Response to Reviewer #2

Review of 'Extreme springs in Swizerland since 1763 in climate and phenological indices' by Noemi Imfeld, Koen Hufkens and Stefan Bronnimann

Based on an existing reconstruction of daily mean temperature and daily precipitation amounts, an overview is given of the variations in sping climate over the period since 1763 for Switzerland. The analysis is based on climatological indices and two phenological indices. In the study, the impact of climate change on these indices is documented and the climatic variability in earlier times is highlighted. A few exceptional springs are discussed in detail and in relation to the atmospheric situation.

The manuscript is well written and a joy to read. There are - as far as I can say - no methodological errors. The main concerns relate to the presentation of results which could be a bit more clear, and the selection and presentation of the climate and phenological indices needs some further thoughts. In addition, the assessment of uncertainty in the results deserves more attention.

My advise to the editor is to accept with minor revisions.

Thank you for your detailed and constructive review of our manuscript. We hope that you find the following response satisfactory.

Main Comments

*) In contrast to the warming spring, the warming of the winter climate has a delaying effect on spring phenology and Wang et al. (2020) argue that existing winter chilling model underestimate the effects of winter chilling, leading to substantial overestimates of the advance of spring phenology under climate change. A similar concern relates to the cherry flowering model and the beech leaf unfolding model. The model used to relate the cherry flowering and beech leaf unfolding is not completely clear to me, but it seems that winter chilling is not part of the equation. Motivate why the winter chilling is left out or explicitly comment on this aspect - if possible with an assessment of the consequence of not using winter chilling.

The description of the phenological model is very terse. As this is an important - and interesting! - part of the work, the description of the model should be expanded a bit to guide the readers through the model that are less well acquainted with these models.

Thank you for this relevant comment. We evaluated different phenological models for the two phenological phases which are all based on daily mean temperatures and which are implemented in the phenor R package by Hufkens et al. (2018). This also includes models with winter chilling (the alternating model and chilling degree days model, see the models in Hufkens et al., 2018). Most models showed largely similar behaviours. We agree, however, that not considering winter chilling may affect the model estimates of phenological dates. Our models are calibrated based on data from 1950 to 2020, however with more observations in the later years. The last around 30 years showed a relevant advancement of spring phenology especially for cherry flowering with likely lower chilling accumulations in winter. Calibrating a model on such data without considering winter chilling might lead to a higher heat requirement as shown in Wang et al. (2020). Further, transferring these models to the past could lead to phenological dates being later in spring. This is hypothetical, and a comprehensive study on such effects would be needed. In this article, we focus, however, on providing a first estimate of phenological phases based on a commonly used processed-based model for the past 258 years. Providing a comprehensive study of phenological models and effects of different models for the past is highly relevant, but out of scope for this article.

To make the section about the phenological application more clear, we will expand the description of the phenological model by including the current Appendix A in Section 3.2. Further, we will discuss the individual terms in the equations of the phenological models in more detail in the same section. Also, we will briefly discuss the effect of model selection (e.g. lack of winter chilling) and model calibration period for estimating phenological dates of the past few hundred years.

*) the selection of climatic indices is strongly biased towards the temperature-related indices. The only two indices which are precipitation-based are the number of Wet Days and Snowfall Days (the latter is a mix between precip and temperature). Although these two indices are relevent, it would be interesting to add indices that relate to droughts or pluvials - like the Consecutive Dry Days or Consecutive Wet Days indices. This would contribute to earlier studies on droughts where a propagating signal from spring drought into summer drought

is observed, and might give some perspective on e.g. the drought in the mid 1940s in central Europe (Brazdil et al. 2016; Hirschi et al. 2013)

We primarily focused on the temperature indices because the precipitation reconstruction shows a much lower reliability than the temperature reconstruction (see Imfeld et al., 2023) and we do not want to promote using the precipitation reconstruction in an imprudent way. We will split Figure 1 into temperature indices and precipitation indices which allows us to increase the size of the individual maps, and we will add the Consecutive Dry Days index since it is also based on wet days and a relevant index for spring weather.

*) In the discussion of the quality of the reconstruction, it was noted that the skill in the temperature reconstruction is higher than that of precipitation. This is perfectly understandable, but what is missing is a view how this uncertainty propagates into the indices. It would have been very nice if the authors would be able to assess the uncertainty in the indices, and therefore in the conclusions. I briefly went through the paper that documents the reconstruction, but I understand that this rerconstruction does not come with an uncertainty estimate in terms of an ensemble? That would have made the assessment of uncertainty not too difficult (it only requires quite a bit of computations). The uncertainty assessment in the manuscript is now based on using various sources reconstructions (like 20CR, ModE-RA and long observational records). The spread in the various reconstructions is demonstrated in fig. 2, but the sometimes large deviations between the estimates is not discussed. Particularly the cherry flowering in the Liestal deserves some attentiuon as the observations show much stronger variability and show for many years much earlier flowering. Can you indicate if this discrepancy relates to the temperature reconstruction or is there an issue with the phenological model?

Thank you for this comment. This comment addresses two different points, a) the uncertainty in the original reconstruction and b) the deviations in the cherry flowering estimates concerning an observational series. We will answer these two points individually.

a) In the article about the data set (Imfeld et al., 2023), we describe in detail how the reconstruction is performed. This includes an ensemble based on the 50 best analogue days which is used in the Ensemble Kalman Fitting for the temperature reconstruction (for details, please refer to the article). However, this is not a true ensemble but an ordered ensemble, where the first analogue day (based on its Gower distance) is a better representation of the historical period than e.g. the second analogue day. A true ensemble would be needed to do a proper uncertainty analysis. However, we propose to show for the phenological estimates of Liestal the 10 best analogue days, including temperature assimilation, in order to depict the uncertainty in the reconstruction, while emphasising that this is not equivalent to a true ensemble.

b) Comparing the reconstructed cherry flowering with the cherry flowering series of Liestal shows a mean bias of 7.36 days as stated in the manuscript. The phenological estimates are based on a phenological model calibrated using cherry flowering observations across Switzerland and then applied to each grid cell of the temperature reconstruction. A direct comparison of the closest grid cell from our reconstructed cherry flowering and the cherry flowering series of Liestal can be biased for several reasons. For example, the exposition and microclimate of the tree (within the 1 km grid cell) can be different leading to e.g. warmer temperatures at the location of the tree and leading to earlier flowering. Also, tree-specific characteristics can lead to a flowering different from when a model is calibrated using all cherry flowering trees in Switzerland likewise leading to biases for this specific tree.

Interesting is also that the GGD and mean temperature reconstructions are spot-on with the 20CR after ±1840, but for the earliest decades, there is a bit of a bias. Can you comment on this?

We assume that this difference in the early years mainly stems from the 20CRv3 reanalysis. The first years of the 20CRv3 reanalysis from 1806 to 1835 are an experimental extension (Slivinski et al., 2021). In our case the results seem to deviate more strongly from observations for the experimental extension of the reanalysis.

Other concerns the authors may want to look at

*) line 95: The group involved in the ETCCDI also prescribed levels of missing data that are allowed in the aggregation of seasonal/annual values. Can you argue why you deviate from their approach by selecting the 10% threshold?

We used the 10% threshold because gaps often occur for several consecutive days, which could affect the seasonal aggregated value by e.g. missing a specific synoptic situation. Note, that this threshold only applies to the Swiss plateau series (Brugnara et al., 2022) and thus, is only relevant for a comparative purpose in Figure 2. The Swiss reconstruction is a spatially and temporally complete data set.

*) Figure 1: Except for the warm spell duration index and cold spell duration index (figs. e and f), the climatologies of the 30-yr periods are quite similar. This has been noted in the text. The figure would be a bit more interesting if you would show one reference period (e.g. your favourite 1871-1900 period) and deviations from this reference for the other periods.

One of the main messages of this Figure is that the 30-year periods are quite similar to each other, except for the last few decades. To show the differences in more detail we added in the Appendix a Figure showing the difference between all periods and the 1871-1900 period.

*) figure 5: it seems that in the figures for the Last Frost Day, something is wrong. I see purple vertical stripes and I wonder if some detail is lacking in figure a?

There is nothing missing in the Figure. Please, see the caption: The vertical purple lines indicate areas where frost (cherry flowering) occurred 15 d later (earlier) than the 1871 to 1900 average.

Smallish concerns

*) line 81: for completeness, you could mention that your definition is considered colder than what is usually called a frost, but also warmer than an ice day (where the Tmax drops below zero)

Thank you, we will add this in a sentence.

*) line 197: the use of two digits for temperature is not in-line with what is used in the rest of the manuscript.

We will correct this.

*) figure 5: the labels of the colour bars are difficult to read with this size of the figure (I needed to zoom to read it). The a) and b) labels are set twice.

Thank you for these comments as well. We will correct them in the revised manuscript.

References

Brugnara, Y., Hari, C., Pfister, L., Valler, V., and Brönnimann, S.: Pre-industrial temperature variability on the Swiss Plateau derived from the instrumental daily series of Bern and Zurich, Climate of the Past Discussions, pp. 1–34, https://doi.org/10.5194/cp-2022-34, 2022.

Hufkens, K., Basler, D., Milliman, T., Melaas, E. K., and Richardson, A. D.: An integrated phenology modelling framework in r, Methods in Ecology and Evolution, 9, 1276–1285, https://doi.org/10.1111/2041-210X.12970, 2018.

Imfeld, N., Pfister, L., Brugnara, Y., and Brönnimann, S.: A 258-year-long data set of temperature and precipitation fields for Switzerland since 1763, Climate of the Past, 19, 703–729, https://doi.org/10.5194/cp-19-703-2023, 2023.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J., McColl, C., Allan, R., Brohan, P., Yin, X., Smith, C., Spencer, L., , Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An evaluation of the performance of the twentieth century reanalysis version 3, Journal of Climate, 34, 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Wang, H., Wu, C., Ciais, P., Peñuelas, J., Dai, J., Fu, Y., and Ge, Q.: Overestimation of the effect of climatic warming on spring phenology due to misrepresentation of chilling, Nature Communications, 11, 4945, 2020.