CHAPTER 2

Operational multiscale predictions of hazardous events

Linus Magnusson[1], C. Prudhomme[1], F. Di Giuseppe[1], C. Di Napoli[1,2] and F. Pappenberger[1]
[1]European Centre for Medium-Range Weather Forecasts-ECMWF, Reading, United Kingdom
[2]University of Reading, Reading, United Kingdom

2.1 Introduction

Providing forecasts that form a useful base for actions ahead of high-impact weather events, involves a cascade of models and predictions targeting different timescales. The production chain includes the estimation of initial and boundary conditions as well as state variables for the atmospheric forecasts, hazard models and impact models. The combination of the models will lead up to warnings and decisions (Zhang et al., 2019). In this chapter we will cover the multi-(time)scale aspect of severe weather forecasting, and the use of these forecasts in hazard forecasting. Based on the experience of these different steps, we discuss the requirements for a weather forecasting system to provide the right ingredients for the forecasting chain.

The notion of multiscale predictability of severe weather builds on the concept of a cascade of processes with different horizontal and temporal scales that determine the final event. It could, for example, be a (1) lightning from a (2) convective cell, that is embedded in a (3) cold front associated with an (4) extra-tropical cyclone, which formed due to (5) large-scale baroclinic conditions. (1) through (5) have very different scales and predictability.

Forecast will always contain a level of uncertainty; it will never be possible for example to predict which second and location a single lightning will strike. It is therefore natural to talk about a level of risk building on the probability for an event to happen. In weather forecasting today such probabilities are usually based on ensemble forecasts. Ensemble forecasts contain a number of scenarios (forecasts) from which the probability for an event to take place can be estimated. Other techniques to estimate uncertainties are often employed in hazard and impact models.

Extreme Weather Forecasting
DOI: https://doi.org/10.1016/B978-0-12-820124-4.00008-6

Fig. 2.1 illustrates examples of ensemble forecasts on different time ranges for a selected event. Starting from a point-in-time long before (beyond the limit of predictability) the event occurred, the probability distribution function (PDF) from the forecast is likely to be very similar to the climatological distribution, with per se a low probability for extreme events. As we approach the event (week(s) before), some slight shifts of the forecasts PDF from the climatology might appear, either because a few members pick up an extreme scenario or most of the members are slightly shifted towards an anomaly due to some large-scale forcing. Closer in time to the event (medium-range, 3 to 10 days ahead) the PDF is skewed towards an extreme solution and in short-range forecasts (0−3 days) the ensemble (hopefully) sharpens around the (the later) observed value. However, if the magnitude of the event is not within the envelope of what the model can simulate, the severity of the event will be missed also by the shortest forecasts. In this chapter we will discuss processes that influence the PDF on different timescales and how to define shortcomings in the prediction system.

The outcome of an extreme event is often determined by several factors, both related to a single meteorological event and past events. For example, river flooding is often caused by a series of precipitation events together with preconditioning hydrological conditions and catchment properties; the life-threatening aspect of a heatwave is a combination of

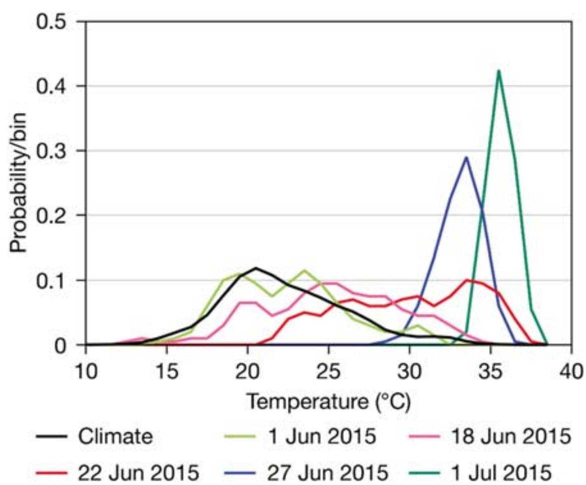

Figure 2.1 Example of evolving probability distribution function. The example is for 2-m temperature in Paris, July 1, 2015. *From ECMWF Newsletter 145.*

the magnitude and length of the period of extreme heat; and coastal flooding can be a combination of storm surge, waves, and precipitation as well as coastal properties. This combination of factors is often referred to as compound events in which the different factors can have different predictability.

How far ahead is a severe weather event predictable to a level where the forecast information is valuable enough to take action? This is a very obvious question from forecast users, but a very complex question to answer from a statistical perspective. First of all, different users have different requirements on the forecast quality and properties needed to take actions and these requirements might be different for different lead times, as forecasting extreme events is important on several timescales ranging from seasonal to very short timescales. Different types of extreme weather events have different predictability and properties and for similar types of events this can vary with season and location and other factors. Hence it is very difficult to give a generic answer to the question above from a user perspective.

In Section 2.3 we discuss sources and barriers for predictability of different events on different timescales. Besides the usefulness aspect of forecast verification, forecast verification is needed to find weaknesses in the prediction system to guide future development. Here one needs to understand the underlying predictability mechanisms for extreme events to identify key processes to be improved, and how such mechanisms vary with the lead time in mind.

Forecasters make use of a series of tools to predict severe weather. For the shortest timescale (0−6 hours) nowcasting techniques are used based on direct interpretation of the most recent observations for ground stations, radar and satellites, and statistically (or artificial intelligence-AI) based movement of the features. For longer lead times, Numerical Weather Prediction (NWP) forecasts are utilized. NWP is based on the determination of the state of the atmosphere that will serve as initial conditions by the use of data assimilation. This state is evolved forward in time by a model that is based on the laws of physics. For short-range forecasts (0−3 days), one uses regional NWP models with high spatial resolution and frequent output. For longer time ranges global models are used as the regional weather is determined by processes far away. As the timescale increases also the remote influences of boundary conditions such as ocean and land conditions start to impact the outcome. The global models are implemented in medium-range (3−10 days), extended-range (1−6 weeks), and up to seasonal (1−12 months) forecasting range.

NWP forecasts are often run to create an ensemble of forecasts. Ensemble forecasts aim to simulate the forecast uncertainties by evolving the uncertainties in the initial state and account for uncertainties in the numerical model. The PDF of the forecast ensemble can be determined under the assumption of the weight of each ensemble member. The output from NWP is used to drive impact models for hydrological conditions including floods and droughts, wildfires, health, etc.

In this chapter we will make use of the ECMWF forecasting system, but the concepts are independent of the choice of system. The chapter is organized as follows: The basics of NWP are outlined in Chapter 1.2 and will not be repeated in this chapter. An example of the prediction of a severe event on different timescales is discussed in Section 2.2. In Section 2.3 factors influencing the predictability for a selection of events are outlined, Section 2.4 discusses hazard modeling, Section 2.5 takes on the evaluation aspect of extremes and finally the chapter is concluded in Section 2.6.

2.2 Example case: 2015 European heatwave

In this section we will use one case of extreme weather to illustrate predictions of different timescales. The selected case is from July 2015, which was dominated by very warm weather in southern and western Europe as can be seen in the monthly mean temperature map in Fig. 2.2. At the same time Scandinavia and northeastern Europe instead saw cooler than normal conditions.

The ECMWF operational forecast system consists of 4 different components aimed for different timescales. In 2021, the four times a day (with shorter lead times for 06UTC and 18UTC, in parentheses) operational forecasts at ECMWF consists of a deterministic forecast (HRES) with 9 km resolution doing out to 10(4) days ahead and an ensemble with 50 perturbed members and 1 unperturbed member all with 18 km resolution going out to 15(6) days. Twice a week (Monday and Thursday 00UTC) ENS is extended out to 46 days with a resolution of 36 km after day 15 to form the extended-range forecasts. With the same configuration, a reforecasts dataset is created based on the same date for the past 20 years and 11 ensemble members. Finally, on the first day every month seasonal forecasts are created out to 7 months ahead with a configuration similar to the extended-range forecasts.

To first examine the predictions of the conditions during the period 29 June to 17 August in the extended-range forecasts, the panels in

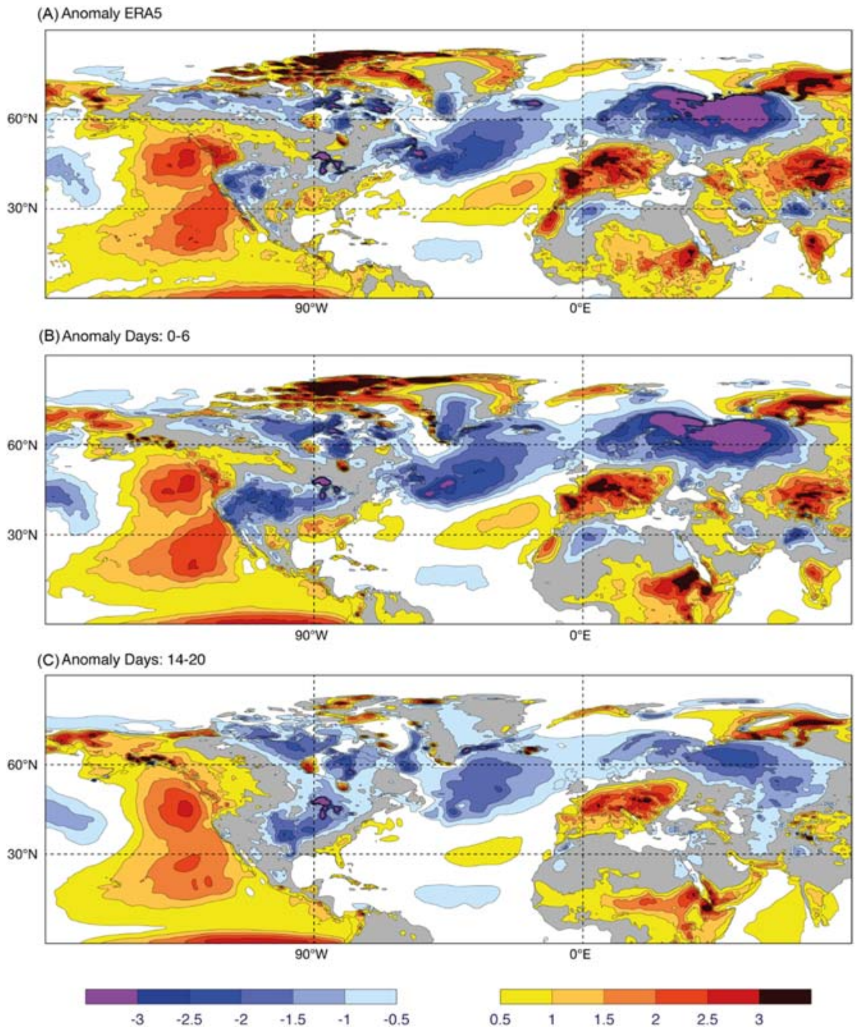

Figure 2.2 Temperature anomaly for the period June 29 to August 17, 2015 in ERA5 (A) and composites of ECMWF extended-range forecasts on lead time 0–6 days (B), and 14–20 days (C).

Fig. 2.2 show composites of 2-m temperature anomalies in ensemble mean forecasts valid the same period as the analysis in the top panel. The forecast anomalies are computed in respect to a 20-year reforecast dataset in order to account for lead time-dependent model biases. The forecasts are based on 7-day averages and are issued at the start of the event (week-1 forecast) and 2 weeks before the start (week-3 forecast). The anomaly patterns were well captured in the composite of week-1 forecasts as expected because of the short lead time. But the patterns were also well captured in the week-3 forecasts. Two likely sources for this skillful prediction in the extended-range are the sea-surface temperature pattern in the northern Atlantic and the soil moisture conditions over Europe. For the latter, as May and June 2015 had been dry in western Europe, the lack of soil moisture for evaporation would amplify a heatwave once the atmospheric conditions are favorable.

To evaluate shorter predictions, we will zoom in on the forecasts for Paris valid 1 July, which turned out to be the most extreme day of the season for Paris in terms of temperature. The observed maximum temperature in central Paris was 39.1 degrees Celsius. The extreme heat was primarily caused by a ridge that developed over south-western Europe at the end of June that advected very warm air northward.

Fig. 2.3 shows the evolution of forecasts for 2-m maximum temperature on 1 June 2015 in Paris, initialized at different times. The plot includes the ensemble distribution from the ECMWF ensemble and the ECMWF high-resolution deterministic forecast. The figure also includes

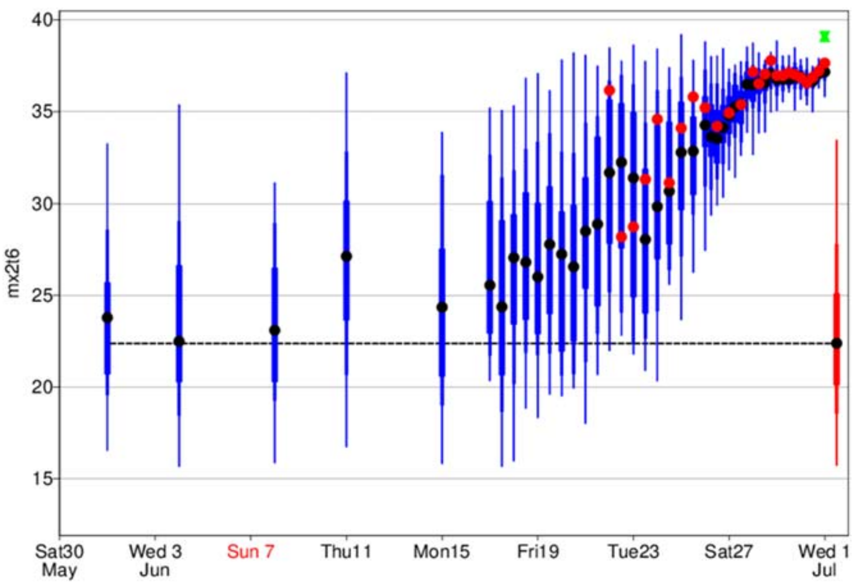

Figure 2.3 Forecast evolution for maximum temperature in Paris on July 1, 2015 in ECMWF ensemble forecasts (blue box-and-whisker), ECMWF HRES forecasts (red dot) and model climatology based on reforecasts (red box-and-whisker). The observed maximum is included in green.

the climatological distribution derived from the ECMWF reforecast dataset. Finally, the plot includes the observation from central Paris. As ECMWF produces ensemble forecasts with different lengths (46 days twice a week, 15 days twice a day and 6 days 4 times a day), we see more frequent forecasts for short lead times before the event.

Starting from one month before the event, one should not expect the forecast to show a strong signal, and indeed the forecast is very similar to the model climatology. But from June 11 and onward all ensemble medians (back dot) are clearly above the median of the model climatology. From about 14 days before the event the ensemble distribution started to gradually shift towards warmer temperatures and during the week before the event the ensemble spread reduced and the forecasts converged to a value well above the 99th percentile of the climatology. However, also the shortest forecast underpredicted the extreme temperature on 1 July. In this section we will discuss some of the processes active during these stages.

Looking at the forecasts initialized around 8–10 days before the event, the ensemble distribution is clearly shifted towards warmer temperatures and is also skewed towards the extreme tail of the distribution. The initialization of this forecast coincides with the time of the first detection of a Rossby wave packet over Western Pacific. The packet first appeared over the Western Pacific around June 22 and started to propagate eastward. The packet reached eastern Atlantic in the last days of June with a positive node (winds from south) over western Europe. By capturing the wave packet propagation, the probability for the ridge over western Europe increased and with that the probability for the heatwave. The effect of the presence of Rossby wave packets on the mid-latitude predictability is documented in, for example, Grazzini and Vitart (2015), who showed that the predictability is increased by the presence of long-lived Rossby wave packets.

One day before the event (30 June 00UTC), the ensemble distribution was narrow and centerd at 37°C. At this point in time the trough over the central Atlantic that later pushed the warm air northward had started to develop, giving a strong confidence in the development of the heatwave. However, when we compare the ensemble and HRES forecast with observed temperatures from stations across Paris, we find that the maximum temperature was underestimated by all individual forecasts, as seen in Fig. 2.4. One plausible explanation is that the model does not include urbanization and can therefore not capture heat islands in cities.

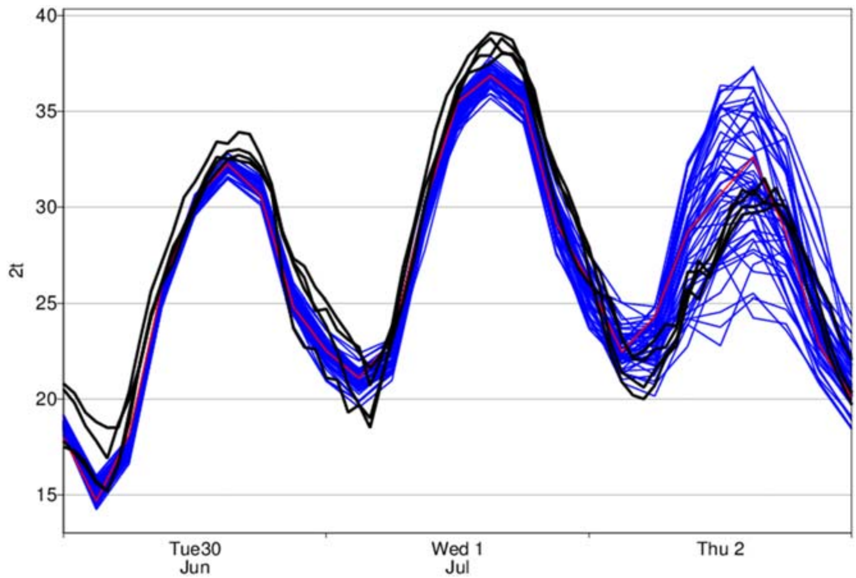

Figure 2.4 2-m temperature forecast for Paris from July 30, 00UTC. ECMWF ensemble members (blue), ECMWF HRES (red), and four different observation stations (black).

But the ECMWF verification of 2-m temperature also shows a more general underestimation of the diurnal cycle over Europe during summer time (Sandu et al., 2020).

To summarize this case, we speculate that the presence of anomalies in soil moisture and Atlantic sea-surface temperatures could have shifted the ensemble towards warmer temperatures on the extended-range timescale. In the medium-range, capturing the Rossby-wave packet seems to have been the key to predicting the meridional flow that brought warm air northward over western Europe. However, short-range forecasts also suffered from an underestimation of the maximum temperatures.

2.3 Key factors of predictability

In this section we aim to discuss sources of predictability for a selection of primary (meteorological) events affecting Europe. Fig. 2.5 lists key factors for predicting a selection of extreme event types on extended-range (2–4 weeks; sometimes referred to subseasonal), medium-range (3–10 days), and short-range (0–3 days) timescales for several types of severe weather. This section is mainly an extraction from Magnusson (2019).

	Extended range (2-4 weeks)	Medium range (3-10 days)	Short range (0-3 days)	Example of ECMWF evaluations
European heatwave	Soil moisture, SST anomalies	Rossby wave packets	Local heating and evaporation	Newsletter (NL) 145, NL 157, NL 161
European cold spells	Teleconnections from MJO, SSW to negative NAO	Transition to negative NAO	Surface inversion, precipitation type	NL 151, NL 155, NL 159
North-western European windstorms	Predict positive NAO	Jet-stream propagation	Timing of development, wind gust parameterisation	NL 139 and NL 147
North-western European extreme rainfall	Predict positive NAO	Presence of atmospheric rivers	Exact position of system, strength of orographic precipitation	NL 163
Southern European extreme precipitation	Negative NAO	Development of cut-off low over Mediterranean	Convection, orographic precipitation	NL 146, NL 150, NL 155, NL 158, NL 166
Severe convection	-	Upper-level trough, position of fronts, CAPE and wind shear	Convective triggering, Organization, lifetime, wind gust parameterization	NL 141, NL 146 and NL 150

Figure 2.5 Summary table of key factors influencing predictions of different timescales. The final column references articles in ECMWF Newsletters. *Available from http://www.ecmwf.int.*

2.3.1 European heatwaves

On the extended-range timescale, the temperature anomalies are modulated both by local predictability drivers such as soil moisture (Ferranti and Viterbo, 2006; Fischer et al., 2007; Vitart and Robertson, 2018), and by the large-scale flow regimes. The former directly affects apportioning between sensible and latent heat fluxes; and in turn cloud cover can also be affected, which itself will provide a negative feedback (in summer) by reflecting back insolation. It is therefore important to correctly initialize the soil conditions in extended-range forecasts (Dirmeyer, 2011). For the large-scale flow, persistent blocks over Europe favor heatwaves. To set up such atmospheric flow conditions in the summer, among other factors, the sea-surface temperature (SST) in north-eastern Atlantic plays a role (Wulff et al., 2017).

To predict onsets of the large-scale flow patterns, medium-range predictability of Rossby Wave Packets (RWP, Wirth et al., 2018) are important (Fragkoulidis et al., 2018). On average the presence of long-lived RWP increases the predictability(Grazzini and Vitart, 2015). However, the prediction of the packet propagation is sometimes affected by uncertain elements such as organized convection and/or rapid cyclogenesis, resulting in bad forecasts of onesets of blocking patterns (Rodwell et al., 2013; Magnusson, 2017). There is also a difference in the processes (advection, subsidence, diabatic processes) leading to heatwaves in different parts of Europe as discussed in (Zschenderlein et al., 2019), and the influence of these processes need to be captured in the medium-range forecasts.

Although there is relatively high predictability for the large-scale flow, short-range forecasts often result in large 2-m temperature errors. During heat waves, the ECMWF model has problems to simulate the amplitude of the diurnal cycle. One reason could be that the model does not currently include urban tiles and hence misses the extra heating due to tarmac and concrete; these factors are discussed in, for example, Hogan et al. (2017).

Hazards that are related to heat waves include human heat stress, wildfires, and meteorological droughts.

2.3.2 European cold spells

Cold spells over Europe are often caused by large-scale flow patterns bringing cold air from north and east, such as Scandinavian blocking and negative North-Atlantic Oscillation (negative NAO, sometimes referred to as Greenland blocking) (Ferranti et al., 2018). These two regimes disrupt the westerly flow towards Europe and replace it with strong meridional flow and often anticyclonic conditions. Under anticyclones, strong surface inversions can form in calm and clear conditions leading to extremely cold wintertime temperatures.

On the extended-range timescale, these large-scale flow patterns have teleconnections from Madden-Julian Oscillation (MJO, Cassou, 2008) and/or sudden-stratospheric warmings (SSW, Baldwin and Dunkerton, 2001). If these precursors are predictable (Vitart, 2014), we should expect some predictability for the flow regimes if the teleconnections are sufficiently captured by the model. The predictive skill of the wintertime regimes and the conditional dependence of skill on MJO was recently presented in Ferranti et al. (2018).

The transition into a blocked flow regimes is often related to Rossby wave breaking (Woollings et al., 2008). The ability of extended-range forecasts to, in a climatological way, capture the link between Rossby wave breaking and formation of blockings has recently been evaluated in Quinting and Vitart (2019). The formation and maintenance of these regimes are also suggested to be linked to diabatic processes related to warm-conveyor belts (Wernli and Davies, 1997; Grams et al., 2011). These onsets are sensitive to small errors in the upstream flow and such a case is discussed in Grams et al. (2018).

Short-range forecasts during cold spells often experience very large (>10 degrees) 2-m temperature errors, and in most cases related to strong surface inversions (add references). The inversions are difficult to simulate in models due to insufficient vertical resolution in the boundary layer. The shallow nature of these inversions leads to large temperature errors for relative small errors in energy fluxes (Day et al., 2020). It could also be the case that physical parameterizations such as vertical diffusion are not perfectly suited for these extreme conditions.

In connection with cold spells, severe weather in the form of heavy snowfall/blizzards can occur. However, the meteorological features causing the extremes can be very different. As an example, proximity to seas or large lakes can give rise to intense, local snow showers on the coasts. With parameterized convection, it is difficult to capture the advection of such showers onto land, and often also the short-range forecasts miss the large amount of snow on the coasts.

Another uncertainty related to precipitation during cold conditions is to predict the precipitation type (Gascon et al., 2018). Here fine boundaries between rain, snow, sleet, and even freezing rain make a huge difference in the severity of the event, even if the precipitation rate is similar. An additional complexity is heavily populated coastal areas with high vulnerability, which magnifies the uncertainty of the impact due the precipitation type.

Hazards related to cold spells are (as mentioned above) blizzards, freezing rain, and human heat stress.

2.3.3 Northwestern European windstorms

While predicting the exact track and timing of Northern European windstorms on the extended-range timescale is impossible, forecasting the increased likelihood of the features is the target on this timescale, and predicting the NAO is key (Donat et al., 2010) as its positive phase favors cyclone tracks towards northwestern Europe. The positive phase of NAO has a statistical teleconnection from enhanced convection in the Indian ocean due to MJO (Cassou, 2008).

As in the case to make medium-range predictions of many other extreme weather types in the mid-latitudes, capturing RWP is also important for windstorms (Wirth and Eichhorn, 2014), in order to predict the risk of downstream developments that can form extreme cyclones. Explosive developments are often associated with upper level divergence by the jet stream (left-exit), and here the key is to capture the phasing of the jet stream and the lower level cyclones.

For shorter timescales, another difficulty for predicting extreme winds is to capture structures that cause wind maximum gusts, such as sting jets (Hewson and Neu, 2015) and embedded convection caused by dry intrusions (Raveh-Rubin and Catto, 2019). As global models still rely on parameterization of wind gusts, this naturally causes uncertainties. But the problem of capturing wind gusts also appears in convection-permitting models (Pantillon et al., 2018).

Apart from the direct damaging effect of the winds, the windstorms can cause high wave events on the sea and storm surges along coasts, and also cause flooding due to heavy rainfall (see below).

2.3.4 Precipitation extremes due to North-Atlantic cyclones

Connected to extra-tropical cyclones are so-called atmospheric rivers: bands of high mean transport moisture that can bring extreme rainfall when ascending over orography (Ralph and Dettinger, 2011). Lavers et al. (2017) showed that using water-vapor flux to trace atmospheric rivers is a good predictor of high precipitation events during positive NAO conditions during European winters. Therefore, the extended-range predictions of these rely on the same mechanism as the windstorms discussed above.

To capture the magnitude of the precipitation over orography, sufficient model resolution is needed together with accurate model microphysics to capture the timescale of the rain-formation (Forbes et al., 2015).

2.3.5 Precipitation extremes in southern Europe

Precipitation extremes in the northern Mediterranean are often connected to large-scale upper level troughs (Nuissier et al., 2011; Raveh-Rubin and Wernli, 2015; Mastrantonas et al., 2020) together with interaction with local orography. Statistically, the precipitation over the Mediterranean is negatively correlated with the NAO (e.g., Trigo et al., 2004; Vergni et al., 2016; Tsanis and Tapoglou, 2019).

Many evaluated cases show that a reasonable signal of extreme precipitation compared to the model climatology appears well into the medium-range, due to the prediction of the large-scale troughs. These large-scale precursors are often part of a Rossby wave packet (Martius et al., 2008) and have been shown to have good predictability (Grazzini, 2007). The extreme precipitation often appears on the eastern side of the trough, connected to strong moisture flux (atmospheric rivers). The extreme precipitation often occurs due to orographic enhancement when the moist air is forced to ascend (Khodayar et al., 2018). The presence of atmospheric rivers is largest in the autumn (Lavers and Villarini, 2013). The predictability of the extremes in the medium-range is dependent on the convective influence in the precipitation extremes, with lower predictability in summer time when the convective part is stronger (Grazzini et al., 2019). An open question is whether the troughs over the Mediterranean have a regime-like behavior and are predictable on the extended-range timescale.

In the short-range forecasts the details in the moisture flux, convective initialization and interactions with orography are important factors to predict the extremes (Gascon et al., 2016). These factors are still difficult for global models where the convection is parameterized and the orography is not sufficiently resolved.

The extreme precipitation can cause flooding events, both as flash floods in coastal mountainous areas and, especially if several events follow each other, more widespread river flooding.

2.3.6 Severe convection

Severe summer-time convection over Europe often results in intense rainfall, hailstorms, severe wind gusts, lightning, and on rare occasions tornadoes. However, on the extended-range timescale it is more difficult to find any key features that would give an early indication of these features. Instead in the medium-range the key feature is to find unstable air that is often ahead of cold fronts. To identify such features, the Extreme Index Forecast Index for CAPE has been found to be a useful approach (Lalaurette, 2003). As convective cells are favored by a vertical wind shear, another EFI index that is a combination of CAPE and the lower tropospheric vertical wind shear has been developed and tested (Tsonevsky, 2015). The atmospheric models cannot explicitly predict lightning, but promising results with a parameterization is discussed in Lopez (2016).

Capturing the true magnitude of wind gusts in connection to convective systems are challenging for the global models as they do not explicitly resolve convection and the associated downdrafts. Instead convective indices calculated from model quantities are still a useful tool such as the EFI products mentioned in the previous paragraph. Another challenge is the variability inside a grid-cell that can be large for precipitation in convective systems. One way to account for this effect is adopted and outlined in Pillosu and Hewson (2017). This method broadens the PDF to account for the variability not resolved by the model, and is dependent on the orography and weather situation (more subgrid scale variability due to mesoscale events compared to synoptic events).

The global models with parameterized convection also have problems to capture the timing of the convection and often miss the extension into the evenings.

Another risk during severe convection is flash floods (pluvial floods). Here the interaction between large-scale forcings, the moisture flux (atmospheric rivers), orography and convective triggering creates favorable conditions, but also the soil wetness before the event plays a role. Dependent on these factors the predictability of the extreme rainfall can be very different. It is also important here to have surface models that correctly model the surface runoff and local storage of water.

2.4 Hazard forecasting

In this section we will discuss how the atmospheric forecasts are used in the hazard forecasting process. We will discuss applications in flood, drought forecasting, heat stress, and wildfire forecasting.

Many hazards are caused by cascading events where for example a sequence of moderate precipitation events causes floods or a blocking system creates dry soil conditions leading to fires. Such cascades are rarely linear like a row of dominoes but more often intrinsically connected with several feedback loops. A blocking system may cause dry soil conditions which lead to a lower evaporation influencing temperature leading to a heatwave which has an impact on human health. This is even more complicated as the different physical processes act and occur on different spatial and temporal scales.

Forecasts are uncertain and such uncertainty is often expressed in ensemble members aiming to approximate the probability distribution of future events. In a connected earth system as described above the cascading hazards start with a probabilistic distribution of the drivers (i.e., precipitation, temperature, and soil moisture) leading to a probabilistic forecast of the hazard (i.e., floods, droughts, fire, and health) modulated through a number of low pass filters such as soil moisture or vegetation growth.

Hazards and disasters are rare and thus have low probability in forecasting. Such low probabilities are usually associated with the tails of climatological distributions. A long range forecast (if reliable) is very close to the climatological distribution and small shifts in the tail of such a forecast could lead to the forecast of a disaster—this is often more complicated as it will require the shift in tails of several variables or even of the same variable at different accumulation times. For example, a flood may be caused by precipitation on wet soil. The same precipitation on a dry soil may lead to the flood not occurring. Therefore not only the precipitation forecasted for a small window matters, but also the accumulated precipitation (in addition to evaporation) will need to be known to understand whether an event occurs or not. It follows that cascading probability distributions for extreme events may narrow or enlarge throughout such a cascade depending on the temporal and spatial scale relevant for the particular physical process and its interaction with the inputs and outputs in question. It is important to note that forecasting the tails of a climatological probability distribution is not equivalent to forecasting an extreme event itself. At shorter lead times the probability distribution of a skillful and reliable forecast will more close center around the extreme event values itself. This chapter investigates the forecasting of individual hazards and the relevance of different scale structures. It demonstrates how meteorological information and evolving probability distributions influence the forecast of natural hazards.

Forecast skill and performance plays a pivotal role in understanding and investigating evolving probability distributions. For example, one may have a "perfect" precipitation forecast, but the snow pack forecasting may be of significant less quality. Therefore, floods caused by rain on snow may have very low skill, which also means that the spatial and temporal relationship between precipitation falling as snow earlier in the season to the precipitation which triggers the flooding is incorrect and all findings relating to scale structures would be a poor guide for future model development. In the worst case, the flood could be still forecasted but for the wrong physical reasons. Thus to understand underlying predictability for extreme events requires the identification of the dominant and relevant processes which lead to this event. Such an example of a "failing" of the physical cascade can also be found with similar event types where predictability is strongly related to seasonal, geographical or other factors. In order to improve forecasts, all earth system models have tuned or calibrated parameters which have often disguised "physical names." This process is designed to ensure that the model is representative of the processes it represents, but can also unduly influence any analysis on scale structures.

2.4.1 Hydrological processes and predictability of flood and droughts

Runoff (or the amount of water at the surface of a catchment) results from the terrestrial water balance components, with gains (precipitation) and losses (from plants and water bodies through evaporation), and from water already stored within the catchment and released to the river.

Water contained in soil and floodplains stores generally reaches the river after a few hours to a few days, but other stores, such as groundwater aquifers, wetlands, lakes, and reservoirs, as well as snow pack and glaciers, release water much slowly, at timescales from weeks to month. Release speed also varies for a same storage family depending on many factors such as soil type, geology, and climate (for water stored at the surface as snow and ice).

Additionally, for long rivers and large catchments, water takes time to move from headwaters upstream to lowlands downstream, so that the movement of flood waves from one part to another of the catchment can be anticipated following heavy rainfall and rise in water levels upstream. This speed depends on catchment topography and morphology, catchment shape, and number of tributaries (Brutsaert, 2005).

Depending whether the catchments water stores are full or empty, additional water from the atmosphere or from upstream will further fill the stores, or go directly to the rivers. Knowing how much and where water is already in the catchment hence provides a very useful indication of future states of the rivers. This is what is called the Hydrological Initial Conditions. They are an important source of predictability for hydrology and river flow forecasting.

The other source of river flow predictability is of course the water balance components (gains and losses, the later depending on both atmospheric and vegetation)—their forecasting performance contributing to forecast skill of river flow at different time ranges.

However, the pace at which the catchment reacts to the atmospheric forcing is variable.

When water stores are full, rain over those areas cannot enter the stores (e.g., through infiltration) and reach the river channel very quickly resulting in a quick rise of river flow—the subcatchment has a fast response to rainfall events. This is typically the case for flood events, with a particular extreme case of flash-flood or pluvial flooding when surplus water does not even reach the river. Conversely, when water stores are empty, rain first replenishes them and does not enter the river channel through surface pathways, but instead, water is slowly released from the stores—the subcatchment has a slow response to rainfall events. This is typically the case for drought events.

Moreover, other water pathways exist, in particular for flood events, when surplus water can come from snow or glacier melt, groundwater aquifers or upstream flood waves. In this case, events are often slower to develop and as a consequence also last longer.

As a result, river flow of fast-responding events (i.e., floods) and/ or catchments tend to evolve primarily according to the atmospheric input, especially precipitation. This means river flow forecasts rely heavily on precipitation forecasts to predict both timing and magnitude of events, and require input at scales consistent with response time—typically subdaily to daily—and main catchment characteristics, typically kilometer.

Conversely river flow in slow-responding events (i.e., droughts) and/or catchments tends to evolve primarily at the storage process speed. This means that river flow forecasts rely on knowledge of both Hydrological Initial Conditions and atmospheric forecasts, with HIC importance stronger at the beginning of the forecast, and atmospheric forcing dominating a longer timescale, varying from days to weeks (Shukla and Lettenmaier, 2011).

The size and number of water stores and geographic location of a catchment and time of the year contribute to define whether river flow is likely to react quickly or not to precipitation input. Urban areas tend to reduce infiltration processes and hence generally are fast-responding.

Because IHC can be monitored or simulated from observation, and because hydrological models account for both atmospheric water balance and terrestrial hydrological processes within the catchments, forecasting river flow in slow-responding catchments or events gain additional predictability from IHC to that of atmospheric forcing when. Typically, drought events might gain predictability at a longer time range even in regions with low predictability of precipitation at extended or seasonal range.

2.4.2 Challenges

As seen above, the way catchments respond to atmospheric forcing depends very much on the landscape, soil, and vegetation, which vary at spatial scales ($\sim$1 km) much finer than that of Numerical Weather Prediction (Bartholmes and Todini, 2005): typically NWP have horizontal resolution larger by a factor between 2 and 5 for limited-area models, and 10 and 40 for global models, also depending on the forecasting range, with extended and seasonal range tending to have coarser horizontal resolution.

This means the need to downscale the information from NWP to match that of hydrological processes as much as possible, in practice to that of the hydrological models used for generating the river flow forecasts. This is particularly true to flood events, and especially flash floods which tend to occur on smaller, steep catchments, or in response to very localized heavy rainfall events.

Postprocessing precipitation forecasts to produce atmospheric forcing representative of the scale of hydrological models (often referred to as downscaling) is one of the key challenges of hydrological forecast, and especially flood forecasting which develops at much smaller scales than droughts; additionally, drought are responding to nonevents (lack of rainfall over an area) hence the discrepancy of scale between NWP and hydrological model in a way disappear.

Postprocessing downscaling techniques range from simple regridding to sophisticated bias-correction techniques (Liu et al., 2008), such as quantile mapping or more complex machine learning techniques. The critical challenge is to correctly predict the location of the rainfall event at the same resolution of the hydrological model. It is also to produce information consistent with the modeling chain, typically as ensemble forecasts and nor probabilistic fields.

In addition, the timing of precipitation events is critical for those fast-responding flood generation processes. This can be challenging for forecasts of lead time beyond the short-range, as such events are less predictable. As a result, forecasts of flood events are often reduced to the short- to medium-range lead time, with extended (subseasonal) to seasonal range forecasts only really possible for predicting likelihood of high flows anomalies or of drought events over a period.

In most parts of the world, rivers are not fully natural systems and are influenced by many human interactions: they include abstraction and discharge to satisfy the various water supply demands, managed man-made reservoirs, embankments and flood defences.

Those human interactions impact differently on flood and drought events: river abstraction have more impact on droughts as tend to exacerbate losses and more modestly floods especially if those are generated by heavy rainfall events which go to the river channel quickly; opposingly, flood defences are by design aiming to minimize floods by retaining as much possible water in the river channel or storing it in flood plains; however, for hydrological models which only model flood flows and not water levels, their impact are minimal.

Reservoirs can impact both floods and droughts, as they act as additional water storage and effectively stop water from upstream to downstream of rivers: for droughts, they can be used as additional water sources for water abstraction, or to compensate for lack of rainfall by releasing pre-agreed water amounts downstream to minimize drought impact (also called environmental flows); during heavy rainfall events, they can be used to temporary store the water and stop the flood wave (e.g., small retention areas on hillslopes or dams constructed along the river). The way a reservoir impacts on floods, droughts or both depends on its size, location and design, as well as its management. Often, reservoirs are not built for the sole purpose of drought and flood mitigation but for other functions, such as hydropower or irrigation. In this case, their use for pure hydrological purpose might conflict with their other uses, requiring complex management and negotiation.

In reality, there is only limited information available on human interactions on river flow, and they are not always included in hydrological models. The most sophisticated models aim to include functions such as abstraction and reservoirs, but they are often associated with simple theoretical rules and thresholds because of lack of detailed and observation-based, information. Artificial structures in rivers are most often neglected especially for distributed models covering large areas, such as continental or global models. In many parts of the world, however, even the location of potential abstraction sources and reservoirs can be not known, and those are simply not accounted for in models.

An additional challenge to modeling the human influence is to also forecast those activities. For irrigation-led abstraction, ideally planting cycle and vegetation needs should be modified according to forecasted atmospheric conditions. In practice this is rarely done and climatological functions are used in forecast mode, ignoring the potential to adapt those rules if exceptionally dry/ hot or wet/cold conditions are forecasted.

Finally, one of the biggest barriers in including human processes in hydrological models is the lack of data available for understanding the processes and building the models. This is especially true when building large-scale models such as global hydrological forecasting systems. In this case, the "natural processes" only are simulated, and forecasts are expressed as differences with extreme events simulated also as natural processes. In this case, the simulated hydrological forecasts are not compared with observations but with model-defined thresholds, aiming to predict the occurrence of a hazard (drought or flood event). In addition to having consistency in the modeled processes, this has also the advantage of minimizing the potential biases in the modeled response compared with real behavior, not only due to absence of human influence processes, but also not appropriate scale or too simplified processes.

2.4.2.1 Type of hydrological, floods and drought forecasting, models

Two main families of hydrological models are being used for forecasting: process-based, which transform atmospheric forcings into river discharge variable by describing the terrestrial processes dominating the generation of flood and droughts through a set of physically plausible equations; and statistical models, which use data and machine learning techniques to produce a set of predefined outputs (e.g., flood or drought event) according to a set of forcings.

Process-based models are generally based on simplified laws of physics, or a set of conceptual processes not necessarily directly measurable (Clark et al., 2017). In their most extreme forms, physically-based models assume that physical parameters of the model are equivalent to measured quantities. When coupled with NWP, they can describe the full water cycle and are often called earth system models. In such cases, they often do not include lateral water transfers at the surface and subsurfaces, and hence need to be coupled with routing models that move water along the river channels to produce river flow, which themselves can then be categorized as flood or drought events. The more processes are included, the more complex the models and the more computational power is required to undertake simulations. For a long time, this has restricted their applications to weather centers, especially when run in coupled mode. When applied off-line, their parameters can be tuned to optimize the river discharge simulations but in practice this is seldom done. The simpler so-called conceptual models are very common for catchment hydrology and require data to optimize the values of their parameters by minimizing errors between observation and simulations. They generally only simulate a few processes with their main output being river discharge, and are extremely data and computationally sparse compared with earth system models. Whilst originally applied to small catchments assumed to have a single rainfall-runoff transformation, they have since been applied successfully larger scales including global, where the spatial domain is divided in smaller units connected through river channels (also called distributed models) and parameters can be attributed a set of default parameters outside catchments with observations based on detailed information on the catchment properties (including soil, geology, vegetation). The choice of the type of process-based model used in a flood or drought forecasting system depends on its aim (forecasts only at points at already identified river gauges or along all river channels), the available data for its training and verification and the computation power available.

Statistical models rely on observational data to define the best set of equations able to reproduce from the forcing data hydrological variables (Guo et al., 2014). They generally can only be applied to locations with data and can require many hundreds of runs to converge to an optimal solution, using machine learning algorithms for example. With the increase of computation power and availability of big datasets, they are now considered with more and more by the research community as alternative to process-based models for hydrological forecasting. Due to their data requirement, direct flood or drought forecasts based on statistical models has so far been limited, but statistic techniques have been used successfully to postprocess hydrological forecasts, correcting for biases in simulations based on near real-time observations in order to produce more realistic forecasts.

2.4.2.2 Improving usefulness of flood and drought forecasting systems

There are a number of challenges in hydrological forecasting (e.g., Cloke and Pappenberger, 2009 for flood). Here we review some key aspects that will impact on the forecasting performance.

2.4.2.3 Hazard thresholds

When forecasting extreme events such as floods and droughts, it is important to agree on the definition of the events to be able to highlight them in the forecast, also referred to the hazard threshold. The same definition should also be used when verifying the forecasts, which can be a challenge when different event definitions are used for putting in place the observatory records.

Generally, the hazard thresholds are defined according to a predefined values, for example associated with a likelihood of occurrence or a regulatory framework (frequency/ return period), a danger level (river flow associated with critical water level) or a known impact (environmental flow now sustained). Theoretically, hazard levels could also be defined to optimize forecast skill (e.g., Sivakumar, 2005), although this is seldom done in hydrology due to lack of observations and often physical levels which can be used to define warnings.

Most often, hazard thresholds are defined according to the climatological behavior of the system, either as absolute values (typically for flood events) or as anomalies (more typical for drought events), and generally from the proxy-observation simulations considered to be follow the same statistical distribution as the forecasts. Anomalies are usually favored to tackle issues of heterogeneity across large spatial domains with different climatology, but could be misleading (e.g., a wetter episode during a dry season might not represent a severe flood event). However, with increase of lead time and evidence of possible drift in the forecasts, such simplification might not be appropriate, and lead time-dependent thresholds based on reforecasts sets are preferable.

Finally, whilst deterministic thresholds are currently used to identify hazard events, it might be conceivable to introduce more complex, probabilistic thresholds especially when applied to ensemble forecasts and defined according to ensemble reforecasts. When applied to ensemble forecasts, hazard thresholds result in defining the likelihood of an event to occur as the percentage of the forecast ensemble to cross that threshold. This information is then passed on to users as an awareness product. This is different from the warnings issued only by Met Services and authorized organizations.

2.4.2.4 Impact forecasting

Increasingly, users ask for hazard forecasts to be complemented by impact forecasts so they can better prioritize their actions: this is the expected consequence the forecasted hazard could have. Impacts are generally defined as categories of risk and level of impact expected.

Typically, impact forecasts are made by comparing the hazard forecast (i.e., the likelihood of an event of a certain severity to occur at a certain place) with information on exposure (i.e., the elements that could be affected by the hazard in that location). For floods, often relevant exposure will depend on the hazard but include population density and vulnerability (e.g schools, nursing homes, etc.), infrastructure (e.g., critical building, road, and rail networks, electric grid), and land use (e.g., agricultural assets, urban areas). Using hazard and exposure categories, risk levels are defined according to pre-agreed thresholds on both hazard and exposure dimensions. For hazards, the severity thresholds can be defined according to the expected timing of the event (next few hours, next few days, new few weeks), its magnitude and the likelihood of occurrence. For exposure, thresholds can be linked to expected human and economical damage from the type of hazard, and existing levels of protection/response.

2.4.2.5 Seamless forecasting

Traditionally, hazard forecasts systems are designed for a given forecast horizon in line with atmospheric forcing forecasts: short-range up to 3 days, medium range up to 1−2 weeks, extended range up to 1−2 months, and seasonal range beyond. The same system can however provide multiple range forecasts in the form, the most often, of one or several products for each of the forecasting ranges. This is important when catering for different decision making processes such as emergency response (hours to 1 or 2 days), emergency planning ($\sim$1 week), and mitigation planning ($>$1 week).

The reason why different products are produced for different forecast ranges is because generally, atmospheric forecasts for different forecast ranges (especially when moving from medium to subseasonal and seasonal range) are produced independently from different models and forecast chains, and often have different horizontal resolution and/or number of members. This means that when a hazard forecasting system provides products for different forecast horizons, those forecasts might not be consistent with each other, which might confuse users and limit uptake.

To address this issue, a seamless forecast has been identified as a natural progression of hazard forecasting. They consist of a unique set of ensemble forecasts providing hazard forecasts at a wide forecast horizon, from nowcasting to subseasonal and even seasonal range. They are created by putting together (or "blending") the atmospheric forecasts with different forecast horizons to create a unique ensemble of atmospheric input over a long forecasting range. Those forcings are then run through the hazard modeling chain to produce a blended hazard forecasting, consistent at all forecast horizons.

There are many challenges in creating seamless forecasts and merging independent ensemble forecasts in a physically plausible way. Horizontal resolution merging can be easily achieved by using downscaling techniques and creating new forecasts all at the finest (hazard) resolution. Merging in time is more challenging. The simplest way is to randomly pair one member from each of the two ensembles, with the seamless forecast created by continuing to the shortest range forecast member series with data from the longest range forecast member. Because of the potential for substantial increase of computing time, generally the final number of ensemble members is kept to one of the two native ensembles. To avoid discontinuities both in space and time, especially for rainfall fields which are important drivers of floods and droughts, more complex pattern matching techniques can be used aiming to pair members with the most similar patterns (typically rainfall cells).

Finally, NWP forecasts produced for different forecast horizons can have different temporal resolution, potentially resulting in seamless forecasts with varying time steps across the forecast range, which can create technical difficulties when running the system and archiving the data.

2.4.3 Fire risk

Fire in the earth system is, simultaneously, a necessary ecological process, a useful tool, a destructive force, and a major source of pollution. Fire modifies the land surface albedo directly through burn scarring and changes in vegetation, and indirectly by depositing black carbon on snow and ice. Smoke can have cooling or warming effects depending on the intensity of the fire, the height at which the smoke is injected, and microphysical interactions with clouds. They are one important component of Earth systems but forecasting wildfires is a complex task as they depend on a completely unpredictable component: the way they start or "the ignition." Fires can start naturally, triggered for example by lightning or through self-combustion (caused by natural heat-generating processes). However, ignition can also be due to human behavior: through intentional acts for forest management (or arson) or simply through negligence. Quite commonly fires are used to encourage regeneration and biodiversity in the forest ecosystem or to replace forest vegetation with agricultural crops. Once a fire is ignited, its spread, sustainability, and difficulty of control is almost exclusively determined by weather conditions (Flannigan et al., 2009). Flames tend to rage out of control if certain soil and atmospheric conditions are met, for example as a consequence of extended drought or favorable wind conditions. Therefore, the prediction of this hazard is like no other as it is inherently stochastic. In many cases, no fire events occur, even when the forecast accurately predicts that conditions are highly favorable. Clearly this peculiarity of fire prediction poses big challenges to the interpretation of model output and the verification of model accuracy.

2.4.3.1 Forecasting fire at different spatial and temporal scales

Fire occurs over a wide range of temporal and spatial scales, from local to global, via many complex, interdependent, and poorly understood processes. Fig. 2.6 summarizes fire modeling as a function of spatial and temporal scales with the three triangles highlighting how different modeling approaches are used depending on how fire is looked at (i.e., as a micro process, a regional or global scale phenomenon) and on which timescales (from hours to decades). Depending on the temporal and spatial scales they focus on, these models all require different inputs and produce only a limited set of information (e.g., heat released, fire behaviors and fire regimes). At very small spatial and temporal scales, the combustion process can be accurately described by a physical model, if fuel and oxygen are

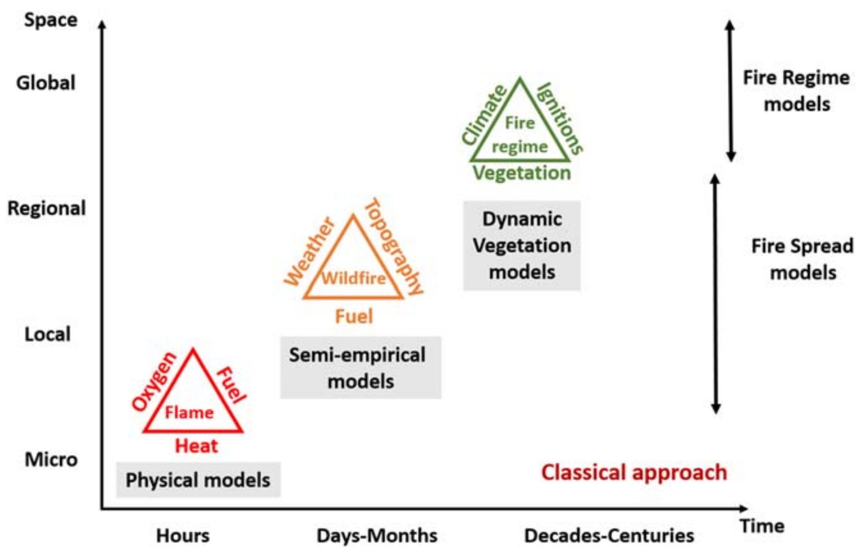

Figure 2.6 State-of-the-art approaches to model fires at different spatial and temporal scales.

provided as input variables. Models of physical combustions can be used to understand the amount and type of particulate produced and how much heat is released during the burning. These models are at the foundation of the description of the burning process, however they are of limited use in an operational context when information such as, fire spread, fire sustainability and control efforts are the main information required. Thus, at the regional to global scale, the aim is to model the relationship between weather and fuel availability as this relationship is the one that ultimately controls fire behavior and its sustainability on a larger scale.

The difficulties of accessing global real-time observations of fuel amount, type and moisture content have led most of the semiempirical models used in fire danger management to only consider observed weather conditions. Due to the improved skills of weather forecasting, the use of numerical weather prediction has also offered in the last years a real opportunity to enhance early warning capabilities (Roads et al., 2005; Mölders, 2008, 2010). Institutions such as Natural Resources Canada (NRC) and the US National Oceanic and Atmospheric Administration (NOAA) have implemented regional fire danger forecasting systems based on their operational weather forecasts (Bedia et al., 2018). The Global Fire Early Warning System is also an international initiative, promoted by the Canadian Partnership for Wildland Fire Science and the United Nation Office for Disaster Risk Reduction, to provide fire danger forecast up to 10 days ahead using the Canadian operational weather forecasting system (http://canadawildfire.ualberta.ca/gfews). Parallel initiatives are promoted by the European Commission under the umbrella of the Copernicus Emergency Management Service (CEMS), namely the European Fire Forecast Information System (EFFIS, http://effis.jrc.ec.europa.eu/) and its global counterpart the Global Wildfire Information System(GWIS, http://gwis.jrc.ec.europa.eu/) (Di Giuseppe et al., 2016, 2020).

2.4.4 Heat stress

Climate, weather and health are closely related. How sensitive the human body is to the outdoor environment is part of everybody's experience. Wind may cool down the human body by taking away its heat whereas sunlight and humidity may warm it up by radiation exposure and limiting sweating, respectively. In each of these situations the human body is continuously challenged to maintain its internal temperature within the range of optimal physiological performance, comfort and health (McGregor and Vanos, 2018). Usually efficient even when the surrounding temperature is very different, the human body's regulatory mechanism may fail when exposed to extreme environmental conditions causing disorders and illnesses (Cheshire, 2016). Intense and prolonged periods of abnormal heat, for instance, induce a thermo-physiological stress on the human body that may result in increased hospital admissions and deaths particularly among the elderlies. Similarly, exposure to cold extremes can depress the immune system and increase susceptibility to flu and colds.

Knowing how *thermal stress* from extremes such as heat waves and cold spells impact human wellbeing is essential to make weather forecasts, usually based on pure meteorological variables such as air temperature, meaningful from a public health perspective. In recent decades indices that assess thermal stress from the energy−thermal exchange between multiple environmental factors and the human body have been proposed. One of these indices is the Universal Thermal Climate Index (UTCI). The UTCI is a state-of-the-art indicator representing the concurrent effect of air temperature, wind speed, water vapor pressure, and short- and long-wave radiant fluxes on human physiology (Jendritzky et al., 2012). The UTCI is expressed on a stress category scale ranging from extreme cold stress to extreme heat stress where each category corresponds to a well-defined set of human physiological responses to the outdoor environment (Błażejczyk et al., 2013).

2.4.4.1 Hazard forecasting

The UTCI is based on an advanced energy-balance model called the UTCI-Fiala model (Fiala et al., 2012). The UTCI-Fiala model simulates the physiological response of an individual to an outdoor environment by coupling the human body's thermoregulation system with the insulation effect from temperature-dependent clothing. The model describes the outdoor environment in terms of four meteorological variables—air temperature, wind speed, relative humidity, and mean radiant temperature (MRT) with the latter being the total radiation from the atmosphere and the ground incident on an individual. The model takes the four variables as input and generates the UTCI which represents in one number (a feel-like temperature expressed in °C) the physiological responses to air temperature, humidity, ventilation and MRT. UTCI forecasts are obtained with a similar procedure (Di Napoli et al., 2021). Forecast outputs of 2-m air temperature, 2-m dew point temperature, solar and thermal radiation from ECMWF IFS model are first retrieved to compute forecasts of relative humidity and MRT, respectively (Di Napoli et al., 2020). The latter are then passed to the UTCI-Fiala model, summarized by a six-order polynomial equation in the four meteorological variables (Brode et al., 2012), for the computation of UTCI forecasts. UTCI outputs are generated at the same time as ECMWF's extended-range forecasts up to 46 days ahead and cover all the globe (land and sea) north of 60°S latitude. Example of an UTCI-forecast is given in Fig. 2.7.

Predicting the UTCI poses two main challenges. First, the uncertainty of UTCI forecasts is inherently linked to the uncertainty associated with the forecasts of its input parameters. The predictability skill of wind speed, for instance, is usually lower than the predictability skill of air temperature. This could negatively affect the overall performance of UTCI forecasts (Pappenberger et al., 2015b). Statistical postprocessing may correct systematic errors in the UTCI input parameters as demonstrated, for instance, for the forecasts of other heat-related indices (Baran et al., 2020). Second, UTCI forecasts have the same spatial resolution as ECMWF IFS forecasts, that is, 9km-by-9km grid for HRES. While such resolution is useful to capture atmospheric circulation patterns linked to thermal extremes, it may be too coarse for assessing the impacts of thermal stress on human health. The excess mortality and morbidity associated with heat and cold extremes are usually observed at urban scales. Downscaling UTCI forecasts at subgrid level may help better represent this (Leroyer et al., 2018).

As for verification, the first challenge is to verify UTCI forecasts against corresponding reanalysis data, for example, ERA5-HEAT, where the latter are used as proxy-observations (Di Napoli et al., 2021). Another

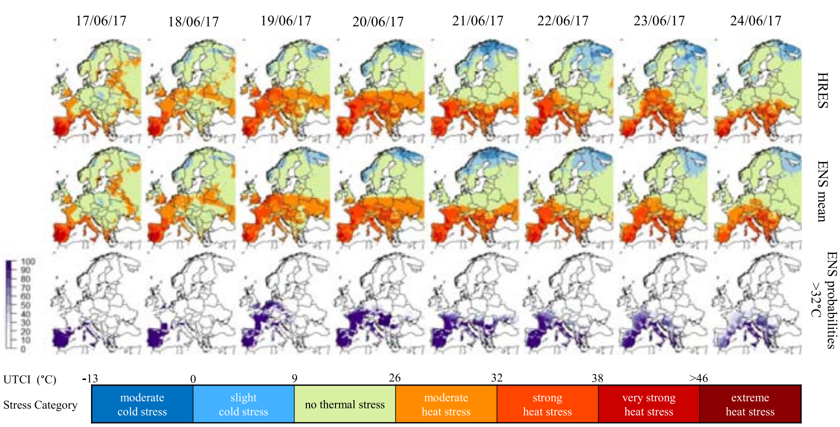

Figure 2.7 UTCI forecasts of the June 2017 heatwave as issued on June 17, 2017.

challenge is to verify the health impacts associated with the thermal extremes predicted by UTCI forecasts. For this to be possible the relationship between the exposure to heat/cold stress and the corresponding health outcomes, that is, deaths or hospital admissions, must be determined. The greatest limitation of this approach lies in the availability of public health data as these are usually not publicly available.

Where available, however, public health data can be matched with climatological data and lead to the definition of a threshold so that, whenever exceeded, an increase of deaths is observed. For instance, the 95th climatological percentile of the UTCI has been demonstrated meaningful for heat health impacts in France as periods of excess mortality corresponds to daily UTCI minima and maxima that are equal to or above that threshold for 3 consecutive days (Di Napoli et al., 2019). In Europe most warning systems for heat hazards use a threshold based on the relationship between mortality records and air temperature to trigger precautionary public health action plans; only a few systems consider other relevant meteorological variables, such as humidity (Casanueva et al., 2019). Although warnings are the responsibility of national health and meteorological services, UTCI forecasts can provide a first indication of detrimental thermal conditions via its stress category scale which was defined to be valid in all climates and seasons. This information could be then combined with how anomalous (e.g., with respect to climatology) the predicted thermal stress is. Heat levels well above seasonal average expose affected populations to a heat stress higher than the one to which they are adapted resulting in extra deaths, as observed in the 2003 European heatwave (Di Napoli et al., 2018).

2.4.4.2 Discussion

Meteorological extremes characterized by thermal (hot and cold) stress represents a serious hazard for human health. The development of indices such as the UTCI linking the exposure of an individual to the outdoor environment with the body's physiological responses has made it possible to predict potentially detrimental conditions using the forecasts of four weather parameters—air temperature, wind speed, relative humidity, and mean radiant temperature.

Defined via a stress category scale meant to be valid in all climates and seasons, UTCI forecasts have the potential to predict thermal extremes across the globe. UTCI forecasts may be improved from several points of view. Statistical postprocessing, for instance, may help reduce forecast uncertainties arising from variables like wind speed that in general show low predictability skill (Di Napoli et al., 2021). As the UTCI is a thermophysiologically defined index, impact-based verification of UTCI forecasts requires comparison against information that is meaningful for human health. This could be mortality records or hospital admissions which, however, are usually not publicly available. Building an open-shared data framework between the meteorological and the epidemiological community is a step that needs to be taken in order to make weather forecasts health-meaningful tools. Outdoor questionnaire surveys involving the general public could also help by assessing whether the thermal stress as predicted by UTCI forecasts corresponds to the thermal stress actually perceived (Lau and Krüger, 2020).

UTCI forecasts may also be combined with forecasts for other hazards potentially lethal for human health. The joint prediction of heat stress and wildfires, for instance, has been proved valuable in disaster risk reduction and emergency response management during a heatwave (Vitolo et al., 2019). UTCI forecasts could also be targeted for areas, called *hotspots*, where historical information from climate reanalysis has revealed the simultaneous or cascading occurrence of heat extremes with respect to wildfires or droughts (Sutanto et al., 2020). Another potential application for UTCI forecasts would be in the definition of a "biocomfort" threshold where a person will potentially experience not only heat stress but also respiratory problems due to air pollution or pollen allergy symptoms (Jacobs et al., 2014).

2.5 Evaluation of hazardous events

A proper evaluation of forecasts serves two purposes: inform the users about the skill of the forecast product and guide the development of the forecasting system. These two purposes require in many cases different evaluation approaches and different types of skill scores. For example, while the value of the decision is the key for the user, the system developer needs to know the quality of all components in the forecasting chain in order to find the weakest links.

Severe events are by definition rare which results in a small sample for verification, especially if we narrow the statistics to a short period (e.g., a season) and/or a region (e.g., Europe). When performing the evaluation of a specific type of severe event, one has to be aware of the trade-off between extending the sample to other seasons/continents and the risk of verifying extremes due to other types of meteorological circumstances.

It is therefore difficult to reach a large enough sample to obtain reliable statistics in verification of extremes. Another complication for statistical verification is that the meteorological conditions leading to hazardous events often are unique in their composition. Examples here are the length of a heatwave leading to hazards in forms of human heat stress and wildfires, the combination of winds and rainfall leading to coastal flooding and a sequence of rainfall events leading to river flooding. Taking such compositions into account further reduces the verification sample.

2.5.1 Observations for evaluation

Finding suitable observations for evaluation of both the meteorological conditions and the hazards is a challenging task. For the evaluation of the meteorological conditions, the spatial scale is often a problem. As the spatial scale of the extreme can be very small, for example in the case of extreme precipitation, a high density of observations is often required. On top of network density is the technical challenge that the instruments need to stand the extreme conditions, and often stations are physically broken during extremes. From an evaluation perspective this could skew the verification. Another technical challenge is if the instrument is designed to measure the value range during the extremes. Examples here are rain gauges that get large errors during high intensities, and remote sensing products like scatterometer winds that get large errors for extreme winds. On top of these challenges, is that the scale captured by the atmospheric model (usually an average over several square km) is different from the scale observed (often point observations). To make a fair evaluation, a postprocessing of the one of the qualities (downscaling of model results or upscaling of observations) is needed (see Pillosu and Hewson, 2017; Ben-Bouallegue, 2020).

For hydrological forecasts, one of the main challenges in verification results from a discontinuity in space: rivers are only on a small fraction of the domain, and even within a same river network, can see large discontinuities in values typically when two rivers merge.

This means that observations typically cannot be interpolated to a field against which verifying the forecasts. Instead, to verify a river forecast, one would ideally need to have observations along all the river network. Even in countries or regions with very dense observation networks, this is not the case, and river gauges can be located hundreds of kilometers apart, with many rivers (or catchments) not gauged at all. Only occasionally gauges are located both upstream and downstream of a confluence, making it difficult to really verify the contributing proportion of the different streams. Even more challenging is to verify events such as floods or droughts, by definition extremes hence not occurring often, and with only limited and incomplete observation.

To verifying the impact of the hazards, news reports and/or event databases can be used. However, event observations are often inconsistent in terms of event definition, as they measure the impact and not directly the magnitude of the physical properties, e.g river flow magnitude. Moreover, different types of events are recorded, not only floods from rivers (typically modeled by hydrological forecasting systems) but also pluvial flooding not entering the river channel (and by definition not modeled by traditional hydrological forecasting systems) or localized flash floods developing at a scale finer than that of the modeling system. This means that the observed events could in fact be very heterogeneous and not consistent with the forecasting systems, and not be a fair comparison dataset against which to verify the forecasts. Moreover, observation-based events are biased according to the network of measurements they come from, whether from heterogeneous river gauges network and human observations. The latter can typically be biased by population density (more events reported in areas of larger population and not necessarily reflecting the true spatial distribution of the event), communication challenges (e.g., through language or technological barriers), timing of the event (e.g., at night) or cultural and societal factors. As a result, false negatives (i.e., no event is recorded whilst one event has occurred) are extremely common in such datasets, increasing the difficulty in defining appropriate verification procedures and metrics. The length and consistency in time of observational records, especially when relying from volunteers, is another challenge, with hydrological volunteering networks being in their infancy compared with well-defined meteorological volunteer networks. Generally, only a few years of data are available, with records obtained from a varying number of sources, hence typically non stationary.

Observing the impact of health hazards poses a variety of challenges. One challenge is the quantification of the impacts via morbidity (e.g., hospital admissions or emergency service activities) or mortality data. Mortality is usually preferred as death records are regularly collected and standardized whereas morbidity data depend on guidelines set by national health systems and may therefore vary from country to country. An important issue with mortality is that the number of deaths that according to the official definition are identified as "heat-related (resp. cold-related) deaths," that is, due directly or significantly to heat (resp. cold), underestimates the actual impact of the hazard. As heat and cold extremes are known to have a potential contributing role also toward cardiovascular and respiratory deaths, these must be considered when assessing the health impacts of thermal stress-related hazards. Mortality and morbidity records are in general collected at high spatial and temporal resolution in national databases which, however, are usually not publicly available. Another challenge is represented by the fact that health impacts are due not only to the exposure to the hazard but also to the vulnerability of the affected population, namely its socioeconomic status and health-care system.

Verifying the fire risk against recorded events poses an additional challenge, as indices are not a physical measure of fire activity but of its potential danger if one were ignited. Therefore, high fire danger, while being correctly forecasted, might not result in active fires if there is no ignition and/or aggressive fire suppression. From the verification point of view, this means that the identification of false alarms is not meaningful, and the verification should mainly rely on hits and misses. Secondly, fires are rare events and, as for any other infrequent phenomena, the verification statistics are heavily influenced by the small number of hits when compared to the total.

As an alternative to use direct observations, simulations obtained by forcing the hazard model (flood, fire, health etc.) with observed atmospheric forcing are generally used as "proxy-observations" to conduct verifications. This has two main advantages: by definition, the simulations are available everywhere without any gap, providing the only opportunity to verify the results over the whole domain. This is most important in forecasting systems covering a large geographical domain such as continental-scale or global. Because the same model is used to create both the proxy-observations and the forecasts, resulting simulations will contain the same systematic biases for example due to missing or simplified processes; in other words, the proxy simulation provides comparable simulations to verify against, which would not necessarily be the case with observations. The main assumption here is that it is not the ability of the modeling system to reproduce the hazardous processes but the forecasting performance. However, it is clear that processes such as delay of the timing of the event, for example, possible through reservoir management will not be captured if they are not included in the model, which might artificially increase the skill score when using proxy data.

2.5.2 Evaluation metrics

An important part of a statistical evaluation process is the choice of verification scores. Verification scores can either evaluate a continuous measure or be based on a discrete event. For verification of extremes, event based scores are often used. These scores are based on counting hits, misses, false alarms for a specified event. However, many traditional verification scores are not well suited for extreme events, as the occurrence of nonevent by definition largely dominates the sample (by an order of magnitude between 10 and 50 or more), depending on the rarity of the event tested. As a result, the scores become difficult to interpret or even misleading. The choice of score for extreme events is, for example, discussed in Stephenson et al. (2008), Ghelli and Primo (2009), and Ben-Bouallegue et al. (2019).

The ultimate question is whether the forecast contains enough information for the user to take preventive actions to reduce the risk of losses due to extreme weather. One simplistic way to answer this question is to evaluate the Potential Economical Values (Richardson, 2000). Albeit building on a simple model of the cost for actions and potential losses, this type of verification can indicate the type of actions for which forecasts are useful. In Magnusson et al. (2014) forecasts for moderate extremes (98th percentile) were verified using this metric, and it was found that only action with a relatively low cost compared to the prevented loss is worth taking based on medium-range forecasts. However, in reality, the preventive actions associated with expectations of extreme weather a week ahead is about preparations and redistribution of resources, and these actions are relatively cheap. It is rather in the day(s) just before the event, the relatively expensive actions need to be taken. With a lead time dependency of the cost/loss ratio, it could well be that the forecasts actually are as (or even more) useful in the medium-range than in the short-range and make multiscale prediction of extreme weather important. Also, forecast quality does not always equal forecast value (Richardson, 2000). A forecast has high quality if it predicts the observed conditions well according to some objective or subjective criteria. It has value if it helps the user to make a better decision in terms of protective actions (Cloke et al., 2017).

One interpretation of the Potential Economical Value model is when the cost of a missed event is very high, for example, in terms of human lives, the deliberate over-forecasting may be justified (Richardson, 2000; Cloke et al., 2017). This means that the decision should be calibrated on the decision level and not on the physical model. This could lead to mis-interpretations of the verification results when developing the physical model, and it is important to keep the evaluations separate.

A common ingredient in many of the evaluation metrics is a benchmark, from which the skill is calculated to be relative to. For example, in the Potential Economical Value discussed above, the value is determined relative to decisions made based on climatology. But other benchmarks could be based on persisting the conditions at the forecast initialization, or based on predictions from a more simplified forecasting system (Pappenberger et al., 2015a). While climatology is preferred as a benchmark to measure the overall level of skill, persistence or a simplified model is preferable to determine the gain in skill by the forecast system. However, a special attention is needed to determine the climatology regarding type of input data (observations or reanalysis), length of sample and seasonality, otherwise the skill estimation can be misleading especially for verification of extremes.

2.6 Conclusion

With all challenges listed above, it is attractive to study cases individually. However, strictly speaking a probabilistic forecast for a single extreme event cannot be verified. This is because there is no such quantity as "true probability distribution" but only the outcome of the event. There is also a risk that one only focuses on the events that appeared, which would skew the evaluation to "hits" and "misses," and ignore "false alarms."

Instead one needs to combine several approaches. In order to get reliable statistics for the verification one needs to lower the threshold for the event and focus on simple (noncompounded) events for which observations exist and are distributed, such as 24-hour precipitation, 10-m mean wind and 2-m temperature. Such verification of extreme weather for ECMWF forecasts was undertaken in Magnusson et al. (2014) and Ben-Bouallegue et al. (2019). The second approach is to evaluate the model climatology for extreme events to verify if the model is able to produce the extremes with the same frequency as in reality (Magnusson et al., 2014). The third approach is to study individual cases, and from such work one can identify important aspects to verify further (Magnusson, 2019), which we have given examples of in Section 3.3.

2.7 Summary

In this chapter we have discussed the prediction of severe weather and related hazards across timescales, and the components involved in a forecasting system. The extended-range predictability of weather extremes comes from long-lived flow patterns in the mid-latitudes together with teleconnections from tropics/stratosphere and boundary conditions such as SST and soil moisture. For medium-range forecasts it is important to capture the large-scale evolution of the atmospheric flow, which requires good global data assimilation. For short-range forecasts the local data assimilation is important as well as physical parameterizations in the model that are suitable for extreme conditions. However, one also needs to keep in mind that uncertainties even in the shortest forecasts cannot be eliminated due to insufficient knowledge about the current state, some physical processes and due to variability inside the grid-box. The barriers of predictability for any of the hazard and impact models in this chapter are more complex and forecast quality on timescales is more intertwined with regional and local properties, thus a more general attribution of the dominant factors influencing predictability at each timescale is more complex. For example, at the long range, the predictability of La Nina or El Nino cannot be directly associated with the predictability of floods or droughts (Emerton et al., 2017). Therefore, process-based understanding of hydrological variability and causality at all space and timescales is still a major challenge for hydrological forecasts (Blöschl et al., 2019), fire and health models are largely challenged in adequately representing human interactions.

Predicting extreme weather several days in advance is clearly a probabilistic problem. The overarching target in ensemble forecasting is to issue as narrow (sharp) PDF as possible whilst maximizing the reliability and keep a desirable consistency in the PDF.

The fundamental question is how to improve the prediction of extreme events on all timescales and where to put the resources in terms of research and computer power. To obtain a good reliability by minimizing frequency biases, one need to simulate the event with the right climatological frequency, that is, the PDF of the model climate needs to be close to the PDF of the true climatology, including the tails. It is common that the magnitude of the simulated extremes are limited by the model resolution, and increased resolution has in the past improved such biases. Here limited-area models play a role to better resolve the extreme; as well as postprocessing techniques/AI to adjust the forecast PDF. However, the frequency bias can also be associated with deficiencies in the model physics connected to the extremes (e.g wind gust parameterization and orographic precipitation), and here improved model physics can help to improve the simulation of the extremes.

To increase the sharpness of the PDF without losing reliability, one needs to decrease the forecast error to allow a more confident ensemble. To do this a key is to reduce the analysis error by improving the components involved in the data assimilation, such as observation usage, modeling of background error statistics, minimization algorithms and also the model used for the first guess forecast. To obtain a reliable ensemble, an accurate simulation of the initial uncertainties is needed as well as simulating the model uncertainties. As predicting extremes in the medium-range is often dependent on resolving the extreme tail of the PDF, a larger ensemble is needed compared to only focusing on the ensemble mean.

Finally, to capture signals from boundary conditions (soil, sea-temperature etc.) on the extended-range timescale we need to include all relevant earth system modeling components. We also need to make sure that the model is capable of simulating the teleconnections from the sources of predictability.

All these points are associated with resources in terms of research and operational constraints (i.e., ensemble configurations and computer power), and the operational forecasting centers need to find a good balance to progress. By evaluating a range of extreme weather events, the current bottlenecks for improving the forecasts can be identified. Current challenges in different parts of prediction systems for high-impact weather has recently been outlined in Majumdar et al. (2021).

The evaluation of the cases here only covers the physical aspects of the forecasts and not the warnings based on the forecasts and the anticipation of the information. These aspects will be evaluated within the WMO/WWRP Hiweather project (Zhang et al., 2019). The future plan is for ECMWF to collaborate with other partners in the project in order to cover the evaluation of the full forecasting chain of hazardous weather events. This would increase our knowledge of the value of the forecast information and shortcomings in the system.

References

Baldwin, M.P., Dunkerton, T.J., 2001. Stratospheric harbingers of anomalous weather regimes. Science 294, 581−584. Available from: https://doi.org/10.1126/science.1063315.

Baran, S., Baran, Á., Pappenberger, F., Ben Bouallègue, Z., 2020. Statistical post-processing of heat index ensemble forecasts: is there a royal road? Q. J. R. Meteorological Soc. 146.

Bartholmes, Todini, 2005. Coupling meteorological and hydrological models for flood forecasting. Hydrol. Earth Syst. Sci. 9, 333−346. Available from: https://doi.org/10.5194/hess-9-333-2005.

Bedia, J., Golding, N., Casanueva, A., Iturbide, M., Buontempo, C., Gutiérrez, J.M., 2018. Seasonal predictions of Fire Weather Index: Paving the way for their operational applicability in Mediterranean Europe. Clim. Serv. 9, 101−110. Available from: https://doi.org/10.1016/j.cliser.2017.04.001.

Ben-Bouallegue, Z., 2020. Accounting for representativeness in the verification of ensemble forecasts, Technical Memorandum 865, ECMWF.

Ben-Bouallegue, Z., Magnusson, L., Haiden, T., Richardson, D.S., 2019. Monitoring trends in ensemble forecast performance focusing on surface variables and high-impact events. Q. J. R. Meteorol. Soc. 145, 1741−1755. Available from: https://doi.org/10.1002/qj.3523. 2019.

Błażejczyk, K., Jendritzky, G., Bröde, P., Fiala, D., Havenith, G., Epstein, Y., et al., 2013. An introduction to the Universal Thermal Climate Index (UTCI). Geographia Polonica 86, 5−10.

Blöschl, G., et al., 2019. Twenty-three unsolved problems in hydrology (UPH) − a community perspective. Hydrological Sci. J. 64 (10), 1141−1158. Available from: https://doi.org/10.1080/02626667.2019.1620507.

Brode, P., Fiala, D., Blazejczyk, K., Holmer, I., Jendritzky, G., Kampmann, B., et al., 2012. Deriving the operational procedure for the Universal Thermal Climate Index (UTCI). Int. J. Biometeorol. 56, 481−494.

Brutsaert, W., 2005. Hydrology: An Introduction. Cambridge University Press, p. 618, ISBN1107268796, 9781107268791.

Casanueva, A., Burgstall, A., Kotlarski, S., Messeri, A., Morabito, M., Flouris, A.D., et al., 2019. Overview of existing heat-health warning systems in Europe. Int. J. Environ. Res. public. health 16, 2657.

Cassou, C., 2008. Intraseasonal interaction between the Madden−Julian oscillation and the North Atlantic oscillation. Nature 455 (7212), 523−527.

Cheshire Jr., W.P., 2016. Thermoregulatory disorders and illness related to heat and cold stress. Auton. Neurosci. 196, 91−104.

Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K. E., et al., 2017. The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrol. Earth Syst. Sci. 21, 3427−3440. Available from: https://doi.org/10.5194/hess-21-3427-2017.

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. J. Hydrol. 375 (3−4), 613−626. Available from: https://doi.org/10.1016/j.jhydrol.2009.06.005. ISSN 0022−1694.

Cloke, H.L., Pappenberger, F., Smith, P.J., Wetterhall, F., 2017. How do I know if I've improved my continental scale flood early warning system? Environ. Res. Lett. 12, 044006. Available from: https://doi.org/10.1088/1748-9326/aa625a. a, b, c.

Day, J.J., Arduini, G., Sandu, I., Magnusson, L., Beljaars, A., Balsamo, G., et al., 2020. Measuring the impact of a new snow model using surface energy budget process relationships. J. Adv. Modeling Earth Syst. 12, e2020MS002144. Available from: https://doi.org/10.1029/2020MS002144.

Di Giuseppe, F., Pappenberger, F., Wetterhall, F., Krzeminski, B., Camia, A., Libertá, G., et al., 2016. The potential predictability of fire danger provided by numerical weather prediction. J. Appl. Meteor. Climatol. 55, 2469−2491. Available from: https://doi.org/10.1175/JAMC-D-15-0297.1.

Di Giuseppe, F., Vitolo, C., Krzeminski, B., Barnard, C., Maciel, P., San-Miguel, J., 2020. Fire Weather Index: the skill provided by the European Centre for Medium-Range Weather Forecasts ensemble prediction system. Nat. Hazards Earth Syst. Sci. 20, 2365−2378. Available from: https://doi.org/10.5194/nhess-20-2365-2020.

Di Napoli, C., Pappenberger, F., Cloke, H.L., 2018. Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI). Int. J. Biometeorol. 62, 1155−1165.

Di Napoli, C., Pappenberger, F., Cloke, H.L., 2019. Verification of heat stress thresholds for a health-based heat-wave definition. J. Appl. Meteorol. Climatol. 58, 1177–1194.

Di Napoli, C., Barnard, C., Prudhomme, C., Cloke, H.L., Pappenberger, F., 2021. ERA5-HEAT: a global gridded historical dataset of human thermal comfort indices from climate reanalysis. Geosci. Data J. 8, 2–10.

Di Napoli, C., Hogan, R.J., Pappenberger, F., 2020. Mean radiant temperature from global-scale numerical weather prediction models. Int. J. Biometeorol. 64, 1233–1245.

Dirmeyer, P.A., 2011. A history and review of the global soil wetness project (GSWP). J. Hydrometeorology 12 (5), 729–749. Available from: https://doi.org/10.1175/JHM-D-10-05010.1.

Donat, M.G., Leckebusch, G.C., Pinto, J.G., Ulbrich, U., 2010. Examination of wind storms over central Europe with respect to circulation weather types and NAO phases. Int. J. Climatol. 30 (9), 1289–1300. Available from: https://doi.org/10.1002/joc.1982.

Emerton, R., Cloke, H.L., Stephens, E.M., Zsoter, E., Woolnough, S.J., Pappenberger, F., 2017. Complex picture for likelihood of ENSO-driven flood hazard. Nat. Commun. 8, 14796. Available from: https://doi.org/10.1038/ncomms14796. ISSN 2041-1723.

Ferranti, L., Viterbo, P., 2006. The european summer of 2003: sensitivity to soil water initial conditions. J. Clim. 19 (15), 3659–3680. Available from: https://doi.org/10.1175/JCLI3810.1.

Ferranti, L., Magnusson, L., Vitart, F., Richardson, D.S., 2018. How far in advance can we predict changes in large-scale flow leading to severe cold conditions over europe? Q. J. R. Meteorological Soc. 144 (715), 1788–1802. Available from: https://doi.org/10.1002/qj.3341.

Fiala, D., Havenith, G., Brode, P., Kampmann, B., Jendritzky, G., 2012. UTCI-Fiala multi-node model of human heat transfer and temperature regulation. Int. J. Biometeorol. 56, 429–441.

Fischer, E.M., Seneviratne, S.I., Vidale, P.L., Luthi, D., Schar, C., 2007. Soil moisture-atmosphere interactions during the 2003 european summer heat wave. J. Clim. 20 (20), 5081–5099. Available from: https://doi.org/10.1175/JCLI4288.1.

Flannigan, M.D., Krawchuk, M.A., de Groot, W.J., Wotton, B.M., Gowman, L.M., 2009. Implications of changing climate for global wildland fire. Int. J. Wildland Fire 18, 483–507. Available from: https://doi.org/10.1071/WF08187.

Forbes, R., Haiden, T., Magnusson, L., 2015. Improvements in IFS forecasts of heavy precipitation. Newsletter 144. ECMWF.

Fragkoulidis, G., Wirth, V., Bossmann, P., Fink, A.H., 2018. Linking northern hemisphere temperature extremes to rossby wave packets. Q. J. R. Meteorological Soc. 144 (711), 553–566. Available from: https://doi.org/10.1002/qj.3228.

Gascon, E., Laviola, S., Merino, A., Miglietta, M., 2016. Analysis of a localized flash-flood event over the central mediterranean. Atmos. Res. 182, 256–268. Available from: https://doi.org/10.1016/j.atmosres.2016.08.007.

Gascon, E., Hewson, T., Haiden, T., 2018. Improving predictions of precipitation type at the surface: description and verification of two new products from the ECMWF ensemble. Weather Forecast. 33 (1), 89–108. Available from: https://doi.org/10.1175/WAF-D-17-0114.1.

Ghelli, A., Primo, C., 2009. One the use of the extreme dependency score to investigate the performance of a NWP model for rare events. Meteorol. Appl. 16, 537–544.

Grams, C.M., Wernli, H., Bottcher, M., Campa, J., Corsmeier, U., Jones, S.C., et al., 2011. The key role of diabatic processes in modifying the upper-tropospheric wave guide: a north atlantic case-study. Q. J. R. Meteorological Soc. 137 (661), 2174–2193. Available from: https://doi.org/10.1002/qj.891.

Grams, C.M., Magnusson, L., Madonna, E., 2018. An atmospheric dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. Q. J. R. Meteorological Soc. 144 (717), 2577–2591. Available from: https://doi.org/10.1002/qj.3353.

Grazzini, F., 2007. Predictability of a large-scale flow conducive to extreme precipitation over the western alps. Meteorol. Atmos. Phys. 95 (3). Available from: https://doi.org/10.1007/s00703-006-0205-8.

Grazzini, F., Vitart, F., 2015. Atmospheric predictability and rossby wave packets. Q. J. R. Meteorological Soc. 141 (692), 2793–2802. Available from: https://doi.org/10.1002/qj.2564.

Grazzini, F., Craig, G.C., Keil, C., Antolini, G., Pavan, V., 2019. Extreme precipitation events over northern italy. part (i): a systematic classification with machine learning techniques. Q. J. R. Meteorological Soc. Available from: https://doi.org/10.1002/qj.36350(ja).

Guo, J., Zhou, J., Lu, J., Zou, Q., Zhang, H., Bi, S., 2014. Multi-objective optimization of empirical hydrological model for streamflow prediction. J. Hydrol. 511, 242–253. Available from: https://doi.org/10.1016/j.jhydrol.2014.01.047. ISSN 0022-1694.

Hewson, T.D., Neu, U., 2015. Cyclones, windstorms and the imilast project. Tellus A: Dyn. Meteorol. Oceanography 67 (1), 27. Available from: https://doi.org/10.3402/tellusa.v67.27128. 128.

Hogan, R., Ahlgrimm, M., Balsamo, G., Beljaars, A., Berrisford, P., Bozzo, A., et al., 2017. Radiation in numerical weather prediction. Technical Memorandum 816, ECMWF.

Jacobs, S.J., Pezza, A.B., Barras, V., Bye, J., 2014. A new 'bio-comfort' perspective for Melbourne based on heat stress, air pollution and pollen. Int. J. Biometeorol. 58, 263–275.

Jendritzky, G., de Dear, R., Havenith, G., 2012. UTCI—why another thermal index? Int. J. Biometeorol. 56, 421–428.

Khodayar, S., Kalthoff, N., Kottmeier, C., 2018. Atmospheric conditions associated with heavy precipitation events in comparison to seasonal means in the western mediterranean region. Clim. Dyn. 51 (3), 951–967. Available from: https://doi.org/10.1007/s00382-016-3058-y.

Lalaurette, F., 2003. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. Q. J. R. Meteorol. 129 (594), 3037–3057. Available from: https://doi.org/10.1256/qj.02.152.

Lau, K.K.L., Krüger E., 2020. Establishing a global database for outdoor thermal comfort survey: A pilot study of standardisation of methodology. In 11th Windsor Conference on Thermal Comfort, 13.

Lavers, D.A., Villarini, G., 2013. The nexus between atmospheric rivers and extreme precipitation across europe. Geophys. Res. Lett. 40 (12), 3259–3264. Available from: https://doi.org/10.1002/grl.50636.

Lavers, D.A., Zsoter, E., Richardson, D.S., Pappenberger, F., 2017. An assessment of the ecmwf extreme forecast index for water vapor transport during boreal winter. Weather. Forecast. 32 (4), 1667–1674. Available from: https://doi.org/10.1175/WAF-D-17-0073.1.

Leroyer, S., Bélair, S., Spacek, L., Gultepe, I., 2018. Modelling of radiation-based thermal stress indicators for urban numerical weather prediction. Urban. Clim. 25, 64–81.

Liu, X., Coulibaly, P., Evora, N., 2008. Comparison of data-driven methods for downscaling ensemble weather forecasts. Hydrol. Earth Syst. Sci. 12, 615–624. Available from: https://doi.org/10.5194/hess-12-615-2008.

Lopez, P., 2016. A lightning parameterization for the ecmwf integrated forecasting system. Monthly Weather. Rev. 144 (9), 3057–3075. Available from: https://doi.org/10.1175/MWR-D-16-0026.1.

Magnusson, L., 2017. Diagnostic methods for understanding the origin of forecast errors. Q. J. R. Met. Soc. 143 (706), 2129–2142. Available from: https://doi.org/10.1002/qj.3072.

Magnusson, L., 2019. ECMWF Severe event catalogue for evaluation of multi-scale prediction of extreme events, Technical Memorandum 851, ECMWF.

Magnusson, L., Haiden, T., Richardson, D.S., 2014. Verification of extreme weather events: Discrete predictands, Technical Memorandum 731, ECMWF.

Majumdar, S.J., et al., 2021. Multiscale forecasting of high-impact weather: current status and future challenges. Bull. Amer. Meteor. Soc. 102, E635–E659. Available from: https://doi.org/10.1175/BAMS-D-20-0111.1.

Martius, O., Schwierz, C., Davies, H.C., 2008. Far-upstream precursors of heavy precipitation events on the alpine south-side. Q. J. R. Meteorological Soc. 134 (631), 417–428. Available from: https://doi.org/10.1002/qj.229.

Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F., Matschullat, J., 2020. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. Submitted Int. J. Climatol.

McGregor, G.R., Vanos, J.K., 2018. Heat: a primer for public health researchers. Public. Health 161, 138–146.

Mölders, N., 2008. Suitability of the Weather Research and Forecasting (WRF) model to predict the June 2005 fire weather for Interior Alaska, 395. Weather. Forecast. 23, 953–973.

Mölders, N., 2010. Comparison of Canadian Forest Fire Danger Rating System and National Fire Danger Rating System fire indices derived from Weather Research and Forecasting (WRF) model data for the June 2005 Interior Alaska wildfires. Atmos. Res. 95, 290–306.

Nuissier, O., Joly, B., Joly, A., Ducrocq, V., Arbogast, P., 2011. A statistical downscaling to identify the largescale circulation patterns associated with heavy precipitation events over southern france. Q. J. R. Meteorological Soc. 137 (660), 1812–1827. Available from: https://doi.org/10.1002/qj.866.

Pantillon, F., Lerch, S., Knippertz, P., Corsmeier, U., 2018. Forecasting wind gusts in winter storms using a calibrated convection-permitting ensemble. Q. J. R. Meteorological Soc. 144 (715), 1864–1881. Available from: https://doi.org/10.1002/qj.3380.

Pappenberger, F., Ramos, M.H., Cloke, H.L., Wetterhall, F., Alfieri, L., Bogner, K., et al., 2015a. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. J. Hydrol. 522, 697–713ISSN 0022-1694. Available from: https://doi.org/10.1016/j.jhydrol.2015.01.024.

Pappenberger, F., Jendritzky, G., Staiger, H., Dutra, E., Di Giuseppe, F., Richardson, D. S., et al., 2015b. Global forecasting of thermal health hazards: the skill of probabilistic predictions of the Universal Thermal Climate Index (UTCI). Int. J. Biometeorol. 59, 311–323.

Pillosu, F., Hewson, T., 2017. New point-rainfall forecasts for flash flood prediction. Newsletter 153, ECMWF.

Quinting, J.F., Vitart, F., 2019. Representation of synoptic-scale rossby wave packets and blocking in the s2s prediction project database. Geophys. Res. Lett. 46 (2), 1070–1078. Available from: https://doi.org/10.1029/2018GL081381.

Ralph, F.M., Dettinger, M.D., 2011. Storms, floods, and the science of atmospheric rivers. Eos, Trans. Am. Geophys. Union. 92 (32), 265–266. Available from: https://doi.org/10.1029/2011EO320001.

Raveh-Rubin, S., Catto, J.L., 2019. Climatology and dynamics of the link between dry intrusions and cold fronts during winter, Part II: Front-centred perspective. Clim. Dyn. 53, 1893–1909. Available from: https://doi.org/10.1007/s00382-019-04793-2.

Raveh-Rubin, S., Wernli, H., 2015. Large-scale wind and precipitation extremes in the mediterranean: a climatological analysis for 1979–2012. Q. J. R. Meteorological Soc. 141 (691), 2404–2417. Available from: https://doi.org/10.1002/qj.2531.

Richardson, D.S., 2000. Skill and relative economic value of the ECMWF Ensemble Prediction System. Q. J. R. Met. Soc. 126, 649–668.

Roads, J., Fujioka, F., Chen, S., Burgan, R., 2005. Seasonal fire danger forecasts for the USA. Int. J. Wildland Fire 14, 1–18.

Rodwell, M.J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., et al., 2013. Characteristics of occasional poor medium-range weather forecasts for Europe. Bull. Amer. Meteor. Soc. 94, 1393–1405.

Sandu, I., Haiden, T., Balsamo, G., Schmederer, P., Arduini, G., Day, J., et al., 2020. Addressing near-surface forecast biases: outcomes of the ECMWF project 'Understanding uncertainties in surface atmosphere exchange' (USURF). Technical Memorandum 875, ECMWF.

Shukla, S., Lettenmaier, D.P., 2011. Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill. Hydrol. Earth Syst. Sci. 15, 3529–3538. Available from: https://doi.org/10.5194/hess-15-3529-2011.

Sivakumar, B., 2005. Hydrologic modeling and forecasting: role of thresholds. Environ. Model. & Softw. 20 (5), 515–519. Available from: https://doi.org/10.1016/j.envsoft.2004.08.006. ISSN 1364–8152.

Stephenson, D.B., Casati, B., Ferro, C.A.T., Wilson, C.A., 2008. The extreme dependency score: a non-vanishing measure for forecasts of rare events. Meteorol. Appl. 15, 41–50.

Sutanto, S.J., Vitolo, C., Di Napoli, C., D'Andrea, M., Van Lanen, H.A.J., 2020. Heatwaves, droughts, and fires: Exploring compound and cascading dry hazards at the pan-European scale. Env. Int. 134, 105276.

Tsonevsky, I., 2015. New EFI parameters for forecasting severe convection. Newsletter 144, ECMWF.

Trigo, R.M., Pozo-Vazquez, D., Osborn, T.J., Castro-Diez, Y., Gamiz-Fortis, S., Esteban-Parra, M.J., 2004. North atlantic oscillation influence on precipitation, river flow and water resources in the iberian peninsula. Int. J. Climatol. 24 (8), 925–944. Available from: https://doi.org/10.1002/joc.1048.

Tsanis, I., Tapoglou, E., 2019. Winter north atlantic oscillation impact on european precipitation and drought under climate change. Theor. Appl. Climatol. 135 (1), 323–330. Available from: https://doi.org/10.1007/s00704-018-2379-7.

Vergni, L., Todisco, F., Di Lena, B., Mannocchi, F., 2016. Effect of the north atlantic oscillation on winter daily rainfall and runoff in the abruzzo region (central italy). Stoch. Environ. Res. Risk Assess. 30 (7), 1901. Available from: https://doi.org/10.1007/s00477-015-1194-2. 191.

Vitart, F., 2014. Evolution of ECMWF sub-seasonal forecast skill scores. Q. J. R. Met. Soc. 140, 1889–1899. Available from: https://doi.org/10.1002/qj.2256.

Vitart, F., Robertson, A.W., 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. Npj Clim. Atmos. Sci. 1, N3. Available from: https://doi.org/10.1038/s41612-018-0013-0.

Vitolo, C., Di Napoli, C., Di Giuseppe, F., Cloke, H.L., Pappenberger, F., 2019. Mapping combined wildfire and heat stress hazards to improve evidence-based decision making. Env. Int. 127, 21–34.

Wernli, H., Davies, H.C., 1997. A lagrangian-based analysis of extratropical cyclones. i: the method and some applications. Q. J. R. Meteorological Soc. 123 (538), 467–489. Available from: https://doi.org/10.1002/qj.49712353811.

Wirth, V., Eichhorn, J., 2014. Long-lived rossby wave trains as precursors to strong winter cyclones over europe. Q. J. R. Meteorological Soc. 140 (680), 729–737. Available from: https://doi.org/10.1002/qj.2191.

Wirth, V., Riemer, M., Chang, E.K.M., Martius, O., 2018. Rossby wave packets on the midlatitude waveguide - a review. Monthly Weather. Rev. 146 (7), 1965–2001. Available from: https://doi.org/10.1175/MWR-D-16-0483.1.

Woollings, T., Hoskins, B., Blackburn, M., Berrisford, P., 2008. A new Rossby wave-breaking interpretation of the north Atlantic oscillation. J. Atmos. Sci. 65 (2), 609–626. Available from: https://doi.org/10.1175/2007JAS2347.1.

Wulff, C.O., Greatbatch, R.J., Domeisen, D.I.V., Gollan, G., Hansen, F., 2017. Tropical forcing of the summer east atlantic pattern. Geophys. Res. Lett. 44 (21), 11,166–11,173. Available from: https://doi.org/10.1002/2017GL075493.

Zhang, Q., Li, L., Ebert, B., Golding, B., Johnston, D., Mills, B., et al., 2019. Increasing the value of weather-related warnings. Sci. Bull. 64 (0), 647–649. Available from: https://doi.org/10.1016/j.scib.2019.04.003.

Zschenderlein, P., Fink, A.H., Pfahl, S., Wernli, H., 2019. Processes determining heat waves across different european climates. Q. J. R. Meteorological Soc. Available from: https://doi.org/10.1002/qj.3599.