LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
DEPARTMENT FÜR PHYSIK
METEOROLOGISCHES INSTITUT
MIM

MIM · Theresienstr.37 · 80333 München

The Editor
Heini Wernli
ACP
Hymex Special Issue

Dr. Christian Keil

Telefon +49 (0)89 2180-4447
Telefax +49 (0)89 280 55 08

Christian.Keil@lmu.de

Meteorologisches Institut
Theresienstr. 37
80333 München
GERMANY

München, 06/10/2020

Revised Manuscript: acp-2020-508
Dependence of Predictability of Precipitation in the Northwestern Mediterranean Coastal Region on the Strength of Synoptic Control

by C. Keil et al.

Dear Editor,

please find attached a revised version of the above submission to ACP. We are grateful for the positive and constructive criticism that the reviewers made on our submission. With their help, we were able to further improve our manuscript. A point-by-point response to the reviewers' comments is included below.

In the track-changed manuscript any revisions are indicated by changing the font colour to blue to show the modifications made in the course of the revision. We hope that the paper is now acceptable.

Yours faithfully,

Christian Keil

Reply to Reviewers' comments

Reviewer 1

This is an interesting paper. The paper aims to use the convective adjustment timescale to examine the synoptic control on convection in the Mediteranean region of France and Italy over a 2-month autumn period using the AROME convection permitting ensemble, and hence determine the nature of the ensemble forecast predictability and performance in the different regimes.

The manuscript is largely clear and produces worthwhile results, but there are some aspects that need some further attention before it should be published. I'll first outline my main concerns before going into some other less crucial detail.

Main points:

1. You say that the domain should not have an area larger than 500x500km as recommended by Wernli and you do meet that criterion, as you say, but by having a domain800km in the west-east direction are you not in danger of incorporating more than one larger-scale wave and therefore more than one regime (which you are trying to avoid)?Your third case (13th Oct) appears to have two areas of precipitation. One in the west that looks predictable and one further east that looks much less predictable. I wonder if you did your statistics for two overlapping domains each 500x300km whether you would get different results for that type of case and a better partitioning. Would it be possible to try that for at least that case? One of your main conclusions is that the strong forcing cases dominate, but is that partly because the domain is too extended and is always likely to capture a strongly forced event which will contribute the most to the timescale calculation?

The key point in the tau_c calculation is to choose a domain size for which the classification of the synoptic control is representative. As already written in the text '*The choice of the location and size of the investigation domain is carefully chosen and represents a compromise between being large enough to have numerous precipitation events giving good statistics, but small enough to comprise a specific and unambiguous meteorological situation in combination with the good coverage of rainfall observations in the Northwestern Mediterranean. If the domain is too large strongly differing meteorological systems may be contained and the results obtained using area averages may be blurred and not representative. However, we believe that the chosen domain … represents a good compromise being at the scale of the Rossby radius of deformation.*' Even before the original submission of the manuscript we divided the domain, as you suggest, into a French (West) and Italian (East) domain, but did not find important differences and decided to use this single Northwestern Mediterranean domain for the present study. Following your suggestion we repeated this for the overlapping domains covering the entire period, but did not find noteworthy differences. For instance, the tau_c values for the third case you mention amount to less than 30 min for all three

domains (NW Med, Western, Eastern). But you are right, if the domain is too large, i.e. larger than the Rossby radius of deformation (roughly 1000km), then you might mis-classify the predominant synoptic control.

Action: None.

2. Many times you say that the ensemble overpredicts the rain in the non-equilibrium cases, but I wonder how much of that is actually an artifact of the gauge interpolation missing rain than an over-prediction by the model. In a showery situation it is very likely the gauges will miss the heaviest rain cores, especially if locally focussed on hills where there are fewer gauges. A good test would be to take a model field, extract the values at the gauge locations and then do the interpolation. I would suspect you will get a lower domain-average value than the original field. Even if you don't try that out, it would still be worth mentioning in the article that a guage interpolation can miss rain when the rain coverage is low and has local spikes.

You are right, rain-gauges are prone to miss the heaviest rain cores in convective weather situations. Here a blend with radar observations yields a more realistic picture (taking into account all the uncertainties inherent in radar observations). We are totally aware of this shortcoming, and wrote in the final section of the original manuscript '*One reason for the apparent overprediction of precipitation during weak control can partly be accounted for by the point type character of rain-gauge measurements that sample the spatial highly heterogeneous and intermittent nature of locally triggered convective precipitation insufficiently. This discrepancy calls for remotely sensed spatial rainfall measurements of high quality, that were not available in the present study.*' To compute the conventional measures RMSE, ROC and reliability we used the nearest neighbour method and extracted the forecast values at the gauge location. In Fig. R1 we show the difference of taking the mean of the model field or the nearest neighbour forecast, respectively, and do not find important differences for the weakly forced case.

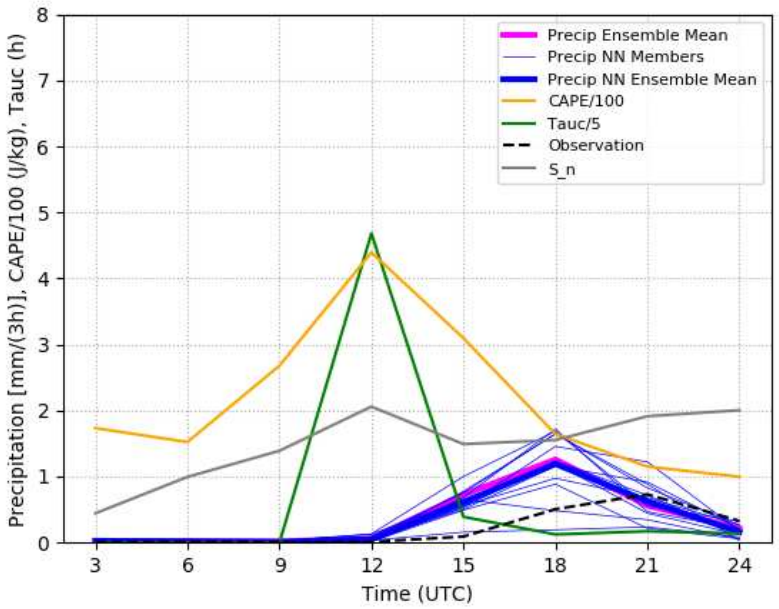

Figure R1: As Fig.6b in the manuscript, additionally depicting the ensemble mean precipitation based on the different calculations (Precip vs Precip NN Ensemble mean).

Action: We added in subsection 2.2. the sentence: We are aware that pointwise rain-gauge measurements can miss rain when the rain coverage is low and has local spikes, typically for weakly forced convective situations.

3. Some aspects of the methodolgy need a bit more clarity. You don't say why you choose a minimum rain amount of 3mm in 3h, which seems somewhat arbitary. A sentence or two about that would be helpful. I know you have a reference, but a few words would still be helpful. Linked to that, do you try to determine whether rain is convective or stratiform in nature? I assume you don't, but then there will sometimes be frontal rainbands that act to lower the timescale because rain occurs along with zero CAPE. You should at least mention this potential difficulty or explain why you think it isn't a problem. I can see it making the convective timescale appear smaller especially going later into the year. Why do you choose 1mm/3h for the standard deviation, but 3mm/3h for the convective timescale calculation? What horizontal scale of Gaussian smoothing do you use? That could potentially have a singificant effect. If the smoothing is done over too large an area the precipitation threshold may not get exceeded anywhere in the scattered convection cases. On the other hand, some degree of smoothing will bring in more locations that may have high CAPE and that will affect (lengthen) the convective timescale. I take it, just for absolute clarity, you do not include any points with rain < 3mm/3h even if CAPE in non-zero?

The rationale of taking 3mm/(3h) stems from earlier work (e.g. Keil et al. 2014, Kühnlein

et al. 2014) using a rain rate of 1 mm/h in raw, deterministic forecasts of convective scale models to separate rainy from non-rainy gridpoints in the tau_c calculation, since dry gridpoints preclude the computation. Due to the availability of 3-hourly data only, we temporally upscaled this threshold value. And yes, no gridpoints below this threshold value are used in the tau_c computation. On the other hand, there is a threshold needed to normalize the STDDEV of precipitation. This threshold is applied on the ensemble mean precipitation, not on individual members. Given the intra-member variability in precipitation forecasts we chose 1mm/(3h) to perform this normalisation.

Yes, the horizontal scale of Gaussian smoothing can have significant effects. Since tau_c represents an environment in which convection occurs, it is necessary to smooth the fields prior to the calculation. In the present work using convective scale models we kept the kernel with a half-width size of 20 GP (50 km) to stay consistent with the body of existing literature in which tau_c has been applied. In this study there is no additional distinction into stratiform or convective nature of precipitation, as was previously done based on CAPE values in e.g. Kober et al. 2014 and Grazzini et al. 2020, since precipitation is predominantly convective, and even frontal rainbands have considerable convective elements in the autumn season in the target region.

Action: The methodology is clarified in Section 2.

4. To be honest I'm not sure about the value of some of the discussion of the individual cases It is useful to see the figures but there is a lot of descriptive text around them that isn't really adding much to the purpose of the paper, it's just describing where rain occurs in that event. A lot of readers will not know the location of regions in France, so it would be better to have some annotation on the figures to point to features instead. The key thing it seems is whether the rain is more or less widespread, and whether the members agree in say region "A" and region "B". Again, I'm not so convinced about the overestimation argument in weak control.

We partly agree and reduced the descriptive text as well as some of the geographical terms in the text. However, we disagree that using region A vs region B increases readability and only kept key geographical terms shown in Figure 1.

Action: Text adapted throughout Section 4.

5. You should explain what you mean by the "ensemble mean FSS". Do you generate an ensemble mean precipitation field and then threshold for the FSS? That wouldn't be a good thing to do because you filter the true ensemble spread and change the frequency biases. Do you calculate the FSS for each member based on those thresholds and then take the average? Again that would not be the best thing to do because you might penalise ensemble spread as much

as ensemble error. Do you threshold each member then take the ensemble mean of the binary probabilities and then apply the FSS? That would be the most sensible of those three options because it is evaluating the final probability field without clipping the distribution. Maybe you do something else? Definately it was good to choose the 95th percentile. Have you looked at other percentiles?

This is a good point and we are grateful that you raised this issue. Originally we calculated the FSS of each member before averaging. Now we follow your recommendation and apply the FSS on the ensemble mean of the binary probabilities. This results in slightly lower FSS values in Figure 13. Inspecting the 95th percentiles in Figure 14, there is an even clearer difference between both weather regimes. In particular, the size of the boxes during weak control is reduced.

Yes, we looked at other percentiles and present the results for the 85th percentiles in Figure R2. There is still a clear, albeit somewhat smaller distinction in spatial forecast quality between both. In particular the variability during strong control increases. Note that there are very low values for the 85th percentiles of observed precipitation during weak control. The use of even lower percentiles is limited by the spatial extent of precipitation during weak control.

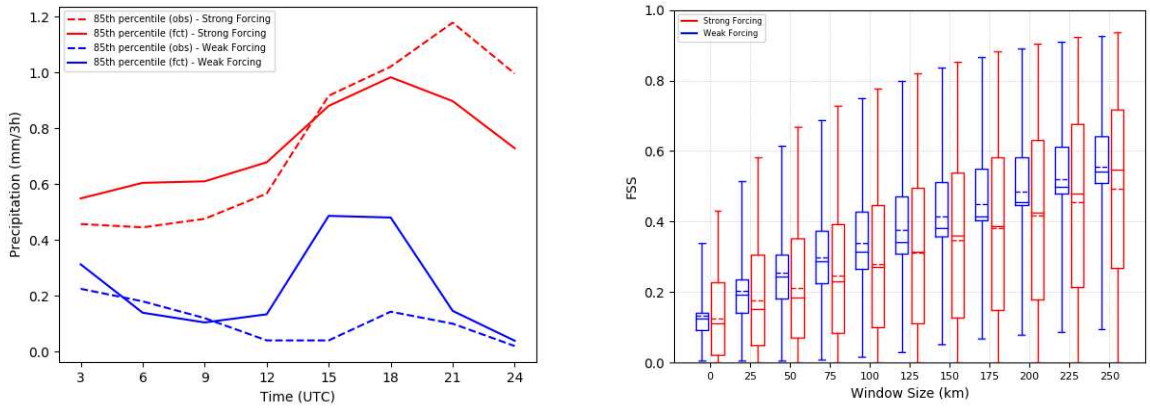

Figure R2: As Fig.14 in the manuscript but for 85th percentiles of precipitation.

Action: New Figures 13 and 14 and text adapted in Section 5.

Other points:

1. You choose a threshold timescale of 3h, but then say it is different for summer and autumn (understandably). If you are most interested in partitioning

out the days with the strongest and weakest synoptic control for evaluation could you just take the highest 30% and lowest 30% and then not have to worry about a timescale threshold. I'm not suggesting you do that here (as the results would be very similar), but it may help further studies of this sort.

Thank you for this interesting suggestion. We will apply this in future studies.

Action: None.

2. When you talk about a "barrier" I assume you mean a stable layer? Sometimes storms form over mountains because of elevated heating and there isn't a clear in-version or organised storms form where there is an inversion but lifting mechanisms reduce it.

Yes, for instance a stable layer or the presence of convective inhibition.

Action: Clarified in the text.

3. In some ways this seems to come down to wether the precipitation is contingous or broken. I wonder if you were to classify that way whether you'd get something similar?

We guess so, and we are curious if you are aware of a suitable measure to classify texture. In the past we used the fractional coverage as predictor, but this quantity does not really describe the texture of the precipitation field.

Action: None.

4. What would a graph of rain against Tau look like?

This is shown in Figure R3. There is no new insight in view of a classification of weather regimes apart the fact that there tends to fall more rain during strong control (i.e. small Tau values).

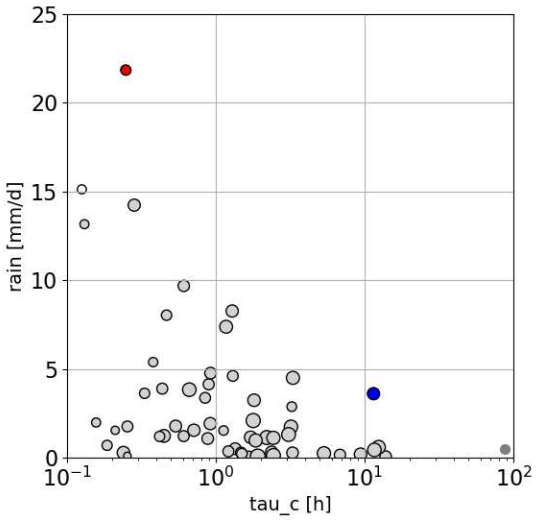

Figure R3: Scatterplot of Tau and rain amount.

Action: None.

5. Figure 2 caption is hard to follow.

Ok.

Action: Clarified.

6. Might be interesting to know something about the spread of Tau and CAPE. Not suggesting you have to include that though.

We leave this for future work.

Action: None.

7. You specify rank correlation values, but not actual correlation values. That leaves me a bit suspicious that they are not as good (closer to zero). Is that the case?

Yes, these amount to 0.34. Since we are primarily interested in the pairing of data we chose the rank correlation, a statistical quantity describing exactly this (Wilks 2011) and waive to show the Pearson correlation value in the text.

Action: None.

8. In line 163 you say "provides a better suitable measure" - but better than what?

Better than, for instance, CAPE values.

Action: Clarified in the text.

9. For the indivuidual cases a few pressure contours would be informative to help set the context for the precipitation.

We presently don't have the data retrieved and think such isobars don't give new insights.

Action: None.

10. Presumably the skill-spread scores are picking up the bias, but also indicating potentially that there are too few members when evaluated at the grid scale, as well as saying the spread isn't sufficient.

Yes, we agree, a 12-member ensemble is at the lower edge in terms of ensemble size.

Action: None.

11. I'm not sure it is a huge surprise that convection linked to mountains is more spatially predictable than convection that is mobile. I wonder what would happen if you just fixed a domain over the Alps and compared that with a flatter region for the convective timescale partitioning?

Yes, we generally agree and changed the text (see reviewer 2 comments, too). In this region it is difficult to fix a flat domain that is large enough to be informative (see above arguments on the domain size).

Action: Clarified.

12. It might be worth also mentioning the paper by Flack et al Flack, D.L.A., Plant,R.S., Gray, S.L., Lean, H.W., Keil, C. and Craig, G.C. (2016), Characterisation of convective regimes over the British Isles. Q.J.R. Meteorol. Soc., 142: 1541-1553.doi:10.1002/qj.2758 This also references the papers you reference prior to 2016 along with Keil and Craig 2011, which you don't reference - which is a surprise. There are some places where the text could be clearer (although in general it is well structured and readable), but it would probably be better to address these after dealing with the points above, which are going to involve changes in the text.

Action: We added these references, thank you.

Reviewer 2

The paper discuss the performance of the AROME-EPS ensemble forecast for precipitation during the SOP of the Hymex Project, in dependency of the predictability of the events, as quantified by the convective adjustment timescale. The argument is scientifically very relevant, addressing the convective-scale predictability of the precipitation for an area interested by severe weather events. The work is well structured and meaningful, and clearly presented. However, I am not convinced of some conclusions, due to the verification process. I think there are some weaknesses in the verification interpretation, which hamper the conclusions to be drawn. Therefore, I recommend to address some issues (described below), in particular in Section 5, before publishing the work.

Detailed Comments

Section 2.3 – Please add a reference for the Relative Operating Characteristics ROC and the reliability diagram. Though a description of these well know tool is not needed in the paper, not all the readers may be familiar with their definition.

We added these references:

Wilks, D.S. (2011) Statistical Methods in the Atmospheric Sciences, 3rd edition. Elsevier Academic Press, Amsterdam, Netherlands.

Jolliffe IT, Stephenson DB. 2011. Forecast Verification: A Practitioner's Guide in Atmospheric Science (2nd edn). John Wiley and Sons: Chichester, UK, doi:10.1002/9781119960003.ch1

Action: Done.

Figure 2: The thin lines are for me unreadable, and particularly their colour. Is it possible to increase the thickness?

The thickness of the thin lines is increased.

Action: Done.

Page 8: A small typo: "southeastern foothills if the Massif Central" should be "of"

Thank you.

Action: Done.

Section 4 – In figure 6 also the spread of the ensemble is shown, by showing the area average precipitation of the members. There is evident that in the

first case the spread is relatively "high", the members being quite different, as also noticed in the discussion of figure 7(b). In the second case, the spread of the precipitation is low, only 2 members having almost no rain, while all the others are close to each other. However, the first case is a predictably one, and the second a less predictable one. The spread, in the case of the precipitation, is not a good indicator, because it depends too much on the amount of precipitation itself: the first case is a predictable one, even if the members are different, because the rain is intense and the differences do not affect the "general performance" of the forecast. I think that, if the ensemble spread is shown, these considerations have to be made explicitly, otherwise the reader may receive a wrong message about the predictability. On top, the spread may be low even when the case is not well predicted, in case the ensemble is overconfident, which seems to be the case of the second case. For this reason, it would be good to have also the average observed precipitation, in figure 6.

Thank you, we agree and completely revised Fig.6 that includes the average observed precipitation and the normalized ensemble spread *S_n* now as well. It is important to distinguish relative (normalized; *S_n*) and absolute ensemble spread. In absolute terms there is a larger variability during strong forcing cases (due to higher rainfall intensities, see e.g. individual members in Fig.6c). The normalized spread *S_n* gives indication on predictability and is evidently larger for the second case pointing towards the below average predictability. Additionally the y-coordinate has been 'rescaled' to highlight the differences between the three prominent cases.

Action: New Figure 6 and corresponding discussion in Section 4.

Figure 7 and related discussion: an overprediction over Genova is noted in the 6-hperiod. Is this an overprediction in absolute sense or a timing problem (e.g. heavy precipitation occurred over Genova in the successive 6 hours?)

Figure R4 indicates that there is no major timing error for IOP16a in this region. In the successive 6 hours (18-24 UTC) the rainfall amounts compare better in the proximity of Genova, whereas rainfall is overestimated further to the east.

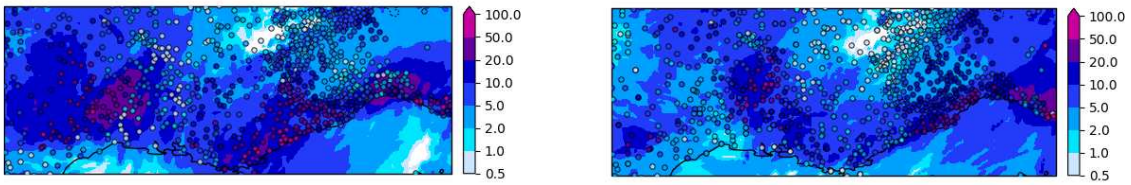

Figure R4: Illustration of 6-hourly precipitation for IOP16a valid 26 October (left) 12-18 and (right) 18-24 UTC.

Figure 11: the mean of the RMS error of the ensemble members is shown and compared with the ensemble spread. Why is not shown instead the RMS error of the ensemble mean, which is the quantity which should be matched (statistically) by the spread? The chosen quantity is for sure higher than the other one, since the ensemble mean has (statistically) lower RMSE than all the members.

Thank you for pointing this out. Actually we computed the RMSE of the ensemble mean and corrected the text.

Action: Caption and text corrected.

Can you motivate a bit more the sentence (pag. 15):"The larger distance of the ROC curve points during strong control indicates the higher absolute spread when 3-hourly (and daily) precipitation accumulations are averaged over the entire SOP1."? Is this related to the point I raised about Section 4?

We clarified this point: "The larger distance of the ROC curve points from the diagonal (resulting in larger concavity) during strong control indicates greater event discrimination when 3-hourly (and daily) precipitation accumulations are averaged over the entire SOP1."

Action: Text changed.

Page 15, about the sentence: "the forecast probabilities are consistently too large relative to the conditional observed relative frequencies. This is an indication of overforecasting equivalent to a wet bias.". The overforecasting in probability/frequency does not indicate a bias in the quantity, but in the probability. Therefore it does not indicate a wet bias, but an overconfidence of the ensemble. The members forecasting an event (e.g. 3mm/3h) are "too many" (therefore producing a too high probability of occurrence) with respect to the observed "probability" (which is the frequency) of occurrence of that event in the sample. I believe that the same overconfidence applies also to the dry areas (here you are considering only the wet areas, since you have a threshold >3mm/3h) and it is not related to an overestimation in the quantity itself.

Thank you for this clarification, we appreciate your comment and agree.

Action: Text changed accordingly.

Page 17, from line 315 to the end of the Section. I am not convinced by the conclusions drawn here by the authors. "Taking this bias into account by using precipitation percentiles results in a superior spatial forecast quality during weakly forced regimes(Fig. 14b). Thus forecasting the location of heaviest precipitation in the afternoon (ex-pressed by the 95th percentiles)

is better during comparably quiescent synoptic-scale atmospheric conditions. This is at first sight an unexpected and surprising result." It is not the scattered nature of the weakly-forced precipitation field, when the isolated in-tense precipitation spots are selected, which gives an impression of skill by upscaling, just because somewhere a spot of precipitation is always available? I do not think that it is possible to conclude that there is a higher quality in the spatial forecast in case of weakly-forced cases based on this result with the FSS. I agree that orography "keeps" the precipitation in place in case of convection, but I am not sure that with this increase of FSS for the 95th percentile can be a prove of skill.

Thank you, however we disagree with your point. By taking the 95th percentiles of precipitation the 5% gridpoints are selected that receive the strongest precipitation, independent to the synoptic control. These upper 5% gridpoints are fairly scattered in both regimes. The FSS of the 95th percentiles gives only information on the position of the strongest precipitation, but *not* on intensities. Considering the 6h accumulated precipitation between 12-18 UTC (the 'convective period') the finding on superior *pure* spatial accuracy of the 95th percentiles coincides with common meteorological sense that deep convection and thunderstorms are strongly linked to orography during weakly forced situations. In contrast, the *timing and position* of heaviest rainfall during strong synoptic control is crucially linked to the cyclone track and only modulated to a minor degree by orography.

Action: We emphasize that we talk on the *pure* forecast location accuracy at various places in the entire manuscript when examining the FSS of the 95th percentiles of precipitation.

Dependence of Predictability of Precipitation in the Northwestern Mediterranean Coastal Region on the Strength of Synoptic Control

Christian Keil[1], Lucie Chabert[1], Olivier Nuissier[2], and Laure Raynaud[2]

[1]Meteorologisches Institut, Ludwig-Maximilians-Universität, Munich, Germany
[2]CNRM (Météo-France & CNRS), 42 avenue G. Coriolis, 31057 Toulouse Cédex, France

Correspondence: Christian Keil (christian.keil@lmu.de)

Abstract. The weather regime dependent predictability of precipitation in the convection permitting kilometric scale AROME-EPS is examined for the entire HyMeX SOP1 employing the convective adjustment timescale. This diagnostic quantifies variations in synoptic forcing on precipitation and is associated with different precipitation characteristics, forecast skill and predictability. During strong synoptic control, which is dominating the weather on 80% of the days in the 2-months period, the domain integrated precipitation predictability assessed with the normalized ensemble standard deviation is above average, the wet bias is smaller and the forecast quality is generally better. In contrast, the pure spatial forecast quality of most intense precipitation in the afternoon, as quantified with its 95th percentiles, is superior during weakly forced synoptic regimes. The study also considers a prominent heavy precipitation event that occurred during the NAWDEX field campaign in the same region, and the predictability during this event is compared with the events that occurred during HyMeX. It is shown that the unconditional evaluation of precipitation widely parallels the strongly forced weather type evaluation and obscures forecast model characteristics typical for weak control.

1 Introduction

The Mediterranean region is affected by intense precipitation events every year particularly during the autumn months. Very high rain amounts and ensuing flash floods can cause widespread damage. Accurate prediction of these precipitation events is crucial to take precautions, warn the public and mitigate potential consequences. The HyMeX (Hydrological Cycle in the Mediterranean Experiment) field campaign was designed to advance the knowledge of Mediterranean heavy precipitation and flash-flooding events, to improve numerical models and to examine the representation and predictability of high-impact weather events (Ducrocq et al., 2014, and references therein).

Precipitation represents a very important yet challenging forecast variable due to the involvement of many atmospheric variables and the role of inherently highly non-linear processes in its formation. Forecasting precipitation with numerical weather prediction models requires among others a sufficiently fine model resolution to explicitly represent important processes like deep convection and a precise description of the microphysical processes leading to precipitation, but also an ensemble approach to quantify the forecast uncertainty.

In the last decade ensemble prediction at the convection permitting kilometric scale has become a standard technique for weather forecasting and provides an important tool to forecast forecast uncertainty. However, the intermittency and spatiotemporal variability of precipitation on the kilometric scale renders the assessment of accuracy even more difficult. Next to the probabilistic approach an evaluation of high-resolution ensemble forecasts of precipitation calls for spatial measures to assess forecast quality.

The predictability of weather in general, and precipitation in particular, is weather regime dependent (Anthes, 1986; Bauer et al., 2015; Yano et al., 2018). The prediction of precipitation is influenced by the synoptic-scale environment and local processes and instabilities. Previous results suggest that there is higher forecast quality and above average predictability, that is lower uncertainty during strong synoptic control. The convective adjustment timescale offers an objective measure to classify weather regimes into strong and weak synoptic control on precipitation (e.g. Keil et al., 2014; Surcel et al., 2017). For instance, Schwartz and Sobash (2019) apply this diagnostic and conclude that forecast quality is related to forcing strength, with higher accuracy in more strongly forced regimes over the conterminous United States.

In the present study we aim to systematically identify different predictability regimes of precipitation in southeast France and northwest Italy during autumn 2012, for which the HyMeX campaign offers an unprecedented transnational observational dataset to validate convective scale ensemble prediction systems (Ducrocq et al., 2014). This period extends from 5 September to 5 November 2012 of which 59 days experienced noteworthy precipitation and includes numerous well studied IOPs of high impact weather situations. Here, forecasts of the kilometric scale AROME-EPS system are evaluated for the first time with neighbourhood methods to examine the spatial distribution of precipitation and to infer on different predictability levels conditional upon the weather regime of the day.

Previously Bouttier et al. (2016) evaluated the AROME-EPS system over the full HyMeX SOP1 period and identified strengths and weaknesses. The impact of initial conditions and model surface perturbations show a significant effect on the ensemble performance. Using a variety of conventional scores like RMS errors and spread-skill relation complemented with the probabilistic measures rank and ROC diagrams applied on near surface variables they found specifically for precipitation an almost negligible impact of direct surface perturbations and a lack of spread. The present study extends this comprehensive work, focuses on ensemble precipitation forecasts over the contiguous 2-months period and adds the aspect of weather regime dependent predictability.

Nuissier et al. (2016) presented a probabilistic evaluation of two convection permitting EPSs for the full HyMeX SOP1 and document a slightly better performance of AROME-EPS forecasts in terms of discriminating behaviour and reliability of 6-hourly rainfall. However, results depend on the choice of the verification domain since small areas suffer from sampling issues and the occurence of precipitation events. The examination of the HyMeX 'golden case' (IOP16a on 26 October 2012, Ducrocq et al., 2014) reveals slightly different predictability levels in two different subdomains in which precipitation is crucially governed by the location and deepening of a surface low pressure system over the Mediterranean Sea controlling the southerly moist low-level flow. Earlier, Hally et al. (2014) investigated the sensitivity of precipitation forecasts in an experimental convective-scale ensemble based on the Meso-NH model to diverse initial and boundary conditions and microphysical uncertainties for two IOPs (IOP6 and IOP7a in southeast France). Since both cases developed under strong synoptic forcing

the impact of atmospheric conditions on the spatiotemporal distribution of precipitation outweighs the one of surface conditions. It is suggested that the specific influence of surface conditions is larger for weakly forced events. Recently, Fourrié et al. (2019) revisited the HyMeX SOP1 period and demonstrate improved rainfall forecasts with a second reanalysis using 24% more additional data in the AROME system. The superior performance is specifically illustrated employing the conventional scores frequency bias and equitable threat score on 24h rainfall accumulations for an intense precipitation event that occurred over Spain and southern France on 29 September 2012 (IOP8).

Beyond HyMeX the downstream impact of synoptic systems on the predictability of high impact weather in the Mediterranean has been one of the science goals of the NAWDEX (The North Atlantic waveguide and downstream impact experiment) campaign in autumn 2016 (Schäfler et al., 2018). One of the NAWDEX highlights represents IOP9 on 13 October 2016 when the 24 h accumulated precipitation in southeast France reached 250 mm ahead of the cyclone Sanchez. This prominent case is included in the present study to be compared with the 2-months HyMeX SOP1.

This article focuses on precipitation predictability examining temporally highly resolved forecasts (3-hourly) of the AROME-EPS and relates differences in ensemble spread and forecast skill to broader environmental characteristics for the entire HyMeX SOP1 period. The remainder of the paper consists of a methods section, followed by a classification of the 2-months period into weather regimes, an illustration of three prominent cases, the verification using classical gridpoint based quality measures, probabilistic metrics and a spatial score to allow for location tolerance, and finishes with conclusions.

2 Model, Data and Metrics

2.1 The convective-scale ensemble

The AROME-EPS used in this present study is based on the AROME forecasting system largely described in Seity et al. (2011) and in Brousseau et al. (2016). It is based on adiabatic, non-hydrostatic equations from the limited-area ALADIN (Aire Limitée Adaptation dynamique Développement InterNational) model (Bénard, 2004; Bubnova et al., 1995). A horizontal resolution of 2.5 km and 60 (HyMeX ensemble) or 90 (Nawdex ensemble) vertical levels are used in this study. AROME shares the same physical parameterizations as the research model Meso-NH (Lac et al., 2018), including a bulk one-moment microphysics scheme following Caniaux et al. (1994), which represents six water species (water vapour, cloud water, rain water, primary ice, graupel and snow). The representation of the turbulence in the planetary boundary layer is based on a prognostic turbulent kinetic energy (TKE) equation combined with a diagnostic mixing length (Bougeault and Lacarrère, 1989). The TKE scheme used in AROME was developed by Cuxart et al. (2000) and the scheme is derived from the full set of equations for second-order moments. At 2.5-km resolution, the deep convection is assumed to be explicitly resolved by the model's dynamics. However the shallow convection requires a parameterization of subgrid effect for which the Pergaud et al. (2009) scheme is used. It is a mass flux scheme based on the eddy diffusivity mass flux (EDMF) scheme (Soares et al., 2004) that parameterizes dry thermals and shallow cumuli.

The AROME-EPS ensemble setup is the following: a) The ensemble comprises 12 members. For the HyMeX period ensemble simulations start at 00:00 UTC (up to 36-h forecast range), whereas ensemble runs are initialized at 21:00 UTC (up

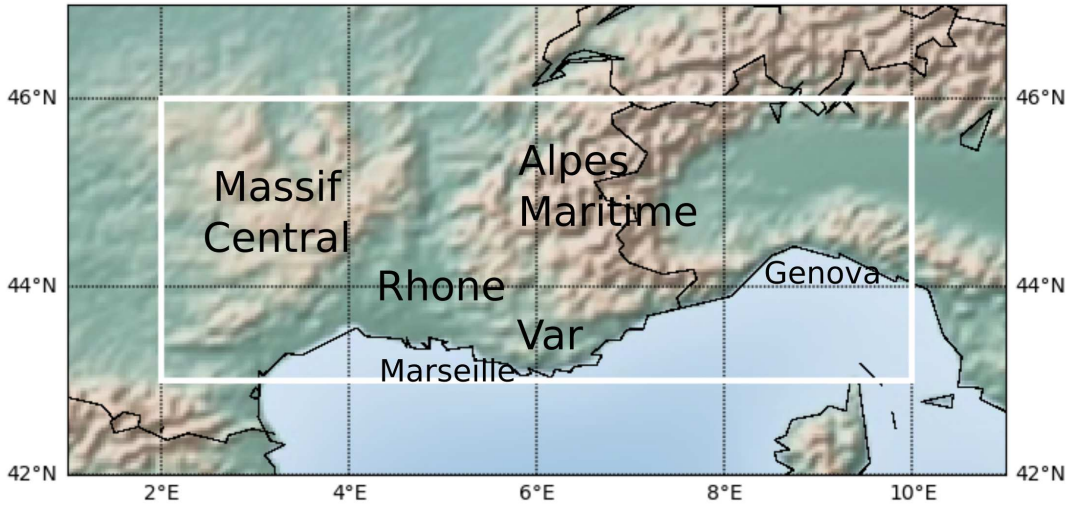

Figure 1. Map illustrating the Northwestern Mediterranean domain (white box) and geographical landmarks used in the text.

to 45-h forecast range) for the Nawdex case. b) In the ensemble simulations, AROME is driven by the global short-range ARPEGE-EPS (Descamps et al., 2014), called hereafter PEARP. Firstly, a subset of 12 members of the PEARP is selected according to the Nuissier et al. (2012) technique. The PEARP 35-member ensemble forecasts are classified by a complete-linkage clustering technique (Molteni et al., 2001). c) The initial conditions are provided by adding downscaled forecast perturbations of the selected PEARP members to the AROME operational analysis (Raynaud and Bouttier, 2017). d) Atmospheric model errors are represented through the so-called SPPT scheme (stochastic perturbation of physics tendencies) described in Bouttier et al. (2012), which simulates the effect of random errors due to the physical parametrizations. e) Finally, random perturbations are added to various parameters of the surface externalisée (SURFEX) surface scheme, including for instance sea-surface temperature, soil moisture and temperature perturbations (Bouttier et al., 2016).

2.2 Domain and observational data

The investigation domain extends across 300 km $\times$ 800 km and encompasses southeastern France and northwestern Italy including the coastal regions of Cote d'Azur and Riviera as well as adjacent mountainous regions of the Massif Central and the Alpes Maritimes (Fig. 1). This region, that is herein called the Northwestern Mediterranean, is prone to heavy precipitation generated by a wide variety of flow conditions including synoptic systems characteristic of Rossby wave breaking at the eastern end of the North Atlantic storm track, modulated by orography and thermal contrasts of the Mediterranean basin as well as calm, conditionally unstable situations requiring trigger mechanisms to generate rainfall (e.g. Ducrocq et al., 2014; Nuissier et al., 2016, and references therein). The choice of the location and size of the investigation domain is carefully chosen and represents a compromise between being large enough to have numerous precipitation events giving good statistics, but small enough to comprise a specific and unambiguous meteorological situation in combination with the good coverage of rainfall

observations in the Northwestern Mediterranean. If the domain is too large strongly differing meteorological systems may be contained and the results obtained using area averages may be blurred and not representative. However, we believe that the chosen domain encompassing 300 km × 800 km represents a good compromise being at the scale of the Rossby radius of deformation. The domain size conforms with the recommendation of Wernli et al. (2009) to use areas smaller than 500 km × 500 km to compute an unequivocal spatial forecast quality value representative of a certain meteorological situation.

As observational data we use 3-hourly rain-gauge observations retrieved from the HyMeX database and hourly rain-gauge observations for the NAWDEX case that are accumulated 3-hourly. The rain-gauge observations are spatially interpolated to the model grid using a linear barycentric interpolation to perform the spatial evaluation. We are aware that pointwise rain-gauge measurements can miss rain when the rain coverage is low and has local spikes, typically for weakly forced convective situations.

2.3 Metrics and measures

Generally, an ensemble of forecasts provides a range of possible scenarios allowing for the estimation of forecast uncertainty. Large deviations of individual ensemble members point towards large forecast uncertainty and low predictability. One method to estimate the predictability of precipitation is the computation of the normalized standard deviation S_n (e.g. Hohenegger et al., 2006; Nuissier et al., 2016). In the present study S_n is calculated at any gridpoint where the 3-hourly ensemble mean precipitation rate exceeds $1mm(3h)^{-1}$, and is subsequently area averaged. Larger S_n values indicate higher ensemble dispersion, larger forecast uncertainty and lower predictability.

A hierarchy of measures is applied to conduct the weather regime dependent verification of precipitation forecasts during the HyMeX SOP1. Following the gridpoint based spread (STDEV) and root mean square error (RMSE) we present two probabilistic measures Relative Operating Characteristics ROC and reliability diagram (Wilks, 2011; Jolliffe and Stephenson, 2012) complemented with the widely used Fractions Skill Score (FSS, Roberts and Lean, 2008) to account for spatial tolerance.

2.4 The convective adjustment timescale

The convective adjustment timescale τ_c constitutes an objective measure to classify weather situations by taking the ratio of convective instability (measured by CAPE) and its removal (expressed by the precipitation rate; Done et al., 2006; Keil et al., 2014). During synoptically forced weather precipitation balances the production of instability generated by, e.g., large-scale ascent, the atmosphere is in equilibrium and the value of the convective adjustment timescale is small (Zimmer et al., 2011). In contrast, when the synoptic forcing is weak, local processes like solar insolation, or the interaction with orography generating convergence lines are necessary to overcome a barrier (e.g. convective inhibition) and release convection. During this non-equilibrium regime large CAPE values can build up before convection is triggered and the convective adjustment timescale amounts to larger values comparable to the synoptic timescale.

The area averaged τ_c value can be used to categorically determine the weather regime of the day (e.g. Keil et al., 2019). Firstly, Gaussian smoothed forecast fields of 3-hourly precipitation rates and most unstable CAPE are taken to calculate τ_c at any gridpoint exceeding $3mm(3h)^{-1}$ in individual members given that a minimum areal coverage of this precipitation rate

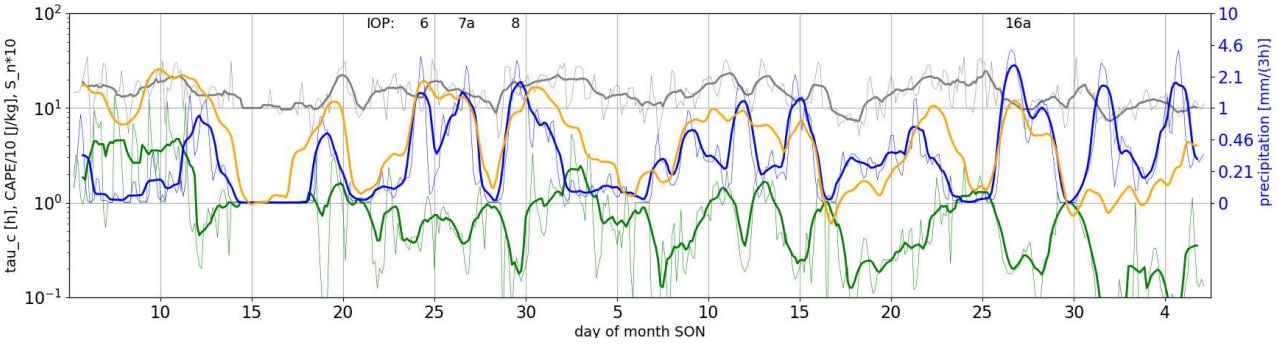

Figure 2. Time series of forecasted area averaged ensemble mean precipitation (blue), convective adjustment timescale (τ_c, green), CAPE (orange), and normalized spread of precipitation (S_n, gray) in the Northwestern Mediterranean Coastal domain for the entire SOP1 period. Thick lines represent temporally smoothed data using a 24-hourly moving average, whereas thin lines indicate the raw 3-hourly data. Note that the y-coordinate has logarithmic scale, CAPE is divided by a factor of 10, S_n multiplied by a factor of 10 and precipitation is labeled on the right hand side to increase readability. ~~The data is plotted using a 24-hourly moving average on the 3-hourly values to increase readability. The thin lines represent the 3-hourly data of precipitation (blue), the convective adjustment timescale (green) and the normalized spread of precipitation (S_n, gray) for reference. Note that the dependent variables are given in logarithmic scale, CAPE is divided by a factor of 10, S_n multiplied by a factor of 10 and precipitation is labeled on the right hand side.~~

is reached. The half-width size of the Gaussian kernel amounts to 20 gridpoints, and a threshold value for precipitation needs to be used, since dry gridpoints preclude any τ_c computation. Both values are set to be consistent with previous work (e.g. Keil and Craig, 2011; Kühnlein et al., 2014; Keil et al., 2014, 2019). Secondly, the ensemble mean of individual τ_c values at any gridpoint is computed, and thirdly, the domain average of this ensemble mean is taken. Finally, if the maximum domain averaged ensemble mean τ_c exceeds a threshold criterion at least once a day, that day is classified to be weakly forced.

In the present study a threshold value of 3 h is chosen to account for the smaller τ_c values occurring in the autumn season of HyMeX SOP1 (see Fig. 3). This is in agreement with the application of τ_c in a maritime environment characterizing convective regimes over the British Isles (Flack et al., 2016). On the other hand a threshold value of 6 h was used to separate mid-latitude precipitation regimes due to dynamic control in a more continental environment in summer (Kühnlein et al., 2014; Keil et al., 2019). ~~This is in contrast to applications of τ_c in the summer season when a threshold of 6 h is reasonable to separate mid-latitude precipitation regimes due to dynamic control. However,~~ Zimmer et al. (2011) argue that the τ_c diagnostic results in a continuous distribution and conclude that a value somewhere between 3 and 12 h clearly distinguishes between different regimes. It turns out that a threshold of 3 h substantially reduces the sampling error giving a distribution of 48 strongly vs 11 weakly forced days during HyMeX SOP1.

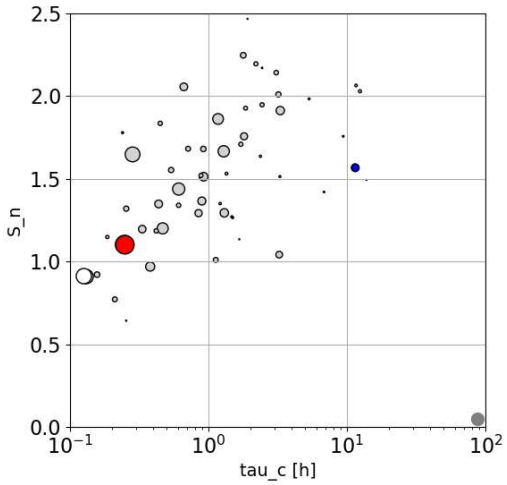

Figure 3. Scatterplot of domain averaged, daily maximum convective adjustment timescale and daily mean normalized standard deviation of precipitation for the entire SOP1. The prominent cases discussed in Section 4 are highlighted: the red circle represents the HyMeX IOP16a, the blue circle the HyMeX 11 September 2012, and the white circle the NAWDEX case. The size of the symbols indicates the daily precipitation accumulation, with the gray circle in the bottom right corner displaying a daily domain integrated rainfall accumulation of 10 mm for comparison.

3 Classification based on the strength of synoptic control

At first glance the timeseries spanning the entire 2-months period shows the variability of weather on daily timescales in autumn 2012 (Fig. 2). The precipitation curve highlights some of the 'golden cases' observed during HyMeX SOP1 (e.g. IOP6 on 24 Sep, IOP7a on 26 Sep, IOP8 on 29 Sep, IOP16a on 26 Oct) with peaks exceeding $2mm(3h)^{-1}$ in domain integrated precipitation. The timeseries of convective instability (CAPE) exhibits large variations, too. High values of spatially averaged CAPE exceeding 100 J/kg mostly concur with the occurence of strong precipitation events (e.g. IOPs 6, 7a, 8 and 16a) pointing towards their predominantly convective character, whereas sometimes maxima do not coincide (e.g. beginning of September). During these episodes convective instability is created by, for instance, solar insolation but cannot be removed by precipitation because of inhibiting factors like capping inversions atop the boundary layer prevent convection initiation. The rank correlation of CAPE and 3-hourly precipitation (and its normalized standard deviation S_n) amounts to 0.44 (and 0.28, respectively) and confirms the limited predictive power of CAPE alone.

Here the convective adjustment timescale τ_c provides a better suitable measure than CAPE to distinguish and to classify weather situations with different synoptic control. Using a categorical threshold of 3 h for the daily maximum area averaged convective adjustment timescale results in 48 strongly and 11 weakly forced days in the Northwestern Mediterranean domain during HyMeX SOP1. Many of the weakly forced cases occur in the first week of the SOP1 (8 to 11 Sept, Fig. 2). After mid-October there are no weakly forced cases anymore suggesting the influence of the seasonal cycle, as decreased solar insolation

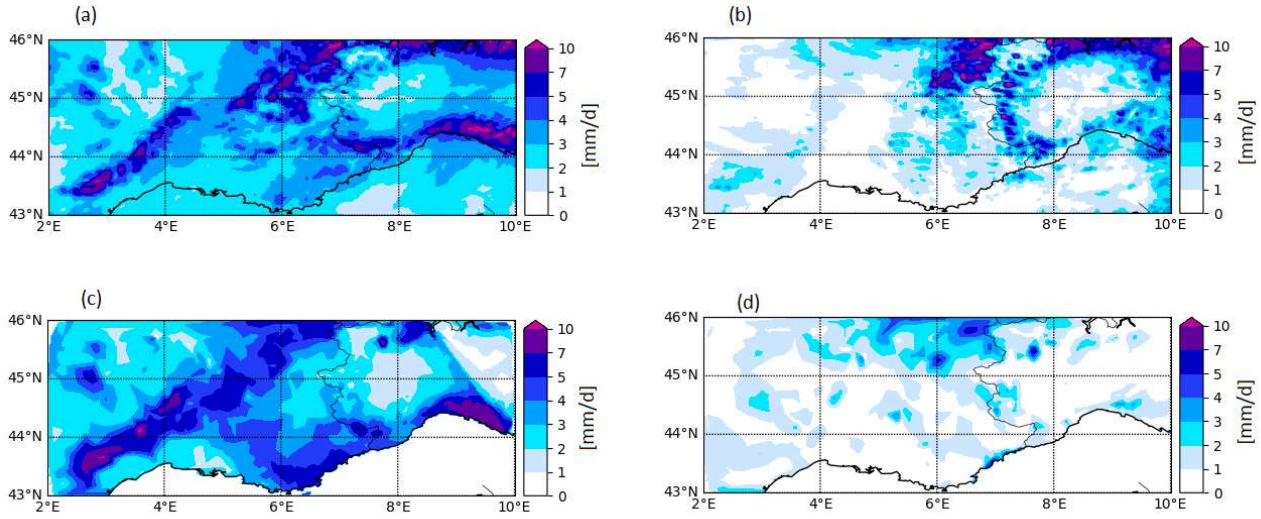

Figure 4. Aggregated mean daily precipitation divided into strong (a,c) and weak (b,d) forcing conditions. Displayed are ensemble mean (a,b) and interpolated rain-gauge observations (c,d) of 24 h precipitation for the 2-months period in autumn 2012.

limits diurnally-driven precipitation. However it is the interplay between the creation of convective instability and its removal by precipitation (both variables make τ_c) that shows the overall decrease in autumn that is strongly modulated by the occurrence of mid-latitude weather systems. During SOP1 τ_c exceeds the threshold value ultimately on 13 October, while area averaged CAPE maxima exceeding 100 J/kg still occur in late October (e.g. on 26 October, IOP16a). A comparison of the timeseries of τ_c and of the normalized standard deviation S_n in Fig. 2 gives a first indication of a connection between both, that is between the weather regime and the forecast uncertainty. Large values of τ_c indicating weakly forced weather conditions correspond with above average values of S_n suggesting below average precipitation predictability.

This relationship and clear dependence of the convective adjustment timescale τ_c and the normalized standard deviation S_n of precipitation is further illustrated in Fig. 3. Large values of τ_c correspond with large S_n of precipitation being a sign of below average predictability. The strongest precipitation events (e.g. IOP16a and the NAWDEX case) occur predominantly at low τ_c values when the normalized ensemble spread of precipitation (S_n) is small, too. Thus, in a domain integrated sense, synoptically forced situations cause higher precipitation accumulations with lower forecast uncertainty. The rank correlation between τ_c and S_n is 0.6 providing statistical evidence that τ_c can be reasonably used as a predictor to classify weather regimes with inherently different precipitation predictability. Moreover, the scatterplot shows that the majority of τ_c values amounts to less than 3 h and a comparison with Fig. 4 in Keil et al. (2019) confirms that the chosen threshold value represents a sensible classification criterion in this specific region at that time of the year.

The different dynamical control shows its fingerprint in the mean spatial distribution of daily rainfall, too (Fig. 4). Apparently, there is more precipitation during strongly forced weather situations than during weakly forced ones. The regions receiving

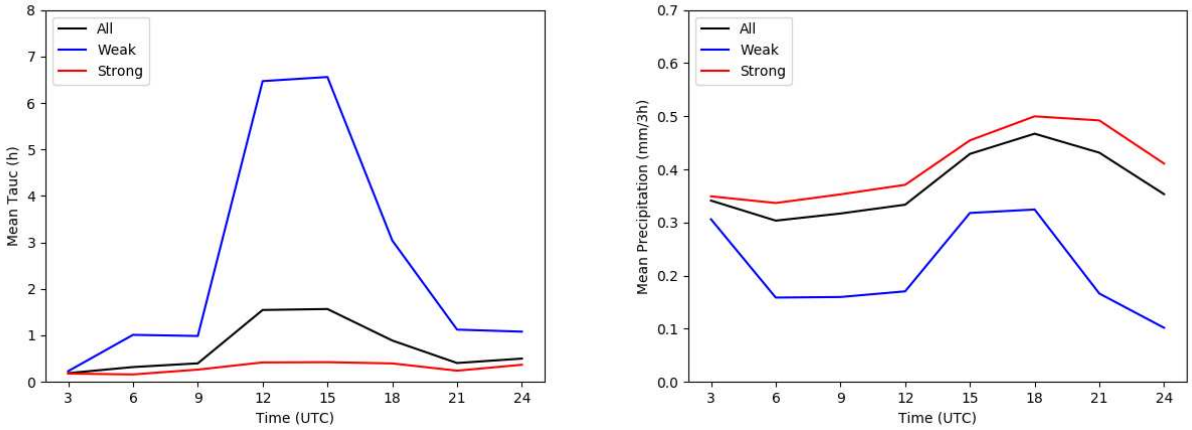

Figure 5. Aggregated diurnal cycle of the ensemble mean convective adjustment time-scale (a) and 3-hourly precipitation (b) averaged over the full SOP1 period (black) and over weakly (blue) and strongly forced weather regimes (red) across the Northwestern Mediterranean domain.

more than 5mm/24h during synoptic control at the southeastern foothills of the Massif Central, the western slopes of the Alpes Maritimes and the Mediterranean coast East of Marseille agree with observations (Fig. 4a,c). In contrast, a spottier distribution of daily rainfall concurrent with an overestimation of precipitation totals becomes evident during weak synoptic control (Fig. 4b,d, and later in Fig. 8).

The aggregated diurnal evolution during weak synoptic control shows the characteristic behaviour with a pronounced diurnal cycle of τ_c peaking in the early afternoon shortly before maximum precipitation rates occur in the late afternoon (Fig. 5). Conversely, during strong forcing conditions there is almost no diurnal pattern in τ_c, precipitation rates are higher and show a weaker amplitude with maxima in the early evening. The unconditional average of τ_c and precipitation is fairly similar to the strongly forced weather type because these flow conditions are prevalent during the HyMeX period thus dominating the diurnal evolution.

4 Three prominent and representative cases

In this section three characteristic cases are presented to highlight the different nature of individual events in detail and to identify hypotheses to be proven in the subsequent systematic evaluation using a hierarchy of measures. The prominent cases comprise HyMeX IOP16a (26 Oct 2012), a typical weakly forced situation during HyMeX (11 Sept 2012) and the NAWDEX Sanchez case (13 Oct 2016). The daily timeseries of precipitation, its normalized standard deviation S_n, ensemble mean CAPE and τ_c clearly depict their different character (Fig. 6). On IOP16a and the NAWDEX case τ_c is always considerably smaller than the threshold criterion (and even less than 1 h, Fig. 6a,c) indicating strong synoptic control (as for IOPs 6, 7a;

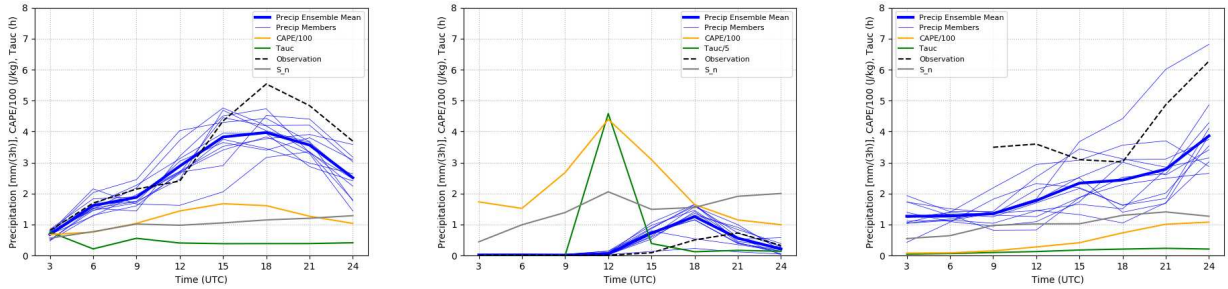

Figure 6. Time series of ensemble mean area averaged precipitation, its normalized standard deviation S_n, ensemble mean CAPE and the convective adjustment timescale τ_c for the prominent cases (a) HyMeX IOP16a, (b) HyMeX 11 Sep 2012 and (c) NAWDEX 13 Oct 2016. Precipitation timeseries of the individual ensemble members highlights intra-ensemble variability. Additionally area averaged rain-gauge observations are shown. Note the different scaling of τ_c on 11 Sept 2012.

not shown) in agreement with Hally et al. (2014) and Nuissier et al. (2016), whereas the temporal evolution of precipitation, CAPE and τ_c on 11 Sept 2012 shows the characteristic behaviour of weakly forced weather situations (Fig. 6b), that is high τ_c values preceeding the strongest rainfall (compare to Keil et al. (2014)). On this weakly forced day the normalized spread S_n is evidently higher and precipitation amounts are overestimated.

4.1 Strongly forced case on 26 October 2012 (IOP16a)

HyMeX IOP16a represents a case of deep convection that developed over the western Mediterranean Sea and affected the coastal regions of France and Italy. The synoptic situation was characterized by a deep upper-level low centered over the Iberian Peninsula moving slowly eastward and fueling slowly propagating Mesoscale Convective Systems in the Northwestern Mediterranean. IOP16a represents a "golden case" enabling us to address the predictability of a high impact weather event (Ducrocq et al., 2014; Nuissier et al., 2016).

Fig. 7 depicts the spatial distributions of 6 h ensemble mean precipitation, its intra-ensemble variability and four individual ensemble members (all 12 to 18 UTC). Three hotspots of precipitation are forecast on 26 October. The spatially largest is located across the Massif Central, where the ensemble mean exceeds 10mm/6h, the intra-ensemble variability (represented by S_n) is small and all members indicate widespread precipitation. At the southeastern foothills of the Massif Central ~~in the Cevennes region~~ the ensemble overestimates the 6 h precipitation (exceeding 20mm/6h) and there is a considerable ensemble spread. There, forecasts of single members diverge (Fig. 7c-f) and show a displacement of heaviest precipitation (e.g. shifted eastward in member 7, very intense and southward in member 8). In the ~~eastern~~ Rhone valley, an area where the ensemble mean indicates more than 5 mm, the intra-ensemble variability is large and individual members fail to predict any precipitation (e.g. member 12).

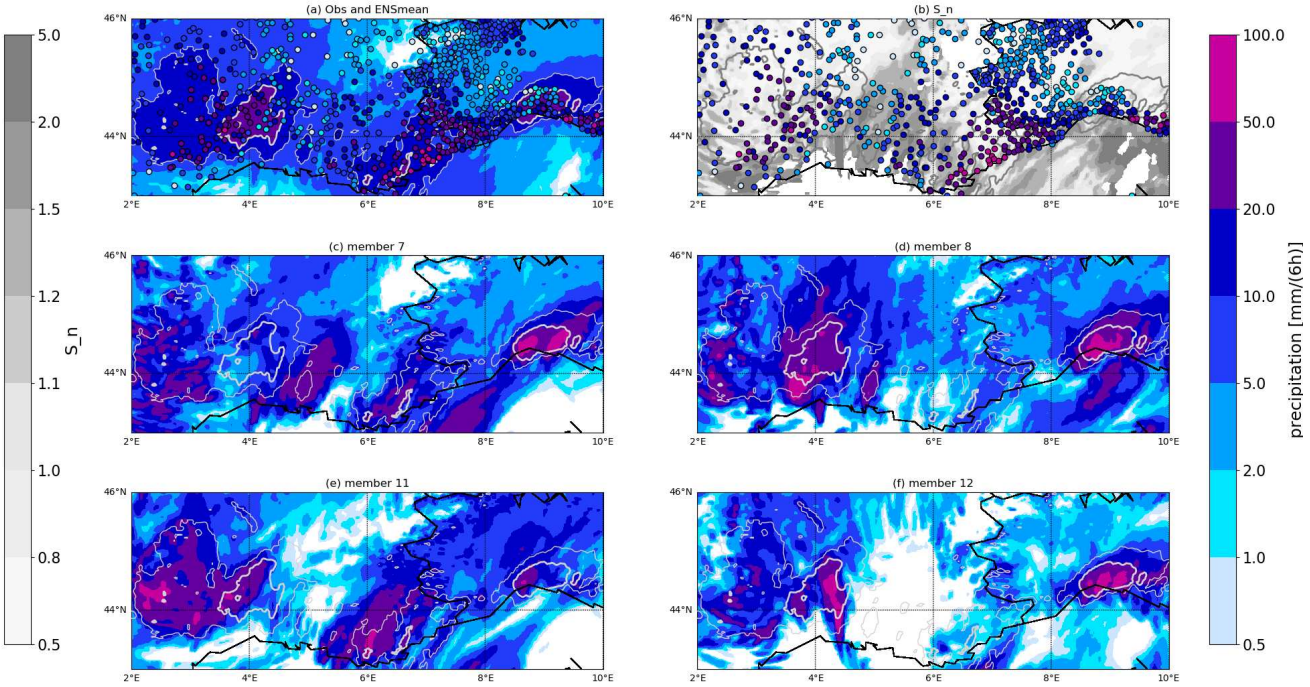

Figure 7. Illustration of 6-hourly precipitation for IOP16a valid 26 October 18 UTC: (a) ensemble mean precipitation forecast (initialized 26 October 00 UTC) and observations (filled circles), (b) normalized ensemble spread (gray) and rainfall observations (filled circles) for reference. Panels (c, d, e and f) show 6-hourly precipitation of selected individual ensemble members, all overlaid with 10 (thin) and 20 mm (thick line) ensemble mean precipitation contours.

A second hotspot of strong precipitation occurs in the Var region where maximum rainfall accumulations are observed (larger than 50mm/6h). There, larger values of S_n point towards a higher intra-ensemble variability that becomes apparent when looking at the rainfall sums of individual members (e.g. Fig. 7e,f). A third heavy precipitation region is forecast close to Genova in Italy. Observations indicate a considerable overprediction in this region with filled circles depicting rain-gauge observations being clearly recognizable (Fig. 7a). However, the hidden circles across large regions of the Massif Central and the Alpes Maritimes point towards the overall good performance of the ensemble mean forecast of precipitation. Interestingly precipitation amounts under strong synoptic control are strongly modulated by orography with highest intra-ensemble variability in the flat regions of the Rhone valley and south of the Massif Central.

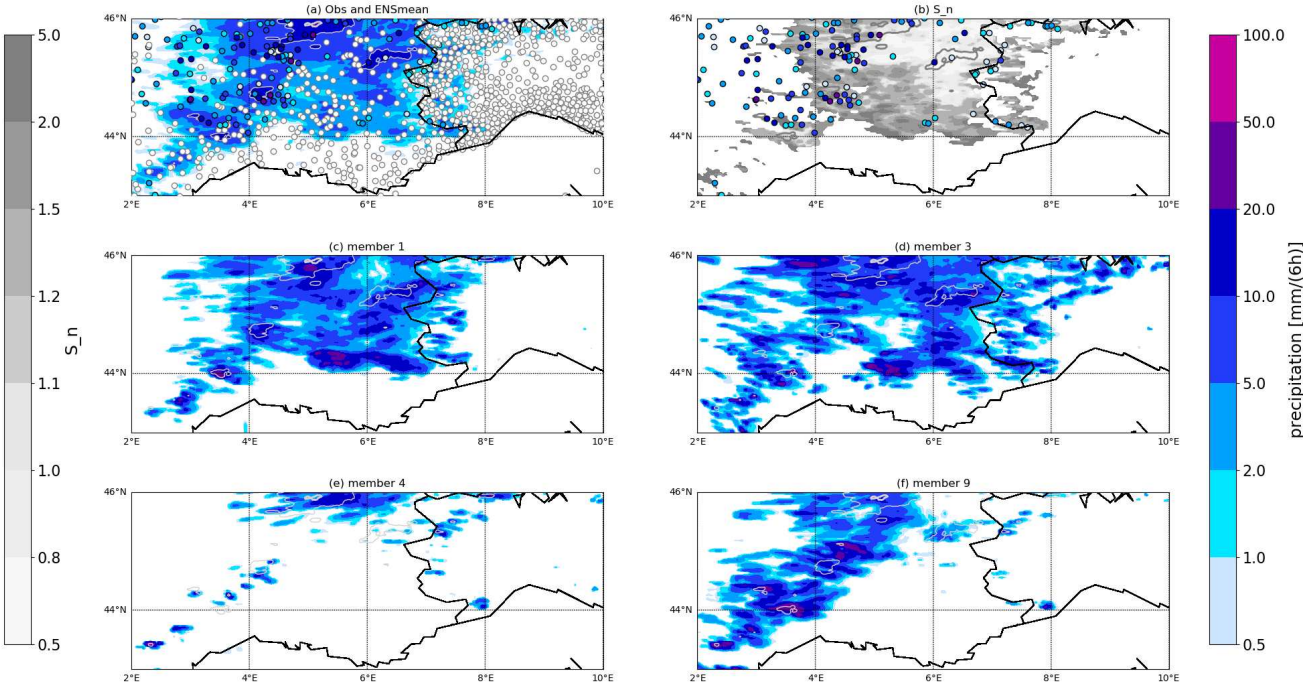

Figure 8. Same as Fig. 7, but for 11 September 18 UTC (initialized 11 September 00 UTC).

4.2 Weakly forced case on 11 September 2012

Weather on 11 September represents a characteristic case of a weakly forced situation during HyMeX. Before noon single convective cells are triggered with very different intensity and location in the individual members (not shown) resulting in very small area averaged rainfall accumulations (Fig. 6). Subsequently, convection intensifies leading to more than 50mm/6h in individual members at different locations (Fig. 8c,d,f). The differences in terms of exact location of heaviest precipitation result in maximum ensemble mean values of less than 20 mm (Fig. 8a). Overall, precipitation is strongest across mountainous regions (a more or less distinct precipitation band extends from southwest to northeast across the Massif Central in all members) with correctly forecast dry conditions south of 44°N in the Var region. Whereas the accumulated precipitation distribution in members 1 and 3 resembles the ensemble mean pattern, other members exhibit big deviations: member 4 forecasts hardly any rainfall, while member 9 forecasts a lot of precipitation in the western part of the domain only. Large S_n values west of 6°E demonstrate this considerable intra-ensemble variability. Overall, the comparison with rainfall observations suggests a notable overestimation (Fig. 8a) and a clear connection to orography during weak control.

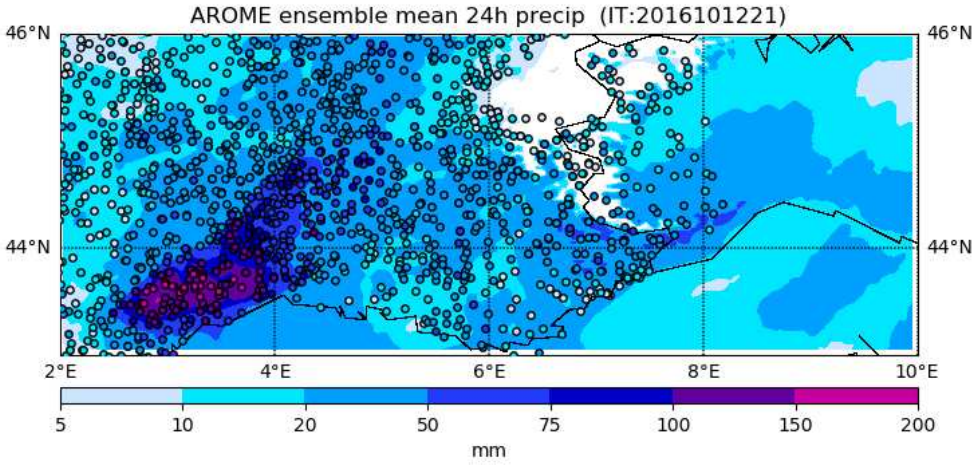

Figure 9. Observed (circles) and forecast (initialized 12 October 21 UTC) ensemble mean accumulated daily precipitation until 14 October 2016 6 UTC.

The weather situation on 11 September is characteristic for the first week of the HyMeX SOP1 period (see Fig. 2) when solar insolation in early autumn is still strong enough to generate convective instability by surface heating resulting in large CAPE and large τ_c values indicating a need for local triggering mechanisms to overcome convective inhibition and to form precipitation (see Fig. 6).

4.3 Heavy precipitation on 13 October 2016 during NAWDEX

The detailed examination of individual heavy precipitation events in the Western Mediterranean region is complemented with one of the most prominent cases in that region that developed downstream of the cyclone Sanchez during the NAWDEX field campaign in autumn 2016 (Schäfler et al., 2018). Fig. 9 shows a good match of forecast and observed 24 h precipitation peaking in the southern foothills of the Massif Central ~~Cevennes region~~ with more than 200 mm. This event is clearly classified as strongly forced regime with an area averaged maximum τ_c of less than 30 min (see Fig. 3 and Fig. 6).

However, and unexpectedly for a case under strong control, individual ensemble members show surprisingly large spatial variability that becomes evident when inspecting the time window of heavy rainfall in the afternoon (between 12 and 18 UTC, Fig. 10). Focussing on the region of heaviest precipitation (20mm/6h in ensemble mean) at the southern foothills of the Massif Central all members exhibit strong precipitation rates individually, whereas in other areas there are large discrepancies (e.g. members 5 and 6, Fig. 10d,e). Within the heaviest precipitation region all members agree well resulting in small S_n values. Above average precipitation variability occurs northward (across the Massif Central) and across the Mediterranean Sea.

In summary, the three selected cases indicate that the heaviest precipitation is co-located with orography during both regimes, that the spatial predictability of precipitation can considerably vary from case-to-case even within one forcing type, and that the precipitation intensity is overestimated during weak control.

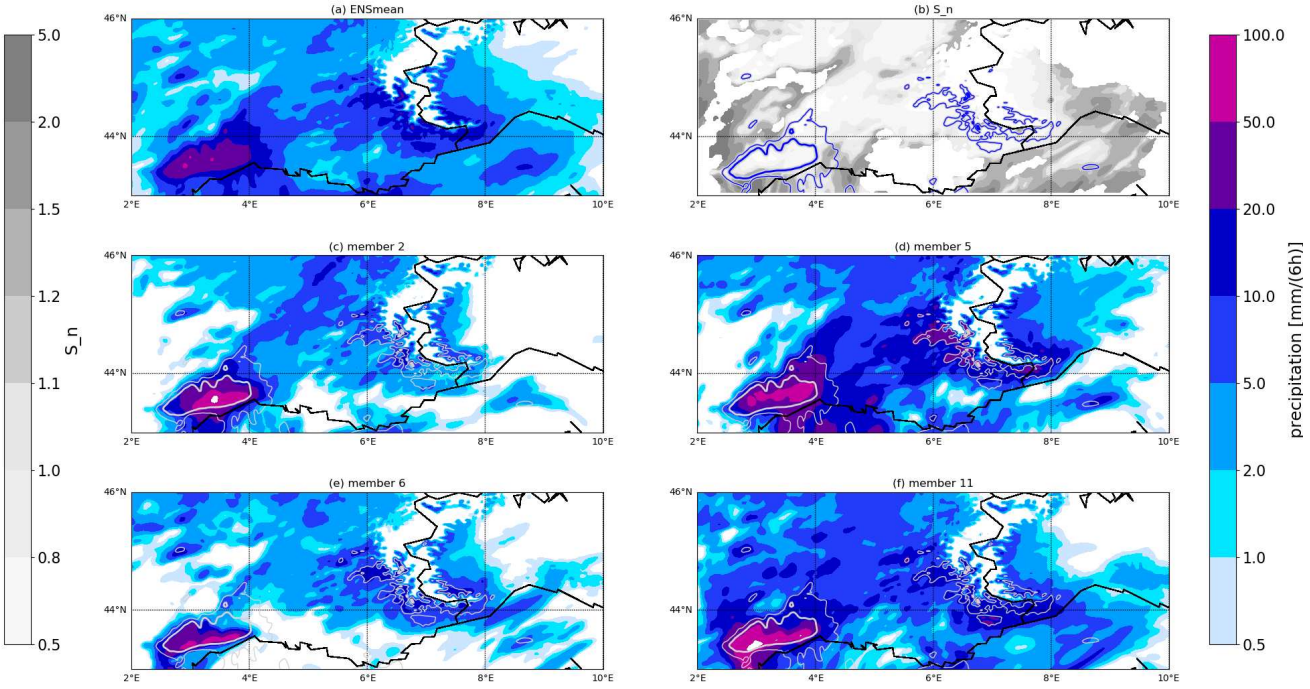

Figure 10. Same as Fig. 7, forecast for 13 October 2016 18 UTC (initialized 12 October 21 UTC).

5 Systematic verification conditional to the strength of synoptic control

Finally, we examine the question how precipitation forecasts during the different weather regimes compare with observations using a hierarchy of measures applied on the full 2-months period. Firstly, we show the mean diurnal evolution of the grid-point based root-mean-square error (RMSE) of ensemble mean 3-hourly precipitation forecasts and rain-gauge observations conditionally averaged on both weather regimes.

During strong control the RMSE exhibits less diurnal variations than during weak control when a typical diurnal cycle is recognizable attaining the highest error during the convective most active period in the afternoon between 12 and 18 UTC (Fig. 11). The magnitude of the error reaches values up to $1.2mm(3h)^{-1}$ in the Northwestern Mediterranean, which is roughly 50% less than found by Bouttier et al. (2016) looking at large parts of Western Europe. Given that rainfall rates during weak forcing amount to only about 60% of the rates during strong forcing (Fig. 5b), the relative error is higher in the weak regime. Likewise, the ensemble spread shows a diurnal cycle and is highest during the convective most active period in the afternoon under weak synoptic control. Since 80% of the days during HyMeX SOP1 are classified as strongly forced weather regime it is not surprising that the regime independent curve follows the strongly forced curve closely thus obscuring the forecast model characteristic during weak control.

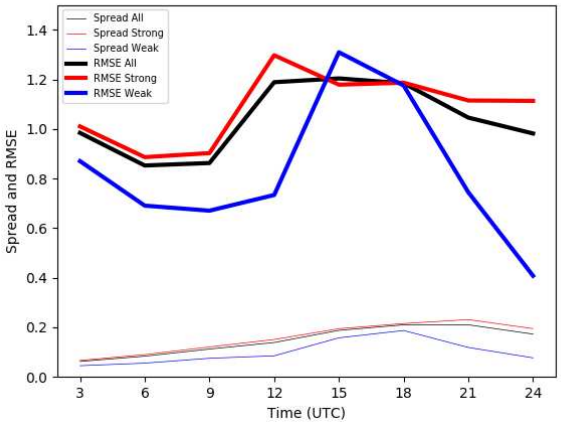

Figure 11. Aggregated time series of the spread (standard deviation) and skill (RMSE of the ensemble mean) averaged over the full SOP1 period (black) and over weakly (blue) and strongly forced weather regimes (red). ~~The RMSE is the ensemble mean of the member RMS forecast error.~~

Secondly, the regime dependent probabilistic performance of the ensemble is investigated using the ROC and reliability diagrams for 3-hourly (Fig. 12a,b) and daily accumulations (Fig. 12c,d). Both probabilistic scores highlight the superior performance during strongly forced weather regimes. ~~A ROC curve closer to the left and upper boundaries displays a greater event discrimination in this weather situation.~~ The larger distance of the ROC curve points from the diagonal (resulting in larger concavity) during strong control indicates ~~the higher *absolute* spread~~ greater event discrimination when 3-hourly (and daily) precipitation accumulations are averaged over the entire SOP1. In this weather regime the AROME-EPS forecasts are generally more reliable, in particular when averaged over 24 hours (Fig. 12d). The calibration functions in the reliability diagrams show that the forecast probabilities are consistently too large relative to the conditional observed relative frequencies. This is an indication of overforecasting equivalent to ~~a wet bias~~ an overconfidence of the ensemble, that ~~. The general wet bias~~ is strongest during weak synoptic control for short (3-hourly) time windows (Fig. 12b). Moreover, the flatness of the calibration function for this weather regime reveals a poor resolution. ~~and an overconfidence~~. Observed relative frequencies depend only slightly on the forecast probabilities and always amount to less than 20% for all forecast probabilities of moderate $3mm(3h)^{-1}$ precipitation rates. Relaxing the temporal exactness and extending the window to daily accumulations improves the reliability, in particular during strong control (Fig. 12d).

Finally, the Fractions Skill Score (Roberts and Lean, 2008; Faggian et al., 2014) is employed to address the double penalty problem inherent in convective scale precipitation forecasts. To compute the FSS we threshold each member first, then take the ensemble mean of the binary probabilities and then apply the FSS. In Fig. 13 the ensemble mean FSS is shown as a function of neighborhood size for absolute rainfall rates ($0.3mm(3h)^{-1}$, a threshold frequently used to separate rain versus no-rain areas, and $10mm(24h)^{-1}$ accumulation) splitted into weather regimes aggregated for the 2-months period. During strong forcing

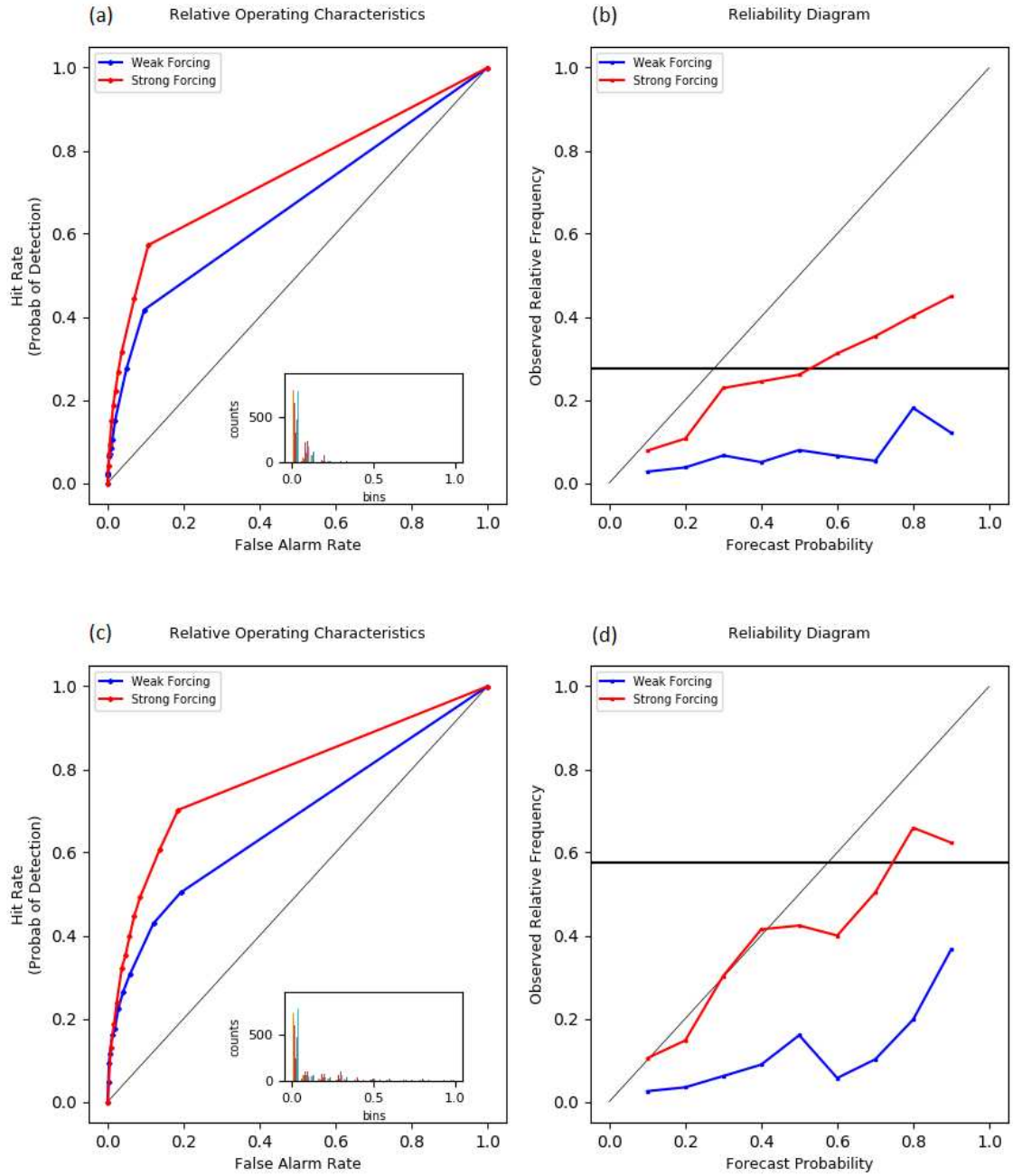

Figure 12. Relative Operating Characteristics (ROC) (a,c) and reliability diagram (b,d) for $3mm(3h)^{-1}$ precipitation rates (a,b) and daily amounts of $10mm(24h)^{-1}$ (c,d) conditional to weather regime aggregated over the entire HyMeX SOP1.

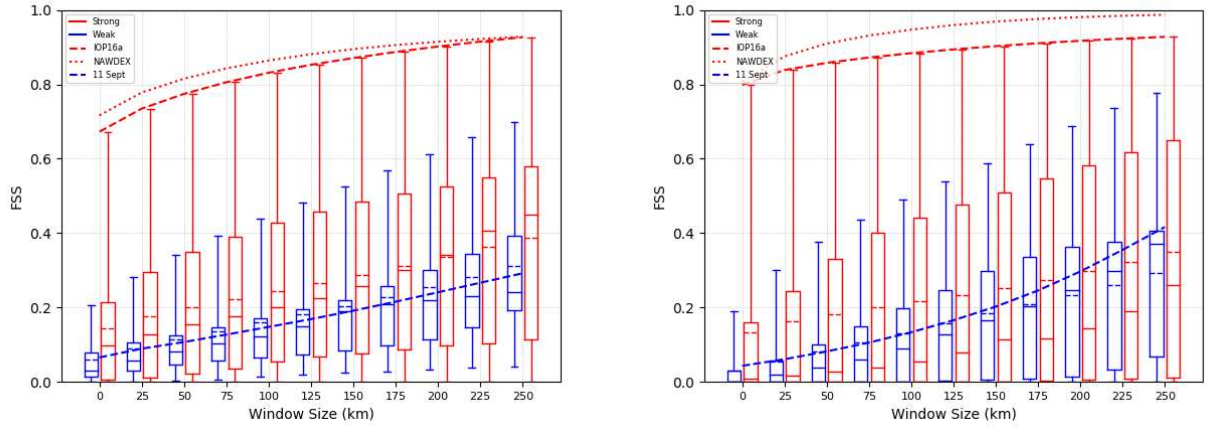

Figure 13. Fraction Skill Score FSS of ensemble mean 3-hourly precipitation vs interpolated rain-gauge observations as a function of neighborhood size conditional to weather regime: (a) $0.3mm(3h)^{-1}$ and (b) $10mm(24h)^{-1}$, all averaged for the entire HyMeX SOP1. The horizontal line (dashed line) within the boxes indicates the median (mean), respectively. The boxes and whiskers are slightly displaced at the discrete window sizes (weak to the left, strong to the right) to increase readability. Additionally, the FSS of the prominent cases is depicted.

the spatial forecast quality of the low rainfall threshold is superior for all neighborhood sizes (Fig. 13a). The skill increases when relaxing the grid point proximity and comparing larger neighborhoods, as expected. The fairly large box sizes and the extension of the whiskers demonstrate the large variability of forecast quality during the 2-months period. The upper quartile of the boxplot is touching upon a FSS value of 0.5 at neighborhood sizes of ~~125 km~~ 150 km during strong forcing. A FSS value of roughly 0.5 is also known as the believable scale (FSS $= 0.5 + f_0/2$, where f_0 is the observed precipitation coverage, see Dey et al., 2014), a scale at which forecasts are deemed reasonably skillful and useful (Roberts and Lean, 2008). Thus, 25% of the time (3-hourly intervals on 48 strongly forced days, i.e. for 96 time windows within SOP1) the forecasts are skillful at a scale of $\mathcal{O}(100\,\mathrm{km})$, which is of the same order as found in previous studies (Clark et al., 2010; Mittermaier et al., 2011; Schwartz and Sobash, 2019; Bachmann et al., 2020), based on FSS and other neighborhood methods. Useful precipitation forecasts are hardly encountered during weak forcing using absolute rainfall rates.

Relaxing the temporal exactness of 3-hourly accumulations towards daily sums confirms previous results. Using a fixed precipitation amount of 10 mm per day reveals that the mean FSS during strong control always lies higher than during weak control, that is on average the spatial forecast accuracy is higher during strongly forced weather situations, as expected (Fig. 13b). The discrepancy between the mean and the median of FSSs ~~during strong forcing~~ suggests that the high threshold of $10mm(24h)^{-1}$ represents rare events with different intermittency characteristics in forecast and observation leading to a skewed distribution.

The inspection of the spatial forecast accuracy of the prominent cases again highlights the large day-to-day variability. Both strongly forced prominent events (IOP16a and NAWDEX) exhibit a very good spatial forecast quality with the FSS reaching

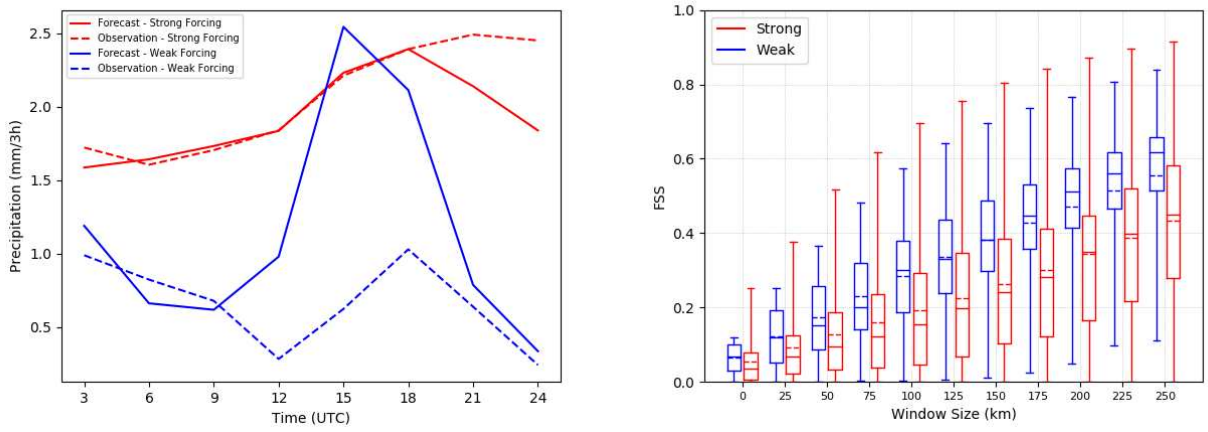

Figure 14. (a) Aggregated diurnal cycle of 3-hourly precipitation values corresponding to the 95th percentiles, both stratified in weakly (blue) and strongly forced weather regimes (red) and averaged over the full SOP1 period. (b) FSS of 95th percentiles of 3-hourly precipitation as a function of neighborhood size conditional to weather regime averaged for the entire HyMeX SOP1 between 12 and 18 UTC.

values larger 0.8 for window sizes larger ~~25 km~~ 50 km tantamount with the highest whiskers (Fig. 13). The NAWDEX case (occurring in 2016) even shows FSS values higher than the highest whiskers found during HyMeX. The excellent forecast performance is mainly caused by the low precipitation threshold ($0.3mm(3h)^{-1}$) and the widespread precipitation occurring on both days. Large parts of the domain receive such precipitation rates and the FSS attains high values. The prominent weakly forced case indicates an average forecast performance (FSS of 11 Sept matches the mean value of this regime) for low rainfall rates separating essentially rain and no-rain regions.

However, taking into account a varying model bias during different weather regimes changes the picture. The *pure* forecast location accuracy neglecting a model bias can be estimated by using percentiles of forecast and observed precipitation amounts. Whereas the amounts corresponding to the 95th percentiles of forecast and observed precipitation agree well during strong forcing (at least until 18 UTC), there is a considerable overprediction during weak forcing (Fig. 14a). This overforecasting is strongest during the convective most active period in the afternoon between 12 and 18 UTC. Taking this bias into account by using precipitation percentiles results in a superior spatial forecast quality during weakly forced regimes (Fig. 14b). Thus forecasting the *pure* location of heaviest precipitation in the afternoon (expressed by the 95th percentiles) is better during comparably quiescent synoptic-scale atmospheric conditions. This is at first sight an unexpected and surprising result. Given the favourable meteorological ingredients for generating deep convection ~~heavy precipitation~~ at this specific geographical region in the autumn season (Grazzini et al., 2020), we hypothesize that well represented steady land surface structures (like orography, particularly) in kilometric scale models provide sufficient trigger mechanisms to initiate convection and serve as a permanent source of precipitation predictability during weak control. The structuring effect of mountains on the location of precipitation has previously been shown in idealized and real-world ensemble simulations of summertime convection in

Central Europe (Bachmann et al., 2019, 2020). In contrast, forecasting the *pure* location of heaviest precipitation with high temporal exactness at forecast horizons of 12 to 24 hours is challenging during transient synoptic-scale weather systems typical during strong control.

6 Conclusions

This study extends prior work documenting the performance of AROME-EPS during HyMeX SOP1 (Bouttier et al., 2016; Nuissier et al., 2016) by the weather regime dependent aspect of precipitation predictability with a special focus on the spatial forecast quality. The convective adjustment timescale τ_c is used to categorically classify every single day within the 2-months period in autumn 2012 into one specific weather type depending on the strength of the synoptic control. From a physical perspective, it is sensible to use variations in forcing (i.e., τ_c), rather than CAPE, as being associated with variations of precipitation characteristics and forecast skill (Schwartz and Sobash, 2019).

Altogether, the ever changing meteorological situations in the Northwestern Mediterranean Coastal region are stratified into 48 strongly and 11 weakly forced days during HyMeX SOP1. All weakly forced, that is locally triggered precipitation events occur before mid-October with a sequence of weakly forced days in the beginning of September. This distribution follows the seasonal cycle and reflects the climatological study of heavy precipitation events in northern Italy showing that weakly forced events occur from mid-May to the end of October with the highest frequency from mid-August to mid-September (Grazzini et al., 2020). Key HyMeX IOPs are classified as strongly forced weather types in agreement with literature (Hally et al., 2014; Ducrocq et al., 2014; Nuissier et al., 2016). Likewise, the prominent heavy precipitation event that occurred during NAWDEX is clearly identified as strongly forced (Schäfler et al., 2018).

A clear connection between the weather regime and (i) the mean diurnal evolution of precipitation, (ii) the mean spatial distribution of daily rainfall, (iii) the precipitation predictability, (iv) the precipitation bias, (v) the probabilistic and (vi) spatial forecast quality is found. During strong synoptic control, which is dominating the weather on 80% of the days during HyMeX SOP1, the domain integrated precipitation predictability assessed with the normalized ensemble standard deviation S_n is above average, the wet bias is smaller and the forecast quality is generally better. Conversely, there is a pronounced diurnal cycle of area averaged precipitation and a considerable intra-ensemble variability in terms of placement of precipitation (i.e. large S_n) during weakly forced weather types consistent with previous results (e.g. Keil et al., 2019; Schwartz and Sobash, 2019; Bachmann et al., 2020). Disregarding the wet bias during weak control by focussing on 95th percentiles of precipitation shows a ~~the unexpected result of~~ superior pure spatial predictability of most intense precipitation in the afternoon during weak control. We hypothesize that a reasonable representation of steady land surface structures (e.g. orography, coast line) in kilometric scale numerical models provide trigger mechanisms to initiate convection during weak control and serve as a source of location predictability for precipitation, given favourable atmospheric conditions in this special geographical region. The important role of orography on precipitation in this region at this season is in agreement with the climatological study of Grazzini et al. (2020), who found that convective precipitation is largely influenced by orography during the frontal uplift with embedded equilibrium deep convection (herein: strong control) as well as non-equilibrium convection (herein: weak control). One reason

for the apparent overprediction of precipitation during weak control can partly be accounted for by the ~~point type~~ pointwise character of rain-gauge measurements that sample the spatial highly heterogeneous and intermittent nature of locally triggered convective precipitation insufficiently. This discrepancy calls for remotely sensed spatial rainfall measurements of high quality, that were not available in the present study.

It is shown that the unconditional evaluation of precipitation widely parallels the strongly forced weather type evaluation and might obscure forecast model characteristics typical for weak control. Such a separation of statistics according to local weather conditions might proof useful to improve physical parameterisations that depend on the weather condition, as, for instance, Bouttier et al. (2012) suggested for the correlation lengths in the stochastic SPPT scheme, to identify a regime dependent impact of certain surface perturbations (Baur et al., 2018) or to enhance nowcasting capabilities (Kober et al., 2014).

Acknowledgements. The authors acknowledge access to the HyMeX database from which rainfall observations were retrieved. LC was supported within the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather" funded by the German Research Foundation (DFG). She was also supported by the Hans-Ertel-Centre for Weather Research. This German research network of universities, research institutes and Deutscher Wetterdienst is funded by the BMVI (Federal Ministry of Transport and Digital Infrastructure).

References

Anthes, R. A.: The General Question of Predictability, in: Mesoscale Meteorology and Forecasting, edited by Ray, P. S., vol. II, pp. 636–656, American Meteorological Society, Boston, MA, https://doi.org/10.1007/978-1-935704-20-1_27, 1986.

Bachmann, K., Keil, C., and Weissmann, M.: Impact of radar data assimilation and orography on predictability of deep convection, Q. J. Roy. Meteor. Soc., 145, 117–130, https://doi.org/10.1002/qj.3412, 2019.

Bachmann, K., Keil, C., Craig, G. C., Weissmann, M., and Welzbacher, C. A.: Predictability of Deep Convection in Idealized and Operational Forecasts: Effects of Radar Data Assimilation, Orography, and Synoptic Weather Regime, Mon. Weather Rev., 148, 63–81, https://doi.org/10.1175/mwr-d-19-0045.1, 2020.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, 2015.

Baur, F., Keil, C., and Craig, G. C.: Soil Moisture - Precipitation Coupling over Central Europe: Interactions between surface anomalies at different scales and its dynamical implication, Q. J. Roy. Meteorol. Soc., https://doi.org/DOI: 10.1002/qj.3415, 2018.

Bénard, P.: On the use of a wider class of linear systems for the design of constantcoefficients semi-implicit time schemes in NWP., Mon. Weather Rev., 132, 1319–1324, 2004.

Bougeault, P. and Lacarrère, P.: Parameterization of orography-induced turbulence in a meso-beta-scale model, Mon. Weather Rev., 123, 1560–1573, 1989.

Bouttier, F., Vie, B., Nuissier, O., and Raynaud, L.: Impact of stochastic physics in a convection-permitting ensemble, Mon. Weather Rev., 140, 3706–3721, 2012.

Bouttier, F., Raynaud, L., Nuissier, O., and Ménétrier, B.: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX, Q. J. Roy. Meteor. Soc., 142, 390–403, https://doi.org/10.1002/qj.2622, 2016.

Brousseau, P., Seity, Y., Ricard, D., and Léger, J.: Improvement of the forecast of convective activity from the AROME-France system, Q. J. Roy. Meteor. Soc., 142, 2231–2243, https://doi.org/10.1002/qj.2822, 2016.

Bubnova, R., Hello, G., Bénard, P., and Geleyn, J.-F.: Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system., Mon. Weather Rev., 123, 515–535, 1995.

Caniaux, G., Redelsperger, J.-L., and Lafore, J.-P.: A numerical study of the stratiform region of a fast-moving squall line. 1. general description and water and heat budgets., J. Atmos. Sci., 51, 2046–2074, 1994.

Clark, A. J., Gallus, W. A., and Weisman, M. L.: Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM, Weather and Forecasting, 25, 1495–1509, https://doi.org/10.1175/2010waf2222404.1, 2010.

Cuxart, J., Bougeault, P., and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations., Quart. J. Roy. Meteorol. Soc., 126, 1–30, 2000.

Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., and Cébron, P.: PEARP, the Météo-France short-range ensemble prediction system, Quart. J. Roy. Meteorol. Soc., pp. 1671–1685, https://doi.org/10.1002/qj.2469, http://dx.doi.org/10.1002/qj.2469, 2014.

Dey, S. R. A., Leoncini, G., Roberts, N. M., Plant, R. S., and Migliorini, S.: A Spatial View of Ensemble Spread in Convection Permitting Ensembles, Mon. Weather Rev., 142, 4091–4107, https://doi.org/10.1175/mwr-d-14-00172.1, 2014.

Done, J. M., Craig, G. C., Gray, S. L., Clark, P. A., and Gray, M. E. B.: Mesoscale simulations of organized convection: Importance of convective equilibrium, Q.J.R. Meteorol. Soc., 132, 737–756, https://doi.org/10.1256/qj.04.84, 2006.

Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayral, P.-A., Belamari, S., Berne, A., Borga, M., Boudevillain, B., Bock, O., Boichard, J.-L., Bouin, M.-N., Bousquet, O., Bouvier, C., Chiggiato, J., Cimini, D., Corsmeier, U., Coppola, L., Cocquerez, P., Defer, E., Delanoë, J., Girolamo, P. D., Doerenbecher, A., Drobinski, P., Dufournet, Y., Fourrié, N., Gourley, J. J., Labatut, L., Lambert, D., Coz, J. L., Marzano, F. S., Molinié, G., Montani, A., Nord, G., Nuret, M., Ramage, K., Rison, W., Roussot, O., Said, F., Schwarzenboeck, A., Testor, P., Baelen, J. V., Vincendon, B., Aran, M., and Tamayo, J.: HyMeX-SOP1: The Field Campaign Dedicated to Heavy Precipitation and Flash Flooding in the Northwestern Mediterranean, Bull. Am. Meteorol. Soc., 95, 1083–1100, https://doi.org/10.1175/bams-d-12-00244.1, 2014.

Faggian, N., Roux, B., Steinle, P., and Ebert, B.: Fast calculation of the Fraction Skill Score, Mausam, 66, 1–12, 2014.

Flack, D. L. A., Plant, R. S., Gray, S. L., Lean, H. W., Keil, C., and Craig, G. C.: Characterisation of convective regimes over the British Isles, Q. J. Roy. Meteor. Soc., 142, 1541–1553, https://doi.org/10.1002/qj.2758, 2016.

Fourrié, N., Nuret, M., Brousseau, P., Caumont, O., Doerenbecher, A., Wattrelot, E., Moll, P., Bénichou, H., Puech, D., Bock, O., Bosser, P., Chazette, P., Flamant, C., Girolamo, P. D., Richard, E., and Saïd, F.: The AROME-WMED reanalyses of the first special observation period of the Hydrological cycle in the Mediterranean experiment (HyMeX), Geosci. Model Dev., 12, 2657–2678, https://doi.org/10.5194/gmd-12-2657-2019, 2019.

Grazzini, F., Craig, G. C., Keil, C., Antolini, G., and Pavan, V.: Extreme precipitation events over northern Italy. Part I: A systematic classification with machine-learning techniques, Q. J. Roy. Meteor. Soc., 146, 69–85, https://doi.org/10.1002/qj.3635, 2020.

Hally, A., Richard, E., and Ducrocq, V.: An ensemble study of HyMeX IOP6 and IOP7a: sensitivity to physical and initial and boundary condition uncertainties, Nat. Hazards Earth Syst. Sci., 14, 1071–1084, https://doi.org/10.5194/nhess-14-1071-2014, 2014.

Hohenegger, C., Lüthi, D., and Schär, C.: Predictability Mysteries in Cloud-Resolving Models, Mon. Weather Rev., 134, 2095–2107, https://doi.org/10.1175/MWR3176.1, 2006.

Jolliffe, I. T. and Stephenson, D. B.: Probability Forecasts, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science (2nd edn), pp. 126–136, John Wiley & Sons, Ltd, https://doi.org/10.1002/9781119960003.ch1, 2012.

Keil, C. and Craig, G. C.: Regime-dependent forecast uncertainty of convective precipitation, Meteorologische Zeitschrift, 20, 145–151, https://doi.org/10.1127/0941-2948/2011/0219, 2011.

Keil, C., Heinlein, F., and Craig, G. C.: The convective adjustment time-scale as indicator of predictability of convective precipitation, Q. J. Roy. Meteor. Soc., 140, 480–490, 2014.

Keil, C., Baur, F., Bachmann, K., Rasp, S., Schneider, L., and Barthlott, C.: Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes, Q. J. Roy. Meteor. Soc., 145, 3102–3115, https://doi.org/10.1002/qj.3607, 2019.

Kober, K., Craig, G. C., and Keil, C.: Aspects of short-term probabilistic blending in different weather regimes, Q. J. Roy. Meteor. Soc., 140, 1179–1188, https://doi.org/10.1002/qj.2220, 2014.

Kühnlein, C., Keil, C., Craig, G. C., and Gebhardt, C.: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation, Q. J. Roy. Meteor. Soc., 140, 1552–1562, 2014.

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., Leriche, M., Barthe, C., Aouizerats, B., Augros, C., Aumond, P., Auguste, F., Bechtold, P., Berthet, S., Bielli, S., Bosseur, F., Caumont, O., Cohard, J.-M., Colin, J., Couvreux, F., Cuxart, J., Delautier, G., Dauhut, T., Ducrocq, V., Filippi, J.-B., Gazen, D., Geoffroy, O., Gheusi, F., Honnert, R., Lafore, J.-P., Brossier, C. L., Libois, Q., Lunet, T., Mari, C., Maric, T., Mascart, P., Mogé, M., Molinié, G., Nuissier, O., Pantillon, F., Peyrillé, P., Pergaud, J., Perraud, E., Pianezze, J., Redelsperger, J.-L., Ricard, D., Richard, E., Riette, S., Rodier, Q., Schoetter, R., Seyfried, L., Stein, J., Suhre, K., Taufour, M., Thouron,

O., Turner, S., Verrelle, A., Vié, B., Visentin, F., Vionnet, V., and Wautelet, P.: Overview of the Meso-NH model version 5.4 and its applications, Geosci. Model Dev., 11, 1929–1969, https://doi.org/10.5194/gmd-11-1929-2018,2018., 2018.

Mittermaier, M., Roberts, N., and Thompson, S. A.: A long-term assessment of precipitation forecast skill using the Fractions Skill Score, Meteorological Applications, 20, 176–186, https://doi.org/10.1002/met.296, 2011.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T.: A strategy for high resolution ensemble prediction. I: Definition of representative numbers and global-model experiments., Quart. J. Roy. Meteorol. Soc., 127, 2069–2094, 2001.

Nuissier, O., Joly, B., Vié, B., and Ducrocq, V.: Uncertainty of lateral boundary conditions in a convection-permitting ensemble: a strategy of selection for Mediterranean heavy precipitation events, Nat. Hazards Earth Syst. Sci., 12, 2993–3011, 2012.

Nuissier, O., Marsigli, C., Vincendon, B., Hally, A., Bouttier, F., Montani, A., and Paccagnella, T.: Evaluation of two convection-permitting ensemble systems in the HyMeX Special Observation Period (SOP1) framework, Q. J. Roy. Meteor. Soc., 142, 404–418, https://doi.org/10.1002/qj.2859, 2016.

Pergaud, J., Masson, V., and Malardel, S.: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction., Bound.-Layer Meteor., 132, 83–106, 2009.

Raynaud, L. and Bouttier, F.: The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts, Q. J. Roy. Meteor. Soc., 143, 3037–3047, https://doi.org/10.1002/qj.3159, 2017.

Roberts, N. and Lean, H.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Mon. Weather Rev., 136, 78–97, 2008.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C. M., Gray, S. L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C. A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K., and Zinner, T.: The North Atlantic Waveguide and Downstream Impact Experiment, Bull. Am. Meteorol. Soc., 99, 1607–1637, https://doi.org/10.1175/bams-d-17-0003.1, 2018.

Schwartz, C. S. and Sobash, R. A.: Revisiting Sensitivity to Horizontal Grid Spacing in Convection-Allowing Models over the Central and Eastern United States, Mon. Weather Rev., 147, 4411–4435, https://doi.org/10.1175/mwr-d-19-0115.1, 2019.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France Convective-Scale Operational Model, Mon. Weather Rev., 139, 976–991, https://doi.org/10.1175/2010mwr3425.1, 2011.

Soares, P. M. M., A., M. P. M., Siebesma, A. P., and Teixeira, J.: An eddy-diffusivity/mass-flux parameterization for dry and shallow cumulus convection, Q. J. Roy. Meteor. Soc., 130, 3055–3079, 2004.

Surcel, M., Zawadzki, I., Yau, M. K., Xue, M., and Kong, F.: More on the Scale Dependence of the Predictability of Precipitation Patterns: Extension to the 2009–13 CAPS Spring Experiment Ensemble Forecasts, Mon. Weather Rev., 145, 3625–3646, https://doi.org/10.1175/mwr-d-16-0362.1, 2017.

Wernli, H., Hofmann, C., and Zimmer, M.: Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique, Weather and Forecasting, 24, 1472–1484, https://doi.org/10.1175/2009waf2222271.1, 2009.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 2011.

Yano, J.-I., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., and Wyszogrodzki, A. A.: Scientific Challenges of

Convective-Scale Numerical Weather Prediction, Bull. Am. Meteorol. Soc., 99, 699–710, https://doi.org/10.1175/BAMS-D-17-0125.1, 2018.

Zimmer, M., Craig, G., Keil, C., and Wernli, H.: Classification of precipitation events with a convective response timescale and their forecasting characteristics, Geophys. Res. Lett., 38, L05 802, 2011.