Anwers to Comments from Referee 1

Answers are written in green font.

This work described in this manuscript extends earlier studies by the same authors, focusing on a thorough analysis of the connection of weather regimes (and their succession) with the predictability of cold-wave days in Central Europe. The analysis shows that more common ('climatological') WR successions tend to be more predictable than uncommon WR successions, while other factors like the number of regime transitions between forecast initialization and valid time did not show a clear association with forecast skill. The paper is interesting, but some clarifications and more evidence for the main conclusion is required as detailed below.

General comment:

The main conclusion of the manuscript is that among the different WR-related explanations of increased/decreased predictability the frequency of WR successions following climatological patterns plays an important role. This conclusion is primarily based on the observation that 61.6% vs. 53.9% of a subselected set of cases follows such climatological patterns. That difference is noticeable but not huge, and given the additional complication due to the subselection criterion (only the most frequent WR successions per WR at the target date are considered), which presumably has the effect of amplifying the observed difference, I feel that more evidence for this conclusion should be provided. Would it be possible, for example, to calculate the Brier score for forecasts with the GL/ScTr WR at initialization time separately for the cases where the WR successions do and do not follow a climatological pattern and test whether the score differences are statistically significant?

Only a limited set of data is available for our analysis. Although the WR proposed by Grams et al. (2017) are calculated for the time period between 1979-2022, the calculations are only based on ERA5. There are no forecasts of the WR available to us. Furthermore, since ECWMF's S2S reforecasts of the 2-meter temperatures are only available bi-weekly and for the lasts 20 winters, this further constraints the sample sizes. Mixing reforecasts with different model versions is something we like to avoid since changes in the model set-up might dilute the results. Having said this, the resulting available sample sizes are in our eyes too small to allow for a sensible significance testing depending on the WR at the initialization and the valid date. Nevertheless, to get an impression, we grouped the reforecasts which follow typical climatological patterns and which not. The results are summarized in the table below. In total, they are 104 reforecasts initialized during GL and 129 reforecasts initialized during ScTr. Of the reforecasts initialized during GL, 54% follow typical climatological pattern. Of the reforecasts initialized during ScTr, 46% follow typical climatological pattern.

GL at initialization				ScTr at initialization			
WR at valid date	# follow. clim. patterns	# follow. not clim. patterns	% follow. clim. patterns	WR at valid date	# follow. clim. patterns	# follow. not clim. patterns	% follow. clim. patterns
AT	4	12	25	AT	3	3	50
ScTr	1	2	33	ScTr	6	7	46
ZO	2	4	33	ZO	8	14	36

AR	6	11	35	AR	3	11	21
EuBL	2	2	50	EuBL	5	3	63
GL	20	11	65	GL	0	11	0
ScBL	7	1	88	ScBL	5	5	50
No	14	5	74	No	29	16	64

Specific comments:

- Section 2.2: Are the ECMWF reforecasts also temporally smoothed (like the observation data), or is that unnecessary due to the subsequent post-processing?

The reforecasts are not temporally smoothed for two reasons. First, we only use the forecast of the specific day of lead time and second, the subsequent postprocessing is already accounting for temporal uncertainties by taking the multi-year daily mean over 19 winters during the mean bias correction. Furthermore, due to the bi-weekly initialization of the reforecasts, the predictions we use are several days apart which makes the application of a running mean difficult. Additionally, for many socio-economic application the forecast of weather at a specific day is more important than a smoothed forecast, especially when it comes to extremes.

- 131: Aren't these just forecast errors of an ensemble mean forecast? I find it strange to call them biases, which to me is a systematic error, while without further aggregation the quantities calculated here contain (a substantial amount of) random forecast errors as well.

Yes, there are forecast errors but there is also a drift of the ensemble mean towards the model's climatology. This drift is in our understanding a bias and it is primarily this drift which we aim to remove from the predictions. Specifically, we substract the mean error (= systematic error = bias) of past forecasts compared to observations from the current predictions.

- Section 3.2, 2nd paragraph: More detail is required for this ERA5-based predictor. Is ERA5 data at the different hours from the day before initialization time used here? Can you briefly describe the preprocessing operations mentioned in 149?

The data at the different hours is retrieved at the day of initialization. We clarified this in the manuscript. I now states: "All meteorological fields are preprocessed by computing the minimum, mean, maximum and variance of each field before model training. This results in four predictors per meteorological variable at each time step. In case of ECMWF's S2S reforecasts, only the ensemble information (minimum, mean and maximum and their variances) instead of each individual ensemble members is taken into account. This leads to four predictors (instead of four predictors times eleven ensemble member) per meteorological variable at each time step. Furthermore, the minimum, mean, maximum and variance of the 2-meter temperature reforecast ensemble, averaged over Central Europe, is added as a predictor. The month is also added in order to account for the seasonality of temperatures and thus the occurrence of cold-wave days in winter."

- 189: I was very confused about this concept of 'hypothetical' forecasts when I read it here, and understood only later that it's not really a forecast, but that the weather regimes on these dates can still be analyzed. Maybe this can already be clarified here.

The paragraphs is changed to "In order to increase the sample size, we use in the following all days of the winters~2000/2001-2019/2020 instead of only the days where ECMWF's S2S reforecasts are initialized. This is done since the WRs on the days in-between initializations can still be analyzed. We refer to these as "hypothetical" forecasts (since no reforecasts are initialized at that date) and investigate them beside the "real" forecasts. For better reading, we use only the term "forecasts" in the following. We assume that the number of days on which each regime is present, 757~in case of GL and 713~in case of ScTr during winters~2000/2001-2019/2020, are similar enough to make a fair comparison. The analyzed WRs are based on ERA5 reanalysis data only."

- 235-236: I don't understand what is meant by 'single actual WR successions', and found this sentence very confusing. This paragraph is generally hard to follow, but it becomes clear what is studied here in connection with Figure 4. The aforementioned sentence, however, could easily be removed without loss of information.

We removed the unclear sentence from the manuscript and referenced Figure 4 in the beginning of the paragraph so the reader can directly refer to it.

- 267-268: I don't understand what is meant by 'without taking persistence of the individual WRs per se into account'. What if the WR at initialization time persists for the 14 days lead time? Please rephrase and/or explain.

We added "If one WR is persistent during all 14~days of the forecast, it is treated as a WR succession of only one WR for simplicity even if it is technically not a succession of WRs." to the manuscript.

- 275: Perhaps clearer to say '..., the number of possible WR successions varies …' √

Typos and language:

72: -> their skill √

130: Therefore -> To this end √

139: Either "the ECMWF S2S reforecast ensemble" or "ECMWF's S2S reforecast ensemble" √

159: Please check this reference, I have never seen a citation with a range of publication years before

The reference is the documentation of the python package I have used (https://skranger.readthedocs.io/en/stable/) and the time range the years of which the author had the copyright. I changed the reference to only the last year o the copyright to avoid confusion.