Reply to the reviewers comments (RC2)

The paper focuses on the cold-air outbreak in early June 1816, during the Year Without Summer, characterized by extraordinarily cold and wet periods in Central Europe. Using the Weather Research and Forecasting (WRF) Model and the 20th Century Reanalysis (20CR) product, the authors perform dynamical downscaling combined with data assimilation of early instrumental observations. The findings highlight the capability of the WRF model to reproduce regional to local meteorological processes and improve the accuracy of simulations when early pressure and temperature measurements are assimilated.

The Year Without Summer, caused by the eruption of Mount Tambora in 1815, has been extensively studied in historical climatology. Previous research often relied on descriptive sources and early instrumental measurements aggregated on a monthly basis. This paper builds on these foundations by providing high-resolution, sub-daily weather simulations, adding significant detail to the understanding of climatic impacts during this period.

The work provides a novel approach to analyzing historical weather events using a combination of dynamical downscaling and data assimilation. This method allows for a more detailed and accurate reconstruction of past weather events than was previously possible, offering new insights into the meteorological conditions of the Year Without Summer. The findings underscore the importance of digitizing early instrumental data and demonstrate the potential of modern numerical models in historical climatology.

The manuscript is well-structured and thorough, presenting a detailed methodology and comprehensive results. The use of both qualitative and quantitative validation against independent historical observations strengthens the credibility of the findings. The careful selection and bias-correction of assimilated data ensure the quality and reliability of the simulations.

We would like to thank the reviewer for the positive feedback and the valuable comments and suggestions for improvement.

While the manuscript is robust, a few areas suggested below could benefit from further clarification:

P4, L105-107 mentions the use of weather diaries and records of eye observations regarding sunshine, cloudiness, precipitation, wind, and other variables. It would be helpful to (1) Specify how the qualitative descriptions are converted into comparable data points and any challenges faced during this process. (2) Clearly state how the digitized eye observations are integrated into the validation process of the simulations. Highlight any specific methodologies used to ensure the reliability of these qualitative data points.

This is a valuable point that we did not mention in the manuscript. We will include some explanations in Sect. 2.1. (P5, L142ff) as follows:

"Furthermore, we use eye observations from selected weather diaries in Switzerland that were recorded in semi-standardized terminology (Auchmann et al., 2012; St. Gallen / Meyer from CHIMES; Hari, 2021). Adjustments included manual re-coding of the available information into classes (e.g. bright, partly cloudy, overcast for cloud cover) or categories (e.g. spray, rain, snow for precipitation types). Refer to Auchmann et al. (2012), Brugnara et al. (2015) or Brönnimann (2023) for information on how to best re-code symbolic or word information, re-classifiy it, and attribute a plausible time to a record, among others.

Given the inherent uncertainties, we consider the final data points as being of qualitative, complementary and relative information which, when taken as a whole, may support or contradict our model outputs."

P5, Line 139-141, The manuscript mentions that the observation series are not homogenized, which could impact the reliability of the assimilated data. A discussion on the potential effects of this and any mitigation strategies would be beneficial.

Thank you for the suggestion. We will include a short explanation at P5 L141.

"Although we assume the raw records to be consistent over a period of only ten days, potential errors must be considered and interpreted when showing the raw data. For assimilation however, a simple correction approach was applied to correct for biases in the measurement series (see Sect. 2.4). In addition, the assimilation algorithms reject values that are too far off from the first-guess simulated value. Hence, negative effects of potentially erroneous values on the assimilation can be considered to be rather small to negligible."

P8, Line 173-174, the manuscript mentions the use of the 20CR as initial and boundary conditions for the downscaling experiments, but it does not specify the exact variables utilized. The distinction between variables used for downscaling and those used for data assimilation is crucial, and the manuscript does emphasize the use of data assimilation for pressure and temperature. However, it lacks clarity on which variables are used exclusively for downscaling. To improve clarity and completeness, the manuscript should explicitly list the variables read from the 20CR for dynamical downscaling. This list should differentiate between variables used for initial and boundary conditions in downscaling and those used in data assimilation.

It is indeed important to provide information on the variables from 20CR which serve as initial and boundary conditions. We will clarify this in Sect. 2.2 of the revised manuscript:

"Initial and boundary conditions of the regional simulations are taken from the 20CR ensemble mean. In particular, these encompass three–dimensional fields of temperature, humidity, geopotential height, pressure, and the horizontal components of wind speed, as well as two-dimensional fields of 2 m temperature and humidity, 10 m wind components, surface and sea level pressure, snow depth, skin temperature, sea surface temperature, a land/sea mask and a sea ice flag. Furthermore, four layers of soil temperature and soil moisture from 20CR are used to initialize the regional model."

P8, Line 182-184, the manuscript states that the WRF model employs three nested, limited-area domains with cell sizes of 27 km, 9 km, and 3 km. These domains are nested to refine the global information to regional and local scales. It could be helpful to describe the process of providing lateral boundary conditions for each nested domain. Specify how the data from the parent domains are used to initialize and drive the simulations of the nested domains. Consider including a diagram that illustrates the nesting process and the flow of lateral boundary conditions from the outermost to the innermost domains. This visual aid would help readers better understand the methodology.

Given the target audience of the journal, we agree on including a few more sentences on the actual flow of information from coarse to fine scales.

"In simple terms, the process of providing initial and lateral boundary conditions for each nested domain is as follows: For model initialization, the nested, smaller domain receives the information from the coarser domain at the horizontal and vertical coordinates that both domains share. For all other coordinates (and for the outermost domain which typically does not share exact coordinates with the reanalysis), this information is spatially interpolated to the finer grid cells. The simulations are then incremented going forward in time until new information from the coarser domain is available, which is fed in at the lateral boundaries of the nested domain."

P8, Line 178, mentions "Here, we mainly use the mean of the 80 ensemble members." It is essential to introduce here how these ensembles are used, specifically whether they used the ensemble mean or individual ensemble members. In the later part of the manuscript, they indicate that the ensemble mean of the 80 ensemble members is primarily used for synoptic analyses and as initial and boundary conditions for the downscaling simulations. Considering the high variability on synoptic scale, can 80 ensemble mean on 6-hour time scale reflect the weather pattern specifically for June in 1816? It would be beneficial to discuss the implications of using the ensemble mean versus individual ensemble members. This could involve addressing the potential smoothing effects and how this choice impacts the simulation results and their interpretation.

Thank you for this suggestion. Whereas the full ensemble of 20CR is only used for veryfing the representation of the summer 1816 in this dataset (see Fig. S5 in the supplement), the other analyses presented in the manuscript are based on the 3-hourly ensemble mean. We will clarify this in the revised manuscript.

We also agree that it is important to explain the effects from using the ensemble mean in this study. In fact, we address this in the results section 3.1 on P12 L277ff, and we give an interpretation of using the ensemble mean versus members on P21 L496ff. In addition, we will insert a sentence about our previous experiences with smoothing effects, as well as related findings in published literature.

"Previous studies for the same region found some deviations in variables such as maximum wind speed from using the ensemble mean versus members, but small smoothing effects in the pressure fields and overall limited benefit from applying an ensemble approach when comparing to station observations (Stucki et al, 2015, 2016). Whereas regarding extreme events, smoothing effects may be more pronounced (e.g. Mahoney et al., 2022), the ensemble mean was found to provide accurate initial and boundary conditions, even if possibly less accurate than individual ensemble memebers (Michaelis and Lackmann, 2013)"

P13, Line 310, while Figure 2 effectively presents data for domains D02 and D03, it skips the outermost domain (D01) and the comparison with the 20CR data. Including D01 in the figure would help demonstrate the first step of the downscaling process, showing how well the WRF model captures large-scale atmospheric patterns compared to the 20CR data. This is crucial for validating the initial downscaling step and ensuring that the model accurately represents the broader atmospheric conditions before refining them in the nested domains. This also helps readers better understand the progressive refinement of the atmospheric data from the global scale (20CR) to the regional (D01) and local scales (D02 and D03), and demonstrates each step of the downscaling process enhances the transparency and credibility of the methodology. By the way, the 500 gph contours can not be observed in Fig2 (b) and (c).

Thank you for this suggestion. Referring also to the first reviewer's comments, we will indicate the borders of the three nested domains in figure 2a and add a panel to the figure showing D01 in the revised version. We will furthermore add the 500 gph contours in all panels.

P17, Line 392, Figure 5 presents meteograms comparing station observations and measurements with WRF model outputs for various variables over a period of time. While it provides comprehensive data, the figure appears too busy, making it difficult to easily interpret the comparisons. I suggest to simplify the figure to make it easier for readers to extract key information. Can group related variables together to provide a clear comparison and emphasize the most important differences between observations and model outputs.

A similar suggestion was made by Referee 1, so we realize that it would be better to reduce the information. For this, we will just keep the Bern location with the full information in one panel, and we will have a second panel with just the most illustrative values, and without the NODA information on it, see the planned figure caption below. Because we think of the mimicked meteograms as a comprehensive illustration, we will place the current figure with the full information in the Appendix. The text will be adapted to follow the new figure.

"Figure 5: Meteograms of station observations and measurements (red colors) for a) Bern (station id be01) with WRF NODA (ligher colors, grey, blue, orange) and DA (darker colors) simulations output for the nearest grid point, for the period between 4 and 12 June 1816 (x-axis).
The top panel in shows observed cloud cover (red squares from 'bright' with no fill to 'mixed' with cross and 'covered' with fill) vs. simulated low, mid- and upper level cloud fraction (larger bars indicate more cloudiness). The second row panel in a) shows simulations of downward short wave flux at ground surface (orange lines; W/m2) and relative humidity at 700 hPa (dashed blue line; permil).
The third row panel shows thrice-daily observations of precipitation vs. simulated precipitation (blue vertical bars, mm). The bottom panel shows observations of wind direction (red vectors at unit length; north is up) vs. simulated wind direction (grey wind vectors; north is up) and velocity (black line and vector length; m/s at 10 m above ground).
Panels b) are as in a), but for selected values for Zurich (zh00), Geneva (ge00) and St. Gall (sg01). Red vertical crosses (dark for observer Escher, light for Feer) indicate the relative humidity measurements on the ground for Zurich. Red vertical bars show precipitation measurements in mm for Geneva; darker red dots for Zurich stand for observer Escher, lighter for Feer. The red tilted cross means 'rain and snow' for St. Gall. Grey plus signs mean confirmed obserations, minus sign mean no record of precipitation found."

P23, in the summary and conclusions, the manuscript briefly touches on the successful application of the methodology for the Year Without Summer but does not delve deeply into the challenges and considerations for applying these methods further back in time. To enhance the manuscript's discussion on the wider application of their methodology in past climates, it would be beneficial to include a more detailed exploration of the challenges and potential strategies for overcoming them. This can be done in the discussion or conclusion sections.

We think that we have already mentioned the challenges and potential in the conclusions section. However, we can make a clearer statement by adding one or two sentences, e.g.

"Given that global or regional gridded datasets may soon appear for periods beyond 1816 and more early measurements will become available, the prospects of soon entering the 18th century with such 4-dimensional studies are very good."

References:

Auchmann, R., Brönnimann, S., Breda, L., Bühler, M., Spadin, R., and Stickler, A.: Extreme climate, not extreme weather: The summer of 1816 in Geneva, Switzerland, Clim. Past, 8(1), 325–335, https://doi.org/10.5194/cp-8-325-2012, 2012.

Brönnimann, Stefan (2023). The weather diary of Georg Christoph Eimmart for Nuremberg, 1695–1704, Clim. Past, 19(7), 1345-1357, https://doi.org/10.5194/cp-19-1345-2023, 2023.

Brugnara, Y., Auchmann, R., Brönnimann, S., Allan, R. J., Auer, I., Barriendos, M., Bergström, H., Bhend, J., Brázdil, R., Compo, G. P., Cornes, R. C., Dominguez-Castro, F., van Engelen, A. F. V., Filipiak, J., Holopainen, J., Jourdain, S., Kunz, M., Luterbacher, J., Maugeri, M., Mercalli, L., Moberg, A., Mock, C. J., Pichard, G., Řezníčková, L., van der Schrier, G., Slonosky, V., Ustrnul, Z., Valente, M. A., Wypych, A., and Yin, X.: A collection of sub-daily pressure and temperature observations for the early instrumental period with a focus on the "year without a summer" 1816, Clim. Past, 11, 1027–1047, https://doi.org/10.5194/cp-11-1027-2015, 2015.

Hari, C.: An Evaluation of Meteorological Observations by Samuel Studer (1807-1818), MSc thesis, University of Bern, Switzerland, 127 pp., https://occrdata.unibe.ch/students/theses/msc/343.pdf, 2021.

Mahoney, K., McColl, C., Hultstrand, D. M., Kappel, W. D., McCormick, B., and Compo, G. P.: Blasts from the Past: Reimagining Historical Storms with Model Simulations to Modernize Dam Safety and Flood Risk Assessment, Bulletin of the American Meteorological Society, 103(2), E266-E280, https://doi.org/10.1175/BAMS-D-21-0133.1, 2022.

Michaelis, A. C. and Lackmann, G. M.: Numerical modeling of a historic storm: Simulating the Blizzard of 1888, Geophys. Res. Lett., 40(15), 4092–4097, https://doi.org/10.1002/grl.50750, 2013.

Stucki, P., Brönnimann, S., Martius, O., Welker, C. S., Rickli, R., Dierer, S., Bresch, D. N., Compo, G. P., Sardeshmukh, P. D., Romppainen-Martius, O., Welker, C. S., Rickli, R., Dierer, S., Bresch, D. N., Compo, G. P., Sardeshmukh, P. D., and Sardeshmukh, A. P. D.: Dynamical downscaling and loss modeling for the reconstruction of historical weather extremes and their impacts – A severe foehn storm in 1925, Bull. Am. Meteorol. Soc., 96(8), 1233–1241, https://doi.org/10.1175/BAMS-D-14-00041.1, 2015.

Stucki, P., Dierer, S., Welker, C. S., Gómez-Navarro, J. J., Raible, C., Martius, O., and Brönnimann, S.: Evaluation of downscaled wind speeds and parameterised gusts for recent and historical windstorms in Switzerland, Tellus. Series A - dynamic meteorology and oceanography, 68(1), 31820, https://doi.org/10.3402/tellusa.v68 .31820, 2016.