

Original research article

Subseasonal predictions for climate services, a recipe for operational implementation

Andrea Manrique-Suñén^a, Lluís Palma^{a,*}, Nube Gonzalez-Reviriego^a,
Francisco J. Doblas-Reyes^{a,b}, Albert Soret^a

^a Barcelona Supercomputing Center (BSC), Carrer de Jordi Girona 29, 08034 Barcelona, Spain

^b Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain



ARTICLE INFO

Keywords:

Subseasonal climate forecasting
Climate services
Operational climate prediction
Climate adaptation
Energy
Agriculture

ABSTRACT

The implementation of operational climate service prototypes, which encompasses the co-design and delivery of real-time actionable products with/to stakeholders, contributes to efficiently leveraging operational climate predictions into actionable climate information by providing practical insight on the actual use of climate predictions. This work showcases a general guideline for implementing an operational climate service based on subseasonal predictions. At this timescale, many strategic decisions can benefit from timely predictions of climate variables. Still, the use of subseasonal predictions is not fully exploited. Here, we describe the key aspects considered to set up an operational climate service from the conception to the production phase. These include the choice of the subseasonal systems, the data sources and the methodology employed for post-processing the predictions. To illustrate the process with a real case, we present the detailed workflow design of the implementation of a climate service based on subseasonal predictions and describe the bias adjustment and verification methodologies implemented. This work was developed in the H2020 S2S4E project, where industrial and research partners co-developed a fully-operational Decision Support Tool (DST) providing 18 months of real-time subseasonal and seasonal forecasts tailored to the specific needs of the renewable energy sector. The operational workflow can be adapted to serve forecast products to other sectors, as has been proved in the H2020 vitiGEOS project, where the workflow was modified to provide downscaled subseasonal predictions to specific locations. We consider this a valuable contribution to future developments of similar service implementations and the producers of the climate data.

Practical implications

Climate services aim to improve society's resilience to climate change by providing useful information about climate to stakeholders and citizens. In recent years, subseasonal climate predictions, covering time ranges from one to several weeks into the future, have demonstrated to have real application capabilities in various strategic sectors (energy, agriculture, water management, disaster preparedness) as well as substantial skill in anticipating extreme events several weeks in advance (such as heatwaves, cold spells, heavy precipitation and cyclones), thus having potential use in decision making and in the activation of emergency measures (White et al., 2021; Domeisen et al., 2022). However, although the capabilities have been demonstrated in various studies, the operational use of such predictions in decision

making is not fully exploited. To provide a practical and realistic demonstration of the usability of the forecast products to users, the co-development and implementation of real-time operational prototypes is a valuable practice. Carrying out this exercise additionally allows the user to routinely test the adequacy of the forecast product in their own decision making context.

This work introduces a general framework for the conceptualisation and implementation of an operational service based on subseasonal predictions. We describe the main steps, from the forecast product definition, the prediction system selection, the available data portals gathering subseasonal forecasts, and tools for setting up, monitoring and testing an operational workflow. The characteristics of two data platforms which collect subseasonal predictions are discussed, namely the World Weather Research Programme (WWRP)/ World Climate Research

* Corresponding author.

E-mail address: lluis.palma@bsc.es (L. Palma).

<https://doi.org/10.1016/j.cliser.2023.100359>

Available online 3 February 2023

2405-8807/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Programme (WCRP) Subseasonal to Seasonal (S2S) Prediction Project database (Robertson et al., 2015) and the Subseasonal Experiment (SubX) (Pegion et al., 2019). These platforms are a valuable asset in the development of subseasonal climate services. Additionally, a practical example of an operational implementation is also presented. This system was developed in the H2020 S2S4E project (Soret et al., 2019) and is the backbone of the subseasonal forecasts products delivered in the project's Decision Support Tool (DST) (<https://s2s4e-dst.bsc.es>). The tool was in production for 18 months issuing real-time subseasonal and seasonal forecasts of essential climate variables and derived products to energy users. In terms of Technology Readiness Levels (TRLs), we estimated that the presented prototype reached a level of TLR7. For part of this period, the system employed forecast data provided by the S2S Real-Time Pilot Initiative. The subseasonal service consisted of weekly updates of forecasts for the four following weeks. The development of climate services covers several aspects, and it should be noted that this work focuses solely on the operational implementation of the real-time system for subseasonal forecasts. The process of identifying users' needs is not addressed here, nor is the design of the platform to visualise the final products. In this work, we present a framework that describes the design of the mechanism to routinely collect, store and post-process the data to generate the forecast product and the conceptualisation behind the pipeline design. The operational workflow presented as an example has been specifically designed to ingest output from the European Center for Medium Range Weather forecasts ECMWF-Ext-ENS system (Vitart et al., 2008). Additionally, since climate predictions at this timescale accumulate substantial systematic errors, we include a bias adjustment of the forecasts at the weekly timescale. Moreover, a quality assessment of the forecast products is also routinely performed. For these processes, hindcast data (forecasts issued in the past) and reanalysis data ERA5 (Hersbach et al., 2020) (as observational reference) are routinely downloaded and stored. The proposed technique for the bias adjustment is the variance inflation (Doblas-Reyes et al., 2005) employing a running window centred on the forecast day. The forecast quality assessment skills scores are computed to be delivered with the forecast product, providing the user with a notion of the quality of the data. Although it was initially conceived for energy, the proposed framework can be adapted to other product definitions, and additional post-processing steps (downscaling, impact indicators) can be incorporated due to its modularity. The same workflow has been adapted to deliver downscaled forecast products to the agricultural sector under the H2020 vitiGEOSS project. Similarly, even when the workflow has yet to be tailored to the user's needs, it can serve as a demonstrator to start conversations with them, making it easier to understand climate services based on subseasonal predictions, their potential capabilities, and their limitations.

Transforming the climate output from climate models into tailored climate information useful for decision-making requires rigorous scientific knowledge of the underpinning models and predictions, as well as a good understanding of the user necessities. The work presented is eminently practical and shows the non-trivial efforts behind an operational prototype that include tasks such as data curation, workflow design and quality checks. The implementation of an operational service prototype is a valuable exercise that supports the continued development of S2S forecasts and related services. The proposed guideline for conceptualising and implementing a downstream pipeline for operational production can be of great utility to climate service providers looking to incorporate S2S predictions. The proposed guideline can be adapted to different products, post-processing techniques or other prediction systems. Some points to consider when employing prediction systems of other characteristics are also discussed. Some direct implications of the implementation of the operational workflow in H2020 projects have been to raise awareness of the usability of subseasonal predictions in the energy and agricultural sectors. As previously mentioned, the framework can be transferred to other sectors that are still unaware of the potential use of subseasonal predictions. Additionally, the real-time operational service made available to users serves as a

test-bed for product validation, thus providing useful feedback to the services providers and the modelling centres.

Subseasonal predictions offer information in a very useful time scale, even with limited skill, there is valuable information that can guide informed decisions. The implementation of real-time prototypes will promote, inform and help improve the service. We, therefore, consider this a valuable experience for future developments of similar service implementations and for the producers of the climate data involved.

1. Introduction

Subseasonal predictions fill the gap between short-range weather forecasts and seasonal climate predictions, providing climate information from 10 days to approximately 2 months ahead. This time range has proved to be very valuable for many sectors such as energy, agriculture, water management or retail, since many strategic decisions are taken within this time frame (White et al., 2017). However, the use of subseasonal predictions by stakeholders is still limited. Advances in the uptake of subseasonal predictions by users will follow from increased skill at this time scale through improvements in the prediction systems (better representation of the physical processes and teleconnections (Merryfield, 2020)), but also from advances in the downstream side of the climate services chain (Buontempo et al., 2014; Hewitt and Stone, 2021). This involves innovation in the design of products, the post-processing and a well suited operational workflow in order to efficiently leverage the climate data produced by the models into actionable climate information. Furthermore, clear communication to the user is key for a good understanding and interpretation of the data to serve as support in decision making (Christel et al., 2018). The implementation of climate service prototypes which develop the co-design and delivery of real-time actionable products with/to stakeholders contributes to the advancement of these downstream processes, by providing a practical insight on the real use of the data. The contribution serves in both ways in the research-to-operations link, as a tool to support the development of forecast products but also as a feedback to the producing centers as the lessons learnt in these activities can guide the systems design and model development. Additionally, by making the predictions available to actual stakeholders in real-time, the ground is set for a real user assessment.

In this regard, this article describes the implementation and setting up of an operational real-time climate service based on subseasonal predictions. It intends to serve as a general guideline for implementing subseasonal operational services but also to provide a real example of a real-time workflow with its detailed specifics and challenges. This work draws from the experience acquired in the execution of the H2020 Subseasonal to Seasonal Predictions for Energy Project, S2S4E (Soret et al., 2019), in which a climate service based on seasonal and subseasonal predictions was co-developed and set up operationally. The service was delivered through a Decision Support Tool (DST), which was in production for 18 months; a snapshot of the interface is illustrated in Fig. 1. In terms of Technology Readiness Levels (TRLs), we estimated that the presented prototype reached a level of TRL7. The whole process was carried out by an interdisciplinary team. In this work we describe the operational implementation of the subseasonal service from a product focused perspective; therefore, the initial process of identifying the user's needs and conceptualising the climate service is not described here.

The article is structured as follows. In Section 2, a general recipe provides the steps for the operational implementation of a climate service based on subseasonal predictions, from the conceptualisation to the production phase. Then, to illustrate the process explained in Section 2 through a real case, the implementation of the subseasonal operational service of the S2S4E Project is presented in Section 3. This service consisted in the real-time provision of forecasts of weekly aggregated essential climate variables based on subseasonal predictions from ECMWF-Ext-ENS system (Vitart, 2004). This section also describes the

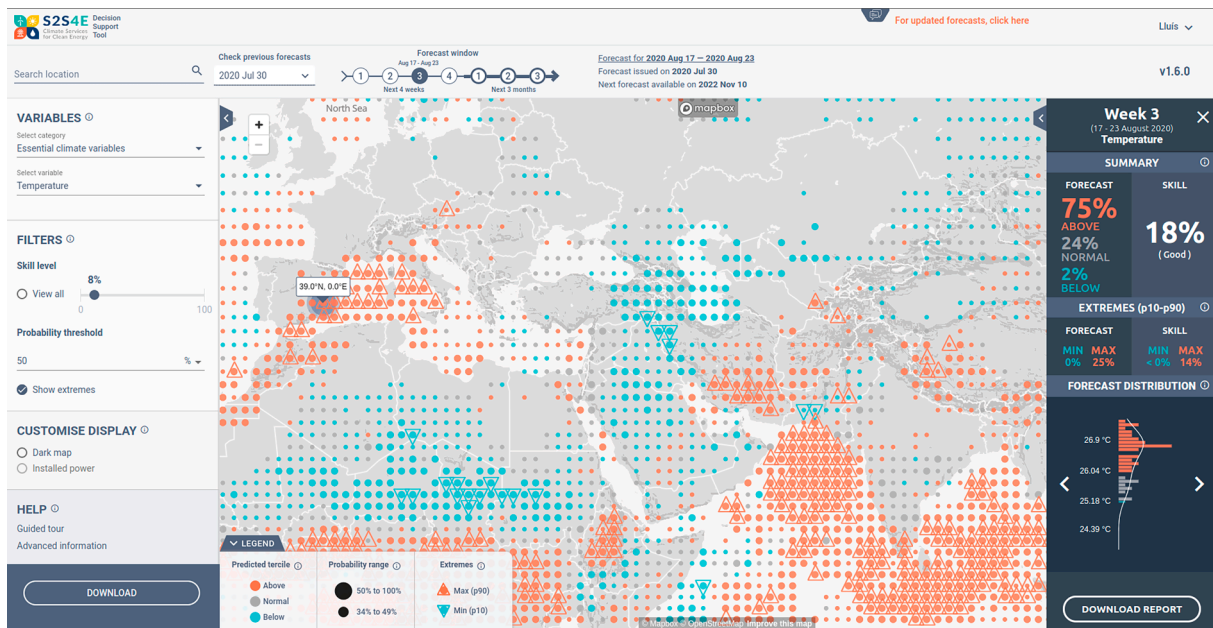


Fig. 1. S2S4E’s Decision Support Tool. Screenshot of the S2S4E’s Decision Support Tool showcasing the sub-seasonal temperature forecast issued on the 30th of July and targeting the third week.

methodologies adopted for the bias adjustment and the skill assessment and details the tailored operational workflow. In Section 4, some challenges regarding the use of other prediction systems are discussed. Some future prospects in the development of subseasonal climate services are discussed in Section 5, and the most relevant conclusions are gathered in Section 6.

2. Operational service design and conceptualisation

Implementing a data-driven climate service requires a co-design that involves data providers/producers and stakeholders, each with a specific point of view and a concrete set of requirements that pay special attention to various aspects. While the user is interested in actionable information, usefulness and user-friendliness, the data producers aim to ensure the scientific accuracy of the information provided, as well as the overall computing performance. For a balance that satisfies all the involved parts, an iterative design method needs to be established, validating every step of the design and implementation and re-doing the necessary steps if the current version of the service is not compliant. Fig. 2 shows a schematic of the different development steps identified;

below, the main aspects to consider at each stage are detailed.

2.1. Forecast product definition

The first step in implementing a service is to define the product to be delivered, based on the input provided by the end users. The product has to respond to the user’s needs while being under the current scientific capabilities, recognising the limitations of the climate predictions. Thus, choices like the variables predicted, the spatial and temporal resolution (forecast aggregation), or the forecasting horizon are crucial when designing a climate service that provides the demanded information to the user while being skilful enough to be helpful in a practical scenario. The issue date and periodicity of a forecast product has to be designed considering both the system’s release frequency of initialisation and the users’ decision making context.

The forecasts are probabilistic, so a sensible way to present a prediction is a probability distribution of the ensemble members for the given variable. A common practice to summarise this probabilistic information is using categories (typically terciles) referred to the past model outcomes in the hindcast period (model climatology). This

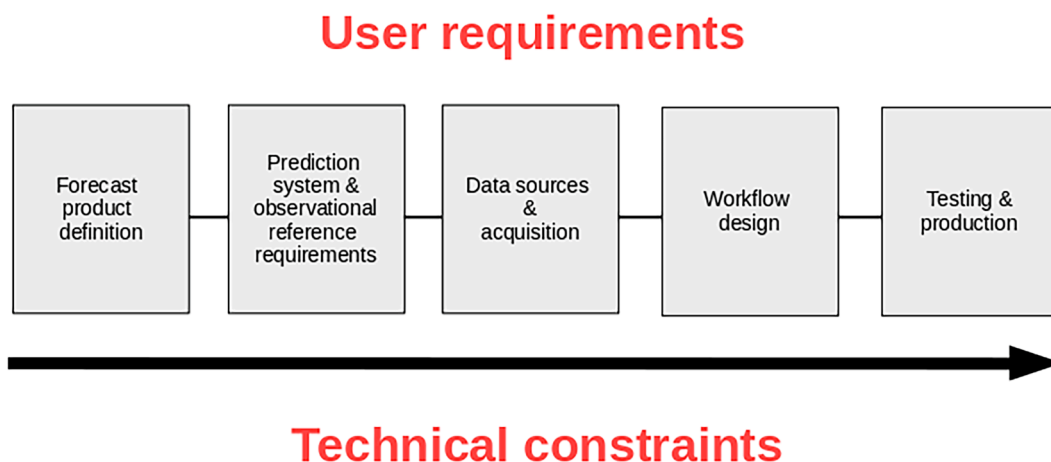


Fig. 2. Operational climate service implementation steps. Scheme exemplifying the steps in the design of a climate service operational implementation.

approach is increasingly criticised by users of climate services, showing the need to co-develop more tailored statistics closely with users. For example, a product of interest may be defined by setting a threshold and providing the probability of surpassing (or not reaching) this value. This threshold should be chosen to fit the user's need, and it can either be an absolute value (e.g. 0 °C – probability of freezing) or statistics from the model climatology (e.g. 90th percentile - probability of extreme event).

Hence, at this stage, a product contemplating the users' needs should be defined, considering the available climate information, the prediction system's characteristics and the potential skill until reaching a balance that meets the requirements while being technically feasible.

2.2. Prediction system and observational reference requirements

As a second stage, the prediction system or systems have to be selected to be the source of the climate forecasts. Several centers across the globe produce subseasonal predictions, some as an extension of the numerical weather prediction integrations and some with coupled systems employed for seasonal predictions. The choice of the most suitable prediction system is made based on many considerations, including the service requirements and the final product definition.

Regarding service requirements, the dates and frequency of initialisation of the system and the forecast horizon are relevant considerations. The initialisation dates of the prediction system are an essential factor to take into account when choosing a system, as this will condition when and how often the service will be updated. Most subseasonal systems are initialised at least once a week, and Thursdays coincide as a common start date day across many systems. Therefore this day could be convenient to facilitate the implementation of a multi-model system or ensure regularity if the prediction system employed has to be changed. Other systems follow a periodic schedule independently of the day of the week (e.g., every five days), which is not practical since users generally demand regularity in a service (same day of the week). Regarding the forecast horizon of the service, typically, subseasonal predictions go up to 4 or 6 weeks. The different subseasonal systems span from 32 to 60 days ahead, so it has to be ensured that the output from the chosen system covers the desired time range of the product.

Regarding the product definition, the main points to consider are the availability of the variables of interest and their temporal and spatial resolution. Not all platforms providing the forecasts have the same characteristics for the same predictions systems, as in some cases, the output variables have been post-processed to reduce the amount of data or to harmonize across systems. The spatial resolution of the subseasonal systems is typically of near 1°. In some prediction systems, these characteristics are not homogeneous over the entire data-set; the temporal or spatial resolution may change across variables. Usually, surface and pressure level variables have different properties to save resources and computing time during model runs.

Other relevant system characteristics when selecting the prediction system are the type of ensemble configuration and the type of hindcast, as both will condition the handling of the data. For the ensemble configuration there are typically two approaches, 'burst' mode and 'lagged' mode. In the 'burst' mode, all the ensemble members are initialised at the same time with slight perturbations in the initial state and/or the physics. In the 'lagged' mode, more frequent initialisations with fewer members are launched, which can be merged to form a larger ensemble called a 'lagged' ensemble. An ensemble obtained from a combination of the two approaches is also possible. In case of the hindcast generation, the producing centers choose between two approaches. Some centers run a hindcast for a fixed period, and this hindcast is used as reference for all forecast runs, until a new version of the system is implemented, when the fixed hindcast should be produced again. Other centers update their hindcast 'on-the-fly', meaning that a new hindcast set is generated with every forecast, ensuring that the exact system version is employed. This approach implies that also the hindcast period is updated (i.e. the 20 years prior to the forecast year).

The type of hindcast has implications on the final workflow; 'on-the-fly' hindcasts need software capable of downloading and processing this data in real-time as well. On the contrary, a fixed hindcast requires a single download without the 'real-time' component. The implications of these aspects on the workflow are addressed in Section 4.

Last but not least, an observational reference is required in order to bias adjust and perform a skill assessment of the predictions, typically a reanalysis or an observational dataset. Any reference system is subject to uncertainty. The selection of the reference dataset will require evaluation according to the specific needs of the users and product. Changes in the reference origin or type can lead to critical differences in skill. To take into account the uncertainty in the reference system, multiple systems could be combined; or some estimate of uncertainty could be included (like the provided for ERA5 by the Ensemble Data Assimilation system (Hersbach et al., 2020)). However, the use and implementation of these methods are not trivial, being a topic of current research. Thus, in the context of the operational implementation, we opted for a single reference dataset. The usage of the reference dataset depends on the system characteristics; in the case of 'on-the-fly' hindcasts, the reanalysis data has to be updated periodically, according to the period covered by the new hindcast.

2.3. Data sources and acquisition

Commonly, there are different sources from which the forecast output of a given prediction system can be retrieved. Although there is a wide variety of options, we can classify them into two categories. On one side, the research/operational institutions which develop and run the models, offer a platform where the output data is available. Whether the center is a research or an operational institution will imply that more resources are destined for one or the other purpose. With this in mind, services provided by operational institutions will be, in general, more reliable as they have more resources dedicated to offering continuous services with strong time forecasting constraints. On the other side, different initiatives are currently emerging, focusing their efforts on gathering and homogenising different systems from various sources into a centralised database. These platforms that collect multiple systems offer some clear advantages. They usually implement a centralised API (Application Programming Interface) that provides access to a collection of datasets by employing the same tools and methodologies and by only changing the request. Besides, files within these platforms commonly share the same data and metadata structure across systems, and, in some cases, they provide a standardised spatial grid. These features can become helpful when expanding the service to different variables or systems. A potential limitation of these kinds of platforms may be that they usually provide only a subset of the data available on the operational centers: not all variables are available and those available may be at lower spatial and temporal resolutions, which can be an issue for some applications. These data portals provide free predictions to promote research or operational service prototypes that can feedback into operational developments. However, the long-term availability or timely disposition of the forecasts may not be guaranteed. Additionally, the provision of the systems depends on each of the producing centers.

In the case of subseasonal predictions, there are two of such platforms: the World Weather Research Programme (WWRP)/ World Climate Research Programme (WCRP) Subseasonal to Seasonal (S2S) Prediction Project (Robertson et al., 2015) and the Subseasonal Experiment (SubX) (Pegion et al., 2019). The S2S Project maintains a valuable database of forecasts and hindcasts from 12 prediction systems for research purposes (Vitart et al., 2017). The predictions are converted to an unified grid of 1.5° × 1.5° to facilitate their analysis and comparison. The predictions available in the S2S database have a 3 weeks delay with respect to real-time, due to commercial constraints, thus limiting their use to research purposes. However, in the Phase II of the S2S Project the importance of offering free forecasts in real-time to research institutions was recognised as a path to explore the usability and promote the uptake

of S2S predictions in various sectors. With this aim, the Real-Time Pilot initiative was launched in 2019 and invited research groups or projects working in climate services for different sectors to participate. On the other hand, the SubX experiment brings together forecasts from 7 prediction systems in real-time, including both research and operational systems (Pegion et al., 2019). The database is publicly available through the International Research Institute for Climate and Society (IRI) Data Library (Kirtman et al., 2017). It includes 17 years of hindcasts and forecasts from seven systems since 2017. Additionally, computed daily climatologies are available.

Regarding data acquisition, most of the portals mentioned earlier rely on the standard HTTP or FTP protocol to allow external access to their datasets, which can be retrieved using the commonly used Wget or SCP tools. Some institutions are moving to more modern cloud-based solutions; for example, the NOAA (National Oceanic and Atmospheric Administration) has started to publish some of their climate datasets in the cloud through Amazon Web Services under the open data program. Cloud services offer unlimited resources, better performance, and improved security protocols over traditional servers. The datasets are then easily accessed through an API. Last but not least, some institutions supply more tailored services, and this is the case of the European Center for Medium Range Weather forecasts (ECMWF), which offers access to the S2S Project dataset through their data portal. This portal has been specifically designed to retrieve climate data, again employing an API, where subsets of the data can be specified and additional options are available (variables, temporal and spatial domains, file format...). Another example is the IRI Data Library which provides, as well as an online data repository of S2S and SubX data, a analysis web-service allowing subsetting and post-processing prior to download reducing the required bandwidth.

2.4. Workflow design

After selecting one or several forecasting systems and their sources, several actions are taken to extract helpful information from the raw data. The transformation applied to the data ranges from simple spatial interpolation techniques or the computation of tailored climate indices to more complex algorithms looking to improve the characteristics of the provided information, such as downscaling and bias adjustment. A considerable part of the processing is simple formatting and minor computations to accommodate the input data to the service's spatial and temporal scales requirements. In addition, to make the probabilistic information more understandable and to evaluate the quality of the predictions, a wide variety of statistics are computed, like tercile categories and skill scores. Finally, the predictions and associated elements to be provided as output have to be saved in an appropriate format. In this step, it is crucial to consider all the different methods applied to the raw data and define a workflow that executes each one of them accurately and efficiently. With this regard, modularity is a valuable property for the post-processing steps, as it implies a more easily adaptable workflow to future user requirements.

Hence, once the data sets needed are selected, and the post-processing steps are defined, a workflow orchestrating the different pieces of the puzzle needs to be designed to routinely retrieve the raw data, compute the operations and transfer the output in its desired format to the user. This process can be more or less complex depending on the volume of the data, the number of sources used to collect it, and the constraints imposed by the user; how much time is the user willing to wait for each update, for example. With this in mind, the final workflow has to contemplate the aspects aforementioned ensuring the quality of the output.

On the technical side, actual implementations of operational services need to handle incidences like delays in the publication of the forecast or corruption of the data originated in the source or during the transfer. For this reason, when designing the system, the software should be able to handle these issues by catching the irregularities and generating the

actions to resolve them in time. With this in mind, workflow managers are a great asset, as they can handle fails and retrials and dependencies between jobs and different machines. There are several open-source workflow managers freely available. For instance, ecFlow (Bahra, 2011); developed and maintained by the ECMWF and tailored, although not limited, to the execution of weather and climate models, Cylc (Oliver et al., 2019); a general purpose workflow engine that orchestrates cycling workflows very efficiently, or Autosubmit (Manubens-Gil et al., 2016); a workflow manager developed at the Barcelona Supercomputing Center to create, manage and monitor climate experiments by using computing clusters, HPC's and supercomputers remotely via ssh.

Besides that, constant checks are needed due to the volume of data processed, making their integration a good practice at every stage of the process, thus ensuring that the resulting data has the expected attributes. Therefore, checks on different attributes like the file size, the contained variables, dimensions and timesteps, and value ranges, should be implemented over every process step. The computing resources needed will grow with the number of checks applied over a dataset and its volume. For this reason, it is essential to verify the data while processing it rather than running a single exhaustive test after-action.

2.5. Testing and production

Once the operational service is implemented, due to its complexity, it is important to verify, before the production phase, that the system is producing the right outcome. For this purpose, a possible strategy is to compare the results against a benchmark, that can be an independent prior study or calculation. In many cases, the operational service originates from research work where the methodology was developed and tested. This provides a test bed to assess the outputs, however it is desirable that the workflow and the tests run in as little common software as possible. Additionally, if other services are available producing similar information, is it valuable to compare both outcomes and try to identify and understand the differences that could be found.

Once an initial verification is done and the service is moved to a production environment, issues and new feedback from the users may arise. Thus, it is advantageous to maintain open the validation loop between users and developers to enhance the service at the production stage. As a consequence, it is advisable to save resources to address possible new developments during this period.

3. Implementation of a climate service based on subseasonal predictions: a practical example

In this section, the steps to implement an operational climate service presented in the previous section are exemplified through the description of a real case, the implementation of the climate service of subseasonal predictions developed within the S2S4E Project. As part of this project, seasonal and subseasonal predictions of essential climate variables and energy related indicators were operationally delivered through an interactive interface, the "Decision Support Tool" (DST), to a wide range of actors in the energy sector (<https://s2s4e-dst.bsc.es>). We here describe the implementation of this operational service: the selection of the prediction system, the strategy for calibration and forecast skill assessment and the design of the operational workflow orchestrating all the processes and their timings, which were adjusted to the producing center's data release schedule. Furthermore, we detail practical aspects related to the treatment of subseasonal predictions and how we addressed the scientific and technical challenges encountered in the process.

3.1. Forecast product definition

The subseasonal climate service consisted in several forecast products of the essential surface climate variables that were deemed most relevant for the sector, i.e. 2 m temperature (mean, maximum and

minimum), 10 m wind speed, mean sea level pressure, precipitation and solar radiation, provided for the forthcoming four weeks. The predictions were given as probabilities of each tercile category for each grid point. The risk of occurrence of extremes was procured as the probability of exceeding (or not reaching) the 90th (10th) percentile; this decision was taken after several discussions with users. Additionally, some tailored indicators for renewable energy generation and electricity demand were calculated and provided. Furthermore, addressing a request raised by users during the co-development, a product of country-averaged predictions (for European countries) for each variable was added. The predictions were updated weekly (on Thursday) and provided as weekly averages up to 4 weeks ahead, starting on day 5 (week 1 days 5–11; week 2 days 12–18; week 3 days 19–25; week 4 days 26–32). The 7-days aggregation period was selected to match Monday to Sunday as this natural week is commonly employed by energy users in their analyses.

3.2. Prediction system requirements and acquisition

This section addresses the following two steps jointly: “Prediction system and observational reference requirements” and “Data sources and acquisition”. In the selection of the prediction system and downloading platform, several requirements imposed by the nature of the service and the constitution of the products were decisive.

The two main databases S2S Project database (Vitart et al., 2017) and SubX (Pegion et al., 2019) were discarded for different reasons. Since the climate service was an operational prototype aiming to offer timely predictions to the users, the availability of real-time forecasts was essential. For this reason, the S2S Project database with 3-week delay in the forecasts was not adequate. On the other hand, the SubX platform does offer real-time predictions from several subseasonal prediction systems, however, the wind forecasts from this database are provided as time averages of the wind components (u and v), instead of the instantaneous values which are required for the calculation of the wind module, necessary to compute a specific wind energy product (wind capacity factor). It should be noted that this drawback is a consequence of a choice of the SubX database to homogenise and save space and that the individual models in their original output may be suitable for our product. As a matter of fact, the first subseasonal prediction system implemented in the S2S4E climate service was NCEP CFSv2 (Saha et al., 2014), which is one of the systems included in SubX database, but alternatively it was downloaded from NOAA’s NOMADS server directly, which provided instantaneous values of the wind components at a higher frequency. At a later stage, the launch of the S2S Project Real-Time Pilot Initiative provided an excellent setting for the S2S4E Project, since this initiative provided access to real-time S2S predictions from all prediction systems in S2S Project to collaborating projects. With this programme in place, the prediction system integrated for S2S4E DST was updated to the ECMWF extended range forecast system ECMWF-Ext-ENS (Vitart, 2004; Vitart et al., 2008). This system was chosen for several reasons: due to its good skill results (Vitart, 2014), to attain coherence with the seasonal predictions (also provided by the S2S4E DST and based on ECMWF seasonal system SEAS5 (Johnson et al., 2019)) and because of the user awareness of ECMWF’s excellent track record. The operational implementation that will be described hereafter is based on predictions from ECMWF-Ext-ENS.

The ECMWF-Ext-ENS system (Vitart, 2004; Vitart et al., 2008) is a prolongation of the ECMWF Integrated Forecast System (IFS) to cover time scales beyond medium-range, producing a forecast up to 46 days ahead. This extension entails a coupling with ocean and ice components to account for the influence of these slowly varying systems. The extended forecasts are run twice a week, every Monday and Thursday. Each forecast run consists of 51 ensemble members to represent model uncertainty. The original resolution is 16 km up to day 15 and 31 km beyond day 15. The predictions provided by S2S are homogenised to a common grid of $1.5^\circ \times 1.5^\circ$. For each real-time forecast, an ‘on-the-fly’

hindcast is computed by initializing the same prediction system with 11 ensemble members, on the same date for the previous 20 years. This approach ensures that the hindcast set is run with the exact same model version as the forecast, thus providing an up-to-date reference to identify systematic model biases and correct them in the real-time forecast during post-processing. ECMWF-Ext-ENS uses this type of hindcast setting due to the frequent cycle upgrades of the IFS, normally every few months. ECMWF-Ext-ENS real-time forecast runs are initialised at 00 UTC on Mondays and Thursdays, and are available a few hours later. The hindcast set corresponding to each initialisation is produced and made available 2 weeks prior to the real-time forecast as represented in the schematic in Fig. 3.

The observational reference employed for both bias adjustment and forecast skill assessment was the ERA5 reanalysis (Hersbach et al., 2020), as this dataset provides a good representation of surface winds suitable for energy applications at the global scale (Ramon et al., 2019).

3.3. Workflow design

This section describes the design of the operational workflow, which has been adapted to the schedule of production of ECMWF-Ext-ENS forecasts and hindcasts depicted in Fig. 3. In particular the timings of the downloading and post-processing tasks are tailored to the release of new data. Additionally, the methodology employed for the bias-adjustment of the predictions and for forecast skill assessment is also described. These processes are also conditioned by the ECMWF-Ext-ENS schedule, the readiness of the ‘on-the-fly’ hindcasts limits the data available to be used in the bias adjustment and quality assessment.

3.3.1. Methodology

The method employed to adjust the systematic mean bias and the spread of the predictions is the variance inflation (Doblas-Reyes et al., 2005). This technique compares the climatological distribution of the prediction system and that of the reference dataset and adjusts the predictions to have the same interannual variance as the observations, while preserving their interannual correlation. Hereafter, this method will be referred to as calibration. It has been successfully tested for seasonal predictions of wind speed (Torrallba et al., 2017), temperature and precipitation (Manzanas et al., 2019). With seasonal predictions, calibration is carried out on monthly averages; in the case of subseasonal predictions, the method has been adapted to work on weekly averages, raising some additional challenges. One of these challenges is the way the climatology is computed to correctly characterize the model and observed climate distributions. In the calibration of a real-time forecast, its associated hindcast is employed to represent the model climatology based on past predictions. The model climatology is confronted against the reference climatology and some adjusting parameters are determined to correct the real-time forecast. To account for the model bias growth with lead-time, the calibration is performed on each forecast week independently. In our operational workflow, both the forecasts and the hindcasts are routinely calibrated with ERA5 reanalysis. In the case of the forecasts, to adjust the product delivered to the user, and in the case of the hindcasts to produce a corrected set of past predictions that will be analysed to provide a robust quality estimate of the forecast product. The skill scores used are fair Ranked Probability Skill Score (RPSS) for the tercile categories and fair Brier Skill Score (BSS) for the probability of extremes (Ferro, 2014). The hindcasts are calibrated with the same methodology as the forecasts but in a leave-one-out cross validation, to ensure no observational information of the adjusted year is included in the process. Similarly, for all the steps mentioned earlier, the whole hindcast period (20 years in the case of the ECMWF-Ext-ENS system) was used to ensure that the probabilistic information provided was robust enough.

To conduct these analyses, a robust statistical software is required. In this case, the R programming language (R Core Team, 2019) is chosen given the number of packages available for environmental sciences

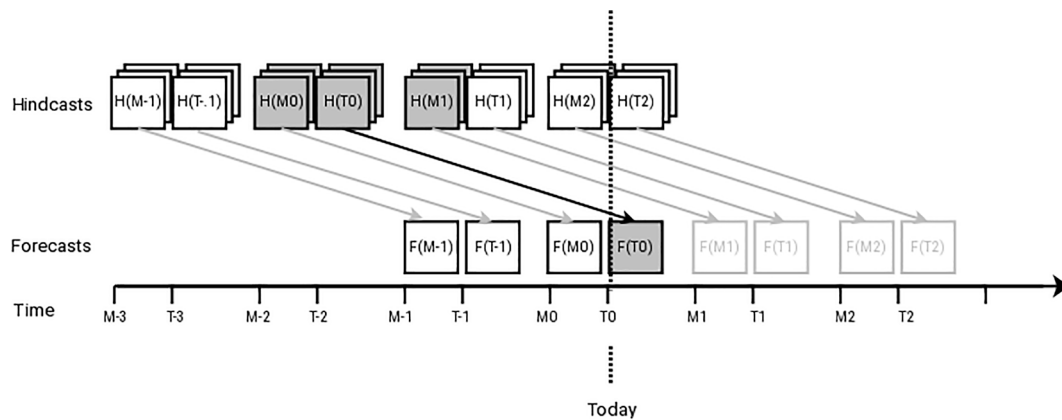


Fig. 3. Schematic representing ECMWF-Ext-ENS real-time schedule for the provision of forecasts and associated hindcasts. Mondays and Thursdays are indicated with 'M' and 'T', while the numbers indicate the corresponding week, being 0 the current week. Positively numerated weeks correspond to the future weeks while negatively numerated weeks correspond to the past weeks. The top row of boxes represents each set of 20 hindcasts associated with every forecast, available 2 weeks before the real-time forecast. Assuming the current date is Thursday of week 0 (T0), the real-time forecast released today is the grey box F(T0). Its associated hindcasts H(T0) were made available on Thursday 2 weeks ago (T-2). The boxes outlined in black represent the data that is available to date, the boxes in light grey represent the data that will be available at future dates. To calibrate today's forecast F(T0), the climatology distribution is computed from three hindcast sets defined by a window of three start dates centered on hindcast H(T0); H(M0), H(T0) and H(M1) (i.e. hindcasts from the previous and following Mondays with respect to the initialisation date are employed to increase the sample).

analysis. Specifically, to construct the operational workflow the R package `startR` (Manubens et al., 2022) has been used to retrieve the data from files, `s2dverification` (BSC-CNS et al., 2022) to calculate anomalies, and `easyVerification` (MeteoSwiss, 2017) to calculate the skill scores. Furthermore, during the S2S4E project, the calibration method has been made available in `CSTools` (Perez-Zanon et al., 2021).

Although ECMWF-Ext-ENS is initialised every Monday and Thursday, our operational service was updated on a weekly basis based on Thursday's forecasts. For the calibration process both Mondays and Thursdays hindcasts were considered, in order to benefit from an increased sample, as tested in Manrique-Suñén (2020). This work showed that the skill scores varied significantly depending on the sample size used to calculate the climatology. The best approach to compute the climatology was to set a running window centered in the start date, to increase sample size while ensuring representativeness of the climate distribution associated with that start date. The size of the window in our operational setting with an 'on-the-fly' hindcast is limited by the readiness of the latests hindcasts. As shown in Fig. 3, in order to calibrate today's forecast F(T0), as well as the hindcast set which was initialised on the same date H(T0) and released 2 weeks before, the hindcasts initialised on the adjacent startdates are already available and can be used in the computation of the climatology. The largest window that can be constructed centered on H(T0) is of seven sets of hindcasts initialized on the 7 start dates from H(M-1) to H(M2). The use of H(T2) is disregarded as including it would allow little time for processing before Thursday's forecast release. The possible windows sizes to compute the climate distribution are therefore of three, five or seven start dates centered on the forecast date T0.

To evaluate the effects that the size of the window have in the forecast skill, both temperature and wind speed were calibrated based on a climatology constructed with: one single set of hindcasts H(T0), and with running windows of three, five and seven start dates around the initialisation date. The comparison of the resulting values of fair RPSS is included in Supplementary Material. This analysis shows that the calibration based on climatology computed from a single start date does not yield a robust adjustment since it degrades skill with respect to the raw predictions ('cal1' in Supplementary Figs. 1–4). These results are in agreement with Manrique-Suñén (2020) where a simple bias correction was employed instead of the calibration method implemented in this work. Employing a window of three start dates ('cal3') improves fair RPSS notably with respect to a single start date ('cal1'), whereas a

further increase to 5 or 7 start dates has no significant effect (Supplementary Figs. 1–4). The larger window of 7 start dates has negative effects on skill in some cases (parts of Asia in Supplementary Fig. 2). The results show that most of the improvement is gained when increasing the window to 3 start dates and only marginal gains are achieved for wind speed when employing 5 start dates. From a practical point of view, a larger window implies a higher computational cost and less margin of operation in case of delays or technical disruptions. For these reasons and given that skill improvements when using 5 start dates were only marginal, the calibration of the real-time forecasts for the operational service was applied based on a climatology computed from a sample of 3 hindcast sets from 3 start dates centered on the initialisation date. This three start date window is highlighted in grey in Fig. 3. The model's climate distribution is therefore constructed from a sample of 660 values (3 start dates \times 11 ensemble members \times 20 years) which is compared against the ERA5 reference climate computed in an analogous manner, in this case from 60 values (3 start dates \times 20 years). From the point of view of representativeness of the climatology, a 3 start date window preserves the seasonal cycle (climatologies span one week). A longer window with more start dates and spanning several weeks may filter out the seasonal cycle. Nevertheless, for specific products, a longer window may be beneficial as the increased sample size provides a better characterisation of the tails of the distribution (i.e. extremes).

With every real-time prediction issued weekly on a Thursday, a probabilistic skill score was also provided, to inform the user of the quality of the forecast product. The fair RPSS and fair BSS skill scores were routinely computed for each new forecast based on the latest eight sets of calibrated hindcast. The use of several start dates increases the sample size in order to yield a robust probabilistic estimate of skill. The skill is assessed on every lead time and grid point independently.

In summary, every real-time forecast released on Thursday is calibrated employing its hindcast and the one of the previous and following Mondays. With this setting, although the forecasts initialised on Mondays are not used (the service forecasts were only updated on Thursdays), all hindcasts (Mondays and Thursdays) are employed for calibration and also for skill assessment, maximising the use of the data available.

3.3.2. Implementation

The operational workflow to generate our forecast product was implemented and controlled via the workflow manager `Autosubmit`

(Manubens-Gil et al., 2016). The schematic in Fig. 4 shows the different processes from the data downloading until the product upload to the platform, which include formatting, calibration, probabilities computation and skill assessment. Different actions are performed every Monday (M) and Thursday (T), days on which new data is available to download. On Thursdays, the real-time forecast is downloaded and the forecast products are computed and issued; additionally a hindcast set is also downloaded (Thursday's actions are included in the light blue box in Fig. 4). On Mondays, a hindcast set is downloaded, and the skill computation takes place (Monday's processes are included in the light brown box in Fig. 4). Note that the hindcast sets are downloaded two weeks before their corresponding real-time forecast.

On Thursday (T0) of the current week, the real-time forecast for the next 4 weeks is downloaded (represented as F(T0) in Fig. 4). After downloading the files, the first step consists on formatting and pre-processing. The original files in GRIB format are converted into NetCDF and the weekly averaging is performed (beginning with day 5 of forecast). Additionally, some quality checks are applied to the data (i.e.

check number and size of files). That same Thursday, the hindcast corresponding to Tuesday in two weeks time H(T2) is downloaded and pre-processed. Formatted files are indicated in orange in Fig. 4. Once the forecast data F(T0) is pre-processed, the next step is the calibration, which is based on a climatology constructed from hindcast sets from 3 start dates to construct the climatology (a running window as explained in the methodology section). In our example, the forecast F(T0) is calibrated employing the three hindcast sets initialised on the same date H(T0), the previous Monday of the same week H(M0) and the following week's Monday H(M1). All of these hindcast sets had been previously downloaded, formatted and pre-processed at earlier dates. The resulting calibrated forecast F(T0) is indicated with a purple box in Fig. 4. Then, the forecast probabilities are computed from the forecast ensemble members employing as percentile thresholds (p33, p66, p10 and p90) which had been previously derived from the corresponding calibrated hindcast H(T0) on Monday of the previous week (M-1). Both the terciles probabilities and probabilities of extremes (probability of exceeding or not reaching p90 and p10) constitute the final forecast products which

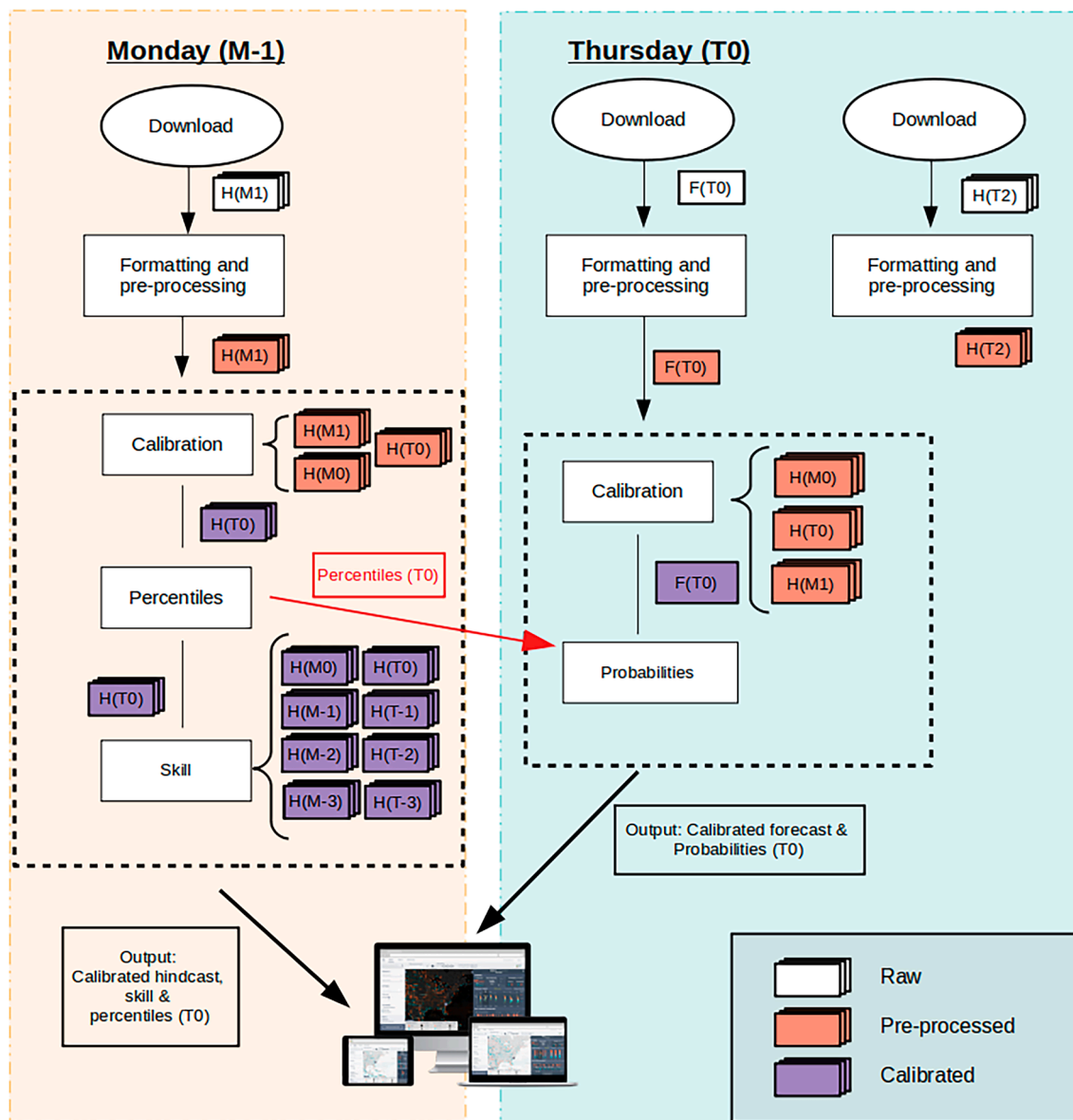


Fig. 4. Operational data workflow based on ECMWF-Ext-ENS prediction system. Every Monday a new hindcast is downloaded and calibrated; the tercile and extremes boundaries are computed as well as the skill scores, these are all stored in the system. Every Thursday the real-time forecast is downloaded and calibrated, then the tercile and extreme probabilities are computed ready to be delivered. Additionally a new hindcast is also downloaded every Thursday.

are uploaded to the platform. Complementing each product, a forecast skill is provided, which was previously calculated, as explained below.

Ten days earlier, on Monday (M-1), both the skill and the necessary percentiles had been already computed for Thursday's forecast product F(T0), based on its corresponding hindcast H(T0). On Monday of the week before (M-1), the hindcast set for Monday H(M1) was published, it was downloaded, formatted and pre-processed. With this hindcast set ready, the calibration of H(T0) could be carried out (employing the 3 sets of hindcasts H(M0), H(T0) and H(M1)). Once the hindcast H(T0) was calibrated, the percentiles that define the tercile categories (p33 and p66) and the extremes of the distribution (p10 and p90) were computed and stored to be used on the current week forecast F(T0) computations. Additionally, the skill computation was performed, assessing the last eight calibrated hindcasts up to H(T0) to produce a robust score (using both Monday's and Thursday's hindcasts). The resulting fair RPSS and fair BSS were stored to be provided along with F(T0) forecast product. Since this computation is done 10 days before the actual forecast is published, there is some room for manoeuvre in case of technical issues and the results can be checked before the actual forecast day. Additionally, taking into account that the computation performed over the hindcast could be expensive in time and resources, this strategy helps to reduce considerably the operations needed on forecast day (Thursday).

3.4. Testing and production

Despite the optimised workflow design and the quality checks included, multiple issues can occur in an operational service, compromising the final product. During the production phase of the S2S4E project, three main types of issues were within the most frequently encountered. The first one, was the delay in the publication of the forecast by the producing center. Although this is not a common issue, there could be delays lasting from hours to days for various reasons. A second issue is the planned or unplanned disruption of the computational resources, leading to failures or delays. A contingency plan needs to be arranged to solve or mitigate these potential issues, this can be as simple as showing a warning message in the tool and/or sending a notification email to the users. Last but not least, on some rare occasions, corrupted data, either generated at its source or during the workflow, was able to pass undetected through the control checks. This issue can affect the correct functioning of the service, implying, on many occasions, a manual inspection of the files and a re-run of the entire workflow.

4. Challenges to extrapolate to other prediction systems

The operational implementation described is designed for ECMWF-Ext-ENS prediction system, therefore the workflow is specifically tailored to its operational schedule of forecast and hindcast configurations. Although many aspects described here apply or can be easily adapted to other subseasonal systems, we will discuss two characteristics which would require slight changes in the approach or at least some attention. These relate to the type of ensemble, (i.e. the use of a lagged ensemble as opposed to the burst ensemble found in ECMWF-Ext-ENS) and to the type of hindcast, (i.e. the fixed hindcast as opposed to the 'on-the-fly' hindcast found in ECMWF-Ext-ENS). These are common configurations found in many subseasonal systems therefore it is worth discussing their implications for the design of the operational workflow.

4.1. Lagged ensemble

Some subseasonal systems have frequent initialisations of a small set of ensemble members. In order to create a larger ensemble for a reliable prediction, runs from different initialisation times have to be pooled together to form a lagged ensemble. This aggregation increases the number of ensemble members to better represent the range of possible outcomes and thus may improve reliability but at the same time it can

deteriorate skill by including information from old initialisations. Analogously, the hindcasts runs follow a similar configuration with typically a reduced number of members or initialisations. The selection of an 'optimal' lagged ensemble has been addressed by some studies (Trenary et al., 2017; Vitart and Takaya, 2021). This choice may be related to the forecast horizon and other properties of the desired product. These type of systems offer more flexibility to create a forecast product, as there is not a pre-defined issue date (i.e. Monday and Thursday), so the forecast can be issued at any chosen dates to better suit the user's needs. Additionally, the number of members of the ensemble can be adapted (more members may characterise better the uncertainty at longer lead times), but consequently these systems require an exhaustive evaluation in order to make a choice of design. An example of a subseasonal system that uses this approach is the NCEP CFSv2 (Saha et al., 2014), whose forecasts have initialisations every 6 h (cycle) launching 4 ensemble members. In the hindcast only one run is launched every cycle 6 h. This configuration can therefore produce 16 ensemble members for forecast by taking the predictions of the last 24 h or increase further the ensemble by adding more initialisations. In order to use a lagged ensemble in an operational service, an additional step has to be added to the workflow which is pooling the different initialisations to create the lagged ensemble. The different runs have to be merged making sure they verify at the expected forecast time, i.e., in the case of NCEP CFSv2 the different lead times need to be adjusted by 6 h. Once the lagged ensemble is created, it can be treated like a burst ensemble and the other post-processing steps can be applied. In the S2S database, the forecast runs are already aggregated daily, thus providing 16 ensemble members per day (4 in hindcasts). The ensemble type has to be considered before implementing the service, as a lagged ensemble can add load to the resources required in terms of design and development of our workflow.

4.2. Fixed hindcast

Many subseasonal systems have a fixed hindcast instead of a 'on-the-fly' hindcast. This hindcast is computed once for a number of years in the past. A fixed hindcast facilitates the workflow described, as the computation of many statistics does not have to be redone with every new forecast. This implies that the skill scores are calculated only once and then when a forecast is launched, the corresponding skill score to that time of the year is presented with it. Equally with the model climatology from which the tercile categories are calculated, these values only need to be computed once and then can be stored as static values to be used to compute the tercile probabilities. Only when the system is upgraded a new hindcast set will be produced, and the skill scores and statistics will need to be re-calculated for the new system.

5. Future prospects

From the experience gathered during the S2S4E project, we analyse three main future aspects which might contribute to improved operational climate services. These aspects relate to the integration of prediction time scales in a seamless way, the combination of multiple prediction systems into a multi-model product, the use of cloud-computing resources and the detection of windows of opportunity.

With the growing number of prediction time scales available for climate services - from numerical weather prediction, subseasonal, seasonal, decadal up to decadal predictions and climate projections - there is an increased amount of information available to suit different decision making contexts. This escalation may imply some overlapping across time scales covered by the different prediction systems, which may entail incoherences in the information issued, caused either by the differences in the prediction systems (Earth components coupling, spatial resolution, etc) or simply in the definition of the forecast products (aggregation period, post-processing, reference climatology, etc). A first approximation toward combining different time-scales predictions

was conducted during the S2S4E project by providing 18 months of real-time subseasonal and seasonal forecasts through the fully-operational Decision Support Tool (DST). During these first tests, the prediction systems used for the subseasonal and seasonal time scales were different, and the post-process applied to bias-adjust each of them was independent. This methodology led to occasional inconsistencies between predictions across the two time scales, generating mistrust among users of the service. Hence, we identify the need to develop post-processing techniques that integrate information from multiple scales into a single and coherent prediction. However, the development of these methodologies should be escorted by efforts in the production upstream of the climate data involved, such as standardised initialisation procedures and data access across prediction systems.

The second point for future developments is the multi-model approach. There are many modelling centers across the world producing S2S predictions. This model diversity is beneficial because the sources of predictability at these timescales come from different mechanisms and may depend on location and season. One model might be better than others at predicting certain patterns. A known strategy to exploit model diversity and produce a joint reliable prediction is to combine several forecasts to construct a multi-model product. However, in the case of subseasonal predictions, this is a very challenging task, again due to heterogeneity of the prediction systems, although some attempts have been done in this direction (Vigaud et al., 2017; Specq et al., 2020). One of the main caveats is that the initialisation date varies with the model, making it non-trivial to aggregate forecasts with different lead times. The diversity in the type of ensembles also complicates the design of such multi-system. In this regard, a protocol aligning the initialisation dates would help in the design of such a system.

Both of the approaches aforementioned imply managing an increasing volume of data, which might come from multiple sources. Thus, access to vast computing resources will be crucial in future implementations of operational climate services. Under this scenario, cloud-computing platforms are a very suitable ally, as they can offer unlimited resources on-demand, delivering improved performance and reliability against traditional clusters. The correct integration of these platforms can lead to better workflows producing more reliable predictions and presenting increased transparency and traceability on the data and methods used, building, as a consequence, more trust in the service provided.

Last but not least, in subseasonal timescales, forecasts may be influenced by climate phenomena or conditions such as ENSO, the MJO, the land surface or the stratosphere. This impact is spatially heterogeneous and intermittent in time, thus providing some strategic periods or windows of opportunity when the subseasonal forecasts may be more skilful (Mariotti et al., 2020). The knowledge a priori of these periods or regions where a subseasonal forecast is more reliable can increase the applicability of subseasonal forecasts. However, skill is often computed not taking into account these prevailing conditions, as it evaluates the average behaviour of the forecast in the past (hindcast). Some studies have attempted to evaluate the skill conditional to these climate phenomena (Lledó and Doblas-Reyes, 2020; Mayer and Barnes, 2021). Yet, this procedure subsets the dataset reducing the available sample and creating less statistically robust estimates of the forecast skill. Thus, finding an optimal way of detecting forecast windows opportunity and their role in forecast uncertainty gathers the attention of researchers, with promising impact in the field of subseasonal predictions.

6. Conclusions and discussion

A general guideline to create a subseasonal climate service has been presented. It contains the main aspects that need to be taken into consideration in order to set up an operational service of climate predictions. These include the characteristics of the systems, the data sources and methodology. Then, a detailed workflow design of a forecast

product implementation from the S2S4E Project is described to provide a real illustration of such process. The service was designed attending to scientific criteria, user requirements and technical limitations; it is therefore the result of a trade-off to attain a well-founded solution in terms of skill, usability and computational resources. The main challenges, in addition to the technical challenges inherent to any operational workflow, (automatisation, efficiency, dealing with delays in input data) were related to the configuration of the subseasonal predictions systems. A useful approach was to plan the timings of the tasks of the workflow to the release schedule in order to distribute the workload and improve efficiency. The scientific challenges relate to the small sample size available for calibration and verification and they were tackled by employing a running window to increment the number of hindcast to construct the climatology, and thus improve its robustness.

This work is the result of the experience gained in the S2S4E Project which had as main output a Decision Support Tool providing operational climate predictions from subseasonal to seasonal time ranges of climate variables and energy related indicators. The engagement with the users was key at all stages and guided the development of the service. The operationalisation of subseasonal predictions as a functional product was one of the main achievements in the project, since the provision of subseasonal forecasts for operational use is still at developing stage. The presented workflow combines scientific rigour and a practical approach to deliver credible and timely information. Additionally, the detailed description presented recognizes and visualises the work behind an operational implementation. The data download and data management are aspects often neglected and their workload is often underestimated. Additionally, despite the high degree of automatisation in the processes, there is the need of some human supervision and maintenance. This work contributes to the development of climate services by sharing the lessons learnt about the implementation of an operational workflow based on subseasonal predictions.

CRedit authorship contribution statement

Andrea Manrique-Suñén: Conceptualization, Methodology, Validation, Writing - original draft. **Lluís Palma:** Software, Methodology, Validation, Writing - original draft. **Nube Gonzalez-Reviriego:** Conceptualization, Writing - review & editing. **Francisco J. Doblas-Reyes:** Writing - review & editing. **Albert Soret:** Writing - review & editing, Project administration, Funding acquisition.

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.

Data availability

The data that support the findings of this study are openly available and they can be found in the respective references.

Acknowledgement

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grants 7767874 (S2S4E) and 869565 (VitiGEOSS). ECMWF-ExtENS real-time predictions used for the operational prototype were provided by the Subseasonal to Seasonal (S2S) Prediction Project's Real-Time Pilot Initiative to S2S4E Project as one of the participating projects. The data can be obtained from the S2S Project database through its two data portals: ECMWF (<https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/>) and CMA (<http://s2s.cma.cn/index>). The ECMWF ERA-5 reanalysis was accessed from Copernicus Climate

Change Service (C3S) Climate Data Store (<https://cds.climate.copernicus.eu/#/home>).

The authors thank Pierre-Antoine Bretonnicre and Margarida Samsó for data downloading and processing, and Núria Pérez Zanón, An Chi-Ho and Nicolau Manubens for technical support. Finally, Isadora J. Christel and Núria Pérez Zanón are acknowledged for fruitful discussions and feedback.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cliser.2023.100359>.

References

- Bahra, A., 2011. Managing work flows with ecflow. *ECMWF Newsletter* 30–32.
- BSC-CNS, Guemas, V., Manubens, N., Perez-Zanon, N., Garcia-Serrano, J., Fuckar, N., Caron, L.-P., Bellprat, O., Rodrigues, L., Torralba, V., Hunter, A., Prodhomme, C., & Menegoz, M. (2022). s2dverification: Set of Common Tools for Forecast Verification. URL: <https://CRAN.R-project.org/package=s2dverification> r package version 2.10.3.
- Buontempo, C., Hewitt, C.D., Doblas-Reyes, F.J., Dessai, S., 2014. Climate service development, delivery and use in Europe at monthly to inter-annual timescales. *Climate Risk Management*, 6, 1–5. <https://doi.org/10.1016/J.CRM.2014.10.002>.
- Christel, I., Hemment, D., Bojovic, D., Cucchiatti, F., Calvo, L., Stefaner, M., Buontempo, C., 2018. Introducing design in the development of effective climate services. *Climate Serv.* 9, 111–121. <https://doi.org/10.1016/J.CLISER.2017.06.002>.
- Doblas-Reyes, F.J., Hagedorn, R., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus A Dyn. Meteorol. Oceanogr.* 57, 234–252. <https://doi.org/10.3402/tellusa.v57i3.14658> <https://www.tandfonline.com/doi/full/10.3402/tellusa.v57i3.14658> <https://www.tandfonline.com/doi/full/10.3402/tellusa.v57i3.14657> <http://tellusa.net/index.php/tellusa/article/view/14658>.
- Domeisen, D.I., White, C.J., Afargane-Gerstman, H., Muñoz, Á. G., Janiga, M.A., Vitart, F., Wulff, C.O., Antoine, S., Ardilouze, C., Batté, L., Bloomfield, H.C., Brayshaw, D.J., Camargo, S.J., Charlton-Pérez, A., Collins, D., Cowan, T., Chaves, M. d. M., Ferranti, L., Gómez, R., González, P.L., Romero, C.G., Infanti, J.M., Karozis, S., Kim, H., Kolstad, E.W., LaJoie, E., Lledó, L., Magnusson, L., Malguzzi, P., Manrique-Suñén, A., Mastrangelo, D., Materia, S., Medina, H., Palma, L., Pineda, L.E., Sftetos, A., Son, S.-W., Soret, A., Strazzo, S., & Tian, D. (2022). Advances in the subseasonal prediction of extreme events: Relevant case studies across the globe. *Bull. Am. Meteorol. Soc.*, -1. URL: <https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-20-0221.1/BAMS-D-20-0221.1.xml>. 10.1175/BAMS-D-20-0221.1.
- Ferro, C.A.T. (2014). Fair scores for ensemble forecasts. *Q.J.R. Meteorol. Soc.*, 140, 1917–1923. URL: <http://doi.wiley.com/10.1002/qj.2270>. DOI: 10.1002/qj.2270.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., & Thépaut, J. (2020). The ERA5 Global Reanalysis. *Q.J.R. Meteorol. Soc.*, (p. qj.3803). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803>. DOI: 10.1002/qj.3803.
- Hewitt, C.D., Stone, R., 2021. Climate services for managing societal risks and opportunities. *Climate Serv.* 23, 100240 <https://doi.org/10.1016/J.CLISER.2021.100240>.
- Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H., Monge-Sanz, B.M., 2019. SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.* 12, 1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>. URL: <https://www.geosci-model-dev.net/12/1087/2019/>.
- Kirtman, B.P., Pegion, K., DelSole, T., Tippett, M., Robertson, A.W., Bell, M., Burgman, R., Lin, H., Gottschalck, J., Collins, D.C., & Others (2017). The subseasonal experiment (SubX). URL: <https://doi.org/10.7916/D8PG249H>.
- Lledó, L., Doblas-Reyes, F.J., 2020. Predicting daily mean wind speed in Europe weeks ahead from mjo status. *Mon. Weather Rev.* 148, 3413–3426. <https://doi.org/10.1175/MWR-D-19-0328.1>. URL: <https://journals.ametsoc.org/view/journals/mwr/148/8/mwrD190328.xml>.
- Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N., Doblas-Reyes, F.J., 2020. Choices in the verification of S2S forecasts and their implications for climate services. *Mon. Weather Rev.* 148, 3995–4008. <https://doi.org/10.1175/MWR-D-20-0067.1>. URL: <https://journals.ametsoc.org/mwr/article/148/10/3995/353477/Choices-in-the-Verification-of-S2S-Forecasts-and>.
- Manubens, N., Ho, A.-C., Perez-Zanon, N., & BSC-CNS (2022). startR: Automatically Retrieve Multidimensional Distributed Data Sets. URL: <https://earth.bsc.es/gitlab/es/startR/> r package version 2.2.0-1.
- Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., Doblas-Reyes, F.J., 2016. Seamless management of ensemble climate prediction experiments on hpc platforms. In: *High Performance Computing & Simulation (HPCS), 2016 International Conference on*. IEEE, pp. 895–900.
- Manzanas, R., Gutiérrez, J.M., Bhend, J., Hemri, S., Doblas-Reyes, F.J., Torralba, V., Penabaz, E., Brookshaw, A., 2019. Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. *Clim. Dyn.* <https://doi.org/10.1007/s00382-019-04640-4>.
- Mariotti, A., Baggett, C., Barnes, E.A., Becker, E., Butler, A., Collins, D.C., Dirmeyer, P.A., Ferranti, L., Johnson, N.C., Jones, J., Kirtman, B.P., Lang, A.L., Molod, A., Newman, M., Robertson, A.W., Schubert, S., Waliser, D.E., Albers, J., 2020. Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Am. Meteorol. Soc.* 101, E608–E625. <https://doi.org/10.1175/BAMS-D-18-0326.1>. DOI: 10.1175/BAMS-D-18-0326.1.
- Mayer, K.J., Barnes, E.A., 2021. Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophys. Res. Lett.* 48 <https://doi.org/10.1029/2020GL092092> URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092092>. arXiv: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL092092>. E2020GL092092 2020GL092092.
- 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., Ferranti, L., Hlynia, T., Kumar, A., Müller, W.A., Rixen, M., Robertson, A.W., Smith, D.M., Takaya, Y., Tuma, M., Vitart, F., White, C.J., Alvarez, M.S., Ardilouze, C., Attard, H., Baggett, C., Balmaseda, M.A., Beraki, A.F., Bhattacharjee, P.S., Bilbao, R., Andrade, F.M.D., DeFlorio, M.J., Díaz, L.B., Ehsan, M.A., Fragkoulidis, G., Grainger, S., Green, B.W., Hell, M.C., Infanti, J.M., Isensee, K., Kataoka, T., Kirtman, B.P., Klingaman, N.P., Lee, J.Y., Mayer, K., McKay, R., Mecking, J.V., Miller, D.E., Neddermann, N., Ng, C. H.J., Ossó, A., Pankatz, K., Peatman, S., Pegion, K., Perlwitz, J., Recalde-Coronel, G. C., Reintges, A., Renkl, C., Solaraju-Murali, B., Spring, A., Stan, C., Sun, Y.Q., Tozer, C.R., Vigaud, N., Woolnough, S., & Yeager, S. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, 101, E869–E896. URL: <https://doi.org/10.1175/BAMS-D-19-0037.1>. 10.1175/BAMS-D-19-0037.1.
- MeteoSwiss (2017). easyVerification: Ensemble Forecast Verification for Large Data Sets. URL: <https://CRAN.R-project.org/package=easyVerification> r package version 0.4.4.
- Oliver, H., Shin, M., Matthews, D., Sanders, O., Bartholomew, S., Clark, A., Fitzpatrick, B., van Haren, R., Hut, R., Drost, N., 2019. Workflow automation for cycling systems. *Comput. Sci. Eng.* 21, 7–21. <https://doi.org/10.1109/MCSE.2019.2906593>.
- Pegion, K., Kirtman, B.P., Becker, E., Collins, D.C., Lajoie, E., Burgman, R., Bell, R., Delsole, T., Min, D., Zhu, Y., Li, W., Sinsky, E., Guan, H., Gottschalck, J., Joseph Metzger, E., Barton, N.P., Achuthavariar, D., Marshak, J., Koster, R.D., Lin, H., Gagnon, N., Bell, M., Tippett, M.K., Robertson, A.W., Sun, S., Benjamin, S.G., Green, B.W., Bleck, R., Kim, H., 2019. The subseasonal experiment (SubX). *Bull. Am. Meteorol. Soc.* 100, 2043–2060. <https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Perez-Zanon, N., Caron, L.-P., Alvarez-Castro, C., Batte, L., Delgado, C., von Hardenberg, J., Lledo, L., Manubens, N., Palma, L., Sanchez-Garcia, E., van Schaeymbroek, B., Torralba, V., & Verfaillie, D. (2021). CStools: Assessing Skill of Climate Forecasts on Seasonal-to-Decadal Timescales. URL: <https://CRAN.R-project.org/package=CStools> r package version 4.0.1.
- R Core Team, 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramon, J., Lledó, L., Torralba, V., Soret, A., Doblas-Reyes, F.J., 2019. What global reanalysis best represents near-surface winds? *Q. J. R. Meteorol. Soc.* 145, 3236–3251. <https://doi.org/10.1002/qj.3616>. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/qj.3616> <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3616> <https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3616>.
- Robertson, A.W., Kumar, A., Peña, M., Vitart, F., 2015. Improving and promoting subseasonal to seasonal prediction. *Bull. Am. Meteorol. Soc.* 96, ES49–ES53. <https://doi.org/10.1175/BAMS-D-14-00139.1>. URL: <https://journals.ametsoc.org/view/journals/bams/96/3/bams-d-14-00139.1.xml>.
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., Ek, M., Meng, J., Yang, R., As, M., Mendez, P.A., Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E., 2014. The NCEP Climate Forecast System Version 2. *J. Clim.* 27, 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>. URL: <https://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-12-00823.1>.
- Soret, A., Torralba, V., Cortesi, N., Christel, I., Palma, L.L., Manrique-Suñén, A., Lledó, L., González-Reviriego, N., Doblas-Reyes, F.J., 2019. Sub-seasonal to seasonal climate predictions for wind energy forecasting, vol. 1222. Institute of Physics Publishing, p. 012009. 10.1088/1742-6596/1222/1/012009.
- Specq, D., Batté, L., Déqué, M., Ardilouze, C., 2020. Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific. *Earth Sp. Sci.* 7 <https://doi.org/10.1029/2019EA001003>.
- 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. Climatol.* 56, 1231–1247. <https://doi.org/10.1175/JAMC-D-16-0204.1>. URL: <http://journals.ametsoc.org/doi/10.1175/JAMC-D-16-0204.1>.
- Trenary, L., Delsole, T., Tippett, M.K., Pegion, K., 2017. A new method for determining the optimal lagged ensemble. *J. Adv. Model. Earth Syst.* 9, 291–306. <https://doi.org/10.1002/2016MS000838>.
- Vigaud, N., Robertson, A.W., Tippett, M.K., 2017. Multimodel Ensembling of Subseasonal Precipitation Forecasts over North America. *Mon. Weather Rev.* 145, 3913–3928. <https://doi.org/10.1175/MWR-D-17-0092.1>. URL: <http://journals.ametsoc.org/doi/10.1175/MWR-D-17-0092.1>.
- Vitart, F., 2004. Monthly Forecasting at ECMWF. *Mon. Wea. Rev.* 132, 2761–2779. <https://doi.org/10.1175/mwr2826.1>.

- Vitart, F., 2014. Evolution of ECMWF sub-seasonal forecast skill scores. *Q.J.R. Meteorol. Soc.* 140, 1889–1899. <https://doi.org/10.1002/qj.2256>. URL: <http://doi.wiley.com/10.1002/qj.2256>.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A.W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.J., Xiao, H., Zaripov, R., Zhang, L., 2017. The subseasonal to seasonal (S2S) prediction project database. *Bull. Am. Meteorol. Soc.* 98, 163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Vitart, F., Buizza, R., Alonso Balmaseda, M., Balsamo, G., Bidlot, J.-R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F., Palmer, T.N., 2008. The new VarEPS-monthly forecasting system: A first step towards seamless prediction. *Q.J.R. Meteorol. Soc.* 134, 1789–1799. <https://doi.org/10.1002/qj.322>. URL: <http://doi.wiley.com/10.1002/qj.322>.
- Vitart, F., Takaya, Y., 2021. Lagged ensembles in sub-seasonal predictions. *Q. J. R. Meteorol. Soc.* 147, 3227–3242. <https://doi.org/10.1002/QJ.4125>.
- White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J., Lazo, J.K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A.J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A.P., Eggen, B., Graham, R., Kjellström, E., Becker, E., Pegion, K., Holbrook, N.J., McEvoy, D., Depledge, M., Perkins-Kirkpatrick, S., Brown, T.J., Street, R., Jones, L., Remenyi, T.A., Hodgson-Johnston, I., Buontempo, C., Lamb, R., Meinke, H., Arheimer, B., & Zebiak, S.E. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol. Appl.*, URL: <http://doi.wiley.com/10.1002/met.1654>. DOI: 10.1002/met.1654.
- White, C.J., Domeisen, D.I.V., Acharya, N., Adefisan, E.A., Anderson, M.L., Aura, S., Balogun, A.A., Bertram, D., Bluhm, S., Brayshaw, D.J., Browell, J., Büeler, D., Charlton-Perez, A., Chourio, X., Christel, I., Coelho, C.A.S., DeFlorio, M.J., Monache, L.D., Giuseppe, F.D., García-Solórzano, A.M., Gibson, P.B., Goddard, L., Romero, C. G., Graham, R.J., Graham, R.M., Grams, C.M., Halford, A., Huang, W.T.K., Jensen, K., Kilavi, M., Lawal, K.A., Lee, R.W., MacLeod, D., Manrique-Suñén, A., Martins, E. S.P.R., Maxwell, C.J., Merryfield, W.J., Muñoz, Á. G., Olaniyan, E., Otieno, G., Oyedepo, J.A., Palma, L., Pechlivanidis, I.G., Pons, D., Ralph, F.M., Reis, D.S., Remenyi, T.A., Risbey, J.S., Robertson, D.J.C., Robertson, A.W., Smith, S., Soret, A., Sun, T., Todd, M.C., Tozer, C.R., Vasconcelos, F.C., Vigo, I., Waliser, D.E., Wetterhall, F., & Wilson, R.G. (2021). Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Am. Meteorol. Soc.*, -1, 1–57. URL: <https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-20-0224.1/BAMS-D-20-0224.1.xml>. 10.1175/BAMS-D-20-0224.1..