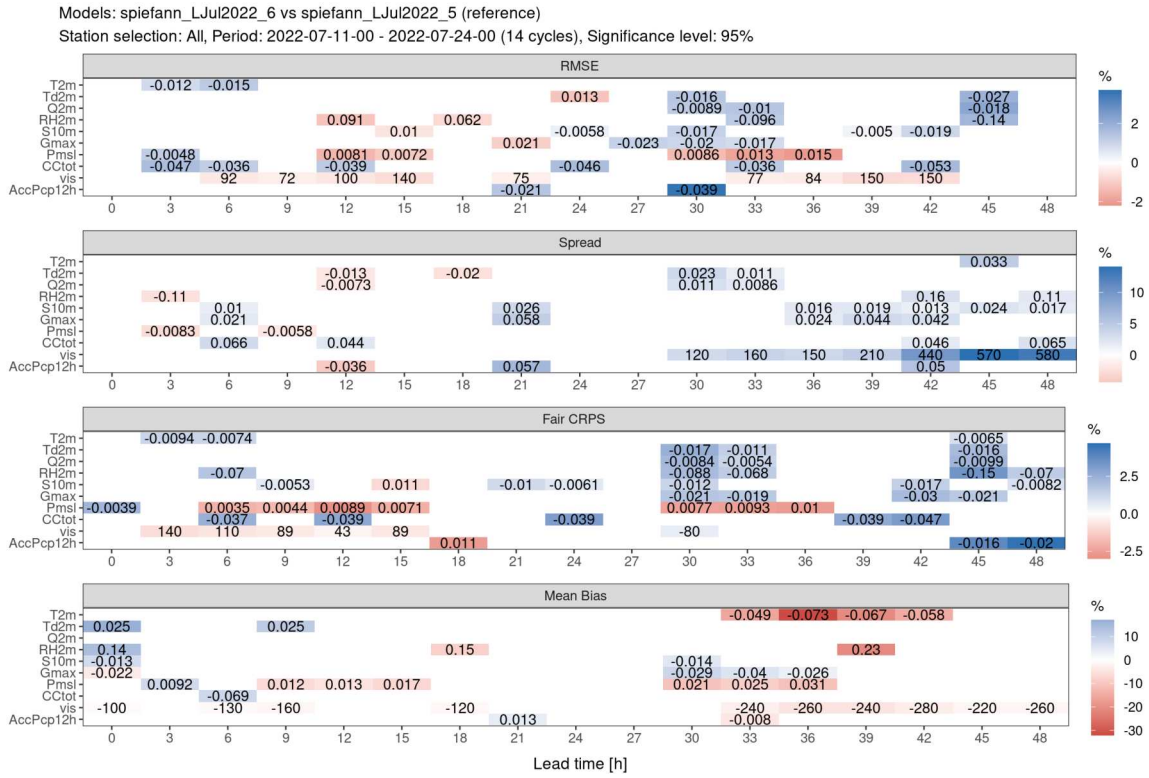


(b) Winter

Figure 35: Surface parameter scorecard comparing experiments SPP_MCET_DP to DPert_DP (reference) over the autumn and winter test periods.

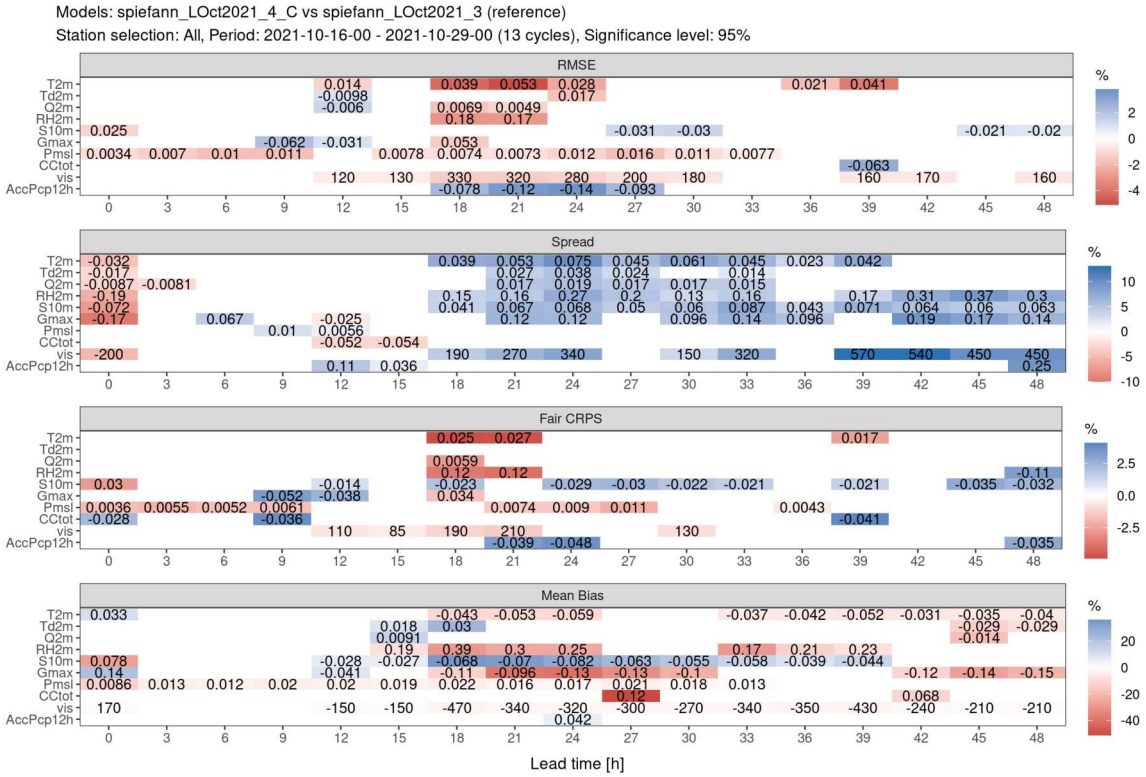


(a) SPP_ATET_SP

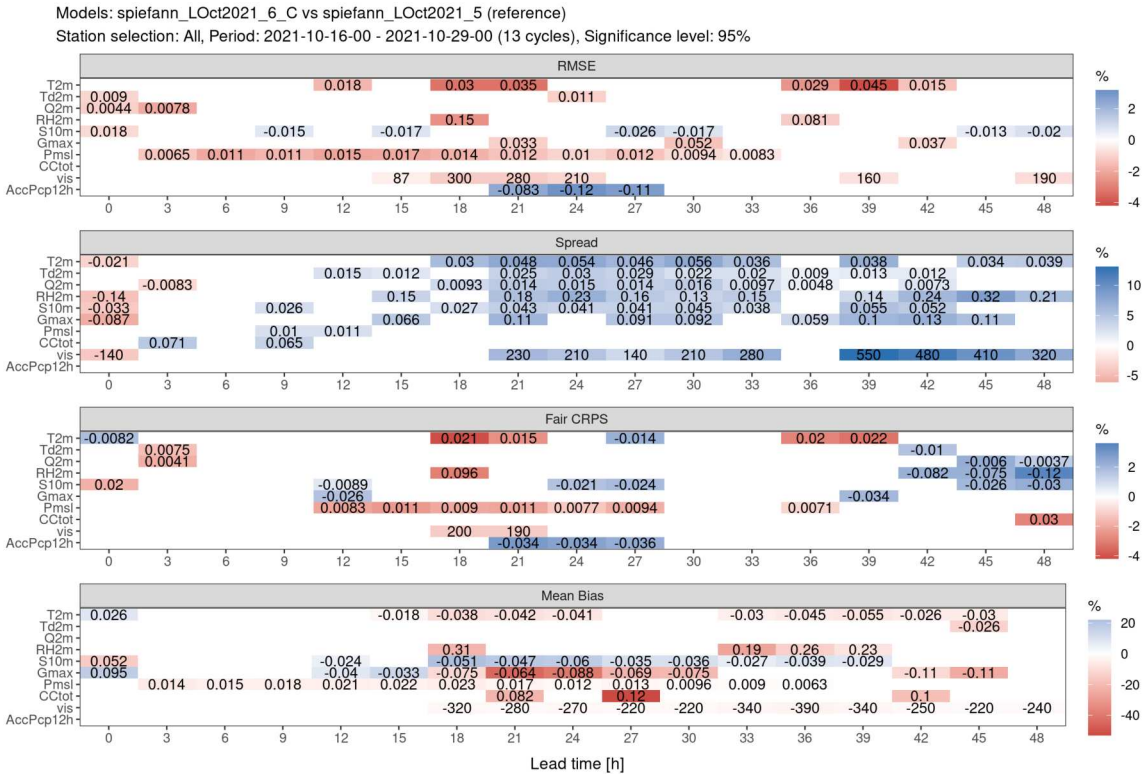


(b) SPP_MCET_SP

Figure 36: Surface parameter scorecard comparing (a) SPP_ATET_SP to SPP_ATET_DP (reference) and (b) SPP_MCET_SP to SPP_MCET_DP over the summer test period. Blue/red indicates an improvement/degradation for the SP ensemble.

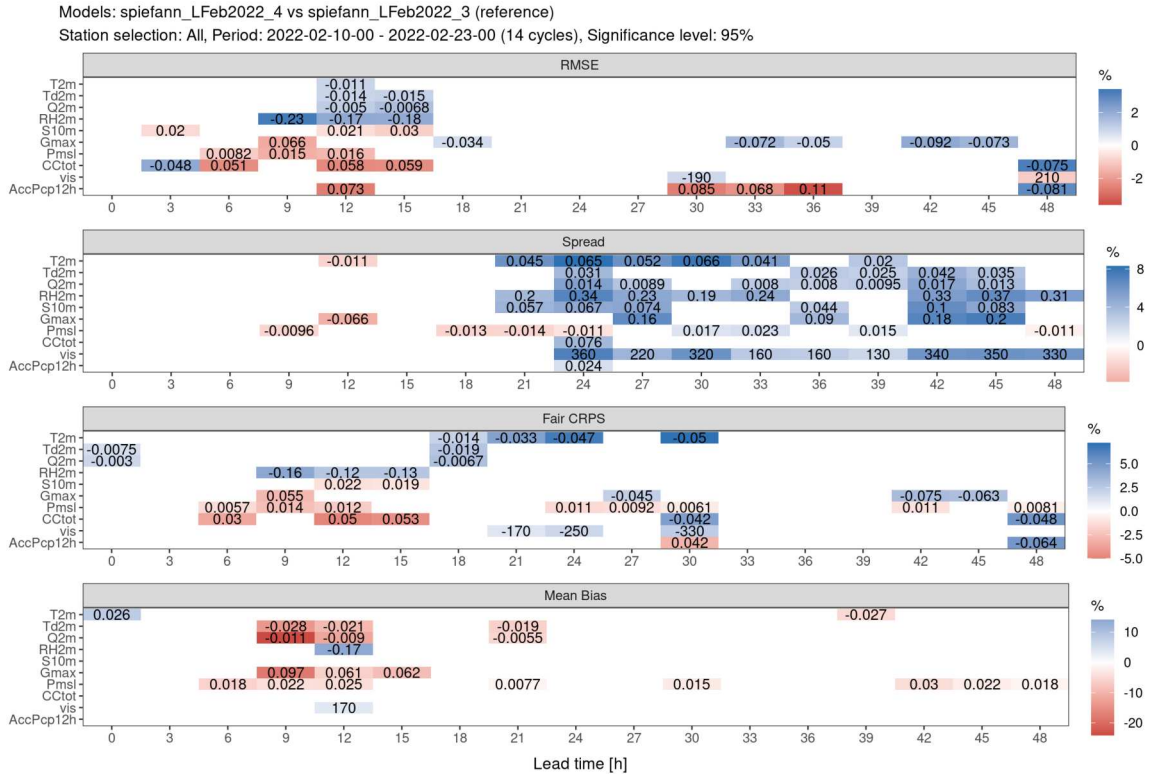


(a) SPP_ATET_SP

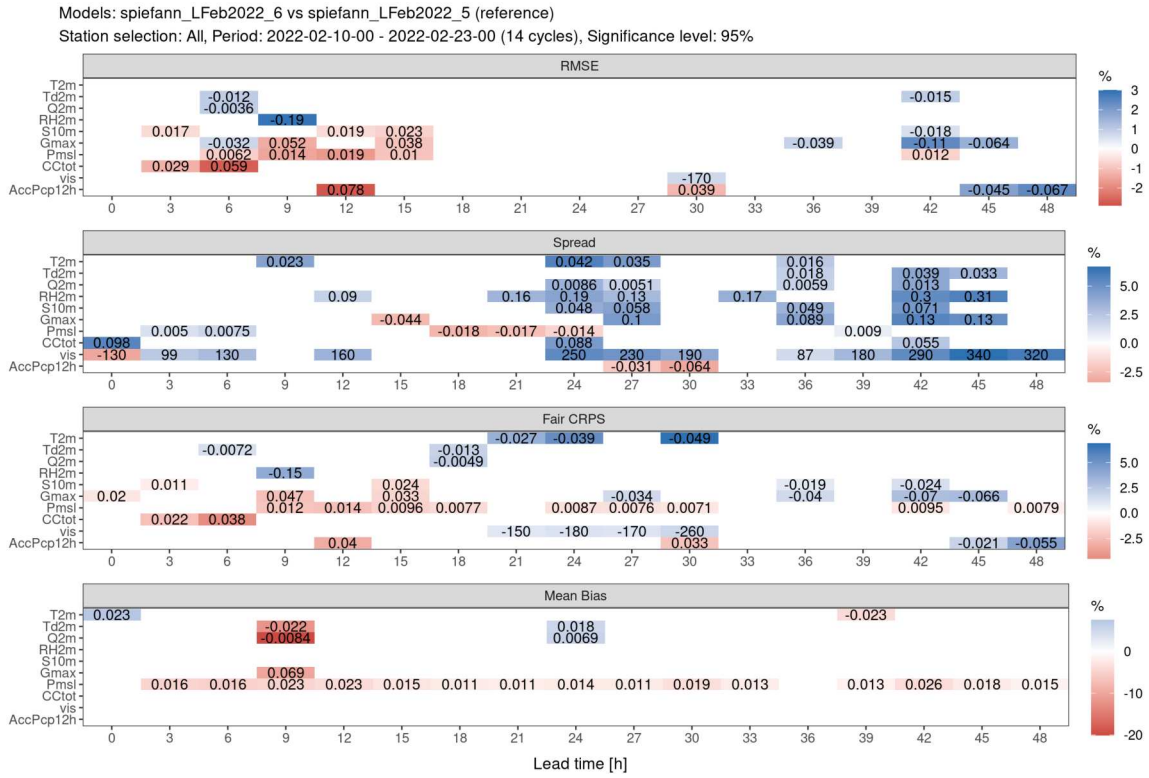


(b) SPP_MCET_SP

Figure 37: Surface parameter scorecard comparing (a) SPP_ATET_SP to SPP_ATET_DP (reference) and (b) SPP_MCET_SP to SPP_MCET_DP over the autumn test period. Blue/red indicates an improvement/degradation for the SP ensemble.

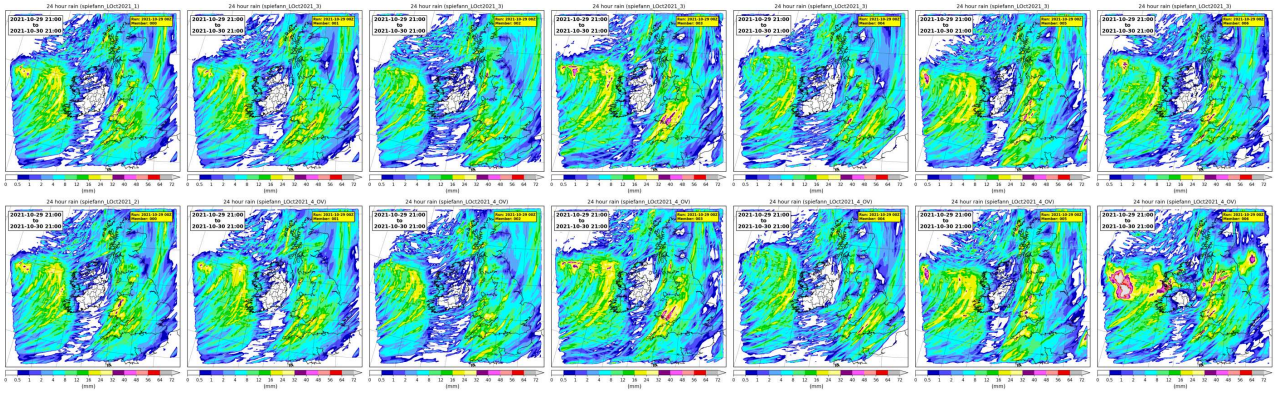


(a) SPP_ATET_SP

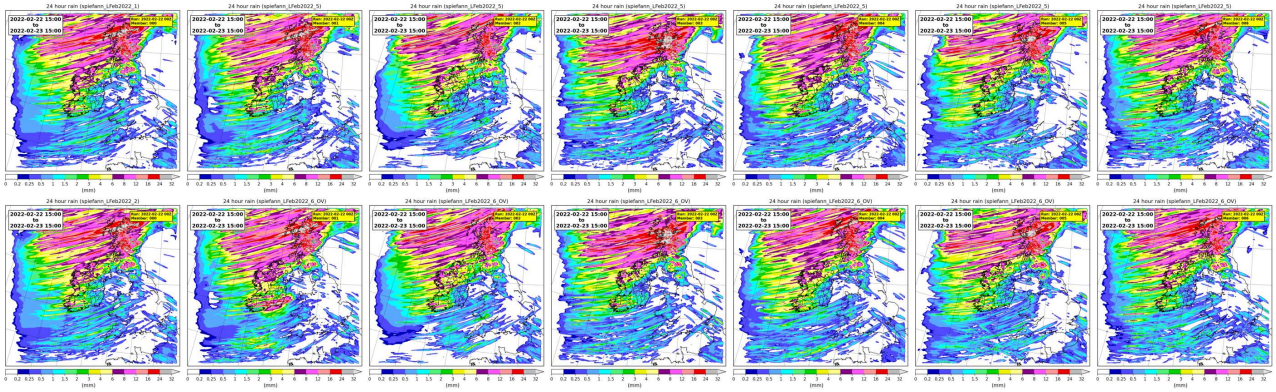


(b) SPP_MCET_SP

Figure 38: Surface parameter scorecard comparing (a) SPP_ATET_SP to SPP_ATET_DP (reference) and (b) SPP_MCET_SP to SPP_MCET_DP over the winter test period. Blue/red indicates an improvement/degradation for the SP ensemble.



(a) 2021/10/29/00+45 forecast



(b) 2022/02/22/00+39 forecast

Figure 39: 24 hour rainfall from (a) the 2021/10/29/00+45 forecast for experiments SPP_ATET_DP (top row) and SPP_ATET_SP_RNM (bottom row) and (b) the 2022/02/22/00+39 forecast for experiments SPP_MCET_DP (top row) and SPP_MCET_SP_RNM (bottom row). Members 0-6 are given from left to right.

References

Bessardon, G., Clancy, C., Daly, C., Darcy, R., Fannon, J., Gleeson, E., Hally, A., Harney, E., McAufield, E., and Whelan, E. (2021). HARMONIE-AROME 43h2.1 upgrade. Met Éireann NWP Note.

Clancy, C., Fannon, J., and Whelan, E. (2022). HARMONIE-AROME 43h2.1 at hectometric scale. Met Éireann NWP Note.

Düben, P. D. and Palmer, T. (2014). Benchmark tests for numerical weather forecasts on inexact hardware. *Monthly Weather Review*, 142(10):3809–3829.

Fannon, J. and Hally, A. (2021). Single precision experiments in HARMONIE-AROME cy43h2.1.1 and harmonEPS-43h2.2. Met Éireann NWP Note.

Feddersen, H. (2021). Running HarmonEPS ensemble members in single precision. ACCORD All Staff Workshop 2021, `http://www.umr-cnrm.fr/accord/IMG/pdf/feddersen_accord_asw2021_singprecmbrs.pdf`.

Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Hämäläinen, K., Kauhanen, J., Ivarsson, K.-I., and Yazgi, D. (2022). Model uncertainty representation in a convection-permitting ensemble — SPP and SPPT in harmonEPS. *Monthly Weather Review*, 150(4):775 – 795.

Hatfield, S., Chantry, M., Düben, P., and Palmer, T. (2019). Accelerating high-resolution weather models with deep-learning hardware. In *Proceedings of the platform for advanced scientific computing conference*, pages 1–11.

Hatfield, S., McRae, A., Palmer, T., and Düben, P. (2020). Single-precision in the tangent-linear and adjoint models of incremental 4d-var. *Monthly Weather Review*, 148(4):1541 – 1552.

HIRLAM (2020). Harmonie 43h2.1 Release Notes. `https://hirlam.org/trac/wiki/ReleaseNotes/harmonie-43h2.1`.

Lang, S. T. K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T., Prates, F., Roberts, C. D., Sandu, I., and Wedi, N. (2021a). More accuracy with less precision. *Quarterly Journal of the Royal Meteorological Society*, 147(741):4358–4370.

Lang, S. T. K., Lock, S.-J., Leutbecher, M., Bechtold, P., and Forbes, R. M. (2021b). Revision of the stochastically perturbed parametrisations model uncertainty scheme in the integrated forecasting system. *Quarterly Journal of the Royal Meteorological Society*, 147(735):1364–1381.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J., and Sandu, I. (2017). Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ecmwf ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143(702):408–422.

Palmer, T. N. (2014). More reliable forecasts with less precise computations: a fast-track route to cloud-resolved weather and climate simulators? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2018):20130391.

Suárez-Molina, D. and Calvo, J. (2021). Very high-resolution experiments at aemet. ALADIN-HIRLAM Newsletter, 16, 106-110.

Váňa, F., Carver, G., Lang, S., Leutbecher, M., Salmond, D., Düben, P., and Palmer, T. (2016). Single-precision IFS (ECMWF newsletter). https://www.ecmwf.int/en/newsletter/148/meteorology/single-precision-ifs.

Váňa, F., Düben, P., Lang, S., Palmer, T., Leutbecher, M., Salmond, D., and Carver, G. (2017). Single Precision in Weather Forecasting Models: An Evaluation with the IFS. *Mon. Wea. Rev.*, 145(2):495–502.

Vignes, O. (2019). Single precision in cycle 43. ALADIN/HIRLAM All Staff Meeting 2019, http://www.umr-cnrm.fr/aladin/IMG/pdf/sp_cy43_olev_asm2019.pdf.