Climate Services 39 (2025) 100594

Contents lists available at ScienceDirect

Climate Services

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

Original research article

Balancing accuracy versus precision: Enhancing the usability of sub-seasonal forecasts

Etienne Dunn-Sigouin [a,b,*], Erik W. Kolstad [a,b], C. Ole Wulff [a,b], Douglas J. Parker [a,b,c,d], Richard J. Keane [c,e]

[a] NORCE Norwegian Research Center AS, Bergen, Norway
[b] Bjerknes Center for Climate Research, Bergen, Norway
[c] School of Earth and Environment, University of Leeds, Leeds, UK
[d] NCAS National Centre for Atmospheric Science, University of Leeds, Leeds, UK
[e] Met Office, Exeter, UK

ARTICLE INFO

Keywords:
Sub-seasonal forecasts
Usability gap
Forecast skill horizon
Climate adaptation
ECMWF
Fractions skill score

ABSTRACT

Forecasts are essential for climate adaptation and preparedness, such as in early warning systems and impact models. A key limitation to their practical use is often their coarse spatial grid spacing. However, another less frequently discussed but crucial limitation is that forecasts are often more precise than they are accurate when their grid spacing is finer than the scales they can accurately predict. Here, we adapt the fractions skill score, a metric conventionally used to quantify spatial forecast accuracy by the meteorological community, to help users navigate the trade-off between forecast accuracy versus precision. We demonstrate how this trade-off can be visualized for daily European precipitation, focusing on deterministic predictions of anomalies and probabilistic predictions of extremes, derived from three years of sub-seasonal forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). Our results show that decreasing precision through spatial aggregation increases forecast accuracy, extends predictable lead times, and enhances the maximum possible accuracy relative to the grid scale, while increased precision diminishes these benefits. Notably, spatial aggregation benefits daily-accumulated forecasts more than weekly-accumulated ones, per unit lead-time. We demonstrate the practical value of our approach in three examples: communicating early warnings, managing hydropower capacity, and commercial aviation planning—each characterized by distinct user constraints on accuracy, spatial scale, or lead-time. The results suggest a different approach for using forecasts; post-processing forecasts to focus on the most accurate scales rather than the default grid scale, thus offering users more actionable information.

Practical implications

Our results address a critical need for more accurate and actionable sub-seasonal forecasts, especially at longer lead times when crucial decisions are made. By adapting the fractions skill score, we illustrate how users can visualize and optimize the trade-off between a forecast's spatial precision and its accuracy. Rather than relying on default high-resolution grids, we show that spatially aggregating forecasts can extend predictability and offer a clearer perspective on potential weather hazards. This approach not only complements existing forecast systems but also provides insights into when and where coarser-scale information is more dependable than finer scales. Ultimately, practitioners gain a practical tool that highlights where and how forecast aggregation pays dividends for planning at longer lead times. We demonstrate the value of the tool using three real-world examples.

In the first example, we show how early warning systems can benefit from the extended lead times offered by spatial aggregation. We demonstrate this using the 2023 Storm Hans case in Norway, which required timely alerts to protect lives and infrastructure. Even with the inherent uncertainty of precipitation forecasts, aggregating them across broader areas yields more robust indications of impending extreme rainfall. This method could allow forecasters to issue warnings earlier, while policy makers and emergency managers stand to gain crucial time to mobilize resources.

The second example focuses on hydropower operations, where decisions are driven by localized hydrological processes but still benefit from a strategic view of precipitation patterns. Because releasing water from reservoirs too early can be costly, operators need maximum confidence in imminent rainfall forecasts. By matching the spatial

[*] Corresponding author at: NORCE Norwegian Research Center AS, Bergen, Norway.
E-mail address: etdu@norceresearch.no (E. Dunn-Sigouin).

https://doi.org/10.1016/j.cliser.2025.100594
Received 11 February 2025; Received in revised form 4 June 2025; Accepted 28 June 2025
Available online 15 July 2025

aggregation scale to the watersheds of interest, hydropower managers can zero in on the most relevant signals. Our analyses highlight how post-processing forecasts at a watershed scale sharpens the focus on potential inflows, thereby supporting economically and environmentally sound reservoir management. Such tailored forecasting helps optimize water releases, reducing both flood risk, infrastructure damage and lost revenue opportunities.

In the third example, commercial aviation stands to benefit from spatial aggregation of forecasts when flights must be scheduled or canceled days in advance to minimize disruption. Spatial aggregation enables major carriers to detect the broader "footprint" of storms like Storm Hans well ahead of time, increasing confidence in decisions regarding flight cancellations, route changes, and resource allocation. Although individual airports lose fine-scale information, airlines can avoid the larger financial losses and passenger inconvenience that arise from last-minute adjustments over the entire network. This balance between lead time and accuracy can make flight networks more robust under uncertain weather conditions, ultimately improving safety and travel reliability while minimizing costs.

Taken together, these examples reveal how our adapted fractions skill score framework opens the door to "scalable" forecasts that users can customize to their unique spatial constraints. By offering a method to systematically aggregate forecasts, decision-makers can glean earlier and more trustworthy signals of potential high-impact events, optimizing their interventions in sectors ranging from disaster management to energy production and transportation. Our method does not eliminate all forecast uncertainties but provides a structured way to capitalize on known forecast strengths. In doing so, it encourages a shift from assuming that higher-resolution forecasts are always better, to aligning forecast precision with the scales that can actually be predicted and those most relevant to users.

1. Introduction

In an era increasingly defined by climate change, the importance of weather and climate forecasts for society has surged, encompassing predictions from days to a decade ahead (Merryfield et al., 2020; White et al., 2022; O'Kane et al., 2023). This shift reflects a broader understanding of the critical role these forecasts play in managing the variable and often extreme environmental conditions caused by climate variability and change, and their integration into society reflects a stronger push for adaptation and preparedness (Goddard, 2016; Trenberth et al., 2016; Coughlan de Perez et al., 2022). Forecasts support a number of international adaptation efforts such as the World Meteorological Organization's Global Framework for Climate Services (Hewitt et al., 2012), the United Nations Early Warnings for All initiative (EW4ALL, WMO, 2022), and the European Union's financial sustainability taxonomy (European-Commission, 2020), and play a critical role in weather and climate services within the private sector (Cusick, 2019; Lam et al., 2023; Price et al., 2024). Forecasts are also used to predict impacts that are societally important but not directly modeled by their systems (Merz et al., 2020). These impact models vary widely in design and what they predict, such as floods, droughts, shipping routes, insurance risk, disease spread, agricultural cycles and renewable energy production (e.g., Torralba et al., 2017; Röösli et al., 2021; Graham et al., 2022; Haupt et al., 2018, 2019a,b).

There is, however, a well-known usability gap between the production of weather and climate information and its use (Lemos et al., 2012; Van den Hurk et al., 2018; Findlater et al., 2021). This gap is often attributed to the limited spatial resolution of forecasts, which often fails to meet the fine-scale precision required by users due to the prohibitive cost of high-resolution modeling.

Another less recognized yet crucial limitation to the practical use of forecasts, which is the focus of this paper, is the rapid decline in forecast skill at finer spatial scales as predictions extend into the future. This degradation in accuracy stems from faster error growth at smaller scales, where predictability is inherently linked to spatial size (Lorenz, 1969; Toth and Buizza, 2019). For instance, while slower large-scale cyclones spanning thousands of kilometers may be predictable over several days, smaller-scale thunderstorms operate on much shorter timescales, with predictability limited to a few hours. Spatial and temporal aggregation can be used to counteract this small-scale error growth by effectively filtering out high-frequency, small-scale noise, thereby enhancing the predictable signal from the lower-frequency, larger-scale circulation. Thus, aggregation helps to extend the limit of predictability, known as the "forecast skill horizon" (Buizza et al., 2015; Buizza and Leutbecher, 2015), but at the cost of spatial or temporal precision.

While temporal aggregation is routine, it is rare to find weather and climate information – either from forecasts or projections – that has been aggregated spatially. For example, sub-seasonal forecasts (often referred to as extended-range or monthly forecasts) are usually presented in the form of weekly aggregated information, such as in the online charts catalogue of the European Centre for Medium-Range Weather Forecasts (ECMWF, https://charts.ecmwf.int/). However, in this case no spatial aggregation is done; the forecast charts are displayed at the model's original grid spacing. Even commonly used daily-aggregated weather forecasts are shown on the default grid spacing despite a well known decline in forecast skill at smaller spatial scales over just a few days. This can lead to a critical mismatch between the apparent precision in forecast products and the underlying accuracy in the data. Such misrepresentation risks undermining effective use of weather and climate information (Nissan et al., 2019; Fiedler et al., 2021).

Two ways of making forecasts better fit user needs are improving their predictive skill at finer spatial scales or helping users more effectively utilize existing skill. While improving forecasts remains a formidable challenge (Bauer et al., 2015; Benjamin et al., 2019), recent advances in machine-learning-based models have made significant strides, offering performance that now rivals traditional dynamical forecast models (Lam et al., 2023; Ben Bouallègue et al., 2024; Price et al., 2024). Despite these breakthroughs, it is widely acknowledged that there are likely intrinsic limits to the forecast skill horizon, which no amount of model improvement can overcome (Lorenz, 1969; Palmer et al., 2014). Thus, a pragmatic strategy involves helping users navigate the inherent trade-off between spatial accuracy and precision, optimizing existing forecasts for their needs.

Various strategies using spatio-temporal aggregation have been suggested (e.g., Gong et al., 2003; Gilleland et al., 2009; Jung and Leutbecher, 2008; Buizza and Leutbecher, 2015; Gehne et al., 2016; Toth and Buizza, 2019; Van Straaten et al., 2020; Young et al., 2020; Rivoire et al., 2023). These focus mostly on quantifying the forecast skill horizon where predictability is small and difficult to exploit in practice. An alternative, more user-oriented method is to use the fractions skill score (Roberts and Lean 2008), which quantifies where forecast predictability is high and usable. This approach, which aggregates forecasts over an increasing number of neighboring grid points, quantifies the trade-off between accuracy versus precision, and can be used to post-process the forecast according to the user's preferred balance. The fractions skill score and a number of other closely related methods stand out for their intuitiveness and practical applicability, yet their use has been largely confined to the meteorological community (e.g., Gilleland et al., 2009; Jolliffe and Stephenson, 2012; Keane et al., 2016; Zhao and Zhang, 2018; Schwartz, 2019; Cafaro et al., 2021).

In this study, we propose a novel methodology, based on the fractions skill score, that realigns forecast capabilities with end-user requirements, thereby enhancing their practical application. The innovative aspect of this method lies in shifting its focus from verifying forecasts for meteorologists to optimizing them for users. The method is applied to sub-seasonal forecasts with lead times ranging from 1 day to several weeks—an essential time-frame for many decision-making processes (Merz et al., 2020; White et al., 2022).

This work is part of the Climate Futures collaboration, an interdisciplinary and intersectoral initiative that, since 2020, has brought together public and private organizations in Norway to co-produce weather and climate prediction-based tools and services. A key case study involved working with Tryg Forsikring, a private insurance company, to incorporate forecasts into their decision-making in order to comply with sustainable finance regulations. The collaboration served two purposes: to inform insurance professionals on the practical limitations of using forecasts, and to maximize the utility of forecasts for insurance impact modeling. Our experience with the insurance sector suggests our approach could be broadly applied to support a wide range of climate adaptation efforts across industries. The intended users are meteorologists and climate scientists who supply forecasts, or industry professionals who can use forecast data but lack forecast expertise.

In the next section we start by introducing the forecast dataset and outlining the methodology, which amounts to spatially aggregating forecasts before calculating commonly used metrics of forecast accuracy. In the results section, we apply the new method to evaluate the trade-off between precision and accuracy in European precipitation forecasts, derived from three years of sub-seasonal forecasts from the ECMWF. We explore how this method can aid users in interpreting deterministic predictions of precipitation anomalies and probabilistic predictions of extremes. To illustrate this, we use the example of Storm Hans, which struck Scandinavia, Northern Europe, and the Baltics in August 2023, with Norway bearing the brunt of its impact. Unusually approaching from the east rather than the west, the storm shattered century-old rainfall records in eastern Norway (Granerød et al., 2023). The resulting extreme rainfall triggered widespread flooding and landslides, severely damaging homes, roads, railways, and bridges, with estimated costs reaching 4 billion Norwegian Krone or 350 million euro (Ekroll, 2023). Over 10,000 insurance claims were filed, and approximately 2,400 people were evacuated—the largest such evacuation in Norway since World War II (KLP, 2023). With extreme rainfall events expected to become more frequent due to climate change (Hanssen-Bauer et al., 2009), storm Hans exemplifies the growing challenges in climate adaptation.

Building on our findings and those of Roberts and Lean (2008), we end the paper by discussing how the method could be applied to forecasts in three different contexts — communicating early warnings, managing hydropower capacity, and commercial aviation planning — each characterized by distinct user-constraints on accuracy, spatial scale, or lead-time. In each case, we enhance forecast utility by post-processing forecasts to focus on the most accurate spatial scales, rather than the default grid scale precision.

2. Data

We use three years (2020–2022) of sub-seasonal forecasts from the ECMWF (Buizza et al., 2018) downloaded from the MARS archive (ECMWF, 2024a). We use bi-weekly initializations on Mondays and Thursdays, for a total of 313 forecasts, each comprising 51 ensemble members running 46 days in the future. The initial 15 lead-time days are higher resolution (0.25° × 0.25° grid spacing) than the last 31 days (0.5° × 0.5°), corresponding to approximately 28 km^2 and 56 km^2 at the equator, respectively. Accompanying each individual forecast is a set of retrospective forecasts. These were initialized on the same calendar day as the forecast over the previous 20 years and consist of 11 ensemble members. Such “hindcasts” provide an estimate of the climatological distribution accompanying each forecast.

The forecast-hindcast pairs correspond to different model versions over time (CY46R1, CY47R1, CY47R2, CY47R3) because the model is updated on the fly and our analysis spans multiple years. Changes in model cycles can influence model biases due to evolving model physics and data assimilation. While our approach (discussed in the next section) focuses on deviations from the model’s climatology, which helps mitigate systematic differences across model cycles, residual biases in the forecasts may remain. However, we do not expect these to qualitatively impact our results.

We focus our analysis on Europe (33°N to 73.5°N and 27°W to 35°E) and on predictions of daily and weekly-accumulated precipitation. A corresponding analysis of daily and weekly-mean 2-m temperature forecasts for two years (2020–2021) is included in the supplementary materials. Forecast skill was verified relative to ERA5 reanalysis (Hersbach et al., 2023) for the same grid, domain, and time period as the forecast. Although ERA5 exhibits known biases, such as a tendency for excessive drizzle (Lavers et al., 2022), it remains a convenient benchmark for verification because its resolution matches that of the forecast. We note, however, that other observational datasets may be used for verification, and that we do not expect this choice to impact our qualitative results. We also extended this analysis to storm Hans in 2023, incorporating additional forecasts and hindcasts initialized between 3 and 7 August alongside ERA5 data.

Finally, to illustrate the spatial scale of the data, we convert its spatial precision from gridpoint units to square kilometers, shown in the y-axis labels of Fig. 1. Specifically, we simply rescale the nominal 28 km^2 area represented by one gridpoint^2 at the equator by the mean cosine of latitude within the domain, consistent with the spherical geometry of Earth’s surface. Consequently, one gridpoint^2 within the European domain corresponds to approximately 15 km^2.

3. Methodology

We assess forecast accuracy as a function of precision and lead time using modified versions of the Fractions Skill Score (FSS, Roberts and Lean, 2008). Here, *accuracy* refers to the skill of the forecast quantified using a skill score, and *precision* refers to the level of spatial aggregation of the forecast. We begin by summarizing the original FSS developed for the meteorological community in Section 3a, followed by our adaptations for end users in sections 3b,c,d. Next, we introduce a modified version of the Extreme Forecast Index (EFI, Lalaurette, 2003), which we use to demonstrate the value of optimizing forecast accuracy during Storm Hans in Section 5.

To facilitate the computation of scores and indices in the following sections, it is useful to first convert the reanalysis verification into the same format as the forecasts and hindcasts. Table 1 summarizes the variables and their dimensions defined in Section 3. Specifically, forecasts $f(m,e,t,i,j)$ are characterized by dimensions of forecast initialization (m), ensemble member (e), lead time (t), latitude (i) and longitude (j). These correspond to a verification $v_f(m,t,i,j)$ from ERA5 reanalysis, where $e=1$ and $t=1$ represents the 24-hour period after the forecast initialization date m. Similarly, hindcasts $h(m,y,e,t,i,j)$ which include a hindcast year dimension (y), correspond to a verification $v_h(m,y,t,i,j)$ with $e = 1$ that spans the past twenty years for each calendar date of forecast initialization m.

3.1. Fractions skill score

The FSS uses binary forecast and verification data to assess the skill of the forecast at different levels of spatial aggregation. Roberts and Lean (2008) developed their method using deterministic forecasts of precipitation, i.e., with only one ensemble-member. First, they converted the forecast f and verification v_f to binary values based on a predefined absolute threshold (e.g., 4 mm). If the precipitation amount exceeded this threshold, the value was set to 1; otherwise, it was set to 0. Next, for each grid point, they averaged surrounding points within a square of length n (this process is referred to hereafter as *aggregation*), yielding an aggregated forecast F and verification V_F (see Eqs. (1) and (2)). These aggregations are not binary, but have fractional values between 0 and 1.

$$F(n,m,t,i,j)=\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}f\left[m,t,i+k-1-\frac{(n-2)}{2},j+l-1-\frac{(n-1)}{2}\right] \quad (1)$$

Table 1
Overview of variables and their dimensions defined in section 3.

Dimension/Variable	Description	Section
m	forecast initialization date	3
e	ensemble member	3
t	lead-time day	3
i	latitude	3
j	longitude	3
y	hindcast year	3
n	spatial aggregation level	3
q	quantile	3
$f(m,e,t,i,j)$	forecast	3
$h(m,y,e,t,i,j)$	hindcast	3
$v_f(m,t,i,j)$	forecast verification	3
$v_h(m,y,t,i,j)$	hindcast verification	3
$F(n,m,t,i,j)$	aggregated forecast	3.1
$V_F(n,m,t,i,j)$	aggregated forecast verification	3.1
$MSE_F(n,m,t)$	mean-square error of aggregated forecast	3.1
$MSE_{REF}(n,t)$	mean-square error of aggregated forecast relative to reference forecast	3.1
$FSS(n,m,t)$	fractions skill score	3.1
$\tilde{f}(m,t,i,j)$	forecast anomaly	3.2.1
$\tilde{v}_f(m,t,i,j)$	forecast verification anomaly	3.2.1
$\tilde{F}(m,t,i,j)$	aggregated forecast anomaly	3.2.1
$\tilde{V}_F(m,t,i,j)$	aggregated forecast verification anomaly	3.2.1
$MSE_{\tilde{F}}(n,m,t)$	mean-square error of aggregated forecast anomalies	3.2.1
$MSE_{R\tilde{E}F}(n,m,t)$	mean-square error of aggregated forecast anomalies relative to reference forecast	3.2.1
$FMSESS(n,t)$	fractions mean-square error skill score	3.2.1
$H(n,m,t,i,j)$	aggregated hindcast	3.2.2
$V_H(n,m,t,i,j)$	aggregated hindcast verification	3.2.2
$F_q(n,m,t,i,j)$	aggregated forecast threshold value for quantile q	3.2.2
$V_q(n,m,t,i,j)$	aggregated verification threshold value for quantile q	3.2.2
$P_{F_q}(n,m,t,i,j)$	aggregated forecast probability for quantile q	3.2.2
$P_{V_q}(n,m,t,i,j)$	aggregated binary verification for quantile q	3.2.2
$BS_q(n,m,t)$	brier-score of aggregated forecasts for quantile q	3.2.2
$BS_{REF_q}(n,m,t)$	brier-score of aggregated forecasts for quantile q relative to reference forecast	3.2.2
$FBSS_q(n,t)$	fractions brier skill score for quantile q	3.2.2
$h_{thresh}(q,m,t,i,j)$	hindcast threshold value	3.3
$fr_f(q,m,t,i,j)$	fraction of forecast ensemble members over threshold	3.3
$EFI(m,t,i,j)$	extreme forecast index	3.3
$v_{h_{thresh}}(q,m,t,i,j)$	verification hindcast threshold value	3.3
$fr_v(q,m,t,i,j)$	binary verification over threshold	3.3
$EVI(m,t,i,j)$	extreme verification index	3.3
$H_{thresh}(n,q,m,t,i,j)$	aggregated hindcast threshold value	3.3
$FR_F(n,q,m,t,i,j)$	fraction of aggregated forecast ensemble members over threshold	3.3
$FEFI(n,m,t,i,j)$	fractions extreme forecast index	3.3

$$V_F(n,m,t,i,j) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} v_f\left[m,t,i+k-1-\frac{(n-2)}{2},j+l-1-\frac{(n-1)}{2}\right] \quad (2)$$

Grid points within the square of length n but outside the domain defined in Section 2 (i.e., Europe) were set to zero. By comparing the mean square error of the aggregated forecast over the domain (Eq. (3)) with that calculated from an aggregated reference forecast for each n, they obtained the FSS (Eq. (4)).

$$MSE_F(n,m,t) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}[F(n,m,t,i,j) - V_F(n,m,t,i,j)]^2 \quad (3)$$

$$FSS(n,m,t) = 1 - \frac{MSE_F(m,n,t)}{MSE_{REF}(n,t)} \quad (4)$$

An FSS value of 1 signifies perfect forecast accuracy relative to the verification data, while an FSS value of 0 or less indicates the forecasts are no better or worse than a reference forecast. The choice of reference forecast is up to the user (e.g., random forecast, climatology or something else).

3.2. Modified fractions skill scores

Next, we modify the original FSS to better adapt it to end-users. The details of the modified scores are described in the next sections and the primary steps can be summarized as:

1. Aggregate the raw forecast and verification fields spatially following Eqs. (1) and (2).
2. Compute a standard grid point-wise score such as Mean Square Error or Brier Score.
3. Compute a skill score by comparing the score to a suitably aggregated reference forecast and average over all spatial grid-points and forecasts.

3.2.1. Fractions mean-square error skill score

The Fractions Mean-Square Error Skill Score (FMSESS) quantifies the accuracy of ensemble-mean forecast anomalies, averaged across all grid points and forecasts, over varying spatial aggregation scales. Unlike the original FSS, which uses binary threshold-based values, the FMSESS incorporates anomalies relative to climatology. This modification offers several benefits to users: (1) it generalizes the widely used mean-square error skill score across multiple spatial scales (Jolliffe and Stephenson, 2012), enabling better comparisons with past studies; (2) it simplifies the interpretation by providing a measure of accuracy independent of threshold; (3) it provides a better estimate of forecast skill by incorporating a mean-bias correction.

Forecast anomalies $\tilde{f}$ are computed by taking the ensemble-mean of the difference between the forecast and the hindcast climatology (Eq. (5)), while verification anomalies $\tilde{v}_f$ are calculated by subtracting the verification climatology from each verification v_f (Eq. (6)).

$$\tilde{f}(m,t,i,j) = \frac{1}{E}\sum_{e=1}^{E}\left[f(m,e,t,i,j) - \frac{1}{Y}\sum_{y=1}^{Y} h(m,y,e,t,i,j)\right] \quad (5)$$

$$\tilde{v}_f(m,t,i,j) = v_f(m,t,i,j) - \frac{1}{Y}\sum_{y=1}^{Y} v_h(m,y,t,i,j) \quad (6)$$

We use a single date m to define the climatologies for each forecast and verification for simplicity. A more robust estimate of the climatology could be achieved by incorporating additional dates centered around the forecast/verification date, as demonstrated by ECMWF's M-climate (ECMWF, 2024b). However, we do not expect this choice to qualitatively affect our main results.

Aggregated forecast and verification anomalies $\tilde{F}$ and $\tilde{V}_F$ are then used to calculate the FMSESS similar to the original FSS (Eqs. (7) and (9)). The aggregated version of the verification climatology (second term on the right-hand size of Eq. (6)) is used as the reference forecast to calculate the reference mean-square error in the FMSESS (Eq. (8)).

$$MSE_{\tilde{F}}(n,m,t) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}[\tilde{F}(n,m,t,i,j) - \tilde{V}_F(n,m,t,i,j)]^2 \tag{7}$$

$$MSE_{R\tilde{E}F}(n,m,t) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}[\tilde{F}(n,m,t,i,j) - \frac{1}{Y}\sum_{y=1}^{Y} V_H(n,m,y,t,i,j)]^2 \tag{8}$$

$$FMSESS(n,t) = 1 - \sum_{m=1}^{M}\frac{MSE_{\tilde{F}}(m,n,t)}{MSE_{R\tilde{E}F}(m,n,t)} \tag{9}$$

3.2.2. Fractions brier skill score

The Fractions Brier Skill Score (FBSS) quantifies the accuracy of forecast extremes, averaged over all grid points and forecasts, across varying spatial aggregation scales. It introduces two key modifications to the original FSS. First, it provides a probabilistic assessment of skill by utilizing an ensemble of forecasts instead of a single deterministic forecast. Second, it uses a threshold value from a predefined quantile based on the hindcast climatology, rather than an absolute threshold. In this study, we demonstrate the method using the 0.1 and 0.9 quantiles, corresponding to dry and wet extremes. Similar to the FMSESS, the FBSS offers distinct advantages to users: (1) it generalizes the widely-used Brier Skill Score across multiple spatial scales (Jolliffe and Stephenson, 2012), enabling better comparisons with past studies; (2) it improves the evaluation of extremes via probabilistic scoring and quantile-based bias correction.

We start by defining a threshold value for extremes for a given quantile q. First, we calculate aggregated hindcast $H(n,m,e,y,t,i,j)$ and verification $V_H(n,m,y,t,i,j)$ for each n following Eqs. (1) and (2). The forecast threshold value $F_q(n,m,t,i,j)$ is then computed for the quantile q from a sample of e ensemble members and y hindcast years in hindcast H. Correspondingly, the verification threshold value $V_q(n,m,t,i,j)$ is computed for the quantile q from a sample of y hindcast years in verification hindcast V_H.

Next, we compute the Brier Score BS_q for a given quantile q. First, we calculate the aggregated forecast $F(n,m,e,t,i,j)$ and verification $V_f(n,m,t,i,j)$ for each n. Then, we compute the forecast probability $P_{F_q}(n,m,t,i,j)$ by determining the fraction of ensemble members e in forecast $F(n,m,e,t,i,j)$ that exceed the threshold value $F_q(n,m,t,i,j)$. Similarly, we compute the binary verification $P_{V_q}(n,m,t,i,j)$ based on whether the verification $V_f(n,m,t,i,j)$ crosses the threshold value $V_q(n,m,t,i,j)$, assigning 1 to values above the threshold and 0 below. The squared difference between the forecast probability and the binary verification is then averaged over all values in the domain (Eq. (10)).

$$BS_q(n,m,t) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}[P_{F_q}(n,m,t,i,j) - P_{V_q}(n,m,t,i,j)]^2 \tag{10}$$

Finally, the FBSS is computed by comparing the Brier Score of the forecast with one calculated from a reference forecast and averaging over all forecasts (Eq. (11)). The reference forecast used in the reference Brier Score is simply the quantile q used to define the threshold ($q = 0.9$ or $q = 0.1$).

$$FBSS_q(n,t) = 1 - \sum_{m=1}^{M}\frac{BS_q(n,m,t)}{BS_{REF_q}(n,m,t)} \tag{11}$$

3.2.3. Statistical significance of skill scores

Statistical significance of the FMSESS and FBSS (Eqs. (9) and (11)) is evaluated using bootstrapping. We generate a distribution of scores for each spatial scale n and lead-time t by resampling the forecasts m 10,000 times with replacement. The null hypothesis is that the score is zero or negative, i.e., less than the climatological reference forecast. A score is considered significantly more skillful (at the 5% level) than the reference if 95% of the resampled distribution is greater than zero.

3.3. Fractions extreme forecast index

The skill scores introduced in the previous sections quantify the accuracy of past forecasts. However, for these insights to inform real-time decision-making, users require a method to apply them operationally. Roberts and Lean (2008) proposed post-processing forecasts via spatial aggregation, enabling users to optimize the balance between accuracy and precision based on a past assessment of forecast skill. A key limitation of spatial aggregation is that it reduces the amplitude of raw forecast values, making them less intuitive for users accustomed to working with unprocessed data. To address this, we propose normalizing the aggregated forecast by an aggregated reference, similar to the procedure to define the Extreme Forecast Index (EFI, Lalaurette, 2003). The EFI measures how extreme a probabilistic forecast is relative to its climatology by comparing the cumulative distributions of the forecast and its corresponding hindcast, and is operationally employed by the ECMWF. In this subsection, we outline the computation of the EFI and then detail our modifications to enhance its applicability for end-users.

To compute the EFI, threshold values are first defined for each quantile q varying from $0 < q < 1$ in steps of Δq. The hindcast threshold $h_{thresh}(q,m,t,i,j)$ is determined as the value of the qth quantile from a sample of e ensemble members and y hindcast years in hindcast $h(m,e,y,t,i,j)$. Then, for each forecast $f(m,e,t,i,j)$, the fraction of ensemble members below the hindcast threshold $h_{thresh}(q,m,t,i,j)$ is computed, called $fr_f(q,m,t,i,j)$. The EFI is derived by summing the difference between the quantile q and the corresponding forecast fractions $fr_f(q,m,t,i,j)$ across all quantiles, normalized by $q(1-q)$, and multiplied by the quantile step Δq (Eq. (12)).

$$EFI(m,t,i,j) = \frac{2}{\pi}\sum_{q=0}^{1}\frac{q - fr_f(q,m,t,i,j)}{q(1-q)}\Delta q \tag{12}$$

EFI values range from −1 to 1, where −1 indicates that the entire cumulative forecast distribution is below the cumulative hindcast distribution, and +1 indicates it is entirely above. An $|EFI| > 0.8$ typically signifies an extreme event (ECMWF, 2024c).

For comparison with the forecast, we define an analogous Extreme Verification Index (EVI) using the verification $v(m,t,i,j)$, verification hindcast $v_H(m,y,t,i,j)$, verification hindcast threshold $v_{h_{thresh}}(q,m,t,i,j)$ and verification fraction $fr_v(q,m,t,i,j)$ in Eq. (13). The verification hindcast threshold is determined as the value of the qth quantile from a sample of y hindcast years in the verification hindcast, and the verification fraction is set to 1 if the verification lies above the threshold and to 0 if it lies below.

$$EVI(m,t,i,j) = \frac{2}{\pi}\sum_{q=0}^{1}\frac{q - fr_v(q,m,t,i,j)}{q(1-q)}\Delta q \tag{13}$$

The Fractions Extreme Forecast Index (FEFI) is computed in the same way as the original EFI except it utilizes the aggregated forecasts $F(n,m,e,t,i,j)$, hindcasts $H(n,m,e,y,t,i,j)$, hindcast threshold $H_{thresh}(n,q,m,t,i,j)$ and forecast fraction $FR_F(n,q,m,t,i,j)$ with the additional aggregation dimension n (Eq. (14)).

$$FEFI(n,m,t,i,j) = \frac{2}{\pi}\sum_{q=0}^{1}\frac{q - FR_F(n,q,m,t,i,j)}{q(1-q)}\Delta q \tag{14}$$

Thus, the FEFI quantifies how extreme the forecast is relative to its climatology across different spatial scales.

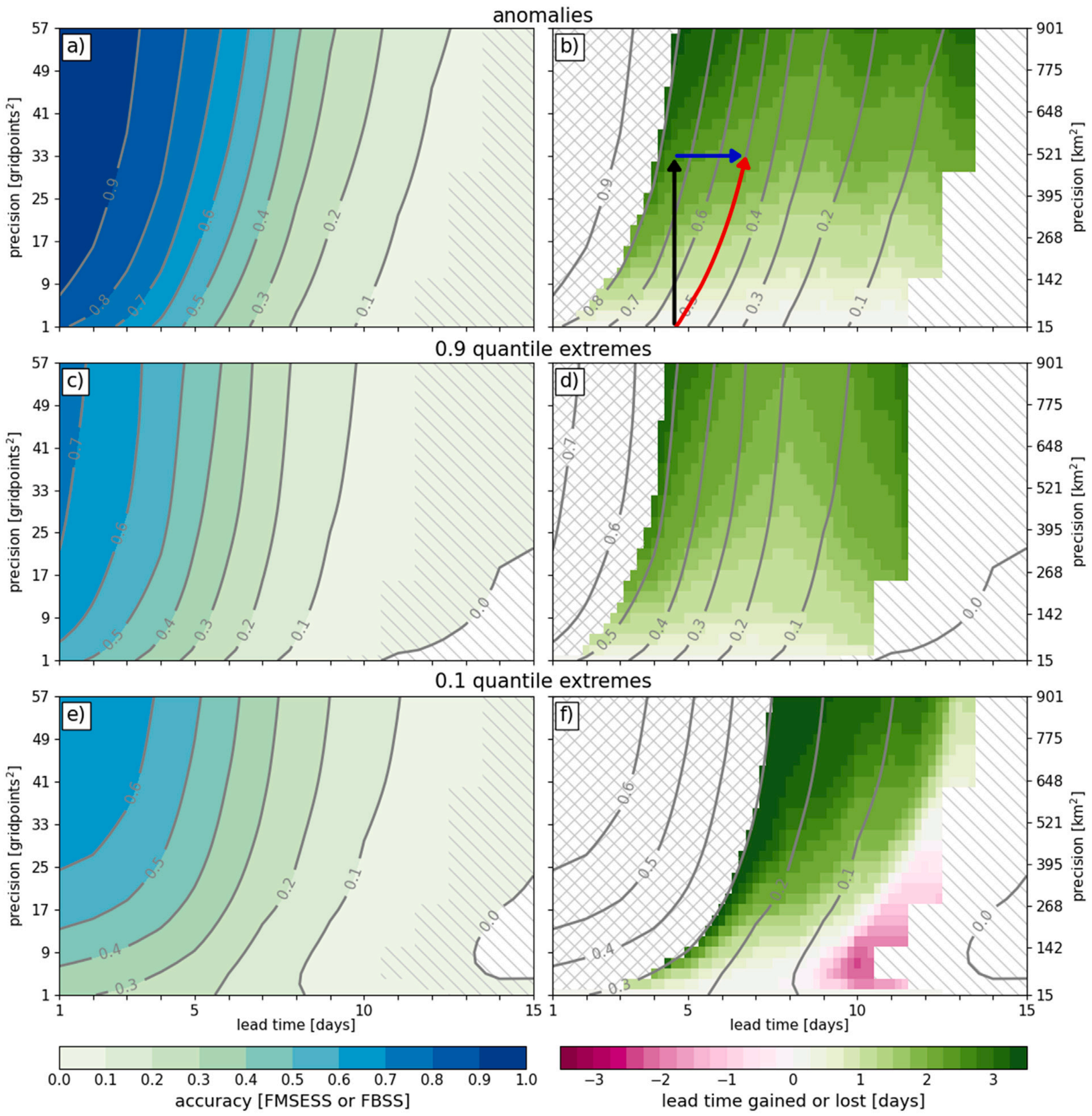

Fig. 1. Forecast accuracy and lead-time gained for daily-accumulated precipitation over Europe as a function of lead time and spatial precision. Left panels show the Fractions Mean-Square Error Skill Score (FMSESS) for anomalies (a) and the Fractions Brier Skill Score (FBSS) for 0.9 and 0.1 quantile extremes in (c) and (e), respectively. Hatching indicates skill scores that are not statistically significant at the 5% level assessed via bootstrapping. Cross-hatching denotes scores surpassing the highest accuracy obtained at the grid scale. Right panels follow the same conventions as the left ones, except that shading represents the lead time gained or lost by spatially aggregating the forecasts. Black, blue and red arrows in (b) illustrate the lead-time gained when spatially aggregating an example forecast at the grid-scale. See Section 4 for further details.

4. Quantifying accuracy versus precision in European precipitation forecasts

In this section, we evaluate the impact of spatial aggregation on sub-seasonal precipitation forecasts. We calculate FMSESS and FBSS for daily and weekly-accumulated precipitation over the entire European domain, examine their regional variations, and compare the results with those for 2-m temperature forecasts. By quantifying the trade-off between spatial accuracy and precision, we offer a clearer understanding of how spatial aggregation influences forecast performance.

Spatial aggregation improves the accuracy of daily-accumulated precipitation forecasts. Fig. 1a shows the FMSESS for daily-accumulated precipitation anomalies across various lead times and spatial scales. At the grid-scale, accuracy is high initially (> 0.8) but decreases with lead-time (< 0.1), with forecasts remaining skillful for up to 10 days (in agreement with Rivoire et al., 2023). Spatial aggregation not only increases accuracy for a given lead time but also extends the forecast skill horizon (as pointed out by Buizza et al., 2015; Buizza and Leutbecher, 2015), indicated by the right-slanted skill contours and significance hatching. This approach also improves accuracy for extreme precipitation, measured by the FBSS for the 0.9 and 0.1 quantiles, although forecasting extremes is generally less accurate than forecasting anomalies (compare Fig. 1a and 1c,e). Notably, low precipitation extremes can show reduced accuracy with spatial aggregation (left-slanted contours close to the grid-scale for lead times greater than 8 days in Fig. 1e). Near the grid-scale, forecasts are more accurate for low precipitation thresholds due to the model's tendency to predict no precipitation. Greater spatial aggregation raises the likelihood of non-zero precipitation thresholds, making the predictions more challenging and reducing accuracy.

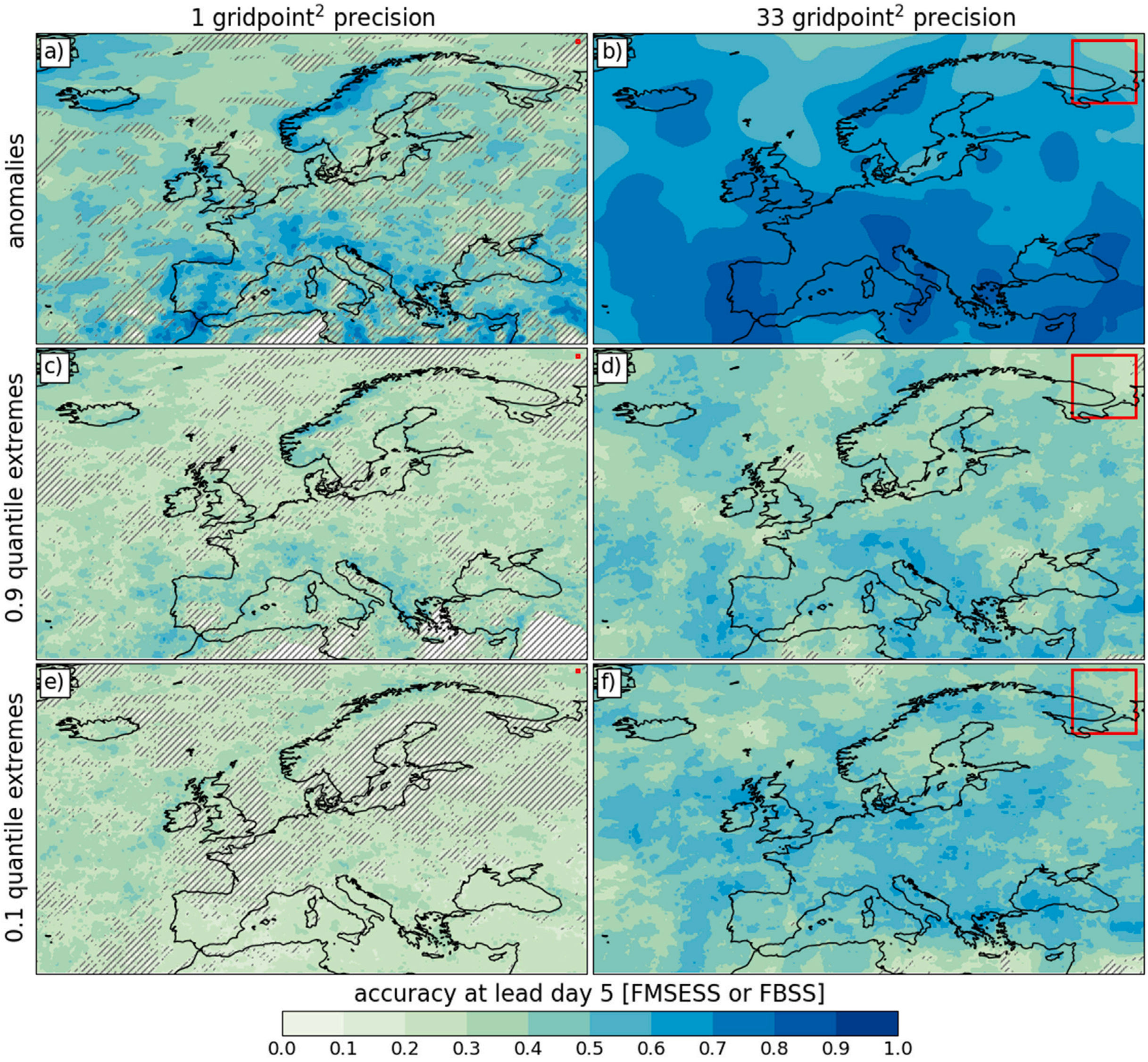

Fig. 2. Forecast accuracy at lead-day 5 for daily-accumulated precipitation over Europe, evaluated at two levels of spatial precision: the grid-scale (left) and 33 gridpoints^2 (right). Shading denotes the Fractions Mean-Square Error Skill Score (FMSESS) for anomalies in (a,b) and the Fractions Brier Skill Score (FBSS) for the 0.9 and 0.1 quantile extremes in (c,d) and (e,f), respectively. Hatching indicates skill scores that are not statistically significant at the 5% level assessed using bootstrapping. Red squares in the upper-right corner of the panels denote the spatial scale of forecast precision.

Reducing precision can extend predictable lead-times of daily-accumulated precipitation by a few days. The green shading on the right-hand side panels of Fig. 1 illustrates the lead-time gained by spatial aggregation. For example, a forecast with a precision of 1 grid point^2 and an accuracy of 0.5 that is spatially aggregated to a precision of 33 grid points^2 represents an accuracy gain of 0.25 (black arrow). This increase in accuracy is equivalent to gain in 2 lead-time days (blue arrow) because the aggregated forecast drops to the same level of accuracy as the grid-scale forecast two days later (red arrow). For low precipitation extremes, reducing precision can also lead to a loss of lead-time relative to the grid-scale (pink shading, Fig. 1f), though this mainly occurs where forecast accuracy is low (< 0.2).

The highest levels of forecast accuracy are only achievable with spatial aggregation. Cross-hatching in Fig. 1b,d,e shows where forecast accuracy exceeds the maximum achievable accuracy at the grid scale, i.e., lead-time $t = 1$. Gains from spatial aggregation are substantial for anomalies (0.8 to 0.95) and high precipitation extremes (0.5 to 0.7), and even greater for low precipitation extremes (0.3 to 0.7). Overall similar results are found for winter and summer only forecasts, where winter generally exhibits higher accuracy (Figs. S1 and S2).

Spatial aggregation also improves regional forecast accuracy of daily-accumulated precipitation. Fig. 2 shows latitude-longitude maps of forecast accuracy for daily-accumulated precipitation at lead-day 5, comparing two spatial precision levels: the grid scale (left) and 33 grid points^2 (right). The FMSESS and FBSS are calculated regionally at each latitude and longitude by omitting the domain average in Eqs. (7) and (10). It is important to note that the spatially aggregated forecasts are displayed with the same grid spacing as the raw forecasts (e.g., 0.25° × 0.25°), but their *effective* spatial resolution is reduced since each grid point represents the aggregate of its neighboring grid points. At the grid scale, forecast accuracy varies regionally, with higher accuracy over mountainous regions like western Norway and the Alps, and is higher for anomalies than extremes (Fig. 2 left). Spatial aggregation increases overall forecast accuracy and extends the forecast horizon, as indicated by the darker shading and reduced hatching in the right-hand versus left-hand panels of Fig. 2. This suggests that the European domain-averaged results in Fig. 1 generally hold regionally. Similar patterns are observed for different lead times and seasons (not shown).

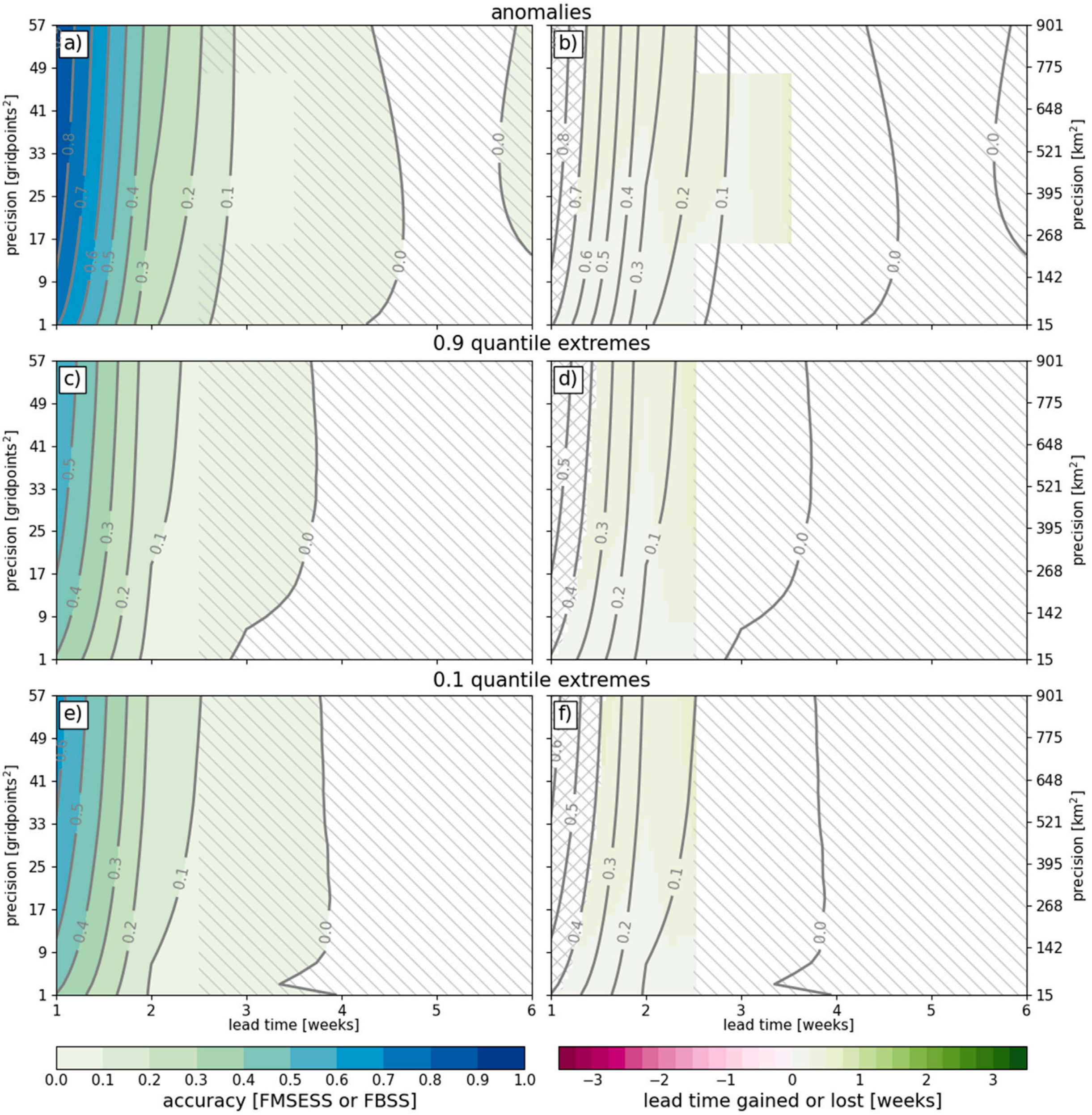

Fig. 3. As in Fig. 1 except for weekly-accumulated precipitation forecasts.

When measured per unit lead-time, spatial aggregation benefits weekly-accumulated precipitation forecasts less than daily-accumulated ones. Fig. 3 shows forecast accuracy for weekly-accumulated precipitation anomalies and extremes across various lead-times and spatial scales. At the grid-scale, forecast accuracy is high in the first week, low in the second, and as skillful as climatology in the third or fourth, consistent with daily-accumulated precipitation forecasts (compare Figs. 3 and 1, left). However, the improvements from spatial aggregation for weekly-accumulated forecasts are marginal compared to those for daily-accumulated forecasts when assessed per unit lead-time: 0.2–0.4 weeks versus 2–3 days (compare the slanted grey accuracy contours in Figs. 3 and 1, right). Regionally, spatial aggregation modestly improves accuracy and extends the forecast skill horizon for anomalies, but less so for extremes (compare left and right-hand panels of Fig. 4).

Surface temperature forecasts, while generally more accurate than precipitation forecasts, benefit less from spatial aggregation. Figure S3 shows forecast accuracy for daily-mean temperature anomalies and extremes across various lead times and spatial scales. Forecast accuracy remains similar with spatial aggregation, in contrast to precipitation, as indicated by the more vertical skill contours and shorter lead-time gains relative to the grid-scale (shading, compare Fig. S3 with Fig. 1). Weekly averaged temperature anomalies and extremes display similar traits, with a forecast skill horizon of 3 weeks across all spatial scales (Fig. S4). More spatially homogeneous temperature fields compared to precipitation fields result in more accurate forecasts at smaller scales, diminishing the benefits of spatial aggregation.

In summary, our results highlight and quantify the fundamental trade-off between accuracy and precision in sub-seasonal precipitation forecasts. Spatial aggregation, which reduces precision, increases forecast accuracy, extends predictable lead times, and enhances maximum possible accuracy compared to the grid scale. Conversely, increased precision tends to diminish these benefits. This trade-off is more important at higher temporal precision (e.g., daily versus weekly aggregation), and for spatially inhomogeneous variables (e.g., precipitation versus temperature). It is important to note that our results are based on

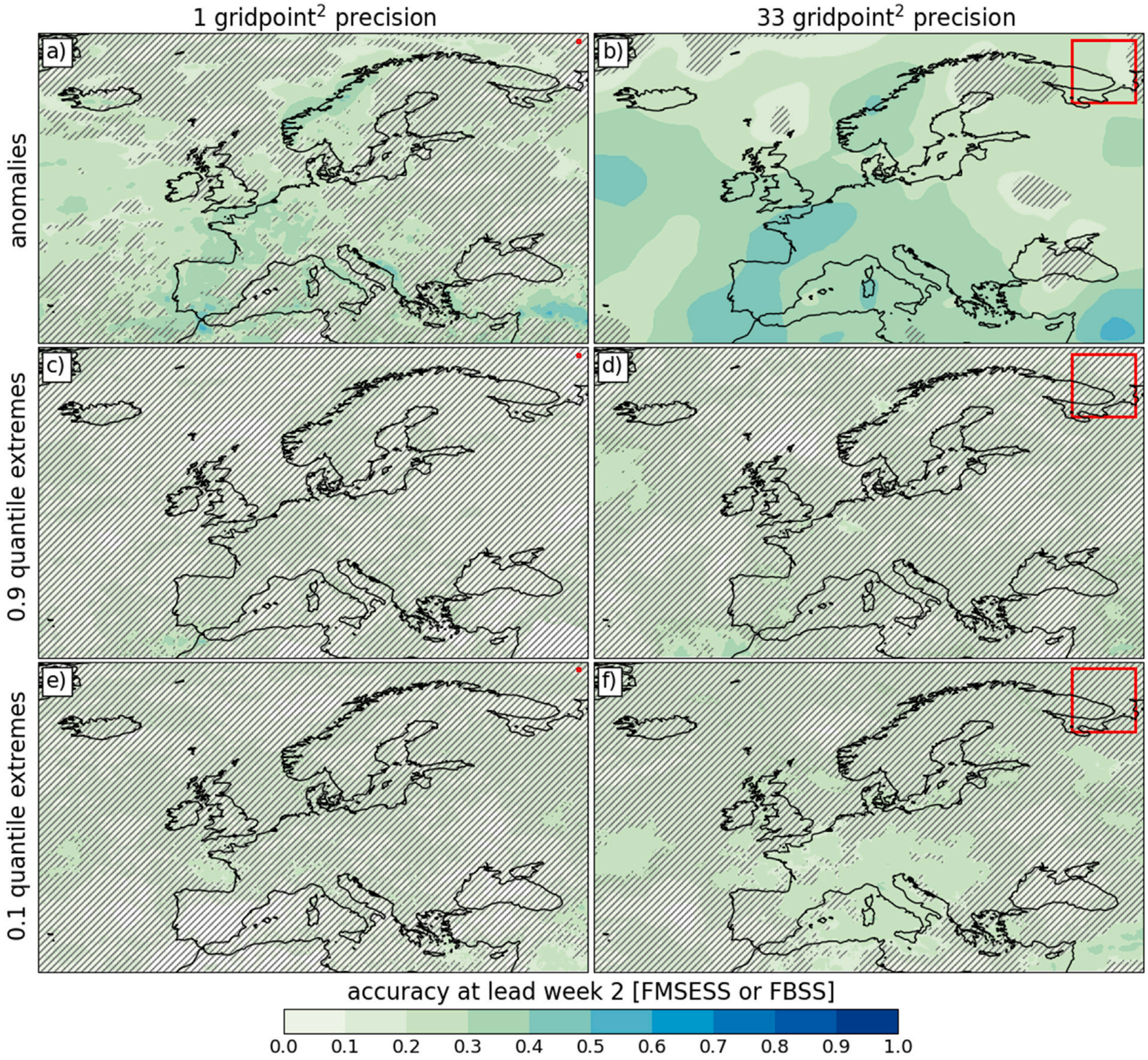

Fig. 4. As in Fig. 2 except for weekly-accumulated precipitation forecasts at lead-week 2.

averages over hundreds of forecasts, but there are windows of opportunity where individual forecasts can predict more accurately and further ahead (Mariotti et al., 2020). While these findings are well-recognized within the meteorological community (Buizza and Leutbecher, 2015; Toth and Buizza, 2019), they are often underappreciated by users who could benefit from them, even potentially leading to the misuse of forecasts (Nissan et al., 2019; Fiedler et al., 2021). In the next section, we demonstrate how users can leverage these findings with three practical examples.

5. Use-cases

In this section, we illustrate how spatial aggregation can make forecasts more usable. Roberts and Lean (2008) proposed post-processing real-time forecasts via spatial aggregation, optimizing spatial precision based on the accuracy of historical forecasts. They envisioned a coarse forecast at longer lead-times, becoming finer as the forecast horizon shortens and predictability increases at smaller scales. However, they did not provide a practical demonstration.

Here, we take their idea further and present three different use-cases for spatial aggregation, schematically illustrated in Fig. 5: optimized accuracy (blue arrow), fixed spatial precision (red arrow), and fixed lead time (black arrow). These use cases are tailored to a smaller Scandinavian domain (see Fig. 6a) and illustrate how the approach can be employed strategically to optimize forecasts given specific user-defined constraints. Rather than replacing existing practices using grid-scale forecasts, our approach offers a complementary perspective, and suggests avenues for further investigation in each of the three examples presented.

5.1. Optimized accuracy

Forecast accuracy often constrains how far ahead decisions can be made. This is particularly true for forecasters, who need to issue early-warnings at extended lead times. Using Storm Hans as an example, which first struck Norway on August 7th 2023, we illustrate how spatial forecast aggregation can give forecasters an earlier indication of extreme precipitation and thus help them issue more timely early warnings.

Fig. 6a–f shows forecasts of ensemble and daily-mean precipitation on August 7th at various lead times, comparing grid-scale precision (left) and spatially aggregated forecasts (right), while Fig. 6g shows the corresponding observed precipitation at the grid-scale. Spatial aggregation is progressively increased with lead time to increase forecast

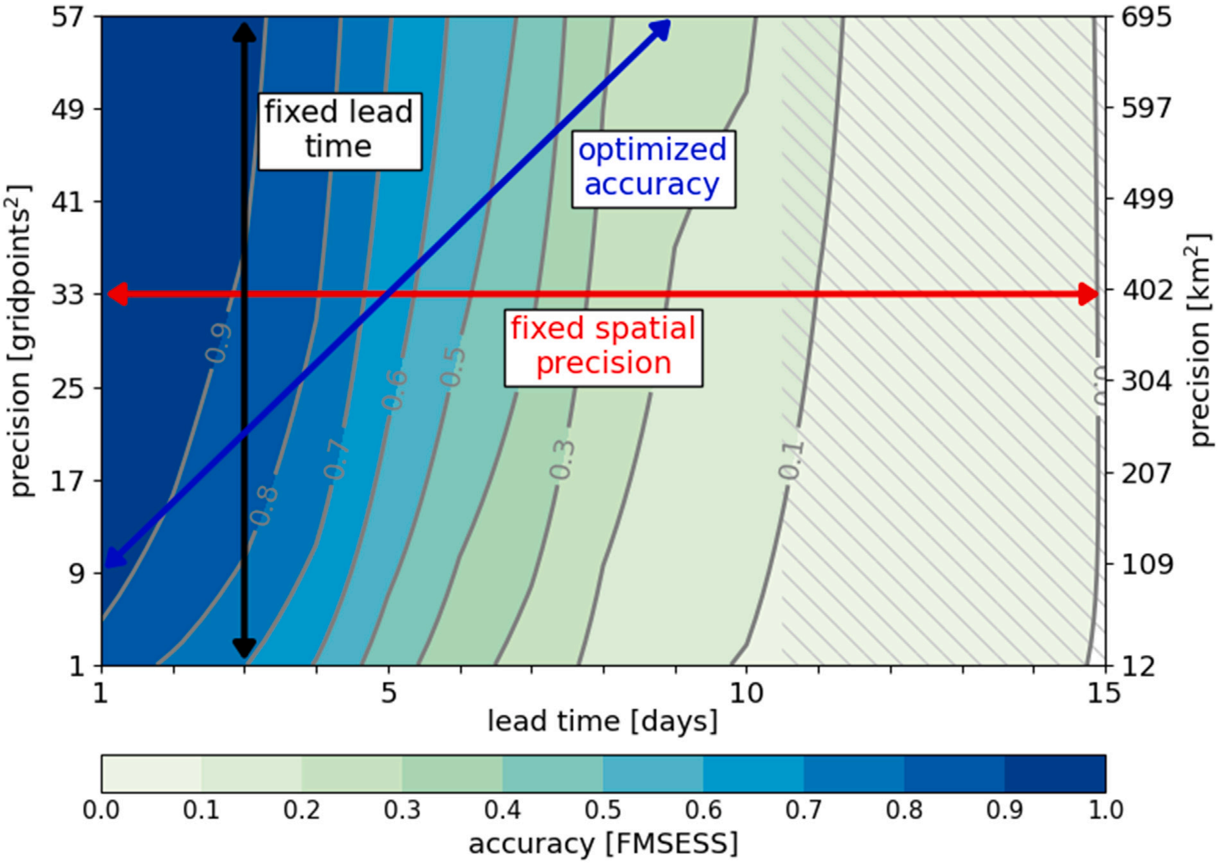

Fig. 5. As in Fig. 1a except for the Fractions Mean-Square Error Skill Score (FMSESS) over the Scandinavian domain shown in Fig. 6. Black, blue and red arrows denote the use-cases described in Section 5.

accuracy (blue arrow in Fig. 5). Spatial aggregation could be used to maintain a constant forecast accuracy by following along a specific contour (e.g., 0.7), but since these do not span several lead-time days, here we 'optimize' accuracy by aggregating diagonally across contours of constant accuracy. Regions with FEFI and EVI values above 0.8, marked by red stippling, highlight areas of extreme precipitation in the forecasts and observations, respectively.

At the grid scale, forecasts capture the general pattern of precipitation over eastern Norway and Sweden, with the FEFI signaling extreme precipitation reasonably well up to lead day 3 (see Fig. 6a,c,e with g). Spatially aggregated forecasts at lead days 1, 3 and 5 achieve accuracy comparable to grid-scale forecasts at lead day 1, as demonstrated by FMSESS and FBSS$_{0.9}$ values (contrast Fig. 6a,c,e with b,d,f). Most importantly, it is difficult to infer from the grid-scale forecast alone that localized extreme rainfall at lead day 5 would evolve into a widespread event across Scandinavia (compare red stippling Fig. 6a and Fig. 6g). In contrast, the spatially aggregated forecast at lead day 5 offers a clearer and earlier indication of the approaching large-scale extreme event (compare red stippling Fig. 6b and Fig. 6g). This is because spatial aggregation filters out small-scale noise and amplifies the larger-scale, predictable signal at longer lead times. Similar results are found up to lead day 7 (Fig. S5).

The Norwegian Meteorological Office issued a red alert for extreme rainfall on August 6th, just one day before Storm Hans struck Norway (Granerød et al., 2023). The decision to issue such alerts is guided by stringent internal procedures, balancing the need for timely warnings against the risk of false alarms. Our findings suggest that incorporating spatial aggregation into existing forecasting workflows could extend the lead time of early warnings for high-impact events like Storm Hans.

5.2. Fixed spatial precision

Forecast users are often constrained by the spatial scales at which they operate. In hydropower, for instance, the size of the watershed dictates both the volume of incoming precipitation and the timing of downstream impacts. When reservoirs near full capacity, Norwegian operators are often forced to discharge water in anticipation of rainfall events (NRK, 2020; TV2, 2024) to avert flooding and infrastructure damage. However, such preemptive measures can be costly since the water could be used to generate higher-priced electricity at a different time. Leveraging the watershed's size to refine precipitation forecasts could help operators optimize decision-making.

By spatially aggregating forecasts to match the watershed's spatial extent (red arrow in Fig. 5), operators could enhance forecast accuracy at the scale that matters most for their decisions. This refinement could enable hydropower managers to decide sooner, and with increased certainty, when and how much to discharge water. Crucially, this approach goes beyond simply spatially averaging the raw forecast over the watershed, since it maintains the default grid-spacing but each grid cell aggregates data from its surrounding neighbors (e.g., Fig. 2). As a result, the method incorporates the spatial uncertainty of precipitation that could occur near but not necessarily over the watershed at longer lead times (e.g., as illustrated by Storm Hans in Fig. 6), thereby extracting a stronger predictable signal from the forecast. This added granularity yields a more nuanced perspective of potential rainfall that could ultimately feed into the basin.

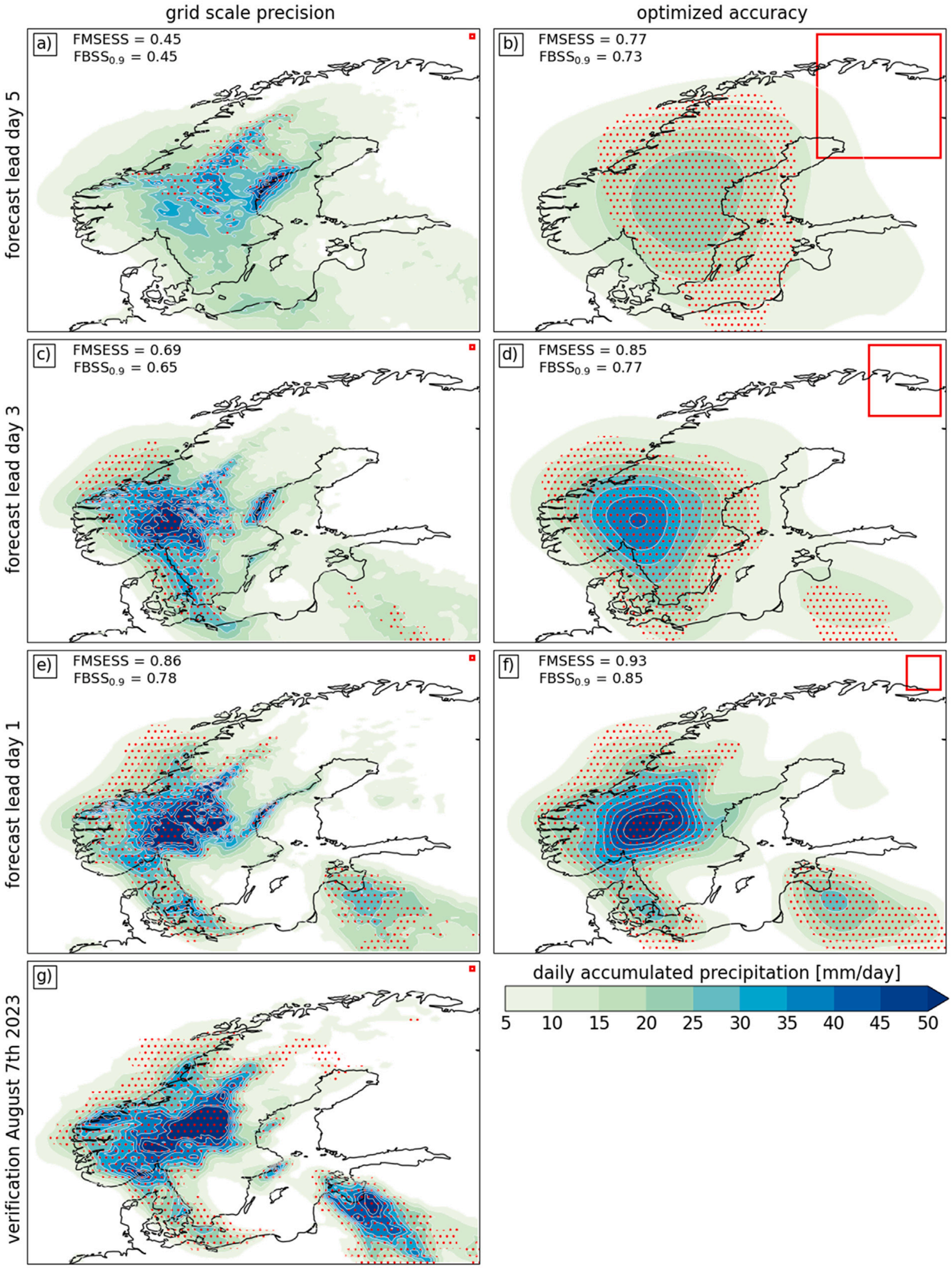

Fig. 6. (a-f) Forecasts of storm Hans on August 7th 2023. Shading denotes the ensemble-mean daily-accumulated precipitation for lead days 5 (a,b), 3 (c,d) and 1 (e,f), shown at the grid-scale (left) and with progressively increasing levels of spatial precision (right). Red squares in the upper-right corner of the panels denote the spatial scale of forecast precision. The FMSESS and FBSS for 0.9 quantile extremes are displayed in the top left hand corner of each panel. (g) Daily-accumulated precipitation during Storm Hans on August 7th 2023 according to data from the ERA5 reanalysis. Red stippling denotes Fractions Extreme Forecast Index (FEFI, panels a-f) and Extreme Verification Index (EVI, panel g) values greater than 0.8, signaling an extreme event.

5.3. Fixed lead-time

Fixed operational lead times can constrain how weather forecasts are applied. In commercial aviation, for instance, large-scale mid-latitude storms, such as Storm Hans (Fig. 6), often disrupt flight schedules across multiple airports and regions over extended periods. Because major carriers maintain extensive route networks on tight schedules, they often decide two to three days in advance whether to cancel, reroute or maintain flights given incoming severe weather (NBC, 2011; nytimes, 2017). These early go/no-go choices are critical for allocating resources and enabling passengers to arrange alternative travel plans. Delaying such arrangements significantly heightens the risk of disruptions for passengers and additional operating costs.

Aggregating forecasts at broader spatial scales offers a practical means of increasing their accuracy at these fixed lead times (black arrow in Fig. 5). As demonstrated by the forecast maps in Fig. 6, airlines could synthesize forecasts across larger geographic areas to gain a more robust sense of whether a major winter system will develop at the time when critical decisions need to be made. Although this aggregated approach drops fine-grained precision at individual airports, it can strengthen confidence in the storm's overall footprint. This would facilitate timely and decisive operational adjustments that span the geographically extensive networks of major carriers.

6. Conclusions and discussion

Despite an abundance of available forecast data, much of it remains underutilized, pointing to a critical usability gap (Lemos et al., 2012; Van den Hurk et al., 2018; Findlater et al., 2021). In this study, we highlight and quantify a crucial yet often overlooked challenge to their practical use: forecasts are often more precise than they are accurate when they are presented on a denser grid spacing than the scales they can accurately predict. This mismatch stems from the loss of accuracy at smaller spatial scales as forecasts extend further out in time (Lorenz, 1969; Toth and Buizza, 2019): a constraint known as the forecast skill horizon (Buizza et al., 2015; Buizza and Leutbecher, 2015). While many meteorologists recognize this trade-off, non-expert users may remain unaware, risking both the missed potential of predictability at larger scales and over-interpretation at finer scales. As forecasts become increasingly important to support climate adaptation and preparedness, users and providers can benefit from recognizing and accounting for the trade-off between forecast accuracy and precision.

It is informative to consider this limitation through the lens of Murphy's (1993) classic framework for "good" forecasts, which comprises three key measures: correspondence between the forecaster's judgment and the delivered forecast (consistency), how well the forecast corresponds to observed conditions (quality), and the practical benefit of the forecast to users (value). Although much attention focuses on how forecast quality impacts value, we argue that issuing forecasts at the default grid spacing, even when forecasters recognize diminished accuracy at those scales, reduces consistency. This lack of consistency, in turn, diminishes value: users unaware of the limitation may make suboptimal decisions, whereas those who are aware must contend with greater complexity in using the forecast. Indeed, Murphy (1993) anticipated this problem, noting that "forecasts and judgments may be inconsistent …in terms of their spatial and/or temporal specificity", and called for practical solutions to address such mismatches.

Here, we modified the original fractions skill score to help users balance the trade-off between forecast precision and accuracy, transforming this metric from its traditional role in verifying spatial forecast accuracy for meteorologists into a tool for optimizing forecasts for end-users. We applied this approach to daily European precipitation forecasts, quantifying the balance for both deterministic predictions of anomalies and probabilistic predictions of extremes, using three years of sub-seasonal data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Our results show that decreasing precision through spatial aggregation increases forecast accuracy, extends predictable lead times, and enhances the maximum possible accuracy relative to the grid scale, while increased precision diminishes these benefits.

We hope that users will employ our approach to optimize forecasts for their specific application. Implementing this involves: (1) identifying the geographic region of interest, (2) verifying past forecasts with a chosen metric, and (3) aggregating real-time forecasts accordingly. We demonstrated the practical value of our approach in three contexts: communicating early warnings, managing hydropower capacity, and commercial aviation planning—each characterized by distinct user-constraints on accuracy, spatial scale, or lead-time. These use-cases showed that focusing on the scales where forecasts are most accurate, rather than the default grid-scale, can offer users more actionable information. Instead of replacing existing practices, our approach offers a complementary perspective, and highlights multiple avenues for further investigation in each of the three examples.

Aggregating forecasts is a well-established practice, yet its implementation is often done by forecast providers rather than end-users. For example, the "ready-set-go" framework (Goddard et al., 2014) links forecasts across timescales to general preparedness levels, from monthly seasonal predictions (ready), to weekly sub-seasonal forecasts (set), and finally to daily weather forecasts (go), with each system having finer spatial and temporal precision as the forecast window shortens. Another approach is to filter forecasts into a few distinct large-scale patterns that are more predictable, called weather regimes (Michelangeli et al., 1995), as done operationally by ECMWF. Both these approaches implicitly involve aggregation, but the scales are set by either the modeling system or the regime classification, not the users. Our approach builds on these approaches but takes a step further, giving the user the ability to tailor forecasts according to their desired spatial scales.

Machine learning models hold considerable promise to improve both forecast accuracy and precision, potentially narrowing the usability gap (Eyring et al., 2024). However, these new approaches are likely not a panacea, and are subject to similar physical constraints which limit conventional models (Ben Bouallègue et al., 2024). It is widely acknowledged that there are fundamental limits to the forecast skill horizon (Lorenz, 1969; Palmer et al., 2014), and emerging evidence suggests that machine learning-based forecasts are not exempt from this constraint (Keane et al. 2025, *Mid-latitude versus tropical scales of predictability and their implications for forecasting*, in review). So, even if machine learning models produce ever finer-scale forecasts, users are still likely to face the challenge posed by the forecast skill horizon, and to require methods to deal with it, as discussed here.

Our findings show that spatial aggregation enhances the accuracy of daily-accumulated precipitation forecasts to a greater extent than weekly-accumulated ones, per unit lead time. This suggests that temporal aggregation compensates to some extent for spatial aggregation, which has implications for how forecasts are used and communicated. For example, sub-seasonal forecasts are often presented as weekly aggregates (https://charts.ecmwf.int/), leveraging time averaging to enhance accuracy at extended lead times. Correspondingly, intuition based on the forecast skill horizon suggests that these predictions should be interpreted on spatial scales larger than the grid-scale. However, our results for both precipitation and temperature show weekly-aggregated forecasts have similar accuracy across spatial scales (Fig. 3 and S4). Thus, the decision not to spatially aggregate these forecasts is sound; grid-scale forecasts appear to be more usable than we expected, provided temporal aggregation is applied.

Further refinements to our approach could enhance its usability and broaden its relevance. For instance, employing alternative fractions skill scores, such as those proposed by Necker et al. (2024), may better assess probabilistic spatial forecast accuracy. Accounting for geographic variations in forecast skill, as shown in Fig. 2, would further improve adaptability across regions and applications. While refining the domain to a smaller area of interest is a straightforward solution, other approaches are possible, such as applying different levels of spatial aggregation in different regions of the domain as suggested by Roberts and Lean (2008). Beyond accuracy, incorporating other metrics of forecast quality, such as reliability and discrimination, could be more relevant to users (Murphy, 1993; Weisheimer and Palmer, 2014). In practice, it is possible to quantify these metrics on the same axes as Fig. 5, swapping out forecast accuracy for alternatives. Finally, the fractions extreme forecast index could be further modified to weight the most extreme quantiles, changing its sensitivity to severe events.

Our results suggest that spatially aggregating weather forecasts could enhance the accuracy of downstream impact models, which are increasingly important for climate adaptation and preparedness (Merz et al., 2020). Optimizing their inputs might be simpler and more effective than improving their basic design. Implementing this approach, however, poses technical challenges. Physics-based impact models, such as those producing hydrological forecasts, require physical consistency between input variables (e.g., rainfall and temperature) which is disrupted by aggregation. On the other-hand, data-driven impact models, which are not constrained by physical laws and are being used in a variety of sectors including insurance, agriculture, and even hydrology, could be trained to capture the relationship between aggregated weather forecasts and impacts. Exploring the feasibility of this approach could be an interesting avenue for further research.

CRediT authorship contribution statement

Etienne Dunn-Sigouin: Writing – original draft, Validation, Project administration, Investigation, Data curation, Writing – review & editing, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Erik W. Kolstad:** Writing – review & editing. **C. Ole Wulff:** Writing – review & editing. **Douglas J. Parker:** Writing – review & editing. **Richard J. Keane:** Writing – review & editing.

Data statement

Sub-seasonal forecasts from ECWMF (ECMWF, 2024a) and ERA5 reanalysis data (Hersbach et al., 2023) are freely available from the MARS archive (https://apps.ecmwf.int/archive-catalogue/) and the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview), respectively. Materials to reproduce the figures in this paper are provided on Github (https://github.com/edunnsigouin/accuracyvsprecision).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT to assist with the writing of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

E.D.S, E.K. and O.W. were supported by the Research Council of Norway, Norway grant 309562 Climate Futures, and E.K. was also supported by the European Union grant 101137847 ACACIA. R.K. was supported by the Centre for Environmental Modelling and Computation (CEMAK) at the University of Leeds. D.P. was supported by the Norad-funded project RAF:23/006 Improving Smallholder Resilience through Customised Climate Services and the WISER Early Warnings for Southern Africa (WISER-EWSA) project, funded by the Met Office as part of the Weather and Climate Information Services (WISER) programme on behalf UK government's Foreign, Commonwealth and Development Office (FCDO). The computations were performed on resources provided by the Sigma2 NS9873K project, as part of the National Infrastructure for High-Performance Computing and Data Storage in Norway. E.D.S acknowledges Håkon Otneim, Sondre Holleland and Bjørnar Jensen for feedback on an early version of the manuscript and Thordis Thorarinsdottir for early discussions which contributed to the main idea for the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cliser.2025.100594.

Data availability

The links to the data and code are contained in the main manuscript and are freely accessible.

References

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525 (7567), 47–55. http://dx.doi.org/10.1038/nature14956.

Ben Bouallègue, Z., Clare, M.C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janoušek, M., Rodwell, M., Pinault, F., Dramsch, J.S., Lang, S.T., et al., 2024. The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. Bull. Am. Meteorol. Soc. 105 (6), E864–E883. http://dx.doi.org/10.1175/BAMS-D-23-0162.1.

Benjamin, S.G., Brown, J.M., Brunet, G., Lynch, P., Saito, K., Schlatter, T.W., 2019. 100 years of progress in forecasting and NWP applications. Meteorol. Monogr. 59, 1–13. http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1.

Buizza, R., Balmaseda, M.A., Brown, A., English, S., Forbes, R., Geer, A., Haiden, T., Leutbecher, M., Magnusson, L., Rodwell, M., et al., 2018. The development and evaluation process followed at ECMWF to upgrade the integrated forecasting system (IFS). European Centre for Medium Range Weather Forecasts, http://dx.doi.org/10.21957/xzopnhty9.

Buizza, R., Leutbecher, M., 2015. The forecast skill horizon. Q. J. R. Meteorol. Soc. 141 (693), 3366–3382. http://dx.doi.org/10.1002/qj.2619.

Buizza, R., Leutbecher, M., Thorpe, A., 2015. Living with the butterfly effect: A seamless view of predictability. ECMWF Newsl. 145, 18–23. http://dx.doi.org/10.21957/x4h3e8w3.

Cafaro, C., Woodhams, B.J., Stein, T.H., Birch, C.E., Webster, S., Bain, C.L., Hartley, A., Clarke, S., Ferrett, S., Hill, P., 2021. Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? Weather. Forecast. 36 (2), 697–716. http://dx.doi.org/10.1175/WAF-D-20-0172.1.

Coughlan de Perez, E., Harrison, L., Berse, K., Easton-Calabria, E., Marunye, J., Marake, M., Murshed, S.B., Zauisomue, E.-H., et al., 2022. Adapting to climate change through anticipatory action: The potential use of weather-based early warnings. Weather. Clim. Extrem. 38, 100508. http://dx.doi.org/10.1016/j.wace.2022.100508.

Cusick, D., 2019. Tech offers a virtual window into future climate change risk. URL: https://www.scientificamerican.com/article/tech-offers-a-virtual-window-into-future-climate-change-risk/, (Accessed 13 August 2024).

ECMWF, 2024a. Meteorological archival and retrieval system (MARS). URL: https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets, (Accessed: 13 August 2024).

ECMWF, 2024b. Section 5.3.1 M-climate, the ENS model climate. URL: https://confluence.ecmwf.int/display/FUG/Section+5.3.1+M-climate%2C+the+ENS+Model+Climate, (Accessed 16 August 2024).

ECMWF, 2024c. Section 8.1.9.2 extreme forecast index - EFI. URL: https://confluence.ecmwf.int/display/FUG/Section+8.1.9.2+Extreme+Forecast+Index+-+EFI, (Accessed: 16 August 2024).

Ekroll, H.C., 2023. Ekstremværet hans vil trolig koste langt mer enn storflommen i 1995. URL: https://www.aftenposten.no/norge/i/76w3oW/ekstremvaeret-hans-vil-trolig-koste-langt-mer-enn-storflommen-i-1995, (Accessed 13 August 2024).

European-Commission, 2020. EU taxonomy for sustainable activities. URL: https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities_en, (Accessed: 13 August 2024).

Eyring, V., Gentine, P., Camps-Valls, G., Lawrence, D.M., Reichstein, M., 2024. AI-empowered next-generation multiscale climate modelling for mitigation and adaptation. Nat. Geosci. 1–9. http://dx.doi.org/10.1038/s41561-024-01527-w.

Fiedler, T., Pitman, A.J., Mackenzie, K., Wood, N., Jakob, C., Perkins-Kirkpatrick, S.E., 2021. Business risk and the emergence of climate analytics. Nat. Clim. Chang. 11 (2), 87–94. http://dx.doi.org/10.1038/s41558-020-00984-6.

Findlater, K., Webber, S., Kandlikar, M., Donner, S., 2021. Climate services promise better decisions but mainly focus on better data. Nat. Clim. Chang. 11 (9), 731–737. http://dx.doi.org/10.1038/s41558-021-01125-3.

Gehne, M., Hamill, T.M., Kiladis, G.N., Trenberth, K.E., 2016. Comparison of global precipitation estimates across a range of temporal and spatial scales. J. Clim. 29 (21), 7773–7795. http://dx.doi.org/10.1175/JCLI-D-15-0618.1.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B., Ebert, E.E., 2009. Intercomparison of spatial forecast verification methods. Weather. Forecast. 24 (5), 1416–1430. http://dx.doi.org/10.1175/2009WAF2222269.1.

Goddard, L., 2016. From science to service. Science 353 (6306), 1366–1367. http://dx.doi.org/10.1126/science.aag3087.

Goddard, L., Baethgen, W.E., Bhojwani, H., Robertson, A.W., 2014. The international research institute for climate & society: why, what and how. Earth Perspect. 1, 1–14. http://dx.doi.org/10.1186/2194-6434-1-10.

Gong, X., Barnston, A.G., Ward, M.N., 2003. The effect of spatial aggregation on the skill of seasonal precipitation forecasts. J. Clim. 16 (18), 3059–3071. http://dx.doi.org/10.1175/1520-0442(2003)016<3059:TEOSAO>2.0.CO;2.

Graham, R.M., Browell, J., Bertram, D., White, C.J., 2022. The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting. Meteorol. Appl. 29 (1), e2047. http://dx.doi.org/10.1002/met.2047.

Granerød, M., Stabell, D., Mjelstad, H., Tajet, H., 2023. Ekstremværet'Hans', ekstremt mye nedbør i deler av Sør-norge 07.-09. August 2023. 26, The Norwegian Meteorological Institute (ed) METinfo.

Hanssen-Bauer, I., Drange, H., Førland, E., Roald, L., Børsheim, K., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S., Sorteberg, A., et al., 2009. Climate in Norway 2100. In: Background information to NOU Climate adaptation (In Norwegian: Klima i Norge 2100. Bakgrunnsmateriale til NOU Klimatilplassing). Norsk klimasenter, URL: https://klimaservicesenter.no/kss/rapporter/kin2100.

Haupt, S.E., Hanna, S., Askelson, M., Shepherd, M., Fragomeni, M.A., Debbage, N., Johnson, B., 2019a. 100 years of progress in applied meteorology. Part II: Applications that address growing populations. Meteorol. Monogr. 59, 1–23. http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-18-0007.1.

Haupt, S.E., Kosović, B., McIntosh, S.W., Chen, F., Miller, K., Shepherd, M., Williams, M., Drobot, S., 2019b. 100 years of progress in applied meteorology. Part III: Additional applications. Meteorol. Monogr. 59, 1–24. http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-18-0012.1.

Haupt, S.E., Rauber, R.M., Carmichael, B., Knievel, J.C., Cogan, J.L., 2018. 100 years of progress in applied meteorology. Part I: Basic applications. Meteorol. Monogr. 59, 1–22. http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-18-0004.1.

Hersbach, H., Bell, B., Berrisford, P., G., B., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2023. ERA5 hourly data on single levels from 1940 to present. Copernicus climate change service (C3S) climate data store (CDS). http://dx.doi.org/10.24381/cds.adbb2d47.

Hewitt, C., Mason, S., Walland, D., 2012. The global framework for climate services. Nat. Clim. Chang. 2 (12), 831–832. http://dx.doi.org/10.1038/nclimate1745.

Van den Hurk, B., Hewitt, C., Jacob, D., Bessembinder, J., Doblas-Reyes, F., Döscher, R., 2018. The match between climate services demands and earth system models supplies. Clim. Serv. 12, 59–63. http://dx.doi.org/10.1016/j.cliser.2018.11.002.

Jolliffe, I.T., Stephenson, D.B., 2012. Forecast Verification: a Practitioner's Guide in Atmospheric Science. John Wiley & Sons, http://dx.doi.org/10.1002/9781119960003.

Jung, T., Leutbecher, M., 2008. Scale-dependent verification of ensemble forecasts. Q. J. R. Meteorol. Soc. 134 (633), 973–984. http://dx.doi.org/10.1002/qj.255.

Keane, R.J., Plant, R.S., Tennant, W.J., 2016. Evaluation of the plant–craig stochastic convection scheme (v2. 0) in the ensemble forecasting system MOGREPS-R (24 km) based on the unified model (v7. 3). Geosci. Model. Dev. 9 (5), 1921–1935. http://dx.doi.org/10.5194/gmd-9-1921-2016.

KLP, K.L.G.F., 2023. Dette læ rte kommunene av ekstremvæ ret hans. URL: https://www.klp.no/artikler/dette-laerte-kommunene-av-ekstremvaeret-hans, (Accessed: 13 August 2024).

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

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al., 2023. Learning skillful medium-range global weather forecasting. Science 382 (6677), 1416–1421. http://dx.doi.org/10.1126/science.adi233.

Lavers, D.A., Simmons, A., Vamborg, F., Rodwell, M.J., 2022. An evaluation of ERA5 precipitation for climate monitoring. Q. J. R. Meteorol. Soc. 148 (748), 3152–3165. http://dx.doi.org/10.1002/qj.4351.

Lemos, M.C., Kirchhoff, C.J., Ramprasad, V., 2012. Narrowing the climate information usability gap. Nat. Clim. Chang. 2 (11), 789–794. http://dx.doi.org/10.1038/nclimate1614.

Lorenz, E.N., 1969. The predictability of a flow which possesses many scales of motion. Tellus 21 (3), 289–307. http://dx.doi.org/10.1111/j.2153-3490.1969.tb00444.x.

Mariotti, A., Baggett, C., Barnes, E.A., Becker, E., Butler, A., Collins, D.C., Dirmeyer, P.A., Ferranti, L., Johnson, N.C., Jones, J., et al., 2020. Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. Bull. Am. Meteorol. Soc. 101 (5), E608–E625. http://dx.doi.org/10.1175/BAMS-D-18-0326.1.

Merryfield, W.J., Baehr, J., Batté, L., Becker, E.J., Butler, A.H., Coelho, C.A., Danabasoglu, G., Dirmeyer, P.A., Doblas-Reyes, F.J., Domeisen, D.I., et al., 2020. Current and emerging developments in subseasonal to decadal prediction. Bull. Am. Meteorol. Soc. 101 (6), E869–E896. http://dx.doi.org/10.1175/BAMS-D-19-0037.1.

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D.N., Domeisen, D.I., Feser, F., Koszalka, I., Kreibich, H., et al., 2020. Impact forecasting to support emergency management of natural hazards. Rev. Geophys. 58 (4), http://dx.doi.org/10.1029/2020RG000704, e2020RG000704.

Michelangeli, P.-A., Vautard, R., Legras, B., 1995. Weather regimes: Recurrence and quasi stationarity. J. Atmos. Sci. 52 (8), 1237–1256. http://dx.doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

Murphy, A.H., 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather. Forecast. 8 (2), 281–293. http://dx.doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

NBC, 2011. Airlines cancel flights before snow flies. URL: https://www.nbcnews.com/id/wbna41502385, (Accessed 09 January 2025).

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., Serafin, S., 2024. The fractions skill score for ensemble forecast verification. Q. J. R. Meteorol. Soc. http://dx.doi.org/10.1002/qj.4824.

Nissan, H., Goddard, L., de Perez, E.C., Furlow, J., Baethgen, W., Thomson, M.C., Mason, S.J., 2019. On the use and misuse of climate change projections in international development. Wiley Interdiscip. Rev.: Clim. Chang. 10 (3), e579. http://dx.doi.org/10.1002/wcc.579.

NRK, 2020. Overfylte vannmagasin gir rekordlave strømpriser: – får betalt for å bruke strøm. URL: https://www.nrk.no/vestland/overfylte-vannmagasin-gir-rekordlave-strompriser_-_far-betalt-for-a-bruke-strom-1.15110896, (Accessed: 08 January 2025).

nytimes, 2017. Airlines take a proactive approach to potential weather woes. URL: https://www.nytimes.com/2017/03/14/business/airline-cancellations-delays-snowstorm.html, (Accessed: 09 January 2025).

O'Kane, T.J., Scaife, A.A., Kushnir, Y., Brookshaw, A., Buontempo, C., Carlin, D., Connell, R.K., Doblas-Reyes, F., Dunstone, N., Förster, K., et al., 2023. Recent applications and potential of near-term (interannual to decadal) climate predictions. Front. Clim. 5, 1121626. http://dx.doi.org/10.3389/fclim.2023.1121626.

Palmer, T., Döring, A., Seregin, G., 2014. The real butterfly effect. Nonlinearity 27 (9), R123. http://dx.doi.org/10.1088/0951-7715/27/9/R123.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T.R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., et al., 2024. Probabilistic weather forecasting with machine learning. Nature 637 (8044), 84–90. http://dx.doi.org/10.1038/s41586-024-08252-9.

Rivoire, P., Martius, O., Naveau, P., Tuel, A., 2023. Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over europe. Nat. Hazards Earth Syst. Sci. 23 (8), 2857–2871. http://dx.doi.org/10.5194/nhess-23-2857-2023.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Weather Rev. 136 (1), 78–97. http://dx.doi.org/10.1175/2007MWR2123.1.

Röösli, T., Appenzeller, C., Bresch, D.N., 2021. Towards operational impact forecasting of building damage from winter windstorms in Switzerland. Meteorol. Appl. 28 (6), e2035. http://dx.doi.org/10.1002/met.2035.

Schwartz, C.S., 2019. Medium-range convection-allowing ensemble forecasts with a variable-resolution global model. Mon. Weather Rev. 147 (8), 2997–3023. http://dx.doi.org/10.1175/MWR-D-18-0452.1.

Torralba, V., Doblas-Reyes, F.J., MacLeod, D., Christel, I., Davis, M., 2017. Seasonal climate prediction: a new source of information for the management of wind energy resources. J. Appl. Meteorol. Clim. 56 (5), 1231–1247. http://dx.doi.org/10.1175/JAMC-D-16-0204.1.

Toth, Z., Buizza, R., 2019. Weather forecasting: What sets the forecast skill horizon? In: Sub-Seasonal To Seasonal Prediction. Elsevier, pp. 17–45. http://dx.doi.org/10.1016/B978-0-12-811714-9.00002-4.

Trenberth, K.E., Marquis, M., Zebiak, S., 2016. The vital need for a climate information system. Nat. Clim. Chang. 6 (12), 1057–1059. http://dx.doi.org/10.1038/nclimate3170.

TV2, 2024. Ikke opplevd maken på 50 år: – ekstremt. URL: https://www.tv2.no/nyheter/innenriks/ikke-opplevd-maken-pa-50-ar-ekstremt/17173589/, (Accessed 08 January 2025).

Van Straaten, C., Whan, K., Coumou, D., van den Hurk, B., Schmeits, M., 2020. The influence of aggregation and statistical post-processing on the subseasonal predictability of European temperatures. Q. J. R. Meteorol. Soc. 146 (731), 2654–2670. http://dx.doi.org/10.1002/qj.3810.

Weisheimer, A., Palmer, T., 2014. On the reliability of seasonal climate forecasts. J. R. Soc. Interface 11 (96), 20131162. http://dx.doi.org/10.1098/rsif.2013.1162.

White, C.J., Domeisen, D.I., Acharya, N., Adefisan, E.A., Anderson, M.L., Aura, S., Balogun, A.A., Bertram, D., Bluhm, S., Brayshaw, D.J., et al., 2022. Advances in the application and utility of subseasonal-to-seasonal predictions. Bull. Am. Meteorol. Soc. 103 (6), E1448–E1472. http://dx.doi.org/10.1175/BAMS-D-20-0224.1.

WMO, 2022. WMO and the early warnings for all initiative. URL: https://wmo.int/activities/early-warnings-all/wmo-and-early-warnings-all-initiative, (Accessed 16 August 2024).

Young, M., Heinrich, V., Black, E., Asfaw, D., 2020. Optimal spatial scales for seasonal forecasts over africa. Environ. Res. Lett. 15 (9), 094023. http://dx.doi.org/10.1088/1748-9326/ab94e9.

Zhao, B., Zhang, B., 2018. Assessing hourly precipitation forecast skill with the fractions skill score. J. Meteorol. Res. 32 (1), 135–145. http://dx.doi.org/10.1007/s13351-018-7058-1.