Evaluation of Alpine-Mediterranean precipitation events in convection-permitting regional climate models using a set of tracking algorithms

Sebastian K. Müller, Cécile Caillaud, Steven Chan, Hylke de Vries, Sophie Bastin, Ségolène Berthou, Erwan Brisson, Marie-Estelle Demory, Hendrik Feldmann, Klaus Goergen, et al.

Evaluation of Alpine-Mediterranean Precipitation Events in convection-permitting Regional Climate Models using a Set of Tracking Algorithms

Sebastian K. Müller[1*†], Cécile Caillaud[2†], Steven Chan[3†], Hylke de Vries[4†], Sophie Bastin[5], Segolene Berthou[6], Erwan Brisson[7], Marie-Estelle Demory[8], Hendrik Feldmann[9], Klaus Goergen[10], Stergios Kartsios, Petter Lind[12], Klaus Keuler[13], Emanuela Pichelli[1], Mario Raffa[14], Merja H. Tölle[15] and Kirsten Warrach-Sagi[16]

[1*]ESP, Earth System Physics, ICTP, The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, Trieste, 34100, Friuli-Venezia-Giulia, Italy.
[2]CNRM, Université de Toulouse, CNRS, Météo-France.
[3]School of Engineering, University of Newcastle.
[4]KNMI, Royal Netherlands Meteorological Institute.
[5]IPSL, Institut Pierre Simon Laplace.
[6]MOHC, Met Office Hadley Centre.
[7]GUF, Goethe University Frankfurt.
[8]ETHZ, ETH Zürich.
[9]KIT, Karlsruhe Institute of Technology.
[10]FZJ, Institute of Bio- and Geosciences (IBG-3, Agrosphere), Forschungszentrum Jülich.
[11]AUTH, Aristotle University of Thessaloniki.
[12]SMHI, Swedish Meteorological and Hydrological Institute.
[13]BTU, Brandenburg University of Technology.
[14]CMCC, Euro-Mediterranean Center on Climate Change.
[15]Universität Kassel, Center for Environmental Systems Research (CESR), Kassel Institute for Sustainability.
[16]UHOH, University of Hohenheim.

Contributing authors: sebastiank.mueller@posteo.net;
†These authors contributed equally to this work.

Keywords: convection-permitting modelling, heavy precipitation events, tracking algorithms

We here inter-compare four different tracking algorithms by applying them onto the precipitation fields of an ensemble of convection-permitting regional climate models (cpRCMs) and on high-resolution observational datasets of precipitation. The domain covers the Alps and the northern Mediterranean and thus we here analyse heavy precipitation events, that are renowned for causing hydrological hazards. In this way, this study is both, an inter-comparison of tracking algorithms as well as an evaluation study of cpRCMs in the Lagrangian frame of reference.

The tracker inter-comparison is performed by comparison of two case studies as well as of climatologies of cpRCMs and observations. We find that that all of the trackers produce qualitatively equal results concerning characteristic track properties. This means that, despite of quantitative differences, equivalent scientific conclusions would be drawn. This result suggests that all trackers investigated are reliable analysis tools of atmospheric research.

With respect to the model ensemble evaluation, we find an encouraging performance of cpRCMs in comparison to radar-based observations. In particular prominent hotspots of heavy precipitation events are well-reproduced by the models. In general most characteristic properties of precipitation events have positive biases. Assuming the under-catchment of precipitation in observations in a domain of such complex orography, this result is to be expected. Only the mean area of tracks is underestimated, while their duration is overestimated. Mean precipitation rate is estimated well, while maximum precipitation rate is overestimated. Furthermore, geometrical and rain volume are overestimated. We find that models overestimate the occurrence of precipitation events over all mountain chains, whereas over plain terrain in summer precipitation events are seen underestimated. This suggests that, despite the convection-permitting resolution, thermally driven thunderstorms are either not triggered or their dynamics still under-resolved. Eventually we find that biases in the spatio-temporal properties of precipitation events appear reduced when evaluating cpRCMs against Doppler radar-based and rain gauge-adjusted observational datasets of comparable spatial resolution, strengthening their role in evaluation studies.

1 Introduction

Moist deep convection in the atmosphere (Stevens, 2005), manifesting in storms of all scales, is responsible for the most severe precipitation events. However, as convection is by nature scarce in space and time, it is challenging to describe its properties, being fluxes of heat, momentum and water, appropriately. It is thus common to estimate convective heavy or extreme precipitation by e.g. the 99th or 99.9th percentile of hourly precipitation rate, or, to apply thresholds on precipitation rate and the frequency of their exceeding is then representative of the frequency of extreme events (Ban et al., 2020; Pichelli et al., 2021). However, statistical analyses in the Eulerian frame of reference remain limited to the description of the time series of each grid cell separately, and no information is retained about the underlying events and their spatial structure. Instead, information about the convective events themselves can be yielded through the application of a tracking algorithm, here referred to simply as tracker. By this, precipitation events are identified and tracked in time, meaning the analysis is transferred from the Eulerian into the Lagrangian frame of reference. Through the use of trackers the precipitation events themselves and their properties are in focus, rather than conditions at specific locations. Many studies (Prein et al., 2017a; Crook et al., 2019; Purr et al., 2019; Guo et al., 2022) have shown the benefit of this idea.

Tracking algorithms were originally developed in order to evaluate precipitation events in numerical weather predictions (Davis et al., 2006a,b; Wernli et al., 2008; Johnson et al., 2013). By the use of a tracker modelled precipitation objects can be compared against observations regarding their spatial structure, intensity, propagation and location. Any model output or observational field that is typically associated with a precipitation event may serve as the tracker input field. Although typically precipitation rate itself is used, also indirect proxies, like outgoing longwave radiation (Morel and Senesi, 2002a; Chen et al., 2019), or mid-level vertical velocity or vorticity, are suitable. On even finer resolutions individual updrafts of convective systems can be analyzed, along with their merging and splitting dynamics (Moseley et al., 2013). Another important application of tracking algorithms is the detection and observation of tropical and extra-tropical cyclones (Neu et al., 2013). Furthermore also droughts are operationally monitored using trackers (Abatzoglou et al., 2017).

The dynamical downscaling approach allows for investigating the impact of climate change on local scales and to derive actionable information for a variety of sectors (Giorgi, 2019, 2020). During the last decades convection-permitting Regional Climate Models (cpRCMs) were established, solving the non-hydrostatic equations of the atmosphere on grids with horizontal grid spacing smaller than 4 km and allowing to turn off flawed parameterizations of deep convection (Prein et al., 2015). At first, year-long integrations of distinct regions were realized (Grell et al., 2000), then decade-long integrations (Rasmussen et al., 2011) and decade-long integrations of entire continents (Prein et al., 2017a), and recently robust ensembles of cpRCMs (Coppola et al., 2020;

Kendon et al., 2021) have been achieved. cpRCMs brought great advances and still need further exploration of their capabilities (Lucas-Picher et al., 2021): significant added value (Rummukainen, 2016; Ciarlo et al., 2020) lies in the representation of precipitation, in particular regarding its diurnal cycle, intensity and extremes (Ban et al., 2015; Kendon et al., 2017; Fumière et al., 2020; Reder et al., 2022) and over complex orography (Reder et al., 2020; Adinolfi et al., 2020). However, due to a more realistic orography, improvements are also found concerning surface temperature (Hohenegger et al., 2008) and mesoscale wind systems (Belušić et al., 2018). Through the application of trackers on cpRCMs, the climate change signal of convective storms can be analyzed in great detail (Prein et al., 2017a,b; Purr et al., 2019). Along with advances in model development, novel observational precipitation datasets, based on Doppler radar measurements, providing comparable spatial and temporal resolution emerged and allow for a rigorous evaluation of cpRCMs. Still, their impact on the evaluation of cpRCMs must be considered carefully (Prein and Gobiet, 2017d).

This present study uses an ensemble of cpRCMs, developed by the CORDEX - Flagship Pilot Study on *Convective phenomena at high resolution over Europe and the Mediterranean* (CORDEX-FPSCONV, Coppola et al. (2020)), conducted on a domain covering the Alps and the northern Mediterranean Sea. This region is renowned for its severe precipitation events (Drobinski et al., 2014) and for being a climate change hotspot (Giorgi, 2006). Several research groups set up trackers in order to evaluate the behaviour of cpRCMs in simulating precipitation events and storms. This study takes advantage of this opportunity and carries out an inter-comparison study on a set of four trackers. They will be applied both to the model ensemble's *evaluation* runs, driven by ERA-Interim reanalysis data, and to a composite of high-resolution observational datasets. Thus the scientific objective of this study is two-fold:

1. Tracker-Inter-comparison: we inter-compare four different trackers in order to find out about their reliability: do different trackers yield the same scientific conclusions?
2. Model-Evaluation: we evaluate an ensemble of convection-permitting regional climate models against observations in the Lagrangian frame of reference by using the trackers, and focus on the following aspects:
   - how good are cpRCMs at simulating the basic properties of precipitation events (intensity, spatial and temporal scales, rain volume)?
   - how good are cpRCMs at simulating the spatial patterns and the annual cycle of basic properties of precipitation events?

The paper is organized as follows. In section 2 we introduce the model ensemble and observational datasets used, as well as the two historic events that serve as case studies. In section 3 we explain the workflow of the tracking algorithms and motivate the setup chosen in order to identify the precipitation events of interest. In section 4 we present our results concerning the tracker

inter-comparison and in section 5 we present results on the model ensemble evaluation. Finally we summarize our findings and give conclusions in section 6.

2 Model Ensemble, Observational Datasets and Historical Events

In this section we briefly describe the CORDEX-FPSCONV ensemble of cpRCMs as well as the composite of datasets of precipitation measurements that in this study serve the input fields for the tracker analyses. Moreover we here introduce two historical heavy precipitation events, that we use as case studies for the tracker inter-comparison.

2.1 Ensemble of convection-permitting Regional Climate Models

The CORDEX-FPSCONV community produced a first-of-its-kind ensemble of cpRCMs for the domain studied herein, that is covering the Alps and the northern Mediterranean (Coppola et al., 2020). Its Eulerian evaluation of precipitation is found in (Ban et al., 2020) and (Pichelli et al., 2021), which we here build upon and extend into the Lagrangian frame of reference. Importantly both studies demonstrate how the cpRCMs reduce model biases in comparison to the driving RCMs. Therein may also be found detailed information on the models. We here analyze the *evaluation* runs, whose boundary conditions are derived from ERA-Interim reanalysis data, through intermediate driving simulations at coarser resolution (RCM) (Ban et al., 2020). The ensemble contains several members using the COSMO-CLMcom and WRF model, which differ in their nesting strategy and physics parameterizations respectively. The model ensemble is summarized in Table 1.

Institute	cpRCM	dx(cpRCM)[km]	Driving RCM	dx(RCM)[km]	RCM domain
AUTH	WRF381BJ (A)	3	WRF	15	EURO-CORDEX
FZJ	WRF381BB	3	WRF	15	EURO-CORDEX
IPSL	WRF381BE (A)	3	WRF	15	EURO-CORDEX
UHOH	WRF381BD	3	WRF	15	EURO-CORDEX
BTU	COSMO-CLM (B)	3	COSMO-CLM	12	EURO-CORDEX
CMCC	COSMO-CLM (B)	3	COSMO-CLM	12	EURO-CORDEX
GUF	COSMO-CLM (B)	3	COSMO-CLM	12	Med-CORDEX
JLU	COSMO-CLM (B)	3	ERAINT	-	-
KIT	COSMO-CLM (B)	3	COSMO-CLM (B1)	25	Europe
ETHZ	COSMO-pompa_5.0 (C)	2.2	COSMO-CLM	12	Europe
CNRM	CNRM-AROME41t1 (C)	2.5	CNRM-ALADIN62 (C1)	12	Med-CORDEX (spectral nudging)
HCLIM-Com	HCLIM38-AROME (D)	3	ALADIN62	12	Europe
KNMI	HCLIM38-AROME (D)	2.5	RACMO	12	Europe
ICTP	RegCM4 (E)	3	RegCM4 (A)	12	Europe
UKMO	UM (F)	2.2	ERAINT	-	-

Table 1 Summary of numerical models used in this study. "dx" denotes the grid spacing of the respective model. Model documentation references: (A) (Powers et al., 2017); (B) (Rockel et al., 2008; Baldauf et al., 2011); (B1) (Keuler et al., 2016); (C) (Caillaud et al., 2021); (C1) (Nabat et al., 2020); (D) (van Meijgaard et al., 2008; Van Meijgaard et al., 2012; Belušić et al., 2020); (E) (Coppola et al., 2021); (F) (Chan et al., 2020)

181

Prior to the tracker analysis we remapped each of the models from their native grid onto the analysis grid *ALP-3i*, using distance weighted average remapping. It is a "regular lat-lon grid", spanning in longitude from 1°E to 17°E in 582 grid cells, and in latitude from 40°N to 50°N in 364 grid cells. This results in a grid spacing of 0.0275° in both latitude and longitude, which translates on average to about 3 km.

2.2 High-resolution observational datasets of precipitation

We use a composite of four observational datasets of precipitation covering France, Germany, Switzerland and Italy respectively, over a common time period from 2001 to end of 2009. Their original spatial resolution is comparable to that of the convection-permitting models, with native grid spacings ranging from 1 to 3 km, and their temporal resolution is hourly. Thus the observational datasets can be neatly compared to hourly precipitation rates of the models. All of the datasets except of one (GRIPHO) are based upon Doppler radar measurements and adjusted with rain gauge measurements. The spatial and temporal resolution of these datasets are the highest currently available for the respective regions. Still, Doppler radar observations are known to systematically underestimate precipitation amounts over mountainous terrain, e.g. through the *shielding* effect (Germann et al., 2022), and underestimate particularly heavy precipitation rates (Schleiss et al., 2020). Further, also rain gauges *under-catch* orographic precipitation and are moreover affected by windy conditions (La Barbera et al., 2002). Furthermore, interpolation methods used to map station data onto regular grids induce an underestimation of high intensities (smoothing effect) and an overestimation of low intensities (moist extension into dry areas) (Isotta et al., 2014). A brief summary of the individual datasets, including their specific spatial resolution and references is given in Table 2.

Abbreviation	Coverage	dx [km]	Reference
COMEPHORE	France	1	(Tabary et al., 2012; Fumière et al., 2020)
RADKLIM_RW	Germany	1	(Winterrath et al., 2018)
RdisaggH	Switzerland	2	(Wüest et al., 2010)
GRIPHO	Italy	3	(Fantini, 2019)

Table 2 Summary of observational datasets of hourly precipitation rate used in this study. dx [km] denotes the original grid spacing.

Prior to the tracker analysis each of the datasets was remapped onto the analysis grid *ALP-3i*, again using distance weighted average remapping. Then we merged them and use their arithmetic mean for regions along the borders

of the nations, where measurements overlap. In this way, both the observations and the models, were mapped onto the same grid.

GRIPHO over Italy and posteriori masking

We here inform about two shortcomings of our analysis and show how we deal with them when interpreting the results.

Firstly, the observational dataset covering Italy, GRIPHO, is based on quality-controlled rain gauge measurements solely. The station density is greater in the north than in the south of Italy, and on average it is estimated to about 1 per $9\times9\,\mathrm{km}^2$. It is then remapped onto a convection-permitting grid with a 3 km grid spacing. In comparison to that, the other datasets are based upon Doppler radar measurements and rain gauges and their original spatial resolution is even finer than that of the analysis grid. We here must expect differences in the spatio-temporal characterization of the precipitation field observed by GRIPHO with respect to the other datasets. Nonetheless, GRIPHO is the most accurate observational dataset available for Italy and in particular the representation of extreme events was found improved (Fantini, 2019), especially over Northern Italy where the station density is higher and where the most extreme precipitation events occur. In terms of domain complexity though, we note that Italy is surrounded by the Mediterranean Sea and the Alps, intersected by the Apennine Mountains and further shows both steep coastlines and a large plain area (Po Valley). Due to this high degree of complexity, which translates into very complex and local interactions, the precipitation events are renowned for being particularly severe and their modelling particularly challenging (see e.g. Morgan, 1973; Buzzi and Alberoni, 1992; Medina and Houze Jr, 2003; Rotunno and Houze, 2007; Panziera et al., 2015; Miglietta et al., 2016; Pichelli et al., 2017).

Secondly, the observations do not cover the entire domain simulated by the models, in particular not the Mediterranean sea. We consider this by *posteriori* applying a mask onto the tracker analyses of models, meaning that only tracks whose centroid is located within the domain of the observations are considered. This implies that in models events entering the observational domain and here particularly those making landfall, are expected to be overestimated in their scales, but little in their averaged properties. Note further that for the Swiss dataset RDisaggH, there is no data available for the period up to June 2003, which is also accounted for through masking.

We account for both of these two shortcomings by presenting the relative biases of the model ensemble not only for the entire domain of observations, but also separately without GRIPHO as well as for GRIPHO exclusively, which we consider representative of the most extreme precipitation events within the domain. By this we account for and understand both, the specific model biases due to the complex Italian domain as well as specific biases associated with the GRIPHO drawbacks. Further, by doing so the overestimation of landfalling tracks can be estimated, because by excluding GRIPHO we also exclude the greatest part of the coastline.

2.3 Historical Heavy Precipitation Events

In the following we introduce the two historical heavy precipitation events that share these characteristics: both occurred along the Mediterranean coastline, both regions affected show steep orographic features and both happened in autumn. Coincidentally they both occurred along the same degree of latitude and the one happened just a little more than one year after the other.

2.3.1 Gard, France in September 2002

The first case study is a heavy precipitation event that occurred in south-eastern France, in the Gard region, during the 8th and 9th of September 2002 (Delrieu et al., 2005; Chancibault et al., 2006). Lasting more than a day, the event was particularly remarkable due to its rain amounts greater than 200 mm within 24 h spread over an area of 5500 km^2. The maximum rain rates of 600–700 mm observed locally by rain gauges are among the highest daily records in the region. The propitious slow-evolving synoptic-scale situation combined an upper-level south-westerly diffluent flow over south-eastern France with a moist and warm low-level south-easterly flow. The rain event can be characterized by three phases (Delrieu et al., 2005): at first a Mesoscale Convective System (MCS) developed over the Gard plains, second a displacement of the MCS toward the Cévennes mountain ridge took place and third, the passage of a cold front with embedded convection swept the convective activity out of the region. This catastrophic event resulted in 24 fatalities and an economic damage estimated at 1.2 billion €(Sauvagnargues-Lesage, 2004).

2.3.2 Carrara, Italy in September 2003

A second case study we carry out by looking at a heavy precipitation event that happened in Carrara, Italy, in September 2003, and which caused severe flash flooding and landslides. Cortopassi and Daddi (2008) investigated how the extensive quarrying activities of the region destabilize the terrain and promote hydro-geological hazards. It may be described as a landfalling convective system. A trough extending from a well structured low pressure centered over Northern Europe, advected hot and humid air from the Mediterranean sea and provided large-scale lifting. At the steep orography of the Apuan Alps the convective instability was triggered and the propagation of the storms was blocked. As a consequence, the region was affected by torrential rain, accumulating up to about 200 mm within a period of only 2 hours. The event claimed two fatalities and caused major damage to the local infrastructure.

3 Trackers

We here describe the basic functionality of the four tracking algorithms investigated in this study, which are referred to as MODE, OSIRIS, DYMECS and celltrack. Their functionalities are summarized in Table 5 and we provide a detailed description of each tracker in the appendix (Section 8). The trackers are completely independent developments and are here applied with setups that are as similar as possible, in order to compare the same precipitation objects. In Section 4, we compare the trackers individually against each other, while in Section 5, we evaluate the model ensemble by the mean of all four trackers, which we refer to as the "tracker ensemble mean".

The 1-hour accumulated precipitation fields (from observations or models) are used as input for all trackers. The principle operations of all trackers investigated include a first step of masking through a specified threshold, followed by a step of clustering in space to form objects and then tracking of those in time to form tracks. Prior to that the input field is smoothed in space. The treatment of cell merging and splitting is done by the allocation of *metatracks*, which can be understood as the smaller branches of merged or split tracks. It is a functionality that is not available in all trackers.

We designed the tracker setup such that it is able to identify precipitation events, that cause high impact weather situations like flash floods. For this reason we chose the (relatively large) precipitation threshold of $5\,\mathrm{mm\,h^{-1}}$. On the other hand we want to investigate the small-scale isolated thunderstorms that cpRCMs are capable of resolving, and to this end, we chose a (relatively small) minimum space-time volume threshold of 100 cells. The input field is smoothed across 3×3 grid cells prior to the analysis.

The common tracker setup is summarized in Table 3:

spatial smoothing	3×3 grid cells
precipitation threshold	5 mm h^{-1}
minimum space-time volume threshold	100 grid cells

Table 3 Summary of the tracker setup.

The characteristic track properties we are investigating in this study are defined in Table 4, with pr describing the precipitation field of a track:

Track Property	Definition
N_T	the number of tracks T identified
OF [$time^{-1}$]	the *Occurrence Frequency*, defined as the number of tracks identified per unit time
OFD [$time^{-1}$ $area^{-1}$]	the *Occurrence Frequency Density*, defined as the number of tracks identified per unit time and unit area
$\overline{pr}^c$ [mm h^{-1}]	the *Mean Precipitation Rate* of a track, with $\overline{\ }^c$ being the mean over all grid cells associated with a track
$max^c(pr)$ [mm h^{-1}]	the *Maximum Precipitation Rate* of a track, with max^c being the maximum over all grid cells associated with a track
D [h]	the *Duration* of a track. (A track occurring only for a single time step will be attributed with 1 hour of duration.)
P_T [m^3]	the *Rain Volume* of a track, given by the integration of its precipitation field
$\overline{A}^D$ [km^2]	the *Mean Area* of a track, averaged over its *Duration*, *D*
Vol [km^2 h]	the *Space-Time Volume* of a track, given by integrating the area of all grid cells, in space and time

Table 4 Definitions of characteristic track properties.

Institute Tracker	space/time smoothing	metatrack at splitting/merging	statistics	boundary treatment	reference
ICTP MTD	yes/yes	no/no	original	smoothed	(Clark et al., 2014)
CNRM OSIRIS	yes/no	yes/yes off/off	smoothed	no	(Morel and Senesi, 2002a)
UKMO DYMECS	yes/no	yes/yes on/on	original	smoothed	(Stein et al., 2014)
KNMI celltrack	no/no	yes/yes off/off	smoothed	no	(Moseley et al., 2013) (Lochbihler et al., 2017)

Table 5 Summary of Trackers investigated in this study: "Institute" denotes the group executing the analysis using "Tracker", representing the abbreviation of the respective tracker. "space/time smoothing" denotes whether the tracker has an option for smoothing the input field prior to the analysis in space or time. "metatrack at splitting/merging" denotes whether the tracker assigns separate *metatracks* when tracks merge or split, with the second line indicating whether this functionality is switch on or off. Column "statistics" indicates whether the statistical properties of the tracks are derived from the original or from the smoothed input field. In "reference" the original description of the tracking algorithm is found.

4 Results on Tracker Inter-Comparison

In this section we inter-compare the four trackers in two steps: first, we compare their performance at analyzing the two historic events, Gard 2002 and Carrara 2003 (subsection 4.1), and second, we compare their climatological properties, derived from the tracker-analyses of the entire 9-year periods of observations and model ensemble (subsection 4.2).

4.1 Tracker Inter-Comparison using Case Studies Gard 2002 and Carrara 2003

We apply the four trackers on the observational dataset and investigate only the region and time periods of the respective historic events. In Figure 1, we show for both events the accumulated total precipitation along with the location of tracks and their respective rain volume. Note that we do not show the full path of propagation for the tracks, as for the stationary precipitation systems investigated here, the paths of propagation appear erratic and the information added does not bring relevant insight. In general we recommend that propagation features of multi-celled convective systems (e.g. distance travelled or propagation velocity) must be interpreted with caution, as the correct identification of their center is challenging. In Table 6, we summarize the properties of all tracks associated with the two historic events.

We first note, that all of our trackers do identify both historic events and attribute several tracks to them. OSIRIS and particularly DYMECS identify more tracks than MTD and celltrack, which can be explained for DYMECS by the allocation of *metatracks* in case of track merging or splitting. We find that the number of tracks identified reflects in the sum of duration and sum of mean track area. For each event, one main track is responsible for the major

Tracker	Gard 2002 / Carrara 2003						
	N_T [−]	$\max^T \overline{pr}^c$ [mm h^{-1}]	$\max^T {}^c(pr)$ [mm h^{-1}]	$\sum^T P_T$ [m^3 e6]	$\sum^T \overline{A}^D$ [km^2 e3]	$\sum^T D$ [h]	$\sum^T Vol$ [km^2 h e3]
MTD	3/4	11.0/9.0	96.0/57.6	5046/210	12/4	44/19	402/20
OSIRIS	10/7	8.8/6.9	85.0/52.0	3396/172	17/7	65/27	283/22
DYMECS	13/6	8.8/7.7	96.0/57.6	4476/240	25/7	141/20	335/24
celltrack	3/4	9.7/7.9	88.9/52.7	4587/190	7/5	45/19	415/22

Table 6 Averaged and integrated track properties, as defined in Table 4, associated with historic events Gard 2002 and Carrara 2003. $\max x^T$ and $\sum x^T$ denote the maximum value and summation of property x over all tracks T.

part of the rain volume and all the trackers agree well on these most severe tracks. Focusing on intensities, lower intensities with OSIRIS can be explained by the calculation of the diagnostics on the smoothed precipitation field.

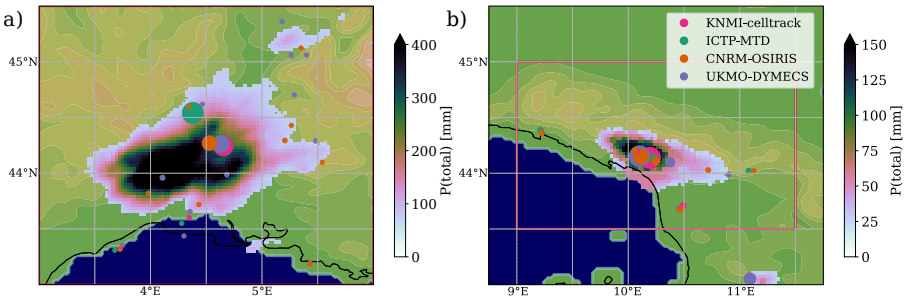

Fig. 1 Historical events Gard 2002 (left panel) and Carrara 2003 (right panel) using the observations. Shading illustrates accumulated total precipitation, P(total) [mm]. Filled circles indicate the location of the centroid of a track and their radius is proportional to their respective rain volume. Filled contours indicate the elevation of the model terrain in intervals of 250 m.

348 Still overall and as listed in Table 7, we find that all trackers agree on the following relations:

Compared to the Carrara 2003 case, the Gard 2002 case is greater in all scales: it is longer-lived, larger in area, and consequently also in Space-Time volume.
Compared to the Carrara 2003 case, the Gard 2002 case is more intense: i.e., its mean and maximum precipitation rates are greater.
As a consequence of both greater scales and intensity, the integrated rain volume of the Gard 2002 case is much greater than that of the Carrara 2003 case.

Table 7 Qualitative attribution of track properties to the two historical events, Carrara 2003 and Gard 2002.

349

350 Thus, all trackers describe the two events with equivalent track properties and
351 moreover the properties attributed and the relations found agree well with
352 the description of the events in literature: the smaller and less intense event,
353 Carrara 2003, is attributed with smaller spatial scales and less intensity than
354 the larger and more intense event, Gard 2002. Based upon these results, the
355 choice of the trackers seems irrelevant, meaning that from any of the trackers'
356 analyses, equivalent scientific conclusions would be derived. In other words,
357 with the differences being only of quantitative nature, the scientific conclusions
358 are found to be independent of the choice of the tracker.

Tracker	Model Ensemble / Observations						
	N_T [year^{-1}]	$\overline{\overline{pr^c}}^T$ [mm h^{-1}]	$\overline{\overline{max^c(pr)}}^T$ [mm h^{-1}]	$\overline{\overline{P_T}}^T$ [m^3 e6]	$\overline{\overline{A^D}}^T$ [km^2 e3]	$\overline{\overline{D}}^T$ [h]	$\overline{\overline{Vol}}^T$ [km^2 h e3]
MTD	2742/2536	8.3/7.8	21.3/17.3	74/56	1.2/1.6	4.4/3.1	8.3/6.8
OSIRIS	3347/2911	7.0/6.7	16.1/13.9	43/40	1.1/1.5	4.1/3.0	5.6/5.6
DYMECS	5095/3618	7.9/7.4	19.6/16.1	43/36	1.3/1.6	3.6/2.7	4.7/4.4
celltrack	2953/2550	7.7/7.3	17.2/14.7	72/54	1.1/1.6	4.5/3.1	8.8/7.1

Table 8 Climatology of track properties derived from both the model ensemble and the observations. In Table 4, definition of the respective properties are given.

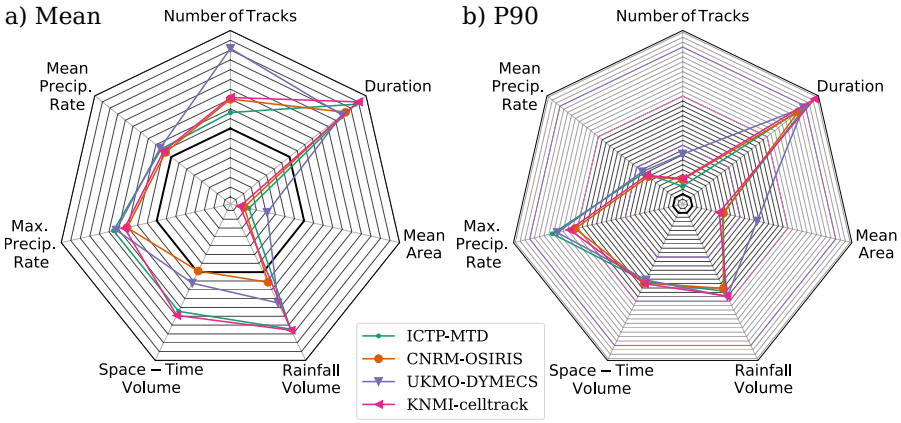

Fig. 2 Relative bias in characteristic of a) the mean and b) the 90th percentile of track properties of each tracker with respect to the observations $\frac{ModEns - Obs}{Obs}$ [%]. Black solid lines are increments of +-5%, with the thick black line representing the tracker ensemble mean of the observations, i.e. 0%.

4.2 Tracker Inter-Comparison using the Climatologies of Model Ensemble and Observations

We continue the tracker inter-comparison over climatological scale through the ensemble of cpRCM simulations and the observations for the common time period 2001-2009. Table 8 shows the climatological means of characteristic track properties of each tracker and, in Figure 2, we show the relative biases of each tracker for the mean and 90th percentile of track properties with respect to the observations.

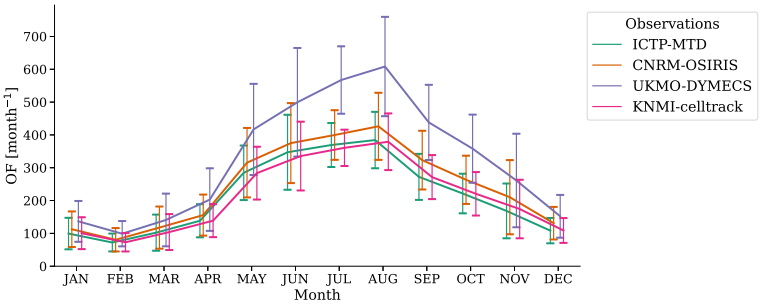

Fig. 3 Annual Cycle of track occurrence frequency, OF [month^{-1}], for all trackers analyzing the observations. Error bars indicate inter-annual variability by the temporal standard deviation.

With respect to characteristic track properties, for both their mean and 90th percentile, all trackers identify the following qualitative biases, shown in Table 9, when comparing the model ensemble against observations:

underestimation	overestimation
mean area	track occurrence
	maximum precipitation rate
	mean precipitation rate
	space-time volume
	rain volume

Table 9 Qualitative biases of characteristic track properties that are consistent across all trackers.

This means that all trackers derive for all properties of precipitation events the same qualitative biases, but these can differ in magnitude.

Looking only at tracker results of the observation-based climatology (Table 8), the characteristic track properties are overall similar between trackers. Particularly mean area, mean and maximum precipitation rates and duration are estimated similarly by the trackers. Some pronounced quantitative differences can still be found and attributed to tracker characteristics: firstly, due to the allocation of *metatracks* at track merging and splitting, the number of tracks is highest with DYMECS, whereas the space-time volume is smallest. Secondly, because of the calculation of the characteristics from the smoothed field, the intensities are lowest with OSIRIS.

Figure 3 shows the annual cycle of track occurrence frequency identified in the observations. For all four trackers, we find that the distribution is unimodal, with a peak in August and a minimum in February. Similarly to what we found for the two single historic events in Section 4.1, a tracker that allocates *metatracks* at splitting and merging (DYMECS) identifies more tracks in total than those that do not (MTD, celltrack and OSIRIS). It also shows greater inter-annual variability. Again, despite the differences of quantitative

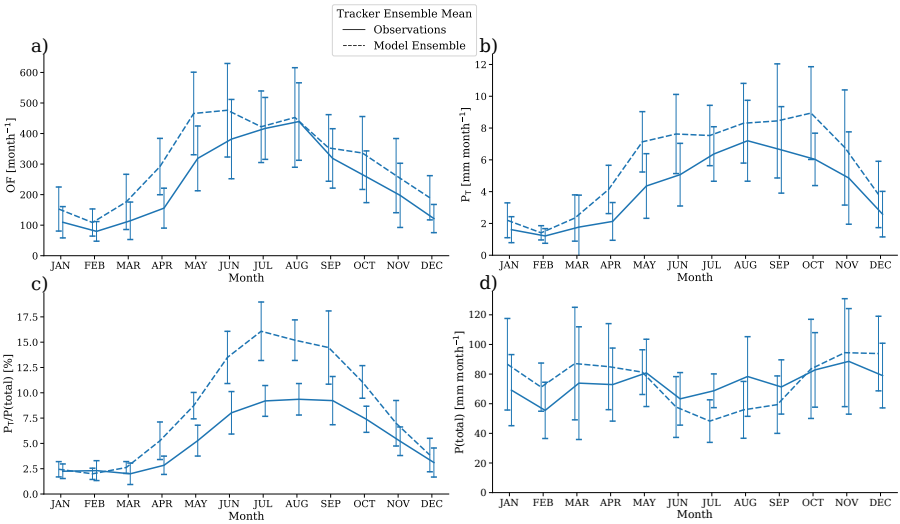

Fig. 4 Annual cycles for the tracker ensemble mean of both the model ensemble (dashed line) and observations (solid line), with panel a) showing track occurrence frequency, OF [month^{-1}], panel b) accumulated precipitation of tracks, P(tracks) [mm month^{-1}], panel c) heavy precipitation fraction, P(tracks)/P(total) [%], and panel d) accumulated total precipitation, P(total) [mm month^{-1}]. Error bars for the model ensemble indicate the temporal standard deviation of the model ensemble mean across years, and likewise for the observations error bars display inter-annual variability by the temporal standard deviation across years.

nature among trackers, the scientific conclusions when comparing climatologies of model ensemble and observations are mainly independent of the choice of the tracker.

5 Results on Model Evaluation

In this section, we evaluate the representation of precipitation events in the cpRCM ensemble. To this end, we use the mean of the tracker analyses (tracker ensemble mean) and compare the entire 9-years periods of the model ensemble against the composite of observations.

In Figure 4a) we show the annual cycle of track occurrence for the tracker ensemble mean, of both the model ensemble and observations. We find that the number of events occurring in spring and fall is overestimated, whereas for July and August the occurrence frequency of tracks in the models is close to that of the observations. The annual cycle of the model ensemble shows two peaks, one in June and one in August, whereas the annual cycle of the observations is unimodal. With respect to the estimate of inter-annual variability, given by the temporal variance across years, we find that the model ensemble exceeds the observations.

Figure 4b) and d) show accumulated precipitation of tracks (P_T) and total accumulated precipitation (P(total)), whereas panel c) shows their fraction. In this domain and time period there is no pronounced annual cycle found in P(total). If anything it is rather the models that show a dry summer w.r.t. a wet winter. In other words, the model ensemble overestimates P(total) in winter and underestimates it in summer. In contrast to that, P_T shows a strong seasonality, with the model ensemble showing a broad peak from May to November and the observations peaking from July to October. Here we find an overestimation of P_T throughout the whole year. Consequently their fraction, P_T/P(total), also shows strong seasonality, again with a peak in summer, and once more we identify a substantial overestimation by the model ensemble. Moreover we from this see that our setup chosen attributes only about 5% (observations) and 10% (model ensemble) of the annual precipitation amount to tracks. This overestimation was already of intense precipitation was already found in (Berthou et al., 2018; Meredith et al., 2020). Eventually we also see that this tracker setup serves well in identifying heavy precipitation events, as the fraction of precipitation amount identified is relatively low.

Figure 5 shows the track occurrence frequency density of the tracker ensemble mean for the model ensemble and observations, as well as their difference. Panel d) shows the normalized difference, i.e. as if there were as many tracks in model ensemble as in observations, and by this emphasizes qualitative differences. Moreover in panel e) we show the difference in track occurrence for different seasons and different elevations. It is evident from observations (Fig. 5b) that track occurrence is strongly correlated to the topography, meaning that the orographic forcing plays a major role for precipitation events to occur; this is well captured by the models as well (Fig. 5a). Prominent *hotspots* of heavy precipitation are the Julian Alps (North-East Italy), the Western Alps (especially the Italian side and the southern Maritime Alps between Italy and France) and the Massif Central (South France). Also Corsica and the Apennines can be identified as hotspots. However, also dry spots, located in the

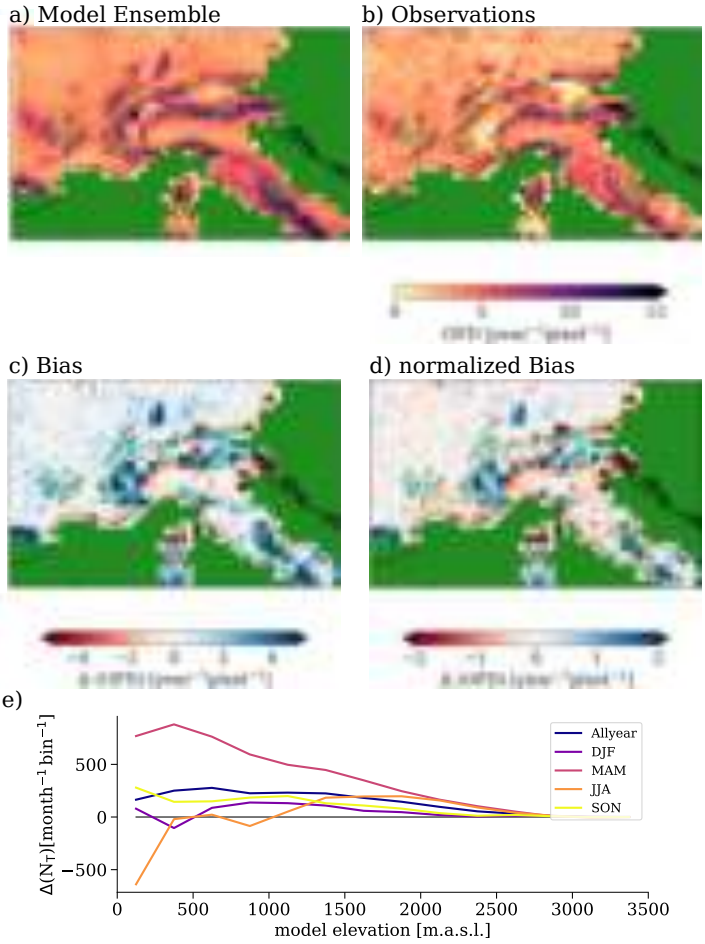

Fig. 5 Panel a) shows the track occurrence frequency density, OFD [month^{-1} pixel^{-1}], of the tracker ensemble mean for the model ensemble and panel b) shows the same for the observations. Panel c) shows their difference and panel d) show the difference, but with model ensemble and observations being normalized by their total number of tracks, respectively. A pixel is here defined as the reference area of $0.36° \times 0.36°$. The green iso-line shows the model elevation at 1000 m.a.s.l.. Panel e) shows the model-observation difference in track occurrence by model elevation.

interior of the Alps, like in Tyrol in Austria, or in the Western Alps are prominent in observations and re-produced well by the cpRCMs. Track occurrence appears overestimated over orography, particularly in the Maritime Alps, the Tyrolean Alps, the Apennines, the Black Forest and to a lesser extent, in the southern Massif Central. In contrast to this, in plains ahead of mountains, like in northern Italy, occurrence frequency is underestimated. We have seen already in the annual cycle of occurrence frequency (Figure 4 a)), that cpRCMs most strongly overestimate track occurrence in spring (MAM) and estimates OF well in summer (JJA). We now in panel e) of Figure 5 identify clearly that

cpRCMs in all seasons overestimate tracks above 1000 m.a.s.l., whereas in summertime, below 1000 m.a.s.l. OF is underestimated. This behaviour may be explained through the following considerations: numerical models easily trigger convection through orographic lifting. However, they appear to struggle to trigger thunderstorms or to resolve complex thunderstorm dynamics over plain terrain in summer, even at convection-permitting resolution (see also Craig et al., 2012; Heim et al., 2020; Prein et al., 2021). On the other hand, observational datasets under-catch rainfall amounts in mountainous regions. Therefore, model performance over orography is expected to be better than it seems. This finding is in line with Lundquist et al. (2019), who propose that well-tuned cpRCMs may outperform observational datasets over complex mountain terrain, in terms of total precipitation amounts.

The statistics of characteristic track properties of the tracker ensemble mean, for the climatologies of both model ensemble and observations, are listed in Table 8. Relative model biases of mean track properties are illustrated in panels a) and b) of Figure 6 and are summarized in Table 10.

Property	Relative Bias (all over)	Relative Bias (w/o GRIPHO-Italy)	Relative Bias (only GRIPHO-Italy)
N_T	+22 %	+43 %	−11 %
$\overline{pr}^c$	+6 %	+3 %	+12 %
$max^c(pr)$	+20 %	+13 %	+36 %
D [h]	+37 %	+29 %	+56 %
$\overline{A}^D$	−27 %	−24 %	−34 %
Vol	+8 %	+0 %	+29 %
P_T	+17 %	+8 %	+43 %

Table 10 Relative biases [%] of characteristic track properties, as defined in Table 4, using the tracker ensemble mean, for the whole domain, as well as without GRIPHO and exclusively for GRIPHO.

We find that the biases of mean track properties are mostly positive (see Figure 6b) and Table 10). Biases of the 90th percentile of track properties are still much larger (see Figure 6d), suggesting that extreme events are strongly overestimated regarding their scales and intensity. Considering the complex orography of the domain investigated (Rotunno and Houze, 2007), in combination with the known issue of under-catchment of orographic precipitation in observations, the positive biases were to be expected. Despite this, we find it important to note that mean precipitation rate of tracks is well-estimated (+6% allover the domain, +3% w/o GRIPHO-Italy). The absolute number of events, maximum precipitation rate, track duration and rainfall volume are considerably overestimated (>17%). Only the mean area of tracks is underestimated. Model biases with respect to GRIPHO-Italy differ from those of the other datasets and regions qualitatively only in terms of number of tracks, showing here an underestimation. It is worth to note that the model spread is particularly large in terms of number of tracks (Figure 6a). For all other

properties we find smaller biases over regions with radar-based datasets, i.e. France, Germany and Switzerland (w/o GRIPHO-Italy) than over Italy (only GRIPHO-Italy). It is particularly the spatio-temporal properties (duration, mean area, space-time volume) and maximum precipitation rate, that show the greatest differences. We assume that the bias reduction in spatio-temporal properties, particularly in mean area, is associated with improvements that the spatially continuous radar measurements ensure. Larger model biases over Italy might be also attributed to some higher degree of complexity not well captured by some or all cpRCMs within the ensemble. Certainly the optimal spatial-temporal representation of precipitation events in radar-based datasets constitutes an advantage in the evaluation of models in a context of Lagrangian analysis. Our findings confirm the key role of observational datasets with comparable spatial resolution in evaluation studies of RCMs (Torma et al., 2015; Prein and Gobiet, 2017d).

The probability density functions in Figure 6 give more detailed insight into the models' behaviour. Looking at track duration, we find that cpRCMs simulate precipitation events of temporal scales longer than 50 hours, that are not found in observations. In turn simulated precipitation events are generally too small regarding their area, whereas we see only a minor overestimation of the distribution's tail. As a combination of the biases of duration and area, the bias of geometrical volume is still positive. It is mostly the overestimation of track duration that causes to the positive biases in geometrical volume and also biases in rain volume are mostly found in the tails of the distribution. In other words, cpRCMs are found to simulate precipitation events of large scales that are not seen in observations. This finding is also reflected in the high biases of the 90th percentile of track properties shown in Figure 6 c) and d).

We in Figure 8 (in the Supplementary Material Section 9) provide the relative biases of mean properties for each model individually and we here would like to address the 2 cpRCM families WRF and CCLM. While the WRF models differ in their physics parameterizations, the CCLM models differ only in their nesting strategy. Among the WRF models the variability in mean biases is considerably large, with e.g. the IPSL-WRF and UHOH being particularly different. In turn the biases among the CCLM model family look much more similar. We from this conclude that physics parameterizations have a large effect on model behaviour and thus can generate greater ensemble variability than differing nesting strategies.

The spatial mapping of model biases in Figure 7 allows us to further understand impacts of the technical shortcoming mentioned in Section 2.2. Note that we in Figure 9 (in the Supplementary Material Section 9) show the mapping of the respective properties for observations and model ensemble. The *posteriori masking* of tracks means that landfalling tracks are overestimated in their spatial and temporal scales. In fact we find that along the coasts rain volume and geometrical volume are overestimated, whereas the other averaged variables appear unaffected. Over Italy we again find the pronounced underestimation of track mean area and overestimation of track duration. We here can speculate

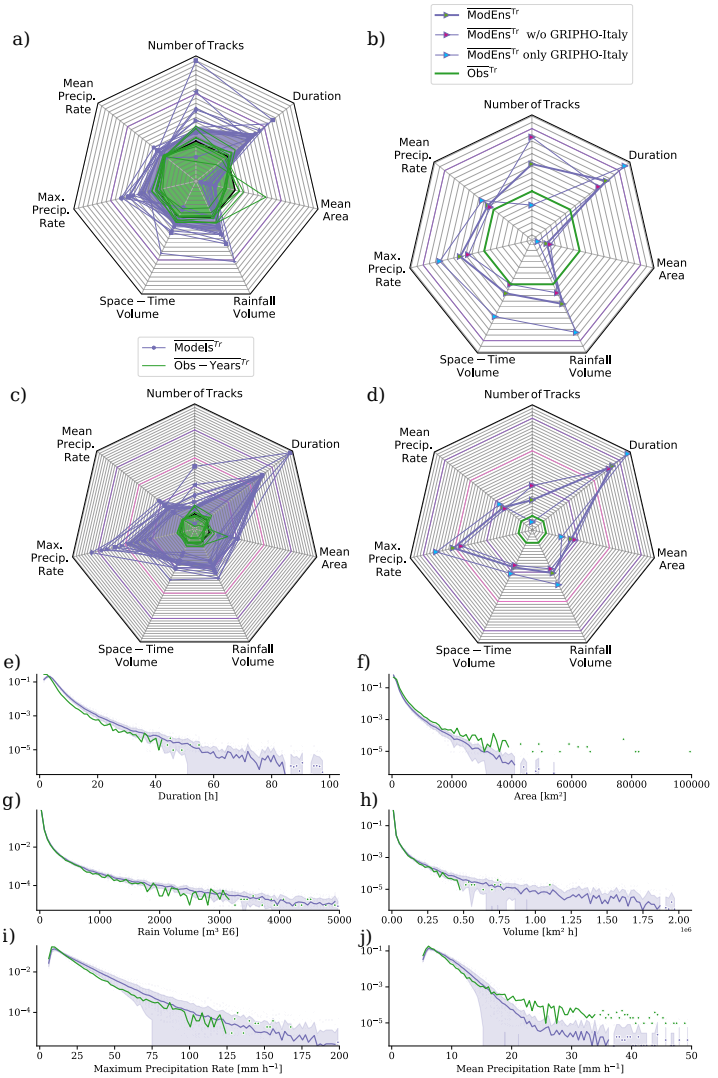

Fig. 6 Panels a) and c): The purple shaded area illustrates the relative bias of a) the tracker ensemble mean of the model ensemble mean and c) the 90th percentile, with respect to the observations: $\frac{\overline{ModEns}^{Tr} - \overline{Obs}^{Tr}}{\overline{Obs}^{Tr}}$ [%], while purple lines indicate the individual models and green lines individual years of the observations. Panels b) and d) shows also $\frac{\overline{ModEns}^{Tr} - \overline{Obs}^{Tr}}{\overline{Obs}^{Tr}}$ [%] for the mean and 90th percentile of characteristic properties, with the Italian dataset GRIPHO excluded as well as for GRIPHO only. Black solid lines are increments of +-5%, with the thick black line denoting 0%. Panels e) to j): probability density functions of Duration [h], Area [km^2], Rain Volume [m^3 E6], Space-Time Volume [km^2 h], Maximum Precipitation Rate [$mm\,h^{-1}$] and Mean Precipitation Rate [$mm\,h^{-1}$].

520 that GRIPHO's interpolation method smoothens the field strongly, enlarging

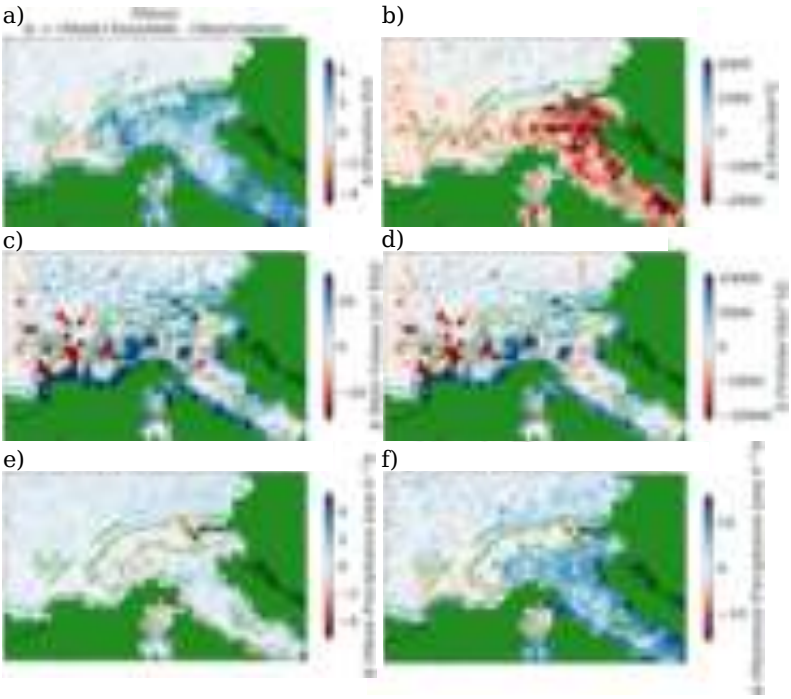

Fig. 7 The spatial biases of the tracker ensemble mean of the model ensemble w.r.t. observations. Panels a) to f): duration [h], (mean) area [km^2], rain volume [m^3 E6], (geometrical) volume [km^2 h], mean and maximum precipitation (rate) [$mm\,h^{-1}$]. Again a pixel is here defined as the reference area of $0.36° \times 0.36°$ and the green iso-line shows the model elevation at 1000 m.a.s.l..

the spatial extent of events. Also positive biases in mean and maximum precipitation are pronounced over Italy, but are not dramatically different from the other sub-regions in the domain.

6 Summary and Conclusions

The present study has a two-fold scientific objective: on the one hand we provide an inter-comparison of tracking algorithms and on the other hand we present an evaluation of an ensemble of convection-permitting regional climate models in terms of Lagrangian precipitation events. We here summarize our findings and give conclusions.

With respect to the tracker inter-comparison (see Section 4) we were able to show through both, the comparison of two historic events and the comparison of climatologies of model ensemble and observations, that all trackers investigated produce equal relations of characteristic track properties and model biases (see Tables 7 and 9 and Figures 1, 2 and 3). Thus all trackers produce qualitatively equal results. In other words, differences among the trackers were found to be only of quantitative nature, which could be addressed to certain specifications of the algorithms. From this we infer that from each tracker analysis the equivalent scientific conclusion would be derived. This result suggests that all trackers investigated are reliable analysis tools of atmospheric research.

The choice of tracker depends here much on whether *metatracks*, allocated when tracks are splitting or merging, are of interest. Further code availability, portability and user support also play a major role.

We find that the setup chosen here, given through smoothing, precipitation rate and volume threshold (Table 3), identifies an abundance of precipitation events all over the domain, of which only a fraction would be considered an extreme event. In our analysis of two historical events, we see that those are represented by several tracks. We recommend to consider that a tracker would identify fewer or only a single track, if thresholds on precipitation rate and minimum volume were raised and smoothing strengthened. The choice of setup depends upon the user-specific application. Certainly though, the most intense events are retained. In turn, reducing thresholds and weakening smoothing will result in a setup that identifies more and greater tracks and a greater fraction of precipitation will be attributed to the events.

With respect to model evaluation (see Section 5) we summarize the following findings. Looking into the spatial representation of precipitation events (see Figure 5), we found that cpRCMs perform well in reproducing *hotspots* of heavy precipitation, which are generally associated with orographic features. At the same time though, cpRCMs appear to overestimate the occurrence of precipitation events over orography. However, the under-catchment of orographic precipitation in radar-based and rain gauge observations (Creutin et al., 1997; La Barbera et al., 2002; Prein and Gobiet, 2017d; Germann et al., 2022) suggests that cpRCMs perform better than it seems. The idea of cpRCMs outperforming observations in complex terrain, particularly in terms of total precipitation amounts, is strongly supported in Lundquist et al. (2019). In contrast to this, we found the occurrence of precipitation events underestimated over plain terrain and ahead of orographic features, particularly in summer. The same model behaviour was found by Prein et al. (2017a) for North America, where the occurrence frequency of MCSs was underestimated in the central

plains but overestimated over the Appalachians. We here assume that, despite of the convection-permitting resolution, complex thunderstorms (e.g. supercells or squall lines) in plain terrain are either not triggered or their dynamics still under-resolved (see also Bryan and Morrison (2012); Pichelli et al. (2017); Prein et al. (2021)). Moreover, the correct prescription of sea surface temperatures is crucial for the intensity and evolution of characteristic landfalling Mediterranean heavy precipitation events (Lebeaupin et al., 2006). Looking into the seasonal representation of precipitation events, we find that cpRCMs overestimate the occurrence of tracks and associated precipitation amounts particularly in late spring (AMJ), and also in fall. In late summer months (JAS) the domain-wide occurrence appears estimated well, as the overestimation in regions over 1000 m.a.s.l. is compensated by the underestimation in regions below 1000 m.a.s.l..

In terms of characteristic properties of precipitation events we found the following biases (listed in Table 10 and illustrated in Figure 6) and give reference to tracker studies using convection-permitting models. The occurrence frequency of events is overestimated with respect to radar-based observations (in line with Clark et al. (2014); Prein et al. (2017a); Caillaud et al. (2021)) and under-estimated over Italy, although the models spread is large around this property. The mean area of tracks is underestimated (in line with (Crook et al., 2019), but in contrast to Caillaud et al. (2021)), while their duration is overestimated (in line with Crook et al. (2019); Purr et al. (2019)). Still, we have identified that both of these biases are particularly pronounced over Italy. In turn, the cpRCMs agree much better with the radar-based observational datasets in terms of track area and duration. The tracks' space-time volume, that is the combination of area and duration, as well as the rain volume, are overestimated. However, we here find considerable impact by the differing representation of landfalling tracks in models and observations, and excluding a major part of the coastline (the Italian sub-region) reduces the biases much. Mean precipitation rates show only small positive biases, with cpRCMs aligning again even better with radar-based observations. Maximum precipitation rate is overestimated in models and here again biases are much reduced when using radar-based observations as benchmark (in line with (Davis et al., 2006b; Prein et al., 2017a; Crook et al., 2019; Caillaud et al., 2021). Still we find that cpRCMs simulate precipitation events of scales and intensities that are not seen in observations, which means an overestimation of extreme event properties. Overall, the results on cpRCM evaluation are encouraging. On the one hand the (mostly) positive biases we find are to be expected, when assuming underestimated precipitation amounts in observations in a region of such complex orography. On the other hand we find that biases of the spatio-temporal properties of precipitation events in cpRCMs appear much reduced when using high-resolution observational datasets, based upon Doppler radar measurements.

7 Declarations

Acknowledgments. The authors gratefully acknowledge the WCRP-CORDEX-FPS on Convective phenomena at high resolution over Europe and the Mediterranean [FPSCONV-ALP-3] and the research data exchange infrastructure and services provided by the Jülich Supercomputing Centre, Germany, as part of the Helmholtz Data Federation initiative. This work has also been supported in part by the Horizon 2020 EUCP (European Climate Prediction System, https://www.eucp-project.eu) project. This EUCP project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 776613. This research is supported by the United Kingdom Natural Environment Research Council (NERC) Changing Water Cycle programme (FUTURE-STORMS; grant: NE/R01079X/1) AUTH acknowledges the support of the Greek Research and Technology Network (GRNET) High Performance Computing (HPC) infrastructure for providing the computational resources of AUTH-simulations (under project ID pr003005) and the AUTH Scientific Computing Center for technical support For JLU simulations computational resources were made available by the German Climate Computing Center (DKRZ) through support from the Federal Ministry of Education and Research in Germany (BMBF) and funding stems from the German Research Foundation (DFG) through grant nr. 401857120. The authors from FZJ gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputer JURECA at Jülich Supercomputing Centre at Forschungszentrum Jülich. The authors thank Meteo-France for providing the COMEPHORE dataset, MeteoSwiss for the RdisaggH dataset. They thank the German Weather Service for providing the RADKLIM dataset (RADKLIM Version 2017.002: Reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW) DOI:10.5676/DWD/RADKLIM_RW_V2017.002). Hylke de Vries wishes to thank Geert Lenderink and Kai Lochbihler for discussions on the use of celltrack. ICTP also acknowledges the CETEMPS, University of L'Aquila, for allowing access to the Italian database of precipitation which GRIPHO is based on.

Data Availability. The observational datasets are available upon request through the institutions listed above. Tracking algorithms OSIRIS and DYMECS are available upon request from the developers respectively. MTD is available at https://dtcenter.org/community-code/model-evaluation-tools-met and celltrack at https://github.com/lochbika/celltrack. The entire tracker analysis is publicly available at https://doi.org/10.5281/zenodo.6949615.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

References

J. T. Abatzoglou, D. J. McEvoy, and K. T. Redmond. The West Wide Drought Tracker: Drought Monitoring at Fine Spatial Scales. *Bulletin of the American Meteorological Society*, 98(9):1815–1820, Sept. 2017. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-16-0193.1. URL https://journals.ametsoc.org/doi/10.1175/BAMS-D-16-0193.1.

M. Adinolfi, M. Raffa, A. Reder, and P. Mercogliano. Evaluation and Expected Changes of Summer Precipitation at Convection Permitting Scale with COSMO-CLM over Alpine Space. *Atmosphere*, 12(1):54, Dec. 2020. ISSN 2073-4433. doi: 10.3390/atmos12010054. URL https://www.mdpi.com/2073-4433/12/1/54.

M. Baldauf, A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, 139(12):3887–3905, 2011.

N. Ban, J. Schmidli, and C. Schär. Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, 42(4):1165–1172, 2015. Publisher: Wiley Online Library.

N. Ban, J. Rajczak, J. Schmidli, and C. Schär. Analysis of Alpine precipitation extremes using generalized extreme value theory in convection-resolving climate simulations. *Climate Dynamics*, 55(1):61–75, 2020. Publisher: Springer.

A. Belušić, M. T. Prtenjak, I. Güttler, N. Ban, D. Leutwyler, and C. Schär. Near-surface wind variability over the broader Adriatic region: insights from an ensemble of regional climate models. *Climate Dynamics*, 50(11-12):4455–4480, June 2018. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-017-3885-5. URL http://link.springer.com/10.1007/s00382-017-3885-5.

D. Belušić, H. de Vries, A. Dobler, O. Landgren, P. Lind, D. Lindstedt, R. A. Pedersen, J. C. Sánchez-Perrino, E. Toivonen, B. van Ulft, and others. HCLIM38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. *Geoscientific Model Development*, 13(3):1311–1333, 2020. Publisher: Copernicus GmbH.

S. Berthou, S. Mailler, P. Drobinski, T. Arsouze, S. Bastin, K. Béranger, and C. Lebeaupin Brossier. Lagged effects of the mistral wind on heavy precipitation through ocean-atmosphere coupling in the region of valencia (spain). *Climate Dynamics*, 51(3):969–983, 2018.

P. Brousseau, Y. Seity, D. Ricard, and J. Léger. Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142(699):2231–2243, 2016. Publisher: Wiley Online Library.

G. H. Bryan and H. Morrison. Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics. *Monthly Weather Review*, 140(1):202–225, Jan. 2012. ISSN 0027-0644, 1520-0493. doi: 10.1175/MWR-D-11-00046.1. URL http://journals.ametsoc.org/doi/10.1175/MWR-D-11-00046.1.

A. Buzzi and P. P. Alberoni. Analysis and numerical modelling of a frontal passage associated with thunderstorm development over the Po valley and the Adriatic sea. *Meteorology and Atmospheric Physics*, 48(1-4):205–224, 1992. ISSN 0177-7971, 1436-5065. doi: 10.1007/BF01029569. URL http://link.springer.com/10.1007/BF01029569.

C. Caillaud, S. Somot, A. Alias, I. Bernard-Bouissières, Q. Fumière, O. Laurantin, Y. Seity, and V. Ducrocq. Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Climate Dynamics*, 56(5):1717–1752, 2021. Publisher: Springer.

S. C. Chan, E. J. Kendon, S. Berthou, G. Fosser, E. Lewis, and H. J. Fowler. Europe-wide precipitation projections at convection permitting scale with the Unified Model. *Climate Dynamics*, 55(3):409–428, 2020. Publisher: Springer.

K. Chancibault, S. Anquetin, V. Ducrocq, and G.-M. Saulnier. Hydrological evaluation of high-resolution precipitation forecasts of the Gard flash-flood event (8–9 September 2002). *Quarterly Journal of the Royal Meteorological Society*, 132(617):1091–1117, Apr. 2006. ISSN 00359009, 1477870X. doi: 10.1256/qj.04.164. URL http://doi.wiley.com/10.1256/qj.04.164.

D. Chen, J. Guo, D. Yao, Y. Lin, C. Zhao, M. Min, H. Xu, L. Liu, X. Huang, T. Chen, and P. Zhai. Mesoscale Convective Systems in the Asian Monsoon Region From Advanced Himawari Imager: Algorithms and Preliminary Results. *Journal of Geophysical Research: Atmospheres*, 124(4):2210–2234, Feb. 2019. ISSN 2169-897X, 2169-8996. doi: 10.1029/2018JD029707. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029707.

J. M. Ciarlo, E. Coppola, A. Fantini, F. Giorgi, X. Gao, Y. Tong, R. H. Glazer, J. A. T. Alavez, T. Sines, E. Pichelli, and others. A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. *Climate Dynamics*, pages 1–22, 2020. Publisher: Springer.

A. J. Clark, R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong. Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Weather and Forecasting*, 29(3):517–542, 2014.

E. Coppola, S. Sobolowski, E. Pichelli, F. Raffaele, B. Ahrens, I. Anders, N. Ban, S. Bastin, M. Belda, D. Belusic, and others. A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics*, 55(1): 3–34, 2020. Publisher: Springer.

E. Coppola, P. Stocchi, E. Pichelli, J. A. Torres Alavez, R. Glazer, G. Giuliani, F. Di Sante, R. Nogherotto, and F. Giorgi. Non-Hydrostatic RegCM4 (RegCM4-NH): model description and case studies over multiple domains. *Geoscientific Model Development*, 14(12):7705–7723, Dec. 2021. ISSN 1991-9603. doi: 10.5194/gmd-14-7705-2021. URL https://gmd.copernicus.org/articles/14/7705/2021/.

P. F. Cortopassi and M. Daddi. Discariche Di Cava E Instabilita Dei Versanti: Valutazione Preliminare Di Alcuni Fattori Significativi Nel Bacino Marmifero Di Carrara (Italia) / Quarrywaste And Slope Instability: Preliminary Assessment Of Some Controlling Factors In The Carrara Marble Basin (Italy). *Italian Journal of Engineering Geology and Environment*, (2008):99–118, Oct. 2008. doi: 10.4408/IJEGE.2008-01.S-08. URL https://doi.org/10.4408/IJEGE.2008-01.S-08.

G. C. Craig, C. Keil, and D. Leuenberger. Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 138(663):340–352, 2012.

J. Creutin, H. Andrieu, and D. Faure. Use of a weather radar for the hydrology of a mountainous area. Part II: radar measurement validation. *Journal of Hydrology*, 193(1):26–44, 1997. ISSN 0022-1694. doi: 10.1016/S0022-1694(96)03203-9. URL https://www.sciencedirect.com/science/article/pii/S0022169496032039.

J. Crook, C. Klein, S. Folwell, C. M. Taylor, D. J. Parker, R. Stratton, and T. Stein. Assessment of the representation of West African storm lifecycles in convection-permitting simulations. *Earth and Space Science*, 6(5):818–835, 2019. Publisher: Wiley Online Library.

C. Davis, B. Brown, and R. Bullock. Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Monthly Weather Review*, 134(7):1772–1784, July 2006a. ISSN 1520-0493, 0027-0644. doi: 10.1175/MWR3145.1. URL http://journals.ametsoc.org/doi/10.1175/MWR3145.1.

C. Davis, B. Brown, and R. Bullock. Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Monthly Weather Review*, 134(7):1785–1795, July 2006b. ISSN 1520-0493, 0027-0644. doi: 10.1175/MWR3146.1. URL http://journals.ametsoc.org/doi/10.1175/MWR3146.1.

G. Delrieu, J. Nicol, E. Yates, P.-E. Kirstetter, J.-D. Creutin, S. Anquetin, C. Obled, G.-M. Saulnier, V. Ducrocq, E. Gaume, and others. The catastrophic flash-flood event of 8–9 September 2002 in the Gard Region, France: a first case study for the Cévennes–Vivarais Mediterranean Hydrometeorological Observatory. *Journal of hydrometeorology*, 6(1):34–52, 2005.

P. Drobinski, V. Ducrocq, P. Alpert, E. Anagnostou, K. Béranger, M. Borga, I. Braud, A. Chanzy, S. Davolio, G. Delrieu, C. Estournel, N. F. Boubrahmi, J. Font, V. Grubišić, S. Gualdi, V. Homar, B. Ivančan-Picek, C. Kottmeier, V. Kotroni, K. Lagouvardos, P. Lionello, M. C. Llasat, W. Ludwig, C. Lutoff, A. Mariotti, E. Richard, R. Romero, R. Rotunno, O. Roussot, I. Ruin, S. Somot, I. Taupier-Letage, J. Tintore, R. Uijlenhoet, and H. Wernli. HyMeX: A 10-Year Multidisciplinary Program on the Mediterranean Water Cycle. *Bulletin of the American Meteorological Society*, 95(7):1063–1082, July 2014. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-12-00242.1. URL http://journals.ametsoc.org/doi/10.1175/BAMS-D-12-00242.1.

A. Fantini. *Climate change impact on flood hazard over Italy*. PhD Thesis, University of Trieste, 2019. URL http://hdl.handle.net/11368/2940009.

Q. Fumière, M. Déqué, O. Nuissier, S. Somot, A. Alias, C. Caillaud, O. Laurantin, and Y. Seity. Extreme rainfall in Mediterranean France during the fall: added value of the CNRM-AROME Convection-Permitting Regional Climate Model. *Climate Dynamics*, 55(1):77–91, 2020. Publisher: Springer.

U. Germann, M. Boscacci, L. Clementi, M. Gabella, A. Hering, M. Sartori, I. V. Sideris, and B. Calpini. Weather Radar in Complex Orography. *Remote Sensing*, 14(3):503, Jan. 2022. ISSN 2072-4292. doi: 10.3390/rs14030503. URL https://www.mdpi.com/2072-4292/14/3/503.

F. Giorgi. Climate change hot-spots. *Geophysical research letters*, 33(8), 2006. Publisher: Wiley Online Library.

F. Giorgi. Thirty years of regional climate modeling: where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, 124(11): 5696–5723, 2019. Publisher: Wiley Online Library.

F. Giorgi. Producing actionable climate change information for regions: the distillation paradigm and the 3R framework. *The European Physical Journal Plus*, 135(5):435, May 2020. ISSN 2190-5444. doi: 10.1140/epjp/s13360-020-00453-1. URL https://link.springer.com/10.1140/epjp/

s13360-020-00453-1.

G. A. Grell, L. Schade, R. Knoche, A. Pfeiffer, and J. Egger. Nonhydrostatic climate simulations of precipitation over complex terrain. *Journal of Geophysical Research: Atmospheres*, 105(D24):29595–29608, Dec. 2000. ISSN 01480227. doi: 10.1029/2000JD900445. URL http://doi.wiley.com/10.1029/2000JD900445.

Z. Guo, J. Tang, J. Tang, S. Wang, Y. Yang, W. Luo, and J. Fang. Object-Based Evaluation of Precipitation Systems in Convection-Permitting Regional Climate Simulation Over Eastern China. *Journal of Geophysical Research: Atmospheres*, 127(1), Jan. 2022. ISSN 2169-897X, 2169-8996. doi: 10.1029/2021JD035645. URL https://onlinelibrary.wiley.com/doi/10.1029/2021JD035645.

R. M. Haralick and L. G. Shapiro. *Computer and robot vision*. Addison-Wesley Pub. Co, Reading, Mass, 1992. ISBN 978-0-201-10877-4 978-0-201-56943-8.

C. Heim, D. Panosetti, L. Schlemmer, D. Leuenberger, and C. Schär. The Influence of the Resolution of Orography on the Simulation of Orographic Moist Convection. *Monthly Weather Review*, 148(6):2391–2410, June 2020. ISSN 0027-0644, 1520-0493. doi: 10.1175/MWR-D-19-0247.1. URL http://journals.ametsoc.org/doi/10.1175/MWR-D-19-0247.1.

C. Hohenegger, P. Brockhaus, and C. Schär. Towards climate simulations at cloud-resolving scales. *Meteorologische Zeitschrift*, 17(4):383–394, Aug. 2008. ISSN 0941-2948. doi: 10.1127/0941-2948/2008/0303. URL http://www.schweizerbart.de/papers/metz/detail/17/56729/Towards_climate_simulations_at_cloud_resolving_sca?af=crossref.

F. A. Isotta, C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, and G. Vertačnik. The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data: CLIMATE OF DAILY PRECIPITATION IN THE ALPS. *International Journal of Climatology*, 34(5):1657–1675, Apr. 2014. ISSN 08998418. doi: 10.1002/joc.3794. URL https://onlinelibrary.wiley.com/doi/10.1002/joc.3794.

A. Johnson, X. Wang, F. Kong, and M. Xue. Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts. *Monthly Weather Review*, 141(10):3413–3425, Oct. 2013. ISSN 0027-0644, 1520-0493. doi: 10.1175/MWR-D-13-00027.1. URL http://journals.ametsoc.org/doi/10.1175/MWR-D-13-00027.1.

E. J. Kendon, N. Ban, N. M. Roberts, H. J. Fowler, M. J. Roberts, S. C. Chan, J. P. Evans, G. Fosser, and J. M. Wilkinson. Do convection-permitting regional climate models improve projections of future precipitation change? *Bulletin of the American Meteorological Society*, 98(1):79–93, 2017. Publisher: American Meteorological Society.

E. J. Kendon, C. Short, J. Pope, S. C. Chan, J. Wilkinson, S. Tucker, P. Bett, G. Harris, and J. Murphy. Update to ukcp local (2.2km) projections. Technical report, United Kingdom Met Office, Exeter, United Kingdom, 7 2021. URL https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/guidance-science-reports.

K. Keuler, K. Radtke, S. Kotlarski, and D. Lüthi. Regional climate change over Europe in COSMO-CLM: Influence of emission scenario and driving global model. *Meteorologische Zeitschrift*, 25(2):121–136, 2016. Publisher: Schweizerbart.

P. La Barbera, L. Lanza, and L. Stagi. Tipping bucket mechanical errors and their influence on rainfall statistics and extremes. *Water Science and Technology*, 45(2):1–9, Jan. 2002. ISSN 0273-1223, 1996-9732. doi: 10.2166/wst.2002.0020. URL https://iwaponline.com/wst/article/45/2/1/9061/Tipping-bucket-mechanical-errors-and-their.

C. Lebeaupin, V. Ducrocq, and H. Giordani. Sensitivity of torrential rain events to the sea surface temperature based on high-resolution numerical forecasts. *Journal of Geophysical Research: Atmospheres*, 111(D12), 2006.

K. Lochbihler, G. Lenderink, and A. P. Siebesma. The spatial extent of rainfall events and its relation to precipitation scaling. *Geophysical Research Letters*, 44(16):8629–8636, 2017. Publisher: Wiley Online Library.

P. Lucas-Picher, D. Argüeso, E. Brisson, Y. Tramblay, P. Berg, A. Lemonsu, S. Kotlarski, and C. Caillaud. Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, 12(6), Nov. 2021. ISSN 1757-7780, 1757-7799. doi: 10.1002/wcc.731. URL https://onlinelibrary.wiley.com/doi/10.1002/wcc.731.

J. Lundquist, M. Hughes, E. Gutmann, and S. Kapnick. Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, 100 (12):2473–2490, Dec. 2019. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-D-19-0001.1. URL https://journals.ametsoc.org/view/journals/bams/100/12/bams-d-19-0001.1.xml.

S. Medina and R. A. Houze Jr. Air motions and precipitation growth in Alpine storms. *Quarterly Journal of the Royal Meteorological Society*, 129(588): 345–371, Jan. 2003. ISSN 1477870X, 00359009. doi: 10.1256/qj.02.13. URL

http://doi.wiley.com/10.1256/qj.02.13.

E. P. Meredith, U. Ulbrich, and H. W. Rust. Subhourly rainfall in a convection-permitting model. *Environmental Research Letters*, 15(3):034031, 2020.

M. M. Miglietta, A. Manzato, and R. Rotunno. Characteristics and predictability of a supercell during HyMeX SOP1. *Quarterly Journal of the Royal Meteorological Society*, 142(700):2839–2853, Oct. 2016. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.2872. URL https://onlinelibrary.wiley.com/doi/10.1002/qj.2872.

C. Morel and S. Senesi. A climatology of mesoscale convective systems over Europe using satellite infrared imagery. II: Characteristics of European mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, 128(584):1973–1995, July 2002a. ISSN 00000000, 00359009. doi: 10.1256/003590002320603494. URL http://doi.wiley.com/10.1256/003590002320603494.

C. Morel and S. Senesi. A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology. *Quarterly Journal of the Royal Meteorological Society*, 128(584):1953–1971, July 2002b. ISSN 00000000, 00359009. doi: 10.1256/003590002320603485. URL http://doi.wiley.com/10.1256/003590002320603485.

G. M. Morgan. A General Description of the Hail Problem in the Po Valley of Northern Italy. *Journal of Applied Meteorology*, 12(2):338–353, Mar. 1973. ISSN 0021-8952. doi: 10.1175/1520-0450(1973)012⟨0338:AGDOTH⟩2.0.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0450(1973)012⟨0338:AGDOTH⟩2.0.CO;2.

C. Moseley, P. Berg, and J. O. Haerter. Probing the precipitation life cycle by iterative rain cell tracking. *Journal of Geophysical Research: Atmospheres*, 118(24):13–361, 2013. Publisher: Wiley Online Library.

P. Nabat, S. Somot, C. Cassou, M. Mallet, M. Michou, D. Bouniol, B. Decharme, T. Drugé, R. Roehrig, and D. Saint-Martin. Modulation of radiative aerosols effects by atmospheric circulation over the Euro-Mediterranean region. *Atmospheric Chemistry and Physics*, 20(14): 8315–8349, 2020. Publisher: Copernicus GmbH.

U. Neu, M. G. Akperov, N. Bellenbaum, R. Benestad, R. Blender, R. Caballero, A. Cocozza, H. F. Dacre, Y. Feng, K. Fraedrich, J. Grieger, S. Gulev, J. Hanley, T. Hewson, M. Inatsu, K. Keay, S. F. Kew, I. Kindem, G. C. Leckebusch, M. L. R. Liberato, P. Lionello, I. I. Mokhov, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, M. Schuster, I. Simmonds, M. Sinclair, M. Sprenger, N. D. Tilinina, I. F. Trigo, S. Ulbrich, U. Ulbrich, X. L. Wang,

and H. Wernli. IMILAST: A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms. *Bulletin of the American Meteorological Society*, 94(4):529–547, Apr. 2013. ISSN 1520-0477. doi: 10.1175/BAMS-D-11-00154.1. URL https://journals.ametsoc.org/doi/10.1175/BAMS-D-11-00154.1.

L. Panziera, C. N. James, and U. Germann. Mesoscale organization and structure of orographic precipitation producing flash floods in the Lago Maggiore region: Orographic Convection in the Lago Maggiore Area. *Quarterly Journal of the Royal Meteorological Society*, 141(686):224–248, Jan. 2015. ISSN 00359009. doi: 10.1002/qj.2351. URL https://onlinelibrary.wiley.com/doi/10.1002/qj.2351.

E. Pichelli, R. Rotunno, and R. Ferretti. Effects of the Alps and Apennines on forecasts for Po Valley convection in two HyMeX cases: Effects of Alps and Apennines on Po Valley Convection Forecasts. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2420–2435, July 2017. ISSN 00359009. doi: 10.1002/qj.3096. URL https://onlinelibrary.wiley.com/doi/10.1002/qj.3096.

E. Pichelli, E. Coppola, S. Sobolowski, N. Ban, F. Giorgi, P. Stocchi, A. Alias, D. Belušić, S. Berthou, C. Caillaud, R. M. Cardoso, S. Chan, O. B. Christensen, A. Dobler, H. de Vries, K. Goergen, E. J. Kendon, K. Keuler, G. Lenderink, T. Lorenz, A. N. Mishra, H.-J. Panitz, C. Schär, P. M. M. Soares, H. Truhetz, and J. Vergara-Temprado. The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Climate Dynamics*, 56(11-12):3581–3602, June 2021. ISSN 0930-7575, 1432-0894. doi: 10.1007/s00382-021-05657-4. URL https://link.springer.com/10.1007/s00382-021-05657-4.

J. G. Powers, J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, D. J. Gochis, R. Ahmadov, S. E. Peckham, and others. The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, 98(8):1717–1737, 2017.

A. Prein, R. Rasmussen, D. Wang, and S. Giangrande. Sensitivity of organized convective storms to model grid spacing in current and future climates. *Philosophical Transactions of the Royal Society A*, 379(2195):20190546, 2021. Publisher: The Royal Society Publishing.

A. F. Prein and A. Gobiet. Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, 37(1):305–327, 2017d. Publisher: Wiley Online Library.

960 A. F. Prein, W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen,
961 M. Keller, M. Tölle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli,
962 N. P. M. Lipzig, and R. Leung. A review on regional convection-permitting
963 climate modeling: Demonstrations, prospects, and challenges. *Reviews of*
964 *Geophysics*, 53(2):323–361, June 2015. ISSN 8755-1209, 1944-9208. doi:
965 10.1002/2014RG000475. URL https://onlinelibrary.wiley.com/doi/10.1002/
966 2014RG000475.

967 A. F. Prein, C. Liu, K. Ikeda, R. Bullock, R. M. Rasmussen, G. J. Holland,
968 M. Clark, and others. Simulating North American mesoscale convective
969 systems with a convection-permitting climate model. *Climate Dynamics*,
970 pages 1–16, 2017a. Publisher: Springer.

971 A. F. Prein, C. Liu, K. Ikeda, S. B. Trier, R. M. Rasmussen, G. J. Holland,
972 and M. P. Clark. Increased rainfall volume from future convective storms in
973 the US. *Nature Climate Change*, 7(12):880–884, 2017b. Publisher: Nature
974 Publishing Group.

975 C. Purr, E. Brisson, and B. Ahrens. Convective Shower Characteristics Simu-
976 lated with the Convection-Permitting Climate Model COSMO-CLM. *Atmo-*
977 *sphere*, 10(12):810, Dec. 2019. ISSN 2073-4433. doi: 10.3390/atmos10120810.
978 URL https://www.mdpi.com/2073-4433/10/12/810.

979 R. Rasmussen, C. Liu, K. Ikeda, D. Gochis, D. Yates, F. Chen, M. Tewari,
980 M. Barlage, J. Dudhia, W. Yu, K. Miller, K. Arsenault, V. Grubišić,
981 G. Thompson, and E. Gutmann. High-Resolution Coupled Climate Runoff
982 Simulations of Seasonal Snowfall over Colorado: A Process Study of Cur-
983 rent and Warmer Climate. *Journal of Climate*, 24(12):3015–3048, June
984 2011. ISSN 0894-8755, 1520-0442. doi: 10.1175/2010JCLI3985.1. URL
985 http://journals.ametsoc.org/doi/10.1175/2010JCLI3985.1.

986 A. Reder, M. Raffa, M. Montesarchio, and P. Mercogliano. Performance
987 evaluation of regional climate model simulations at different spatial and
988 temporal scales over the complex orography area of the Alpine region.
989 *Natural Hazards*, 102(1):151–177, May 2020. ISSN 0921-030X, 1573-0840.
990 doi: 10.1007/s11069-020-03916-x. URL http://link.springer.com/10.1007/
991 s11069-020-03916-x.

992 A. Reder, M. Raffa, R. Padulano, G. Rianna, and P. Mercogliano. Character-
993 izing extreme values of precipitation at very high resolution: An experiment
994 over twenty European cities. *Weather and Climate Extremes*, 35:100407,
995 Mar. 2022. ISSN 22120947. doi: 10.1016/j.wace.2022.100407. URL https:
996 //linkinghub.elsevier.com/retrieve/pii/S2212094722000019.

997 R. E. Rinehart and E. T. Garvey. Three-dimensional storm motion detection
998 by conventional weather radar. *Nature*, 273(5660):287–289, May 1978. ISSN
999 0028-0836, 1476-4687. doi: 10.1038/273287a0. URL http://www.nature.

com/articles/273287a0.

B. Rockel, A. Will, and A. Hense. The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, 17(4):347–348, 2008. Publisher: Berlin: Borntraeger, c1992-.

R. Rotunno and R. A. Houze. Lessons on orographic precipitation from the Mesoscale Alpine Programme. *Quarterly Journal of the Royal Meteorological Society*, 133(625):811–830, Apr. 2007. ISSN 00359009, 1477870X. doi: 10.1002/qj.67. URL https://onlinelibrary.wiley.com/doi/10.1002/qj.67.

M. Rummukainen. Added value in regional climate modeling. *WIREs Climate Change*, 7(1):145–159, Jan. 2016. ISSN 1757-7780, 1757-7799. doi: 10.1002/wcc.378. URL https://onlinelibrary.wiley.com/doi/10.1002/wcc.378.

S. Sauvagnargues-Lesage. Retour d'expérience sur la gestion de l'événement de Septembre 2002 par les services de Sécurité Civile. *La Houille Blanche*, 90(6):107–113, Nov. 2004. ISSN 0018-6368, 1958-5551. doi: 10.1051/lhb:200406015. URL https://www.tandfonline.com/doi/full/10.1051/lhb%3A200406015.

M. Schleiss, J. Olsson, P. Berg, T. Niemi, T. Kokkonen, S. Thorndahl, R. Nielsen, J. Ellerbæk Nielsen, D. Bozhinova, and S. Pulkkinen. The accuracy of weather radar in heavy rain: a comparative study for denmark, the netherlands, finland and sweden. *Hydrology and Earth System Sciences*, 24(6):3157–3188, 2020. doi: 10.5194/hess-24-3157-2020. URL https://hess.copernicus.org/articles/24/3157/2020/.

T. Stein, R. Hogan, K. Hanley, P. Clark, C. Halliwell, H. Lean, J. Nicol, and R. Plant. The three-dimensional microphysical structure of convective storms over the southern United Kingdom. *Monthly Weather Review*, 142:3264–3283, 2014.

B. Stevens. ATMOSPHERIC MOIST CONVECTION. *Annual Review of Earth and Planetary Sciences*, 33(1):605–643, May 2005. ISSN 0084-6597, 1545-4495. doi: 10.1146/annurev.earth.33.092203.122658. URL https://www.annualreviews.org/doi/10.1146/annurev.earth.33.092203.122658.

P. Tabary, P. Dupuy, G. L\rqHenaff, C. Gueguen, L. Moulin, O. Laurantin, C. Merlier, and J.-M. Soubeyroux. A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over France: methodology and first results. *IAHS Publ*, 351:255–260, 2012.

C. Torma, F. Giorgi, and E. Coppola. Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps: ADDED VALUE OF RCM OVER COMPLEX TERRAIN. *Journal of Geophysical Research: Atmospheres*, 120(9):3957–3972, May 2015. ISSN

2169897X. doi: 10.1002/2014JD022781. URL http://doi.wiley.com/10.1002/2014JD022781.

E. van Meijgaard, L. Van Ulft, W. Van de Berg, F. Bosveld, B. Van den Hurk, G. Lenderink, and A. Siebesma. *The KNMI regional atmospheric climate model RACMO, version 2.1.* KNMI De Bilt, The Netherlands, 2008.

E. Van Meijgaard, L. Van Ulft, G. Lenderink, S. De Roode, E. L. Wipfler, R. Boers, and R. van Timmermans. *Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe.* Number KVR 054/12. KVR, 2012.

H. Wernli, M. Paulat, M. Hagen, and C. Frei. SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136(11):4470–4487, Nov. 2008. ISSN 1520-0493, 0027-0644. doi: 10.1175/2008MWR2415.1. URL http://journals.ametsoc.org/doi/10.1175/2008MWR2415.1.

T. Winterrath, C. Brendel, M. Hafer, T. Junghänel, A. Klameth, K. Lengfeld, E. Walawender, E. Weigl, and A. Becker. RADKLIM Version 2017.002: Reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW). *Deutscher Wetterdienst (DWD)*, 2018.

M. Wüest, C. Frei, A. Altenhoff, M. Hagen, M. Litschi, and C. Schär. A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *International Journal of Climatology*, 30(12): 1764–1775, 2010. Publisher: Wiley Online Library.

List of Figures

List of Tables

8 Tracker Descriptions

We here provide detailed descriptions of the tracking algorithms and supplementary material.

8.1 MTD

The Method for Object-Based Diagnostic Evaluation (MODE) time domain tool (MTD) is part of the Model Evaluation Tools (MET, see https://dtcenter.org/community-code/model-evaluation-tools-met). It is developed, maintained and made freely available by the Developmental Testbed Center and used here by ICTP. The toolbox comprises various analysis tools developed for the evaluation of numerical atmospheric models. (Davis et al., 2006a,b) first introduce the basic methodology of MODE and demonstrate the advantages of an object-based evaluation in numerical weather predictions. Later on the tracking of objects in time was added and the capability of MTD in describing the characteristic properties of rain systems in both simulations and observations on a continental scale was explored in (Clark et al., 2014). Finally (Prein et al., 2017a) applied MTD in order to identify mesoscale convective systems in convection-permitting climate simulations of North America.

The algorithm can be summarized as follows:

1. the field is smoothed by convolution in space, across a radius (here: 1 grid cell, which means smoothing across 3×3 grid cells) and in time across a number of time steps (here: +-0).
2. a precipitation threshold (here: $5\,\mathrm{mm\,h^{-1}}$) is applied and thereafter only grid cells exceeding the threshold are considered
3. adjacent cells in space and time are clustered to form objects.
4. a minimum volume threshold is applied (here: 100 grid cells), meaning that all object that are too small will be dropped.

The output of the analysis is a set of tracks, that represent precipitation events, along with information about their respective location, scale, intensity and propagation.

The location of a track is given through the geometrical centroid across all grid cells associated with the track in space and time. Due to computational limitations the tracker only processes periods of 10 days at a time. Since we are here looking at events with time scales much shorter than that, we don't expect the analysis being deteriorated much due to this. The mask for comparison of observations against models was applied after the analysis. All statistical properties presented here are derived from the raw un-smoothed precipitation field, whereas the grid cells associated with the event are identified from the smoothed field. Along the boundaries and 1 smoothing radius inwards the input field is set to zero before applying the smoothing.

8.2 OSIRIS

The precipitating system detection and tracking algorithm used by CNRM is based on the algorithm developed at CNRM (Morel and Senesi, 2002a,b) applied in precipitation nowcasting at Meteo-France and for the evaluation of the numerical weather prediction model AROME (Brousseau et al., 2016). It has also been recently used in an evaluation study of CNRM-AROME on Mediterranean Heavy Precipitation Events (Caillaud et al., 2021). The 1-hour accumulated precipitation fields are used as input of the tool and the method can be summarised as follows:

1. Smoothing : first, each grid cell is replaced by a weighted average of the 3×3 adjacent grid cells and second, a Gaussian filter is applied with a small standard deviation (0.5) allowing for a slight smoothing;
2. Detection of the precipitating systems every hour with a minimum surface of 20 km^2 and contiguous grid points exceeding several intensity thresholds (here: 5 $mm\,h^{-1}$);
3. Tracking of system trajectories by identifying links between systems at different time steps according to overlapping and correlation conditions. The overlapping condition uses the velocity of the cells calculated between different time steps with a minimum recovery rate of 15 %. The correlation condition is based on spatial correlation calculation between cells at different time steps present in a research box around the cell, with a minimum correlation of 0.4;
4. Minimum volume threshold applied (here: 100 grid cells),
5. Diagnostics: each cell is schematized as an ellipse: centre of gravity, length of the minor axis and the major axis, angle and the main characteristics of each trajectory can be calculated (location, duration, area, mean and maximum intensity, velocity, ...). The different characteristics are calculated on the smoothed field.

A further description of the algorithm can be found in (Caillaud et al., 2021).

8.3 celltrack

The tracker (celltrack) used by KNMI is described in detail in (Lochbihler et al., 2017) and is inspired by the work of (Moseley et al., 2013). By default celltrack does not use prior smoothing of the input field. To make celltrack comparable to the other trackers, the input field was smoothed using a 3×3 box-smoothing. This smoothed input field is used in all subsequent steps. First individual cells above a precipitation threshold (5 $mm\,h^{-1}$) are diagnosed, not considering a specific minimum area. These cells are subsequently linked into tracks. A six-fold iterative advection correction is implemented using advective velocities derived on a coarse-grained grid. After the linking, various track types can be diagnosed (e.g., single "clean" tracks, mergers, splits, etc) following a specific taxonomic classification (Lochbihler et al., 2017). The optional diagnostics of sub-cell and mainstream detection are not used in this study. The Fortran code is available on GitHub.

8.4 DYMECS

The precipitation system detection and tracking algorithm was originally developed for sub-hourly radar and forecast model precipitation data (Stein et al., 2014). Since then, it has been applied to hourly climate model data with resolutions as coarse as 25km (Crook et al., 2019).

The algorithm is divided into two parts: the detection of objects-of-interest for each image and the tracking of these objects-of-interest between consecutive images. The detection algorithm is based on the "local table method" (Haralick and Shapiro, 1992), labelling pixels-of-interest by line-by-line scanning. The tracking component is based on the windowed cross-correlation between consecutive precipitation images (Rinehart and Garvey, 1978). Windowed correlations between consecutive images are computed, and velocities between images estimated. Objects identified in the previous image are then moved by those estimated velocities. Areal overlap of objects between the two images are computed. If the object overlapping fraction exceeds 0.6, the overlapping objects are considered part of the same track, with splitting and merging allowed if there are multiple overlapping. If two or more objects can be traced back to a single object in the previous image, a split occurs with the object with higher area overlap retaining the same track identifier (i.e., metatrack) and the other objects labelled as new tracks. For the opposite case of two or more objects from the previous image tracing to a single object in the current image, a merge occurs, and retains the identifier with the track with the highest overlap; other merged objects and their identifiers cease to exist.

Smoothing is not originally part of the algorithm. This is added for this study, using the same Gaussian blurring approach as in (Caillaud et al., 2021). Smoothing is applied to the detection phase only, affecting only pixel labelling without changing the underlying precipitation intensities.

The code is written in MATLAB/OCTAVE and is available from the Met Office upon request.

9 Supplementary Material

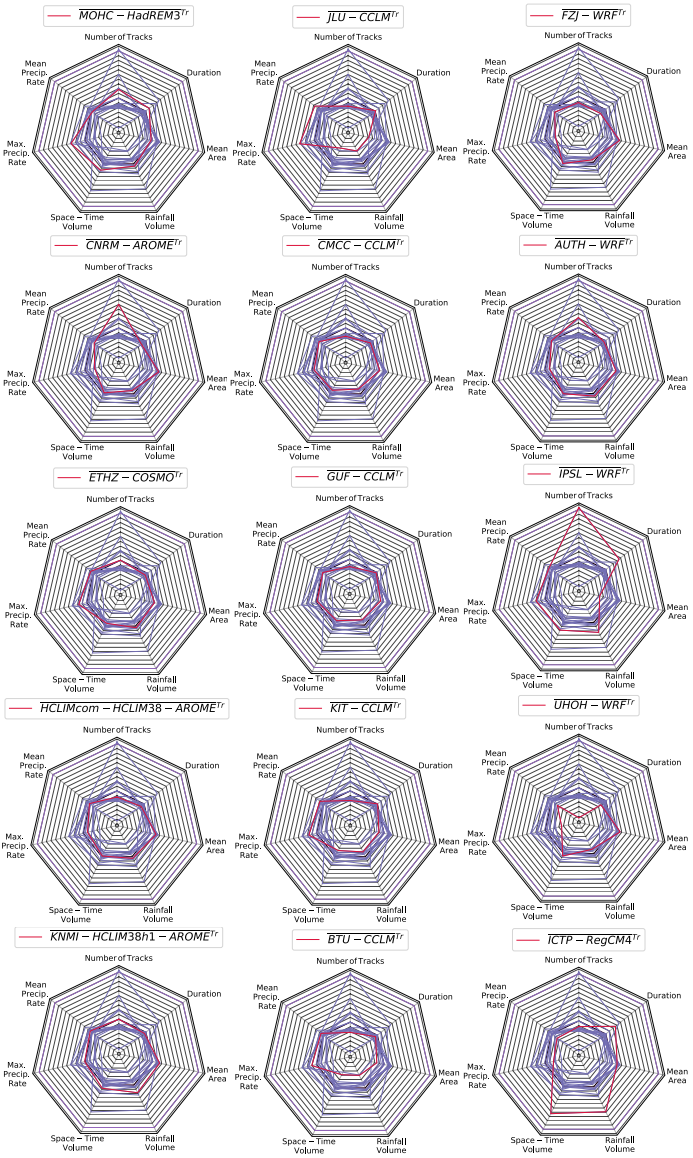

Fig. 8 Relative Biases of each model w.r.t. to the model ensemble mean using the tracker ensemble mean: $\frac{\overline{Model}^{Tr} - \overline{ModEns}^{Tr}}{\overline{ModEns}^{Tr}}$[%]. Black solid lines are increments of +-5%, with the thick black line representing the tracker ensemble mean of the model ensemble, i.e. 0%.

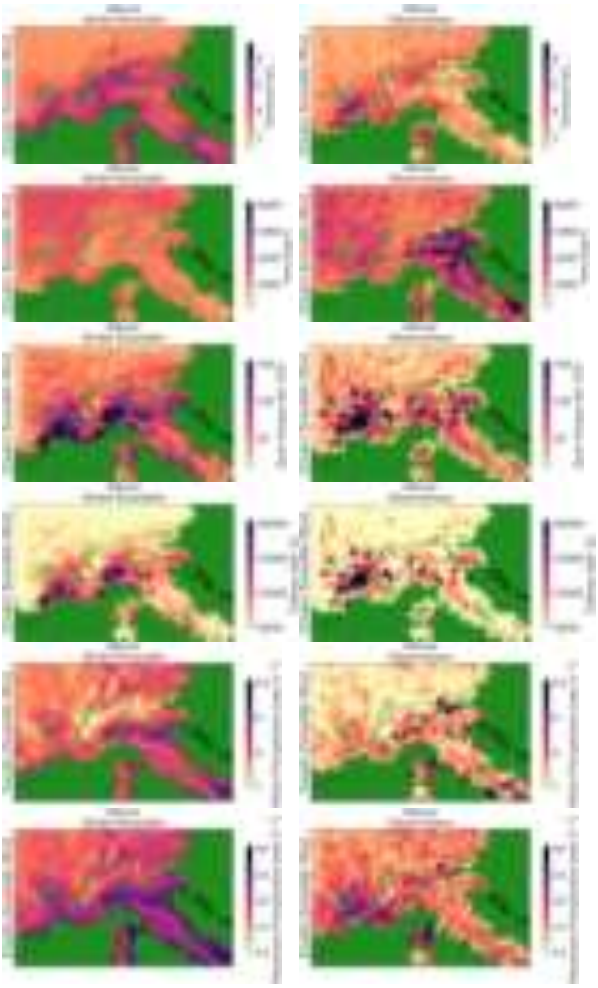

Fig. 9 Spatial mapping of characteristic track properties using the tracker ensemble mean, for the model ensemble (left column) and the observations (right column), From top to bottom: Duration [h], Area [km^2], Rain Volume [m^3 E6], (geometrical) Volume [km^2 h], Mean Precipitation Rate [$mm\,h^{-1}$] and Maximum Precipitation Rate [$mm\,h^{-1}$].