Advances in the subseasonal prediction of extreme events:

Relevant case studies across the globe

Daniela I. V. Domeisen

Christopher J. White

Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

Hilla Afargan-Gerstman

ETH Zurich, Institute for Atmospheric and Climate Science, Zurich, Switzerland

Ángel G. Muñoz

International Research Institute for Climate and Society (IRI), Columbia University’s Climate School, and The Earth Institute at Columbia University, New York, U.S.A

Matthew A. Janiga

Naval Research Laboratory, Monterey CA, U.S.A

Frédéric Vitart

European Centre for Medium-Range Weather Forecasts, Reading, UK

C. Ole Wulff

ETH Zurich, Institute for Atmospheric and Climate Science, Zurich, Switzerland / NORCE

Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

Early Online Release: This preliminary version has been accepted for publication in Bulletin of the American Meteorological Society, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-20-0221.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2022 American Meteorological Society
Salomé Antoine

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Constantin Ardilouze

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Lauriane Batté

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Hannah C. Bloomfield

Department of Meteorology, University of Reading, UK / School of Geographical Sciences, University of Bristol, UK

David J. Brayshaw

Department of Meteorology, University of Reading, UK

Suzana J. Camargo

Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, U.S.A

Andrew Charlton-Pérez

Department of Meteorology, University of Reading, UK

Dan Collins

Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, U.S.A

Tim Cowan
Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, Australia / Bureau of Meteorology, Melbourne, Australia

Maria del Mar Chaves

Climate Simulations and Predictions, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy / now at: University of Bologna, Bologna, Italy

Laura Ferranti

European Centre for Medium-Range Weather Forecasts, Reading, UK

Rosario Gómez

Organismo Internacional Regional de Sanidad Agropecuaria, San Salvador, El Salvador

Paula L.M. González

NCAS/Department of Meteorology, University of Reading, UK / International Research Institute for Climate and Society, The Earth Institute, Columbia University, U.S.A

Carmen González Romero

International Research Institute for Climate and Society (IRI). Columbia University’s Climate School, and The Earth Institute at Columbia University. New York, U.S.A

Johnna M. Infanti

Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, U.S.A

Stelios Karozis

National Centre for Scientific Research "Demokritos", Greece
Hera Kim

*School of Earth and Environmental Sciences, Seoul National University, South Korea*

Erik W. Kolstad

*NORCE Norwegian Research Center, Bjerknes Center for Climate Research, Bergen, Norway*

Emerson LaJoie

*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, U.S.A*

Llorenç Lledó

*Barcelona Supercomputing Center (BSC), Barcelona, Spain*

Linus Magnusson

*European Centre for Medium-Range Weather Forecasts, Reading, UK*

Piero Malguzzi

*CNR-ISAC, Bologna, Italy*

Andrea Manrique-Suñén

*Barcelona Supercomputing Center (BSC), Barcelona, Spain*

Daniele Mastrangelo

*CNR-ISAC, Bologna, Italy*

Stefano Materia

*Climate Simulations and Predictions, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy*
Hanoi Medina

Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, U.S.A

Lluís Palma

Barcelona Supercomputing Center (BSC), Barcelona, Spain

Luis E. Pineda

Yachay Tech University, School of Earth Sciences, Energy and Environment, Hda. San José s/n y Proyecto Yachay, Urcuquí, Ecuador

Athanasios Sfetsos

National Centre for Scientific Research "Demokritos", Greece

Seok-Woo Son

School of Earth and Environmental Sciences, Seoul National University, South Korea

Albert Soret

Barcelona Supercomputing Center (BSC), Barcelona, Spain

Sarah Strazzo

Embry Riddle Aeronautical University, Daytona Beach, FL, U.S.A

Di Tian

Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, U.S.A

*Corresponding author: Daniela I.V. Domeisen, daniela.domeisen@env.ethz.ch
ABSTRACT

Extreme weather events have devastating impacts on human health, economic activities, ecosystems, and infrastructure. It is therefore crucial to anticipate extremes and their impacts to allow for preparedness and emergency measures. There is indeed potential for probabilistic subseasonal prediction on timescales of several weeks for many extreme events. Here we provide an overview of subseasonal predictability for case studies of some of the most prominent extreme events across the globe using the ECMWF S2S prediction system: heatwaves, cold spells, heavy precipitation events, and tropical and extratropical cyclones. The considered heatwaves exhibit predictability on timescales of 3-4 weeks, while this timescale is 2-3 weeks for cold spells. Precipitation extremes are the least predictable among the considered case studies. Tropical cyclones, on the other hand, can exhibit probabilistic predictability on timescales of up to 3 weeks, which in the presented cases was aided by remote precursors such as the Madden-Julian Oscillation. For extratropical cyclones, lead times are found to be shorter. These case studies clearly illustrate the potential for event-dependent advance warnings for a wide range of extreme events. The subseasonal predictability of extreme events demonstrated here allows for an extension of warning horizons, provides advance information to impact modelers, and informs communities and stakeholders affected by the impacts of extreme weather events.
Capsule summary. An assessment and comparison of the subseasonal predictability of case studies of the most prominent extreme weather events on a global scale: heatwaves, cold spells, precipitation extremes, and cyclones.

1. Subseasonal prediction of extreme events

Extreme weather events pose threats to humans, infrastructure, and ecosystems. In a changing climate, many extremes are projected to increase in strength, frequency, and/or duration, and it is therefore increasingly important to anticipate extreme events and their impacts as early as possible. A successful prediction several weeks in advance will benefit stakeholders’ decision making for emergency management (White et al. 2017; Merz et al. 2020; White et al. 2021). Indeed, there is increasing potential for probabilistic subseasonal prediction on timescales of several weeks for extreme events (Vitart 2014; Vitart and Robertson 2018; Robertson et al. 2020). Increased predictability can arise from remote drivers or long-lived precursor patterns that are conducive to the occurrence of extreme events. These drivers include tropical precursors such as the Madden-Julian Oscillation (MJO) (e.g. Vitart and Molteni 2010; Rodney et al. 2013) and El Niño Southern Oscillation (ENSO) (e.g. Domeisen et al. 2015), surface interactions with snow cover (e.g. Cohen and Jones 2011) or sea ice (e.g. Sun et al. 2015), the upper atmosphere (e.g. Domeisen et al. 2020b; Domeisen and Butler 2020), or a combination of predictors (Muñoz et al. 2015, 2016; Doss-Gollin et al. 2018; Dobrynin et al. 2018). A better understanding of these precursors can contribute to increased predictability. At the same time, improvements in the prediction of extremes arises from progress in the performance of prediction systems through advancements in process representation, coupling, and parameterization, as well as model resolution (Bauer et al. 2015). Merryfield et al. (2020) recommended an assessment of the predictability of historical high-impact weather events...
as a way forward to demonstrate the potential benefits of subseasonal to seasonal (S2S) forecasts. Here we discuss extreme event predictability based on a state-of-the-art subseasonal prediction system and a range of precursors for selected case studies of high-impact extremes in Europe, Africa, Asia, Australia, as well as South, Central, and North America for the most prominent extreme events on a global scale: heatwaves, cold spells, heavy precipitation events, and both tropical and extratropical cyclones. The following sections provide a brief overview of the physical drivers and potential for predictability for these extreme events, while the subsequent sections dive into the specific case studies.

a. Heatwaves

Heatwaves over land have devastating impacts on human health and ecosystems (Campbell et al. 2018; Yang et al. 2019), agriculture (Brás et al. 2021), and energy demand (Auffhammer et al. 2017; Bloomfield et al. 2020). Over the past decades, heatwaves have significantly increased in frequency and intensity (Perkins et al. 2012) with further increases predicted for the future (Watanabe et al. 2013; Lopez et al. 2018), largely due to anthropogenic global warming (Stocker 2014; Shiogama et al. 2014). Heatwaves are commonly characterized by temperature and duration thresholds (Russo et al. 2014), in addition to humidity and diurnal temperature cycle characteristics for applications to human morbidity and mortality (e.g. Raymond et al. 2020).

Heatwaves are often associated with persistent anticyclonic circulation patterns (Li et al. 2015; Freychet et al. 2017) that can sometimes be identified as blocking (Pfahl and Wernli 2012; Schaller et al. 2018; Brunner et al. 2018; Carrera et al. 2004; Dong et al. 2018; Li et al. 2019; Yeo et al. 2019), long-lived Rossby Wave Packets (RWPs, Wirth et al. (2018)), which can contribute to predictability (Fragkoulidis et al. 2018; Grazzini and Vitart 2015), or quasi-stationary wave trains (Enomoto 2004; Kim et al. 2018; Li et al. 2019). These patterns can be triggered or enhanced
by remote effects. For instance, sea surface temperature (SST) anomalies in subtropical and extratropical ocean basins can help induce European and North American heatwaves (Wulff et al. 2017; Duchez et al. 2016; McKinnon et al. 2016; Hartmann 2015), and East Asian heatwaves can be triggered by the North Atlantic Oscillation (NAO), Ural blocking, and diabatic heating in the eastern Mediterranean (Yasui and Watanabe 2010; Jian-Qi 2012; Wu et al. 2016; Gao et al. 2018; Li et al. 2019).

These remote forcings can enhance the predictability of heatwaves. Recent research has indeed shown potential for the extended-range prediction of heatwaves on sub-seasonal to seasonal timescales (Kueh and Lin 2020; Koster et al. 2010; Luo and Zhang 2012; Pepler et al. 2015; Tian et al. 2017; Wulff and Domeisen 2019). In addition, heatwaves can also be exacerbated by land-atmosphere feedbacks (e.g. Fischer et al. 2007; Mueller and Seneviratne 2012; Miralles et al. 2014; Hauser et al. 2016; Seneviratne et al. 2010; Berg and Sheffield 2018; Tian et al. 2016, 2018) and improvements in soil moisture initialization can therefore increase the predictability of heatwaves (Ferranti and Viterbo 2006; Dirmeyer et al. 2018; Bunzel et al. 2018).

b. Cold spells

Cold spells can affect electricity production (Beerli et al. 2017; Gruber et al. 2021; Doss-Gollin et al. 2021) and demand (Cradden and McDermott 2018; Bloomfield et al. 2018, 2020), human mortality (Charlton-Perez et al. 2019, 2021), and agriculture (Materia et al. 2020a). Similar to heatwaves, cold spells are often defined by temperature and duration thresholds (de Vries et al. 2012). Like heatwaves, cold spells can be related to atmospheric blocking and hence model biases in blocking frequency can impair predictions at lead times beyond two weeks (Hamill and Kiladis 2014; Quinting and Vitart 2019). Predictability can be gained from tropical drivers such as the MJO, and model performance can be enhanced by capturing the predictable signal of
large-scale weather patterns such as the NAO at the extended range (Ferranti et al. 2018). Blocking associated with the negative phase of the NAO can also be induced through sudden stratospheric warming (SSW) events (Thompson et al. 2002; Lehtonen and Karpechko 2016; Charlton-Perez et al. 2018; Domeisen 2019), which can induce cold spells both over land (Kolstad et al. 2010) and ocean (Afargan-Gerstman et al. 2020). However, not all regions gain predictability skill from stratospheric forcing (Domeisen et al. 2020b; Materia et al. 2020a).

c. Precipitation events

Heavy precipitation events can lead to flooding as well as land- or mudslides, and they are often accompanied by strong winds and low temperatures, the combination of which can be detrimental to humans, agriculture and infrastructure (Zscheischler et al. 2020). Heavy precipitation events are projected to become more frequent in many regions (Donat et al. 2016; Prein et al. 2017) due to anthropogenic climate change (Westra et al. 2013; Zhang et al. 2013; Li, Chao et al. 2021). Similar to temperature extremes, rainfall extremes arise through persistent atmospheric conditions, which can be triggered or maintained by large-scale forcing (e.g. from ENSO and the MJO (Jones et al. 2004; Kenyon and Hegerl 2010; Muñoz et al. 2015)), atmospheric blocking (Lenggenhager and Martius 2019), or monsoon systems (Zhang and Zhou 2019).

Precipitation extremes tend to be less predictable than temperature extremes such as warm and cold spells (de Andrade et al. 2019). The ability of a prediction system to predict rainfall extremes beyond deterministic timescales is related to the simulation of the connection between precipitation and its large-scale forcing such as ENSO and the MJO (Vigaud et al. 2017; Specq et al. 2020) or atmospheric rivers (DeFlorio et al. 2019). Regions with strong ENSO teleconnections exhibit better predictability of rainfall extremes, as for example, in Australia (King et al. 2020) or the
southwestern U.S. (Gershunov 1998; Pan et al. 2019), if ENSO is correctly simulated (Bayr et al. 2019). Interference of drivers on multiple timescales can further modulate the intensity, occurrence and predictability of precipitation extremes (Muñoz et al. 2015, 2016).

d. Tropical Cyclones and Medicanes

Tropical and extratropical cyclones impact human lives and livelihoods and lead to large environmental impacts and economic losses (Camargo and Hsiang 2015; Hsiang 2010; Hsiang and Narita 2012). Anthropogenic climate change affects various properties of tropical cyclones (TC), in particular their intensity, as well as the precipitation and storm surge associated with these events (Knutson et al. 2019, 2020). While individual cyclones’ genesis, tracks and intensity are not predictable beyond deterministic timescales, large-scale drivers can provide predictability in a probabilistic sense on S2S timescales. On seasonal timescales, ENSO modifies the characteristics of TC frequency, intensity and tracks (e.g., Vitart et al. 2003; Lin et al. 2017; Nicholls 1979; Evans and Allan 1992). On subseasonal timescales, TC activity is enhanced (decreased) during and after an active (suppressed) MJO (e.g. Camargo et al. 2019), especially in the southern hemisphere (e.g. Hall et al. 2001; Camargo et al. 2009), allowing for successful statistical forecasts (Leroy and Wheeler 2008). Recently, the performance of dynamical models for forecasting TCs on subseasonal timescales has significantly improved (Camp et al. 2018; Camargo et al. 2019; Robertson et al. 2020; Vitart et al. 2010; Camargo et al. 2021). A successful example is cyclone Hilda, which made landfall in northwestern Australia and was predicted 3 weeks in advance (Gregory et al. 2019). However, this success is not consistent across models, and is likely linked to a successful prediction of the MJO (Vitart 2017; Lee et al. 2018, 2020).

In addition to tropical cyclones, we also consider medicanes (“Mediterranean Hurricanes”), rare intense and high-impact extratropical cyclones in the Mediterranean region (Ulbrich et al. 2009;
Cavicchia et al. 2014; Mylonas et al. 2018; Flaounas et al. 2021). These events occur on average 1.6 times / year (Flaounas et al. 2015) and can lead to severe damage in coastal areas associated with flooding and high winds.

2. Data and Methods

To evaluate the subseasonal prediction of the above extreme events we use both forecasts and hindcasts (historical forecasts) from the extended-range operational ensemble prediction system (Vitart et al. 2008) from the European Centre for Medium-Range Weather Forecasts (ECMWF), which is part of the S2S database (Vitart et al. 2017). The prediction system includes coupling with the ocean and sea ice (Buizza et al. 2017). The atmospheric model has a horizontal resolution of approximately 36 km and 91 vertical levels with a model lid at 0.01 hPa (at the time of data download for this study). Where available, that is, for case studies after June 2015, forecasts from the prevailing model version were used (cycles 43R1, 43R3 and 45R1); these ensemble forecasts consist of 51 members. For the case studies using hindcasts, the 11-member hindcast ensemble from model cycle 46R1 was used. Both forecasts and hindcasts are initialized twice weekly.

The target weeks are selected for each case study individually based on the week of the most extreme anomalies. Since the forecasts are only initialized twice weekly, it is not always possible to find a forecast that is initialized exactly the day before week 1. Week-1 lead time for a specific case study is therefore chosen such that the target week lies directly on or after the initialization, that is, the forecast is initialized either on the first day of week 1 or up to two days earlier. The additional forecast lead weeks (weeks 2 - 4) then lie exactly adjacent to week 1.

To compute anomalies for the subseasonal predictions, a 7-day mean climatology is computed based on the 11-member ensemble hindcasts initialized for the same lead time for the corresponding available 20-year hindcast period. For example, for the California heatwave on 23 July 2018, the
corresponding week-1 climatology is based on the ensemble mean of the hindcast ensemble initialized on 23 July for each year from 1998 to 2017. The climatology is computed for each lead week separately, yielding a lead-time dependent climatology. Anomalies for the predictions are then computed by subtracting the model climatology from each ensemble member. For the earlier case studies, the climatology is computed over a 19-year hindcast period excluding the year of the case study to simulate an operational prediction setting. Anomalies for reanalysis are computed in a consistent way, by subtracting the daily mean climatology computed from reanalysis data for the same years that are used for computing the hindcast climatology for each case study. The use of anomalies for the model and reanalysis with respect to their respective climatologies provides a simple bias correction.

The temperature predictions are verified against the 2m temperatures from ERA5 reanalysis (Hersbach et al. 2020), as temperatures are well represented in reanalysis. Precipitation can show greater biases in reanalysis (Alexander et al. 2020), hence precipitation is verified against observational datasets from the Australian Water Availability Project (AWAP) 5 km daily gridded rainfall analysis (Jones et al. 2009) and the CPC Global Unified Gauge-Based Analysis of Daily Precipitation (Chen et al. 2008).

The temperature extremes case studies compare the probability density functions (PDFs) of the ensemble members for different lead weeks. Tercile limits (below-normal, normal, and above-normal, as well as the 10th and the 90th percentiles) are computed with respect to the lead time-dependent model climatology, based on 11 hindcast members. For the rainfall extremes, forecast performance is assessed by measuring the forecast system’s association and discrimination attributes, using the Spearman correlation coefficient (Wilks 2019) and the area under the Relative Operating Characteristic (ROC, Wilks 2019) curve for the above-normal category, respectively. The Spearman correlation is a non-parametric measure of how in-phase the forecasts
and observations are (correlation values of 1 indicate perfect association), and the ROC area for the above-normal category measures how well the forecast system discriminates between the above-normal and the other tercile-based categories, with values at 50% indicating a discrimination as good as that of climatology-based forecasts, and values above (below) 50% indicating better (worse) discrimination than climatology-based forecasts. The precipitation forecasts are calibrated according to a pattern-based Model Output Statistics approach using canonical correlation analysis (CCA; Tippett et al. (2008)), implemented via PyCPT, a set of Python libraries interfacing the Climate Predictability Tool (Muñoz 2020; Muñoz and Coauthors 2019; Mason et al. 2021), using IRI’s “NextGen” forecast approach (Muñoz and Coauthors 2019; WMO 2020). To obtain a robust sample size, these metrics were computed using all 8 initializations (20 years per initialization) available for the months and target dates listed in Table 1, conducted independently for each rainfall extreme case study. For example, for the Guatemala case study (see next section), eight 20-year-long hindcasts were used, corresponding to all initializations available for June 1998-2017, providing a total of 160 hindcast weeks to compare against the corresponding 160 weeks of observed rainfall. For additional details see Materia et al. (2020a).

For evaluating the model performance for the cyclones, their observed tracks are compared against the probability of cyclone occurrence given by the probability of a cyclone passing within 300 km of each grid point using the ECMWF tracker (Vitart et al. 1997) from the 51-member ensemble of the prediction system. The observed tropical cyclones data are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010). The observed track for the medicane is obtained from the ECMWF operational analysis. The medicane is further evaluated using Convective Available Potential Energy (CAPE), an indicator of atmospheric instability, which is a necessary condition for the development of severe weather events. CAPE has been found to
be a prominent indicator and potential predictor for tropical cyclones (Huang and Liang 2010; Lee and Frisius 2018; Mylonas et al. 2018) but has not been prominently used for medicanes.

3. Extreme event case studies

This section presents specific case studies for the four types of extremes. The case studies were selected based on their extreme nature and societal impacts. While this selection should not be seen as a complete assessment of model performance or inter-comparison of predictability between event types or within the same event type, these case studies serve as a representative selection of extreme events and their predictability, which can translate into timescales of emergency preparedness (White et al. 2021). Table 1 provides an overview of the timing and location of each case study.

a. Heatwaves

We first examine the predictability of four extreme heatwaves in North America, Europe, and East Asia between 2013 and 2019 (Fig. 1). The first two heatwaves are part of the extreme Northern Hemisphere heatwave in summer 2018, when heatwaves simultaneously affected North America and Eurasia. We focus on the week of July 23-29, 2018, when temperatures over California reached 51°C in Death Valley. California monthly mean temperatures for July surpassed the previous record set in 1931 (NOAA 2018) as heatwaves also occurred earlier that month. Similarly in Europe, the seasonal mean was strongly affected as the heat arrived in two waves, one from mid-May to mid-June and the second from mid-July to the beginning of August.

The model successfully predicts the concurrent 2018 heatwaves for the target period 3 weeks ahead in terms of the spatial structure of the anomalies for both considered regions, although with reduced amplitudes, meaning that most ensemble members remain well below the observed
anomalies (Fig. 1a-d). For Europe, at lead times of 2 weeks, 49 out of 50 ensemble members exceed the upper third of the climatological distribution (Fig. 2b). The forecast probability for the upper tercile is still 86% at lead times of 3 weeks and reduces to 60% for lead week 4, but with a long tail of the distribution towards extreme heat. For California, the model also predicts the extreme heat with some confidence out to 4 weeks (Fig. 2a). The 2-week lead forecast yields the most confident prediction, with 29% of ensemble members predicting temperatures above the 90th percentile, and 78% predicting above normal temperatures. Interestingly, although the 3-week lead forecast distribution is still shifted towards above normal temperatures, it is arguably the weakest prediction, with only 12% of members predicting temperatures above the 90th percentile, as compared to 24% for week 4.

Generally, California / western U.S. heat waves tend to be associated with high pressure over the Great Plains, low pressure off the California coast, and warm moist air transport from the south. There has been an increasing trend in this type of humid heatwave in recent years due to warming ocean temperatures (Gershunov and Guirguis 2015). When present, this ocean-atmosphere pattern can lead to higher predictability of heat waves, although forecast accuracy over the western U.S. and California is on average lower relative to other U.S. regions (Gershunov and Guirguis 2012; Ford et al. 2018; Kornhuber et al. 2019). However, July 2018 was atypical in that it was characterized by a wave-7 pattern (Kornhuber et al. 2019) associated with a strong and persistent region of high temperatures over much of the U.S. in the first half of July, and high pressure anomalies off the coast of and over the western U.S. in the last two weeks of July. Land - atmosphere and vegetation feedbacks are further suggested to have played a role in the 2018 heatwave, especially over central Europe (Liu et al. 2020; Sinclair et al. 2019; Albergel et al. 2019). Finally, the event was made more likely due to anthropogenic climate change (Yiou et al. 2019).
Less than a year after the devastating 2018 heatwave, another series of heatwaves affected the United States in 2019. In late May 2019 (we here consider the week of May 24 - 30), an early season heatwave affected the southeastern U.S., tied to a wavy jet stream pattern with anomalously high (low) pressure over the southeastern (southwestern) U.S. (Liberto 2019). The model captures the temperature anomalies at 3-week lead time, but it notably underestimates the extreme temperature anomalies (Fig. 1e,f), which is also found in the NCEP CFSv2 model (Luo and Zhang 2012). This underestimation is evident in the ensemble spread (Fig. 2c).

A further devastating heatwave was observed in East Asia in August 2013. The heatwave persisted for over two weeks from late July to mid-August, resulting in severe socio-economic losses in the region (Duan et al. 2013; Sun et al. 2014; Li et al. 2019). South Korea experienced the hottest summer nights and the second hottest summer days since 1954 (Min et al. 2014). In western Japan, daily maximum temperature records were broken or tied at 143 weather stations (JMA 2013), many of which were broken again during the 2018 heatwave. The extreme persistence and severity of the event resulted from the combination of a westward extension of the North Pacific subtropical high (Jing-Bei 2014; Li et al. 2015) and a zonal wave train (Yeo et al. 2019) resembling the circumglobal teleconnection (Ding and Wang 2005).

For the considered target week of 5-11 August 2013, a warm anomaly of over 4°C was observed in the large metropolitan areas of eastern China, while the heatwave extended to the Korean peninsula and Japan (Fig. 1g). The temperature anomaly was larger in the urban areas than in rural areas (Wang et al. 2017), possibly due to the urban heat island effect. The temperature distribution is well captured by the model over land at a 3-week lead time, though the magnitude is slightly underestimated, while the warm anomaly over the eastern China Sea is not reproduced (Fig. 1h). When initialized four weeks before the target period on July 15, more than a third of the ensemble members point to below normal temperatures, although twenty percent already
predict temperatures above the 90th percentile (Fig. 2d). However, starting at the 3-week lead time, essentially all ensemble members predict above normal temperatures, and only one ensemble member at 2-week lead time predicts temperatures below the 90th percentile. More importantly, the ensemble-mean of these initializations quantitatively well captures the observations (i.e., individual ensemble members are well centered about the observed value). This result indicates that the 2013 East Asia heatwave is quantitatively well predicted by the model at a maximum lead time of three weeks.

b. Cold spells

Several examples of extreme cold spells in Europe are studied in this section. We start with a cold spell in eastern and southeastern Europe in late winter and early spring of 2003 (Levinson and Waple 2004) that preceded a record-breaking summer heatwave. The month of February was the coldest on record in Albania and Macedonia, and temperatures in southeastern Europe were between -2°C and -5°C below normal for much of February and early March (Dittmann et al. 2004). The target week of April 3-9 (Fig. 3a) marked the end of this cold period, but was cold enough that the month of April registered record minimum temperatures in the Baltic region, the Danube watershed, and part of Italy and the Balkans (Dittmann et al. 2004). The extreme cold was associated with atmospheric blocking over the UK leading to southward advection of cold air masses from the Arctic, reaching southeastern Europe on April 7. The temperature contrasts between the frigid air mass and the southern Adriatic Sea caused strong convective precipitation, with heavy snowfall along the coasts of western Greece, Albania and southern Italy.

The model predicts the cold anomaly in central Europe (Fig. 3b), though with a southeastward shift and smaller anomalies than observed. The ensemble starts encompassing the observed anomaly at the 3-week lead time (March 19 initialization, Fig. 4a), indicating a 51% probability of
temperatures in the lower tercile for the target week, and a 29% chance of temperatures below the
tenth percentile. At the 2-week lead time, the confidence about the occurrence of cold weather is
clearly increased, with 72% of the ensemble members indicating temperatures below normal, and
53% below the 10-percentile threshold.

Another cold spell preceding a hot summer occurred in late February / early March 2018 in
central and western Europe after an otherwise mild winter. The cold wave was likely linked to a
major SSW event in mid-February 2018, which enhanced the probability of the negative NAO and
Greenland blocking during the peak of the cold event (Kautz et al. 2020). The SSW itself was
anticipated 10 days ahead (Karpechko et al. 2018) – a typical predictability timescale for SSWs
(Domeisen et al. 2020a). Knight et al. (2021) identified the extreme MJO event of January 2018
as an important driver of this SSW.

The blocking associated with this cold spell shows predictability in the ECMWF system (Ferranti
et al. 2019). The forecast initialized on February 12, 2018, the day of the SSW event (a lead time of
around 3 weeks), captures the cold anomaly over central Europe and part of the British Isles, but the
anomaly is significantly underestimated (Fig. 3c,d). Already at 4 weeks lead time (initialization on
February 5) the most likely category is the below normal tercile (with 54% of ensemble members)
for temperature over western Europe (Fig. 4b). Further analysis using North Atlantic weather
regimes suggests that the sequence of weather regimes before and during the cold spell (positive
NAO, blocking, followed by negative NAO, as documented in Kautz et al. (2020)) were correctly
anticipated by the model from the February 12 start date (not shown).

Another cold spell linked to atmospheric blocking occurred in winter 2016/2017 (Fig. 3e). The
block over Europe brought warm air to Scandinavia and Arctic air to eastern–central Europe in
the second week of January (Magnusson 2017). A cut-off low developed, causing exceptionally
low temperatures in the Balkan Peninsula as well as snowfall in Greece and southern Italy with
significant socioeconomic impacts due to the long duration of the event (Anagnostopoulou et al. 2017). The following week (16-22 January 2017), central Europe was affected by further cold air advection due to a tripole in surface pressure, with high pressure from the UK towards the Black Sea, and low pressure in the western Mediterranean and to the north of Scandinavia. This tripole was consistent with quiescent, cold and dry conditions over central Europe in the region of the anticyclone (Fig. 3e).

The forecast issued on January 2 (3-week lead time) already indicates an enhanced probability of below normal temperatures (Fig. 3f). Four weeks before the event, the probability for temperatures in the lower tercile already reaches 45% and increases to 63% (89%) at 3 (2) weeks before the event (Fig. 4c). The ensemble clearly narrows towards the observed anomaly at shorter lead times. The probability of temperature anomalies below the 10th percentile increases closer to the event, from 18% (4 weeks before), to 29% (3 weeks before), and finally to 64% 2 weeks before the event.

The cold spell produced a peak in electricity demand, particularly in France, where most of the heating is powered by electricity. The concomitant low wind speeds led to a lower than normal wind power generation, and several nuclear power plants in France were under maintenance (RTE 2017). This combination caused a high-risk situation for France’s energy system that could have been better managed given the forecasts, for example through a postponement of the planned maintenance operations in the nuclear power plants.

Another extreme cold spell occurred in late 2010. From late November to early December 2010, Germany and France recorded the coldest December in 40 years, while in the United Kingdom this was the coldest December in 100 years (Fig. 3g). December 2010 was characterized by an unusually strong negative NAO (Maidens et al. 2013) with strong cold air advection from northern Europe and Siberia (Prior and Kendon 2011). The cold anomaly over land was accompanied by a marine cold air outbreak (MCAO, according to the MCAO index used in Afargan-Gerstman
et al. (2020)) in the Norwegian and the Barents Seas. MCAOs can have devastating impacts on marine infrastructure and offshore activities, for example by creating favorable conditions for the formation of polar lows (Rasmussen 1983; Kolstad et al. 2009; Noer et al. 2011; Landgren et al. 2019). Indeed, a polar low was detected in satellite imagery in the Norwegian Sea off the coast of Norway on the 25th of November 2010, two days before our selected target date, based on the STARS database of polar lows (http://polarlow.met.no/), but no records regarding damages from this polar low have been found. Although the occurrence of cold air outbreaks in the North Atlantic and over northern Europe is often associated with stratospheric weak polar vortex events (e.g., Kolstad et al. 2010; Afargan-Gerstman et al. 2020), this event is unlikely to have been driven by the stratosphere, possibly reducing its predictability.

Cold anomalies had been predicted for northern Europe 3 weeks earlier by the hindcast initialized on November 11, however the prediction clearly underestimates the magnitude of the observed event (Fig. 3g,h). Hindcasts for lead times beyond 3 weeks (initialization on Nov 4) already provide an indication of the cold anomaly, with probabilities around 20% for temperatures below the 10th percentile. Hindcasts initialized at lead times of 2 and 3 weeks capture the below normal temperatures with a probability of above 90% and 50%, respectively (Fig. 4d). Hence, although the probability of a cold extreme is significantly increased already 3 weeks before the event, the magnitude of the extreme event is only captured at 2-weeks lead time.

c. Precipitation events

In this section we focus on four events with anomalous precipitation in Central and South America, Europe, and Australia. The first considered event is analyzed in the context of a volcanic eruption, as an example of using subseasonal forecasts for compound events, where the possibility of heavy rainfall was of concern. Guatemala’s Volcán de Fuego, a stratovolcano, erupted on June
3rd 2018, killing at least 113 people, while more than 300 remained unaccounted for (Program 2018). Ash plumes and pyroclastic flow material affected communities up to 25 km away from the volcano. The pyroclastic flows produced lahars (i.e., mudflow or debris flow) intermittently for several weeks, leading to evacuations of the nearby communities and displacing thousands of Guatemalans, destroying infrastructure and damaging crops. Overall, the eruption impacted 1.2 million Guatemalans, and cost more than U.S. D$219 millions (CEPAL 2018; CONRED 2018; WorldBank 2018).

The impacts could have been worse if precipitation, which typically peaks in the region in June, had been higher. Intense or persistent rainfall events (a) tend to make lahar viscosity thinner, which sustains the flow of pyroclastic debris for a longer duration, potentially causing more damage; (b) can remobilize unconsolidated pyroclastic deposits, causing post-eruption lahars; (c) can displace hanging slabs of solidified mud, debris and boulders down steep slopes, with the potential to destroy infrastructure and kill people; and (d) tend to interfere with evacuation, search and rescue, cleaning, and rebuilding operations. Due to the activities deployed at the time in Guatemala by the Columbia University World Project “Adapting Agriculture to Climate Today, for Tomorrow” (IRI 2018), the International Research Institute for Climate and Society and INSIVUMEH – the Guatemalan national meteorological agency – started working together immediately after the eruption to provide calibrated subseasonal rainfall forecasts from the prediction system to the National Government and a wide variety of local institutions.

Calibrated rainfall NextGen forecasts (Muñoz and Coauthors 2019) initialized on June 4 indicated low chances of exceeding the weekly median for the following four weeks for most of Guatemala (compare to Fig. 5a,b; Fig. 6a,b), and further analysis for the location of interest helped INSIVUMEH advise government institutions on evacuation, search and rescue, and cleaning and rebuilding operations. Subsequent weekly forecast updates confirmed the original expected out-
comes. These results build evidence on the advantages of using real-time subseasonal rainfall forecasts to help decision makers during and after volcanic eruptions, and potentially other seismologic and compound environmental events. Using a combination of forecasts at multiple timescales is suggested to be an optimal practice in these cases, consistent with the “Ready-Set-Go” approach (Goddard et al. 2014).

Another event of interest occurred in January 2016, when a series of heavy precipitation events affected Northwestern South America, leading to widespread flooding in coastal northern Ecuador, especially in the Province of Esmeraldas. The flood displaced 120 families, left one casualty, and was the largest such event in 20 years (Davies 2016). The flooding was associated with an early onset of the heavy rainfalls and severe mesoscale convective systems (MCSs) that would normally not be expected until annual precipitation peaks in April / May (Mohr and Zipser 1996; Bendix et al. 2009). On January 25, convective storms developed into a MCS with an extent of around 250 km over the western Andes foothills of the Esmeraldas river basin, a region of abundant low-level moisture bounded by the Andes. This heavy precipitation event was favored by interactions between the very strong El Niño event and an unusually persistent MJO in phases 2 and 3 (Pineda et al. 2021).

Weekly ensemble-mean rainfall anomaly hindcasts represent the spatial pattern of the anomalous precipitation extreme over the catchment over all lead times (Fig. 5c,d), with the best event identification for week 3 initialized on 28th Dec 2015 (i.e., the week 3 anomaly was closer to the observations as compared to week 2 (not shown)). For the Esmeraldas river basin the ROC scores for week 3 range from 0.5 to 0.6 (Fig. 6c), indicating low to modest discrimination of the above-normal rainfall on January 25th. The Spearman-rank correlations range from -0.25 to 0.25 (Fig. 6d); thus, based on the hindcast, the model performance is limited for the region where the extreme rainfall occurred at a lead time of 3 weeks. However, the positive precipitation anomaly
of more than one standard deviation averaged over the grid points closest to the catchment was captured for all lead times of 1-3 weeks (Pineda et al. 2021). Therefore, the use of the S2S rainfall forecast could have provided decision-makers with useful information about the onset of this extreme precipitation event. A timely uptake of the available forecasts 2-3 weeks in advance by the National Met-Hydro Service could have allowed for an early warning for this catastrophic flood event.

Another heavy precipitation event affected northwestern Italy (Piedmont and Liguria) in the period from 21 - 25 November 2016. Over these 5 days, more than 50% of annual precipitation was recorded in several areas, with peaks above 600 mm (ARPA Liguria 2017; ARPA Piemonte 2017). Severe damage was caused by river floods with flow-rate return times up to 200 years, and widespread occurrence of shallow landslides (Cremonini and Tiranti 2018). This episode developed in the middle of a persistent drought affecting most of central and western Europe in 2016/2017 (García-Herrera et al. 2019). The precipitation anomaly is underestimated by the model and exhibits a misplaced maximum for the forecast initialized on 7 November 2016 for week 3 (lead times 15–21 days, Fig. 5e,f). However, the positive anomaly over northwestern Italy is reproduced more than 2 weeks in advance. Positive anomalies were also correctly located in the Western Mediterranean region. These anomalies are significantly different at the 10% level from the ensemble climatology according to a Wilcoxon–Mann–Whitney test (not shown).

The large-scale mid-tropospheric configuration leading to this precipitation event was characterized by a persistent low pressure anomaly over the Iberian Peninsula, surrounded by areas of high pressure extending from the North Atlantic to Eastern Europe (ARPA Piemonte 2017). This dipole in pressure anomalies favors meridional moist advection across the complex orography downstream, leading to heavy precipitation in the Mediterranean in this season (e.g., Buzzi et al. 2014; Khodayar et al. 2018). The anomalous persistence of the large-scale pattern likely favored
the predictability of the event (Vitart et al. 2019). Although the verification scores of the week-3 forecasts for this area (Fig. 6e,f) indicate, on average, a relatively low predictive performance, the sufficiently correct representation of the atmospheric dipole in the extended range may have enhanced the predictability of precipitation for this event. Similarities are found with the historical Piedmont 1994 flood (Davolio et al. 2020), when heavy precipitation was triggered by a similar but less persistent large-scale pattern.

The last precipitation extreme considered here investigates extreme rainfall, strong winds and below normal daytime temperatures over tropical northeastern Australia in early February 2019. The event caused wide-spread infrastructure damage, coastal inundation to homes, and destroyed over 500,000 livestock, predominantly beef cattle (losses were in the dark green areas in Fig. 5g). The total economic loss was estimated at $5.68 billion AUD (Deloitte 2019). The extreme rainfall was associated with a quasi-stationary monsoon depression that lasted around 10 days, with weekly rainfall totals above 1000 mm in some locations, maximum temperatures of 8-12°C below average, and sustained winds between 30 to 40 km/h (Bureau of Meteorology 2019). The event was associated with an active MJO that stalled over the western Pacific (Cowan et al. 2019). Even though most of the predictability in extreme austral summer precipitation for northeastern Australia comes from equatorial Pacific SSTs (King et al. 2014), ENSO conditions were neutral and likely did not contribute to this event. Consistent with the neutral ENSO conditions, the Australian Bureau of Meteorology issued a monthly rainfall outlook for February with little indication of the impending event. Only in the week prior to the event, the Bureau’s dynamical prediction system, the Australian Community Climate Earth-System Simulator-Seasonal version 1 (ACCESS-S1), predicted a more than doubled likelihood of extreme rainfall (Cowan et al. 2019).

The operational real-time forecasts initialized on 17 January 2019 (i.e., a week 3 forecast) confirm the above analysis (Fig. 5h). The region with the highest observed rainfall accumulations (blue
box in Fig. 5g) has a ROC score between 0.4 and 0.6, indicating low model performance (Fig. 6g). Likewise, wide-spread Spearman-rank correlations of between 0 and 0.25 (Fig. 6h) provide further evidence that the week 3 forecast does not predict the extreme rainfall week. This confirms separate results from eleven S2S models that suggest the rainfall event’s very extreme nature could not be predicted with certainty more than a week ahead (not shown).

d. Cyclones

We here analyze the subseasonal predictability of four cyclones (three tropical cyclones and one medicane). While all selected tropical cyclones occurred in different regions, all were associated with an active MJO, as discussed below.

As a first case we investigate tropical cyclone (TC) Claudia (Fig. 7a) in the western part of the Australian basin classified as a severe TC in the Australian scale. TCs in the western part of the Australian basin represent an important challenge to the oil industry since the majority of Australian oil rigs are located in this region. Therefore, the predictability of tropical cyclones a few weeks in advance in western Australia has important economic value, as well as societal impact in the case of landfall. Climatologically, 5.2 cyclones occur in that sub-basin per season, with 2.6 reaching severe TC intensity and 1.2 making landfall in Australia (Chand et al. 2019). The Australian TC season typically lasts from November to April, with a peak in January to March. Claudia’s characteristics (e.g., lifetime, latitude of genesis, maximum intensity and dissipation) were very typical of western Australia TCs (Chand et al. 2019). Claudia developed over Indonesia’s Maluku Island on 4 January 2020 and moved south-westward along the northwestern coast of Australia for about 2 weeks (including a period as a tropical depression) (Fig. 7a,b). It reached a peak intensity of 968 hPa (140 km/h) on January 13.
The prediction system initialized on 30 December 2019 predicted probabilities of up to 40% for a TC north-west of Australia for lead times of 15-21 days (week 3) (Fig. 7b) – significantly higher than the climatological probability (about 5%) for this season. Although the observed TC track is located slightly north of the area of maximum probability, this result suggests that the forecast could have provided a useful early warning for this TC. While other models from the S2S database also predicted an increased risk of TC activity in this region, the multi-model ensemble probability of TC strike was only around 10-20%. Claudia coincided with an exceptionally intense MJO (3 standard deviations above climatology of the RMM index (Wheeler and Hendon 2004)) over the Maritime Continent and warm SST anomalies over the eastern Indian Ocean. This combination is likely to have contributed to make this intense and long-lasting tropical cyclone more predictable than usual.

Another recent example of a well-predicted system is cyclone Belna (Fig. 7c) just a few months earlier. Belna formed to the north of the Mozambique channel and eventually moved southward. Cyclones occur in the channel on average twice per year (Kolstad 2021). Over recent years, multiple tropical cyclones made landfall in that region (Idai and Kenneth in 2018/19 and Chalane, Eloise, Guambe and Iman in 2020/21), leading to devastating floods in Mozambique and neighboring countries (Emerton et al. 2020).

For cyclone Belna (Fig. 7c), the model prediction initialized on 18 November predicts a probability of cyclone occurrence of up to 30% in the Mozambique Channel at the remarkable lead time of four weeks (Fig. 7d). On 5 December, 17 days after forecast initialization, the system was upgraded to a tropical storm and named. On 7 December it attained hurricane intensity, and a day later it passed near the Mayotte Islands in the northernmost part of the Channel. It made landfall in Madagascar on 9 December, to the east of the predicted path (Fig. 7d), and it dissipated over land two days later. A reason for the successful long-range prediction of Belna is likely the
strong MJO envelope within which Belna formed (letter B in Fig. 8c), although the MJO was not successfully predicted thereafter. The model forecast (Fig. 8d) indicates enhanced convection in that area, particularly in early December when Belna developed. The very intense TC Ambali (marked "A" in Fig. 8c) also formed near the MJO envelope just to the east of Belna.

Another TC associated with an intense MJO event occurred during a period of unusually high TC activity in the West Pacific. In early June 2015, an MJO convective envelope developed over the Indian Ocean, intensified and propagated eastward reaching an amplitude of 2.58 in the Realtime OLR MJO Index (ROMI) (Kiladis et al. 2014). Only two other MJO events during June and July in the period 1979-2018 reached this amplitude. This MJO event provided favorable conditions for TC formation leading to the genesis of typhoons Linfa, Chan-hom (Fig. 7e), and Nangka over the Western North Pacific, exemplified by the observed OLR anomalies and MJO-filtered OLR anomalies (Fig. 8a). Typhoons Linfa, Chan-hom, and Nangka (denoted by letters C, L, and N) in late June and early July formed soon after the passage of the MJO envelope. All three storms would go on to make landfall; Chan-hom was responsible for the second highest damages (1.5 billion U.S.D) in the West Pacific that season (Camargo 2016). Additional TCs in both the Indian Ocean and West Pacific were associated with this MJO event (Fig. 8a).

The ensemble forecast initialized on June 15, 2015 (0000Z, Fig. 7f) indicates the increased probability of a TC during week 4 of the forecast (valid July 7-13) in this area. The tracks of typhoons Linfa, Chan-hom, and Nangka (from west to east) overlap this area of enhanced TC formation probability. The forecast also captures the eastward propagation of the MJO envelope (Fig. 8b), although the MJO amplitude is weaker than observed.

As a last case we investigate a medicane, specifically the Mediterranean Cyclone 2018 - M02 Zorbas (Fig. 7g). The medicane developed on September 27, 2018 in the eastern Mediterranean Sea between Sicily and Southern Greece and gradually intensified, developing characteristics of a...
tropical cyclone. As for many medicanes, its origin was related to a potential vorticity streamer (Miglietta et al. 2017). On September 29, the storm made landfall at peak intensity in Kalamata, Peloponnese, Greece, with a pressure of 989 hPa and sustained winds of 120 km/h (approx. 33 m s$^{-1}$). The event was associated with a Dvorak number of T4.0 (Service 2019; ECMWF 2019), corresponding to a marginal category 1 hurricane.

The initialization on September 13, 2018 predicts a region of formation shifted to the west compared to the actual area of event formation (Gulf of Sirte, Libya) (Fig. 7h). While the low probability of formation is an indication of the difficulty of predicting such a rare event, the climatological probability of cyclone formation in the model in this region is less than 1%, hence the displayed chance of a cyclone in this region is clearly above the expected probability. In addition, the prediction shows low probability for the event to follow the observed path (black line) towards Greece. One of the reasons for the limited predictability of the event was likely the uncertainty in the initial conditions near an upper-level jet streak over the Gulf of Saint Lawrence (Portmann et al. 2019).

However, predictability may potentially be improved using CAPE (see section 2). For an initialization of the model as early as August 30, 2018 and a validation on September 26, 2018, very high values of CAPE are found in the formation region of medicane Zorbas (Fig. 8). Hence, CAPE provides evidence of a medicane 3-4 weeks prior to its formation. Further analysis is needed to assess the full predictability capabilities of CAPE for medicanes.

4. General Discussion and Outlook

We have here demonstrated subseasonal predictability for selected case studies of some of the most prominent and impactful extreme events globally, namely heatwaves, cold spells, precipitation events, and cyclones. Heatwaves tend to be the most predictable among the extreme events
considered. The prediction system can often anticipate the anomalous temperature 3-4 weeks in advance, though often with a reduced amplitude. Cold spells also often show an indication of predictability, generally at lead times of 2-3 weeks. Precipitation events tend to be less predictable, but if the large-scale circulation associated with a large-scale driver (e.g., an active MJO) is successfully captured, predictability of 2-3 weeks can be obtained. For tropical cyclones, their formation region and tracks can often be anticipated 3 weeks in advance provided a successful prediction of strong MJO events. Furthermore, CAPE shows promise for indicating tracks and formation regions for extratropical cyclones. Note that these conclusions are based on the here documented case studies, and although the predictability and conclusions obtained here agree with other published results, it is likely that individual events may be much more or less predictable depending on the region, type, and amplitude of the event. Therefore, in addition to differences in predictability between different types of extremes there are important differences in predictability within the same event type. In the here demonstrated case studies, these inter-event differences hint at different processes and precursors responsible for forcing, modulating, or amplifying certain extreme events of the same type, including remote drivers such as the MJO.

We would like to emphasize that the case studies presented here do not represent a comprehensive evaluation, hence the predictability shown for these events may differ from a systematic evaluation across a larger number of events. Hence, while this study only investigates a limited number of extreme events as case studies, systematic studies of inter-event differences in predictability will be required to better understand the role of the identified drivers. In particular, extreme events with a common remote driver could be cross-compared in order to more clearly evaluate the driver’s role (or, in fact, its absence). These studies should also include an investigation of false alarms, that is, extreme events triggered by remote drivers and predicted in the model that do not verify in observations.
An improved process understanding of the drivers of extremes and their representation in prediction systems as well as the development of post-processing techniques will continue to significantly benefit the subseasonal prediction of extreme events. On the other hand, even with significant model improvements, many extremes will retain an inherent unpredictability related to the chaotic nature of the climate system. Still, understanding why and when certain extreme events are more predictable than others will help to identify and use windows of opportunity, that is, atmospheric states with enhanced predictability. Event-based and region-specific knowledge of the level of predictability of the relevant processes and the related extreme events will significantly benefit stakeholders and users of extreme weather data.

While this study has focused on a single prediction system from the ECMWF, an increasing number of multi-model studies for the prediction of specific extremes are currently becoming available (e.g. Li et al. 2021; Materia et al. 2020b; Domeisen et al. 2020b), highlighting inter-model differences rather than inter-event differences, which were the focus of this study. Furthermore, bias correction and calibration methodologies that refine the forecast’s statistical properties based on a reference period will further enhance these forecasts. In this study we used anomalies in order to correct the systematic bias and model drift, keeping in mind that this might affect the chance of the model to predict, for example, hot versus cold spells, especially for longer lead times. However, region- and process-specific biases and drifts are likely still present in our analysis. In addition, standard bias-correction applied here is "unfair" (Risbey et al. 2021), since it uses observed data that would not be available to a real-time forecast: in fact, in several cases the observations used for the climatology occur after the forecast starts, and the hindcast therefore contains artificial skill. This can be misleading for users who must take decisions using real forecasts, which are likely to exhibit lower forecast skill than what is commonly shown in research studies.
In addition, a wider range of model evaluation and bias correction techniques are available, with the most relevant choices depending on the variable and on the desired characteristics for the output (see Torralba et al. (2017) and Manzanas et al. (2019) for a comparison of methodologies for seasonal predictions and Wernli et al. (2008); Dorninger et al. (2018) for forecast evaluation techniques on deterministic timescales). Although some standard methods and tools are starting to be used more widely (Muñoz 2020; Muñoz and Coauthors 2019), implementation at subseasonal timescales is non-trivial and requires a robust climatological reference to be successful (Manrique-Suñén et al. 2020). One of the challenges is the limited amount of model data available for the reference period (short hindcast periods and few ensemble members). Examples of implementation of bias-correction methodologies for subseasonal predictions can be found in Monhart et al. (2018) and Manrique-Suñén et al. (2020). These statistical adjustments are of particular importance in sectoral applications (Materia et al. 2020a; DeMott et al. 2021; DiSera et al. 2020), when S2S predictions are used as input in impact models to calculate sector-relevant indicators or derived variables (e.g., energy production or agricultural yield (White et al. 2021)). As S2S predictions increasingly make their way into risk-based decision-making contexts, a continued development and assessment of subseasonal models, calibration techniques, and combination with other tools will significantly benefit these applications (Goddard et al. 2014; White et al. 2021).

Lastly, it remains difficult to quantify the economic value of S2S forecasts. In fact, even for very skillful forecasts, there can be significant economic losses that depend on factors beyond the forecasts themselves, involving the emergency response and preparedness of the affected region. However, it is clear that skillful forecasts on sub-seasonal to seasonal timescales can indeed add economic value, as has been shown for both temperature and cyclone predictions (Dorrington et al. 2020; Emanuel et al. 2012).
In summary, this work is meant to showcase the importance of subseasonal forecasts in the development and improvement of a large variety of climate services. Therefore, it is difficult to homogenize across event type, forecast quality metrics, and prediction format (deterministic versus probabilistic). By their own nature, distinct events in different locations of the world require different verification tools, and time aggregations must be meaningful to users. This study goes towards this direction by starting to address the recommendations for advancing the S2S forecast verification practices recently highlighted by Coelho et al. (2019): Appropriate verification methods to deal with extreme events, novel verification measures specifically adapted for S2S forecasts, and enlargement of the sample size to address sampling uncertainties. All of these techniques are meant to build knowledge about the strengths and weaknesses of forecasts, and eventually increase confidence in S2S products among forecasters and users (Coelho et al. 2018).

As the performance of prediction models for extreme events at subseasonal lead times continues to increase with improvements in the understanding of extreme events and their representation in models, the here documented extreme events can be viewed as demonstrations and examples of this progress, which reaches far beyond these case studies, contributing to build or strengthen (depending on the case) a robust ecosystem of climate services (Goddard et al. 2020).

Acknowledgments. Support from the Swiss National Science Foundation through project PP00P2_170523 to D.D., H.A.-G. and O.W. is gratefully acknowledged. In addition, support from the Swiss National Science Foundation through project PP00P2_198896 to D.D. is gratefully acknowledged. H.A.-G. also acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 891514. Á.G.M., C.G.R. and D.P. were partially supported by the Columbia World Project “ACToday”, at Columbia University in the City of New York (https://iri.columbia.edu/actoday/);
Á.G.M. was also partially supported by the NOAA grant NA18OAR4310275. A.M.-S., L.L., L.P., H.B., P.G., A.C.-P., D.B. and A.S. acknowledge funding from the European Union’s Horizon 2020 research and innovation programme under Grant 7767874 (S2S4E). SJC acknowledges support of NOAA S2S project NA16OAR4310079. TC is supported by the Northern Australian Climate Program (NACP). DT and HM are supported by the National Institute of Food and Agriculture Hatch project (Accession No. 1012578). S.-W.S. and H.K. are supported by the Basic Science Research Program through the National Research Foundation of Korea (2017R1E1A1A01074889). L.E.P acknowledges support of Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEPRA grant). We acknowledge the use of the CSTools software package (Perez-Zanon et al. 2019) and R (R Core Team 2015) to produce figures 2 and 4.

Data availability statement. ERA5 reanalysis data was obtained from the Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home. The ECMWF S2S model data was obtained through the MARS archive (https://apps.ecmwf.int/datasets/data/s2s/). CPC Global Unified Precipitation data were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, U.S.A, from their Web site at https://www.psl.noaa.gov/thredds/catalog/Datasets/cpc_global_precip/catalog.html. Australian precipitation data from the Australian Water Availability Project (AWAP) is available on request from the Bureau of Meteorology at http://www.bom.gov.au/climate/austmaps/metadata-daily-rainfall.shtml. The satellite image for tropical cyclone Claudia was captured by NOAA-20 satellite’s IITS instrument [https://www.nesdis.noaa.gov/content/tropical-cyclone-claudia-loses-strength-it-moves-away-australia]. The satellite image for cyclone Belna was obtained from https://en.wikipedia.org/wiki/Cyclone_Belna [NASA: https://worldview.earthdata.nasa.gov/]. The satellite image for typhoon Chan-Hom was ob-
tained from https://en.wikipedia.org/wiki/Typhoon_Chan-hom_%282015%29 [SSEC/CIMSS, University of Wisconsin–Madison]. The satellite image for medicane Zorbas is a MODIS image captured by NASA’s Terra satellite (EOSDIS Worldview) from https://commons.wikimedia.org/wiki/File:Zorbas_2018-09-29_0912Z.jpg. The ECMWF CAPE data for studying medicane Zorbas were obtained from the IRI/LDEO Climate Data Library (https://iridl.ldeo.columbia.edu/SOURCES/.ECMWF/.S2S). Observed tropical cyclone data are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010) at https://climatedataguide.ucar.edu/climate-data/ibtracs-tropical-cyclone-best-track-data.

References


Auffhammer, M., P. Baylis, and C. H. Hausman, 2017: Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the united states. Proceedings of the National Academy of Sciences, 114 (8), 1886–1891.


Bureau of Meteorology, 2019: Special Climate Statement 69 — an Extended Period of Heavy Rainfall and Flooding in Tropical Queensland.


Ferranti, L., L. Magnusson, F. Vitart, and D. S. Richardson, 2018: How far in advance can we predict changes in large-scale flow leading to severe cold conditions over europe? *Quarterly Journal of the Royal Meteorological Society*, **144** (715), 1788–1802.


cyclone genesis prediction and MJO in the S2S dataset. Wea. Forecasting, 33, 967–988, doi:
10.1175/WAF-D-17-0165.1.

Lee, M., and T. Frisius, 2018: On the role of convective available potential energy (CAPE) in
tropical cyclone intensification. Tellus A: Dynamic Meteorology and Oceanography, 70 (1),

Lehtonen, I., and A. Y. Karpechko, 2016: Observed and modeled tropospheric cold anomalies
associated with sudden stratospheric warmings. Journal of Geophysical Research: Atmospheres,


Leroy, A., and M. C. Wheeler, 2008: Statistical prediction of weekly tropical cyclone activity in


Li, J., T. Ding, X. Jia, and X. Zhao, 2015: Analysis on the extreme heat wave over China around
Yangtze River region in the summer of 2013 and its main contributing factors. Advances in
Meteorology, 2015.

Li, M., Y. Yao, D. Luo, and L. Zhong, 2019: The linkage of the large-scale circulation pattern to a
long-lived heatwave over Mideastern China in 2018. Atmosphere, 10 (2), 89.

Li, Chao, Zwiers, Francis, Zhang, Xuebin, Li, Guilong, Sun, Ying, and Wehner, Michael, 2021: Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models. *Journal of Climate*, **34** (9), 3441–3460.


Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Ob-


Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the

national-climate-201807.

Noer, G., Ø. Saetra, T. Lien, and Y. Gusdal, 2011: A climatological study of polar lows in the

Pan, B., K. Hsu, A. AghaKouchak, S. Sorooshian, and W. Higgins, 2019: Precipitation prediction
skill for the West Coast United States: From short to extended range. *Journal of Climate*, **32**, 
161–182.

Pepler, A. S., L. B. Díaz, C. Prodhomme, F. J. Doblas-Reyes, and A. Kumar, 2015: The ability of
a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet

Perez-Zanon, N., and Coauthors, 2019: CSTools: *Assessing Skill of Climate Forecasts on Seasonal-
to-Decadal Timescales*. URL http://CRAN.R-project.org/package=CSTools, r package version 2.0.0.

Perkins, S., L. Alexander, and J. Nairn, 2012: Increasing frequency, intensity and duration of

Pfahl, S., and H. Wernli, 2012: Quantifying the relevance of atmospheric blocking for co-located
temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophysical
Research Letters*, **39** (12).


Risbey, J. S., and Coauthors, 2021: Standard assessments of climate forecast skill can be misleading. 

*Nature Communications*, **12** (1), 4346–14.


Specq, D., L. Batté, M. Déqué, and C. Ardidouze, 2020: Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific. Earth and Space Science, 7 (9), e2019EA001 003.


Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Clim. Atmos. Sci., 1*, 3, doi:10.1038/s41612-018-0013-0.


LIST OF TABLES

Table 1. Overview of the case studies evaluated in this study. 64
Table 1. Overview of the case studies evaluated in this study.

<table>
<thead>
<tr>
<th>Location / target region</th>
<th>Forecast target period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HEATWAVES</strong></td>
<td></td>
</tr>
<tr>
<td>Western U.S. (California) (235 - 250°E, 32 - 48°N)</td>
<td>23-29 July 2018</td>
</tr>
<tr>
<td>Central / northeastern Europe (10 - 20°E, 50 - 60°N)</td>
<td>23-29 July 2018</td>
</tr>
<tr>
<td>Southeastern U.S. (92 - 70°W, 25 - 45°N)</td>
<td>24-30 May 2019</td>
</tr>
<tr>
<td>East Asia (eastern China, Korea, Japan) (105 - 130.5°E, 30 - 40.5°N)</td>
<td>5-11 August 2013</td>
</tr>
<tr>
<td><strong>COLD SPELLS</strong></td>
<td></td>
</tr>
<tr>
<td>Southeastern Europe (10.5 - 30°E, 37.5 - 54°N)</td>
<td>3-9 April 2003</td>
</tr>
<tr>
<td>Central / northern Europe (12.5°W - 30°E, 37.5 - 65°N)</td>
<td>26 February - 3 March 2018</td>
</tr>
<tr>
<td>Southwestern Europe (France) (4.5°W - 7.5°E, 43.5 - 49.5°N)</td>
<td>16-22 January 2017</td>
</tr>
<tr>
<td>Northern Europe (UK, Germany, Scandinavia) (10°W - 30°E, 45 - 65°N)</td>
<td>27 November - 3 December 2010</td>
</tr>
<tr>
<td><strong>PRECIPITATION EVENTS</strong></td>
<td></td>
</tr>
<tr>
<td>Volcán de Fuego, Guatemala (91 °W, 14.5 °N)</td>
<td>18-24 June 2018</td>
</tr>
<tr>
<td>Northwestern Ecuador (79 °W, 0 °N)</td>
<td>21-27 January 2016</td>
</tr>
<tr>
<td>Northwestern Italy (6.5 - 10°E, 43.5 - 46.5°N)</td>
<td>21-27 November 2016</td>
</tr>
<tr>
<td>Northeastern Australia (138°-147°E, 18°-22°S)</td>
<td>31 January - 6 February 2019</td>
</tr>
<tr>
<td><strong>CYCLONES</strong></td>
<td></td>
</tr>
<tr>
<td>Western Australia: Cyclone Claudia (no landfall)</td>
<td>5 January 2020 (formation) / 18 January 2020 (dissipation)</td>
</tr>
<tr>
<td>Mozambique Channel: Cyclone Belna (landfall: Madagascar)</td>
<td>2 December 2019 (formation) / 9 December 2019 (landfall)</td>
</tr>
<tr>
<td>Western North Pacific: Typhoon Chan-hom (landfall: China)</td>
<td>29 June 2015 (formation) / 11 July 2015 (landfall)</td>
</tr>
<tr>
<td>Mediterranean: Medicane Zorbas (landfall: Peloponnese, Greece)</td>
<td>27 September 2018 (formation) / 29 September 2018 (landfall)</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1. **Heatwaves**: (a,c,e,g) 2m temperature anomalies for the target week (indicated in the panel titles) from ERA5 data and (b,d,f,h) predicted by the ECMWF week 3 forecasts (hindcasts prior to 2016), initialization dates indicated in panel titles. (a,b) California heatwave, (c,d) European heatwave, (e,f) U.S. heatwave, (g,h) East Asia heatwave. White boxes indicate the averaging areas used for Fig. 2. All case studies use model version CY45R1, except for the East Asia heatwave, which uses CY46R1. 67

Fig. 2. **Heatwaves**: The PDF distribution of the predicted 2m temperature anomalies from the model ensemble averaged over the target week (indicated in table 1) for the heatwave case studies, averaged over the white boxes in Fig. 1 and initialized at (panels from left to right) 4, 3, and 2 weeks before the start of the target week. (a) California heat wave 2018, (b) European heat wave 2018, (c) southeastern U.S. heat wave 2018, and (d) east Asia heatwave 2013. Tercile limits (below-normal: blue, normal: gray, and above-normal: red) are computed with respect to the lead time - dependent model climatology. Values above the 66th percentile (below the 33rd percentile) are represented by red (blue) shading. Grey shading represents values between these terciles. The yellow dots indicate the ensemble members that were used to construct the PDF (51 for forecasts, 11 for hindcasts) 68

Fig. 3. **Cold spells**: Same as Figure 1 but for the cold spell case studies: (a,b) Southeastern Europe cold spell in 2003 (model version CY46R1), (c,d) central/northern European cold spell in 2018 (model version CY43R3), (e,f) France cold spell in 2017 (model version CY43R1), (g,h) northern European cold spell in 2010 (model version CY46R1) 69

Fig. 4. **Cold spells**: Same as Figure 2 but for the cold spell case studies: (a) Southeastern Europe cold spell in 2003, (b) European cold spell in 2018, (c) France cold spell in 2017, and (d) the northern European cold spell in 2010 70

Fig. 5. **Precipitation events**: Accumulated precipitation anomalies over the target week (week 3, indicated in the panel titles) for (a,c,e,g) observations and (b,d,f,h) the ECMWF model prediction (initialization date indicated in the panel title). (a,b) Guatemala, (c,d) western Ecuador (e,f) northwestern Italy, and (g,h) northeastern Australia. The blue boxes or dots, respectively, indicate the target location for each case study, as indicated in Table 1. Observations are from (a,c,e) CPC and (g) AWAP 71

Fig. 6. **Precipitation extremes**: Predictability scores for week 3, (a,c,e,g) assessed through the area under the ROC curve for the above-normal category, and (b,d,f,h) Spearman’s rank correlation coefficient. The results were interpolated to the CPC unified grid. For details of the scores see section 2. (a,b) Guatemala, (c,d) western Ecuador (e,f) northwestern Italy, and (g,h) northeastern Australia. The blue boxes or dots are as in Figure 5 72

Fig. 7. **Cyclones**: Satellite images at a time close to the maximum intensity of the storms for (a) cyclone Claudia on January 13, 2020 [NOAA], (b) cyclone Belna on December 7, 2019 [NASA], (c) typhoon Chan-hom on July 10, 2015 [SSEC/CIMSS, University of Wisconsin–Madison], and (d) medicane Zorbas (2018M02) on September 29, 2018 [MODIS NASA]. Probability of cyclone occurrence for (b) Claudia initialized on 30/12/2019 for lead times of 15–21 days, (d) Belna initialized on 18/11/2019 for lead times of 22–28 days, (f) Chan-hom initialized on 15/06/2015 for lead times of 22–28 days, and (h) medicane Zorbas initialized on 13/09/2018 for lead times of 0–32 days. Black lines indicate the observed cyclone tracks during the verification period, and the names of the cyclones corresponding to the tracks are indicated. The different choice of lead times for the case studies refers to the furthest lead time for which the events were possible to be predicted. 73
Fig. 8. **Cyclones:** Outgoing longwave radiation (OLR) anomalies (shaded, W m$^{-2}$) and MJO-filtered OLR anomalies (red contours, every 15 W m$^{-2}$ for negative values) from (a,c) observations averaged over 0-10°N and 0-10°S with tropical cyclone tracks (black lines) and names (first letter of the cyclone name in red circle) and (b,d) ECMWF ensemble forecasts initialized on 15/06/2015 and 18/11/2019. MJO-filtering is performed using a wavenumber-frequency filter that selects for wavenumbers 0-9 and periods of 20-100 days. MJO-filtered OLR was calculated by padding the forecast with observations prior to initialization following the methodology described in Janiga et al. (2018). (e) CAPE (J kg$^{-1}$) from the ECMWF ensemble forecast initialized on 30/08/2018, valid on 26/09/2018.
Fig. 1. Heatwaves: (a,c,e,g) 2m temperature anomalies for the target week (indicated in the panel titles) from ERA5 data and (b,d,f,h) predicted by the ECMWF week 3 forecasts (hindcasts prior to 2016), initialization dates indicated in panel titles. (a,b) California heatwave, (c,d) European heatwave, (e,f) U.S. heatwave, (g,h) East Asia heatwave. White boxes indicate the averaging areas used for Fig. 2. All case studies use model version CY45R1, except for the East Asia heatwave, which uses CY46R1.
**Fig. 2.** **Heatwaves:** The PDF distribution of the predicted 2m temperature anomalies from the model ensemble averaged over the target week (indicated in table 1) for the heatwave case studies, averaged over the white boxes in Fig. 1 and initialized at (panels from left to right) 4, 3, and 2 weeks before the start of the target week. (a) California heat wave 2018, (b) European heat wave 2018, (c) southeastern U.S. heat wave 2019, and (d) east Asia heat wave 2013. Tercile limits (below-normal: blue, normal: gray, and above-normal: red) are computed with respect to the lead time - dependent model climatology. Values above the 66th percentile (below the 33rd percentile) are represented by red (blue) shading. Grey shading represents values between these terciles. The yellow dots indicate the ensemble members that were used to construct the PDF (51 for forecasts, 11 for hindcasts). The extremes above the 90th (below the 10th) percentile are hatched and their probabilities are indicated by red (blue) numbers, respectively. The purple dashed line represents the anomaly in ERA5 reanalysis averaged over the target week.
Fig. 3. **Cold spells:** Same as Figure 1 but for the cold spell case studies: (a,b) Southeastern Europe cold spell in 2003 (model version CY46R1), (c,d) central / northern European cold spell in 2018 (model version CY43R3), (e,f) France cold spell in 2017 (model version CY43R1), (g,h) northern European cold spell in 2010 (model version CY46R1).
Fig. 4. Cold spells: Same as Figure 2 but for the cold spell case studies: (a) Southeastern Europe cold spell in 2003, (b) European cold spell in 2018, (c) France cold spell in 2017, and (d) the northern European cold spell in 2010.
Fig. 5. **Precipitation events:** Accumulated precipitation anomalies over the target week (week 3, indicated in the panel titles) for (a,c,e,g) observations and (b,d,f,h) the ECMWF model prediction (initialization date indicated in the panel title). (a,b) Guatemala, (c,d) western Ecuador (e,f) northwestern Italy, and (g,h) northeastern Australia. The blue boxes or dots, respectively, indicate the target location for each case study, as indicated in Table 1. Observations are from (a,c,e) CPC and (g) AWAP.
Fig. 6. **Precipitation extremes:** Predictability scores for week 3, (a,c,e,g) assessed through the area under the ROC curve for the above-normal category, and (b,d,f,h) Spearman’s rank correlation coefficient. The results were interpolated to the CPC unified grid. For details of the scores see section 2. (a,b) Guatemala, (c,d) western Ecuador (e,f) northwestern Italy, and (g,h) northeastern Australia. The blue boxes or dots are as in Figure 5.
Fig. 7. **Cyclones:** Satellite images at a time close to the maximum intensity of the storms for (a) cyclone Claudia on January 13, 2020 [NOAA], (c) cyclone Belna on December 7, 2019 [NASA], (e) typhoon Chan-hom on July 10, 2015 [SSEC/CIMSS, University of Wisconsin–Madison], and (g) medicane Zorbas (2018M02) on September 29, 2018 [MODIS NASA]. (b,d,f,h) Probability of cyclone occurrence for (b) Claudia initialized on 30/12/2019 for lead times of 15–21 days, (d) Belna initialized on 18/11/2019 for lead times of 22–28 days, (f) Chan-hom initialized on 15/06/2015 for lead times of 22–28 days, and (h) medicane Zorbas initialized on 13/09/2018 for lead times of 0–32 days. Black lines indicate the observed cyclone tracks during the verification period, and the names of the cyclones corresponding to the tracks are indicated. The different choice of lead times for the case studies refers to the furthest lead time for which the events were possible to be predicted.
Fig. 8. **Cyclones:** Outgoing longwave radiation (OLR) anomalies (shaded, W m\(^{-2}\)) and MJO-filtered OLR anomalies (red contours, every 15 W m\(^{-2}\) for negative values) from (a,c) observations averaged over 0-10\(^\circ\) N and 0-10\(^\circ\) S with tropical cyclone tracks (black lines) and names (first letter of the cyclone name in red circle) and (b,d) ECMWF ensemble forecasts initialized on 15/06/2015 and 18/11/2019. MJO-filtering is performed using a wavenumber-frequency filter that selects for wavenumbers 0-9 and periods of 20-100 days. MJO-filtered OLR was calculated by padding the forecast with observations prior to initialization following the methodology described in Janiga et al. (2018). (e) CAPE (J kg\(^{-1}\)) from the ECMWF ensemble forecast initialized on 30/08/2018, valid on 26/09/2018.