Weather and Climate Extremes 18 (2017) 65–74

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities

Jana Sillmann[a,*], Thordis Thorarinsdottir[b], Noel Keenlyside[c], Nathalie Schaller[a], Lisa V. Alexander[d], Gabriele Hegerl[e], Sonia I. Seneviratne[f], Robert Vautard[g], Xuebin Zhang[h], Francis W. Zwiers[i]

[a] CICERO, Center for International Climate Research, Gaustadalleen 21, 0349, Oslo, Norway
[b] Norwegian Computing Center, Gaustadalléen 23, 0373, Oslo, Norway
[c] Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Allegt. 70, 5020, Bergen, Norway
[d] Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, UNSW, Sydney, NSW, Australia
[e] Geosciences, University of Edinburgh, Edinburgh, United Kingdom
[f] ETH Zurich, Zurich, Switzerland
[g] LSCE IPSL, Unversité Paris-Saclay, Orme des Merisiers, 91191, Gif sur Yvette Cedex, France
[h] Environment and Climate Change Canada, Toronto, OT, Canada
[i] Pacific Climate Impact Consortium, Victoria, BC, Canada

ABSTRACT

Weather and climate extremes are identified as major areas necessitating further progress in climate research and have thus been selected as one of the World Climate Research Programme (WCRP) Grand Challenges. Here, we provide an overview of current challenges and opportunities for scientific progress and cross-community collaboration on the topic of understanding, modeling and predicting extreme events based on an expert workshop organized as part of the implementation of the WCRP Grand Challenge on Weather and Climate Extremes. In general, the development of an extreme event depends on a favorable initial state, the presence of large-scale drivers, and positive local feedbacks, as well as stochastic processes. We, therefore, elaborate on the scientific challenges related to large-scale drivers and local-to-regional feedback processes leading to extreme events. A better understanding of the drivers and processes will improve the prediction of extremes and will support process-based evaluation of the representation of weather and climate extremes in climate model simulations. Further, we discuss how to address these challenges by focusing on short-duration (less than three days) and long-duration (weeks to months) extreme events, their underlying mechanisms and approaches for their evaluation and prediction.

1. Introduction

The recent Fifth Assessment Report of the Intergovernmental Panel on Climate Change affirmed that our climate and its extremes are changing (IPCC 2013). Reliable predictions of extremes are needed on short and long time scales to reduce potential risks and damages that result from weather and climate extremes (IPCC, 2012; Seneviratne et al., 2012). Understanding, modeling and predicting weather and climate extremes is identified as a major area necessitating further progress in climate research and has thus been selected as one of the World Climate Research Programme (WCRP) Grand Challenges, which is hereafter referred to as the Extremes Grand Challenge. The WCRP Extremes Grand Challenge (Zhang et al., 2014; Alexander et al., 2016) is organized around four overarching research themes: Document (focusing on observational requirements), Understand (focusing on the relative roles of different spatial scales and their interactions), Simulate (focusing on model reliability and improvement), and Attribute (focusing on unraveling the contributors to extreme events). Underlying all research themes is a focus on four core types of extreme events: Heavy Precipitation, Heatwaves, Droughts, and Storms.

As part of the implementation of the WCRP Extremes Grand Challenge, and in particular contributing to the Understand and Simulate themes, a workshop on "Understanding, modeling and predicting weather and climate extremes" was held (see http://www.wcrp-climate.org/extremes-modeling-wkshp-about). It brought together international experts and early career scientists from the weather, climate and statistical sciences to discuss the main theoretical and modeling challenges and opportunities around extreme events. The workshop focused on various processes underlying weather and climate extremes, and how an improved understanding of these processes may lead to advances in their

* Corresponding author.
E-mail address: jana.sillmann@cicero.oslo.no (J. Sillmann).

https://doi.org/10.1016/j.wace.2017.10.003
Received 3 April 2017; Received in revised form 10 August 2017; Accepted 30 October 2017
Available online 4 November 2017

simulation and prediction. More details of the workshop format, talks and participants are given in Appendix A.

The purpose of this paper is to provide an overview for a wider audience of climatologists and statisticians on the current state of knowledge, the prevailing challenges and the potential ways forward based on the expert discussions form the workshop supplemented with current literature. We conclude with a summary of the key points, as well as ideas for future research and opportunities for cross-disciplinary collaborations.

2. Scientific challenges

To better serve local and national climate adaptation planning and decision-making, there is a clear need for improved understanding and prediction of extreme weather events. This is a cross-community challenge that requires collaboration between global programs such as the WCRP and the World Weather Research Programme (WWRP). One example is WWRP's HiWeather project (www.wmo.int/pages/prog/arep/wwrp/new/high_impact_weather_project.html) that aims to build resilience to high-impact weather events by improving their forecasts and predictability across temporal and spatial scales. Other good examples of collaborative research programs funded by the European Commission are projects such as EUPORIAS (European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales), SPECS (Seasonal-to-decadal climate Prediction for the improvement of European Climate Services) and EUCLEIA (European Climate and Weather Events: Interpretation and Attribution). Building on the insights gained from these (and many other) projects, we address current challenges and opportunities for scientific progress in various aspects of understanding and predicting weather and climate extremes.

As illustrated in the conceptual Fig. 1, the development of an extreme event depends on some or all of the following: a favorable initial state, the presence of large-scale drivers, and positive local feedbacks, as well as stochastic processes (noise). We therefore structured the scientific challenges into large-scale drivers of extreme events (section 2.1) and local-to-regional feedback processes of extreme events (section 2.2) that we need to be better understand to improve the prediction of extremes (section 2.3) and to assess model performance by process-based evaluation of climate extremes (section 2.4).

2.1. Large-scale drivers of extreme events

Our understanding of the mechanisms that lead to the occurrence of

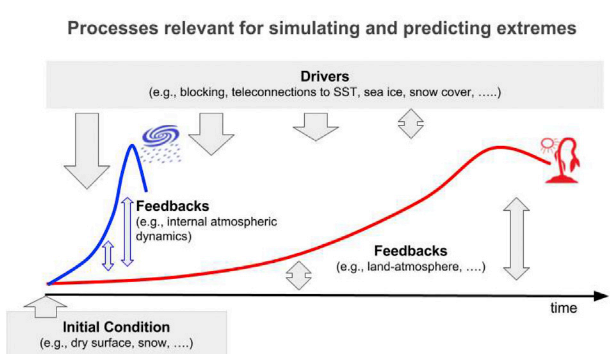

Fig. 1. The development of an extreme event depends on a favorable initial state, the presence of large-scale drivers, and positive feedbacks, as well as stochastic processes (noise). The relative importance of these factors varies for different types of extremes. For example, feedbacks for short lived events (blue) like convective storms are typically associated with unstable atmospheric dynamics, whereas longer duration events (red) like heatwaves or droughts typically involve soil moisture - atmosphere interaction. External factors like global warming can influence extremes through these various factors. For example, the increased water vapor in a warmer atmosphere can enhance convective feedbacks, or increased surface evaporation might amplify heat waves and droughts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extreme events will be the basis to assess their predictability and enable their prediction using model simulations. It is convenient and attractive to try to separate dynamic (i.e., circulation induced changes) and thermodynamic (i.e., temperature induced changes) processes when diagnosing mechanisms. However, the separation is artificial as forced dynamical changes are ultimately caused by thermodynamic processes. For instance, changes in temperature have a direct impact on the hydrological cycle (i.e., Clausius-Clapeyron relationship, Held and Soden (2006)) but can also have an indirect impact in extreme precipitation through changing circulation patterns (e.g., displacement of circulation systems). Despite this obvious deficiency, the separation between dynamical and thermodynamic changes is used particularly in event attribution studies to better understand the underlying processes contributing to a specific extreme event (Mitchell et al., 2016, Vautard et al., 2016, Yiou et al., 2017) or in a recent study by Pfahl et al. (2017) to better understand regional changes in extreme precipitation. Two examples from the event attribution where such a separation proved useful were the extreme precipitation event in the UK in winter 2014/15 (Schaller et al., 2016) and the European heatwave in summer 2003 (Mitchell et al., 2016). In these studies, it was illustrated that both dynamical and thermodynamic processes (e.g., changes in atmospheric patterns and soil moisture) can be respectively relevant for generating an extreme event. It is important to mechanistically assess the physical characteristics of the extreme event in terms of what processes are mainly driving the event occurrence and whether they are affected by anthropogenic forcing (e.g., Hauser et al. (2016)).

In Europe, for instance, extreme temperatures can occur during atmospheric blocking conditions, whereas the processes driving the temperature extremes differ for summer (local processes) and winter (advection of cold air) (Pfahl and Wernli, 2012). In addition, processes can even differ within a season, such as described in (Brunner et al., 2017) for spring, but even this may change with global warming (Cassou and Cattiaux, 2016). There seems to be a distinct regional dependency of the relationship between blocking anticyclone locations and the corresponding surface heat or cold extreme event as illustrated in Fig. 2 (Bieli et al., 2015).

Furthermore, the weakening of the equator-to-pole temperature gradient due to global warming, particularly in boreal summer, is associated with a decrease in eddy kinetic energy (EKE), a measure of transient wave activity (Lehmann and Coumou, 2015). This may lead to more persistent summer weather and enhanced anti-cyclonic flow regimes in some regions. The European summers of 2003 and 2010 are good examples in which high-amplitude quasi-stationary waves were associated with extreme heat waves (Coumou et al., 2015). Climate models need to be evaluated based on their performance simulating such kinds of underlying large-scale processes to be able for us to have confidence in their representation of related surface extremes.

This said, climate models can have large biases in some regions and may not be able to simulate key dynamical patterns such as atmospheric blocking or other weather regimes, jet stream position and intensity, tropical dynamics and teleconnections, or stratosphere-troposphere connections. A key challenge is to evaluate and improve models by targeting key processes that are relevant for a realistic, or at least a sufficient, representation of extremes. Approaches to improve them include developing theories and hierarchies of models to untangle complex processes, further increasing model resolution and using novel approaches for parameterizing sub-grid scale processes.

A combination of high-resolution simulations with lower resolution ensemble simulations would be beneficial to study effects of internal variability and better quantify the signal-to-noise ratio, particularly for precipitation extremes (Palmer, 2014). Such efforts require an international collaboration to pool resources for coordinated high-resolution modelling on a global scale (such as in PRIMAVERA, https://www.primavera-h2020.eu/ or HighResMIP (Haarsma et al., 2016)). Detecting, and even predicting the changes in large-scale circulation is a major challenge to be overcome in order to better predict changes in the odds of

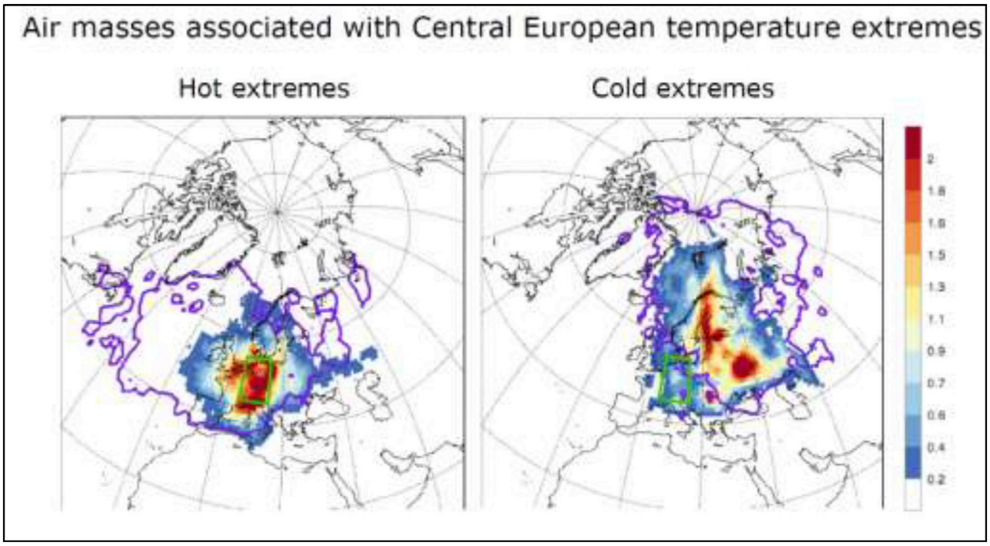

Fig. 2. Trajectory densities (number of trajectories per area of 1,000 km^2) four days prior to a hot or cold event at the location marked with a green rectangle in Central Europe. The contour line (purple) refers to the distribution of trajectories seven days before the warm or cold events, representing a density level of 0.2. Adapted from Bieli et al. (2015). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extreme events. This has raised scientific interest in the past decade or so, but the topic remains open and challenging (Shepherd 2014).

Lastly, probabilistic event attribution, the effort to probabilistically ascertain the mechanisms responsible for changes in the climate related to specific events, is a topic that depends largely on the ability of models to simulate extremes. Current attribution studies generally attempt to identify if the odds of the occurrence of a specific type/class/category of (extreme) event are changed by enhanced greenhouse gas forcing. However, the framing of the attribution question is crucial and affects the result in terms of spatial and temporal scales under consideration (Trenberth et al., 2015; Stott et al., 2016; Angélil et al., 2014) and further sensitivities arise related to choice of observational product and model (Angélil et al., 2016; Bellprat and Doblas-Reyes, 2016).

Another challenge with probabilistic event attribution is to clearly identify the part of the signal that is due to a change in dynamics (which as stated above is complex to unravel). Analog-based methods for this kind of identification now exist (Vautard et al., 2016; Yiou et al., 2017), which need to be developed in various contexts and in combination with evaluation of model-simulated changes in circulation. In order to identify thermodynamic changes, a circulation analogue method can also be used for attribution conditional to the large-scale flow (Cattiaux et al., 2011; Yiou and Cattiaux, 2013, 2014). However, a potential drawback of this method is that some events may not have enough good analogues, loosening the flow-conditioning, and therefore the identification of a thermodynamic signal. In addition, analogues are defined on past conditions and thus might not be representative of future ones. Recent trends in analogues may give an indication of changes in the atmospheric dynamics itself although this remains to be explained and disentangled from long-term natural variability.

2.2. Local-to-regional feedback processes and drivers of extreme events

In addition to large-scale drivers, understanding specific processes acting at local to regional scales is also essential to understand the evolution of extreme events. For instance, in early summer additional mechanisms, such as soil moisture conditions, are relevant for the evolution of a heat wave (Seneviratne et al., 2010; Quesada et al., 2012). Therefore, controlled experiments are planned to study the effects of circulation, soil moisture and sea surface temperature (SST) in the formation of heat waves in the future coordinated experiment called ExtremeX. For soil moisture, a similar set-up as the one from the GLACE (Koster et al., 2004) and GLACE-CMIP5 (Seneviratne et al., 2013) experiments is considered. Further examples of such relevant processes and feedbacks are subject of the SNOWGLACE experiment (http://uni.no/en/uni-climate/climate-services/snowglace/) that studies the impact of snow on sub-seasonal-to-seasonal forecast by "realistic" snow initialization (Orsolini et al., 2013).

A misrepresentation of feedback mechanisms in the models can be an important source of uncertainty for future projections, e.g., for heat waves or droughts. For instance, soil moisture and associated processes and feedbacks are very uncertain in current generation climate models (e.g., Taylor et al., (2012); Orlowsky and Seneviratne (2013); Mueller and Seneviratne (2014); Stegehuis et al., (2013)). Constraining soil moisture in model experiments may be as important as other factors, such as climate sensitivity and cloud feedbacks, for generating or enhancing extremes on a local-to-regional scale (see also Hurk et al. (2016)). Feedback mechanisms over land can be reflected, for instance, in changes in the shape of the temperature distribution, not only in a shift in the mean (Douville et al., 2016). This is often associated with a stronger warming of climate extremes (Seneviratne et al., 2016).

Satellite records can be very useful for statistical analysis and process understanding in conjunction with simple mechanistic models to help interpret this kind of measurement (Beck et al., 2017; Papagiannopoulou et al., 2016). The improved process understanding could then be applied to benchmark complex models. For instance, remote sensing data were linked to simplified mechanistic models to study coinciding periods of "dry soil" and "high temperature" (Miralles et al., 2014). A clear spatial correlation between antecedent soil moisture (drier) conditions and heat wave temperatures was shown for the European heatwaves of 2003 and 2010 (see Fig. 3). However, correlation does not necessarily imply causation and, thus, needs to be carefully investigated.

Furthermore, high-resolution climate model simulations are required to study feedback-driven preconditioning of extreme events, e.g., feedback of soil moisture/snow on circulation patterns, such as shown for small-scale thunderstorms over Lake Victoria and their changes in the future (Thiery et al., 2016). A process analysis to separate "lake" and "land" events and looking at thermodynamic versus meso-scale dynamic contribution to these storms showed that the dynamic processes dominate in current climate conditions, but also implied changes in this relationship with climate warming. Utilizing this process understanding in developing early warning systems could help prevent loss of life.

To realistically represent small-scale processes (e.g., convective storms, orographic rain), some recent climate change experiments have employed km-scale (e.g., 1.5 km resolution) regional models over

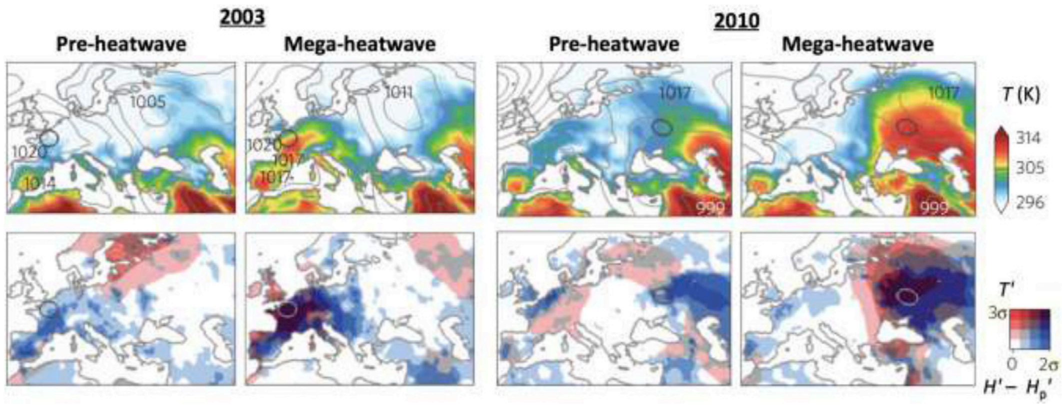

Fig. 3. Air temperature and soil moisture in Europe during recent mega-heatwave summers. Data for 10-day pre-heatwave and mega-heatwave periods in 2003 and 2010. Top: Average afternoon near-surface air temperature (T, K) and mean sea-level pressure (hPa) from ERA-Interim reanalysis. Bottom: Co-variability of the T anomalies (T′, σ) in red, and the anomalies in the contribution of soil moisture deficit to the surface sensible heat flux (H' – Hp', σ) in blue. The areas of 200 km radius around two locations, Trappes (France, 2003) and Voronezh (Russia, 2010) are marked in white contours. For details see Miralles et al. (2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Europe. Similar convection permitting models have already been widely used in numerical weather forecasting. These experiments show that while mean changes in precipitation are not affected, the very fine resolution is useful to better simulate summer precipitation, in particular convective events (i.e., hourly precipitation, extremes, steep orography) (Ban et al., 2014; Kendon et al., 2014, Prein et al., 2017). So far there are only a few studies that investigate climate extremes from long time periods or a large number of cases using such high-resolution model experiments. To be able to compare model results and assess the robustness of present and future climate projections, coordinated high-resolution (i.e., convection permitting) modelling experiments are urgently needed. Some experiments of this type are now being undertaken within the new framework of EURO-CORDEX (Jacob et al., 2014) and a flagship project study dedicated to convection with high resolution models has started (Coppola and Sobolowski, 2017). However, there are some limitations associated with these high-resolution simulations, such as high computational costs for relatively short time slice simulations, dependency on the driving global climate model and multi-nesting of domains. Some of these can possibly be overcome with future high-performance computing developments.

Going from annual to monthly or daily to sub-daily scales in the analysis of extremes will help to better understand the temporal variability and changes in extremes. For instance, we need to reduce systematic biases in model simulations related to the seasonal cycle of precipitation and evapotranspiration, and to better simulate the diurnal cycle of precipitation. The benefit of going to higher spatial resolution versus having a larger number of ensembles to study the effect of variability and a robust ensemble statistics is intensely debated and requires further research. For some types of extreme events, such as convective precipitation, the gains from increased resolution (i.e. 3 km and less) are obvious, but for other types of extremes, such as droughts or heat waves, it is not.

However, the main limiting factor for the analysis of such small-scale processes remains the availability of observational data. Collecting high-frequency sub-daily precipitation datasets over long time periods will be extremely useful for attribution of, for example, damaging convective rainfalls. In many regions of the world (e.g., Africa, South America, Asia) even daily temperature and precipitation data are non-existent or not publicly accessible, and the situation is even worse for other variables, such as wind or soil moisture measurements in most locations (Alexander et al., 2016). This challenges any significant progress in process understanding and model evaluation, particularly when trying to underpin results from high-resolution model experiments with corresponding observations. Besides that, the value of single measurements (i.e., case studies of single events) versus robust statistics over a series of similar events or regions for model evaluation needs further investigation. Particularly, the assessment of trends in changes of extremes is hampered by the limited data availability and quality (see also the WCRP workshop on Data Requirements http://www.wcrp-climate.org/index.php/extremes-data-wkshp-about).

2.3. Predictability of extremes

The long-term goal of climate prediction and projection is to provide society with useful information about future changes in weather and climate. User needs must therefore be considered when determining and developing the products and climate services for them. Thus, it is also most appropriate to understand the predictability of "user-relevant" extremes. Understanding predictability also improves confidence for future projection. Defining user-relevant weather and climate extremes requires a climate services approach involving a dialogue between users and providers, because what forecasters currently provide is far from what some users demand (see, for example, http://www.euporias.eu/). To that end the World Meteorological Organization (WMO) set up the Global Framework on Climate Services (GFCS) that, for example, provides a worldwide mechanism for coordinated actions to enhance the quality, quantity and application of climate services. Under this umbrella, there are index definitions from e.g., CCl/WCRP/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI; https://www.wcrp-climate.org/etccdi) and the ET on Sector-specific Climate Indices (ET-SCI; http://www.wmo.int/pages/prog/wcp/ccl/opace/opace4/ET-SCI-4-1.php) that could be complemented with indices used by weather forecasters.

Communicating predictions is a challenge. In general, changes in extremes are communicated in terms of their frequency, probability of occurrence, and intensity. In addition, users are often interested in impact-related parameters, such as flood level, heat stress, and water availability. Two metrics that are often used in attribution studies are the "Fraction of attributable risk (FAR)" and the 'risk ratio (RR)', where FAR might be highly sensitive to the level of signal-to-noise ratio. Additionally, FAR is meaningful only when an external factor has resulted in the increase in the frequency of extreme events. As a result, RR has been suggested as an alternative in a recent report of the US National Academies (NAS, 2016). In this context, RR is just a ratio of frequencies of occurrence, and does not include any measure for damage, vulnerability or exposure, which are included in more comprehensive risk definitions (e.g., Oppenheimer et al. (2014)). It is also critically important to convey the uncertainties and skill levels in the prediction of indices and metrics to the users (Bhend et al., 2017). For this setting, Gneiting et al. (2007) state that the prediction goal should be to maximize the sharpness of the prediction subject to calibration, or reliability. Reliability is a joint property of the prediction and the observed event in that predicted probabilities should match the observed frequencies of an event. Reliability is the most important measure of prediction usefulness for decision-making (Weisheimer and Palmer, 2014) and the EUPORIAS and SPECS projects showed that it is the key measure of probabilistic prediction skill for users. Sharpness, however, refers to the amount of prediction uncertainty and is a property of the probabilistic prediction only. The principle proposed by Gneiting et al. (2007) thus says that given two equally reliable prediction methods, the less uncertain method should be preferred.

The predictability of extremes can be understood in terms of the factors controlling their development (see Fig. 1) and depends on the relative importance or contribution of these factors. On seasonal timescales, much of the skill in predicting temperature and precipitation extremes arises from skill in predicting seasonal mean temperature and precipitation, and the consistent large-scale drivers causing shifts in temperature and precipitation distributions. On these short timescales, for example, in some regions most of the skill in predicting large-scale drivers is derived from the El Niño-Southern Oscillation (ENSO) (van Oldenborgh et al., 2005), while changes in forcing emerge as sources of skill even within a decade (e.g., Hanlon et al. (2013)). Another potential source of skill might be associated with sea ice variability. For example, winter NH variability has recently been linked to autumn and early winter sea ice anomalies through both tropospheric and stratospheric pathways, which are partially captured by models (García-Serrano et al., 2015, 2016; King et al., 2016). Using the temperature of the previous month has recently been shown to be a better predictor than common large-scale patterns (e.g., the North Atlantic Oscillation) for predicting month-to-month persistence of winter and summer temperature extremes over Europe (Kolstad et al., 2015). This process is not yet fully understood, but could be related to land-atmosphere feedbacks associated with soil moisture or snow, as highlighted in the previous section.

Numerical climate prediction is developing rapidly and providing new possibilities to predict user-relevant extremes. For example, there is now some consensus that wintertime variations in the NAO can be predicted a season in advance, because of models improvements, increased computer power, better observations and data assimilation methods (Scaife et al., 2014; Weisheimer et al., 2017; Athanasiadis et al., 2017). There is also a growing interest in subseasonal forecasts, and a wealth of data from operational predictions is available via the WWRP/WCRP Subseasonal to Seasonal (S2S) Prediction Project (Vitart et al., 2017). Our improving ability to simulate and predict the Madden Julian Oscillation (Lee et al., 2017) is a step towards realizing sub-seasonal prediction of weather extremes in both the tropics and extra-tropics (Matsueda and Takaya, 2015). Multi-year dynamical predictions also show promise. For example, there is skill in predicting Atlantic tropical cyclone variations (Smith et al., 2010; Caron et al., 2015), drought conditions over the Sahel (Mohino et al., 2016; Sheen et al., 2017), and wild fire wildfire probabilities in southwestern North America (Chikamoto et al., 2017) a few years in advance.

Case studies may be the most effective means to estimate predictability of very rare extreme events and to assess model performance. Examples include seasonal predictions of the 2003 European dry summer and the 2013/2014 NH cold winters. In the first case, Fig. 4 shows the predictability of the event was more related to *in situ* processes that helped maintain the dry surface anomalies occurring at the beginning of the summer, rather than to remote teleconnection effects (Weisheimer et al., 2011; Prodhomme et al., 2016). In the second case, tropical processes seem to be the most relevant cause for the 2013/14 cold winter (Watson et al., 2016). Since models suffer from large systematic errors that can adversely affect the prediction skill of climate and weather extremes, case studies of this type can be very useful to identify the key processes that need to be better represented to enhance prediction skill. Two key areas for future research can be identified based on these two examples: ocean-atmosphere interaction in the tropics, and land surface

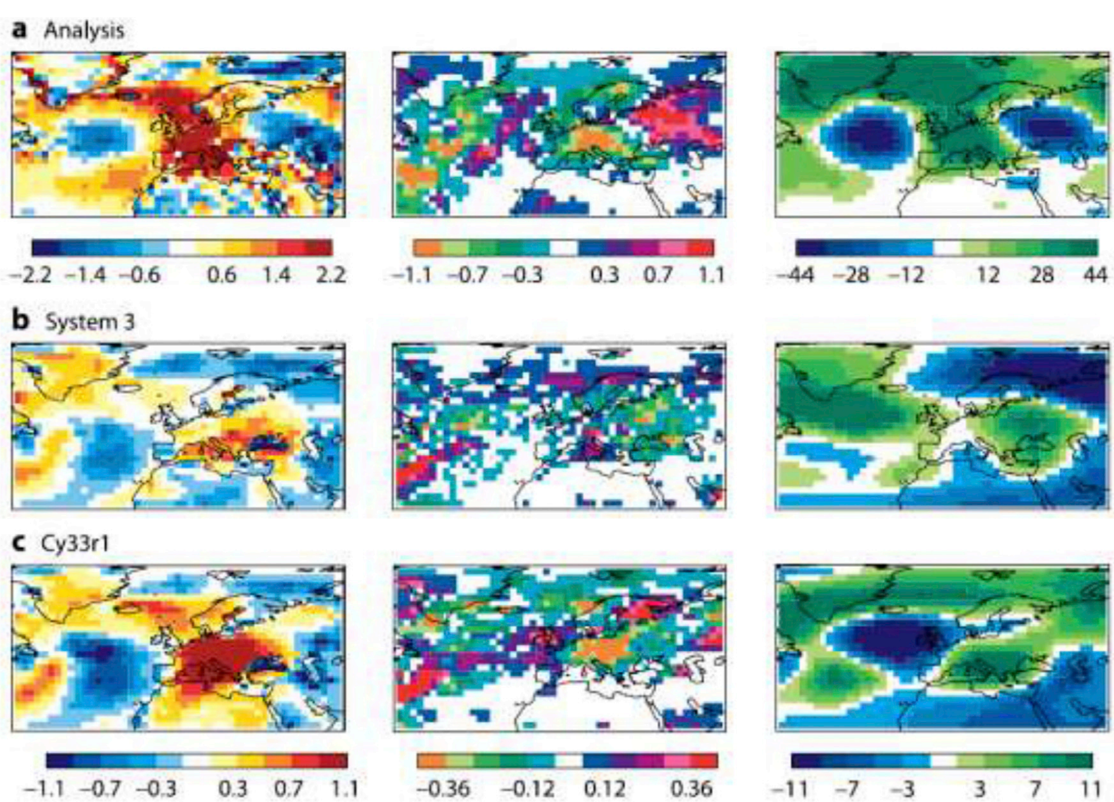

Fig. 4. Anomalies of 2 m-surface temperature (T2m), precipitation and 500 m geopotential height (Z500) in (a) the verification dataset, (b) the ECMWF's forecasting system S3 (operational since 2007) and (c) the updated seasonal re-forecast system (CY33R1) with sophisticated formulations of land surface hydrology, radiation and convection. The combination of all three processes proved key to the successful retrospective predictions of the record heat wave in Europe in June-July-August 2003. Adapted from (Weisheimer et al., 2011).

processes. There is currently also active research on the assessment of predictability arising from extra-tropical SST and Arctic sea ice changes on short and long-timescales (e.g., the NordForsk GREENICE project; www.greenice.no).

Given the limited number of events and data availability, one way forward to improve our understanding to enhance prediction skill of climate and weather extremes would be to focus on mechanisms (as conceptualized in Fig. 1). The greatest predictability is expected for events for which large-scale thermodynamic changes dominate or for which dynamical process are linked to predictable climate variability (such as ENSO). The predictability of extreme events can be assessed through analyzing both dynamic and thermodynamic factors controlling their initiation and evolution, and by quantifying the contributions from local feedbacks versus remote effects. Overall, it is important to perform numerical experiments to confirm case study analyses. Not to be forgotten in this context, is the important issue of bias corrections on the simulation of extremes, which has important implications for use in impact modeling (e.g., Sippel et al. (2016)). A necessary basis for any bias correction, however, is an adequate evaluation of extreme events in model simulations (see Sec. 2.4). Finally, our ability to predict extremes or improve the predictability of such will be hampered by making assumptions about linearity between processes, i.e. that non-linear interaction between mean and variability can be neglected.

2.4. Model performance and process-based evaluation of climate extremes

When evaluating extremes in model simulations, both statistics as well as underlying processes should be considered in conjunction. There should be an emphasis on evaluating the contributing mechanisms to determine realism and to assess systematic model errors. In general, models should avoid biases in the frequency, amplitude and duration of events (e.g., Hanlon et al. (2013)) and should be evaluated for multiple types of extremes. Mechanisms can be evaluated in terms of thermodynamics, large-scale circulation, and local feedbacks. While it is important to analyze models in terms of their performance in simulating individual events to understand model performance, such analyses might not be informative on the general skill of the model to predict such events given the important role of internal variability.

Internal variability represents a major challenge to the evaluation of extremes and the assessments of their trends (e.g., Sillmann et al. (2014)). The latter can be addressed, for instance, by spatial aggregation. Recent work shows that model agreement on the forced response of precipitation and temperature extremes is higher than widely recognized (Fischer et al., 2013). Also, large-scale changes in extreme temperature show clear evidence of a human influence (Bindoff et al., 2013; Morak et al., 2013, Kim et al., 2016). Similarly, observations show evidence of a human influence on changes in precipitation extremes, specifically a widespread long-term increase in intensity (see Min et al. (2011), Zhang et al., 2013). This daily precipitation intensification is consistent across observations and model hierarchies.

To evaluate if models simulate changes in extreme events for the right reason, multiple events that are less intense but similar in nature can be used to determine the dominant drivers of a class of extreme events. This process can further serve to evaluate to what extent models are able to reproduce these mechanisms for moderate extremes (e.g., Krueger et al., (2015)). While the general weather situation leading to warm and cold spells appears reasonably well simulated by models, various issues have been raised. This includes the overestimation or underestimation of precipitation amounts due to parameterized convection (e.g., Stephens et al. (2010)); the reliability of trends in observations (e.g., Donat et al. (2016)); heterogeneities and gridding issues with observations (e.g., Herold et al. (2016)); and model deficiencies in the representation of driving processes (e.g., atmospheric blocking, boundary layer dynamics and land-atmosphere interactions) (e.g., Vial and Osborn, (2012)).

It is often claimed that regional climate models (RCMs) better represent extreme events than global models. In a study focusing on the evaluation of dynamical or dynamically-influenced phenomena, Whan et al. (2016) find that the added value partly depends on the RCM's ability to generate additional internal variability. In particular, RCMs appear to model the effects of atmospheric blocking on minimum temperature reasonably well over the large CORDEX North America domain. Model evaluation in this context should reflect whether large-scale circulation influences on extremes are well simulated, and whether they improve biases in simulating, for instance, extreme precipitation or intense storms.

Feature-based methods provide a potentially powerful way of evaluating particular events and these are commonly employed in numerical weather prediction. One option is the method of Object-based Diagnostic Evaluation that uses the identification of objects and their evaluation with respect to different features such as location, intensity, shape, area or orientation (Mittermaier and Bullock, 2013). These objects could be, for example, large-scale circulation patterns, clouds or small-scale feedback processes. However, in particular for small-scale processes, it may be challenging to find adequate gridded data sets.

When assessing the general performance of model ensembles, the focus should be on a probabilistic rather than a deterministic assessment of model ensembles in terms of their ability to represent climatological statistics as the entire ensemble distribution reflects the ability to simulate the climate realistically. Probabilistic assessment can be performed using so-called proper scoring rules that assign a numerical score to each prediction-observation pair (Gneiting and Raftery, 2007). Average scores over many such pairs can then be used to rank competing models. One example is the root mean square error (RMSE), which is often used as an evaluation metric in climate model comparison studies (e.g., Sillmann et al. (2013)). However, it evaluates only the average of the predictions against the observation and, thus, does not account for the prediction uncertainty or the uttermost tails (i.e., extremes) of a distribution.

Alternatively, fair scores favor optimal ensembles and can be chosen to evaluate specific features (e.g., mean, variance) of the ensemble (Ferro, 2014). Commonly, multiple scores are needed to identify which features are not being simulated correctly. If the focus of the evaluation is on the simulations of extreme events, additional considerations need to be taken into account. Here, weighted scores that put emphasis on extremes should be used (Lerch et al., 2015). However, the scores may hide key information such as whether a model has a negative or a positive bias. Important open questions on this topic to be addressed in future research include: Are existing scores sensitive to differences in the underlying processes? What ensemble sizes are needed to detect differences? How can we handle observation error or lack of observations? What other forecast evaluation methods are useful for assessing model performance? Such investigations would benefit from close collaboration of climate and weather forecast modelers and statisticians (see also Thorarinsdottir et al. (2014) and Benestad et al. (2017)).

Model evaluation is inevitably dependent on the quality and availability of the underlying reference datasets (e.g., reanalysis or observations) and evaluation methods should account for uncertainty in the observations. While large datasets already exist for the evaluation of model performance for a broad suite of extremes, both for case studies and aggregates of moderate extremes, new data sources (e.g., satellite or remote sensing data) and variables should be explored and exploited to improve model evaluation. The community is encouraged to perform analyses to identify models that represent a particular mechanism well or poorly, and quantify its effect on the simulation of extremes (e.g., stratosphere-troposphere interaction). Results should be confirmed by coordinated sensitivity experiments where key processes can be identified and their representation can be improved.

3. Addressing the scientific challenges on different time scales

Weather and climate extreme events naturally cluster into two classes or categories related to the temporal scales on which they can occur, and which involve different models, processes, and research questions.

Therefore it is sensible to focus on short-duration (less than three days) and long-duration (weeks to months or even years) extreme events, their different mechanisms and differing approaches to evaluation and prediction. The following five questions provide guidance for addressing scientific challenges:

a) What are relevant definitions of extremes on the respective time scales?
b) What are the necessary observations and model output requirements to analyze these extremes?
c) What are the processes driving these extremes and their changes?
d) How do we best evaluate these extremes (including relevant processes)? (i.e., is the model right for the right reason)
e) What are relevant sources for predictability of these events that can support the attribution, prediction and projection of these extremes?

While it is possible to partly address the first three questions directly based on our knowledge as summarized below, the last two questions suggest future research directions to improve our understanding of extremes and associated underlying mechanisms.

3.1. Short-duration extreme events

Short-duration Extremes (SDE) are here defined as rare hazardous meteorological phenomena occurring over time scales up to 3 days with the relevant processes indicated in Fig. 1 (blue curve). SDEs may include (i) Convective events leading to heavy precipitation, hail, lightning, tornadoes, violent downdrafts; (ii) Extra-tropical cyclones leading to wind storms, storm surges, extreme precipitation (rainfall or snowfall), freezing rain; (iii) Anticyclones leading to fog and air pollution, cold outbreaks, long-lived heat waves and extended cold spells; (iv) Tropical cyclones.

A conceptual framework to study long-term changes in SDEs includes the challenges of (i) detection of trends in the frequency and intensity of SDEs, (ii) attribution of changes in frequency or intensity to various anthropogenic or natural forcings, and (iii) projection of their evolution into the future. Each of these challenges is difficult because observational data is scarce, unevenly distributed and often not long and/or homogeneous enough for trend analysis. Further, some phenomena (e.g., hail, fog and lightning) are of a scale that is too small for climate models and long-term simulations to represent. However, a few areas exist for which knowledge, data and models are sufficient to provide us with results within a timeframe of about 2 years. For instance, progress in understanding, modelling and attribution could be achieved for sub-daily precipitation events in the limited regions where high quality, long-term, hourly precipitation measurements are available in many different types of climate (Westra et al., 2014; Blenkinsop et al., 2017). With suitable data rescue efforts substantially more data could become available in the future (Brunet and Jones, 2011).

3.2. Long-duration extreme events

Long-duration Extremes (LDE) are here defined as events lasting longer than 3 days with the relevant processes illustrated in Fig. 1 (red curve). Examples include drought, heat waves, cold spells and floods caused by persistent rainfall, but also extreme low Arctic sea ice extent, increased storminess and wildfire seasons. Processes that drive and influence the frequency, duration and intensity of LDE include, for instance, tropical SST forcing (e.g., associated with ENSO), Arctic sea ice changes, stratospheric conditions, land-surface (e.g., soil moisture, snow cover) conditions and land-atmospheric feedbacks, as well as external anthropogenic and natural drivers (e.g., atmospheric composition, land-use changes, solar forcing).

Therefore, relevant sources for predictability of LDE could be SSTs, soil moisture, snow cover, stratospheric conditions, vegetation, greenhouse gases, and aerosols. Important aspects to investigate are whether the relevant sources of predictability are linear (additive) or not, and whether predictability can be understood in terms of thermodynamic, large-scale precursors, and local feedbacks. The relevance of different sources may be time scale dependent. For instance, the influence of soil moisture may not be the same for seasonal forecasts and long-term projections.

As for the SDE in section 3.1, the types of events we can analyze often depend on data availability. Measurements of daily temperature and precipitation are the best data we have currently for assessing extremes, but not for all regions. However, there are other relevant observed variables to define some of the extremes mentioned, such as soil moisture, humidity and maximum wind speed (surface, 850 h Pa, 200 h Pa) that are not as readily available. In some cases we could address this issue through coordinated digitization initiatives while in other cases measurement networks would need to be expanded (Alexander et al., 2016; Seneviratne et al., 2010). Surface data in particular are very important for the evaluation of satellite and model-based product data (e.g., Sapiano and Arkin (2010); Sillmann et al., (2013)), but in some cases observation-driven model output data products are better than indirect measurements (e.g., satellite-based products for snow and soil moisture (Beck et al., 2017). Furthermore, data assimilation uses integration techniques that can give smaller uncertainty than for each individual measurement (e.g., Vila et al. (2009)). Model evaluation requires case studies and systematic studies with large samples to identify mechanisms, model biases and to avoid over-interpreting aspects unique to a particular event. Examination of individual cases is challenging to generalize across numerous cases and requires the development of diagnostics that can be reliably applied across multiple situations.

4. Conclusion and outlook

A workshop as part of the implementation of the WCRP Extremes Grand Challenge has led to the discussions and outcomes presented in this paper. We have outlined an urgent need for better observations and model evaluation tools that are specifically suited to the analysis of extremes. Developing these requires a dedicated cross-community effort. Many of the processes underpinning the evolution of weather and climate extremes are yet not fully understood. Coordinated model experiments should be set-up for exploring dynamic and thermodynamic drivers of extremes. As both large-scale circulation and local-to-regional feedback processes are important for the generation of extreme events (see Fig. 1), these phenomena should be studied in conjunction to improve process understanding and predictability of extremes. Extreme events occurring at temporal and spatial scales much smaller than that of current state-of-the-art climate models are generally difficult to predict, but there is certainly potential for long-duration extremes (particularly on monthly or seasonal scales). Future studies are needed to assess the benefits of going to higher spatial resolution versus having a larger number of ensembles to study the effect of variability and robust ensemble statistics for various types of extremes. Joint research and model development across scales involving climate and numerical weather prediction models will be crucial to make progress in our understanding of what level of predictability can be expected and achieved for extreme events.

While this overview focuses on the physical science challenges in understanding and predicting extreme weather and climate events, the socio-economic and ecological challenges associated with this topic are equally important. In the context of climate services, the advances in the physical sciences are essential to serve public and private sectors, such as renewable energy, (re-)insurance, health, agriculture, infrastructure planning, which are strongly impacted by extreme weather and climate events. Cross-community research activities in collaboration with relevant stakeholders can stimulate new and exciting research questions that can pave the way to build more resilience to the risk of extreme events and future climate change. In this context, a new Future Earth initiative (http://www.e3s-future-earth.eu/index.php/Main/Home) for a Knowledge-Action-Network (KAN) on Emergent Risks and Extreme

Events is proposed to integrate research aspects from Future Earth, the Disaster Risk Community (e.g., Integrated Research on Disaster Risk (IRDR)) and the WCRP Extremes Grand Challenge to advance interdisciplinary research on reducing risks from climate-related disasters.

Acknowledgements

We thank all M-CLIX workshop participants for their contributions and the very engaged and insightful discussions that enabled us to write this paper. The workshop was supported by funding from WCRP, the Norwegian Environment Department, and the Norwegian Research Council through projects ClimateXL (grant 243953), MCLIX (grant 240751) and HDwave (grant 243814). NK acknowledges support from the NordForsk GREENICE Project (grant 61841). This work contributes to the World Climate Research Programme Grand Challenge on Weather and Climate Extremes.

Appendix A

The M-CLIX workshop was held in Oslo, Norway from October 5 to 7, 2015. It was structured in a poster and four oral sessions with presentations and discussions related to large-scale drivers of extreme events (session A), local-to-regional drivers and feedbacks (session B), predictability of extremes (session C), and model performance (session D). Workshop presentations from the participants are publicly available through the WCRP workshop website (https://www.wcrp-climate.org/extremes-modeling-wkshp-agenda-presentations).

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.wace.2017.10.003.

References

Alexander, L.V., Zhang, X., Hegerl, G., Seneviratne, S.I., 2016. Implementation Plan for WCRP Grand Challenge on Understanding and Predicting Weather and Climate Extremes – the "Extremes Grand Challenge". Version, June 2016 available from: https://www.wcrp-climate.org/images/documents/grand_challenges/WCRP_Grand_Challenge_Extremes_Implementation_Plan_v20160708.pdf.

Angélil, O., Stone, D.A., Tadross, M., Tummon, F., Wehner, M., Knutti, R., 2014. Attribution of extreme weather to anthropogenic greenhouse gas emissions: sensitivity to spatial and temporal scales. Geophys. Res. Lett. 41 (6), 2150–2155.

Angélil, O., Perkins-Kirkpatrick, S., Alexander, L.V., Stone, D., Donat, M.G., Wehner, M., Shiogama, H., Ciavarella, A., Christidis, N., 2016. Comparing regional precipitation and temperature extremes in climate model and reanalysis products. Weather Clim. Extrem. 13, 35–43.

Athanasiadis, P.J., Bellucci, A., Scaife, A.A., Hermanson, L., Materia, S., Sanna, A., Borrelli, A., MacLachlan, C., Gualdi, S., 2017. A multisystem view of wintertime NAO seasonal predictions. J. Clim. 30 (4), 1461–1475.

Ban, N., Schmidli, J., Schär, C., 2014. Evaluation of the convection-resolving regional climate modelling approach in decade-long simulations. J. Geophys. Res. Atmos. 119, 7889–7907.

Bellprat, O., Doblas-Reyes, F., 2016. Attribution of extreme weather and climate events overestimated by unreliable climate simulations. Geophys. Res. Lett. 43 (5), 2158–2164.

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615.

Benestad, R., Sillmann, J., Thorarinsdottir, T.L., Guttorp, P., Mesquita, M.D.S., Tye, M.R., Uotila, P., Fox Maule, C., Thejll, P., Drews, M., Parding, K.M., 2017. New vigour involving statisticians to overcome ensemble fatigue. Nat. Clim. Change 7, 697–703.

Bhend, J., Mahlstein, I., Liniger, M.A., 2017. Predictive skill of climate indices compared to mean quantities in seasonal forecasts. Q. J. R. Meteorol. Soc. 143, 184–194. https://doi.org/10.1002/qj.2908.

Bieli, M., Pfahl, S., Wernli, H., 2015. A Lagrangian investigation of hot and cold temperature extremes in Europe,. Q. J. R. Meteorol. Soc. 141 (686), 98–108.

Bindoff, N.L., et al., 2013. Detection and attribution of climate change: from global to regional. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 867–952.

Blenkinsop, S., Lewis, E., Chan, S.C., Fowler, H.J., 2017. Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. Int. J. Climatol. 37, 722–740. https://doi.org/10.1002/joc.4735.

Brunet, M., Jones, P., 2011. Data rescue initiatives: bringing historical climate data into the 21st century. Clim. Res. 47 (1–2), 29–40.

Brunner, L., Hegerl, G.C., Steiner, A.K., 2017. Connecting atmospheric blocking to european temperature extremes in spring. J. Clim. 30 (2), 585–594.

Caron, L.-P., Hermanson, L., Doblas-Reyes, F.J., 2015. Multiannual forecasts of Atlantic U.S. tropical cyclone wind damage potential. Geophys. Res. Lett. 42, 2417–2425.

Cassou, C., Cattiaux, J., 2016. Disruption of the European climate seasonal clock in a warming world. Nat. Clim. Change 6, 589–594.

Cattiaux, J., Vautard, R., Yiou, P., 2011. North-Atlantic SST amplified recent wintertime European land temperature extremes and trends. Clim. Dyn. 36 (11–12), 2113–2128.

Chikamoto, Y., Timmermann, A., Widlansky, M.J., Balmaseda, M.A., Stott, L., 2017. Multi-year predictability of climate, drought, and wildfire in southwestern North America. Sci. Rep. 7 (1), 6568.

Coppola, E., Sobolowski, S., 2017. Convective Phenomena at High Resolution over Europe and the Mediterranean. The Joint EURO-CORDEX and Med-CORDEX Flagship Pilot Study. Geophysical Research Abstracts, Vol. 19, EGU2017–2547.

Coumou, D., Lehmann, J., Beckmann, J., 2015. The weakening summer circulation in the Northern Hemisphere mid-latitudes. Science 348 (6232), 324–327.

Donat, M.G., Alexander, L.V., Herold, N., Dittus, A.J., 2016. Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. J. Geophys. Res. Atmos. 121 (19), 11174–11189.

Douville, H., Colin, J., Krug, E., Cattiaux, J., Thao, S., 2016. Midlatitude daily summer temperatures reshaped by soil moisture under climate change. Geophys. Res. Lett. 43 (2), 812–818.

Ferro, C., 2014. Fair scores for ensemble forecasts. Q. J. R. Meteorol. Soc. 140 (683), 1917–1923.

IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, and New York, NY, USA, p. 582.

Fischer, E.M., Beyerle, U., Knutti, R., 2013. Robust spatially aggregated projections of climate extremes. Nat. Clim. Change 3 (12), 1033–1038.

García-Serrano, J., Frankignoul, C., Gastineau, G., De La Càmara, A., 2015. On the predictability of the winter Euro-Atlantic climate: lagged influence of autumn Arctic sea ice. J. Clim. 28 (13), 5195–5216.

García-Serrano, J., Frankignoul, C., King, M., Arribas, A., Gao, Y., Guemas, V., Matei, D., Msadek, R., Park, W., Sanchez-Gomez, E., 2016. Multi-model assessment of linkages between eastern Arctic sea-ice variability and the Euro-Atlantic atmospheric circulation in current climate. Clim. Dyn. 1–23.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat. Methodol. 69 (2), 243–268.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.

Haarsma, R.J., Roberts, M.J., Vidale, P.L., Senior, C.A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N.S., Guemas, V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L.R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M.S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., Von Storch, J.-S., 2016. High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. Geosci. Model Dev. 9 (11), 4185–4208.

Hanlon, H.M., Hegerl, G.C., Tett, S.F., Smith, D.M., 2013. Can a decadal forecasting system predict temperature extreme indices? J. Clim. 26 (11), 3728–3744.

Hauser, M., Orth, R., Seneviratne, S.I., 2016. Role of soil moisture vs. Recent climate change for the 2010 heat wave in Western Russia. Geophys. Res. Lett. 43, 2819–2826.

Held, I.M., Soden, B.J., 2006. Robust responses of the hydrological cycle to global warming. J. Clim. 19, 5686–5699.

Herold, N., Alexander, L.V., Donat, M.G., Contractor, S., Becker, A., 2016. How much does it rain over land? Geophys. Res. Lett. 43 (1), 341–348.

Hurk, B., Kim, H., Krinner, G., Seneviratne, S.I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M.J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D.M., Weedon, G.P., Ellis, R., Hagemann, S., Mao, J., Flanner, M.G., Zampieri, M., Materia, S., Law, R.M., Sheffield, J., 2016. LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome. Geosci. Model Dev. 9, 2809–2832.

IPCC, 2013. Climate change 2013: the physical science basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 1535.
Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., 2014. EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg. Environ. Change 14 (2), 563–578.
Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C., Senior, C.A., 2014. Heavier summer downpours with climate change revealed by weather forecast resolution model. Nat. Clim. Change 4 (7), 570–576.
Kim, Y.-H., Min, S.-K., Zhang, X., Zwiers, F., Alexander, L.V., Donat, M.G., Tung, Y.-S., 2016. Attribution of extreme temperature changes during 1951-2010. Clim. Dyn. 46 (5–6), 1769–1782.
King, M.P., Hell, M., Keenlyside, N., 2016. Investigation of the atmospheric mechanisms related to the autumn sea ice and winter circulation link in the Northern Hemisphere. Clim. Dyn. 46, 1185.
Kolstad, E.W., Sobolowski, S.P., Scaife, A.A., 2015. Intraseasonal persistence of european surface temperatures. J. Clim. 28 (13), 5365–5374.
Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C., Kanae, S., Kowalczyk, E., Lawrence, D., 2004. Regions of strong coupling between soil moisture and precipitation. Science 305 (5687), 1138–1140.
Krueger, O., Hegerl, G.C., Tett, S.F., 2015. Evaluation of mechanisms of hot and cold days in climate models over Central Europe. Environ. Res. Lett. 10 (1), 014002.
Lee, J.Y., Fu, X., Wang, B., 2017. Predictability and prediction of the Madden-Julian oscillation: a review on progress and current status. In: The Global Monsoon System: Research and Forecast, pp. 147–159.
Lehmann, J., Coumou, D., 2015. The influence of mid-latitude storm tracks on hot, cold, dry and wet extremes. Sci. Rep. 5, 17491.
Lerch, S., Thorarinsdottir, T.L., Ravazzolo, F., Gneiting, T., 2015. Forecaster's Dilemma: Extreme Events and Forecast Evaluation arXiv preprint arXiv:1512.09244.
Matsueda, S., Takaya, Y., 2015. The global influence of the madden–Julian oscillation on extreme temperature events. J. Clim. 28, 4141–4151.
Min, S.-K., Zhang, X., Zwiers, F.W., Hegerl, G.C., 2011. Human contribution to more-intense precipitation extremes. Nature 470 (7334), 378–381.
Miralles, D.G., Teuling, A.J., Van Heerwaarden, C.C., de Arellano, J.V.-G., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7 (5), 345–349.
Mitchell, D., Davini, P., Harvey, B., Massey, N., Haustein, K., Woollings, T., Jones, R., Otto, F., Guillod, B., Sparrow, S., 2016. Assessing mid-latitude dynamics in extreme event attribution systems. Clim. Dyn. 1–13.
Mittermaier, M., Bullock, R., 2013. Using MODE to explore the spatial and temporal characteristics of cloud cover forecasts from high-resolution NWP models. Meteorol. Appl. 20 (2), 187–196.
Mohino, E., Keenlyside, N., Pohlmann, H., 2016. Decadal prediction of Sahel rainfall: where does the skill (or lack thereof) come from? Clim. Dyn. 1–20.
Morak, S., Hegerl, N., Christidis, 2013. Detectable changes in temperature extremes. J. Clim. 26, 1561–1574.
Mueller, B., Seneviratne, S.I., 2014. Systematic land climate and evapotranspiration biases in CMIP5 simulations. Geophys. Res. Lett. 41 (1–7) https://doi.org/10.1002/2013GL058055.
National Academies of Sciences, E., and Medicine, 2016. Attribution of Extreme Weather Events in the Context of Climate Change. National Academies Press.
Oppenheimer, M., Campos, M., Warren, R., Birkmann, J., Luber, G., O'Neill, B., Takahashi, K., 2014. Emergent risks and key vulnerabilities. In: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part a: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1039–1099.
Orlowsky, B., Seneviratne, S.I., 2013. Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections. Hydr. Earth Syst. Sci. 17, 1765–1781. https://doi.org/10.5194/hess-17-1765-2013.
Orsolini, Y., Senan, R., Balsamo, G., Doblas-Reyes, F., Vitart, F., Weisheimer, A., Carrasco, A., Benestad, R., 2013. Impact of snow initialization on sub-seasonal forecasts. Clim. Dyn. 41 (7–8), 1969–1982.
Palmer, T., 2014. Build high-resolution global climate models. Nature 515 (7527), 338.
Papagiannopoulou, C., Miralles, D.G., Depoorter, M., Verhoest, N., Dorigo, W., Waegeman, W., 2016. Discovering relationships in climate-vegetation dynamics using satellite data. In: Proceedings of AALTD 2016: Second ECML/PKDD International Workshop on Advanced Analytics and Learning on Temporal Data.
Pfahl, S., Wernli, H., 2012. Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. Geophys. Res. Lett. 39 (12).
Pfahl, S., O'Gorman, P.A., Fischer, E.M., 2017. Understanding the regional pattern of projected future changes in extreme precipitation. Nat. Clim. Change 7, 423–427.
Prein, A.F., Rasmussen, R.M., Ikeda, K., Liu, C., Clark, M.P., Holland, G.J., 2017. The future intensification of hourly precipitation extremes. Nat. Clim. Change 7, 48–52.
Prodhomme, C., Doblas-Reyes, F., Bellprat, O., Dutra, E., 2016. Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. Clim. Dyn. 47 (3–4), 919–935.
Quesada, B., Vautard, R., Yiou, P., Hirschi, M., Seneviratne, S.I., 2012. Asymmetric European summer heat predictability from wet and dry southern winters and springs. Nat. Clim. Change 2 (10), 736–741.
Sapiano, M.R.P., Arkin, P.A., 2009. An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. J. Hydrometeorol. 10 (1), 149–166.
Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J., Williams, A., 2014. Skillful long-range prediction of european and north american winters. Geophys. Res. Lett. 41, 2014GL059637.
Schaller, N., Kay, A.L., Lamb, R., Massey, N.R., Van Oldenborgh, G.J., Otto, F.E., Sparrow, S.N., Vautard, R., Yiou, P., Ashpole, I., 2016. Human influence on climate in the 2014 southern England winter floods and their impacts. Nat. Clim. Change 6 (6), 627–634.
Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. Earth-Sci. Rev. 99 (3), 125–161.
Seneviratne, S.I., Donat, M., Pitman, A.J., Knutti, R., Wilby, R.L., 2016. Allowable CO2 emissions based on regional and impact-related climate targets. Nature 529, 477–483. https://doi.org/10.1038/nature16542.
Seneviratne, S.I., Nicholls, N., Easterling, D., Goodess, C.M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., 2012. Changes in Climate Extremes and Their Impacts on the Natural Physical Environment, Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, pp. 109–230.
Seneviratne, S.I., Wilhelm, M., Stanelle, T., Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M.E., Meier, A., Brovkin, V., 2013. Impact of soil moisture - climate feedbacks on CMIP5 projections: first results from the GLACE-CMIP5 experiment. Geophys. Res. Lett. 40 (19), 5212–5217.
Sheen, K.L., Smith, D.M., Dunstone, N.J., Eade, Rosie, Rowell, D.P., Vellinga, M., 2017. Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescales. Nat. Commun. 8.
Shepherd, T.G., 2014. Atmospheric circulation as a source of uncertainty in climate change projections. Nat. Geosci. 7, 703–708. https://doi.org/10.1038/ngeo2253.
Sillmann, J., Kharin, V., Zhang, X., Zwiers, F., Bronaugh, D., 2013. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. J. Geophys. Res. Atmos. 118 (4), 1716–1733.
Sillmann, J., Donat, M.G., Fyfe, J.C., Zwiers, F.W., 2014. Observed and simulated temperature extremes during the recent warming hiatus. Environ. Res. Lett. 9, 064023 https://doi.org/10.1088/1748-9326/9/6/064023.
Sippel, S., Otto, F., Forkel, M., Allen, M., Guillod, B., Heimann, M., Reichstein, M., Seneviratne, S., Thonicke, K., Mahecha, M.D., 2016. A novel bias correction methodology for climate impact simulations. Earth Syst. Dyn. 7 (1), 71.
Smith, D.M., Eade, R., Dunstone, N.J., Fereday, D., Murphy, J.M., Pohlmann, H., Scaife, A.A., 2010. Skilful multi-year predictions of Atlantic hurricane frequency. Nat. Geosci. 3 (12), 846–849. https://doi.org/10.1038/ngeo1004.
Stegehuis, A., Vautard, R., Ciais, P., Teuling, R., Jung, M., Yiou, P., 2013. Summer temperatures in Europe and land heat fluxes in observation-based data and regional climate model simulations. Clim. Dyn. 41, 455–477.
Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. J. Geophys. Res. Atmos. 115 (24). D24211.
Stott, P.A., Christidis, N., Otto, F.E., Sun, Y., Vanderlinden, J.P., van Oldenborgh, G.J., Vautard, R., von Storch, H., Walton, P., Yiou, P., 2016. Attribution of extreme weather and climate-related events. Wiley Interdiscip. Rev. Clim. Change 7 (1), 23–41.
Taylor, C.M., de Jeu, R.A.M., Guichard, F., Harris, P.P., Dorigo, W.A., 2012. Afternoon rain more likely over drier soils. Nature 489, 423–426. https://doi.org/10.1038/nature11377.
Thiery, W., Davin, E.L., Seneviratne, S.I., Bedka, K., Lhermitte, S., van Lipzig, N.P., 2016. Hazardous thunderstorm intensification over Lake Victoria. Nat. Commun. 7.
Thorarinsdottir, T., Sillmann, J., Benestad, R., 2014. Studying statistical methodology in climate research, eos. Trans. Am. Geophys. Union 95 (15), 129–129.
Trenberth, K.E., Fasullo, J.T., Shepherd, T.G., 2015. Attribution of climate extreme events. Nat. Clim. Change 5 (8), 725–730.
van Oldenborgh, G.J., Balmaseda, M.A., Ferranti, L., Stockdale, T.N., Anderson, D.L., 2005. Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period. J. Clim. 18 (16), 3250–3269.
Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., Van Oldenborgh, G.J., Schaller, N., 2016. Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events. Environ. Res. Lett. 11 (11), 114009.
Vial, J., Osborn, T.J., 2012. Assessment of atmosphere-ocean general circulation model simulations of winter northern hemisphere atmospheric blocking. Clim. Dyn. 39 (1–2), 95–112.
Vila, D.A., de Goncalves, L.G.G., Toll, D.L., Rozante, J.R., 2009. Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. J. Hydrometeorol. 10 (2), 533–543.
Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A.W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D., Xiao, H., Zaripov, R., Zhang, L., 2017. The subseasonal to seasonal (S2S) prediction project database. Bull. Amer. Meteor. Soc. 98, 163–173.

Watson, P.A., Weisheimer, A., Knight, J.R., Palmer, T., 2016. The role of the tropical West Pacific in the extreme Northern Hemisphere winter of 2013/2014. J. Geophys. Res. Atmos. 121, 1698–1714.
Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. J. R. Soc. Interface 11.96 (2014), 20131162.
Weisheimer, A., Doblas-Reyes, F.J., Jung, T., Palmer, T., 2011. On the predictability of the extreme summer 2003 over Europe. Geophys. Res. Lett. 38 (5).
Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D.A., Palmer, T., 2017. Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. Q. J. R. Meteorol. Soc. 143, 917–926.
Westra, S., Fowler, H., Evans, J., Alexander, L., Berg, P., Johnson, F., Kendon, E., Lenderink, G., Roberts, N., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. Rev. Geophys. 52 (3), 522–555.
Whan, K., Zwiers, F.W., Sillmann, J., 2016. The influence of atmospheric blocking on extreme winter minimum temperatures in North America. J. Clim. 29, 4361–4381. https://doi.org/10.1175/JCLI-D-15-0493.1.
Yiou, P., Cattiaux, J., 2013. Contribution of atmospheric circulation to wet north European summer precipitation of 2012. Bull. Am. Meteorol. Soc. 94 (9), S39.
Yiou, P., Cattiaux, J., 2014. Contribution of atmospheric circulation to wet Southern European winter of 2013. Bull. Am. Meteorol. Soc. 95 (9), S66.
Yiou, P., Jézéquel, A., Naveau, P., Otto, F.E.L., Vautard, R., Vrac, M., 2017. A statistical framework for conditional extreme event attribution. Adv. Stat. Clim. Meteorol. Oceanogr. 3, 17–31.
Zhang, X., Wan, H., Zwiers, F.W., Hegerl, G.C., Min, S.K., 2013. Attributing intensification of precipitation extremes to human influence. Geophys. Res. Lett. 40 (19), 5252–5257.
Zhang, X., Hegerl, G., Seneviratne, S., Stewart, R., Zwiers, F., Alexander, L., 2014. WCRP Grand Challenge: Science Underpinning the Prediction and Attribution of Extreme Events available at: https://www.wcrp-climate.org/images/documents/grand_challenges/GC_Extremes_v2.pdf.