

Author's response

To nhess-2019-174 Submitted on 31 May 2019

Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux

Peter Stucki, Paul Froidevaux, Marcelo Zamuriano, Francesco Alessandro Isotta, Martina Messmer, and Andrey Martynov

Contents

- List of relevant changes made in the manuscript
- Marked-up manuscript version
- Response to Reviewer 1
- Response to Reviewer 2
- Response to Reviewer 3

30-09-2019, Peter Stucki

List of relevant changes made in the manuscript

(page number in brackets refers to the pdf with markups below)

Abstract

- Contribution of (embedded) convection added (p.1)

Introduction

- Enhanced definition of 'cyclonic' and 'cyclonic moisture flux' (p. 2)

Data and Methods

- Extended description of RrecabsD gridded reconstruction of precipitation (p. 4)
- Number of stations in Switzerland that are assimilated into 20CR for 1876, 1910, 2005 (p. 4)

Downscaling of 2005 case / Evaluation of setup

- Rationale for extended initialization period for WRF simulation (p. 5)
- Rephrasing of results from Fig. 2 wrt. CombiPrecip gridded dataset (p. 6)
- Reasoning for additional experiments after evaluation of the standard WRF setup (p. 6, 7)
- Extended description of WRF configurations (p. 7)
- Extended description of evaluation measures (MAE, EMD, Theil-Sen slopes; p. 9)
- Extended discussion of setup evaluation (p. 10)

Analyses and simulations of three cases

- Extended description of SLP and isobaric pressure fields for cyclone tracking (p. 10)
- Short description of Wernli and Schwerz 2006 cyclone tracking method (p.11)
- Extended analyses of Fig. 4, enhanced caption Fig. 4 (p.14.15)
- Added panels to Fig. 5 with 9-km domain and convective contribution (p. 15)
- Extended analyses of Fig. 5 (p. 15, 16)
- Extended analyses of Figs. 7-9, color key introduced (p.18, 19, 20)
- Rephrased interpretation of analyses Figs. 7 -10 (p. 22)
- New analyses of new Fig. 11 concerning innermost domains (p.22, 23)

Conclusions

- Extended explanation of 'small-scale adaptation' and influence of large-scale flow (p. 24)
- Enhanced conclusion on increasing variability in the ensemble when going back in time (p. 25)
- Inclusion of single-member analyses in new Fig. 11 (p. 25)

Appendix

- Fig. A2 redrawn with maximum and minimum members

References

- A number of additional references given

Figures and Tables

- Fig. 1 redrawn (no edges cut)
- Fig. 2 includes member number
- Fig. 4 redrawn (panels g- i)
- Fig. 5 redrawn (new panels of 9-km domain, WRF-ERA-Interim added)
- Figs. 7-9 redrawn with color key, enhanced raster image quality
- New Fig. 11
- Fig. A2 redrawn with max. and min. member

Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux

Peter Stucki^{1,2}, Paul Froidevaux^{2,4}, Marcelo Zamuriano^{1,2}, Francesco Alessandro Isotta⁴, Martina Messmer^{1,3,5}, Andrey Martynov^{1,2}

5 ¹Oeschger Centre for Climate Change Research, University of Bern, Bern, 3012, Switzerland

²Institute of Geography, University of Bern, Bern, 3012, Switzerland

³Climate and Environmental Physics, Physics Institute, University of Bern, Bern, 3012, Switzerland

⁴Federal Office for Meteorology and Climatology MeteoSwiss, Zurich-Airport, 8058, Switzerland

⁵now at: School of Earth Sciences, The University of Melbourne, Melbourne, Victoria, Australia

10

Correspondence to: Peter Stucki (peter.stucki@giub.unibe.ch)

Abstract. In June 1876, June 1910 and August 2005, northern Switzerland was severely impacted by heavy precipitation and extreme floods. Although occurring in three different centuries, all three events featured very similar precipitation patterns and an extra-tropical storm following a cyclonic, so-called Vb trajectory around the Alps. Going back in time from the recent to the historical cases, we explore the potential of dynamical downscaling a global reanalysis product from a grid size of 220 km to 3 km. We use the full, 56-member ensemble provided in the reanalysis and a regional weather model to investigate sensitivities of the simulated precipitation amounts to a set of differing model configurations. The best-performing model configuration in the evaluation, featuring a 1-day initialization period, is then applied to assess the sensitivity of simulated precipitation totals to cyclonic moisture flux along the downscaling steps. The analyses show that cyclone fields and tracks are well defined in the reanalysis ensemble for the 2005 and 1910 cases, while deviations increase for the 1876 case. In the downscaled ensemble, the accuracy of simulated precipitation totals is closely linked to the exact trajectory and stalling position of the cyclone, with slight shifts producing erroneous precipitation, e.g., due to a break-up of the vortex if simulated too close to the Alpine topography. Simulated precipitation totals only reach the observed ones if the simulation includes continuous moisture fluxes of $>200 \text{ kg m}^{-1} \text{ s}^{-1}$ from northerly directions, and high contributions of (embedded) convection. Misplacements of the vortex and concurrent uncertainties in simulating convection, in particular for the 1876 case, point to limitations of downscaling from coarse input for such complex weather situations and for the more distant past. On the upside, single (contrasting) members of the historical cases are well capable of illustrating variants of Vb cyclone dynamics and features along the downscaling steps.

15

20

25

30

Deleted: so called

Deleted: These setups are evaluated by combining spatial verification metrics, inter-subjective visual inspection and an objective similarity measure.

Deleted: setup

Deleted: and moderate spectral nudging

Deleted: To reproduce the extreme events,

Deleted: accurate

Deleted: are required.

Deleted: with

Deleted: a well-reasoned selection of reanalysis members for downscaling may be adequate in cases where the driving large-scale features in the atmosphere are well known

1 Introduction

Floods are among the most damaging natural hazards worldwide (Bevere et al., 2018); they affect more people than any other natural hazard (CRED 2019). The costliest flood event in Switzerland of the last decades occurred in 2005 (Hilker et al. 2009); it caused fatalities and led to heavily damaged infrastructure (Bezzola et al. 2008). This event was well documented and subsequently, a range of publications analyzed the flood-inducing meteorological conditions (e.g., Frei 2005; Beniston 2006; MeteoSwiss 2006; Bezzola and Hegg 2007; Zängl 2007a; Bezzola and Hegg 2008; Hohenegger et al. 2008; Jaun et al. 2008; Langhans et al. 2011; Stucki et al. 2012; Messmer et al. 2017).

35

Deleted: BAFU 2005

On a synoptic scale, the associated extratropical cyclone mainly followed the classical so-called Vb cyclone track after Köppen (1881) and van Bebber (1891), see also Hofstätter et al. (2016). Cyclones on a Vb track are associated with heavy to extreme precipitation over Central Europe, and particularly north of the Alps (Hofstätter et al., 2016, 2018; Nissen et al., 2013). Most of the cyclones following the Vb trajectory are generated in the western Mediterranean region, and most of them pass the Golf-of-Genoa region (Hofstätter and Blöschl, 2019; Messmer et al., 2015). For this, they are also called Genoa Lows at this stage; Bezzola et al. 2008). They take up moisture over the Ligurian Sea, then propagate eastward to the Adriatic Sea and recurve northward (Hofstätter et al., 2018; Hofstätter and Blöschl, 2019; Messmer et al., 2015; Pfahl, 2014; Ulbrich et al., 2003). Regarding the pattern of mid-tropospheric geopotential heights, the 2005 case has also been characterized as a pivoting cut-off low (PCO; Stucki et al. 2012; Froidevaux and Martius 2016) referring to the recurving track of the system around the Alps while its axis of symmetry turns from meridional to zonal; see also Awan and Formayer (2017) for a general description of cut-off lows and their influence on extreme precipitation in the European Alps. Furthermore, quasi-stationarity (i.e., stalling) of the system over northern Italy was important: In the cyclonic circulation around the Vb cyclone, large quantities of warm and humid Mediterranean air were led over and around the Eastern Alps. This dynamic mechanism is described as "cyclonic moisture flux" hereafter. With the term "cyclonic", we refer to the pattern of the moisture flux streamlines, forming a (closed) anticlockwise rotation around the cyclone center. Downstream, the cyclonic moisture flux impinged onto the slopes of north-eastern Switzerland from sector North (Hohenegger et al. 2008; Froidevaux and Martius 2016; Messmer et al. 2017). The intensity of the integrated water vapor transport (IVT) was estimated to exceed $300 \text{ kg m}^{-1} \text{ s}^{-1}$ (Froidevaux and Martius 2016). Hence, IVT was found to be an important precursor for severe floods in Switzerland (cf. Kelemen et al., 2016, for a European summer flood in 2013). Precipitation occurred in two peak episodes in the afternoons of 21 and 22 August 2005, respectively, when stratiform upslope orographic precipitation was locally enhanced by embedded convection (Hohenegger et al. 2008; Langhans et al. 2011).

Although the impact of the 2005 event was very severe, it was not unique in a historical context. Several studies found similar spatial distributions of damage and precipitation, as well as similar synoptic-scale weather patterns. Two such analog cases occurred in June 1910 and June 1876 (Röthlisberger 1991; Pfister 1999; Frei 2005; Stucki et al. 2012). For instance, their similarities were analyzed on a synoptic scale using the Twentieth Century reanalysis (20CR; Compo et al. 2011), and classified as PCO type 1 (2005, 1910) and type 2 (1876), where the 1876 case features a more north-westerly flow towards the Swiss Alps (Stucki et al. 2012).

However, two options for hydro-meteorological analyses have not been considered so far to learn from these historical cases. The first option is the systematic use of a reanalysis ensemble to assess sensitivities of the severe weather with regards to determining factors such as cyclone trajectories or IVT. The second option is using the global reanalysis products for dynamical downscaling to meso-scale resolutions, i.e. the nesting of limited-area weather models into larger-scale models in several refinement steps (von Storch et al., 2000; von Storch and Zorita, 2019). In fact, 20CR has proven to be a valuable input dataset for downscaling heavy-precipitation and windstorm events over the Central Alps back to the 19th century. Stucki et al. (2018) showed that downscaling the ensemble mean is not only computationally cheaper, but can be seen as a minimum-error and thus natural approach in well-represented areas and distinctive synoptic flow conditions. For an extreme flood in 1868, they found a small smoothing effect of the associated cyclone, which induced southerly moisture flux, i.e. perpendicular to the Alpine

Deleted: (often

Deleted: BAFU

Deleted: cyclonic

Deleted: is

range. In contrast, Hohenegger et al. (2008) used a limited-area ensemble prediction system to assess potential benefits for precipitation forecasting. They found that member-to-member variability tends to have a larger effect than resolution. Other studies point to limited benefit from the full ensemble and recommend using a set of well-chosen members (Jaun et al. 2008; see also Horton and Brönnimann, 2018, for statistical downscaling of precipitation fields).

For dynamical downscaling, there are manifold options regarding the configuration of the limited-area weather models, including the choice of adequate reanalysis products as input datasets, initialization spans (so called spin-up), spatial extent and resolution of the simulation domains, or model physics (e.g., Prein et al., 2015). To date, only 20CR covers the 1876 case, at the expense of a coarse grid size of $2^\circ \times 2^\circ$ in the horizontal, while finer-resolved reanalysis products have been used to downscale cases after 1900 (e.g. Brugnara et al. 2017). Regarding the initialization, long spin-ups would allow soil moisture and similarly slow-adapting model variables to reach an equilibrium, at the expense of potentially losing control over the simulation with time. In turn, short spin-up times constrain the potential evolution close to the large-scale input. For Vb cyclones, relatively short spin-ups were chosen by Hohenegger et al. (2008) and Messmer et al. (2015, 2017). Regarding spatial resolution, cloud-resolving and convection-permitting grid sizes equal to or lower than about 3 km are necessary to reproduce the precipitation of the 2005 case (Zängl 2007b; cf. Prein et al., 2015). Typical setups of limited-area models mostly include explicit production of precipitation in the innermost domain, while convection is parameterized for the coarser domains. Further options are one- versus two-way nesting or nudging in one or more domains.

In this study, we assess the 2005, 1910 and 1876 cases in two ways. First, we aim to find a setup of the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008) that is adequate for dynamical downscaling from 20CR and for our cases. For this, we use the 2005 case as testbed because of the large amount of observations available for verification. Second, we apply the chosen setup to all three cases, aiming to investigate relevant atmospheric features that induce heavy precipitation, and to assess the inevitably increasing uncertainty along the downscaling steps and among the ensemble members as we go back in time from 2005 to 1910 to 1876.

The article is organized as follows. Data and models are introduced in Section 2. The experiments with different model setups are described in Section 3. The synoptic and meso-scale reconstructions of the three cases are presented in Section 4. A summary and conclusive remarks are given in Section 5.

2 Data and Models

2.1 Observation-based precipitation datasets

All observation-based precipitation datasets used in this study come from the Federal Office for Meteorology and Climatology MeteoSwiss. Precipitation totals derived from all three observation-based products are shown in Figure A1 in the Appendix. The first dataset are observations of daily precipitation totals. These measurements are quality checked and homogenized according to Fülleemann et al. (2011).

CombiPrecip, the second product, is also a gridded dataset. It results from a geo-statistical combination of rain-gauge observations and radar images. It covers the entire Swiss territory for the period 2005 to present at high spatial and temporal resolutions. The hourly precipitation accumulation is available as a running sum updated every 10 minutes on a 1-km grid.

The third MeteoSwiss product is RrecabsD, a prototype dataset specifically calculated for this study with a statistical reconstruction technique. The procedure was previously used to reconstruct monthly and daily precipitation in the Alps for different scopes (Isotta et al., 2019; Masson and Frei, 2016; Schiemann et al., 2010; Schmidli et al., 2002). It involves a Principal Component Analysis (PCA) of a high-resolution grid dataset in a calibration period defined between 1981 and 2010 and an Optimal Interpolation of PCA scores from long-term station data. The high-resolution dataset used for the PCA is RhiresD (MeteoSwiss, 2013) on a grid-size of 2.2 km, with daily precipitation totals retrieved by spatial interpolation of rain-gauge measurements within the Swiss borders. In RrecabsD, the focus is on spatial consistency by using all station measurements available both in the respective days in 1910 or 1876 and in a consistent part of the calibration period (1981 - 2010), which is accordingly slightly reduced by eliminating the days with gaps at one or more of the chosen stations. The three datasets described above integrate observational information and are used as reference. They are, like other dataset, affected by uncertainties and errors, which were analyzed in detail in the provided references and in several applications.

2.2 Reanalysis datasets

The Twentieth Century Reanalysis dataset version 2c (20CR; Compo et al., 2011) is used for synoptic analyses and as initial and boundary conditions for the downscaling experiments. 20CR is a global atmospheric reanalysis with a 2-degree spatial grid (approx. 220 km over Europe), 28 vertical hybrid-sigma pressure levels and 6-hourly temporal resolution going back to 1851. Only surface pressure observations are assimilated. Over the time period of the 1876 case, a number of four stations within Switzerland are used in the assimilation (Bern, Sion, Grand St. Bernard, and Geneva). This number grows to six for the 1910 case (including also Zurich and Basel), and to 34 for the 2005 case. The 20CR ensemble mean and 56 members (in fact, 1st-step deviations from the ensemble mean) are available.

The ERA-Interim (Dee et al. 2011) and CERA-20C reanalyses (Laloyaux et al. 2016) are used for comparisons of synoptic fields to 20CR. ERA-Interim (CERA-20C) has a horizontal grid size of approx. 80 km (125 km), 37 (37) pressure levels, 6-hour (3-hour) temporal steps and reaches back to 1979 (1901). ERA-Interim is also used as initial and boundary conditions for downscaling the 2005 case to compare with downscaling based on 20CR.

2.3 Regional model WRF

The non-hydrostatic Weather Research and Forecasting model (WRF-ARW Version 3.7.1; Skamarock et al. 2008) is used for dynamical downscaling of the 2005, 1910 and 1876 cases. An initial model setup is applied to downscale the 2005 case, and a number of nine follow-up setups are used to investigate differences among the setups regarding the representation of precipitation totals, and to select a best setup to downscale the 1910 and the 1876 cases.

Deleted: we

Deleted: ;

Deleted: t

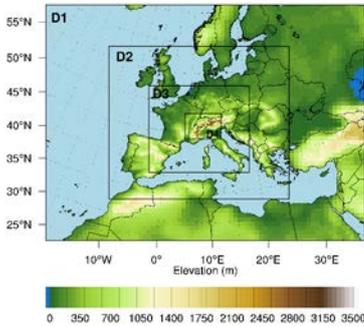


Figure 1: Nesting of the WRF regional model into 20CR, where D1 to D4 refer to the simulation domains with 81-km, 27-km, 9-km and 3-km grid sizes. Shading indicates model topography for each domain in m a.s.l.

5 Here, we describe the initial, standard setup, which is used to downscale each of the 56 ensemble members in
 20CR. Model initialization is set to 11 August 2005, while the largest precipitation totals were observed on 22
 August 2005. This is based on the original assumption that an ample spin-up time of more than a week is desirable
 for the inner domains to reach some internal equilibrium. For example, the accumulation of soil moisture can
significantly contribute to enhanced convection and precipitation, and should therefore be considered (Zbinden,
 10 2005; MeteoSwiss, 2006; Bezzola and Hegg, 2007; Cioni and Hohenegger, 2019). The horizontal setup consists of
 four nested domains with grid sizes of 81, 27, 9 and 3 km. The innermost domain covers much of the Alpine bow
 to avoid complex terrain at the boundaries (Figure 1). The vertical setup consists of 60 eta levels (with a top level
 of 50 hPa) to capture fine-scale features of vertical lifting and condensation. The Thompson microphysics scheme
 (Thompson et al. 2008) is used for bulk microphysical parameterization after a first check against the Morrison
 15 scheme (Morrison et al., 2009; not shown). Additionally, we use the Yonsei University (YSU) planetary boundary
 layer (PBL) scheme (Hong et al., 2006) for resolving turbulent fluxes with a complex orography effects' correction
 to the finest domain (Jiménez et al., 2012). The Kain Fritsch cumulus parameterization is used in the larger domains
 (Kain, 2004), and turned off in the innermost domain. Moderate spectral nudging (corresponding to a wavelength
 of about 1500 km) is applied to temperature, wind and geopotential fields above the PBL in the outermost domain
 20 to ensure consistency with the large scale forcing (von Storch et al., 2000). The downscaling output is stored in
 hourly resolution.

Deleted: , and t

3 Downscaling of the 2005 case

3.1 Downscaling with one initial and nine modified setups

25 All 56 ensemble members of 20CR are downscaled with a standard WRF setup (see Sect. 2) in the first step. This
 is done for the 2005 event, as only for this case, simulations can be verified using a state-of-the-art spatial
 reconstruction of precipitation totals, which is CombiPrecip in our case. For the verification, we focus on a control
 area in north-eastern Switzerland (see the small box in Figure A1 in the Appendix), the region where most of the
 precipitation fell, and on precipitation totals over 48 hours starting from 21 August 2005 06 UTC, i.e. the two-day
 period with highest precipitation intensities (MeteoSwiss, 2006). In fact, precipitation was concentrated over north-

eastern Switzerland in the 2005 case, with gradients from the Alpine crests towards the Swiss Plateau. This spatial distribution was also important for the subsequent flooding (Bezzola and Hegg, 2005). Results show a general underestimation of the accumulated precipitation in the control area (Figure 2). Furthermore, CombiPrecip shows high precipitation rates over most of the analyzed period. In contrast, the downscaled ensemble has only two periods of very high precipitation rates (at around 12 and 36 hours after initialization), and obviously, there is too little precipitation between these two simulated peak episodes. In addition, the spread of precipitation totals in the ensemble is large. For instance, the median member (#15) underestimates the mean accumulated precipitation in Northern Switzerland by a factor of four: approx. 20 mm/48 h versus 80mm/48 h in CombiPrecip. Only one member (#24) reaches around 50 mm/48 h, while the lowest member (#23) produces only around 10 mm/48 h.

In all, we cannot be satisfied with these results yet. On the one hand, we aim to investigate particularly flood-inducing features of Vb-cyclones. For this, a certain variability in the ensemble is helpful and necessary. For example, we can find and assess (non-)decisive features by means of opposing ensemble members. On the other hand, we also need to ensure that our downscaling experiment delivers plausible results, especially regarding precipitation intensities and patterns. For this, the deviations from the observations must not become too large in the ensemble. A somewhat smaller spread of the simulated precipitation for the 2005 case would also increase our confidence that the simulation of historical events will produce reasonable and valid results.

In short, we would have expected less underestimation and smaller deviations with this downscaling configuration (cf. Coppola et al., 2018, their figure 5, for estimations of accumulated precipitation over the Alps from a multi-model ensemble). One possible reason for the large variability is the very long spin-up time of 10 days, which may let the simulations run too freely, i.e. independently from the synoptic reanalysis data.

Deleted: , and

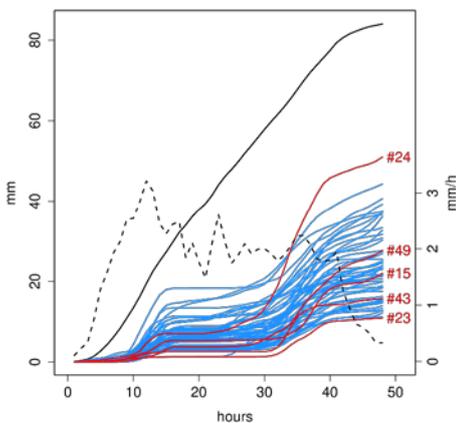


Figure 2: Cumulative totals of hourly precipitation (mm, left y-axis) over 48 hours starting at 2005-08-21 07 UTC (x-axis), as simulated in 56 members (blue lines) that are downscaled from 20CR. Mean values are given for a control area in north-eastern Switzerland. Ensemble members (#23, #43, #15, #49, #24) representing the quartiles of the distribution are highlighted in red. For comparison, cumulative totals in the CombiPrecip dataset is shown (black line), as well as the hourly time series in CombiPrecip (dashed line; mm on the right y-axis).

Accordingly, a second set of experiments is done with decreasing spin-up periods from 10 to 7, 5, 3, and 1 days (see also Table 1 and Table 2 for details). With a 1-day spin-up, both early and late onsets of the most intense precipitation in different regions of the Central Alps are still captured. To save computational costs, these tests are done with a subset of ten members that cover the full range of precipitation variability from the original setup (not shown). In terms of precipitation over northern Switzerland, the best results are achieved with a spin-up time of 1 day (see [Sect. 3.2](#) below for details of the full evaluation).

Deleted: with modified setups, that is,

10 **Table 1: Abbreviations and description of setup experiments.***

NAME	INITIALIZATION	OTHER MODIFICATIONS
sp10	<u>2005-08-11 00 UTC</u>	initial setup, see Sect. 2
sp7	<u>2005-08-14 00 UTC</u>	
sp5	<u>2005-08-17 00 UTC</u>	
sp3	<u>2005-08-19 00 UTC</u>	
sp1	<u>2005-08-20 06 UTC</u>	
sp1_dom	<u>2005-08-20 06 UTC</u>	larger domains, grid ratio=5
sp1_cu	<u>2005-08-20 06 UTC</u>	cu_physics=11,
sp1_cu_mp	<u>2005-08-20 06 UTC</u>	cu_physics=11, mp_physics=95
sp1_cu_nudg	<u>2005-08-20 06 UTC</u>	cu_physics=11, nudging also 2nd domain
sp1_cu_nest	<u>2005-08-20 06 UTC</u>	cu_physics=11, two-way nesting

Deleted: Initialization

*Changes are indicated with respect to the original setup with a 10-day spin-up period (sp10): The abbreviation sp5 indicates a change in the spin-up period to 5 days, mp indicates a change of the microphysics scheme, cu a different cumulus scheme, nudg indicates nudging in two domains, nest indicates two-way nesting, and dom stands for a larger domain. See text for details.

Deleted: culumus

In a third set of experiments, potential enhancements are explored by applying further modifications to this last setup. The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience. Concretely, we test larger domains since the high-resolution domain does not cover the entire extent of the cyclone. For this test, we also increase the grid ratio from three to five. Another test involves changing the cumulus scheme from Kain Fritsch to Multi-scale Kain Fritsch (Kain, 2004). Multi-scale Kain Fritsch contains a grid resolution dependence, which may improve the location and intensity of precipitation in high-resolution simulations (Zheng et al., 2016). Then, the microphysics scheme is changed from Thompson et al. (2008) to Ferrier et al. (1995). The latter has been associated with higher precipitation totals over the USA (Schwartz et al., 2010). Among other differences concerning hydrometeors, ice, graupel and hail are represented in less detail. Note that the Morrison scheme (Morrison et al., 2009) had been discarded after a check with the original setup (Section 2.3). Furthermore, nudging in the two outermost domains is applied to test for the effect of

Deleted: W

Deleted: moderate

keeping the simulation close to initial conditions. Finally, one-way is compared to two-way nesting, allowing also for exchange of information from finer to coarser domains (cf. Bowden et al., 2012).

3.2 Evaluation of ten WRF model setups for precipitation totals

We use three evaluation methods to determine an overall best-performing model setup for downscaling our cases.

5 Again, we focus on the control area in north-eastern Switzerland and on precipitation totals over 24 and 48 hours starting from 20 August 2005 06 UTC. This is because we are particularly interested in finding a setup that would ideally produce correct precipitation maxima in the correct areas on the regional and local scales. CombiPrecip is used as the reference, i.e. observation-based dataset (see above and Sect. 2). Recall that CombiPrecip also includes radar information. Both CombiPrecip and the WRF simulation are bilinearly interpolated to a 6-km horizontal grid for comparability and to reduce single-cell effects.

Deleted: s

Deleted: again

Table 2: Evaluation of WRF setups for dynamical downscaling.*

VER	sp10	sp7	sp7 _mp	sp5	sp3	sp1	sp1 _dm	sp1 _cu	sp1 _cu_nd	sp1 _cu_ns
MAE 24h	14.50	14.87	14.96	14.27	16.90	13.38	13.30	16.14	15.32	16.22
MAE 48h	27.75	28.69	28.70	26.89	25.70	25.44	27.45	24.45	28.79	26.99
BOX 24h	0.90	0.90	0.74	0.86	1.00	0.64	0.70	0.94	1.37	1.05
BOX 48h	0.59	0.58	0.46	0.54	0.67	0.43	0.43	0.62	1.05	0.73
Rank	5	5	4	3	8	1	2	5	10	9

VIS	sp10	sp7	sp7 _mp	sp5	sp3	sp1	sp1 _dm	sp1 _cu	sp1 _cu_nd	sp1 _cu_ns
24 h	10	8	10	13	7	14	13	8	1	4
48 h	10	9	13	12	4	15	13	8	1	4
Rank	5	6	4	3	8	1	2	7	10	9

EMD	sp10	sp7	sp7 _mp	sp5	sp3	sp1	sp1 _dm	sp1 _cu	sp1 _cu_nd	sp1 _cu_ns
24 h	0,231	0,279	0,282	0,214	0,273	0,117	0,135	0,11	0,146	0,13
48 h	0,159	0,171	0,209	0,13	0,116	0,096	0,12	0,099	0,12	0,12
Rank	8	9	10	7	5	1	4	1	5	3

*Spatial verification measures for evaluating the performance of ten different WRF simulation setups (columns; see Table 1 for abbreviations) against observations (CombiPrecip fields) of 24-h and 48-h precipitation totals in a rectangle box over north-eastern Switzerland (see Figure A1), each calculated over a subset of 10 ensemble members. The top section shows two verification (VER) measures, i.e. (i) mean absolute error (MAE) of simulated versus observed precipitation totals (mm/24 h or mm/48 h), and (ii) mean absolute deviation of the simulated box ratios from the observation-based box ratio (specific rows denoted with BOX). For instance, the value of 0.9 (0.59) for sp10 and 24-hour (48-hour) precipitation totals indicates the mean absolute deviation from the observation-based value of 1.99 (2.26) in CombiPrecip. The middle section shows the scores from visual inspection (VIS), and the bottom section considers the spatial distribution of precipitation inside the box (EMD; see text for details). Ranks are added to all sections; ties are set equal.

Deleted: 4

The first evaluation is based on two spatial verification measures, i.e. mean absolute error (MAE) of the simulated versus CombiPrecip precipitation inside the control area, and a metric using box ratios (see section VER in Table 2). The MAE measures the average distance between forecast and observation, and is preferred over RMSE because it is more resistant to outliers, and over correlation coefficients because we are more interested in accuracy than linear association (Joliffe and Stephenson, 2012). Box ratio means the ratio of mean precipitation in the control area, i.e. the small box w.r.t. a larger box that encompasses Switzerland (see Figure A1 in the Appendix). The box ratio indicates how much of the precipitation is simulated in the correct region when compared to CombiPrecip. The box ratio in CombiPrecip for the 1-day (2-day) mean precipitation is 1.99 (2.26). In simple words, the mean observed rainfall was about twice as intense in the control area over north-eastern Switzerland compared to all of Switzerland. For the evaluation, we calculate the mean absolute deviation of the simulated box ratios from the box ratio in CombiPrecip.

The second evaluation is based on visual inspection of the simulated precipitation totals, with a focus on the highest amounts of precipitation, that is, the 4th quartile. Again, the according precipitation totals in CombiPrecip are used as a reference. The inter-subjective judgement by the authors yields two points per downscaled ensemble member for a ‘good’ match, one point for a ‘fair’ match, and zero points for ‘mismatch’ (see section VIS in Table 2). A good match is achieved in case the simulation places the highest quartile of precipitation in the correct regions when compared to the spatial patterns in CombiPrecip.

To contrast the subjective judgements, the third evaluation uses the Earth Mover’s Distance (EMD; Rubner et al. 1998, 2000) as a purely objective metric of similarity. The EMD, sometimes known as the Wasserstein distance, is typically used for pattern recognition in digital image processing, and has as well been applied in atmospheric sciences, e.g., to pollutant concentrations, top-of-atmosphere radiation fluxes, time series of wind maxima, or precipitation and climate indices (Düsterhus and Hense 2012; Baker et al. 2013; Farchi et al. 2016; Düsterhus and Wahl 2018). Intuitively, it measures the cost (mass times distance) of turning one pile of dirt in one area into a second, reference pile with the same overall mass and covering the same area. For our case, this means that the precipitation fields are normalized, and hence, the EMD considers only the relative patterns of precipitation. Specifically, the EMD indicates how well the simulated spatial distribution of the precipitation matches the observed distribution on a 6-km grid inside the control area. EMD was chosen over typical feature-based methods (e.g., the SAL method by Wernli et al., 2008) because it yields one number, involves less subjective choices and thresholds, and emphasizes the relative pattern. Section EMD in Table 2 gives the median distance for each setup and for 24-h and 48-h precipitation totals.

The evaluation shows substantial differences in overall performance and ranking of the ten WRF setups. In the first place, Theil-Sen slope estimates are calculated over sp10, sp7, sp5, sp3, and sp1 for all measures; they are all negative (MAE 24 h: -0.2; MAE 48 h: -0.70; BOX 24 h: -0.03; BOX 48 h: -0.03; VIS 24 h: -0.75; VIS 48 h: -1.13; EMD 24 h: -18.5; EMD 48 h: -17.9). Although the trends are not significant in Mann-Kendall tests (or not clearly attributable, due to the small sample), the negative slopes indicate that performance generally increases with decreasing spin-up time. We infer from this that the Vb cyclone should already be located within the outermost WRF domain at the time of initialization. This allows the WRF model to better track the evolution of the storm system. In the end, the setup with the shortest spin-up is the best ranked. Secondly, the Ferrier and Thompson

microphysics schemes perform similarly well. This might be because 2005 is a summer case with rather high temperatures and hence, the variety in hydrometeors is not very large. Therefore, the Thompson microphysics is selected, which is commonly used in studies on simulating precipitation in complex terrain (e.g., [Pieri et al., 2015](#); [Parodi et al., 2017](#)). Thirdly, further experiments based on a one-day spin-up do not result in better overall performance: neither changing the cumulus scheme, applying two-way nesting, nudging of shorter wavelengths nor using a larger innermost domain result in a better representation of the precipitation over Northern Switzerland during the 2005 event.

Overall, we do not find enhanced simulation performance by modifying an number of WRF configurations in these last experiments: the best ranked WRF setup in all three evaluations is still the standard setup with a one-day spin-up (sp1 in Table 2). This means that our model configuration is robust and expedient for our purposes, and in the following, this setup is used for the simulations of the 2005, 1910 and 1876 cases.

For a comparison with better-resolved input data, the selected setup is contrasted to a downscaling experiment with the same simulation setup and WRF version, but with initial and boundary conditions from the ERA-Interim dataset. The experiment yields an EMD value of 0.2, and 48-hour precipitation totals of up to 350 mm (Figure A2 in the Appendix). While the EMD value in this simulation is in the range of the best downscaled 20CR members, precipitation is much higher. We infer from this that downscaling from 20CR reproduces the relative distribution of precipitation equally well, while the higher intensities may be attributed to the better spatial resolution of moisture variables in the ERA-Interim reanalysis. In addition, specific humidity at 1000 hPa is higher in 20CR over the Alps, but higher over Eastern Europe in ERA-Interim (not shown). Hence, the advection of moist air to the central Alps is arguably stronger in ERA-Interim.

4 Analyses and simulations of the 2005, 1910 and 1876 cases

4.1 Precipitation, cyclone fields and tracks in the 20CR ensemble

In this section, we analyze how well the three cases are represented in the 20CR members on a synoptic scale (Figures 3 and 4). For this, we compare data from 20CR to data from ERA-Interim (only available for 2005) and CERA-20C (available for 2005 and 1910). Specifically, we analyze the large-scale patterns of precipitation totals during the most intense phases (21 - 22 August 2005, 13 - 14 June 1910 and 11- 12 June 1876). Furthermore, we investigate the synoptic setting and intensity of the associated cyclone in the ensemble members.

For the analyses, we use both sea level pressure (SLP) and mid-tropospheric pressure fields. SLP fields inform about the quality of the assimilation process in 20CR, and the isobaric pressure fields (at 500 hPa here) tell about the derivation of upper-air variables from the SLP information in 20CR. Combining SLP and isobaric levels has been found useful for cyclone tracking (Hofstätter and Blöschl, 2019). The cyclone tracks shown in Figure 4a, b, and c are reconstructed as follows: In a first step, absolute minima of the 500-hPa geopotential height are inventoried. Then, the cyclone centers (here, the absolute minima of geopotential height) closest to Corsica on 22 August 2005, 13 June 1910 and 11 June 1876 are selected. In a third step, cyclone tracks are reconstructed every six hours backward and forward in time by selecting the closest cyclone position and starting from the three selected cyclones. The tracks are terminated if the cyclone position jumps over more than approximately 1200 km in six hours. For example, an absolute minima of the 500 hPa geopotential height (here, a cyclone center) exists in 36 of

Deleted: I

the 56 ensemble members over southern England on August 20, 00UTC (Figure 4a). One day later, all 56 members contain a cyclone center over southern France. For comparison to the mid-tropospheric level, cyclone fields (Figure 4d, e, and f) and tracks (Figure 4g, h, i) are also calculated for SLP according to Wernli and Schwierz (2006; see also Welker and Martius, 2015). The algorithm detects cyclone fields in terms of a finite area around a regional SLP minimum, that is, by a closed SLP contour line. The regional SLP minima for each cyclone life cycle are stored as cyclone tracks, and the presence or absence of a cyclone is represented in a binary field for each grid point and time step.

Inferring from Figure A1 in the Appendix, as well as from analyses of supra-national rain gauge measurements (Frei 2006; Stucki et al. 2012) or model simulations (Langhans et al. 2011, for the 2005 case), most precipitation is expected to accumulate over north-eastern Switzerland, and to reach well into Austria and south-eastern Germany along the Alpine bow during these three cases. A second area of heavy precipitation is expected to stretch from the south-eastern Alps into the Dinarides mountains. From these similarities, it can be assumed that also the synoptic fields of precipitation look similar for all three cases. Similarity is also presumed regarding the location and intensity of the rain-associated cyclones and cyclone tracks, as they strongly determine whether heavy precipitation is advected to the expected regions along the north-eastern Swiss Alps.

For 2005, both ERA-Interim and CERA-20C produce a center of heavy precipitation (up to 50 mm /24 h) in the expected region (Figure 3a and b), and tongues of heavy precipitation reach east and southeast along the Alpine bow and along the Adriatic coast. In comparison, 20CR shows only one coherent, but larger center of precipitation that is shifted towards the south-east and has lower intensities, while representative for larger grid boxes (up to 35 mm/24 h; Figure 3c). Variability in precipitation totals among the 56 members of 20CR is larger (interquartile range of approx. 10 to 15 mm/24 h) than among the ten members of CERA-20C (interquartile range of up to approx. 10 mm/24 h; Figure 3d and e). The cyclone fields of the 2005 case, as well as the associated cyclone tracks south of the Alps, are well defined in the 20CR ensemble: shifts on only a couple of grid cells occur (Figure 4a, d and g). For instance, 46 members show a cyclone track at 10° E, 44° N for 21 August 2005 18 UTC. Differences to the cyclone fields in ERA-Interim are also mostly within the range of the 20CR members.

For 1910, the centers and tongues of heavy precipitation have a similar location to 2005 in CERA-~~20C~~; although intensities are lower (up to 30 mm/24 h). 20CR also shows similar centers of heavy precipitation, while intensities in 20CR are clearly lower (Figure 3f and g). In contrast, the variability among the members is higher in the CERA-20C dataset (up to approx. 10 – 15 mm/24 h compared to below 10 mm/24 h in 20CR; Figure 3i and j). Compared to the 2005 case, the range of calculated cyclone tracks and cyclone fields for 1910 becomes larger in 20CR, and encompasses three or sometimes even more grid points (Figure 4b, e and h). A number of 56 cyclone tracks are found at two grid points (10° E, 44° N and 12° E 44° N) for 13 June 1910 18 UTC.

Deleted: storm track

Deleted: storm track

Deleted: storm track

Deleted: 20C,

Deleted: storm track

Deleted: storm track

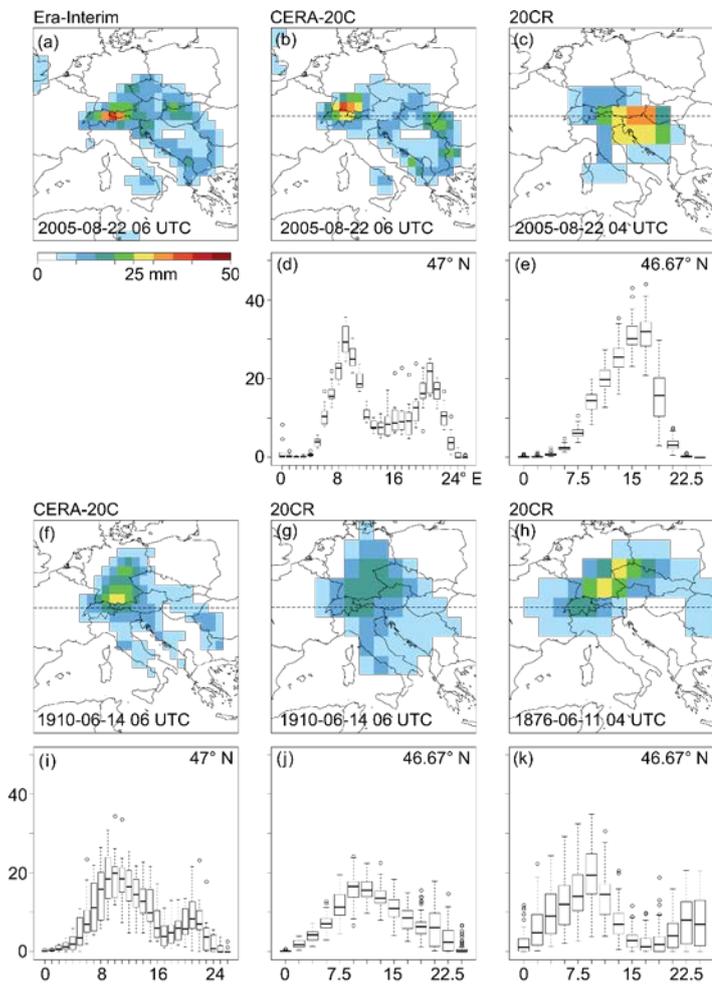


Figure 3: Daily precipitation totals (mm/24 h; color shade) as calculated for 22 August 2005 (a, b, and c), 14 June 1910 (f and g) and 11 June 1876 (h) from ERA-Interim (a), CERA-20CR (b and f) and 20CR (c, g, and h). Slightly differing time steps are due to differing temporal resolutions of the reanalyses products. Boxplots below the map panels show, respectively and where applicable, the variability of daily precipitation totals of all ensemble members in grid boxes along cross sections at 47° N (CERA-20CR; d and e) or 47.67° N (20CR; e, j, and k). The horizontal lines in (b, c, f, g, and h) delineate these cross sections.

For 1876, only 20CR is available (Figure 3h and k) to assess the representation of synoptic precipitation fields. Compared to 1910 and 2005, the center of heavy precipitation is located more to the north-east of Switzerland, over south-eastern Germany. Intensities are higher than for 1910 and lower than for 2005. Whereas the cyclones pass across the Ligurian Sea and Northern Italy in the 2005 and 1910 cases, the bulk of the ensemble takes a more northerly path in the 1876 case (Figure 4c, f and i). A number of 25 members show a cyclone track at two grid

points just south of Switzerland (10° E, 46° N and 12° E 46° N). And while a small part of the members tracks towards the north(-east) on 12 and 13 June 1876, the rest shows a south-eastward propagation along the Adriatic Sea.

Overall, the analyses at synoptic scales (Figures 3 and 4) show that differences among the 20CR members are substantially smaller over the region of interest (Southern and Central Europe) than over other regions of the North Atlantic / European sector (Figure 4 d, e and f); this corresponds to the relatively high density of assimilated stations over Central Europe (not shown; see Compo et al. 2015). The main fields of precipitation are approximately co-located in all three datasets (Figure 3). Variability in the 20CR ensemble is comparable to CERA-20C for the 2005 and 1910 cases. 20CR shows overall lumpier spatial patterns of heavy precipitation and lower values due to the coarser horizontal grid, and a potential displacement of the precipitation field for the 1876 case. As expected, the uncertainty, in terms of disagreement between the 20CR ensemble members, becomes increasingly larger when going back in time. For instance, the cyclone fields and cyclone tracks over the Alpine area are only a little less well defined for 1910 compared to 2005, but much less for 1876. Among others, this is shown by the number of co-located cyclone tracks (in terms of pressure minima) in the 20CR ensemble. The algorithm detects 56 co-located cyclone tracks at a grid point over northern Italy on 21 August 2005 18 UTC (Figure 4 g). For the time step 13 June 1910 18 UTC (Figure 4 h), a number of 56 cyclone tracks are detected at two adjacent grid points, whereas only 17 co-locations are found for 10 June 1876 18 UTC (Figure 4 i). From these analyses, we infer a very good to satisfactory positioning of cyclone tracks and cyclone fields in 20CR for the 2005 and 1910 cases, but not necessarily for 1876. This means that the boundary and initial conditions appear to be captured in 20CR for the 2005 and 1910 cases, while 1876 shows two or even more potential developments of the cyclone.

Deleted: u

Deleted: storm track

Deleted: storm track

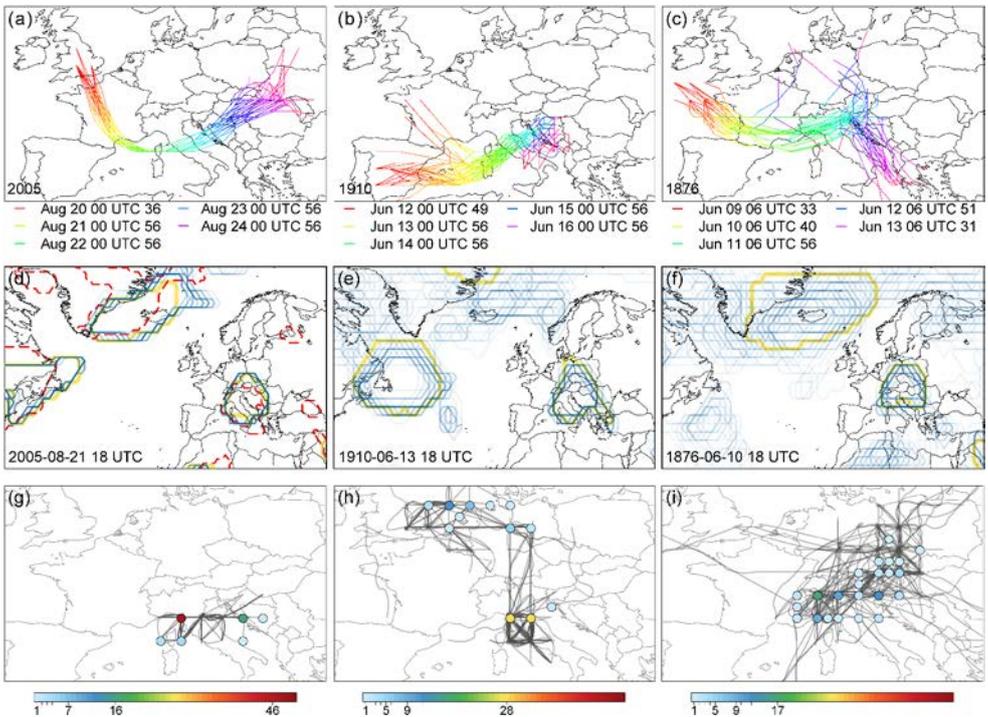


Figure 4: Synoptic situations as depicted in the 20CR ensemble for the 2005, 1910 and 1876 cases. (a, b, c) Cyclone tracks for each ensemble member of 20CR at a mid-tropospheric, i.e. the 500-hPa level. The color of the lines corresponds to the time steps indicated below the panels. In addition, the number indicates for how many of the 56 members a cyclone position could be reconstructed at a certain time step. (d, e, f) Objectively identified cyclone fields (calculated for sea level pressure; using Wernli and Schwerz, 2006; see also Welker and Martius, 2015) in the 20CR ensemble (blue contours) and 20CR mean (yellow contours) at the time steps of (d) 21 August 2005 18 UTC, (e) 13 June 1910 18 UTC, and (f) 10 June 1876 18 UTC. The color scheme, ranging from light blue to dark blue, indicates in how many of the 56 ensemble members a cyclone is detected in the respective grid cell. The red broken lines in the 2005 panel indicate the cyclone field as calculated from ERA-Interim. (g,h,i) Cyclone centers identified for sea level pressure at (g) 21 August 2005 18 UTC, (h) 13 June 1910 18 UTC, and (i) 10 June 1876 18 UTC, the same time steps as in (d, e, f). Colored dots and the tick marks in the color key indicate the number of cyclone tracks located at a specific grid point at the respective time steps. Greyscaled lines mark the cyclone tracks over the period of 48 h before to 48 h past the respective time steps. Darker (lighter) grey shades indicate more (less) cyclone tracks along a certain path.

Deleted: at time steps when they are similarly located over the Alps.

Deleted: As in middle panels, but for cyclone tracks (grey lines). Colored dots indicate the number of cyclone tracks located at a specific grid point at the respective time steps in d, e, and f

4.2 Precipitation, cyclone tracks and moisture transport along the downscaling steps

In this section, we examine how well the three Vb cases are represented when downscaling the global information from 20CR to a 3-km horizontal grid using the WRF regional model.

Deleted: after

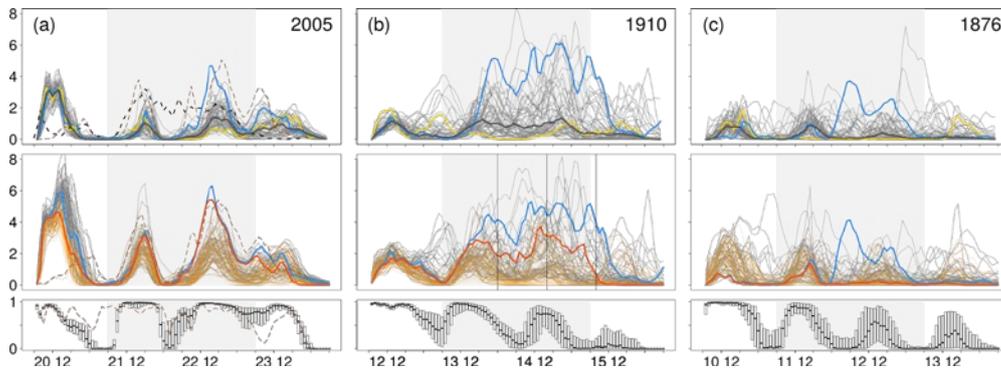


Figure 5: Time series of simulated precipitation over northern Switzerland for (a) 20 August 2005 06 UTC to 24 August 2005 06 UTC, (b) 12 June 1910 06 UTC to 16 June 1910 06 UTC, and (c) 10 June 1876 06 UTC to 14 June 1876 06 UTC. Dark grey lines indicate mean hourly precipitation (mm/h) within the control area over northern Switzerland for the 56 downscaled ensemble members, as simulated for the 3-km domain (upper panels) and the 9-km domain (middle panels). In the upper panels, the thick black (yellow, blue) solid lines mark the median (minimum-precipitation, high-precipitation) members, as selected for Figs. 6 to 9 and 11. In the middle panels, thin orange lines show the contribution of the convection parametrization to the total precipitation in the 9-km domain. The thick dark orange lines indicate convective precipitation for the selected high-precipitation members. The bottom panels show the proportion of convective precipitation with respect to total precipitation in the 9-km domain. In (a), the dashed black line shows the corresponding hourly precipitation calculated from CombiPrecip, and the dashed brown lines show the precipitation from the ERAI-WRF experiment with 4 domains; the convective contribution is shown with points in the middle panel. Black vertical lines in (b) indicate the instances in time selected for Figure 11. Grey shadings mark the most intense 48-h periods of precipitation according to Figure A1.

In a first analysis, we address the temporal variability of the simulated precipitation. Figure 5 shows time series of aggregated precipitation in the control area over northern Switzerland. In the 2005 case (Figure 5a), two distinct peaks occur on 21 and 22 August 2005 around 18 UTC. This evolution is very much in line with Hohenegger et al. (2008; their figure 8); even the increase during the second peak episode is very similar. As already seen with the standard downscaling configuration (Figure 2), intensities are mostly underestimated when compared to CombiPrecip. Arguably, too little precipitation is produced between the simulated peak episodes, although more intermittent precipitation was also observed in a (smaller) area over eastern Switzerland (Hohenegger et al., 2008), and CombiPrecip may as well be imperfect over the control area. Two high-precipitation episodes are also simulated for 1910 and 1876 (Figure 5b and c), although variability among the members increases regarding the timing and intensities of precipitation. For instance, the ensemble interquartile range is smallest for 2005 (around 1.5 mm/h in the peak episode; note that this is smaller than in Hohenegger et al., 2008) and becomes larger for the earlier cases (around 2 mm/h on 14 June 1910 18 UTC, and around 2.5 mm/h on 11 June 1876 18 UTC). For all cases, the ensemble shows most precipitation peaks in the afternoon. This would be in agreement with an enhancing effect by (embedded) convection. To investigate this effect, we turn to the second finest domain with a 9-km horizontal grid. Whereas convection is explicitly simulated for the 3-km domain, it is parameterized in the 9-km domain, resulting in WRF model variables of non-convective (RAINNC) versus convective (RAINNC) contributions to the total precipitation (the shallow convection variable RAINSH is turned off). For the 2005 case

Deleted: red

Deleted: T

Deleted: , and the dark blue line in (c) marks the selected high-precipitation member for Figs. 6 and 9

Deleted: , including a potentially enhancing effect by embedded convection.

Deleted: , and it also agrees with CombiPrecip, except for potentially producing too little precipitation

and in the 9-km domain, convective precipitation is the largely dominant process during the afternoons (Figure 5a), reaching nearly 100 percent in all the members and on all days of the event. The proportion of convective precipitation is smaller during other times of the day and varies more in the ensemble. The same pattern is found for the 1910 and 1876 cases, although the proportion of convective precipitation is mostly smaller and variability in the ensemble is higher (Figures 5b and c).

Peaks of precipitation are also simulated during the initialization period of each case, which is in line with observations (not shown; cf. Stucki et al., 2012). However, the peak on 20 August 2005 is too prominent compared to CombiPrecip (Figure 5a). The convection-driven peak is simulated for all members including the minimum-precipitation member, while precipitation intensity appears more realistic in the WRF-ERA1 simulation. One explanation is that in the 9-km domain, the coarser interpolation from 20CR input data results in too high temperatures over the Alps for this specific case. In turn, this produces excess water vapor in the first hours of the 20CR-WRF simulation (not shown).

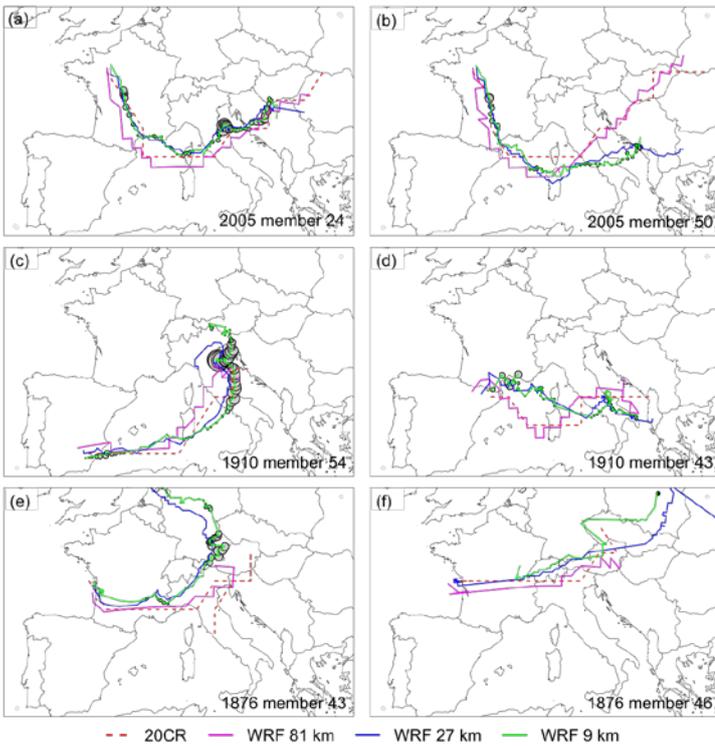


Figure 6: Cyclone tracks, calculated for 500 hPa levels, for the cases of 2005 (top), 1910 (middle) and 1876 (bottom) and for members producing maximum (left) and minimum (right) precipitation totals over northern Switzerland. The cyclone tracks in 20CR and WRF at 81-, 27- and 9-km grid sizes are shown in red, magenta, blue and green, respectively. Filled circles indicate the mean hourly precipitation over northern Switzerland at the time step when the cyclone was centered at the respective grid point in the WRF 3-km domain. The circle diameters grow linearly with the precipitation amounts; the largest circle (in c) represents 6.6 mm/h.

Deleted: storm track

Next, we examine whether the found features and variabilities in the ensemble reflect differences that are already present in the 20CR members (Figure 3) or if they appear along the downscaling steps. Concretely, we search for flow features that help to systematically distinguish members with low or high precipitation simulated in the correct region (i.e. in the control area over north-eastern Switzerland). For this, we compare the simulated 48-hour precipitation totals with RrecabsD. We use the ratio of precipitation in the simulation versus the reconstructions, and the EMD between simulation and reconstruction to assess the similarity of the spatial distribution of precipitation.

Deleted: se

For illustration, the panels in Figure A2 in the Appendix show (i) a maximum member in terms of simulated precipitation (a near-maximum member is chosen for the 1876 case because the maximum member does not show plausible patterns, cf. Fig. 5c), and (ii) a minimum member in terms of lowest precipitation totals in the control area. Indeed, the two contrasting members are exemplary for the large variability of the simulation results. Throughout the ensemble, we find members that largely misestimate the precipitation totals (the range is around 20 to 160 percent for the 2005 and 1910 cases, and 5 to 70 percent for the 1876 case, not shown), while others produce precipitation at the wrong place, but also a number of members that produce quite accurate spatial patterns and precipitation totals compared to observations and the RrecabsD reconstruction.

Figure 6 delineates the corresponding evolution of the cyclone tracks in selected ensemble members that yield maximum (a) or minimum (b) precipitation for the 2005 case. The cyclone track for the maximum-precipitation member follows closely the original cyclone track in 20CR in each downscaling step. During the peak episode, the cyclone center is located just above the Adriatic coast of northern Italy. Moreover, the multiple circles at the same location (Figure 6a) indicate quasi-stationarity of the cyclone. In contrast, the minimum-precipitation member has a cyclone track in the 27-km domain that clearly departs from the original 20CR cyclone track: Instead of recurving to the north over Italy, it keeps propagating eastward. The high-resolution domains (9-km; and 3-km, not shown) then represent refinements of these patterns without significant changes. In the 1910 case, the cyclone tracks for the maximum-precipitation member (Figure 6c) also show the vicinity to 20CR in all downscaling steps, the recurving to the north and the same location during the stalling, i.e. peak episode. In contrast, the minimum-precipitation member shows a more southerly and eastward track after it reaches Italy (Figure 6d). That is, the tracks of 20CR and the coarsest downscaling step never turn towards the north, thus, making it more difficult to bring precipitation towards the target area. In contrast to the 2005 and 1910 cases, the algorithm has difficulties to detect clear cyclone tracks along the downscaling steps for the 1876 case (Figure 6e and f). In addition, the found tracks run just south of or even across Switzerland, hence on a much more northerly path than for the other two cases. Such tracks do no longer represent a typical Vb trajectory.

Deleted: Remarkably, t

Deleted: storm track

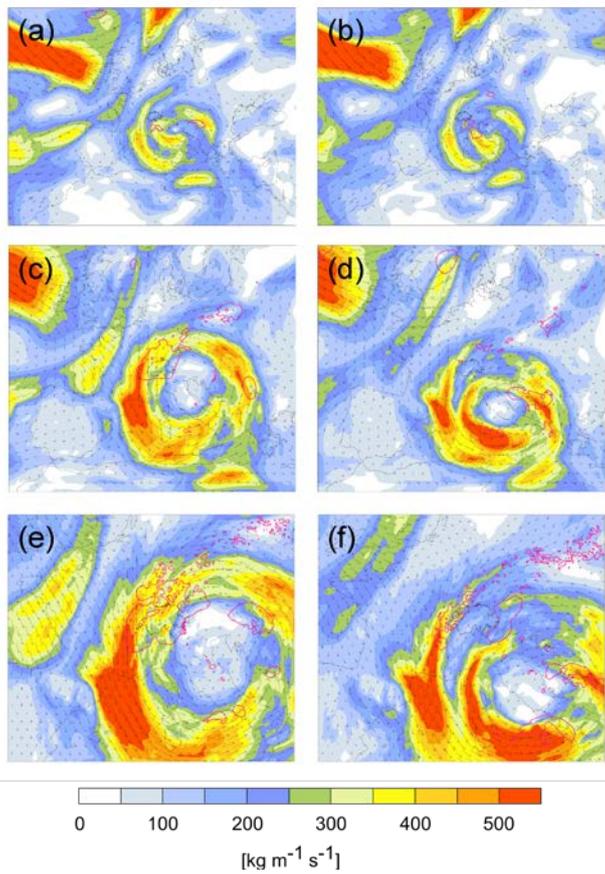


Figure 7: IVT ($\text{kg m}^{-1} \text{s}^{-1}$) on 22 August 2005 18 UTC for in the WRF 81-km (a, b), 27-km (c, d), and 9-km (e, f) domains and for a member producing maximum (#24; a, c, e) and minimum (#50; b, d, f) precipitation totals. Smoothed hourly precipitation is shown with pink contours of 0.2 and 1mm/h.

5 The panels in Figure 7 (for the 2005 case), Figure 8 (for 1910), and Figure 9 (for 1876) illustrate the link between the exact position of the cyclone track and the moisture transport in terms of IVT, showing variations of synoptic to meso-scale features along the downscaling steps for the simulated peak times. In the 2005 case, the IVT patterns of the vortex located south-east of Switzerland are very similar among the two contrasting members in 20CR (not shown), and small differences appear in the 81-km domain (Figure 7a and b). This changes in the 27- and 9-km domain (Figure 7c and d): In line with the cyclone track in Figure 6, the IVT vortex of the minimum-precipitation member is clearly shifted towards the south. The location of the cyclone center strongly determines the intensity of the moisture flux and precipitation over the Alps: In the maximum-precipitation member in Figure 7e, moisture is transported all the way around the Alpine chain. For the minimum-precipitation member in Figure 7f however, the circle of intense IVT around the cyclone center is shifted to the south and is moreover partly interrupted over the northern Alps / Switzerland, arguably because a lot of the moisture already precipitates upstream, i.e. over the Dinarides mountains in the Balkans.

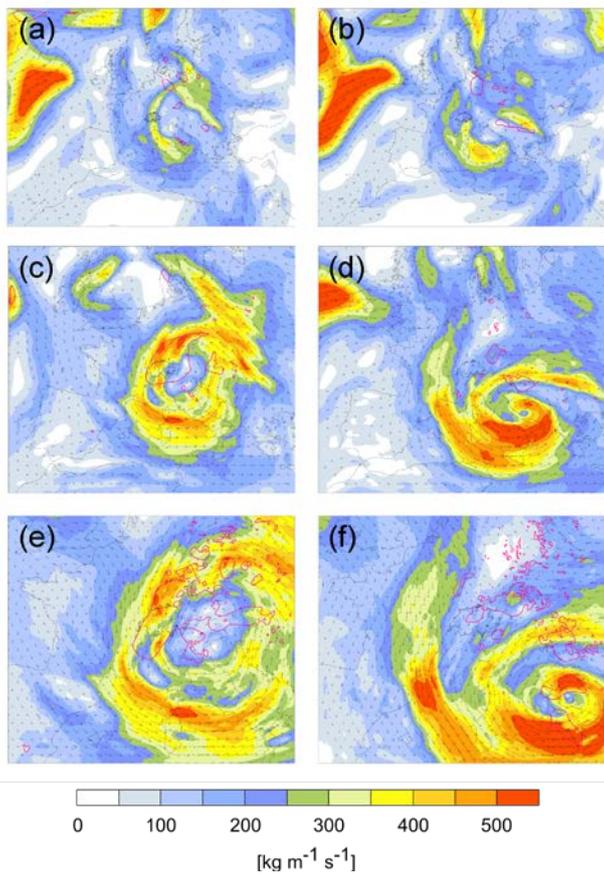
Deleted: spatial

Deleted: IVT

In the 1910 case (Figure 8), the corresponding patterns of the IVT vortices are very similar to the 2005 case, although showing lower intensities over the Alps in 20CR (not shown) and the 81-km domain. The maximum-precipitation member shows northerly winds and intense moisture flux / precipitation over Switzerland, associated with the cyclonic IVT pattern surrounding the cyclone center (Figure 8c and e). In the minimum-precipitation member however, the cyclone center and associated cyclonic IVT pattern are shifted south-eastwards, such that the intense IVT misses the Central Alps (Figure 8d and f).

5 In the 1876 case (Figure 9), the maximum-precipitation member features hardly any structures of a vortex in 20CR (not shown) and in the 81-km WRF domain, and the IVT vortex appears broken and misplaced at higher resolutions. A center of the cyclone is located over northern Germany and induces intense moisture transport in a westerly flow. Hence, some areas in northern Switzerland receive intense precipitation, although it is not anymore associated with a classical Vb cyclone. In the minimum-precipitation member, the cyclone center is also located to the north, but also to the west of Switzerland (Figure 9d and f). Again, Switzerland is on the south side of the vortex, within southwesterly moisture flux reaching into western Switzerland only, and missing most of the Central Alps.

Deleted: es



15 **Figure 8: As in Figure 7, but for 14 June 1910 18 UTC, and for different maximum (#54) and minimum (#43) members.**

Furthermore, Figure 10 demonstrates that indeed, precipitation intensity over northern Switzerland was closely related to the intensity and direction of the moisture transport towards the Alps during the heavy-precipitation phase of the three cases. In the 2005 and 1910 cases (Figure 10a and b), IVT intensities of more than $200 \text{ kg m}^{-1} \text{ s}^{-1}$ are advected from Northwest to Northeast (i.e. directed towards the northern side of the Alps). This is concurrent with average precipitation rates of up to 8 mm/h , whereas precipitation rates become clearly lower with decreasing IVT and with other inflow angles, as seen in the 1876 case (Figure 10c). Similar results were found by Froidevaux and Martius (2016).

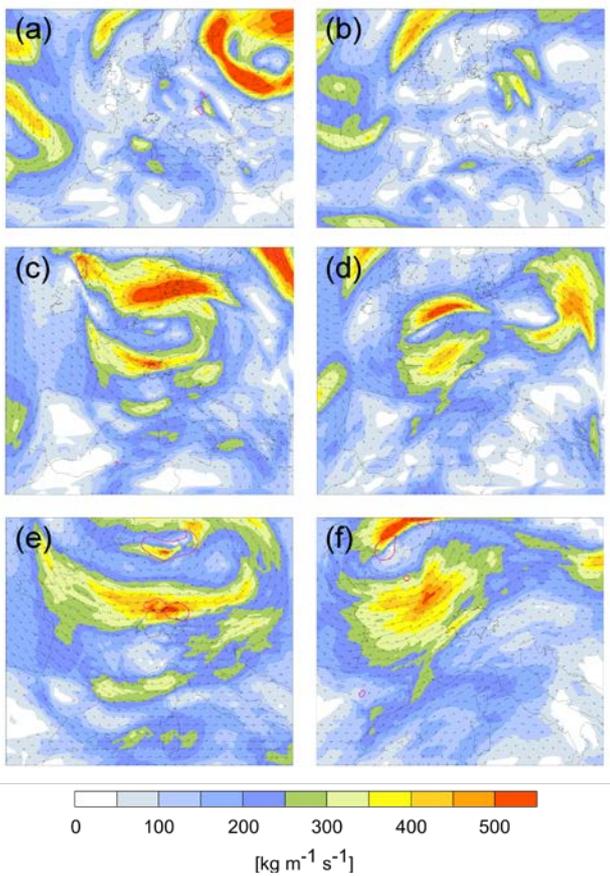


Figure 9: As in Figure 7, but for 13 June 1876 00 UTC, and for different maximum (#43) and minimum (#46) members.

In summary, the 2005 and 1910 cases behave similarly along the downscaling steps in the simulation, whereas the 1876 case deviates in a range of aspects. In the 2005 and 1910 cases, Vb cyclones exist for all members in 20CR (Figure 4, Figure 6), and the cyclone centers cross the WRF domains at 81-, 27- and 9-km grid sizes. In the 3-km domain, the trajectory of the cyclone centers typically passes southwards of the domain, and a clear cyclonic

circulation is systematically present, corresponding to the position of the cyclone in the 9-km domain (Figures 7, 8 and 9). This can be expected, as the cyclone is larger than the two smallest domains, which hampers shifting of cyclone centers. Moreover, we find that the cyclones with centers that stall over a specific location of Northern Italy / the Adriatic Sea are associated with more intense precipitation over north-eastern Switzerland. The maximum-precipitation simulation for 1910 produces even larger totals than observed (Figure 5, see also Figures 8, 10 and A2 in the Appendix). This may indicate that under slightly different atmospheric conditions, e.g. with longer stalling of the cyclone at a particularly unlucky location, the real cases could have had even worse impacts.

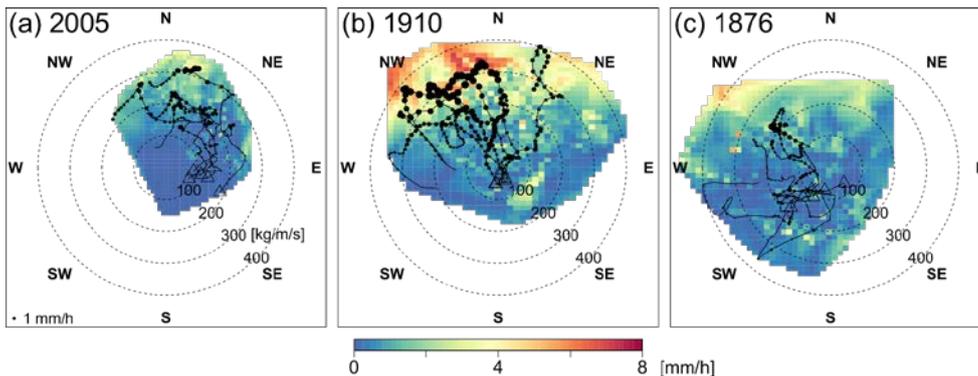


Figure 10: Hourly precipitation (mm/h; color shade) as a function of IVT ($\text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1}$; dashed circles) over north-eastern Switzerland during 48-h periods starting at (a) 21 August 2005 06 UTC, (b) 13 June 1910 06 UTC, and (c) 11 June 1876 06 UTC. Both precipitation and IVT are averages over the control area in north-eastern Switzerland in the 3-km WRF domain. The 480 black dots represent 48 different time steps for a subset of ten members (see Section 3.1). The size of the dots represents precipitation intensity; the location of the dots on the radial diagram represents IVT: the azimuth represents the direction of IVT and the radial component its intensity. All time steps of a same member are connected by a line, with the last time step marked by a triangle. The corresponding two-dimensional interpolation of the precipitation intensity is shown in color. The color maps hence represent the mean precipitation intensity as a function of IVT intensity and direction.

In contrast, members with more southerly tracks do not produce heavy precipitation in this region. With a displacement to the south, the moisture transport does no longer provide a northern inflow towards the Alps, which then inhibits the orographic lifting along the Alps (Figures 7, 8 and 9; and Figure A2 in the Appendix). Hence, the moisture removal from the atmosphere is limited, leading to less precipitation in general and especially in the target area.

Too northerly tracks are not helpful in generating plausible precipitation patterns either. The 1876 case shows that the vortex structure is destroyed as soon as the cyclone centers are located too close to the Alps. Instead of intense moisture advection from a sector North, advection on the north side of the Alps shifts to a sector West or even South. This behavior of the model can be explained by the interaction of the Alpine orography with atmospheric circulation. Indeed, Vb cyclone trajectories are typically initiated by deepening upper-level troughs, which finally cut off from the westerly flow when passing over the Alps (e.g. Awan and Formayer, 2017). The interaction of upper-level troughs with the Alpine orography have been described in detail (Buzzi and Tibaldi 1978; Aebischer

Deleted: Remarkably, t

Deleted: particularly

Deleted: the right place

Deleted: , which affects production of potential vorticity (PV) in the downscaling process. In fact, the Alps have three different impacts on a PV streamer coming from the west, and they all tend to generate lee cyclogenesis over the Gulf of Lions / the Ligurian Sea

and Schär 1998; Kljun et al. 2001); the underlying processes include flow splitting and lee cyclogenesis, with further amplifications of the cyclone formation by frontal retardation and latent heat release due to orographic lifting. The combination of these processes implies that the cyclones are formed on the lee of the right side of the Alps, typically over the Ligurian Sea. In 20CR, however, the Alpine orography is very coarse, smoothed and reaches only about 1000 m a.s.l. (cf. Stucki et al., 2012). Hence, the influence of the Alps on the large-scale flow is limited in 20CR. Given also that the 1876 case is least confined by pressure observations, this allows untypical cyclone tracks in many 20CR members. Once accounting for a more and more realistic orography throughout the downscaling steps with WRF, the high-resolution runs may thus end up in a compromise simulation - driven both by the WRF model physics and by the 20CR input flow. In other terms, the large-scale flow forced from 20CR might not be compatible with the orography of the high-resolution domains.

To conclude the analyses, Figure 11 illustrates how a certain combination of cyclone tracks and cyclonic moisture flux translates into a specific weather situation at the surface. For this, we select the maximum-precipitation member (#54) for the 1910 case and show an early, mid-, and late instance of the heavy-precipitation period (cf. Figure 5), and we compare it with findings regarding the 2005 case.

For 14 June 1910 00 UTC, the 3-km downscaling shows patches of heavy precipitation along the Alps and Alpine foothills of northern Switzerland (Figure 11a). Many of them appear in banded structures, similar to findings from the 2005 case (Bezzola and Hegg, 2007; Langhans et al., 2011). The structures are generally oriented parallel to the Alpine bow and reach from south-eastern Germany into central Switzerland. Surface winds in the control area come from sector North to North-Northwest and weaken upwind of the Alpine barrier. Concurrent areas of convection appear in the 9-km simulation (Figure 11b), and the pressure gradient along the Alpine rim shows the Vb low-pressure system over the Adriatic Sea. At the same time, the IVT vortex just starts to show intense moisture transport from the northeast towards Switzerland (not shown, cf. Figure 7 for a later instance). In all, this indicates persistent airflow upon orography and orographic lifting, as documented for the 2005 case (Zbinden, 2005; Bezzola and Hegg, 2007) and visible in CombiPrecip and wind observations for the evening of 21 August 2005 (not shown). On 14 June 1910 16 UTC, the center of the low-pressure system is located just south of Switzerland (see red dot in Figure 11d). Accordingly, the northerly cross-Alpine flow is substantially stronger, and heavy precipitation becomes most intense along the northern Alpine ranges (Figure 11c). Again, this shift of the heavy precipitation into the Alpine ranges with enhanced northerly flow is in line with analyses of the 2005 case (Zbinden, 2005; MeteoSwiss, 2006; Bezzola and Hegg, 2007), and with CombiPrecip and wind observations for 22 August 2005 (not shown). The 9-km simulation shows the associated areas with intense convection reaching into Switzerland from the northeast (Figure 11d). At this stage, the cyclonic flow forms a distinct arc that stretches from the eastern to the central Alps.

On 15 June 1910 08 UTC, the SLP minimum has crossed the Alps to the northeast of Switzerland (Figure 11f). In the 3-km simulation, heavy precipitation occurs in an area of southwesterly winds along the Alps, the Swiss Plateau, and towards the east, while the flow remains north-westerly at higher elevations and towards the (north)west (Figure 11e). Concurrently, the SLP fields in the 9-km simulation indicate higher pressure from the west, before precipitation intensifies along the eastern Alps, while it finally eases over Switzerland. This is again analog to a late stage of the 2005 case (MeteoSwiss, 2006; Bezzola and Hegg, 2007; Figure 11f), e.g. visible in CombiPrecip and wind observations for around midnight on 22 August 2005 (not shown).

Deleted:). The first impact is flow splitting and generation of PV banners, both at the small scale (isolated mountain peaks) and meso-scale (on the sides of the main alpine ridge). On the right side of the Alpine ridge (southwestern French Alps), the PV banner is positive, creating a low-level positive anomaly that, in turn, combines with the upper-level PV-streamer to create a cut-off. The second impact is frontal retardation, which increases temperature contrasts and creates an upper-level low-pressure center over the cold-air outbreak. This again amplifies the cut-off. The third impact is latent heat release by orographic lifting. This further increases the horizontal temperature contrasts and pressure gradients, and therefore further amplifies cyclogenesis

In the end, all simulated instances are associated with heavy precipitation over northern Switzerland, with slight changes in the inducing weather dynamics. In the first instance, banded convection and orographic lifting both contribute to intense precipitation. The second instance, corresponding to the peak precipitation, is associated to stronger northerly winds and a distinct cyclonic moisture flux around and over the Alps. The last instance is linked to a shift towards westerly advection and increasing pressure. While the early stage of the SLP cyclone track calculated for member 54 is not very typical (no cyclogenesis in the classic Genoa region, see Figure 6c), the surface analyses of the two innermost domains during the heavy precipitation phases show that the simulation produces realistic near-surface weather dynamics at local scales, and they can clearly be related to the circulation and features of a Vb cyclone.

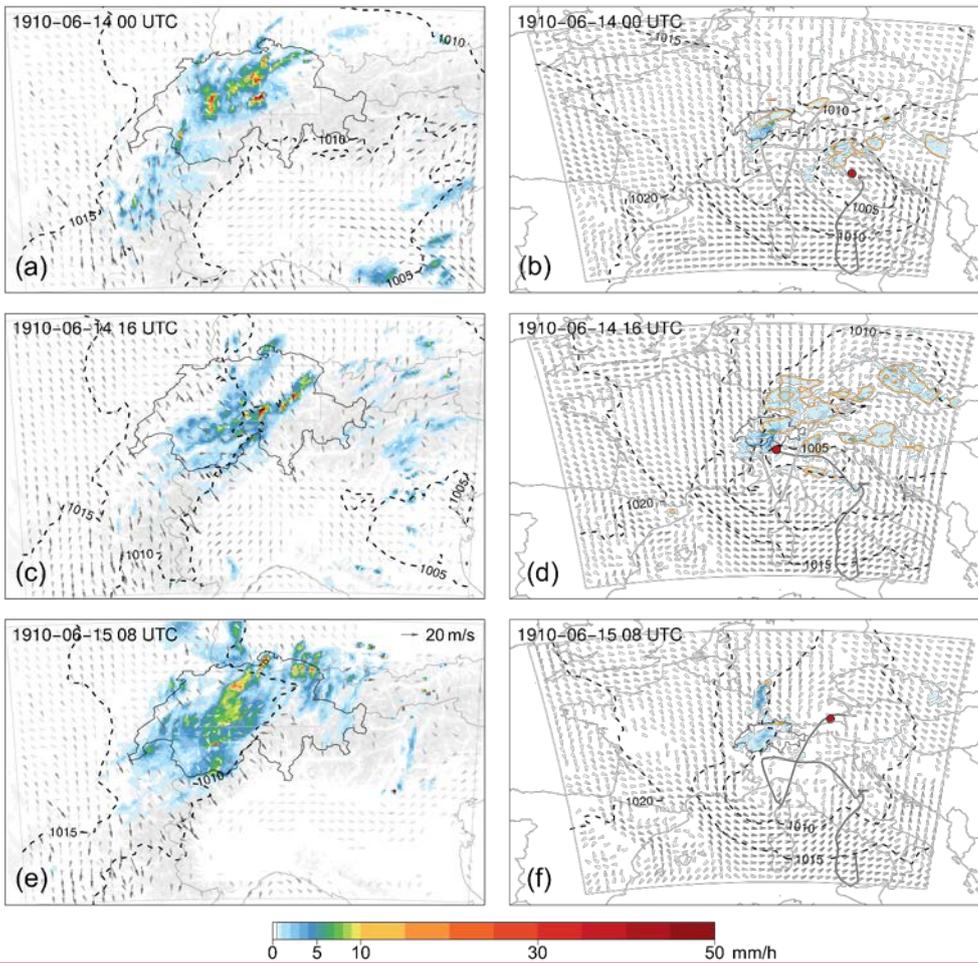


Figure 11: The 1910 event (as shown in the downscaled member #54) in the 3km WRF domain (left) and corresponding 9km WRF domain (right) and for three instances in time corresponding to heavy precipitation over northern Switzerland, that is at (a, b) 14 June 1910 00 UTC (c, d) 14 June 1910 16 UTC, and (e, f) 15 June 1910 08 UTC. Shown are hourly precipitation (mm/h; color shade), 10-m wind field (m/s, light grey vectors darken with increasing velocity).

see reference vector of 20 m/s in panel e) and SLP (dashed contours, in 5-hPa increments). In (b, d, and f), orange contours show the contribution of the convection parameterization to the total precipitation (contours at 2 and 5 mm/h). The dark (light) grey lines in the 9-km domain indicate the smoothed cyclone track starting on 13 June 1910 00 UTC and until the shown instance. The red dot marks the cyclone center (i.e. pressure minimum) at the respective instance in time.

5 Summary and conclusions

In this study, we have assessed the potential of dynamical downscaling from 20CR input to 3-km grid sizes for three well-known Vb cyclones that led to heavy precipitation and flooding in (north-eastern) Switzerland in August 2005, June 1910 and June 1876. In particular, we have analyzed the sensitivity of the produced precipitation totals in a control area in north-eastern Switzerland to (i) the setup of the regional weather model and to (ii) the representation of moisture flux in the 20CR ensemble and along the downscaling steps.

Regarding the configuration of the regional weather model (WRF in our case), we find that the downscaling procedure with a standard setup results in mostly underestimated precipitation totals, and a large variability among the ensemble members. This has led to a series of experiments to test the sensitivity of precipitation totals to a range

of differing model setups. For our purposes, we have found that short spin-up periods (encompassing around 24 hours before the heavy-precipitation episode) are preferable over long spin-up periods, which would allow (partial) adaptation of small-scale and slow-reacting variables in the model, such as soil moisture, for instance. In our experiments, precipitation totals in the ensemble become less variable and more realistic if the cyclones are already

present in the outermost model domain at model initialization; if not, the simulation runs too freely. Other than that, substantial changes of standard physics options do not increase model performance, be it cumulus or micro-

physics, or two-way nesting. Given that the parameterizations are turned off for the 9-3-km downscaling step, comparable outcomes with differing physics schemes point to the importance of the larger-scale atmospheric flow for producing the heavy precipitation. Although we find no relevant enhancements from nudging in smaller domains in our test experiments, nudging smaller domains could still be beneficial for other specific studies. In the

simulations of the cyclonic vortex, the largest deviations along the downscaling steps occur in the 27-km domains. The increasing variability of the simulations in these domains might be explained by the fact that no nudging is applied, while it is in the larger, 81-km domains. Although going back far in time, we have only analyzed a very

small number of events – many more cases would be needed to reach robust recommendations on how to configure a model for Vb cases. Nevertheless, we have demonstrated that that one can achieve a relatively best configuration

for the desired application with a well-thought series of experiments. In our context, the EMD has proven to be a valuable and intuitively understandable tool for spatial verification of the simulated precipitation fields with observation-based reconstructions. In fact, our EMD analysis results in similar rankings as obtained from the more common spatial verification scores and metrics (MAE, box ratios; see

Table 2) or from intersubjective, visual analyses of the precipitation patterns. Regarding the representation of precipitation and related variables in the 20CR ensemble, we find that 20CR delivers a well-confined ensemble for the 2005 case. Given the coarser horizontal grid sizes and lower vertical resolution, it compares well with other long-term reanalysis products. The 1910 case is also comparably well defined in the 20CR ensemble, whereas the 1876 case shows more uncertain developments of the cyclone fields in

Deleted: W

Deleted: features

Deleted: For good results,

Deleted: should be

Deleted: the

Deleted: traditional

the ensemble. This gradually increasing uncertainty when going back in time is also found for precipitation-related variables along the downscaling steps. For instance, the dynamical downscaling procedure captures the peak episodes of all three cases, although gradually less well going back in time. Furthermore, the accuracy of precipitation totals is closely linked to the exact cyclone track and the exact location of the vortex when it comes to stalling. Concretely, this location should be over northern Italy, or just off the northern Adriatic coast for best simulation results with regards to the intensity and spatial distribution of precipitation totals over north-eastern Switzerland.

Ensemble members that do not follow such a trajectory produce erroneous precipitation totals in the control area, where too southerly tracks generally produce too little precipitation, and too northerly tracks lead to a break-up of the associated vortices because of interaction with the Alpine (model) topography. This is found to be a decisive element, because the exact (stalling) location of the vortex strongly influences the cyclonic moisture transport around the Alps and the exact inflow angle from a sector North to the Central Alps. In fact, IVT intensities of >200 kg m⁻¹ s⁻¹ or even more from the right direction are needed to reproduce the extreme events. Interestingly, we have found a range of members that produce more precipitation than observed and reconstructed for the 1910 case. We infer that with a slightly different, hence ideal constellation of the cyclonic vortex to produce heavy precipitation over northern Switzerland, e.g., a longer stalling at the right location, the 1910 floods could have had even worse impacts.

Misplacements of the vortex increase in the ensemble from the 2005 to the 1910 and 1876 cases. While the patterns and dynamics can be reproduced for the 2005 case and, a bit less well, for the 1910 case with downscaling from 20CR, the variability of the cyclone fields and tracks becomes very large in the 20CR ensemble for the 1876 case. As a consequence, we find synoptic patterns in some members that are substantially different from the 2005 case, e.g., with some cyclone tracks that do not anymore follow a typical Vb path anymore. Furthermore, the increasing uncertainties in the ensemble going back in time are also due to the decreasing quality and amount of assimilated pressure data in 20CR. For illustration, the total number of stations assimilated in 20CR in the year 1876 is 218 (Compo et al., 2015). This number grows to 377 in the year 1910 and to 9251 in the year 2005. Of course, this uncertainty propagates into our downscaled ensemble. In turn, this means that with the 1876 case, we may have reached the limits of downscaling from the current 20CR (version 2c with a 2° by 2° horizontal grid) for such complex weather situations. The WRF regional model requires more accurate locations and intensities of input variables, like cyclone fields and moisture transport, to properly reproduce such sensitive Vb cases. On the upside, we have shown that despite of these deficiencies, single ensemble members, even from the early cases, can be used to analyze and illustrate local-scale weather dynamics, as well as sensitivities of the precipitation over northern Switzerland to the evolution of the associated Vb vortex.

The question remains whether a full ensemble needs to be downscaled to gain such insights. Generally speaking, the benefit from downscaling all 20CR members is that we obtain a full set of propositions for local weather patterns during historical events. In terms of impact and intensities (in our case the local precipitation totals over northern Switzerland), the spread between these propositions is very large, reflecting the strong uncertainty inherent to the process of downscaling over a wide range of scales (here from 200 to 3 km). Using ensemble members hence has allowed us to (i) compare members with observations and select realistic runs, and to (ii) relate the differences among the members in local weather to a different evolution on larger scales. In hindsight however, the limitations

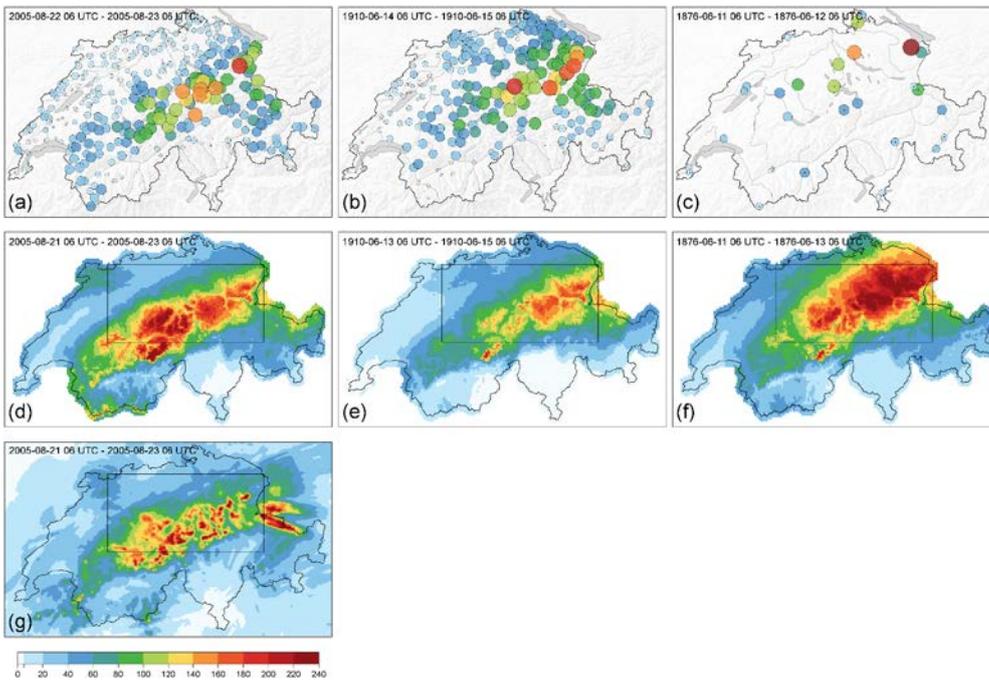
Deleted: are often not realistically located

Deleted: To some extent, this might be attributed to the slightly different synoptic flow pattern in the 1876 case (e.g. in terms of PCO type 2 versus type 1 in the other cases, as stated in Stucki et al. 2012).

Deleted: with the 1876 case

of downscaling and the potential ranges of the precipitation-related variables may as well be predictable from the input data to some extent. In our case, the well (or, in contrast, badly) confined cyclone tracks and fields in 20CR for the 2005 and 1910 (1876) cases give a good indication regarding the prospects of success for dynamical downscaling. This means that in a case where the driving atmospheric dynamics on a large scale can be anticipated, the chances of a good reproduction of the local patterns and intensities with accordingly selected ensemble members are high (cf. Stucki et al. 2015). A second option would be to save computational costs by downscaling to an intermediate scale in the first place, assess the relevant dynamics in this domain, and then do the full downscaling with a well-reasoned selection of members. In our case, the largest deviations from the initial conditions often appear in the 27-km domain (the largest domain without nudging), if not already present in the 20CR member. This means that downscaling to the first non-nudged domain could be sufficient to assess if an ensemble detects a cyclone well. In such a way, future studies may minimize the computational efforts for downscaling from a coarsely resolved reanalysis ensemble.

Appendix A



15 **Figure A1:** Upper panels: Daily precipitation totals (mm/24 h, indicated by color shade and circle sizes) from measurements over Switzerland, starting on (a) 22 August 2005 06 UTC, (b) 14 June 1910 06 UTC, and (c) 11 June 1876 06 UTC. Middle panels: Reconstructions of 48-h totals starting at (d) 21 August 2005 06 UTC, (e) 13 June 1910 06 UTC, and (f) 11 June 1876 06 UTC, as derived from RrecahSD data. Lower panel: Reconstruction of (g) 48-h totals starting at 21 August 2005 06 UTC as derived from CombiPrecip. The rectangle inset shows the smaller box (i.e., control area) used for the VarRatio, EMD and precipitation totals; the full panel shows approximately the larger box used.

Deleted: p

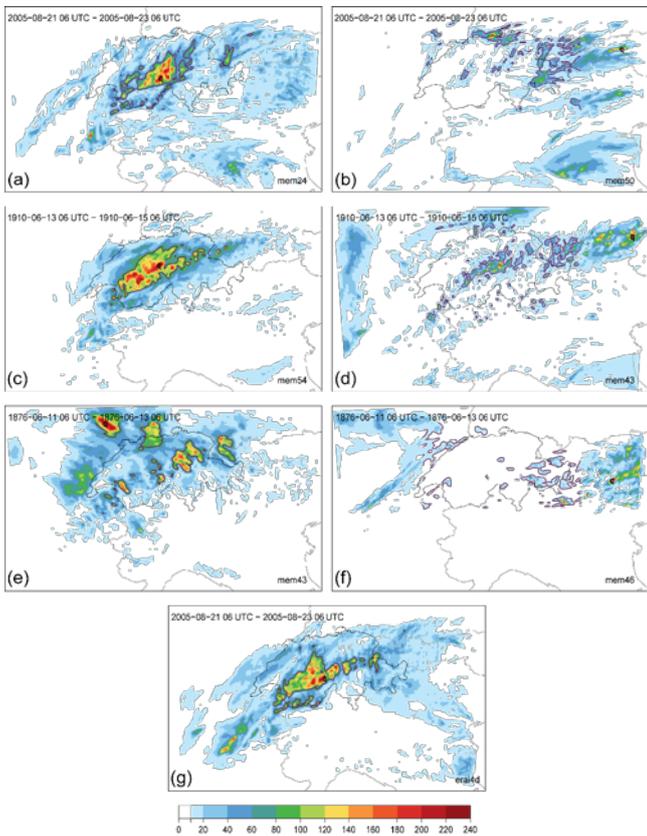


Figure A2: 48-h precipitation totals (time periods as in Figure A1 d, e, and f) for (a,c,e) maximum-precipitation members and (b, d, f) minimum-precipitation members for (a,b) the 2005, (c,d) the 1910 and (e,f) the 1876 cases, and for (g) the same simulation setup, but with downscaling ERA-Interim for the 2005 case. Precipitation totals may be larger than 240 mm in the panels.

5

- Deleted: 2
- Deleted: good (accurate
- Deleted:
- Deleted: and EMD)
- Deleted: bad (low
- Deleted:
- Deleted: , high EMD)

Author contribution

PS, PF, MZ, MM and AM designed the experiments. MZ and MM carried them out. FAI contributed RrecabsD. PS, PF, and MZ produced the figures and tables. All authors contributed to interpretation of the analyses, particularly the spatial verification, as well as writing or reviewing the manuscript.

10

Acknowledgements

PS and PF have been supported by the Oeschger Centre for Climate Research, University of Bern. MZ has received support from the Federal Commission for Scholarships for Foreign Students through the Swiss Government

Excellence Scholarship (ESKAS No. 2015.0793) for the academic year(s) 2015-2018/19. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office.

Competing interests

The authors declare that they have no conflict of interest.

References

- 10 Aebischer, U. and Schär, C.: Low-Level Potential Vorticity and Cyclogenesis to the Lee of the Alps, *J. Atmos. Sci.*, 55(2), 186–207, doi:10.1175/1520-0469(1998)055<0186:LLPVAC>2.0.CO;2, 1998.
- Awan, N. K. and Formayer, H.: Cutoff low systems and their relevance to large-scale extreme precipitation in the European Alps, *Theor. Appl. Climatol.*, 129(1), 149–158, doi:10.1007/s00704-016-1767-0, 2017.
- Baker, L. H., Gray, S. L. and Clark, P.: Idealised simulations of sting-jet cyclones, *Q. J. R. Meteorol. Soc.*, 15 140(678), 96–110, doi:10.1002/qj.2131, 2013.
- van Bebber, W. J.: Die Zugstrassen der barometrischen Minima nach Bahnenkarten der Deutschen Seewarte für den Zeitraum von 1870–1890, *Meteorol. Zeitschrift*, 8, 361–366, 1891.
- Beniston, M.: August 2005 intense rainfall event in Switzerland: Not necessarily an analog for strong convective events in a greenhouse climate, *Geophys. Res. Lett.*, 33(5), 1–5, doi:10.1029/2005GL025573, 2006.
- 20 Bezzola, G. R. and Hegg, C.: Ereignisanalyse Hochwasser 2005, Teil 1 - Prozesse, Schäden und erste Einordnung, Bern, Switzerland., 2007.
- Bezzola, G. R. and Hegg, C.: Ereignisanalyse Hochwasser 2005, Teil 2 - Analyse von Prozessen, Massnahmen und Gefahregrundlagen, Bern, Switzerland., 2008.
- Bezzola, G. R. (BAFU), Hegg, C. (WSL) and Koschni, A. (WSL): Hochwasser 2005 in der Schweiz, 25 Synthesebericht zur Ereignisanalyse, Bern, Switzerland., 2008.
- [Bowden, J. H., Otte, T. L., Nolte, C. G. and Otte, M. J.: Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling, *J. Clim.*, 25\(8\), 2805–2823, doi:10.1175/JCLI-D-11-00167.1, 2012.](#)
- Brugnara, Y., Brönnimann, S., Zamuriano, M., Schild, J., Rohr, C. and Segesser, D. M.: Reanalyses shed light on 30 1916 avalanche disaster, *ECMWF Newsl.*, (151), 28–34, 2017.
- Buzzi, A. and Tibaldi, S.: Cyclogenesis in the lee of the Alps: A case study, *Q. J. R. Meteorol. Soc.*, 104(440), 271–287, doi:10.1002/qj.49710444004, 1978.
- [Cioni, G. and Hohenegger, C.: Effect of Soil Moisture on Diurnal Convection and Precipitation in Large-Eddy Simulations, *J. Hydrometeorol.*, 18\(7\), 1885–1903, doi:10.1175/jhm-d-16-0241.1, 2017.](#)
- 35 Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y.,

- Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Mauder, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D. and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. R. Meteorol. Soc.*, 137(654), 1–28, doi:10.1002/qj.776, 2011.
- 5 [Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Allan, R. J., McColl, C., Yin, X., Vose, R. S., Matsui, N., Ashcroft, L., Auchmann, R., Benoy, M., Bessemoulin, P., Brandsma, T., Brohan, P., Brunet, M., Comeaux, J., Cram, T. A., Crouthamel, R., Groisman, P. Y., Hersbach, H., Jones, P. D., Jonsson, T., Jourdain, S., Kelly, G., Knapp, K. R., Kruger, A., Kubota, H., Lentini, G., Lorrey, A., Lott, N., Lubker, S. J., Luterbacher, J., Marshall, G. J., Mauder, M., Mock, C. J., Mok, H. Y., Nordli, O., Przybylak, R., Rodwell, M. J., Ross, T. F., Schuster, D., Srnec, L., Valente, M. A., Vizi, Z., Wang, X. L., Westcott, N., Woollen, J. S., and Worley, S. J.: The International](#)
- 10 [Surface Pressure Databank version 3. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Accessed 17 July 2019, doi: 10.5065/D6D50K29, 2015.](#)
- 15 [Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda, M., Belusic, D., Caldas-Alvarez, A., Cardoso, R. M., Davolio, S., Dobler, A., Fernandez, J., Fita, L., Fumiere, Q., Giorgi, F., Goergen, K., Güttler, I., Halenka, T., Heinzeller, D., Hodnebrog, Jacob, D., Kartsios, S., Katragkou, E., Kendon, E., Khodayar, S., Kunstmann, H., Knist, S., Lavín-Gullón, A., Lind, P., Lorenz, T., Maraun, D., Marelle, L., van Meijgaard, E., Milovac, J., Myhre, G., Panitz, H. J., Piazza, M., Raffa, M., Raub, T., Rockel, B., Schär, C., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Stocchi, P., Tölle, M. H., Truhetz, H., Vautard, R., de Vries, H. and Warrach-Sagi, K.: A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean, Springer Berlin Heidelberg., 2018.](#)
- 20 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and
- 25 performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137(656), 553–597, doi:10.1002/qj.828, 2011.
- Düsterhus, A. and Hense, A.: Advanced information criterion for environmental data quality assurance, *Adv. Sci. Res.*, 8, 99–104, doi:10.5194/asr-8-99-2012, 2012.
- Düsterhus, A. and Wahl, S.: Advanced score for the evaluation of prediction skill, in EGU General Assembly Conference Abstracts, vol. 20, p. 6894., 2018.
- 30 Farchi, A., Bocquet, M., Roustan, Y., Mathieu, A. and Quérel, A.: Using the Wasserstein distance to compare fields of pollutants: Application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident, *Tellus, Ser. B Chem. Phys. Meteorol.*, 68(1), doi:10.3402/tellusb.v68.31682, 2016.
- Ferrier, B.: A Double-Moment Multiple-Phase Four-Class Bulk Ice Scheme. Part I: Description, *J. Atmos. Sci.*, 51(2), 249–280, doi:https://doi.org/10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2, 1994.
- 35 Frei, C.: August-Hochwasser 2005: Analyse der Niederschlagsverteilung, *Arbeitsberichte der MeteoSchweiz*, (211), 1–5, 2005.
- Frei, C.: Eine Länder übergreifende Niederschlagsanalyse zum August Hochwasser 2005: Ergänzung zu Arbeitsbericht 211, *Arbeitsberichte der MeteoSchweiz*, (213), 10, 2006.

- Froidevaux, P. and Martius, O.: Exceptional moisture transport towards orography : a precursor to severe floods in Switzerland, 142, 1997–2012, doi:10.1002/qj.2793, 2016.
- Füllemann, C., Begert, M., Croci-Maspoli, M. and Brönnimann, S.: Digitalisieren und Homogenisieren von historischen Klimadaten des Swiss NBCN – Resultate aus DigiHom, Arbeitsberichte der MeteoSchweiz, (236), 5 2011.
- Hilker, N., Badoux, A. and Hegg, C.: The swiss flood and landslide damage database 1972-2007, Nat. Hazards Earth Syst. Sci., 9(3), 913–925, doi:10.1002/asl.183, 2009.
- Hofstätter, M., Chimani, B., Lexer, A. and Blöschl, G.: A new classification scheme of European cyclone tracks with relevance to precipitation, Water Resour. Res., 52(9), 7086–7104, doi:10.1002/2016WR019146, 2016.
- 10 Hofstätter, M., Lexer, A., Homann, M. and Blöschl, G.: Large-scale heavy precipitation over central Europe and the role of atmospheric cyclone track types, , 38(December 2017), doi:10.1002/joc.5386, 2018.
- [Hofstätter, M. and Blöschl, G.: Vb Cyclones Synchronized With the Arctic-/North Atlantic Oscillation, J. Geophys. Res. Atmos., 124\(6\), 3259–3278, doi:10.1029/2018JD029420, 2019](#)
- Hohenegger, C., Walser, A., Langhans, W. and Schär, C.: Cloud-resolving ensemble simulations of the August 15 2005 Alpine flood, Q. J. R. ... , 134, 889–904, doi:10.1002/qj. 2008.
- Horton, P. and Brönnimann, S.: Impact of global atmospheric reanalyses on statistical precipitation downscaling, Clim. Dyn., 52(9), 5189–5211, doi:10.1007/s00382-018-4442-6, 2019.
- Isotta, F. A., Begert, M. and Frei, C.: Long-Term Consistent Monthly Temperature and Precipitation Grid Data Sets for Switzerland Over the Past 150 Years, J. Geophys. Res. Atmos., 124(7), 3783–3799, 20 2019, doi:10.1029/2018JD029910, 2019.
- Jaun, S., Ahrens, B., Walser, A., Ewen, T. and Schär, C.: A probabilistic view on the August 2005 floods in the upper Rhine catchment, Nat. Hazards Earth Syst. Sci., 8(2), 281–291, doi:10.5194/nhess-8-281-2008, 2008.
- Jiménez, P. A. and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind in the wrf model, J. Appl. Meteorol. Climatol., 51, 300–316, doi:10.1175/JAMC-D-11-084.1, 2012.
- 25 Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, J. Appl. Meteorol., 43(1), 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.
- [Joliffe, I. T. and Stephenson, D. B.: Forecast Verification, Wiley-Blackwell Publishing, Inc, Chichester., 2012.](#)
- Kelemen, F. D., Ludwig, P., Reyers, M., Ulbrich, S. and Pinto, J. G.: Evaluation of moisture sources for the Central European summer flood of May/June 2013 based on regional climate model simulations, Tellus A Dyn. Meteorol. 30 Oceanogr., 68(1), 29288, doi:10.3402/tellusa.v68.29288, 2016.
- Kljun, N., Sprenger, M. and Schär, C.: Frontal modification and lee cyclogenesis in the alps: A case study using the alpeX reanalysis data set, Meteorol. Atmos. Phys., 78(1–2), 89–105, doi:10.1007/s007030170008, 2001.
- Köppen, W.: Die Zugstrassen der barometrischen Minima in Europa und auf dem nordatlantischen Ocean und ihr Einfluss auf Wind und Wetter bei uns, Mitth. der Geogr. Gesellschaft Hambg., 1, 76–97, 1881.
- 35 Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K. and Janssen, P.: A coupled data assimilation system for climate reanalysis, Q. J. R. Meteorol. Soc., 142(694), 65–78, doi:10.1002/qj.2629, 2016.
- Langhans, W., Gohm, A. and Zängl, G.: The orographic impact on patterns of embedded convection during the August 2005 Alpine flood, Q. J. R. Meteorol. Soc., 137(661), 2092–2105, doi:10.1002/qj.879, 2011.

- Masson, D. and Frei, C.: Short Communication Long-term variations and trends of mesoscale precipitation in the Alps : recalculation and update for 1901 – 2008, *Int. J. Climatol.*, 36, 492–500, doi:10.1002/joc.4343, 2016.
- McClellan, D. (UNISDR) and Guha-Sapir, D. (CRED): 2018 REVIEW OF DISASTER EVENTS, 6, 2019.
- Messmer, M., Gómez-Navarro, J. J. and Raible, C. C.: Climatology of Vb cyclones, physical mechanisms and their impact on extreme precipitation over Central Europe, *Earth Syst. Dyn.*, 6(2), 541–553, doi:10.5194/esd-6-541-2015, 2015.
- Messmer, M., Gómez-Navarro, J. J. and Raible, C. C.: Sensitivity experiments on the response of Vb cyclones to sea surface temperature and soil moisture changes, *Earth Syst. Dyn.*, 8(3), 477–493, doi:10.5194/esd-8-477-2017, 2017.
- 10 MeteoSwiss: Starkniederschlagsereignis August 2005, *Arbeitsberichte der MeteoSchweiz*, 211, 63pp, 2006.
- MeteoSwiss: Documentation of MeteoSwiss grid-data products: Daily precipitation (final analysis): RhiresD, 1–4, doi:OFEV2014, 2013.
- Morrison, H., Thompson, G. and Tatarskii, V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes, *Mon. Weather Rev.*, 137(3), 991–1007, doi:10.1175/2008MWR2556.1, 2009.
- 15 Nissen, K. M., Ulbrich, U. and Leckebusch, G. C.: Vb cyclones and associated rainfall extremes over Central Europe under present day and climate change conditions, , 22(6), 649–660, doi:10.1127/0941-2948/2013/0514, 2014.
- Parodi, A., Ferraris, L., Gallus, W., Maugeri, M., Molini, L., Siccardi, F. and Boni, G.: Ensemble cloud-resolving modelling of a historic back-building mesoscale convective system over Liguria : the San Fruttuoso case of 1915, 455–472, doi:10.5194/cp-13-455-2017, 2017.
- Pfahl, S.: Characterising the relationship between weather extremes in Europe and synoptic circulation features, *Nat. Hazards Earth Syst. Sci.*, 14(6), 1461–1475, doi:10.5194/nhess-14-1461-2014, 2014.
- Pfister, C.: *Wetternachhersage: 500 Jahre Klimavariationen und Naturkatastrophen (1496-1995)*, Verlag Paul Haupt, Bern., 1999.
- 25 Pieri, A. B., von Hardenberg, J., Parodi, A. and Provenzale, A.: Sensitivity of precipitation statistics to resolution, microphysics, and convective parameterization: A case study with the high-resolution WRF climate model over Europe, *J. Hydrometeorol.*, 16(4), 1857–1872, doi:10.1175/JHM-D-14-0221.1, 2015.
- Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Keller, M., Tölle, M., Gutjahr, O., Feser, F. and 30 Brissson, E.: Reviews of Geophysics A review on regional convection-permitting climate modeling : Demonstrations , prospects , and challenges, *Rev. Geophys.*, 53, 323–361, doi:10.1002/2014RG000475., 2015.
- Röthlisberger, G.: *Chronik der Unwetterschäden in der Schweiz*, Berichte der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft WSL, 330, 122, 1991.
- Rubner, Y., Tomasi, C. and Guibas, L. J.: A metric for distributions with applications to image databases, in *Sixth International Conference on Computer Vision*, pp. 59–66., 1998.
- 35 Rubner, Y., Tomasi, C. and Guibas, L. J.: Earth mover’s distance as a metric for image retrieval, *Int. J. Comput. Vis.*, 40(2), 99–121, doi:10.1023/A:1026543900054, 2000.
- Schiemann, R., Liniger, M. A. and Frei, C.: Reduced space optimal interpolation of daily rain gauge precipitation in Switzerland, 115, 1–18, doi:10.1029/2009JD013047, 2010.

- Schmidli, J., Schmutz, C., Frei, C., Wanner, H. and Schär, C.: Mesoscale precipitation variability in the region of the European Alps during the 20th century, *Int. J. Climatol.*, 22(9), 1049–1074, doi:10.1002/joc.769, 2002.
- [Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J., Coniglio, M. C. and Wandishin, M. S.: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather Forecast.*, 25\(1\), 263–280, doi:10.1175/2009WAF2222267.1, 2010.](#)
- Skamarock, W. C., Klemp, J. B., Gill, D. O., Barker, D. M., Duda, M. G., Wang, W. and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Tech. Note, (NCAR/TN-475+STR), 2008.
- von Storch, H. and Zorita, E.: The History of Ideas of Downscaling—From Synoptic Dynamics and Spatial Interpolation, *Front. Environ. Sci.*, 7, 21, doi:10.3389/fenvs.2019.00021, 2019.
- von Storch, H., Langenberg, H. and Feser, F.: A Spectral Nudging Technique for Dynamical Downscaling Purposes, *Mon. Weather Rev.*, 128(10), 3664–3673, doi:10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2, 2000.
- Stucki, P., Rickli, R., Brönnimann, S., Martius, O., Wanner, H., Grebner, D. and Luterbacher, J.: Weather patterns and hydro-climatological precursors of extreme floods in Switzerland since 1868, *Meteorol. Zeitschrift*, 21(6), 531–550, doi:10.1127/0941-2948/2012/368, 2012.
- Stucki, P., Brönnimann, S., Martius, O., Welker, C., Rickli, R., Dierer, S., Bresch, D. N., Compo, G. P. and Sardeshmukh, 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, doi:10.1175/BAMS-D-14-00041.1, 2015.
- Stucki, P., Bandhauer, M., Heikkilä, U., Rössler, O., Zappa, M., Pfister, L., Salvisberg, M., Froidevaux, P., Martius, O., Panziera, L. and Brönnimann, S.: Reconstruction and simulation of an extreme flood event in the Lago Maggiore catchment in 1868, *Nat. Hazards Earth Syst. Sci.*, 18(10), 2717–2739, doi:10.5194/nhess-18-2717-2018, 2018.
- SwissRe: Natural catastrophes and man-made disasters in 2017: a year of record-breaking losses., 2018.
- Thompson, G., Field, P. R., Rasmussen, R. M. and Hall, W. D.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, *Mon. Weather Rev.*, 136, 5095–5115, doi:10.1175/2008MWR2387.1, 2008.
- Ulbrich, U., Bru, T., Fink, A. H. and Leckebusch, G. C.: The central European floods of August 2002 : Part 2 – Synoptic causes and considerations with respect to climatic change, 58(4), doi:10.1256/wea.81.03A, 2003.
- Wernli, H. and Schrier, C.: Surface Cyclones in the ERA-40 Dataset (1958–2001). Part I: Novel Identification Method and Global Climatology, *J. Atmos. Sci.*, 63(10), 2486–2507, doi:10.1175/JAS3766.1, 2006.
- [Wernli, H., Paulat, M., Hagen, M. and Frei, C.: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Mon. Weather Rev.*, 136, 4470–4487, doi:10.1175/2008MWR2415.1, 2008](#)
- Welker, C. and Martius, O.: Large - scale atmospheric flow conditions and sea surface temperatures associated with hazardous winds in Switzerland, *Clim. Dyn.*, 44, 1857–1869, doi:10.1007/s00382-014-2404-1, 2015.
- Zängl, G.: Interaction between Dynamics and Cloud Microphysics in Orographic Precipitation Enhancement A High-Resolution Modeling Study of Two North Alpine Heavy-Precipitation Events, *Mon. Weather Rev.*, 135, 2817, doi:10.1175/Mwr3445.1, 2007a.

Zängl, G.: To what extent does increased model resolution improve simulated precipitation fields? A case study of two north-Alpine heavy-rainfall events, *Meteorol. Zeitschrift*, 16(5), 571–580, doi:10.1127/0941-2948/2007/0237, 2007b.

Zbinden, P.: 2005 *Annalen Annales Annali, MeteoSchweiz*, 2005.

- 5 Zheng, Y., Alapaty, K., Herwehe, J. A., Del Genio, A. D. and Niyogi, D.: Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) model with an updated Kain-Fritsch scheme, *Mon. Weather Rev.*, 144(3), 833–860, doi:10.1175/MWR-D-15-0005.1, 2016.

Interactive comment on “Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux” by Peter Stucki et al.

Anonymous Referee #1

Received and published: 27 June 2019

General comments:

The authors downscale three historic flood events caused by so-called Vb-cyclones and analyse the conditions under which the high-resolution simulations capture precipitation over Switzerland best. They compare model set-ups with different initialization periods, parametrizations and nudging. The most important factor determining the result with respect to precipitation is the correct representation of the cyclone path.

The article is well written and structured. It addresses three extreme precipitation events and is therefore within the scope of the journal. The authors systematically analyse different model configurations and initial data sets which allows them to identify influencing factors of higher and lower importance. The result is a good reference for other scientists applying dynamical downscaling to study extreme precipitation events. It also shows the pros and cons of using 20CR reanalysis for initializing regional model simulations, as it turns out that deviations to observations increase for events lying further in the past.

There is however one aspect that is missing in the paper. The authors downscale to a precipitation resolving resolution of 3 km. As a motivation they cite Zängl 2007 and Prein et al. 2015 who analysed the 2005 event (page 3, line 29). The simulations performed by Stucki et al. would allow to study the effect of convection resolving simulations versus simulations with convection parametrization also for the two other cases. The paper would strongly benefit from adding an additional section which discusses the effect of the last downscaling step in more detail.

We thank the reviewer for the clear and helpful comments. Specifically, we appreciate the valuable note suggesting analyses on the effect of the last downscaling step and the associated convection. This is indeed an important aspect that we have neglected in the previous version. To address the suggestion, we have now redrawn Figure 5 and added two panel rows showing the contribution of convective precipitation in the 9-km domain, where precipitation is parameterized. This gives insights in the process of (embedded) convection during the events. In addition, we add a new Figure 11. It shows surface weather simulated for a historical maximum-precipitation member at three instances in time. With this, we can better show how shifts in the cyclonic flow and moisture transport translate into regional to local surface weather and precipitation patterns.

Specific comments:

p.2 l.4: The cited article (BAFU 2005) is missing in reference section.

We thank the reviewer for the hint. It is in fact Bezzola et al. 2008, which is available in the reference list.

p.2 l.12: Cyclogenesis of Vb cyclones can be outside the Mediterranean region but most of them pass the Golf-of-Genoa region (e.g. Messmer et al., 2015).

We have adopted the suggested wording, and we agree that a broader definition is more adequate. It reads now as follows: “Cyclones on a Vb track are associated with heavy to extreme precipitation over Central Europe, and particularly north of the Alps (Hofstätter et al., 2016, 2018;

Nissen et al., 2013). Most of the cyclones following the Vb trajectory are generated in the western Mediterranean region, and most of them pass the Golf-of-Genoa region (Hofstätter and Blöschl, 2019; Messmer et al., 2015). For this, they are also called Genoa Lows at this stage; Bezzola et al. 2008)”

Table 2: Does a positive mean absolute error means that the model initialized with 20CR produces more precipitation than observed and that initialization with ERA40 (p.11) overestimates precipitation even further? This is probably not the case as it would be in contrast to Fig. 2. Please define what a positive MAE value indicates. From looking at Fig. 2 I also don't understand how MAE48h can be 27.75 for sp10 when averaging over 10 ensemble members. Shouldn't the error be much higher?

We have added a short explanation of MAE in Sect. 3.2. MAE gives absolute values, that is, the deviation or distance from the observation – regardless of its sign. The bias is actually negative. We also checked the calculations and come to the same results.

The new text reads as follows: “The MAE measures the average distance between forecast and observation, and is preferred over RMSE because it is more resistant to outliers, and over correlation coefficients because we are more interested in accuracy than linear association (Joliffe and Stephenson, 2012).”

p.11 l.28: Please give a short explanation how cyclone fields are calculated (1-2 sentences). To understand what is shown one shouldn't need to read another paper. Please also clarify if cyclone fields and cyclone tracks are both calculated at SLP.

We agree that this needs to be clarified. We have inserted a short explanation, and we have adapted the caption in Figure 4. The new text reads as follows: “The algorithm detects cyclone fields in terms of a finite area around a regional SLP minimum, that is, by a closed SLP contour line. The regional SLP minima for each cyclone life cycle are stored as cyclone tracks, and the presence or absence of a cyclone is represented in a binary field for each grid point and time step.”

p.12 l.17 +others: You use storm track and cyclone track synonymously. This is confusing as you only show cyclone tracks. The term storm track is mostly used for the variance at synoptic time scales (e.g. Hoskin and Hodges, 2002: New perspectives on the Northern Hemisphere winter storm track. Journal of Atmospheric Sciences, vol. 59, Issue 6, pp.1041-1061). I suggest that you change "storm track" to "cyclone track" everywhere in the manuscript.

We fully agree; the suggestion has been adopted.

Fig.4 first row: Genoa cyclogenesis is not visible on these figures.

We agree. Shown are cyclone tracks at the 500-hPa level in the 2° by 2° reanalysis. We have made this clear in the caption.

Fig.4 middle row: There is a time step indicated in each panel but the caption suggests that different time steps are combined. This is confusing.

We have rephrased the caption and we emphasize now that it summarizes the ensemble at one specific time step. The caption reads as follows: “. (g,h,i) Cyclone centers identified for sea level pressure at (g) 21 August 2005 18 UTC, (h) 13 June 1910 18 UTC, and (i) 10 June 1876 18 UTC, the same time steps as in (d, e, f). Colored dots and the tick marks in the color key indicate the

number of cyclone tracks located at a specific grid point at the respective time steps. Greyscaled lines mark the cyclone tracks over the period of 48 h before to 48 h past the respective time steps. Darker (lighter) grey shades indicate more (less) cyclone tracks along a certain path”

Fig.4 last row: Please redraw. It is almost impossible to see the cyclone tracks. In addition, the caption states “as in middle panels” even though g, h and i show a sequence of days while the d, e and f show single time steps. I also don’t understand why the numbers on the colour bars differ for g, h and i.

The panel is redrawn, with larger and darker lines, a smaller map area and larger circles. In addition, we have rephrased the caption. We also describe the last panel better in the text now, which in fact supports our conclusions. This is also in line with a suggestion by Reviewer 2.

Fig. 5a: There is a very prominent precipitation peak before the event starts. The text does not give any explanation for this peak.

This is true. We have now addressed the peaks in the text, including that they are convection-driven. We have also added the simulation results from WRF-ERA-Interim. We cannot fully conclude on the reasons for the overestimation during the spin-up, but found that at the very beginning of the model initialization, the interpolated temperature field is coarser in 20CR-WRF for the 9-km domain, which leads to areas with high temperatures over the Alps and excess moisture production in the first hours of the simulation. This mechanism is now summarized in the text.

Figs 7, 8 and 9: Are the black dots supposed to indicate the land sea mask? This does not come out neither in print nor on my screen. Please redraw.

This is a common issue for all reviewers. Apparently, this happened during rasterization of the original vector format to PNG format, such that it can be inserted in the submitted Word document. We still have trouble to produce a good PNG file, but hope that the image quality is now sufficient. The PDF to be submitted for print production should be in good quality.

p.24 last paragraph: The decreasing differences to observations with time in 20CR are probably due to the increasing quality/amount of data that is assimilated in 20CR. It would be good to mention how the input data for the 20CR product has changes between 1976 and 1910.

This is a good suggestion. We inserted the according numbers from the ISPD database in the conclusion; they are really quite illustrative.

Fig. A2 caption: “(time periods as in Figure 2)”. Fig. 2 only shows 2005 Fig. A2 shows all 3 episodes.

We thank the reviewer for this remark; it should refer to Fig. A1, of course.

Technical corrections:

p.9 l.11: Value in the table is 0.59 the value in the table captions is 0.49 (sp10 48-hour precip). Which one is correct?

We thank the reviewer for this hint. The value in the table is correct; we have corrected the one in the text.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2019-174>, 2019.

Further changes made

- **Additional references: Hofstätter and Blöschl, 2019; Zbinden, 2005 (Annals of MeteoSwiss); Cioni and Hohenegger, 2019; Coppola et al., 2018; Compo et al, 2015; Joliffe and Stephenson, 2012; Wernli et al., 2008.**
- **Figure A2 redrawn with the maximum- and minimum-precipitation members. This makes more sense since we focus on these in the text.**

Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to
model physics and cyclonic moisture flux

Peter Stucki
et al. nhe-
2019-174

There are 3 main objectives of this paper (i) to find a setup of the WRF model that is adequate for dynamical downscaling from 20CR, (ii) to investigate sensitivity of heavy precipitation to cyclonic moisture flux and (iii) to assess the uncertainty along the downscaling steps and among the ensemble members for historical cases. The paper is well written but I'm concerned about the balance of the paper. The paper is dominated by the technical aspects of performing downscaling for historical events (with poorly-defined motivation for performing model tests and testing ensemble suitability) and contains too brief analysis of the cyclonic moisture flux to achieve objective (ii) (see general comments below). If the points below are addressed this paper would be suitable for publication in NHESS.

We thank the reviewer for pointing to some important aspects and potential flaws of the manuscript. We think that while addressing the concerns, we provide now a more balanced manuscript regarding the technical and dynamical parts of the article. The motivation, or rather the rationale, of doing the series of experiments is now described in a more explicit and clear way. Considering also comments by Reviewer 1, we have better defined and extended the analyses of cyclonic moisture flux with new figure panels, a new figure and extended text. In all, we are grateful for the review, which supported us in improving the quality of the manuscript regarding some crucial aspects and clarity.

General comments

1. I'm concerned that there is no motivation given for testing sensitivity to the convection, microphysics schemes or nesting in WRF. What was the reason for performing these simulations? Did the authors have a hypothesis that they wanted to test? What are the differences between the schemes? The authors conclude that there is no difference in performance when changing the cumulus scheme or nesting but do not analyse this result. Did they expect to see a difference? If so, why are the results insensitive to these choices? What is the conclusion of these experiments and how general are they, i.e. would the same hold for other historical cases or are they specific to these cases? If the conclusions are case specific, then perhaps this analysis could be reported in an appendix?

We see the need for a better reasoning about the experiments. To begin with, we state more clearly why we started with ample spin-up time, then why are not satisfied with the standard setup, and then derive why we go with decreasing spin-up. Regarding the specific question of 'sensitivity' to a number of schemes and settings, we now state more clearly that "The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience." In turn, this means that we

do not aim to analyze each tuning option in detail.

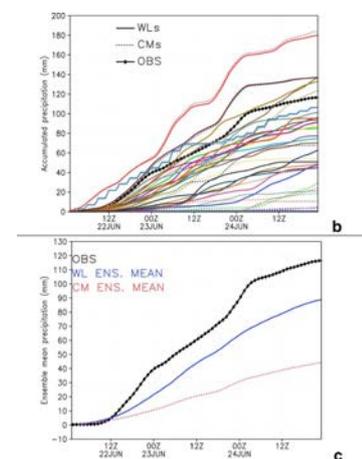
Some of this concern could also stem from the fact that we use the term ‘sensitivity’ in the title, which may provoke expectations of a thorough testing of many schemes and options. In a narrow sense, the term might only be acceptable for the more detailed tests regarding the spin-up. We use ‘sensitivity’ in a wider sense, where we also allude to checking the other tuning options in less detail as well as to the dynamic sensitivities addressed by contrasting members, for instance. In fact, we have considered replacing the term ‘sensitivity’ in the title and elsewhere (e.g. Implications from ..., the role of ..., etc.). We have not found a concise alternative term that would include all these meanings, and we think that the meaning in this article should become clear when reading the abstract.

2. There is some confusion in the paper over what constitutes a ‘good’ ensemble spread. The answer to this depends on the hypothesis being tested. At some points in the paper the authors claim the ensemble is good or bad by examining spread in the precipitation totals (figure 2, table 2) concluding that the 10-day forecast runs ‘too freely’ because the precipitation accumulation spread is large. However, later in the paper they examine the spread in cyclone tracks (figure 4) and conclude that there is a ‘good’ spread in storm track position for the 2005 and 1910 cases but not the 1876 case (meaning smaller track position differences in the ensemble). If the focus of the paper is to test the sensitivity of precipitation accumulation over Switzerland to cyclonic moisture flux, some spread in precipitation accumulation is surely necessary? However, spread in precipitation that occurs due to factors such as cyclone position presumably need to be minimised? Is this the rationale for the later measure of ensemble suitability? If so, why is precipitation spread used in the early analysis of ensemble spread? There appears to be some inconsistency in the analysis of ensemble suitability in the paper which needs to be clarified.

We agree that we need to address in more detail what we mean by a ‘good spread’, and the comment has helped a lot in this respect. We discussed the trade-off between ‘plausible runs’ and ‘necessary spread’. In the end, we think that we can justify our line of thoughts well in an additional paragraph in Sect. 3.1.

In addition, we include a reference (Coppola et al., 2018, <https://doi.org/10.1007/s00382-018-4521-8>) to a thorough test of multi-model ensembles, where the inclined reader can actually look up how state-of-the-art models perform in a comparable Alpine setting, see their Figure 5 to the right.

The new text reads now as follows: “In all, we cannot be satisfied with these results yet. On the one hand, we aim to investigate particularly flood-inducing features of Vb-cyclones. For this, a certain variability in the ensemble is helpful and necessary. For example, we can find and assess (non-)decisive features by means of opposing ensemble members. On the other hand, we also need to ensure that our downscaling experiment delivers plausible results, especially regarding precipitation intensities and patterns. For this, the deviations from the observations must not become too large in the ensemble. A somewhat smaller spread of the simulated precipitation for the 2005 case would also increase our confidence that



the simulation of historical events will produce reasonable and valid results. In short, we would have expected less underestimation and smaller deviations with this downscaling configuration (cf. Coppola et al., 2018, their figure 5, for estimations of accumulated precipitation over the Alps from a multi-model ensemble).”

3. Analysis relating to the sensitivity of precipitation to cyclonic moisture fluxes (figures 7-10) is described in just 19 lines. This is rather brief for 4 figures, especially given that one of the major objectives of the paper is to examine ‘sensitivity to cyclonic moisture flux’ (title). The analysis must be expanded to provide a better balance between the technical aspects and the scientific hypothesis testing analysis and to achieve objective (ii).

We agree on this comment, which is similar to concerns by Reviewer 1, and it has made us review and revise the manuscript in several ways.

The discussion of Figure 10 is actually done later in the text (see also reply to specific comment 25). We have introduced more references to Figure 10, for instance, to enhance the context of our findings.

To be more balanced between technical and dynamic aspects of sensitivity, and to address the innermost domains better (see comments by Reviewer 1), we have now redrawn Figure 5 and added two panel rows showing the contribution of convective precipitation in the 9-km domain, where precipitation is parameterized. In addition, we add a new Figure 11. It shows surface weather simulated for a historical maximum-precipitation member at three instances in time. With this, we can better show how shifts in the cyclonic flow and moisture transport translate into regional to local surface weather and precipitation patterns. An according clause is added in the Abstract.

Specific comments

1. Page 1, line 22: What is ‘moderate’ spectral nudging?
Moderate refers to wavelengths of 1500 km; this is only defined in the text. We see that the term is not commonly used and decided to remove the information from the abstract, and instead focus on the most robust result, the initialization period. This is also in accordance to the comment by Reviewer 3.
2. Page 2, line 20: How is the moist air ‘let’ around the Eastern Alps? Do you mean advected?
The sentence is rephrased in active voice with ‘flowed over and around the Eastern Alps’. Note that in agreement with Reviewer 3, the dynamics of ‘cyclonic moisture flux’ is now better defined here.
3. Page 4, line 28: The authors state that they use a ‘consistent part of the calibration period, which is accordingly slightly reduced’. I’m unclear what the consistent part of the calibration refers to. Please could the authors expand on this?
We have modified the sentence to improve clarity. The stations used are the one that were available in the days of interest in 1910 or 1876 and at the

same time during most of the calibration period. We repeat the calibration period (1981- 2010, given also two sentences above). The calibration period is slightly reduced to the days when all chosen stations were available (the method needs to have the same set of stations in the calibration period, in our case from 1981-2010, and the reconstructed period, i.e. the days in 1910 or 1876).

4. Page 5, line 5. It would be useful to know if any of the assimilated surface pressure observations were located in Switzerland.
We thank the reviewer for the hint. We have introduced the numbers here, and come back to the total number of assimilated stations in the conclusions. This is also in line with a suggestion by Reviewer 1.

5. Figure 1: The numbers on the colour bar have been cut off.
We thank the reviewer for the remark. We have replaced Figure 1.

6. Page 6, line 26 and page 16, line 18: The authors refer to ‘two peak episodes’ but in figures 2 and 5 the CombiPrecip dataset does not show 2 peak episodes. Instead there is continuous high precipitation rates over a 30hr period.
We were taking the perspective from the simulation, but agree that this is a misleading phrasing. We rephrased on Page 6 to ‘Furthermore, CombiPrecip shows high precipitation rates over most of the analyzed period. In contrast, the downscaled ensemble has only two periods of very high precipitation rates (at around 12 and 36 hours after initialization), and obviously, there is too little precipitation between these two simulated peak episodes’. A similar phrasing is employed on Page 16. We think that it fits also better with the revisions suggested by Reviewer 3.

7. Figure 2: The right-hand axis does not have any units. Also, it is not clear what the red numbers represent.
We thank the reviewer for pointing this out. We have redrawn Figure 2, and have added a # to mark the quartile members, and have also added it in the caption for clarity.

8. Page 7, line 1: It is not surprising that a 10-day forecast exhibits large spread in the ensemble. However, this is not necessarily a bad thing if the cyclone tracks are similar but with differing moisture flux as they would still be able to test the sensitivity of precipitation totals to moisture flux. Therefore, I don’t think it is sensible to examine the suitability of the ensemble by looking at the spread in precipitation as is done in table 2.
We suppose Figure 2 is meant instead of Table 2. This comment helped a lot for clarifying our rationale for the downscaling, and the follow-up series of experiments. Following also the comment on this topic by Reviewer 3, we now explain this in more detail in the analysis and interpretation of Figure 2. Later in the text, we also explain why we are not satisfied with the results and how we came up with the follow-up experiments. See also our reply to comment 10.

9. Page 7, line 16: Can the authors be more specific about the section containing the full evaluation. Currently they say it is 'below', below where?
We have introduced 'Sect. 3.2'.
10. Page 10, lines 11-12: There is no motivation given for testing sensitivity to the convection, microphysics schemes or nesting in WRF. What was the reason for performing these simulations? Did the authors have a hypothesis that they wanted to test? What are the differences between the schemes?
We suppose it is Page 8 instead of 10. There, we now give the reasoning for the last series of experiments: 'The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience.' **We thank the reviewer for this remark; we think we are much clearer now about the purposes of the three series of experiments. See also our response to Reviewer 1 on this issue.**
11. Page 10, line 27: The authors claim that there is a systematic improvement with decreasing lead time. However, this is difficult to detect in the spatial verification statistics shown in table 2.
We see that detection by eye might be difficult. Therefore, we have included the calculated Theil-Sen slopes in the text for better documentation. The new text reads now as follows: "Theil-Sen slope estimates are calculated over sp10, sp7, sp5, sp3, and sp1 for all measures; they are all negative (MAE 24 h: -0.2; MAE 48 h: -0.70; BOX 24 h: -0.03; BOX 48 h: -0.03; VIS 24 h: -0.75; VIS 48 h: -1.13; EMD 24 h: -18.5; EMD 48 h: -17.9). Although the trends are not significant in Mann-Kendall tests (or not clearly attributable, due to the small sample), the negative slopes indicate that performance generally increases with decreasing spin-up time"
12. Page 11, line 4: The authors conclude that there is no difference in performance when changing the cumulus scheme or nesting but do not analyse this result. Did they expect to see a difference? If so, why are the results insensitive to these choices? What is the conclusion of this experiment?
We agree that the explanations are not very detailed in the manuscript. This is mainly because, as we state in the text now, the goal of these last experiments is not to achieve a robust sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal configuration of the WRF model. We have inserted some explanations of why a different setup might change the precipitation pattern and indicate some more literature. Some of our reasoning is given below in more detail. The microphysics parameterization describes the hydrometeors (cloud water, raindrops, ice, snow, graupel, etc.), the number concentration, size, fall speed etc. These descriptions differ between the chosen parameterizations. These differences play a role in cloud development and hence, also in precipitation amounts and patterns. As the microphysics parameterization is responsible for all the precipitation in the innermost domain (no cumulus parameterization), changes in this parameterization can certainly affect precipitation amounts and patterns. Compared to

Thompson, the ice particles (ice, graupel and hail) are described in less detail in the Ferrier parameterization. The reason, why the difference is rather small might be that we only look at summer events, where ice particles are less important. We have included a sentence about this argumentation in the manuscript.

The two cumulus parameterizations that have been used show differences in scale-dependence. The Kain-Fritsch scheme is one of the most commonly used parameterizations over Europe. The scale-dependent scheme is designed to improve the realizations in the so-called grey zone between 10 and 5 km.

Since we do not employ a cumulus parameterization in the inner nest, the changes in the outer domains seem to be relatively small, when changing the cumulus scheme. This might be because the description of the convection in this case is not triggered by fine scale structures, but mainly by lifting along the orography. We have included a short description of the differences in the two cumulus parameterizations.

Two-way nesting can be helpful, if an event can be modified by the small-scale features that are resolved in the innermost domain. In such a case, the result of the innermost domain can be transported to the coarser domains as well. In a Vb event one or two-way nesting might not make a big difference, as the cyclone is steered by large-scale atmospheric flow.

13. Page 11, line 28: Here the authors present figures 4d-f and 4g-l but do not analyse these figures. If they are not referred to in the text should they be removed?

The references can be found on p11 bottom and p12 top, then again on p12|18 for Figure 4a, d and g, on p12|26 for Figure 4b, e and h, and on p14L8 for Figure 4c, f and i. We see that more references could be helpful, though, and introduced some more in the text where appropriate.

14. Figure 3: The right-hand edge of the figure has been cut off. It is also not clear what cross-section figures 3i-k are for. Could the cross-section be added to figures 3f-h respectively?

We do not see the cut in our submission. However, we have redrawn Figure 3, adding all cross-sections to simplify referencing, and have rephrased the caption for more clarity.

15. Page 14, line 18 and elsewhere: The authors conclude that the ensemble spread becomes increasingly larger when going back in time. Although this is an intuitive result, it is not possible to conclude this from 3 points only. More case studies would be needed to confirm this.

We did not intend to make a general claim here. The whole paragraph is meant as a summary / discussion of the synoptic analyses before we go on. This is now clearly stated at the top of the paragraph.

16. Page 14, lines 12-22: The authors do not refer to any figures in this analysis section. Which figures are used? Is this where the analysis of 4d-f and 4g-l is performed?

See also the reply to comment 14. We include now more specific references to the respective Figures, and add some more text to guide the reader.

17. Figure 4: Why is a different domain used in figures d-f? Is the ensemble track position agreement in the North-Atlantic relevant? It appears as though the track agreement over Switzerland is similar for all 3 cases, is this correct?
The larger area is chosen to show the relatively good performance of 20CR in the region of interest in comparison to other regions in this area. We have extended the phrasing to ‘Overall, the analyses at synoptic scales (Figures 3 and 4) show that differences among the 20CR members are substantially smaller over the region of interest (Southern and Central Europe) than over other regions of the North Atlantic / European sector (Figure 4 d, e and f); this corresponds to the relatively high density of assimilated stations over Central Europe (not shown; see Compo et al. 2015).’
18. Figure 4: I do not know what figures 4g-i are showing. Please explain these figures in the text.
The Figure panels have been redrawn, according to a comment by Reviewer 1, and the caption has been extended. As requested, we now analyze the three panels in more detail. We thank the reviewer for this comment; we think the revisions make our points much clearer now.
19. Figure 4: These are quite complex figures. Are the country outlines important? Perhaps they could be removed? Or only Switzerland included?
We would like to keep the country borders as is because they can be useful for localizing the features. We used some lighter grey shades where appropriate, though.
20. Page 16, line 18: The authors say that the model ‘agrees’ with the CombiPrecip precipitation. How did they come to this conclusion? The time evolution of the CombiPrecip appears to lie outside the ensemble spread for a large part of the timeseries implying poor agreement.
Please see the reply to comment 6.
21. Page 18, line 13: Why is the fact that the storm track for max precipitation in 20CR and downscaled simulations is similar ‘remarkable’? Did the authors expect to see large differences in the position of the storm track? Doesn’t the similarity indicate that the track of the cyclone is the primary control on precipitation accumulations over Switzerland?
We have removed the word in the revised manuscript.
22. Figures 7, 8 and 9: These figures are of very poor quality. They do not contain lat/lon, a colour bar or continent outlines. This makes the analysis impossible to follow.
Similar comments are made by all reviewers. We think that the loss of quality was introduced when converting from vector format to PNG, and during the upload process. In the revised version, we have a better resolution of the PNG file, and readability of the PDF for print is good in our view. Note also that the color key had been missing.
23. Page 19, lines 5-15: Analysis of figures 7-9 is described in just 13 short lines. Is it therefore justified to include all 18 figure sub-panels?

This is an eye-opening comment. We have realized that the analyses needs more details, and we have extended the paragraph accordingly. We think that it links also much better with the additional analyses suggested by Reviewer 1.

24. Figure 10: As far as I can tell both the colours and size of dots represent the precipitation intensity. Are both methods needed?

Yes, we prefer to keep both elements. The co-authors found in the internal review that comprehension of the Figure content is easier.

25. Page 20, lines 8-13: These lines describe figure 10. This is a complex diagram and the analysis of it is rather brief (6 lines). Given that one of the major objectives of the paper is to examine ‘sensitivity to cyclonic moisture flux’ (title) the analysis should be expanded.

We agree that the text is rather short at this point. In fact, we come back to Figure 10 in a later discussion paragraph. Following the suggestions also earlier in the review, we introduced a specific reference. Note that according to the comment by Reviewer 1, the analysis of ‘cyclonic moisture flux’ in the two innermost model domains is extended, including a new Figure 11.

26. Page 21, line 12: The authors describe the act that one of the ensemble members produces higher precipitation for the 1910 event than those observed as ‘remarkable’. I’m not sure why this is remarkable. The purpose of the ensemble is to represent the range of plausible situations given the large-scale flow conditions so if all of the ensemble members underpredicted the observed precipitation totals then this would be a poor ensemble. Perhaps I have misunderstood something here?

We can follow this argumentation. Hence, the word ‘remarkable’ may not be appropriate here, and we omit it in the revised version of the manuscript.

27. Page 23, lines 1-8: While the discussion of PV streamers is interesting, it is not a result of this paper, so it should not be in the results section.

Throughout the article, discussion and interpretation follow the presentation of results. However, we have rephrased the discussion without PV streamers. In accordance to a similar comment by Reviewer 3, we have rephrased the discussion without PV streamers. The aim of the paragraph is to mention the underlying physical processes and interaction between the Alpine orography and upper-level troughs. These physical processes can be referred to without introducing the notion of PV, which is indeed not introduced in our study. In fact, this has allowed to focus more on explaining the untypical results from the 1876 case.

The new text reads as follows: “This behavior of the model can be explained by the interaction of the Alpine orography with atmospheric circulation. Indeed, Vb cyclone trajectories are typically initiated by deepening upper-level troughs, which finally cut off from the westerly flow when passing over the Alps (e.g. Awan and Formayer, 2017). The interaction of upper-level troughs with the Alpine orography have been described in detail (Buzzi and Tibaldi 1978; Aebischer and Schär 1998; Kljun et al. 2001); the underlying processes include flow splitting and lee cyclogenesis, with further amplifications of the

cyclone formation by frontal retardation and latent heat release due to orographic lifting. The combination of these processes implies that the cyclones are formed on the lee of the right side of the Alps, typically over the Ligurian Sea. In 20CR, however, the Alpine orography is very coarse, smoothed and reaches only about 1000 m a.s.l. (cf. Stucki et al., 2012). Hence, the influence of the Alps on the large-scale flow is limited in 20CR. Given also that the 1876 case is least confined by pressure observations, this allows untypical cyclone tracks in many 20CR members. Once accounting for a more and more realistic orography throughout the downscaling steps with WRF, the high-resolution runs may thus end up in a compromise simulation - driven both by the WRF model physics and by the 20CR input flow. In other terms, the large-scale flow forced from 20CR might not be compatible with the orography of the high-resolution domains.”

28. Page 23, line 19: Whether short spin-up periods are ‘preferable over long spin-up periods’ depends on what you are trying to optimise and is not a general result. I think the objective in this study was to minimise spread in the ensemble tracks so as test sensitivity of precipitation to moisture flux rather than track position. Another objective may well have resulted in a different optimal spin-up period.

We agree. Accordingly, we introduced ‘for our purposes’.

29. Page 23, line 20: What are slow-reacting features?

We changed the line to ‘slow-reacting variables like soil moisture’. Note that we come back here to something we introduced in Section 2.3.

30. Page 23, line 20: Again ‘good results’ depends on what you are trying to achieve. Small differences in the ensemble will occur if the cyclones are already present in the outermost model domain. Is that the point?

We agree that we should be more specific here and expand on the suggestion of the reviewer in the revised manuscript.

31. Page 24, line 7: I do not think you can conclude that uncertainty increases gradually when going back in time using 3 case studies only.

In fact, we do not think that we generalize here, given the context of the paragraph, and the first word of the sentence being ‘This’, that is the uncertainty in our cases. Note however the added sentence in the conclusions: ‘Although going back far in time, we have only analyzed a very small number of events – many more cases would be needed to reach robust recommendations on how to configure a model for Vb cases.’

32. Page 24, line 9: Similarly, concluding that dynamical downscaling is less accurate going back in time is difficult using 3 case studies only. There are many other factors that would increase the uncertainty for specific case studies.

See the reply to comment 31.

33. Page 24, lines 14-23: This is an excellent summary and it would be nice to see a more in-depth analysis in the main body of the text to support these conclusions.

We tried to corroborate these conclusions with the extended analyses of the ‘cyclonic moisture flux’, including the new Figure 11 and an additional clause in the Abstract.

34. Page 24, line 26: How do you conclude that the 20CR tracks are not ‘realistically located’ for the 1876 case? Are you stating that 20CR produces unrealistic tracks, or simply that the uncertainty in the position of the track is large for this case potentially because it is a complex situation?

Realistic is indeed the wrong term here; we thank the reviewer for pointing this out. We rephrased it to ‘variability becomes very large’, and we add that some cyclone tracks do not follow the classical Vb path anymore. Note also the extended explanations of uncertainty.

Further changes made

- **Additional references: Hofstätter and Blöschl, 2019; Zbinden, 2005 (Annals of MeteoSwiss); Cioni and Hohenegger, 2019; Coppola et al., 2018; Compo et al, 2015; Joliffe and Stephenson, 2012; Wernli et al., 2008.**
- **Figure A2 redrawn with the maximum- and minimum-precipitation members. This makes more sense since we focus on these in the text.**

Review of nhess-2019-174:

“Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux”

by

Peter Stucki, Paul Froidevaux, Marcelo Zamuriano, Francesco Alessandro Isotta, Martina Messmer, Andrey Martynov

Recommendation: major revisions

The authors present three cases of extreme precipitation along the Alps associated with the classical synoptic situation Vb. They utilize downscaling of different reanalysis products for their analysis, where they employ WRF in a nested method to go down to 3 km resolution in the innermost domain. The authors test several sensitivities of their results with respect to the simulation setup, addressing both numerical as well as physical changes. They find that the lead time is among the most crucial parameters. While I find the manuscript well written, I struggle to see a clear motivation and conclusion for the study. The motivation is probably also somewhat difficult, because the manuscript tries to address many different questions at the same time: (1) What is the best downscaling setup for the cases in question. (2) What lead to the extreme precipitation. (3) Description of the three individual cases and comparison. The treatment of all these topics makes the paper sometimes difficult to follow. Regarding the conclusions, similar arguments apply and the authors sometimes appear to state opinions/speculations that are not necessarily solidly grounded in the material they presented. However, after responding to some of my major concerns, I believe this manuscript is acceptable for publication in NHESS.

We thank the reviewer for the comments and helpful, concrete suggestions for many aspects and important details. Many of the comments made us rethink our argumentation, and we found ways to better explain a topic or a reasoning based on this review. At times, this needed a new paragraph, at other times a couple of extra words were sufficient. With all these efforts taken, we think that the reviewer contributed largely to an enhanced, more specific and clearer manuscript. We hope that he / she will also more easily find our general line of thought in the revised manuscript - from first finding a setup, and then applying it to assess precipitation and moisture flux in the three cases.

General Comments:

In the title, I am not sure what the authors really mean with the “cyclonic moisture flux”. Is it a moisture flux going in a cyclonic direction or the moisture flux associated with a cyclone? Given this ambiguity, I encourage the authors to further clarify this aspect in their title as well as throughout the manuscript to help the reader making a clear link.

Indeed, the comment has made us re-think about what is meant by ‘cyclonic’ and ‘cyclonic moisture flux’. We come to the conclusion that both aspects (cyclonic rotation and association to a cyclone) are relevant in our case, although we mainly refer to it as ‘anti-clockwise’. To be clearer, we’ve added an explicit description of the specific dynamic mechanism of what we call a ‘cyclonic moisture flux’. This can be found in the Introduction section.

We also see that the term ‘cyclonic moisture flux’ is indeed not commonly used in publication titles or abstracts. However, we have not found a less specific, but still correct and concise term that would substitute it in the title. For this, and because it is also a relatively important term in our article, we would like to keep it there. We think that most readers may have some notion of what is meant when they read the title, and will be ready to read up about it in the introduction.

Given that there are only three cases, and the fact that a majority of the reasoning for the downscaling is based on the most recent case, the authors should be more cautious about general statements on downscaling procedures, as the results are highly sensitive to the case(s) at hand. While the technicalities that were overcome by the group are certainly impressive, it is still not clear to me how generic these results can be treated. In order to make a more general claim about the downscaling for Vb situation, one would need to explore many more cases to arrive at a firm conclusion. The authors should thus make it clear that this study can at most give an indication what one might need to test in order to arrive at a more general conclusion.

According to the reviewers’ suggestion, we have tried to avoid the impression of making general conclusions throughout the manuscript. This is mostly done by adding some words like ‘for our purposes’, ‘in our case’, ‘our experiments’, etc. We have also added a clear statement in the conclusions: ‘Although going back far in time, we have only analyzed a very small number of events – many more cases would be needed to reach robust recommendations on how to configure a model for Vb cases. Nevertheless, we have demonstrated that one can achieve a relatively best configuration for the desired application with a well-thought, logical series of experiments.’

What made the authors pick a 10-day spin-up time? It seems excessively long for the investigation of such a regional and meso-scale influenced precipitation event. At the end, the authors arrive at a 1-day spin-up time anyway, but the vastness of the parameter space is not sufficiently motivated, similar to some of the other sensitivity tests.

The list of sensitivities is extremely exhaustive, ranging from resolution to resolution ratios over spin-up time to parameterizations and model domains. The enormous parameter space is rather difficult to grasp and all results will primarily be in relation to

the 2005 case, with general deductions being rather limited due to the specifics of the case. In general, it would aid the reader if the authors more clearly state their working hypotheses as well as the reasoning for their choices and expectations. This will make it more straight forward to follow the ensuing arguments.

Based on these comments, see also the specific comment below, we now state more clearly in Section 3.1 what (i) our initial expectation was (10-day spin-up allows soil moisture and other variables to reach a partial equilibrium), (ii) that unsatisfactory first results led to experiments with reduced spin-up, and (iii) when the best setup was found, checking its robustness by testing additional changes of the model configuration. This reasoning is now also better underpinned with references.

Also, we state more clearly that ‘The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience’. In addition, we have introduced more literature and some short explanations of why we check against another (technical) configuration. In the Abstract, we focus on the result concerning the initialization time only because it is best documented in the manuscript.

The authors often refer to reproducing “correct” precipitation amounts. What is meant by correct? Presumably compared to observations, though the authors list several observations that are used. In addition, all of these “observations” also rely on some sort of downscaling and gridding, as data voids need to be filled. The authors, however, do not provide a detailed analysis of the representativeness of these observations. They refer to other studies that addressed these to some extent, but given the specifics of the case studies, the authors should also comment on the validity of the observations before comparing the model simulations to the data in order to claim “correctness”.

We think that the term ‘correct’ describes well what we are striving for in the experiments, and therefore prefer to keep it in the manuscript. To be more specific about what we mean by correct, we have tried to include a reference where possible throughout the manuscript. For instance, we insert ‘that would ideally produce’, ‘i.e., in the control area’, ‘compared to CombiPrecip’.

We are aware that the interpolations (as well as the original measurements) are affected by uncertainties and errors. The uncertainties and errors were analyzed in detail as described in the publications mentioned in the manuscript as well as in many applications (e.g. at MeteoSwiss). In-depth knowledge about the strengths and deficiencies of each datasets allows MeteoSwiss to make comparisons always considering the performance of the datasets used as reference. A detailed description of the representativeness of the observation for each case would replicate already published results and go beyond the scope of the present study. In section 2.1, we add a comment about our awareness about the uncertainties and errors of our datasets (and the consequential appropriate use).

The fact that more than one dataset is used is due to their availability (e.g. CombiPrecip is not available for the 1910 case) and the differentiated comparison with grid data and time series.

For the validation of, for example, precipitation, it has proved useful to use feature-based detections that consider location, shape, and timing. Why have the authors not considered more such verification tools for the study at hand? It appears the method referred to as EMD is in fact such a measure, though it appears confusing why the authors use a visual inspection for a quantitative comparison. The reasoning for the choices and omission of other tools should be clearly motivated.

The same as the reviewer, we see EMD as a kind of feature-based measure. In fact, we tested other measures, mostly using the spatialVx package for R (<https://ral.ucar.edu/projects/icp/references.html>), including SAL by Wernli and Schwierz, 2006. Advantages of the EMS were in our view: smaller number of subjective decisions (thresholds, smoothing options etc.), one instead of three measures, no failures of actual feature detection, a measure for the relative distribution of precipitation totals. As the reviewer suggests, we have summarized these reasons in the text.

Regarding the ‘eyeball inspection’, it seems to be a common and valid option and complement to machine techniques for pattern recognition, see

**https://www.cawcr.gov.au/projects/verification/#Standard_verification_methods
<https://www.swpc.noaa.gov/sites/default/files/images/u30/Spatial%20Forecast%20Verification.pdf>**

We do agree on the point that more reasons are required to justify the selection of measures and scores, though. Therefore, we have introduced some short justifications in Section 3.2.

For the philosophical concluding paragraph on page 25 not much hard evidence has been provided in the manuscript for the claims put forward. It thus reads more like a written piece of opinion than a well and quantitatively justified conclusions.

We suppose that the reviewer alludes to the use of ‘propositions of possible weather’. We see this rather as an intuitive and informative description of what the ensemble provides in the end, and think this could be appropriate in a final paragraph. Apart from that, we do not see any other philosophical conclusions on page 25. We rather think that we describe options that have been tested already, or paths that could be taken to balance computational costs and meaningfulness of downscaling an ensemble.

Specific Comments:

The page (P) and line (L) numbers refer to the ones in the manuscript.

P1 L23: “to the cyclonic” and see comment above about the ambiguity of “cyclonic moisture flux”.

Please refer to our reply to the general comments above.

P1 L28: “accurate directions” with respect to what? What is the reference?

We agree that ‘accurate’ is too vague here and have changed it to ‘northerly’.

P8 L1: The precipitation data is interpolated. Can the authors please clarify if the interpolation was carried out in such a way that the total precipitation was unaffected by

the interpolation? Depending on the kind of interpolation, the results for the totals can deviate.

Please refer to our reply to the general comments above (fourth comment). We added some sentences in Section 2.1 to illustrate our awareness about the interpolated datasets.

P11 L12: The authors speculate on the differences between ERA-Interim and 20CR in terms of moisture distribution, though the authors could provide direct evidence for their claim by investigating differences between ERA-Interim and 20CR fields for the case at hand in more detail.

Following this comment, we have added some more information about a short investigation of specific moisture at 1000 hPa.

P11 L22: Why did the authors chose to identify and track cyclones using geopotential at 500 hPa? This seems rather unconventional and needs further motivation.

We agree that it is a good idea to extend the explanation of our choices. Concretely, we introduce the tracking paragraph with: ‘For the analyses, we use both sea level pressure (SLP) and mid-tropospheric pressure fields. SLP fields inform about the quality of the assimilation process in 20CR, and the isobaric pressure fields (at 500 hPa here) tell about the derivation of upper-air variables from the SLP information in 20CR. Combining SLP and isobaric levels has been found useful for cyclone tracking (Hofstätter and Blöschl, 2019).’

Fig. 4: The cyclone tracks for ERA-Interim look very edgy. In order to compare them better to the other plots, a grid that is not fixed to the grid spacing of the data could be beneficial, which is most often done in other cyclone track algorithms, see also comment above about cyclone track determination in this manuscript.

We agree on the ‘edginess’, but would prefer to leave the visualization as is in Figure 4d. In our view, it reflects the nature (the coarse resolution) of the underlying data; a smoothed version would rather mask it. In addition, we see the ERAI line as a rather complementary information in the plot. Note also that for the same reason, we did not interpolate the grid points in Fig. g – i, and included the original grid points in the new Figure 11b, d and f. For other variables however, we see that smoothing is necessary, e.g. for cyclone tracks or SLP.

Fig. 7, 8, and 9: I find these figures not very legible. Maybe this is due to the downgrading of the figure quality for the review process, but otherwise the readability of the information of these figures needs to be significantly improved. In particular the arrows are not very visible.

We fully agree on that, and all reviewers are of the same opinion. Unfortunately, the quality loss was introduced by the conversion to PNG format. This should mostly be enhanced with the current PNG, and the PDF for print is fine as well, in our view. Note also that the color key had been missing.

P22 L21: The authors should explain how PV is produced in the downscaling process, as

this appears to be crucial in their arguments.

In accordance to a similar comment by Reviewer 2, we have rephrased the discussion without PV streamers. The aim of the paragraph is to mention the underlying physical processes and interaction between the Alpine orography and upper-level troughs. These physical processes can be referred to without introducing the notion of PV, which is indeed not introduced in our study. In fact, this has allowed to focus more on explaining the untypical results from the 1876 case. The new text reads as follows:

“This behavior of the model can be explained by the interaction of the Alpine orography with atmospheric circulation. Indeed, Vb cyclone trajectories are typically initiated by deepening upper-level troughs, which finally cut off from the westerly flow when passing over the Alps (e.g. Awan and Formayer, 2017). The interaction of upper-level troughs with the Alpine orography have been described in detail (Buzzi and Tibaldi 1978; Aebischer and Schär 1998; Kljun et al. 2001); the underlying processes include flow splitting and lee cyclogenesis, with further amplifications of the cyclone formation by frontal retardation and latent heat release due to orographic lifting. The combination of these processes implies that the cyclones are formed on the lee of the right side of the Alps, typically over the Ligurian Sea. In 20CR, however, the Alpine orography is very coarse, smoothed and reaches only about 1000 m a.s.l. (cf. Stucki et al., 2012). Hence, the influence of the Alps on the large-scale flow is limited in 20CR. Given also that the 1876 case is least confined by pressure observations, this allows untypical cyclone tracks in many 20CR members. Once accounting for a more and more realistic orography throughout the downscaling steps with WRF, the high-resolution runs may thus end up in a compromise simulation - driven both by the WRF model physics and by the 20CR input flow. In other terms, the large-scale flow forced from 20CR might not be compatible with the orography of the high-resolution domains.”

P23 L24: How can the authors conclude that “nudging smaller domains can still be beneficial”? Has any evidence been provided in this study to support such a claim?

We agree that this could be understood as a claim. Therefore, we rephrased the sentence to: ‘Although we find no relevant enhancements from nudging in smaller domains in our test experiments, nudging smaller domains could still be beneficial for other specific studies.’

P24 L1: The authors should be more specific what they are referring to with “traditional spatial verification scores”.

Regarding traditional (or standard) methods, we referred to <https://www.wmo.int/pages/prog/arep/wwrp/new/jwgfvr.html>, among others. However, we see that the term might be confusing, and substituted it by 'more common', and we name our measures again.

Further changes made

- **Additional references: Hofstätter and Blöschl, 2019; Zbinden, 2005 (Annals of MeteoSwiss); Cioni and Hohenegger, 2019; Coppola et al., 2018; Compo et al, 2015; Joliffe and Stephenson, 2012; Wernli et al., 2008.**
- **Figure A2 redrawn with the maximum- and minimum-precipitation members. This makes more sense since we focus on these in the text.**