UNIVERSITAT DE BARCELONA

High-quality observations for improved seasonal predictions

Implications for the wind energy sector

Jaume Ramon Gamon

UNIVERSITAT DE BARCELONA

HIGH QUALITY OBSERVATIONS FOR IMPROVED SEASONAL PREDICTIONS. IMPLICATIONS FOR THE WIND ENERGY SECTOR

Jaume Ramon Gamon
PhD Thesis
Universitat de Barcelona

Supervised by: **Llorenç Lledó** and **Albert Soret**



Barcelona, 2021

High-quality observations for improved seasonal predictions. Implications for the wind energy sector.

Memòria presentada per optar al grau de doctor
per la Universitat de Barcelona.

Programa de doctorat en física

Autor: Jaume Ramon Gamon

Directors: Llorenç Lledó Ponsati
Albert Soret Miravet

Tutor: Joan Bech Rustullet

UNIVERSITAT DE BARCELONA

High-quality observations for improved seasonal predictions.
Implications for the wind energy sector.

PhD Thesis
Barcelona, 2021

Author: Jaume Ramon Gamon

Cover photos: Abel and Miriam. Santa Fe del Montseny and Tossa de Mar, 2015.

The research leading to this PhD thesis has been carried out at the Earth Sciences department of the **Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS)**.

The author acknowledges funding support from Ministerio de Ciencia e Innovación (MICINN) **Formación de Personal Investigador (FPI)** grant BES-2017-082216 (“Ayudas para contratos predoctorales”). He also acknowledges funding support from projects **INDECIS** (cofunded by the H2020 ERA-net ERA4CS, GA 690462), **S2S4E** (project funded by the Horizon 2020 framework of the European Commission, GA 776787) and **NEWA** (funded by the Spanish Ministerio de Economía y Competitividad (MINECO) and by the European ERA-NET Plus NEWA).

A Margarita, a Vicentica, i a totes
aquelles que no pogueren estudiar...

... però que tant m'han ensenyat.

Acknowledgements

Abans d'entrar en detall, em permetran introduir unes paraules d'agraïment en la meua llengua. Primerament, cal dir que este treball no haguera sigut possible sense l'ajuda dels meus directors i supervisors. M'agradaria agrair en primer lloc a Llorenç, a qui li dec tantíssim, i qui m'ha instruït i guiat en tot aquest viatge. Perquè el dia que ens vam conéixer i em vas dir que *"la distribució de vent a València és bimodal"* no sabia que et convertiries en una peça clau en la meua carrera acadèmica. Gràcies per haver estat sempre al meu costat —o més ben dit, en diagonal. Per favor, considera este treball com a teu també. A més, vull agrair profundament l'ajuda de Paco i Albert, per regalar-me una miqueta de la seua gran experiència. Gràcies als tres.

Seguidament vull donar les gràcies a tres companyes a qui he tingut el plaer de conéixer estos anys en el BSC, i que han sigut d'importància primera per a mi. A Vero, qui em va rebre en el meu primer dia al BSC —els nervis d'aquell dia no em permeten recordar res més. Tampoc sabia que s'aniria a convertir en una persona de referència per a mi. També vull donar-li gràcies a Núria, per haver-me proporcionat tantíssimes ferramentes de treball, haver-me ensenyat sobre observacions d'estacions, i per tants moments d'alegria. I com no, a Miriam, per tants i tants atacs de riure junts —a vegades davant d'una pantalla—, per les castanyades, per les coques de Sant Joan... gràcies per haver creat este esperit d'unió al departament.

Agrair a tots els companys i companyes del departament de Ciències de la Terra del BSC amb qui he pogut col·laborar. Gràcies sobretot als meus companys de Earth System Services, per fer de l'oficina un lloc on m'he pogut sentir còmode treballant. A Andrea, per tants cafés junts. A Nube, pels seus consells. A Isa, Marina i Diana, per l'ajuda en la creació de la web de les torres. I a Francesc, per compartir el meu *frikisme*.

També vull agrair als companys del grup de CES, liderats per Kim. Gràcies a vosaltres una persona amb nocions molt bàsiques d'informàtica ha pogut desenvolupar la seua tesi en un centre de supercomputació. Gràcies a Pierre-Antoine i Marga, feu una feina d'un valor immens. A Albert Vila i Adrià,

pel suport informàtic quasi immediat. Agrair també la feina d'An-Chi i Nico, havent proporcionat les ferramentes d'R. I també, al departament d'operacions del BSC, per haver-me donat suport informàtic, així com un lloc físic de treball.

Vull donar gràcies a Stefano i Silvio de CMCC, per permetre'm fer una estància a Itàlia que mai va poder ser. M'he sentit molt privilegiat de poder entaular discussions interessants amb César Azorín, José Guijarro, Enric Aguilar i Chris Ferro. I agrair tots els comentaris i revisions rebudes en els tres articles que ací es presenten. Persones anònimes que han contribuït a incrementar la qualitat d'este treball.

Gràcies també a tots aquells amb qui he pogut passar tants moments de diversió. A 'la gent del tupper', al *morning coffee*, i sobretot a aquells amb qui em quedava cada divendres mitja horeta després de dinar per fer alguna partida a jocs de taula: a Saskia, a Lluís, a Marcus... gràcies per haver-me fet sentir tan a gust amb vosaltres. Molt especialment vull agrair a Carlos, Cristina, Jesús i Marina, probablement amb els qui més temps he passat durant estos anys a Barcelona. Companys de pis que acaben convertint-se en bons amics. També a Bala i Bianca, per compartir el seu temps lliure amb mi.

També vull agrair a tot el professorat que ha contribuït a la meua formació. Molt especialment, vull agrair al professor Enric Valor, de la Universitat de València; i a la professora Maria Rosa Soler i el doctor Àlex Montornès de la Universitat de Barcelona. Ells em van introduir en el món de la météo i l'eòlica. Als meus companys de classe de física, a la penya l'Espardenya: Victòria, Sílvia, Míriam, Arcadi, Andrea, Miguel, Josep, Núria, Óscar i Joan, que bé m'ho passe en eixes MEFs!

Finalment, agrair a la meua família tot el suport que m'han donat. Tan lluny però sempre presents. A Tere i Doménec, els meus pares, el meu pensament. A les iaies, a les quals va dedicada esta tesi. A Josep Maria, qui em va injectar la curiositat científica. I a tots els tios/es, cosins/nes i demés gent d'Estivella, la meua terra, la meua gent, les nostres nits a la fresca.

Sis-cents metres d'alçada. Casat amb l'Albarda i
no hi ha mort que els separe. Tot i que són més
alts el puntal de l'Abella o l'Alt del Pi, no hi fa res.
No tenen la seua màgia...

J. M. Blasco en *Estivella, un encontre*. 1991

Siempre acabamos llegando a donde nos esperan.

J. Saramago en *El viaje del elefante*. 2008

E quando pensi che sia finita
è proprio allora che comincia la salita.

A. Venditti en *Che fantastica storia è la vita*. 2003

Contents

List of Figures

List of Tables

Acronyms

AMV Atlantic Multidecadal Variability.

CDF Cumulative Density Function.

CMIP Coupled Model Intercomparison Project.

CRPS Continuous Ranked Probability Score.

DCPP Decadal Climate Prediction Project.

ECV Essential Climate Variables.

ENSO El Niño/Southern Oscillation.

GFCS Global Framework for Climate Services.

GHGs Greenhouse Gases.

IPO Interdecadal Pacific Oscillation.

MJO Madden-Julian Oscillation.

ML Machine Learning.

RPS Ranked Probability Score.

S2S2D Sub-seasonal to Seasonal to Decadal.

SST Sea Surface Temperature.

WCRP World Climate Research Program.

WMO World Meteorological Organization.

Extended abstract

Multiple initiatives are being implemented to mitigate and, in the worst-case scenarios, adapt society to climate change. The vast majority consider renewable energies key to accomplishing a necessary transition from fossil fuels to clean energies. The electricity system, in particular, is facing a significant transformation, being it more dependent on renewable production and, subsequently, on meteorological factors like wind speed or solar radiation. The prediction of anomalies of meteorological variables is well-established and trusted from minutes to days ahead, and so is the amount of renewable generation. Beyond those time scales, seasonal predictions start to produce beneficial results in anticipating the amount of generation months in advance, but their quality is still far from that offered by weather forecasts. In this regard, the climate community is advancing towards better seasonal predictions, both from the perspective of climate modelling and its post-processing.

To further increase the value of seasonal predictions, climate services have recently appeared to make climate information —sometimes deemed challenging to digest— more understandable and practical for non-experienced users. Climate services facilitate the integration of seasonal predictions into the renewable industry. The wind power industry, for example, employs seasonal predictions not only to advance the future availability of the wind resource but also to schedule maintenance activities in wind farms. A better understanding of the opportunities of seasonal predictions allows wind energy users to identify gaps and report specific needs. This PhD thesis looks into those user needs to improve the quality of seasonal predictions for wind speed.

More specifically, the enhancement of seasonal predictions is achieved from the perspective of wind observations. We first focus on wind records measured at tall meteorological towers, a non-standard type of climate data widely used within the wind industry. We identify, retrieve and collect climate records from 222 tall tower locations distributed worldwide. After unifying the data format

and performing an exhaustive quality control, specifically designed for this type of wind data, we release the dataset under the name of The Tall Tower Dataset. The data collection is made publicly accessible through a data web portal. We later explore reanalysis datasets to quantify how they differ from the true observed wind speeds. We consider five global reanalyses and describe their agreements and discrepancies in representing surface wind speeds. By comparing reanalysis data against winds from The Tall Tower Dataset, we conclude that representativeness errors in reanalyses can be large sometimes, to the extent not to trust gridded estimates in specific areas. We also conclude that ERA5 shows the closest wind speed estimates to those observed at the tall towers.

Once wind observations are characterised, and their quality is ensured to be sufficiently high to produce robust results, they are used to enhance seasonal predictions. The hybrid seasonal forecasts provided in this work allow predicting near-surface wind speeds at a point scale — e.g. wind farm location. Those forecasts rely on the information of the large-scale atmospheric circulation, summarised in the state of the four main Euro-Atlantic Teleconnections. In general, hybrid predictions show skill at lead times two and three, while dynamical predictions do not. Another aspect that is improved is the skill assessment of seasonal predictions. We illustrate the strong dependency of the score estimates, namely the Brier Score, on the choice of the observational reference. This has implications in, for example, the selection of the best prediction system among a set of possible candidates. To solve this issue, we consider two methodologies already proposed in the literature and apply them to seasonal predictions for wind speed. We evaluate their strengths and weaknesses to end up recommending the use of the observation-error-corrected scoring rules.

Resum ampliat

Són múltiples les iniciatives que s'estan implementant per mitigar i, en el pitjor dels escenaris, adaptar la societat al canvi climàtic. La gran majoria consideren les energies renovables claus per tal d'acomplir una transició necessària des dels combustibles fòssils a les energies netes. El sistema elèctric, per exemple, està patint una transformació significativa, ja que és molt més depenent de la producció renovable i, per tant, de factors meteorològics com la velocitat del vent o la radiació solar. La predicció d'anomalies de variables meteorològiques està ben assentada i es confia en ella en escales temporals des d'uns quants minuts a dies vista. Més enllà, les prediccions estacionals comencen a produir resultats beneficiosos pel que fa a l'anticipació de la quantitat de producció a mesos vista, però la seva qualitat encara està lluny de la què ofereixen els pronòstics meteorològics. En aquest sentit, la comunitat climàtica avança cap a la producció de prediccions estacionals millorades, tant des de la perspectiva de la modelització climàtica com del seu post-processat.

Per tal d'incrementar encara més el valor de les prediccions estacionals, els serveis climàtics han aparegut recentment per tal de fer la informació climàtica —de vegades considerada difícil de digerir— més entenible i pràctica per a usuaris inexperts. Els serveis climàtics faciliten la integració de les prediccions estacionals en la indústria renovable. La indústria eòlica, per exemple, utilitza les prediccions estacionals no solament per predir la disponibilitat futura del recurs eòlic, sinó també per planificar el manteniment dels parcs. Un millor enteniment de les oportunitats que ofereixen les prediccions estacionals permet als usuaris de la indústria eòlica identificar manques i reportar necessitats específiques. La present tesi doctoral investiga aquestes necessitats a fi de millorar la qualitat de les prediccions estacionals de vent.

Més concretament, la millora de les prediccions estacionals s'aconsegueix des de la perspectiva de les observacions de vent. Primerament, ens centrem en dades de vent mesurades en torres

meteorològiques altes, un tipus de dades climàtiques que no són estàndard però que s'utilitzen àmpliament dins de la indústria eòlica. Identifiquem, recuperem i col·lectem dades climàtiques de 222 torres altes localitzades arreu del món. Després d'unificar el format de les dades i portar a terme un exhaustiu control de qualitat de les mateixes, específicament dissenyat per a aquest tipus de dades, publiquem el conjunt de dades sota el nom de The Tall Tower Dataset. La col·lecció de dades està disponible de forma pública a través d'un portal de dades. Després, explorem dades de reanàlisi per tal de quantificar la diferència d'aquestes respecte de les dades reals observades. Considerem cinc reanàlisi globals, i descrivim les principals similituds i diferències entre ells a l'hora de representar la velocitat del vent en superfície. Comparant les dades de reanàlisi amb les dades del Tall Tower Dataset, concloem que els errors de representativitat són grans en els reanàlisi, fins al punt de no poder confiar en aquestes estimacions de vent en àrees específiques. També concloem que ERA5 mostra les estimacions de vent més properes a aquelles observades a les torres altes.

Una vegada les observacions de vent han sigut caracteritzades, i s'ha assegurat que la seva qualitat és suficientment alta per produir resultats robustos, poden ser utilitzades per tal de millorar les prediccions estacionals. Les prediccions estacionals híbrides proporcionades en aquest treball permeten predir la velocitat del vent a prop de la superfície a escala puntual (per exemple, a un parc eòlic). Aquestes prediccions es basen en la informació de la circulació general atmosfèrica, resumida en l'estat de les quatre teleconnexions Euro-Atlàntiques principals. En general, les prediccions híbrides mostren habilitat per predir el clima futur (també anomenat *skill*) a dos o tres mesos vista, cosa que les prediccions dinàmiques de velocitat de vent no fan. Un altre aspecte que s'ha millorat és l'avaluació de l'*skill* de les prediccions estacionals. Il·lustrem la forta dependència de les estimacions dels valors d'*skill* amb l'elecció de la referència observacional. Això té implicacions en, per exemple, la selecció del millor sistema de predicció entre un conjunt de possibles candidats. Per solventar-ho, considerem dos metodologies que ja han estat proposades en la literatura i les apliquem a les prediccions estacionals de velocitat de vent. Avaluem els seus punts forts i febles, per acabar recomanant l'ús de les mètriques que han sigut degudament redefinides per incorporar l'error observacional.

Chapter 1

Introduction

1.1 Renewable energies: current scene and future targets

The European Union aims at becoming climate neutral by 2050: as much carbon would have to be absorbed as released into the atmosphere. The so-called European Green Deal (European Commission, 2019) is an ambitious transformation that needs that all sectors of our economy take action in order to reduce the magnitude and effects of climate change. The energy sector, for example, is required to continue with the decarbonisation for it to be based largely on renewable sources.

The European Green Deal is in line with other national initiatives like the compromise of the United States to reach net zero emissions economy-wide by no later than 2050[1], introducing renewable sources for energy production in key sectors like transportation or industry.

In 2019, the share of renewables in gross final energy consumption stood at 19.7% in the European Union (Figure 1.1 left). This percentage has doubled since 2004, mirroring the growth in installed power capacity and the rising investments in clean energies. This is indeed the way forward to meet the 38–40% target set by the Green Deal agreement, which has been confirmed by the European Union[2]. Among renewable energies, wind power has recently become the main source of electricity generation (Figure 1.1 right), overtaking hydropower which has remained steady since the 1990s.

Nevertheless, the integration of renewable energies into the electricity mix for consumption poses some risks. Unlike conventional power plants, the power output from renewable plants cannot be fully controlled because it is inter-

[1] https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/22/fact-sheet-president-biden-sets-2030-greenhouse-gas-pollution-reduction-target-aimed-at-creating-good-paying-union-jobs-and-securing-u-s-leadership-on-clean-energy-technologies/

[2] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_3541

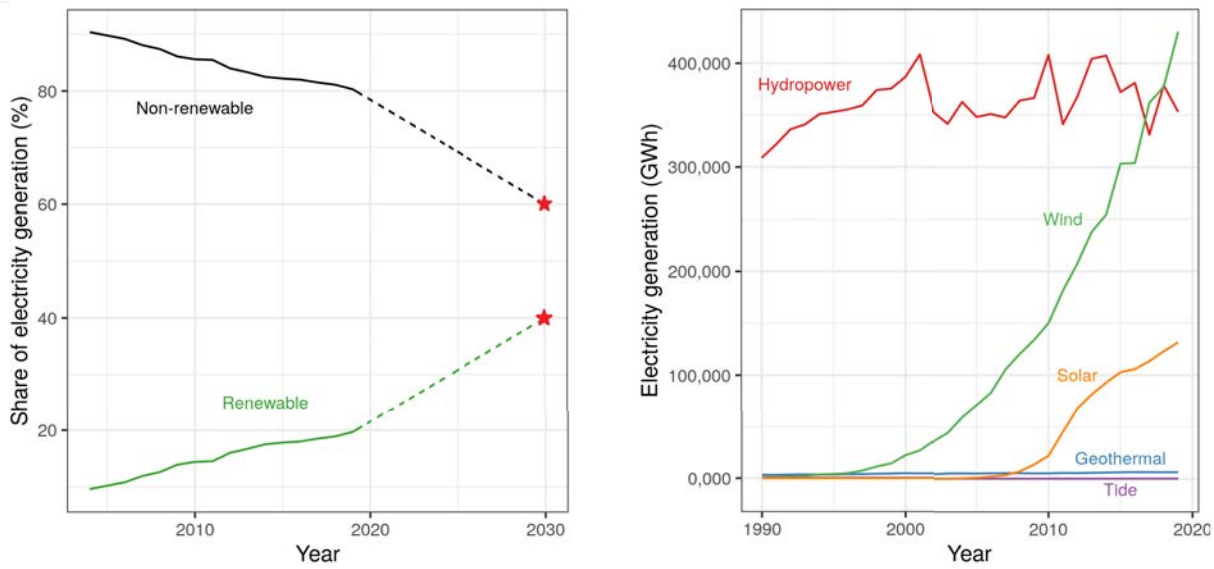


Figure 1.1: Left: yearly evolution of the share of renewables and non-renewables in gross final energy consumption (in percentage) in the European Union. The two red stars indicate the Green Deal's target by 2030. Right: yearly evolution of electricity generation (in GWh) by the different renewable sources in the European Union. Source data: Eurostat and International Energy Agency.

mittent and highly dependent on meteorological factors like wind speed or solar radiation. Moreover, nowadays, it can be stored only for hours or a few days at maximum. Prolonged lack of renewable resource may then threaten the availability of the renewable energy supply, as it occurred in parts of Europe during winter 2016-2017 (see the detailed case study in Box I). Since the shortage of renewable production is backed up by fossil-fuel-based energy, a total disconnection of conventional power plants is currently inconceivable without timely and accurate anticipation of the future availability of the renewable resource.

1.2 Climate prediction

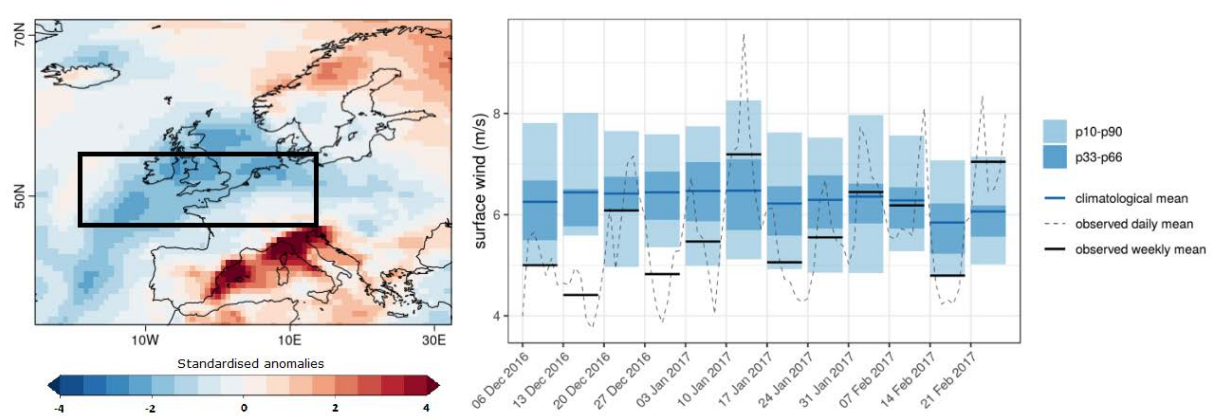
The prediction of the amount of future renewable production can be much specified with the prediction of the anomalies of Essential Climate Variables (ECV) like wind speed or solar radiation. These anomalies can be predicted at a wide range of timescales: from a few minutes ahead to daily or monthly averages and up to seasonal or centennial scales. At a weekly to decadal timescales, forecasts of future wind speeds benefit many stakeholders in the wind energy sector. These are referred to as climate predictions (Figure 1.2), which can be used to schedule maintenance activities months in advance, anticipate weekly to decadal total generation or advance electricity prices at the different timescales.

Box I: Wind speed variability and its implications in energy production

Around mid-January 2017, a cold wave affected large areas in central and southern Europe. Cold spells are common in winter and usually increase electricity demand to power the heating of houses. The energy system is already prepared to absorb this peak demand to guarantee the correct electricity supply. That winter, however, a combination of several climatic factors threatened the electricity supply leading to a high-risk situation (S2S4E, 2020).

The situation was especially critical in France, where the cold event found several nuclear reactors shut down to maintain and check some of their components (APPLICATE, 2020). In a country where around 70% of the total electricity generated comes from nuclear power, searching for alternative power sources became a priority at that moment. Renewables (which may represent up to 18%), and in particular wind power, could have been an asset to cover the demand. However, wind speeds were weaker than average in France during that week and in five out of the six previous (see figure below), so the renewable energy supply was low too. In the end, the electricity system could cope with the high demand thanks in part to stock energy and the imports from neighbouring countries. Nuclear reactors were back to production a few days later and winds recovered by the end of January.

With this example, we merely intend to illustrate the direct impact of climate on energy and the fact that renewable energy production is shaped by atmospheric variability (Bloomfield et al., 2016). Further persistence of this combination of low temperatures and weak winds could have collapsed the electricity system in France. Could not have this anomalous climate event been foreseen weeks or months before? Accurate anticipation of future conditions of ECV such as temperature or wind speed appears to be vital to estimating future renewable energy production, planning maintenance activities, anticipating revenues, and ultimately guaranteeing the balance between supply and demand.



Left: surface wind speed anomaly field for the 17th to 23rd of January 2017 in Europe. The black rectangle indicates the region where the weekly anomalies shown in the right plot are computed. Right: weekly anomalies of surface wind speed averaged over the region [5°W, 12°E; 47°N, 54°N] compared to the corresponding observed weekly values. In both cases, data is sourced from the ERA-Interim reanalysis and the reference period is 1980-2017. Source: S2S4E (2020).

1.2.1 Modelling the Earth System: dynamical vs empirical models

Accurate climate models are the crucial element to produce high-quality climate predictions. They can be divided into dynamical and empirical models, both equally employed in the climate prediction arena.

Dynamical climate models build on a set of equations to represent the climate system as a whole and its processes. The so-called *primitive equations* are initialised with climate observations (*initial value* problem), which represent the initial state of the climate system. Then, the model is run towards the future to simulate the evolution of the climatic fluctuations.

Since many elements of the Earth system are represented in the dynamical climate models, they are formally referred to as climate prediction systems. The atmosphere, for example, is included analogously as weather forecast models do. Additional components of the climate system such as the ocean, the land surface or the sea ice are also incorporated depending on the forecast horizon being targeted. Indeed, the latest developments in seasonal to decadal forecasting involve fully coupled models of the ocean and atmosphere and the two-way interaction between them (Haarsma et al., 2020).

In contrast to dynamical models, empirical climate predictions are based on past experience, or in the modern era, on a statistical-mathematical relationship between observational data of current and past climate states. Empirical predictions originated before dynamical forecasts, more than a hundred years ago, with fairly simple techniques, like persistence (i.e., current weather/climate is predicted repeatedly over a period of time). Current empirical forecasts' methods evolved into more sophisticated methods like regression analysis (Wilks (2011), chapter 7) or analogs (Van Den Dool, 1994), and have been shown to be as accurate as dynamical predictions (Turco et al., 2017; Wang et al., 2017). Section 1.5.3 expands on the different statistical methods applied to empirical seasonal predictions.

1.2.2 The S2S2D scales: from subseasonal to decadal predictions

Climate forecasting relies on the *predictability*[3] provided by the wide variety of climatic fluctuations that exist at the Sub-seasonal to Seasonal to Decadal (S2S2D) scales. In dynamical climate models, the initial state of such oscillations is provided to the climate model as initial conditions. However, their information is gradually lost as the forecast time increases (Figure 1.3). That is why climate predictions at the longest time scales also benefit from the extended predictability given by external forcings that other elements of the Earth system exert as boundary conditions (e.g. anomalies of sea

[3] Defined as "*the extent to which an informative prediction is possible if an optimum procedure is used*" (Doblas-Reyes et al., 2013).

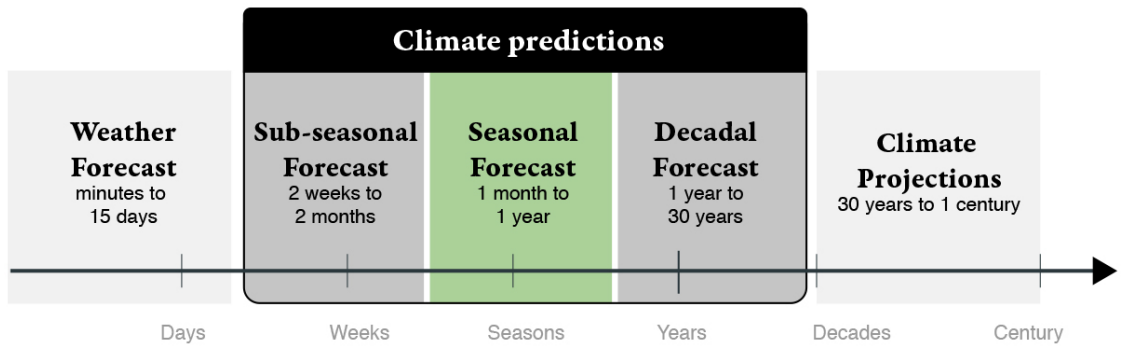


Figure 1.2: Time horizon of different short to medium range forecasts, climate predictions and climate projections. Seasonal forecasts are highlighted as they are the forecasts used in this thesis. Adapted from Torralba (2019).

ice extent or sea surface temperature). Therefore, it is necessary to subdivide climate predictions according to their main sources of predictability and configure the climate models appropriately. The subdivision yields three categories: sub-seasonal, seasonal and decadal predictions (Figure 1.2).

Sub-seasonal predictions extend out weather forecasts by targeting all the processes that occur on a timescale of two to eight weeks. At these timescales, the effect of the atmospheric initial conditions has substantially decreased, but the influence of some of the slow-moving components of the Earth system like the ocean is still not dominant. However, some climatic oscillations, and in particular the Madden-Julian Oscillation (MJO), have been reported to provide significant predictability in the tropics and extratropics (Zhang, 2013; Lledó and Doblas-Reyes, 2020). Other potential sources of predictability at the sub-seasonal scale are the soil moisture (Koster et al., 2011), the sea ice state (Jeong et al., 2013), the stratosphere–troposphere interactions (Domeisen et al., 2020) and tropical–extratropical teleconnections (Vitart et al., 2015).

At the other end of the spectrum of climate predictions, decadal forecasts encompass predictions on annual, multi-annual to decadal timescales. Predictability at these timescales is given by decadal climatic oscillations like the Atlantic Multidecadal Variability (AMV) (Christensen et al., 2013) or the Interdecadal Pacific Oscillation (IPO) (Salinger et al., 2001), which are the main modes of variability in the Atlantic and Pacific, respectively. External forcings from Greenhouse Gases (GHGs), aerosols and volcanoes also provide predictability to decadal forecasts (van Oldenborgh et al., 2012), thanks to which the future state of climate variables can be inferred. Decadal prediction studies have focused largely on tem-

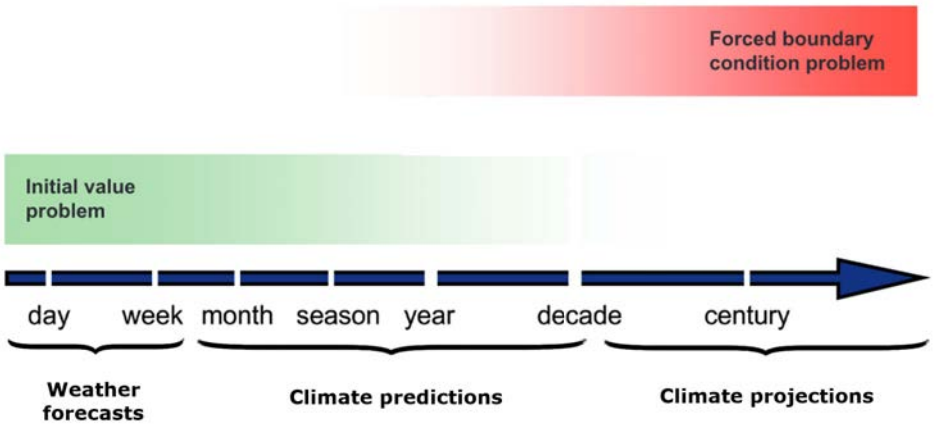


Figure 1.3: Schematic representation of the impact of the initial conditions and boundary forcings on weather forecasts, climate predictions and climate projections. Adapted from Intergovernmental Panel on Climate Change (2013)

perature, whose annual mean variations have been proven to be predictable over much of the globe (Wu et al., 2019). There is currently less skill in predicting wind speed or precipitation, although progress is expected to be made as a consequence of the Decadal Climate Prediction Project (DCPP) (Boer et al., 2016) of the World Climate Research Program (WCRP) and other initiatives and investigations.

The seasonal timescale falls between the sub-seasonal and decadal timescales, i.e., between more than one month and slightly longer than one year. In this prediction range, the information for the prediction comes from both initial conditions and external forcings (Figure 1.3). The sources of predictability involve long-lived phenomena that have much longer variability timescales than those in sub-seasonal and include details on the state of the ocean (Kushnir et al., 2006), sea ice (Guemas et al., 2016) and land surface (Prodhomme et al., 2016b). The main mode of variability at the seasonal timescale, and hence the main source of predictability, is the El Niño/Southern Oscillation (ENSO), a two-way ocean-atmosphere interaction in the tropical Pacific whose fluctuations on the Sea Surface Temperature (SST) impact the climate conditions worldwide (Doblas-Reyes et al., 2013). Although seasonal predictions have been developed for more than 30 years (Cane et al., 1986), their quality is still improvable. In this regard, seasonal predictions will be the focus of this PhD thesis.

1.2.3 Addressing the uncertainties in climate prediction

During recent decades, much progress has been made to understand the climate system

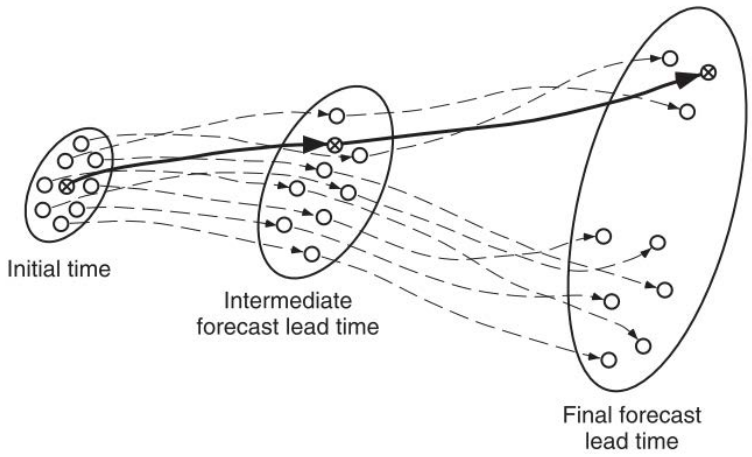


Figure 1.4: Schematic representation of the concepts of probabilistic and deterministic forecasts. The solid line represents the evolution of the deterministic forecast, whose initial state (initial time) is known with certainty. The dashed lines represent the evolution of the different ensemble members of the probabilistic forecast, originating from various perturbations of the initial state. The growing size of the ellipses respond to the deviation of the ensemble trajectories, illustrating the increasing uncertainty from the initial to the final forecast time. Source: Wilks (2011).

and the processes occurring at the S2S2D scales (Merryfield et al., 2020). The climate community has devoted efforts to translating all this knowledge into climate models to generate predictions. Despite the continuous attempts, the accuracy of those predictions is rather limited due to the atmosphere's non-linearity and its chaotic character. Consequently, any user of climate prediction data has to be aware of their associated uncertainty which, in practice, comes mainly from errors in the climate model and its initialisation.

As stated by Slingo and Palmer (2011), it is essential to distinguish between two types of model error: the uncertainty coming from the imperfect knowledge of the climate system and the uncertainty arising from the sub-grid scale phenomena that, although understood, are misrepresented in the climate model due to its limited resolution (representativeness errors). Circumventing such constraints requires significant investments in computational resources (Shukla et al., 2010).

Aside from model error, a considerable proportion of uncertainty comes from the observations employed as initial conditions since they are imperfectly known. In 1963, Lorenz showed that the uncertainty in the initial conditions, however small, leads to a particular uncertainty in the forecasts after a period, responding again to the climate system's chaotic nature (Lorenz, 1963). In other words, minimal perturbations in the initial conditions are amplified by chaotic processes so that diverging forecasts are obtained.

Considering such uncertainty is the basis for employing probabilistic forecasts and ensemble prediction systems.

In a probabilistic forecast, predictions indicate the probability of occurrence of a specific event rather than a single value. This is possible because probabilistic predictions provide multiple outcomes (realisations) for every predicted event, resulting from the several calculations generated after performing minimal perturbations to the initial conditions and/or model parametres (Figure 1.4). This set of realisations or ensemble members generates a distribution that allows for inferring probability estimates to each predicted event. For example, a prediction of 70% chance of above-average rain in October mirrors that 70% of the realisations indicated above-normal rain. Thus, although most of the ensemble members indicate above-average rain in October, one cannot be fully certain that the event will eventually occur.

1.3 The importance of developing a climate service

Understanding the probabilistic nature of climate predictions can sometimes be challenging for the vast majority of users from outside the climate prediction field. For instance, in the wind energy sector, most of the decisions the users deal with are yes/no measures (e.g. shut down a wind turbine above/below a certain wind speed threshold), thus not being used to work with forecast probabilities and what they imply. Besides, other barriers make climate information difficult to digest for non-experienced users. One of them is the lack of a common and widely accepted terminology between climate scientists and user communities, making the language sometimes very technical and misleading. Another instance is the low quality of climate predictions compared to weather forecasts, being the latter much more reliable and integrated into the decision-making processes. Lastly, climate predictions are issued at global and coarse scales, hindering their usage from anticipating anomalies at a fine scale (e.g. power plant scale).

Fortunately, many initiatives have appeared to increase understanding and *value* of climate information and make it relevant to different users (McNie, 2007; Buontempo et al., 2014). All these actions fall under the umbrella of climate services. Climate services translate and interpret the climate information generated by climate scientists into tailored products and tools that can be integrated into practical applications for both society and industry. Such is the importance of climate services that the WMO motivated the creation of the GFCS, a coordinated effort to strengthen the engagement between scientists and users for the integration of climate information in users' decision-making processes (Hewitt et al., 2012). The implementation of the GFCS has five pillars (Figure 1.5) or components:

1. **Observations and Monitoring:** collection

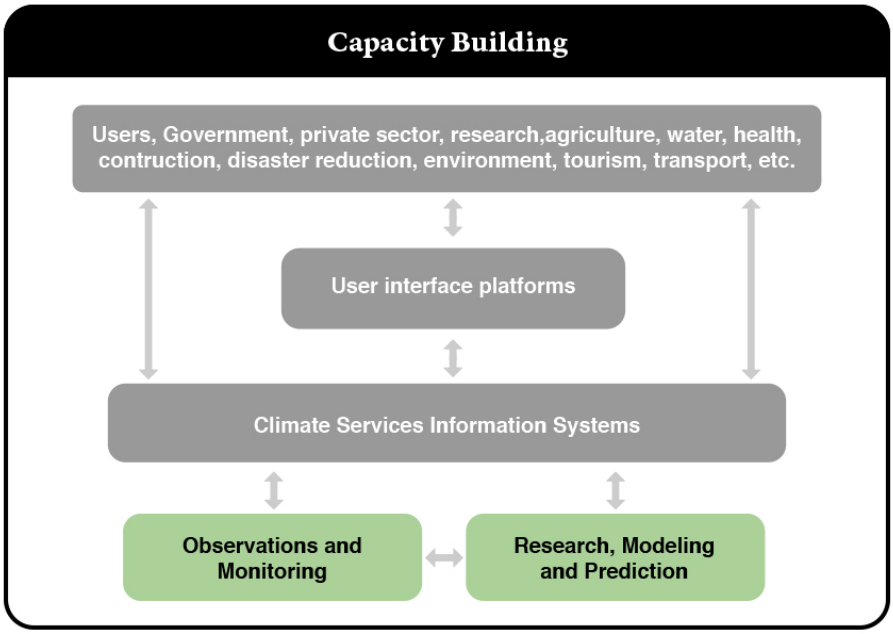


Figure 1.5: The five pillars of the GFCS and the interactions between them and the priority areas/sectors. This thesis will focus on the two pillars highlighted in green. Adapted from WMO.

of climate data and development of standards to be used in their generation.

2. **Research, Modeling and Prediction:** further understanding of the climate system dynamics and change.
3. **Climate Services Information Systems:** generate, protect and distribute climate information according to the users' needs.
4. **User Interface Platforms:** monitor users' requirements by providing a forum for dialogue.
5. **Capacity Building:** provide support for the effective development of climate services.

This thesis will focus on pillars 1 and 2. Firstly, the collection of observations of various types and adequate quality and format is required to create effective climate services. Observations are utilised in almost all climate analyses: from attribution studies (i.e., unveil the drivers of a particular event of interest) to preconstructive renewable resource assessments. Secondly, advancing in the art of climate prediction and in particular seasonal forecasting is essential to boost the integration of renewables in the electricity mix. High-quality seasonal predictions are undoubtedly beneficial to many activities within the renewable industry that require previous planning, as well as to anticipate the future availability of the renewable resource. For example, operations involving

a wind farm's maintenance need to be scheduled either avoiding periods of strong winds for safety reasons or during prolonged times with weak winds and minimal production (Brower et al., 2012). Besides, having estimates of renewable energy production weeks or months in advance can help decide whether they are sufficient to satisfy the electricity demand or if alternative power supply sources are needed (Foley et al., 2012). For example, the impact of the cold spell in the electricity system that occurred in France during January 2017 (see Box I) could have been minimised with the correct utilisation of sub-seasonal and seasonal predictions.

1.4 A wide range of wind observations

1.4.1 On-site measurements: a close look to reality

Around the globe, millions of weather observations are recorded each day, by both human observers and automated instruments. To make them usable for later climate analyses or forecasting, these weather observations undergo a quality control process, and are finally stored in datasets to become climate observations.

The Global Historical Climatology Network (GHCN) and the Integrated Surface Dataset (ISD), both hosted by the National Oceanic Atmospheric Administration (NOAA), are two of the largest archives containing point observations from more than 100,000 and 35,000 land surface stations, respectively, multiple variables and spanning up to 175 years (Menne et al., 2012; Smith et al., 2011). Wind series such as those represented in Figure 1.6 can be accessed online through a unique access point[4], and unified formats and quality control. In fact, having such type of common attributes is essential to make a dataset profitable for observational data analyses.

The World Meteorological Organization (WMO) through the Instruments and Methods of Observation Programme (IMOP) has coordinated the establishment of standards and conventions to be followed by the climate community (WMO, 2017). Those guidelines concern the installation and monitoring of measuring stations, as well as the analysis, quality control and storage of those observations. The benefits of such coordination efforts are varied: from facilitating the usage of observational data for climate analyses and forecasts to ensuring fair comparisons against different sets of data.

Besides, observations should be accompanied by the so-called metadata: an up-to-date documentation of a scientific, technical, operational, and administrative nature that involves the observing station and aims at enhancing knowledge of the climate observations (Aguilar et al., 2003). The metadata can be very help-

[4] In the case of GHCN: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn

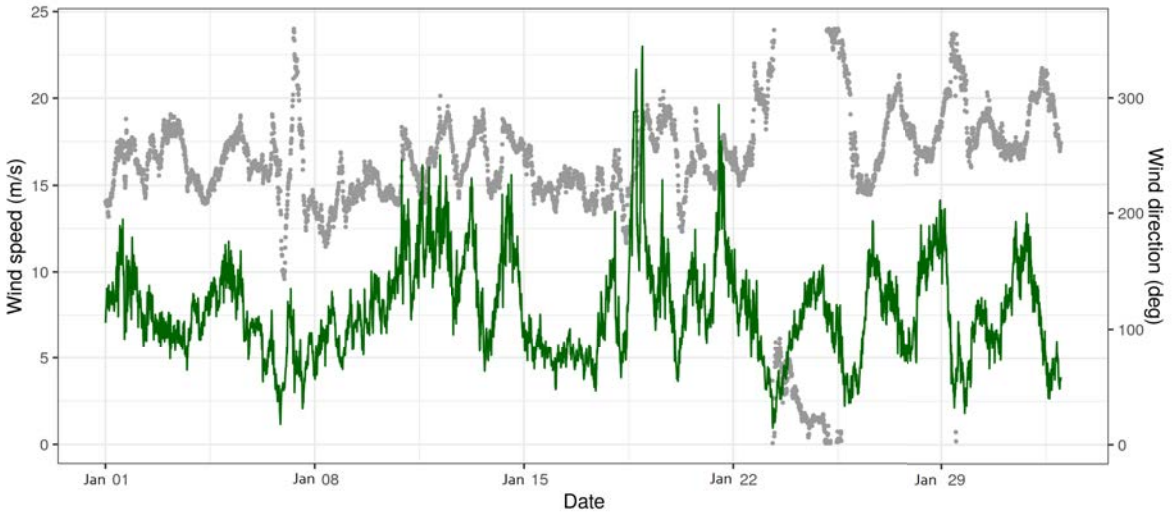


Figure 1.6: Time series of wind speed (green lines) and wind direction (grey dots) at Lindenberg meteorological observatory, Germany [52.17°N, 14.12°E] in January 2007. The peak wind speed observed on the 18th of January corresponds to the passing of windstorm Kyrill. Source data: GHCN and Frank Beyrich (DWD).

ful to identify quality issues —as well as their sources— in the wind series. The importance of a having a complete metadata has triggered the creation of specific initiatives like the Climate and Forecasting (CF) metadata conventions to promote the processing and sharing of climate data files with a unified metadata information (Gregory, 2003).

Apart from land surface stations, other observing systems such as weather radars, ocean buoys, aircrafts or ships equipped with meteorological sensors provide us with climate observations (Figure 1.7). Remote sensing systems are becoming increasingly popular due to their easy installation and cheap maintenance. LiDAR (based on laser detection) and SoDAR (measures the scattering of sound waves) instruments are well known for wind data users, and emerged as efficient alternatives to replace on-site measurements (Lang and McKeogh, 2011). Orbiting the Earth, satellites measure a wide range of climate variables from surface to the upper levels of the atmosphere. The most recent satellite missions have focused on improving accuracy of measurements, to make it comparable to that reported by surface sensors. Still, the recently launched wind-lidar satellite *Aeolus* is reporting systematic errors in the range of 1-5 m/s for tropospheric wind speeds (Baars et al., 2020; Martin et al., 2021).

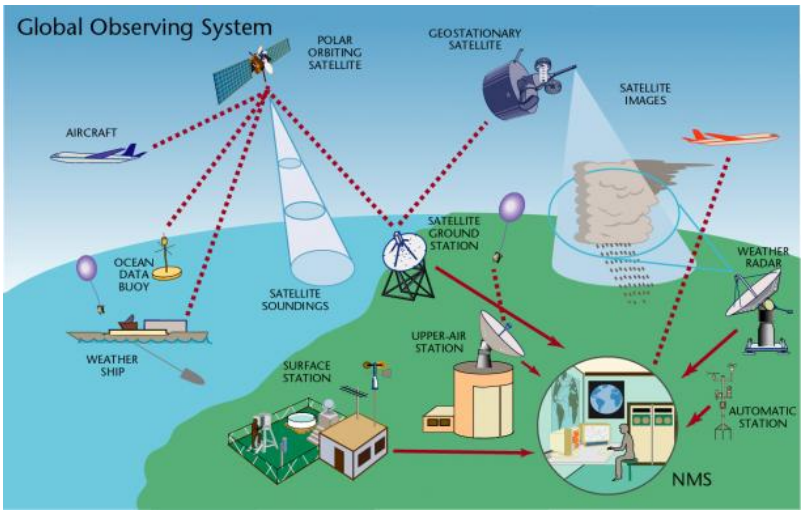


Figure 1.7: The varied types of individual surface- and space-based observing systems. Source: WMO.

1.4.2 Reconstruction of the past climate by reanalyses

Reanalysis products provide physically consistent and continuous reconstructions of spatio-temporal fields of climate variables (Figure 1.8). A reanalysis is generated by combining a numerical model with the assimilation (ingestion) of historical observations (Fujiwara et al., 2017). In this way, most of the error coming from the historical model run is corrected utilising real observations.

The result of such modeling is delivered on a global or regional grid with time series spanning more than 30 years, which makes reanalyses sometimes advantageous over station data. Indeed, reanalyses can provide knowledge of the past atmospheric conditions in remote places where no station data is available, or where long homogeneous records are lacking (Thorne and Vose, 2010).

Reanalyses are a common source of historical data for wind-derived analyses. Attribution studies rely on reanalysis data to provide knowledge about the magnitude and causes of anomalous wind events (Lledó et al., 2018). The wind power industry uses those products in pre-constructive wind farm assessments (Tammelin et al., 2013), as well as to estimate the long-term wind-energy production (Jude and Leseney, 2017). Reanalysis values are also used to assess the quality of gridded forecasts such as seasonal predictions (e.g. Clark et al. (2017)).

1.4.3 Uncertainties in wind observations

Since wind is highly variable in space and time, every observed wind value is inherent to con-

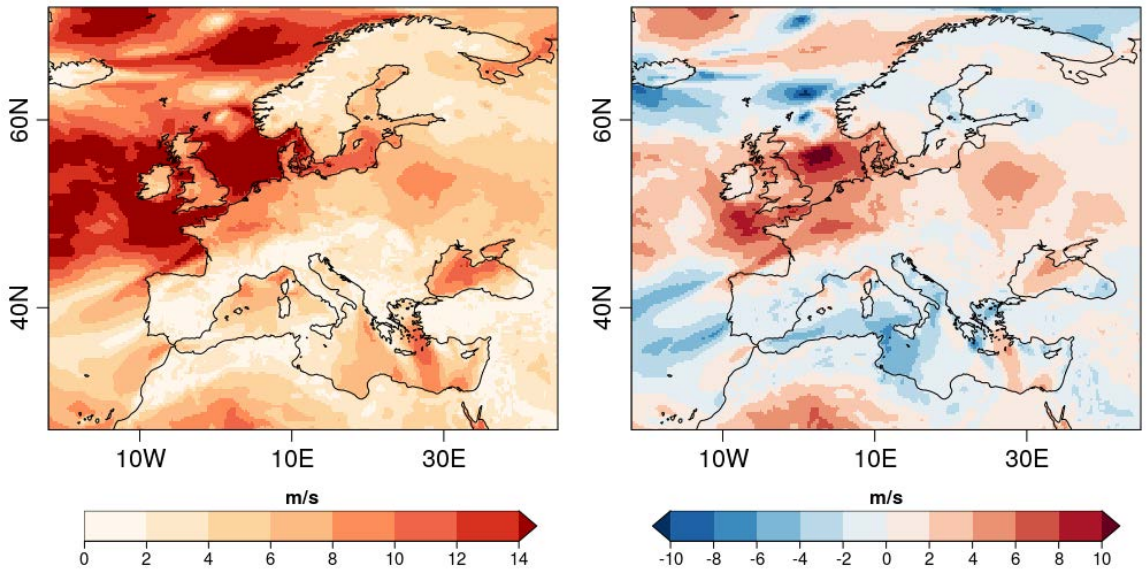


Figure 1.8: Reconstruction of the wind speed field (left) and anomaly field (right) by the ERA5 reanalysis for January 2007. Anomalies are calculated with respect to January mean wind speeds over the period 1981-2020. The passing of various windstorms across northern Europe produced wind speed anomalies of up to +10 m/s over the North Sea.

taining a certain degree of uncertainty. Not all wind observations are the same and some are considered of higher quality than others, but if their quality is known and demonstrable, they can be used appropriately (Table 1.1).

On-site measurements are possibly the most accurate wind data we can obtain. The extent to which they deviate from the *true value* (i.e., the value that would be measured in the absence of error) is diverse depending on the instrument, but usually lesser than that for reanalysis. Therefore, it is little wonder that in-situ measurements are considered perfect in some analyses (e.g. Gubler et al. (2020); Molina et al. (2021)), despite being subject to containing some degree of uncertainty, coming mainly from instrumental errors (WMO, 2017).

The assimilation of uncertain observations by an imperfect model in the generation process of a reanalysis leads necessarily to uncertain response values. Since each reanalysis producing centre uses its own configuration there is an added component of uncertainty coming from (i) the different assimilation methods used in the generation of a reanalysis to nudge fields towards observed values (see Fujiwara et al. (2017) for a review), (ii) the varied parametrisation schemes implemented, and specially those for surface levels (e.g. boundary layer or cloud physics), and (iii) the changing density and quality of the assimilated observations, which may produce inconsistencies leading to spurious low-frequency trends sometimes dif-

Table 1.1: Summary table of the different types of wind datasets employed in this thesis.

	In-situ observations	**Reanalyses**	**Seasonal predictions**
Native grid	Irregular	Gaussian or regular	Gaussian or regular
Timespan	Up to centennial scale	30-50 years typically but also centennial scale	Up to 40 years normally
Time resolution	Sub-secondly	Up to 1-hourly	Up to 6-hourly
Spatial resolution	$\sim$1 m	$\sim$10 km	$\sim$100 km
Main source of uncertainty	Instrumental error	Representativeness error	Model error

ficult to distinguish from the real wind speed trends (Dee et al., 2011a).

However, all these errors originated during the creation of a reanalysis could be small —and even negligible— when compared against the representativeness errors (Janjić et al., 2018). Given the gridded nature of reanalysis data, each datum represents a unique wind estimate for an area of hundreds of square kilometres —suppose a grid cell with horizontal dimensions 30 km x 30 km, which is the finest resolution of the currently available global reanalyses. Hence, those areas with special topographic features at much finer scales (e.g. mountainous areas or coastlines) within a grid cell are not properly represented by the model, and neither are the local-scale winds induced by this fine topography (e.g. anabatic and katabatic winds). The misrepresentation of those winds generates systematic errors (biases) that could be significant (Molina et al., 2021). Being aware of such limitations of reanalysis data is crucial, so that the representativeness error in reanalysis will be investigated further in Chapter 4 of this thesis.

1.5 Searching for accurate seasonal predictions

Advances towards better seasonal predictions have been achieved from two different angles. On one hand, climate scientists have expanded their understanding of the climate processes to eventually incorporate them into seasonal prediction systems. On the other hand, numerous studies have focused on producing a wide range of post-processings to be applied

on the prediction systems' output. These post-processings include the adjustment of the biases of the model output, the improvement of the techniques for the forecast quality assessment and the development of statistical methods to be combined with dynamical predictions.

1.5.1 The need for bias-adjustment

An important issue affecting the quality of seasonal predictions is the model-specific biases that, unlike weather forecast models, grow more strongly in a fully coupled system (Slingo and Palmer, 2011). These biases in the output of a climate model result from the inability of the climate models to numerically reproduce all the relevant processes responsible for climate variability, the initialisation of the primitive equations, and the coarse grids in which seasonal predictions are computed (Table 1.1).

Those biases can become large, to the extent that the distribution of probable outcomes of the seasonal forecast is notably different from that of the observed distribution (Figure 1.9). This inconsistency makes seasonal forecasts less useful, and does not fulfil an acceptable accuracy requirement for their adoption in specific fields like the wind energy sector (Torralba, 2019). The magnitude of those biases can be inferred using the so-called hindcasts of a seasonal prediction system, i.e. retrospective forecasts made with the same configuration as the real-time forecasts. Since observations are available for the hindcast's timespan, hindcasts are compared against an observational reference dataset (e.g. station-based or reanalysis). Once characterised, the biases in the seasonal predictions can be corrected. Over the years, many methods have been proposed to adjust prediction biases (Torralba et al., 2017b; Manzanas et al., 2019), which affect not only to the mean value, but also to other moments of the wind distribution (Marcos et al., 2018).

In particular for seasonal predictions for wind speed, three different bias-adjustment approaches have been proposed in the literature: simple bias correction (Leung et al., 1999), quantile mapping (Themeßl et al., 2011) and calibration (Doblas-Reyes et al., 2005). All these post-processing methods address different aspects of forecast error and produce seasonal forecasts with similar statistical properties to the observed distribution.

Firstly, simple bias correction adjusts predictions to have an equivalent standard deviation and mean to that of the observations:

$$y_{i,j} = (x_{i,j} - \overline{x})\frac{\sigma_{ref}}{\sigma_e} + \overline{o}. \qquad (1.1)$$

In Equation 1.1, $y_{i,j}$ refers to the adjusted forecast for each year i and member j, $x_{i,j}$ is the raw forecast for year i and member j, $\overline{x}$ represents the mean value of all raw forecasts for all years i and members j. σ_{ref} and σ_e refer to the standard deviations of the observations — understood as a reference— and the forecast

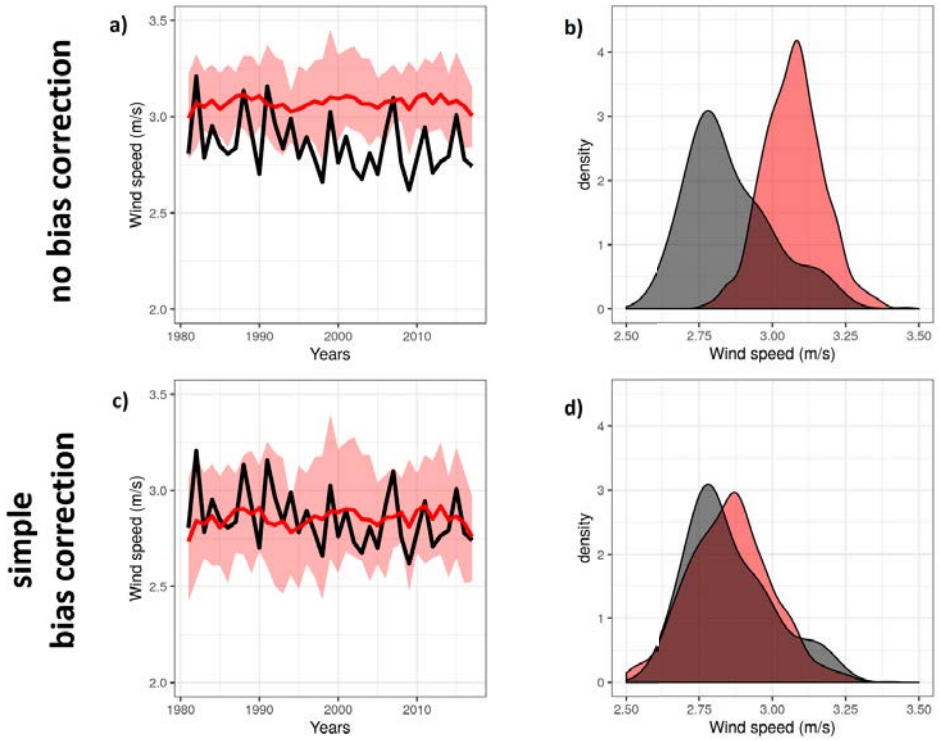


Figure 1.9: Surface wind speed seasonal predictions for December-January-February from the SEAS5 seasonal prediction system at a location in the east Pacific during the 1981-2017 period. (a-b) depict the raw seasonal predictions (i.e., without performing any bias correction procedure), while (c-d) represent the same predictions after applying a simple bias correction. Reds in (a,c) represent the ensemble members of the prediction, in particular, the mean (solid line) and the range of the ensemble (shadow). Black solid lines in (a,c) represent the observed seasonal wind speeds, sourced from the ERA5 reanalysis. Red and black curves in (b,d) represent the probability distribution of the predicted and observed (ERA5) values, respectively. Predictions were initialised on the 1st of November.

—given by the ensemble members—, respectively. Finally, $\overline{o}$ is the climatological value of the observations. It is common practice that the reference observations o_i are used later in the verification of the bias-adjusted seasonal predictions. In such cases, and to ensure the fairness of the verification, it is highly advisable considering cross-validation techniques (Wilks, 2011).

Unlike simple bias correction, quantile mapping adjusts all moments of the forecast distribution so that seasonal forecasts have an equivalent Cumulative Density Function (CDF) to that of the observations. Calibration techniques, on the other hand, encompass a set of methodologies that adjust forecast probabilities to resemble the observed frequencies. However, all these post-processings do not necessarily lead to an increase in the quality of the predictions, especially when cross-validation is applied. Adjusting for one metric can lead to degradation

for another one. The following subsection elaborates on how the quality of seasonal predictions is measured and why an estimate of it is of primary importance whenever a seasonal prediction is delivered.

1.5.2 Measures of the quality of a seasonal prediction

The forecast quality assessment (Jolliffe and Stephenson, 2012) is a fundamental step in seasonal prediction because a forecast product is useless without an estimate of its quality (or skill) that is based on past performances. The hindcasts of a seasonal prediction system are compared against an observational reference dataset so that good seasonal prediction systems exhibit a close correspondence to that observational reference.

Many verification scores have been proposed over the years to assess different aspects of the predicted wind speed distribution adequately (see Jolliffe and Stephenson (2012) for a review). The ensemble mean correlation (based on Pearson's correlation) is one of the most used in forecast quality assessment. It measures the linear association between predictions and observations, with a particular feature sometimes considered an advantage: it ignores systematic biases, which can be pretty large in seasonal forecasts. Other scores are defined specifically for probabilistic forecasts. The most commonly used is the Ranked Probability Score (RPS), and its continuous version the Continuous Ranked Probability Score (CRPS). Both scores measure the quality of the entire predicted probability distribution by measuring its distance to the observations, with the difference that the RPS discretises the probability distribution by transforming it into ordered categories.

All these scores assume the reference observation values to be perfect, known with absolute certainty. However, as discussed in Section 1.4.3, all observations are inherent to containing some degree of error. Not accounting for such error has several implications such as the diverging verification results obtained depending on which observational reference dataset is used (Gómez-Navarro et al., 2012). This leads to an ambiguity for which some solutions have been proposed. One option is to re-build the verification scores to account for observational uncertainty (Ferro, 2017). However, its implementation into seasonal forecast verification has not yet been made. Chapter 6 of this thesis will tackle this issue.

1.5.3 Statistical methods applied to seasonal prediction

The limitations and pitfalls of dynamical modelling in climate predictions has favoured the adoption and development of statistical techniques for empirical seasonal forecasting. A wide range of approaches have been proposed, with varying complexity, but all achieving com-

parably good results. Statistical methods can be applied to predicting seasonal anomalies (Rust et al., 2015), the assessment of climate extremes (Ortizbeviá et al., 2011) or the downscaling of data from coarse to fine grids (Huang et al., 2014). The selection of the most appropriate method highly depends on the climate variable being targeted as well as on the purpose of the study.

Linear regression is the most simple statistical method, and it is easily understandable from the point of view of the simple and multiple linear regression (Wilks, 2011). Much of the relationships between different sets of climate variables can be described linearly so that simple and multiple linear regressions seek to summarise that relationship between two or more climate variables, respectively. Analytically, this consists of the computation of a set of coefficients for the linear equation relating the two sets of variables, being the least-squares method the most commonly used for fitting such coefficients. In particular, least-squares regression looks for the coefficients that minimise the sum (or, equivalently, the average) of the squared errors.

All the traditional methods like regression may sometimes appear less efficient, especially when there is plentiful data to deal with. The current climate data archives are growing rapidly thanks to the increasing data acquisition by satellites and the massive climate modelling projects like the Coupled Model Intercomparison Project (CMIP) (Taylor et al., 2012). As a result, climate scientists have recently begun to explore new computationally efficient techniques to incorporate this new information into climate analyses. Machine Learning (ML) techniques are advancing in the art of climate prediction (Bonavita et al., 2020), helping in the creation of coherent and computationally tractable models. Nevertheless, the development of such methodologies is still in its infancy.

Chapter 2

PhD thesis within the BSC-ES department

This PhD thesis has been conducted within the Earth Sciences Department of the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-ES). The department is directed by Dr. Francisco J. Doblas-Reyes and its mission is to develop and implement global and regional models and data solutions for air quality and climate forecasting as well as their applications. The present thesis deals with the topic of climate prediction, which can be sub-divided into a wide variety of research lines aimed at enhancing the quality of those forecasts (Figure 2.1). All three types of climate predictions are studied at the BSC-ES department, and for each group, different components of the Earth System are targeted. The improvement in the quality of climate prediction is achieved from various perspectives, which include not only the tuning of the prediction systems but also the application of post-processing methods.

In particular, this study will target seasonal predictions for wind speed to eventually provide results in the form of a climate service. Part of the work presented here has been conducted within the framework of two European projects: INDECIS[1] and S2S4E[2]. Furthermore, this research expands on the advances of two PhD thesis carried out at BSC-ES. Torralba (2019) constituted the first PhD thesis of the department aimed at developing a climate service for the wind energy sector. They collect, assess and ultimately recommend the most appropriate techniques for delivering the best seasonal predictions for wind speed. These methodologies regard bias adjustment, forecast verification and dynamical-empirical seasonal prediction. Besides, Lledó (2020) gain insight into the wind speed variability and its sources at sub-seasonal and seasonal timescales. Having said that, the present PhD thesis uses the knowledge provided by these two theses, and continues with the ultimate goal of improving the quality of seasonal predictions. The use of high-quality observations will be the key element introduced to achieve successful results.

[1] www.indecis.eu

[2] www.s2s4e.eu

Earth System Component	Perspective	Subseasonal	Seasonal	Decadal	Reference
Troposphere	Climate modeling		●		Prodhomme et al. (2016); Tintó-Prims et al. (2019)
	Data assimilation and initialisation			●	García-Serrano et al. (2015); Solaraju-Murali et al. (2019)
	Forecast verification	●	●	●	Manrique-Suñén et al. (2019); Manubens et al. (2018); Ramon et al. (in prep.); Delgado-Torres et al. (in prep.)
	Teleconnections	●	●		Lledó and Doblas-Reyes (2020); Mezzina et al. (2020)
	Bias-adjustment		●	●	Torralba et al. (2017); Ramon et al. (2021)
	Observations		●		Ramon et al. (2019)
	Extremes		●		Donat et al. (2019)
	Artificial Intelligence		●		Gómez-González et al. (in prep.)
	Weather regimes	●	●		Cortesi et al. (2021); Torralba et al. (2021)
Stratosphere	Climate modeling		●		Palmeiro et al. (2020)
Ocean	Climate modeling		●		Falls et al. (submitted)
Sea Ice	Climate modeling		●		Cruz-García et al. (2019)
	Data assimilation and initialisation		●		Guemas et al. (2016)

Figure 2.1: Summary of the research lines at the ES department devoted to improving climate predictions. For each research line, a set of representative publications is listed. This PhD dissertation deals with the research lines highlighted in green.

Chapter 3

Objectives and structure of this work

The overall goal of this thesis is to improve the quality of seasonal predictions for the wind energy sector from the perspective of the climate observations. Even though observations are routinely employed in seasonal prediction, little attention has been traditionally paid to their opportunities to understand the coming climatic conditions better. The aim is addressed through the following objectives:

Objectives

1. **Adjust** seasonal predictions to the **local scale** by combining information from point observations.
2. Account for the **uncertainty** in the **observational reference** in seasonal **forecast quality assessment**.
3. Evaluate the **quality** and quantify the error in **observational sources**, to later use them in the verification and enhancement of seasonal predictions.
4. **Bring the advances** to the wind energy **users**.

Resulting from this work, an observational dataset as well as novel methods and approaches will be delivered. Since the wind energy industry is one of the biggest users of the var-

ied types of wind data, our results ultimately provide knowledge in the form of a climate service. While some outputs are already available, others will shape future climate services. With our conclusions and recommendations, we intend to guide individuals and organisations on their choices among alternative courses of action to increase benefits and reduce expenses. Therefore, we are indirectly increasing the value of seasonal forecasts.

This work is presented in the form of a compendium of three publications. These articles appear numbered in three different high-impact journals, namely, *Quarterly Journal of the Royal Meteorological Society*, *Earth System Science Data* and *Environmental Research Letters*.

Chapter 3 includes the scientific paper in which we present the Tall Tower Dataset, a collection of non-standard climate observations from 222 tall towers distributed worldwide. This chapter also describes the quality-control process carried out over the dataset, as well as the result of such exercise. Chapter three deals with the third objective.

Chapter 4 also focuses on the third objective. In the article included there, the error in reanalysis data is quantified by means of a comparison and a later verification of five widely used global reanalyses. We eventually select the product that best approximates near-surface wind speeds to those observed in-situ at 77 instrumented tall towers.

Chapter 5 tackles the first objective. Here we describe the methodology to produce hybrid seasonal predictions for wind speed. By combining observations from the Tall Tower Dataset and the reanalyses with dynamical predictions of the synoptic state of the atmosphere, we aim to provide accurate seasonal predictions at a fine scale.

Chapter 6 evaluates the impact of observational error in the context of seasonal forecast quality assessment, i.e. the second objective of this thesis. We discuss various approaches to account for such error in the verification scores and study the implications in a model ranking exercise. Will the best seasonal prediction system stand out as so regardless of the observational reference used to verify it?

In each of the above-mentioned chapters, we analyse in what way a climate service is provided. We describe the potential usage of each output in wind-related studies, and within the wind power industry. By bringing our results to the wind energy users (fourth objective), the non-experienced audience can also benefit from the advances of the present thesis.

Chapter 7 presents the conclusions of this work. The main scientific contributions and the ideas for the continuation of this study are outlined in this chapter.

Chapter 4

Winds blowing at 100 metres

Objective

Create a quality-controlled dataset containing wind data measured at tall meteorological masts. The height of these structures ranges between 20 to more than 400 metres so that the vertical wind profile in the lowermost levels of the troposphere can be characterised. Since modern wind turbines reach heights around 100 metres, a data collection with such qualities would be of high value for the wind energy sector.

Methodology

- Collect climate data from sparse existing datasets.
- Create and apply a quality control suite to ensure the high quality of wind data.

Results

- The Tall Tower Dataset contains data from 222 tall tower locations.
- The timespan of the series is diverse, ranging from 2 to 34 years.

- The 95.2% of the original data passed successfully the 18 quality control tests.
- A simple exercise demonstrates that our quality control suite is effective in detecting bad data.
- The Tall Tower Dataset is publicly accessible at talltowers.bsc.es

Conclusions

- The wind industry and the research academy are the target users of the Tall Tower Dataset.
- There is still a general reluctance to share climate data. However, recent initiatives like the Tall Tower Dataset aim at boosting the open data.

Publication

- Ramon, J., Lledó, L., Pérez-Zañón, N., Soret, A., and Doblas-Reyes, F. J. (2020). The Tall Tower Dataset. A unique initiative to boost wind energy research. *Earth System Science Data*, 12:429–439, doi: 10.5194/essd-12-429-2020

Earth Syst. Sci. Data, 12, 429–439, 2020
https://doi.org/10.5194/essd-12-429-2020

Open Access Earth System Science Data

The Tall Tower Dataset: a unique initiative to boost wind energy research

Jaume Ramon[1], Llorenç Lledó[1], Núria Pérez-Zanón[1], Albert Soret[1], and Francisco J. Doblas-Reyes[1,2]

[1]Barcelona Supercomputing Center (BSC), c/ Jordi Girona, 29, Barcelona 08034, Spain
[2]ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain

Correspondence: Jaume Ramon (jaume.ramon@bsc.es)

Received: 22 July 2019 – Discussion started: 10 September 2019
Revised: 20 January 2020 – Accepted: 21 January 2020 – Published: 18 February 2020

Abstract. A dataset containing quality-controlled wind observations from 222 tall towers has been created. Wind speed and wind direction measurements covering the 1984–2017 period have been collected from existing tall towers around the world in an effort to boost the utilization of these non-standard atmospheric datasets, especially within the wind energy and research fields. Observations taken at several heights greater than 10 m above ground level have been retrieved from various sparse datasets and compiled in a unique collection with a common format, access, documentation and quality control. For the last, a total of 18 quality control checks have been considered to ensure the high quality of the wind records. Non-quality-controlled temperature, relative humidity and barometric pressure data from the towers have also been obtained and included in the dataset. The Tall Tower Dataset (Ramon and Lledó, 2019a) is published in the repository EUDAT and made available at https://doi.org/10.23728/b2share.136ecdeee31a45a7906a773095656ddb.

1 Introduction

Renewable energies have experienced the fastest growth among all electricity sources in the last few years (OECD/IEA, 2018, 2019). Together with solar photovoltaic, the wind power sector is leading this development, and the number of new wind farms and the installed capacity is currently facing an important increase worldwide (WindEurope, 2018; AWEA, 2019).

With higher shares of electricity generation depending on wind speed conditions, it is crucial to advance understanding of wind speed conditions at heights between 50 and 150 m above ground – where current wind turbines are installed – and at multiple timescales ranging from turbulence to mesoscale circulations, seasonal to decadal oscillations and climate change impacts. To characterize these features, high-quality meteorological observations are needed.

Vast numbers of surface wind measurements taken at the standard height of 10 m above surface level do already exist, and efforts have been made to compile the existing surface wind observations (Lott, 2004; Dunn et al., 2012; Klein Tank et al., 2002; Lucio-Eceiza et al., 2018a, b). However, meteorological data at turbine hub heights are much scarcer than surface observations. To take those measurements, a tall tower or met mast needs to be installed and instrumented. The basic structure of these masts consists of a high vertical tower reaching heights of 100 to 200 m above ground with several platforms distributed along the vertical structure. It allows the placement of several wind sensors (i.e. anemometers and wind vanes) at different heights so that the vertical wind shear can be profiled. In addition, it is also typical to install several horizontal booms at each measuring height oriented to different directions. Thus, more than one sensor per measurement level can be installed to correct or replace data from one of these redundant sensors in case it is affected by a technical failure or by the wind shadow produced by the mast itself. The physical structure of a tall tower, as well as a typical instrumentation layout, is illustrated in Fig. 1.

Recently, the usage of remote sensing devices to measure atmospheric profiles has increased as an alternative to the tall tower in situ measurements. Atmospheric lidars, for example, are becoming more popular due to their easy installa-

Published by Copernicus Publications.

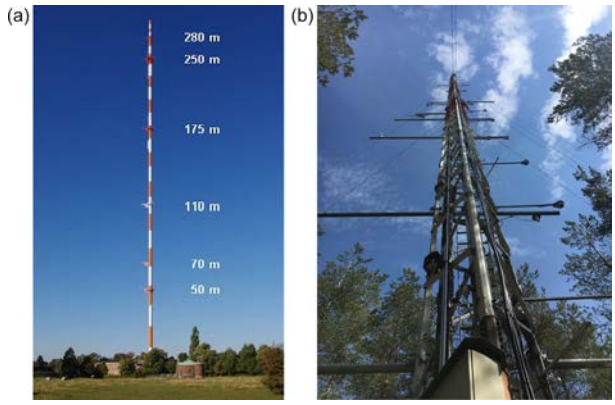


Figure 1. **(a)** Measuring levels at University of Hamburg meteorological mast (Germany). Source: https://icdc.cen.uni-hamburg.de (last access: February 2020).**(b)** Arrangement of the instrumentation in booms at Hyytiälä forest met tower (Finland). Courtesy of Jesús Yus-Díez.

tion and maintenance when compared to tall towers. However, the lack of historical lidar data limits their utilization in long-term assessment studies. One more example of that new trend is the lidar-based satellite Aeolus, which was launched by the European Space Agency in 2018, and has just started acquiring profiles of Earth's wind on a global scale (https://www.esa.int/Applications/Observing_the_Earth/Aeolus, last access: February 2020).

Hub-height observations are widely used in different initiatives to (a) evaluate the wind resource characteristics and derive wind power generation estimates (Brower et al., 2013); (b) study local wind shear, turbulence and the dynamics of the planetary boundary layer, PBL (Li et al., 2010); (c) enhance or verify reanalysis products (Ramon et al., 2019; Decker et al., 2012); (d) correct meteorological forecasts (Baker et al., 2003) and climate predictions (Torralba et al., 2017); or (e) calibrate and verify wind atlas products (e.g. Troen and Petersen, 1989; Fernando et al., 2018; Tammelin et al., 2013).

Most of the existing met masts are owned by private companies mainly from the wind energy industry. Wind energy companies need to take those measurements prior to the construction of a new wind farm to characterize the wind speeds in the area and eventually ensure the return of the initial investment. In addition, some local effects such as topographic channelling, sea breezes, turbulence or vertical wind shear must be inferred because they can have a substantial impact on the electricity production (Hansen et al., 2012). Since the maintenance costs of these large and complex structures are rather expensive, the energy industry typically takes measurements for a relatively short period (1 or 2 years usually). Then the towers are decommissioned, so the lack of long records of tall tower data reduces the possibilities to study, for example, wind variability at seasonal to decadal timescales. In addition, private companies are usually reluctant to share the tall tower data with third parties, obstructing their further usage even more.

Fortunately, many of the initiatives from (a) to (e) also take tall tower measurements for their research and then the data are usually made freely accessible for non-commercial purposes. Derived from these diverse efforts devoted to boosting the utilization of tall tower records, there exist various sparse datasets containing measurements from instrumented towers. Regrettably, they are often difficult to find or access, and the lack of coordination in terms of formats, metadata, data access and quality control (QC) hinder their usability outside the owner institution.

The INDECIS (http://www.indecis.eu/, last access: February 2020) project is making attempts to collect existing non-standard meteorological observations, among other efforts. In this paper, a dataset is presented, and the QC of the wind data is further detailed. The reader is referred to Ramon and Lledó (2019b) to find complete information on the identification and collection of towers, data formatting and documentation. Section 2 of this article describes the main features of the dataset, as well as the data characteristics. The QC software suite is defined in Sect. 3. Then, a wrap-up of the results after running the QC checks is presented in Sect. 4. The benchmark experiment carried out to test the robustness of the QC software is shown in Sect. 5. Finally, conclusions are presented in Sect. 6.

2 Tall Tower Dataset description

The Tall Tower Dataset (Ramon and Lledó, 2019a) is a unique collection of data from 222 tall towers resulting from an exhaustive process of identification of existing masts and their later data retrieval. Figure 2 presents the global distribution of the sites, which is highly heterogeneous. Most of the masts are located in Asia (51 %), mainly clustered in Iran, resulting from a national campaign aimed to boost renewable energies at a country level. Then, tall towers appear more spatially distributed over North America (23 %) and Europe (16 %), mirroring the important deployment of wind power that is taking place in those regions. Africa (8 %), Oceania (1 %) and Antarctica (1 %) follow. Unfortunately, it has been hard to retrieve data from South America, so no records from this area can be found in the Tall Tower Dataset.

The height above the surface where the top sensor is located for each tower is also depicted in Fig. 2. On the one hand, masts placed in historical observatories (i.e. often having more than 20 years of data) tend to be short, with heights ranging between 18 and 50 m above the ground and usually consisting of one measuring level at the top of the pole. Two examples are the American masts in Utqiaġvik (formerly known as Barrow) and Mauna Loa. On the other hand, modern towers often reach 100 to 200 m of altitude. Indeed, most of the masts in northern Europe have been installed during

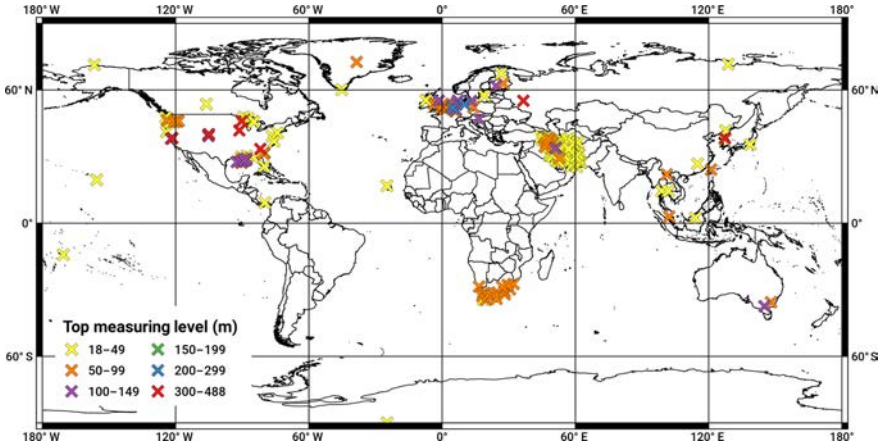


Figure 2. Global distribution of the 222 tall tower locations within the Tall Tower Dataset. Colours indicate the top measuring level for each tower. Further information can be found in Sect. S1 of the Supplement.

the last 15–20 years and are generally taller than 80 m, usually reaching 150 to 200 m. However, the tallest structures are located in the USA, reaching the exceptional height of 500 m, allowing the placement of sensors at that height. The top anemometer at Walnut Grove tall tower in California is at 488 m above ground level. The number of measuring levels in these masts is almost always higher than three, and up to eight in the case of the FINO met masts.

A list of the towers included in the Tall Tower Dataset, as well as their main characteristics such as the owner institution, country, geographic coordinates or specific recording periods, can be found in Sect. S1 of the Supplement. The record lengths and other structural features such as height or instrumentation are quite diverse as they depend on the purpose they were designed for. Most of the towers are typically installed to provide in situ observations for experimental field campaigns within the research or industry fields. In this case, the tall towers are commonly referred to as meteorological masts or met masts, and they represent up to the 87 % of all the tall towers in the dataset. However, other sensors are installed over marine platforms (11 %) or at the top of lighthouses (1 %) to monitor the coastal weather conditions. Finally, 1 % of the towers are instrumented communication transmitters that take meteorological measurements at several platforms along with the antenna. Concerning the location, almost 80 % of these tall towers are found inland while the other 20 % are placed offshore.

Information indicating the representative features mentioned above is included in the dataset within the corresponding site metadata, which have been standardized for all the sites. This material was sometimes confusing, sparse or even missing in the datasets distributed by the owner data centres, especially when it comes to the conventions in which the initial data were prepared. For example, if the time zone in which the time stamps were delivered was not specified, it could be challenging to discern whether they are provided in local time or Coordinated Universal Time (UTC). Another example concerns the data units, which were not explicitly stated in a few cases either. In both of these confusing situations, the data provider was contacted to confirm the original convention. Further information on the diverse standards in which the data were provided as well as the final conventions employed in the Tall Tower Dataset can be found in Ramon and Lledó (2019b).

The time span of the 222 time series is depicted in Fig. 3a. First, we split the series according to their time resolution, which varies from every 10 min to once per hour. Most of the series, i.e. a total of 172, provide 10 min averaged data, meeting the WMO standard (WMO, 2007) for estimating mean wind speeds. The other 50 masts report 15 min, 20 min, 30 min or hourly data. Information on how these averages have been taken is hardly ever available. The fact is that resulting aggregated values vary depending on whether averages are taken over the horizontal wind components or speed and direction modules independently. WMO (2007) does not prefer one option over the other, as it may depend on the application or available instrumentation. Even though the effects of this choice are rather small, especially for higher wind speeds, it represents an additional source of uncertainty for the values themselves.

The total coverage of the Tall Tower Dataset ranges from 1984 to 2017. While the 90 % of the series span less than 20 years, 3 % cover 30 or more years. The precise beginning and end of recording periods can be found in Sect. S1 of the Supplement for each tall tower. Nevertheless, several of these masts have been recently installed, and measurements are currently operational. Missing data periods – 12.1 % of the dataset – appear sometimes embedded within the series.

Concerning the data retrieval process, the initial efforts focused on collecting the largest number of wind observations possible. Those records have been complemented with temperature, relative humidity and surface pressure data also

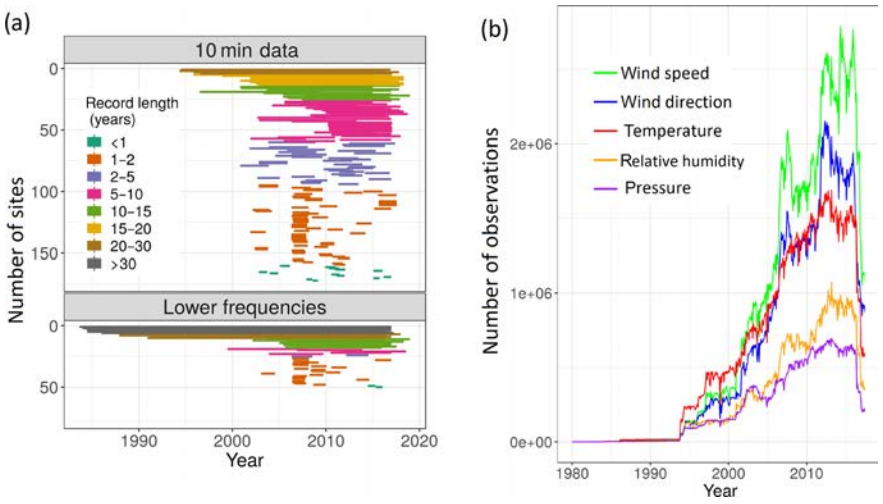


Figure 3. **(a)** Time coverage of the 222 tall towers depending on whether they report 10 min data (top) or lower resolutions up to 1 h data (bottom) and **(b)** time evolution of the total number of observations within the Tall Tower Dataset for wind speed (green), wind direction (blue), temperature (red), relative humidity (orange) and pressure (purple).

measured at the different platforms along the tower. The time evolution of the number of these five variables is plotted in Fig. 3b. Most of the data fall within the 21st century, with a significant increase at the beginning of the millennium. Up to 2.7 million wind speed records have been retrieved for one single month, i.e. December 2015, which constitutes the month with the maximum amount of wind speed data. In the case of wind direction, the month with the highest number of records is October 2012 (2.1 million measurements). A decrease in the number of observations has been noticed from 2017 onwards. Generally, some of the data providers prefer to keep the most recent data and release them once measurements are preliminarily checked for gross errors. Temperature, relative humidity and pressure are not always available. We note that the fewest records correspond to barometric pressure, which is usually measured only at surface level (i.e. 2 m above ground level).

3 The Quality Control Software Suite for Tall Towers (QCSS4TT)

To ensure the high quality of tall tower wind data and guarantee the accuracy of any result derived from these records, a QC procedure needs to be carried out. The scientific literature has devoted efforts to QC wind data taken at surface stations (e.g. Dunn et al., 2012; Lucio-Eceiza et al., 2018a, b). However, no QC software has been specifically designed to tackle the same problem with tall tower observations, whose features vary considerably when compared to surface wind data (e.g. measurements are taken at higher altitudes, the spatial density of stations is considerably lower). Unique measuring techniques, such as the parallel measurements at different platforms along the mast or sensor redundancy at a given height, can also be taken into account to complement and enhance the typical QC.

After a review process of the existing QC routines, a set of 18 sequential QC tests (two preliminary and 16 main tests) has been selected and designed to be performed over wind measurements. The Quality Control Software Suite for Tall Towers (QCSS4TT) designed here is applied to all the wind speed and wind direction data within the Tall Tower Dataset, regardless of whether they were previously quality controlled or not by the providing institution. A general description of the QCSS4TT is presented below in this section. The software is fully described in Sect. S2 in the Supplement.

The QC tests within the QCSS4TT are all intra-station checks, as they do not compare series from nearby tall towers. QC routines ingest entire time series of winds at a specific heights, whose time frequencies vary between 10 min and 1 h. The recommended sequence for the application of the QC tests is presented in Fig. 4. Checks are grouped in five categories depending on the purpose they were designed for. The two preliminary checks are designed to detect gross manipulation errors. Then, the 16 main QC tests ensure the limits, spatio-temporal and internal consistency of the wind speed and wind direction time series. We note that the routines can be run independently, with the exception of the quartile occurrences and isolated pass tests which feed on the output of other tests within the QCSS4TT. After deciding the appropriate order, the tests have been applied over the Tall Tower Dataset according to the flux diagram in Fig. 4.

The QCSS4TT starts with two preliminary tests. Firstly, the time stamp check is carried out during the data formatting process and ensures that all the time stamps are included in the dataset and equally sampled according to the tower reporting frequency. Those time stamps that are either du-

Table 1. Main QC tests summary. The abbreviations ws, wd and tmp stand for wind speed (in metres per second), wind direction (in meteorological degrees) and temperature (in degrees Celsius), respectively. The reader is referred to Sect. S2 in the Supplement for detailed information on each of the tests.

QC	Fail	Suspect	Remarks
Plausible values	ws $\notin [0, 113.2)$ wd $\notin [0, 360]$	ws $\in (75, 113.2]$	
Difference between extreme values of the distribution		$\max(\text{ws}) - \max_{2\text{nd}}(\text{ws}) > \max_{2\text{nd}}(\text{ws})$	Runs iteratively until the condition is not satisfied
Persistence test		if $A = \{x_t, \ldots, x_{t+60}\}$ is a set of 60 consecutive values, $\max(A) - \min(A) < 0.7$ for ws and $\max(A) - \min(A) < 5$ for wd	Skips calms (ws $\leq 0.5\,\text{m s}^{-1}$)
Flat line	Six or more consecutive ws values all equal, or 40 or more consecutive wd values all equal.	Three or more consecutive ws values all equal, or 20 or more consecutive wd values all equal	
Icing	$\max(\text{tmp}) < 0$ and $\max(\text{ws}) = 0$ for a 4 d period or longer.		
Abnormal variations		$\sigma_i \notin [\overline{\sigma} \pm 4 \cdot \text{sd}(\sigma)]$, where σ is the distribution generated by the standard deviations of all the 30 d periods within a time series and σ_i is the standard deviation of the ith 30 d period	Only for ws. Disabled when more than 50 % of data within a 30 d period are missing
Systematic errors		$m_i \notin [\overline{m} \pm 4 \cdot \text{sd}(m)]$, where m is the distribution generated by the means of all the 30 d periods within a time series and m_i is the mean of the ith 30 d period	Only for ws. Disabled when more than 50 % of data within a 30 d period are missing
Quartile occurrences	Given a period of time, all the observations fall above or below the first, second or third quartiles of the ws distribution (see Table S2 of the Supplement)	Given a period of time, all the observations fall above or below the first, second or third quartiles of the ws distribution (see Table S2 of the Supplement)	
Rate of change	$\text{ws}_{t+1} - \text{ws}_t \geq 3 \cdot \text{IQR}$	$2 \cdot \text{IQR} \leq \text{ws}_{t+1} - \text{ws}_t < 3 \cdot \text{IQR}$	IQR is the interquantile range of the ws distribution
Step test	$\text{ws}_{t+1} - \text{ws}_t \geq 20$		
Repeated sequences	Repetition of a sequence of 20 or 30 consecutive ws values, or repetition of a sequence of 30 consecutive wd values.		The values of the sequences need not be all equal. The maximum allowable length of the ws sequence depends on the decimal places of the data
Tower shadow		ws values falling in the wake of the vertical structure	Only works if the tower contains redundant anemometers at the same level
Vertical ratios	Ratio between parallel ws observations at different levels exceeds 30 units	Ratio between parallel ws observations at different levels exceeds 15 units	Skips ws lower than $1\,\text{m s}^{-1}$
Isolated pass	Unflagged ws or wd values are surrounded by sequences of erroneous or missing data (see Table S3 of the Supplement)	Unflagged ws or wd values are surrounded by sequences of suspect data (see Table S4 of the Supplement)	This QC test needs to be run after the other routines
Occurrences of 0's and 360's	The occurrence of 0's represents more than 30 % of ws values, or the occurrence of 0's and 360's represents more than 30 % of wd values		Does not flag individual records but the entire time series
Internal consistency	ws $= 0$ and wd $\neq$ NA		

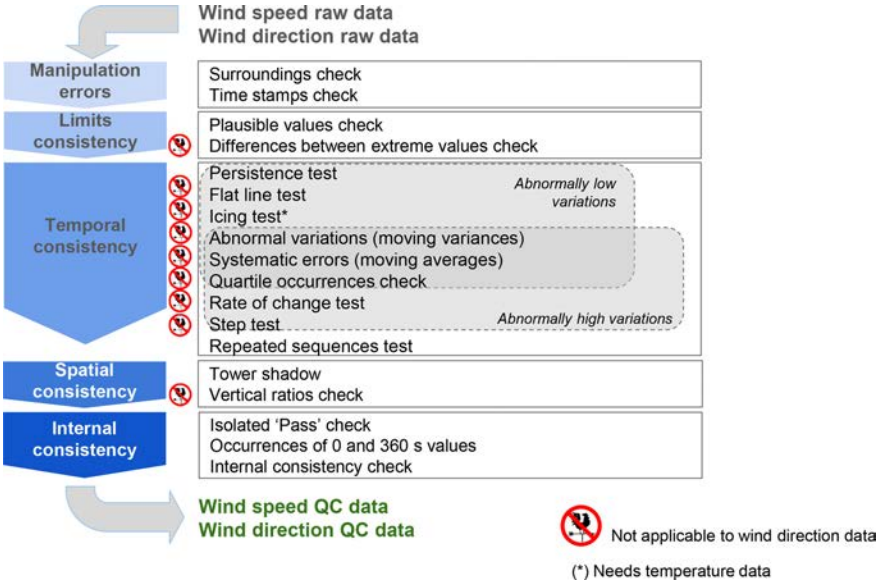


Figure 4. Summary of the workflow of the QCSS4TT routines applied over wind data within the Tall Tower Dataset.

Table 2. Flag level definitions.

Flag value	The observation...
0	has not been evaluated by three or more QC tests (partly QCed)
1	has passed all QC tests successfully
2	is deemed suspect
4	has failed at least one of the tests
5	is a calm wind
9	is missing

plicated or wrongly spaced have been discarded, and those missing have been included, setting the corresponding record to "not available" (NA). Secondly, the surroundings check is performed by detecting nearby elements that could potentially perturb the wind flow and then produce unreal records. To do so, detailed descriptions of the encircling area of the mast as well as its possible changes over time are required.

Then, the 16 main QC tests follow. A summary can be found in Table 1 and complete information can be found in Sect. S2 in the Supplement. Most of them are standard checks typically performed over wind and other Essential Climate Variables such as temperature or precipitation. However, we propose two new QC tests (the so-called tower shadow and vertical ratios checks, respectively) here to guarantee the spatial consistency of the data by considering the special characteristics of the tall tower measurements since classic interstation comparisons appear challenging due to poor spatial density of sites.

After running the QCSS4TT, a natural number (hereafter referred to as QC flag or flag; see Table 2) is attached to each observation according to its nature and/or level of confidence. To decide which flag should be assigned to each observation, different threshold values have been set for each of the QC routines. The threshold selection has been based on the World Meteorological Organization (WMO) standards (WMO, 2007; Aguilar et al., 2003), QC software manuals (IOSS, 2017) or state-of-the-art bibliography (e.g. Jiménez et al., 2010). Many of these standards, if not all, have been developed specifically for surface winds (i.e. 10 m winds), whose features vary importantly when compared with winds observed at higher altitudes, such as those from the present work. After a preliminary test of the thresholds over the wind data within the Tall Tower Dataset, it was noted that some tests considerably overestimated the amount of erroneous data (also known as Type I errors; see Hubbard et al., 2004). The WMO allows adjustment of some of the fixed-value limits proposed in the WMO (2007) to reflect singular climate conditions more accurately. As the QCSS4TT aims to clean data from towers located all over the world regardless of the prevailing climate conditions in the area, thresholds need to be adjusted manually to not deem wrong the general and particular climate features observed in a wide variety of world climates. It is also vital to take into account that this sensitive experiment should reduce the number of Type I errors without increasing the number of invalid data that have been accepted by the tests (also referred to as Type II errors).

Based on these thresholds and the nature of the individual wind records, six different categories have been defined (Table 2), and each datum is flagged appropriately. The quality of a record is inferred automatically by checking if it passes all the tests successfully (flagged as "1"), passes the tests but might need further checks such as a visual inspection (here-

after referred to as *suspect* and marked as "2") or fails at least one of the tests (flagged as "4"). When an observation is not considered suspect or wrong by any of the QC tests, additional levels may indicate that the observation was not evaluated by three or more tests (indicated as "0") or corresponds to a calm period ("5"). Finally, missing values are flagged uniformly (categorized as "9").

Wind records flagged as "4" are deemed to be erroneous data and thus unreliable. They have been removed by changing the original record to NA. Suspect data as well as those observations that have not been evaluated by all the QC tests remain unaltered because they might be potentially correct and usable for some applications. But in case the user prefers to impose their own level of restriction, we also include the raw data jointly with the flag values resulting from the quality controlling. Therefore, the data user is able to filter the raw data based on the flag values. Still in those cases, we strongly discourage the usage of data marked as erroneous ("4").

4 Results of the application of the QCSS4TT

The QCSS4TT has been applied sequentially over the Tall Tower Dataset according to the flux diagram in Fig. 4. We present here the global results obtained from the quality controlling of the Tall Tower Dataset, as well as a summary of the performance of the main tests.

As stated in Sect. 3, the surroundings check needs detailed original metadata of the tower location. Unfortunately, this valuable information is not always available so the surroundings check cannot be carried out over all the tower sites. The unique case when this QC test confirms that a series of wind speeds were disturbed by the surrounding forest occurs at Wallaby Creek met mast. After running the main QC routines, long sequences of wind speeds measured at the lowermost level of this met mast – placed at 10 m above surface – have been flagged as wrong. A close look at the site metadata reveals that the canopy well exceeds the 10 m height during the whole recording period, considerably reducing the observed wind speeds. Hence, all the individual observations of the Wallaby Creek 10 m wind series have to be used with caution, even those that have not been considered problematic by other tests.

Then, the main QC routines have examined each of the 240 371 908 wind speed and wind direction values individually in the Tall Tower Dataset and flagged them accordingly. After this process, 228 780 679 values (95.2 % of the total data) passed successfully all the checks and can be considered reliable. Conversely, 6 827 880 observations (2.8 %) have been considered erroneous by at least one of the QC tests. They have been replaced by NA, increasing the total number of missing data from 12.1 % to 14.6 %. A total of 1.8 % of the dataset is flagged as suspect. Some of the QC tests, particularly those that compute period-aggregated statistics such as moving averages or variances, require a minimum amount of data. Due to this constraint, 0.2 % of the data have not been evaluated by three or more QC tests to avoid the computation of such statistics with reduced sample sizes. Records identified as calm (i.e. wind speeds under 0.5 m s^{-1}) have also been skipped on purpose by a small group of tests, i.e. those that compute quotients between pairs of simultaneous observations. However, calm wind records can be trusted as they successfully passed all the other QC checks. The percentage of calm wind values is highly dependent on the geographical location of the tall tower. Met masts located in Southeast Asia contain the largest percentage of calm wind records, reaching up to 24 % of the total data.

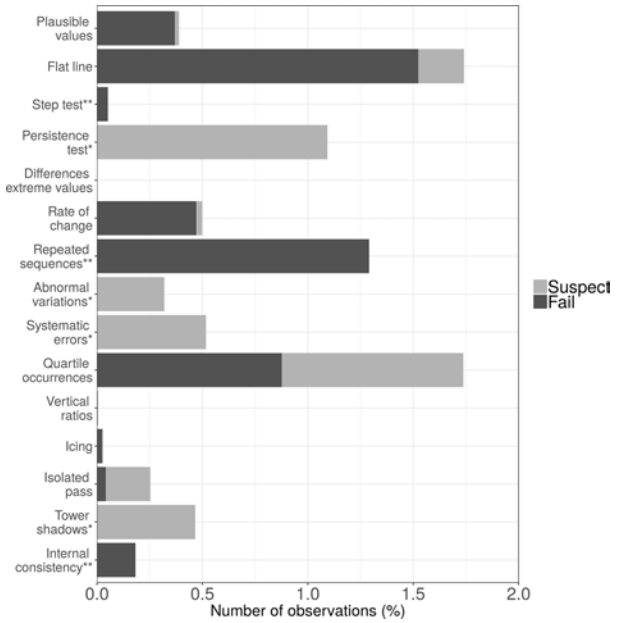


Figure 5. Percentage of data flagged as "fail" (dark) and "suspect" (light) by 15 of the main QC tests within the QCSS4TT. Asterisk (*) indicates that the QC test only flags data as suspect. Double asterisk (**) denotes that the QC test only flags data as erroneous.

The amount of data flagged by each test is considerably different, as can be noticed in Fig. 5, which depicts the percentage of data flagged as erroneous and suspect by the main QC tests. We note that both the flat line and quartile occurrences checks have flagged the largest amount of data (1.74 %). The former detected the most substantial amount of erroneous data (1.52 %), followed by the repeated sequences and quartile occurrences tests (1.29 % and 0.88 %, respectively). The vertical ratios check has detected very few erroneous or suspect records, and the difference of extreme values test has flagged no data. The occurrences of 0 and 360 values tests is not included in Fig. 5 since this test does not flag individual records but the entire time series according to their quality. Results for this QC show that no wrong or suspicious time series have been detected after the inspection of

Table 3. Percentage of the detected errors (%) depending on the proportion of data that were set to missing. The percentages of missing data are approximately 20 %, 10 %, 5 % and 0 % (NA-free). The table differentiates between land and offshore locations.

	NA-free	5 % missing	10 % missing	20 % missing
Land sites	40.3	40.2	39.8	39.1
Offshore sites	40.0	39.8	39.4	38.8
Total	40.2	40.0	39.6	39.0

the frequency of appearance of null wind speeds and 0 and 360 wind direction values.

5 How reliable is the QCSS4TT?

The performance of the QCSS4TT needs to be assessed. Here, a benchmark experiment has been specifically designed to test the ability of the QCSS4TT in detecting wrong values. In the following, the preparation of the experiment and its results are described.

The setup of the experiment consists of generating a set of presumably QC-free time series where a set of errors will be purposely introduced later on. The time series have been extracted at 50 randomly selected points from the ERA5 reanalysis (Copernicus Climate Change Service , C3S) global grid at hourly frequency, thus meeting the requirements of the QCSS4TT concerning the time resolution. These time series span the 10-year period from 2007 to the end of 2016, which constitutes the time range with the largest number of records within the Tall Tower Dataset (Fig. 3b). To better emulate the features of the tall tower data, we retrieve two parallel series at each of the 50 points. These wind speeds are those provided at 10 and 100 m, respectively.

The set of 50 series is replicated fourfold. Three of these four groups of series are firstly modified by introducing missing data at random, either by erasing data individually or removing sequences of records. The percentages of missing data in these series are approximately 5 %, 10 % and 20 %. The introduction of missing records emulates the frequently observed sporadic sensor failures and no data periods within the wind speed series. Finally, the remaining group of series is left with no datum set to missing.

The error "seeding" process is carried out following the methodology in Hubbard et al. (2004), where the performance of a set of basic QC tests for temperature and precipitation data is assessed. In this publication, a subset of 2 % of the total data is selected to be modified by introducing an error of magnitude:

$$E_{ix} = \sigma_x r_i, \quad (1)$$

where σ_x is the standard deviation of the time series x and r_i is a randomly selected number generated using a uniform distribution ranging from -3.5 to $+3.5$ specifically for the ith observation. Once the errors are inserted, the QCSS4TT is executed.

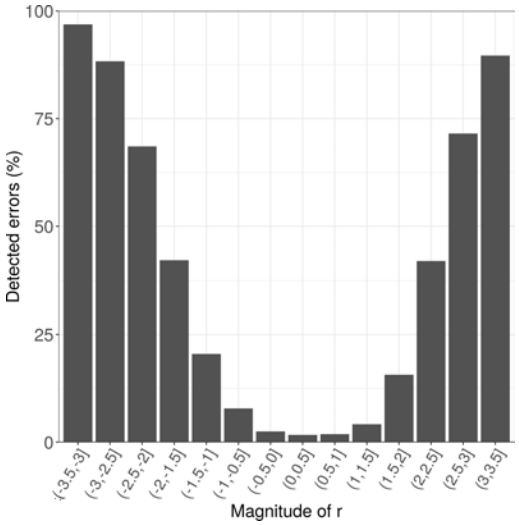


Figure 6. Percentage of detection of seeded errors as a function of the magnitude of the random number r.

Table 3 summarizes the differences in the error detection depending on whether missing data are introduced or not into the wind series for both continental and oceanic locations. It is worth noting that the QCSS4TT shows a slight sensitivity to missing data, reducing the percentage of detected errors when the percentage of missing records increases. This decrease might be attributed to the fact that some QC tests are deactivated when a period with very sparse data is encountered. No important differences are noted between onshore and offshore sites, as the percentages of identified errors are quite similar. Results that follow in this section are presented for the set of series containing approximately 10 % of missing data, which is close to the average of absent records within the Tall Tower Dataset (12.5 %).

The QCSS4TT has detected on average nearly 40 % of all the seeded errors (see Table 3). Indeed, this result is at the average of the percentage of detection observed for precipitation data in Hubbard et al. (2004), which was 30 %–40 % for complex terrain sites and 40 %–50 % for the other locations.

At this stage of the experiment, it is important to study the role of the random number r, and particularly its magnitude, which subsequently influences the size of the error E. Values of r close to zero will introduce smaller errors, which will be less likely to be caught by any of the QC checks. Figure 6 presents the percentage of detection as a function of the r values, which have been grouped in intervals of 0.5 units. We note that the QC tests detect most of the biggest errors. However, the percentage of detection decreases as the magnitude of r does, as we expected. Thus, the smallest errors are

usually skipped by all the QC tests. Indeed, this result mirrors the conservative philosophy employed in the threshold selection of the checks.

Finally, it has been observed that Type I errors have been made in 8 % of the total data, corresponding mainly to suspect flagging.

6 Code and data availability

Records from 181 out of the 222 tall towers within the Tall Tower Dataset (Ramon and Lledó, 2019a) are publicly accessible through the following EUDAT repository: https://doi.org/10.23728/b2share.136ecdeee31a45a7906a773095656ddb. Data from the other 41 tall towers are not directly provided for download within the Tall Tower Dataset but can be downloaded from specific tower websites (mostly after registration). Links to all the 222 tall tower original sources are provided in a data catalogue website at http://talltowers.bsc.es (last access: February 2020), which also facilitates browsing the Tall Tower Dataset and visualizing the main climatic features of all the data.

The QCSS4TT code as well as an explanatory vignette on how to use it is also available via GitLab at: https://earth.bsc.es/gitlab/jramon/INDECIS-QCSS4TT (Ramon and Lledó, 2019c).

7 Conclusions

Hub-height wind data are vital to assess the local wind flow features at heights ranging from 20 to 120 m, where wind turbines are located. Nonetheless, the wind industry is not the only user of these observations; the research academy is also interested in retrieving hub-height winds for their studies such as PBL experiments or the verification of climate products. Unfortunately, these non-standard climate data appear sparsely, and the lack of standardized formats, quality and metadata jeopardizes their further usage. This is the first time efforts were devoted to gathering the most substantial possible amount of existing data measured at tall towers around the world and perform an exhaustive QC assessment to eventually make them publicly available for non-commercial purposes in a standard format and access point.

Wind speed, wind direction, temperature, pressure and relative humidity observations measured at different heights on 222 tall towers – owned mainly by public institutions such as universities, meteorological weather services or research centres – have been retrieved from sparse archives, compiled in a unique collection, quality controlled – in the case of wind speed and wind direction data – and released under the name of the Tall Tower Dataset. Data from of 181 of these sites are stored in the EUDAT data repository and can be publicly accessed. Records from the other 41 towers are not available there since the authors of the Tall Tower Dataset do not own the observations and the data providers do not grant rights to share with third parties. Although some initiatives such as the Climate Data Store (https://cds.climate.copernicus.eu/, last access: February 2020) are starting to boost the free utilization of climate observations, there is still some reluctance, mainly in Europe, to contribute to open initiatives that include data in public external archives, thus hindering their further usage.

To guarantee the reliability of the wind measurements, a QC software suite has been designed and applied over the Tall Tower Dataset, and the erroneous data have been removed. Some of the QC functions are coded to deal simultaneously with huge amounts of data so that the computation costs may be high, especially when considering high-resolution data. After the application of the QCSS4TT, the vast majority of the dataset (i.e, the 95.2 % of the wind data) passed all the tests successfully.

A benchmark experiment based on Hubbard et al. (2004) has been designed to assess the efficacy of the QCSS4TT in detecting wrong wind speed data. The exercise is based on the detection of a set of seeded errors introduced in 100 wind hourly time series at 50 randomly selected locations obtained using the ERA5 reanalysis. On average, 40 % of these seeded errors have been identified, even though the magnitude of the error is sometimes close to zero and therefore difficult to detect. This result agrees with those obtained by the previously mentioned publication, thus assuring the reliability of the QCSS4TT results. We do not perform any analogue experiment for wind direction data since the nature of these data requires a more complex exercise.

Even though some tall towers have been decommissioned recently due to several different reasons, most of the locations within the Tall Tower Dataset continue taking measurements that could be added to the collection in the near future. In addition, the authors of this work are open to receive useful input on new tower locations not included in Sect. S1 of the Supplement and whose data could be potentially added to the Tall Tower Dataset in future updates. Enlarging the collection of these non-standard climate data and increasing the density of stations may allow, for instance, further quality checks by means of inter-station comparisons with nearby tall towers.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/essd-12-429-2020-supplement.

Author contributions. JR retrieved and formatted the tall tower data, produced the QCSS4TT code and carried out the benchmark experiment. He also wrote the first draft of this manuscript. This work has been done under the supervision of LL, who conceived the research, gave advice on the data collection, and assisted JR in several IT issues and code debugging. NPZ came up with the design of the benchmark experiment as well as some of the graphics

to visualize the results. AS and FJDR also supervised this work and facilitated the retrieval of data. All authors contributed to the analysis of the results and to the writing and editing of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. The Tall Tower Dataset is made available in good faith to be used for non-commercial purposes. In no event will the authors be liable to any user or third party for any damage or loss resulting from any use or misuse of these data.

Acknowledgements. Authors acknowledge the funding support from INDECIS co-funded by the H2020 ERA-net ERA4CS (GA 690462) and the MICINN grant BES-2017-082216 ("Ayudas para contratos predoctorales"). Exceptional thanks are given to Enric Aguilar for his useful advice on the QC production. The authors are grateful to Marina Conde, Isadora Jiménez, Pierre-Antoine Bretonnière and Javier Vegas for their technical support at different stages of this work. We also acknowledge the ECMWF ERA5 reanalysis production centre.

Finally, the authors would like to acknowledge all the principal investigators and owner institutions of the 222 tall towers within the Tall Tower Dataset for sharing their valuable climate data. We thank the National Data Buoy Center (NDBC), the Earth System Research Laboratory (ESRL) and the Physical Monitoring Program of the Smithsonian Tropical Research Institute (STRI). Special thanks are given to all the contributors to the national mast database by the Renewable Energy and Energy Efficiency Organization in Iran. NCAR/EOL provided BERMS Aspen data under the sponsorship of the National Science Foundation. We are grateful to Jinkyu Hong and Young-San Park (Korea Meteorological Administration) for facilitating Boseong mast data retrieval. Authors would like to thank the Bonneville Power Administration (BPA), the CESAR observatory and Fred Bosveld (KNMI) for providing Cabauw data, the South Africa Weather Service (SAWS), the Wind Atlas for South Africa project (WASA) and the United Kingdom Met Office (UKMO). Many thanks to Lena Kozlova (University of Exeter) for sharing Cabo Verde tall tower data. We acknowledge to The Crown Estate, the BMWi (Bundesministerium für Wirtschaft und Energie, Federal Ministry for Economic Affairs and Energy), the PTJ (Projektträger Jülich, project executing organization), the data contributors to the AsiaFlux database, Ingo Lange (University of Hamburg), Laszlo Haszpra (Hungarian Met Service), Jan Schween (University of Cologne) and Frank Beyrich (Deutscher Wetterdienst) for facilitating acquisition of University of Hamburg, Hegyhátsál, Jülich and Lindenberg mast data, respectively. We also acknowledge Rolf Neubert (University of Groningen), Met Éireann, the NREL National Wind Technology Center (NWTC, Jager and Andreas, 1996), Anna Rutgersson (Uppsala University), Gil Bohrer (The Ohio State University), and the NoordzeeWind B.V. (NZWBV) and its (sub)contractors. Thanks are also given to Ole Ziemer (Nukissiorfiit Hovedkontoret) for providing data and Kurt S. Hansen (DTU) for putting us in contact with the data providers. Park Falls tower data (Davis et al., 2003) was provided by AmeriFlux, whose funding was provided by the U.S. Department of Energy's Office of Science. Thanks are given to the FLUXNET community, who provided valuable data from different masts, and also to Christy Schultz (GMD Met – NOAA) for allowing us access to South Pole mast data. Credit is also given to the sources of the Tumbarumba met mast data and the Vielsalm data manager Anne de Ligne and data provider Tanguy Manise.

Financial support. This research has been supported by the H2020 ERA-net ERA4CS (grant no. 690462).

Review statement. This paper was edited by Kirsten Elger and reviewed by two anonymous referees.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J.: Guidelines on climate metadata and homogenization, World Meteorological Organization, p. 55, 2003.

AWEA: AWEA US Wind Industry Annual Market Report Year Ending 2018, Tech. rep., AWEA, 2019.

Baker, I., Denning, A. S., Hanan, N., Prihodko, L., Uliasz, M., Vidale, P. L., Davis, K., and Bakwin, P.: Simulated and observed fluxes of sensible and latent heat and CO2 at the WLEF-TV tower using SiB2.5, Glob. Change Biol., 9, 1262–1277, https://doi.org/10.1046/j.1365-2486.2003.00671.x, 2003.

Brower, M. C., Barton, M. S., Lledó, L., and Dubois, J.: A study of wind speed variability using global reanalysis data, Tech. Rep. May 2013, AWS True Power, available at: https://www.awstruepower.com/assets/A-Study-of-Wind-Speed-Variability-Using-Global-Reanalysis-Data.pdf (last access: February 2020), 2013.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), Tech. rep., ECMWF, 2017.

Davis, K., Bakwin, P. S., Yi, C., Berger, B. W., Zhaos, C., Teclaw, R. M., and Isebrands, J. G.: The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, Glob. Change Biol., 115, 1278–1293, https://doi.org/10.1029/2009JD012832, 2003.

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., and Bosilovich, M. G.: Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations, J. Climate, 25, 1916–1944, https://doi.org/10.1175/JCLI-D-11-00004.1, 2012.

Dunn, R. J. H., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., and Vose, R. S.: HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011, Clim. Past, 8, 1649–1679, https://doi.org/10.5194/cp-8-1649-2012, 2012.

Fernando, H., Mann, J., Palma, J., Lundquist, J., Barthelmie, R., BeloPereira, M., Brown, W., Chow, F., Gerz, T., Hocut, C., Klein, P., Leo, L., Matos, J., Oncley, S., Pryor, S., Bariteau, L., Bell, T., Bodini, N., Carney, M., Courtney, M., Creegan, E., Dimitrova, R., Gomes, S., Hagen, M., Hyde, J., Kigle, S., Krishnamurthy, R., Lopes, J., Mazzaro, L., Neher, J., Menke, R., Murphy, P., Os-

wald, L., Otarola-Bustos, S., Pattantyus, A., Rodrigues, C. V., Schady, A., Sirin, N., Spuler, S., Svensson, E., Tomaszewski, J., Turner, D., van Veen, L., Vasiljević, N., Vassallo, D., Voss, S., Wildmann, N., and Wang, Y.: The Perdigão: Peering into Microscale Details of Mountain Winds, B. Am. Meteorol. Soc., 100, 799–820, https://doi.org/10.1175/bams-d-17-0227.1, 2018.

Hansen, K. S., Barthelmie, R. J., Jensen, L. E., and Sommer, A.: The impact of turbulence intensity and atmospheric stability on power deficits due to wind turbine wakes at Horns Rev wind farm, Wind Energ., 15, 183–196, https://doi.org/10.1002/we.512, 2012.

Hubbard, K. G., Goddard, S., Sorensen, W. D., Wells, N., and Osugi, T. T.: Performance of quality assurance procedures for an applied climate information system, J. Atmos. Ocean. Technol., 22, 105–112, https://doi.org/10.1175/JTECH-1657.1, 2004.

IOSS: Manual for Real-Time Quality Control of Wind Data, May, 2017.

Jager, D. and Andreas, A.: NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data); NREL Report No. DA-5500-56489., Tech. rep., National Renewable Energy Laboratory, https://doi.org/10.5439/1052222, 1996.

Jiménez, P. A., González-Rouco, J. F., Navarro, J., Montávez, J. P., and García-Bustamante, E.: Quality assurance of surface wind observations from automated weather stations, J. Atmos. Ocean. Tech., 27, 1101–1122, https://doi.org/10.1175/2010JTECHA1404.1, 2010.

Klein Tank, A. M., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A. F., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, Int. J. Climatol., 22, 1441–1453, https://doi.org/10.1002/joc.773, 2002.

Li, Q. S., Zhi, L., and Hu, F.: Boundary layer wind structure from observations on a 325 m tower, Journal of Wind Engineering and Industrial Aerodynamics, 98, 818–832, https://doi.org/10.1016/j.jweia.2010.08.001, 2010.

Lott, J. N.: The quality control of the Integrated Surface Hourly Database, paper presented at the 14th Conference on Applied Climatology, Am. Meteorol. Soc., Seattle, Wash, 10–16, 2004.

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., and Beltrami, H.: Quality control of surface wind observations in Northeastern North America. Part I: Data management issues, J. Atmos. Ocean. Tech., 35, 163–182, https://doi.org/10.1175/JTECH-D-16-0204.1, 2018a.

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., Beltrami, H., and Conte, J.: Quality control of surface wind observations in northeastern North America. Part II: Measurement errors, J. Atmos. Ocean. Tech., 35, 183–205, https://doi.org/10.1175/JTECH-D-16-0205.1, 2018b.

OECD/IEA: Global Energy & CO_2 Status Report 2017, Tech. Rep. March, IEA, available at: https://www.iea.org/ (last access: February 2020), 2018.

OECD/IEA: Global Energy & CO2 Status Report 2018, Tech. rep., available at: https://www.iea.org/ (last access: February 2020), 2019.

Ramon, J. and Lledó, L.: The Tall Tower Dataset https://doi.org/10.23728/b2share.136ecdeee31a45a7906a773095656ddb, 2019a.

Ramon, J. and Lledó, L.: The Tall Tower Dataset. Technical Note, Tech. rep., Barcelona Supercomputing Center – Centro Nacional de Supercomputación, Barcelona, available at: https://earth.bsc.es/wiki/lib/exe/fetch.php?media=library:external:technical_memoranda:technical_report_talltower_database_v2.pdf (last access: February 2020), 2019b.

Ramon, J. and Lledó, L.: The INDECIS-QCSS4TT, GitLab repository, available at: https://earth.bsc.es/gitlab/jramon/INDECIS-QCSS4TT/, 2019c.

Ramon, J., Lledó, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J.: What global reanalysis best represents near-surface winds?, Q. J. Roy. Meteorol. Soc., 145, 3236–3251, https://doi.org/10.1002/qj.3616, 2019.

Tammelin, B., Vihma, T., Atlaskin, E., Badger, J., Fortelius, C., Gregow, H., Horttanainen, M., Hyvönen, R., Kilpinen, J., Latikka, J., Ljungberg, K., Mortensen, N. G., Niemelä, S., Ruosteenoja, K., Salonen, K., Suomi, I., and Venäläinen, A.: Production of the Finnish Wind Atlas, Wind Energ., 16, 19–35, https://doi.org/10.1002/we.517, 2013.

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., and Davis, M.: 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, 2017.

Troen, I. and Petersen, E. L.: European Wind Atlas, Risoe National Laboratory, available at: http://orbit.dtu.dk/files/112135732/European_Wind_Atlas.pdf (last access: February 2020), 1989.

WindEurope: Offshore Wind in Europe. Key trends and statistics 2017, Tech. rep., available at: https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Offshore-Statistics-2017.pdf (last access: February 2020), 2018.

WMO: Guide to the Global Observing System. Third edition, Tech. rep., World Meteorological Organization, Geneva, available at: http://www.wmo.int/pages/prog/www/OSY/Manual/488_Guide_2007.pdf (last access: February 2020), 2007.

Table S1: List of tall towers within the Tall Tower Dataset. The ISO ALPHA-2 code has been utilised to present the country where the tower is located. Latitudes and longitudes are shown in decimal degrees. POR stands for Period Of Record.

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
42361	Shell International EP	US	-92.49	27.55	200507	201612 *	37, 122	hourly	122
42362	Enven Energy Corporation	US	-90.65	27.80	200507	201612 *	37, 122	15minutely, hourly	122
42363	Shell International EP	US	-89.22	28.16	200507	201606 *	37, 122	15minutely, hourly	122
42364	Shell International EP	US	-88.09	29.06	200709	201612 *	37, 122	15minutely, hourly	122
42365	Shell International EP	US	-89.12	28.20	201201	201311 *	37, 122	hourly	122
42369	BP Inc	US	-90.28	27.21	201005	201612 *	2, 60	20minutely	60
42370	BP Inc	US	-90.54	27.32	201005	201211 *	2, 79	20minutely	79
42375	BP Inc	US	-88.29	28.52	201005	201612 *	2, 61	20minutely	61
42394	Shell International EP	US	-89.24	28.16	201409	201612 *	2, 100	hourly	100
42887	BP Inc	US	-88.50	28.19	200911	201612 *	2, 48	20minutely	48
Abadan	SATBA	IR	48.31	30.45	200709	200908 *	2, 10, 30, 38, 40	10minutely	40
Abadeh	SATBA	IR	52.25	31.09	200606	200711†	2, 10, 30, 38, 40	10minutely	40
Abarkooh	SATBA	IR	53.66	31.30	200608	200801†	2, 10, 30, 38, 40	10minutely	40
Abhar	SATBA	IR	49.39	36.11	200706	200907†	2, 10, 30, 38, 40	10minutely	40
Afriz	SATBA	IR	58.96	33.45	200608	200802†	2, 10, 30, 38, 40	10minutely	40
Agh Ghala	SATBA	IR	54.47	37.11	200607	200710†	2, 10, 30, 38, 40	10minutely	40
Ahar	SATBA	IR	47.22	38.59	200811	201504†	2, 10, 30, 38, 40	10minutely	40
Ardakan	SATBA	IR	54.27	32.59	200609	200802†	2, 10, 30, 38, 40	10minutely	40

* Operational

† Decommissioned

[1] TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Asfestan	SATBA	IR	47.60	37.93	200503	200602†	10, 20, 30, 40	10minutely	40
BAO	ESRL	US	-105.00	40.05	200706	201607†	2, 10, 100, 300	10minutely	300
Bardkhoon	SATBA	IR	51.49	27.98	200606	200802†	2, 10, 30, 38, 40	10minutely	40
Barrow	ESRL	US	-156.61	71.32	198801	201605*	2, 10, 18, 20	10minutely, hourly	20
Barzook	SATBA	IR	51.14	33.81	201506	201601†	2, 60, 80, 98, 100	10minutely	100
Barro Colorado Island	Princeton Environmental Institute	PA	-79.85	9.17	200112	201710	2, 20, 42, 48	15minutely, hourly	48
Behabad	SATBA	IR	56.12	31.78	200606	200801†	2, 10, 30, 38, 40	10minutely	40
Old Aspen	UCAR	CA	-106.20	53.63	200210	200912	18, 36, 37, 38	30minutely	38
Binalood	SATBA	IR	59.39	35.99	200212	200309†	10, 30	10minutely	30
Bojnoord	SATBA	IR	57.25	38.14	200608	200805†	2, 10, 30, 38, 40	10minutely	40
Bonab	SATBA	IR	46.03	37.4	200607	200710†	2, 10, 30, 38, 40	10minutely	40
Boroojen	SATBA	IR	51.31	31.97	200606	200711†	2, 10, 30, 38, 40	10minutely	40
Boseong	Yonsei University	KR	127.35	38.27	201404	201610*	10, 20, 40, 60, 80, 100, 140, 180, 220, 260, 300	10minutely	300
Braschaat	INBO	BE	4.52	51.31	199512	201412*	41	30minutely	41
BURL1	NBDC	US	-89.43	28.91	198402	201612*	13, 14, 38	hourly	38
Butler Grade	Bonneville Power Administration	US	-118.68	45.95	200208	201804*	31, 45, 62	10minutely	62
bygl1	NOAA's National Ocean Service	US	-90.42	29.79	200502	201612*	2, 15, 31	10minutely	31

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Cabauw	KNMI	NL	4.93	51.97	198602	201703*	2, 10, 20, 40, 80, 140, 200	10minutely, 30minutely	200
Cape Point	South African Weather Service	ZA	18.48	-34.35	200701	201311*	30	hourly	30
Cardington	UKMO	GB	-0.42	52.10	200405	201303*	10, 25, 50	10minutely	50
Chabahar	SATBA	IR	60.66	25.33	200807	200912†	2, 10, 30, 38, 40	10minutely	40
Chaldoran	SATBA	IR	44.45	39.05	200607	200710†	2, 10, 30, 38, 40	10minutely	40
Changbaishan	Institute of Applied Ecology	CN	127.72	41.70	200212	200511	2, 32	30minutely	32
Chinook	Bonneville Power Administration	US	-119.53	45.83	200601	201611†	50	10minutely	50
CHLV2	NBDC	US	-75.71	36.91	198408	201606†	22, 23, 43	hourly	43
CVO	Cape Verde Atmospheric Observatory	CV	-24.87	16.85	201110	201807*	30	10minutely	30
Davarzan	SATBA	IR	56.81	36.27	200607	200803†	2, 10, 30, 38, 40	10minutely	40
Dehake Saravan	SATBA	IR	62.67	27.14	200606	200712†	2, 10, 30, 38, 40	10minutely	40
Deilaman	SATBA	IR	49.91	36.88	201001	201012†	2, 10, 30, 38, 40	10minutely	40
Delgan	SATBA	IR	59.46	27.49	200608	200712†	2, 30, 38, 40	10minutely	40
Delvar	SATBA	IR	51.05	28.84	200609	200801†	2, 10, 30, 38, 40	10minutely	40
DESW1	NBDC	US	-124.48	47.67	198408	201612*	31, 39	hourly	39
Docking Shoal	Centrica	GB	0.65	53.16	200606	200908	5, 20, 30, 60, 70, 80, 88, 90	10minutely	90
Eghlid	SATBA	IR	52.62	30.89	200606	200805†	2, 10, 30, 38, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Enjilavand	SATBA	IR	50.67	34.94	201105	201207†	2, 30, 38, 40	10minutely	40
Esfaryen	SATBA	IR	57.40	37.05	200608	200803†	2, 10, 30, 38, 40	10minutely	40
Eshtahard	SATBA	IR	50.69	35.73	200807	200912†	2, 10, 30, 38, 40	10minutely	40
Fadashk	SATBA	IR	58.79	32.78	200608	200802†	2, 10, 30, 38, 40	10minutely	40
Falideh	SATBA	IR	49.40	36.81	200207	200403†	10, 20, 30, 40	10minutely	40
Fino1	Fino Project	DE	6.59	54.01	200401	201710*	20, 30, 33, 40, 50, 60, 70, 80, 90, 100	10minutely	100
Fino2	Fino Project	DE	13.15	55.01	200707	201711*	30, 31, 32, 40, 42, 50, 51, 52, 62, 70, 71, 72, 82, 91, 92, 99, 102	10minutely	102
Fino3	Fino Project	DE	7.16	55.20	200909	201711*	23, 28, 29, 30, 40, 50, 55, 60, 70, 80, 90, 95, 100, 106	10minutely	106
fmoa1	NOAA's National Ocean Service	US	-88.02	30.23	200810	201612*	18, 31, 36	10minutely	36
fsnm2	NOAA's National Ocean Service	US	-76.53	39.22	201604	201612*	40, 42	10minutely	42
Fuji Hokuroku	NIES	JP	138.76	35.44	200512	200911*	1, 35	30minutely	35
FWYF1	NBDC	US	-80.10	25.59	199106	201612*	11, 29, 44	hourly	44
Ganje	SATBA	IR	49.46	36.86	200207	200310†	10, 20, 30, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Gardaneh Almas	SATBA	IR	48.67	37.59	200906	201009†	2, 10, 30, 38, 40	10minutely	40
Ghadamgah	SATBA	IR	59.01	36.06	200609	200803†	2, 10, 30, 38, 40	10minutely	40
Ghoroghchi	SATBA	IR	51.00	33.59	201305	201408†	2, 40, 60, 80, 100	10minutely	100
Ghorveh	SATBA	IR	47.75	35.18	200810	200912†	2, 10, 30, 38, 40	10minutely	40
Goodnoe Hills	Bonneville Power Administration	US	-120.55	45.78	200201	201804*	15, 59	10minutely	59
Greater Gabbard MMX	Innogy SE - SSE Renewables	GB	1.90	51.86	201205	201501	3, 23, 25, 43, 45, 62, 64	10minutely	64
Greater Gabbard MMZ	Innogy SE - SSE Renewables	GB	1.92	51.94	200509	201412	42, 52, 62, 72, 82, 84, 88	10minutely	88
Gunfleet Sands	Development Back of Japan -Marubeni Corporation -Dong Energy	GB	1.20	51.73	200201	200711†	61	10minutely	61
Gwangneung Deciduous Forest	Seoul National University	KR	127.15	37.75	200312	200811*	20, 40	30minutely	40
Gwynt Y Mor	UK Green Investment Bank - Stadtwerke München GmbH - Siemens AG - Innogy SE	GB	-3.51	53.48	200509	201412†	24, 25, 44, 45, 64, 82, 85, 90	10minutely	90
Hadadeh	SATBA	IR	54.73	36.25	200608	200802†	2, 10, 30, 38, 40	10minutely	40
Haft Chah	SATBA	IR	52.43	27.72	201002	201107†	10, 30, 38, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Halvan	SATBA	IR	56.30	33.96	200607	200802†	2, 10, 30, 38, 40	10minutely	40
Hamburg University	Hamburg University	DE	10.10	53.52	200401	201812*	10, 50, 110, 175, 250, 280	10minutely	280
Hegyhatsal	Hungarian met service	HU	16.65	46.96	199408	201611*	10, 48, 82, 115	hourly	115
Hendijan	SATBA	IR	49.77	30.12	201004	201110†	0, 2, 30, 38, 40	10minutely	40
Hesarak	SATBA	IR	51.32	35.80	201102	201201†	2, 10, 30, 38, 40	10minutely	40
Hormozgan University	SATBA	IR	56.44	27.26	201402	201601†	2, 10, 30, 38, 40	10minutely	40
Hoseinieh	SATBA	IR	48.18	30.80	200711	200908†	2, 10, 30, 38, 40	10minutely	40
Humber Gateway	E.ON	GB	0.27	53.64	200910	201210†	3, 19, 34, 52, 68, 70, 86, 88, 90	10minutely	90
Hyytiala	Helsinki university	FI	24.29	61.85	199512	201710*	2, 4, 17, 34, 50, 67, 74, 125	10minutely	125
Ijmuiden	ECN	NL	3.44	52.85	201111	201603*	27, 58, 90	10minutely	90
Inner Dowsing	UK Green Investment Bank	GB	0.44	53.13	199908	200802†	2, 16, 41, 43	10minutely	43
Jangal	SATBA	IR	59.21	34.70	200607	200803†	2, 10, 30, 38, 40	10minutely	40
Jask	SATBA	IR	58.11	25.69	200608	200709†	2, 10, 30, 38, 40	10minutely	40
Javim	SATBA	IR	54.09	28.19	200606	200711†	2, 10, 30, 38, 40	10minutely	40
Jirandeh	SATBA	IR	49.78	36.71	200303	200407†	10, 20, 30, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Juelich	Research Center Juelich - Institute for Energy and Climate research (IEK-8)	DE	6.22	50.93	201110	201712*	10, 20, 30, 50, 80, 100, 120	10minutely	120
Kaboodar Ahang	SATBA	IR	48.75	35.35	200607	200710†	2, 10, 30, 38, 40	10minutely	40
Kahak Garm-sar	SATBA	IR	52.32	35.12	200607	200802†	2, 10, 30, 38, 40	10minutely	40
Kahrizak	SATBA	IR	51.32	35.47	200708	200903†	2, 10, 30, 38, 40	10minutely	40
Kennewick	Bonneville Power Administration	US	-119.12	46.10	200201	201804*	24, 37	10minutely	37
Kentish Flats	Vatenfall AB	GB	1.09	51.46	200210	200501†	2, 13, 20, 35, 50, 65, 80	10minutely	80
Kerend Gharb	SATBA	IR	46.19	34.43	201204	201407†	2, 40, 60, 78, 80	10minutely	80
Khaf	SATBA	IR	60.31	34.49	200707	200903†	2, 10, 30, 38, 40	10minutely	40
Khalkhal Bafrajerd	SATBA	IR	48.57	37.54	201109	201410†	2, 10, 30, 38, 40	10minutely	40
Khalkhal Eilkhichi	SATBA	IR	48.25	37.63	200906	201103†	2, 10, 30, 38, 40	10minutely	40
Khash	SATBA	IR	61.06	28.10	200606	200712†	2, 10, 30, 38, 40	10minutely	40
Khomein	SATBA	IR	50.16	33.80	200607	200709†	2, 10, 30, 38, 40	10minutely	40
Kohein	SATBA	IR	49.71	36.34	201105	201504†	2, 40, 60, 78, 80	10minutely	80
Korit	SATBA	IR	56.95	33.44	200607	200801†	2, 10, 30, 38, 40	10minutely	40
Langrood	SATBA	IR	50.23	37.26	200607	200804†	2, 10, 30, 38, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Larijan	SATBA	IR	52.22	35.98	201006	201105†	2, 10, 30, 38, 40	10minutely	40
Latman	SATBA	IR	51.23	35.77	200708	200808†	10, 30, 40	10minutely	40
Likak	SATBA	IR	50.12	30.86	201009	201106†	2, 10, 30, 38, 40	10minutely	40
Lindenberg	Deutscher Wetterdienst	DE	14.12	52.17	199901	201701*	10, 20, 40, 60, 80, 98	10minutely	98
London Array	E.ON - Caisse - Dong Energy - Masdar	GB	1.39	51.59	200412	201012	16, 20, 29, 32, 57, 77, 78, 82	10minutely	82
Lootak Zabol	SATBA	IR	61.39	30.73	200606	201001†	2, 10, 30, 38, 40	10minutely	40
lopl1	Louisiana Offshore Oil Port	US	-90.03	28.89	201108	201612*	40, 58	15minutely	58
Lutjewad	Gronigen university	NL	6.35	53.40	200012	201701	2, 7, 40, 60	10minutely	60
Mae Klong	National Institute of Advanced Industrial Science and Technology	TH	98.84	14.58	200212	200411*	42, 45	30minutely	45
Mahidasht	SATBA	IR	46.73	34.39	200606	200709†	2, 10, 30, 38, 40	10minutely	40
Mahshahr	SATBA	IR	49.09	30.58	200709	200908†	2, 10, 30, 38, 40	10minutely	40
Malin Head	Met Éireann	IE	-7.33	55.35	198801	201712*	1, 2, 22	10minutely, hourly	22
Manjil	SATBA	IR	49.40	36.74	200402	200411†	10, 20, 40	10minutely	40
Marvdasht	SATBA	IR	52.92	29.98	200606	200711†	2, 10, 30, 38, 40	10minutely	40
Mauna Loa	ESRL	US	-155.58	19.54	199101	201605*	2, 10, 20, 40	10minutely, hourly	40
Mayan	SATBA	IR	46.05	38.09	200607	200801†	2, 10, 30, 38, 40	10minutely	40
Megler	Bonneville Power Administration	US	-123.88	46.27	200210	201804*	53	10minutely	53

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Meshkin Shahr	SATBA	IR	47.73	38.27	200811	201003†	2, 10, 30, 38, 40	10minutely	40
mhrn6	NOAA's National Ocean Service	US	-74.16	40.64	201505	201612*	46	10minutely	46
Mil Nader	SATBA	IR	61.16	31.09	201009	201203†	2, 10, 30, 38, 40	10minutely	40
Mir Javeh	SATBA	IR	61.44	29.03	200905	201008†	2, 10, 30, 38, 40	10minutely	40
Mir Khand	SATBA	IR	49.40	36.67	200207	200310†	10, 30, 40	10minutely	40
Moalleman	SATBA	IR	54.57	34.87	200608	200802†	2, 10, 30, 38, 40	10minutely	40
Moghar	SATBA	IR	52.18	33.57	200606	200711†	2, 10, 30, 38, 40	10minutely	40
Nahavand	SATBA	IR	48.21	34.27	200607	200709†	2, 10, 30, 38, 40	10minutely	40
Namin	SATBA	IR	48.38	38.38	200607	200712†	2, 10, 30, 38, 40	10minutely	40
Nanortalik	DTU	DK	-45.23	60.14	200706	200906	2, 10, 30, 41, 49	10minutely	49
Naselle Ridge	Bonneville Power Administration	US	-123.80	46.42	201002	201804*	30	10minutely, 5minutely	30
Nikooye	SATBA	IR	49.53	36.31	200911	201206†	2, 10, 30, 38, 40	10minutely	40
Nir	SATBA	IR	47.98	38.03	201305	201411†	2, 40, 60, 78, 80	10minutely	80
NOAH	FoundOcean	GB	-1.49	55.14	201209	201403	2, 35, 52, 69, 86, 101, 103	10minutely	103
Nosrat Abad	SATBA	IR	60.16	29.81	200606	200712†	2, 30, 38, 40	10minutely	40
NWTC M2	NREL	US	-105.23	39.91	199609	201701*	2, 10, 20, 50, 80	10minutely	80

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
NWTC M4	NREL	US	-105.23	39.91	201201	201604*	3, 10, 15, 26, 30, 50, 76, 80, 88, 100, 131, 134	10minutely	134
NWTC M5	NREL	US	-105.23	39.21	201208	201705*	3, 10, 15, 30, 38, 41, 55, 61, 74, 80, 87, 100, 105, 119, 122, 130	10minutely	130
Obninsk	Institute of Experimental Technology	RU	36.60	55.11	200712	201604*	2, 8, 121, 301	hourly	301
Oestergarnsholm	Uppsala university	SE	18.98	57.43	200306	201412*	7, 29	10minutely	29
Ohio State University	Ohio State University	US	-84.71	45.56	200701	201707*	2, 34, 46	30minutely	46
Egmond aan zee	ECN	NL	4.39	52.61	200508	201012*	2, 21, 70, 116	10minutely	116
Palangkaraya	Hokkaido University	ID	114.04	2.35	200112	200511*	42	hourly	42
Papooli	SATBA	IR	50.06	36.08	200907	201011†	2, 10, 30, 38, 40	10minutely	40
WLEF	ESRL	US	-90.27	45.95	200301	201711*	2, 30, 122, 396	hourly	396
Pasoh	Kyoto University	MY	102.30	2.97	200212	200911*	53	30minutely	53
Puijo	Finnish Meteorological Institute	FI	27.65	62.91	200510	201512*	75	10minutely	75
Qianyanzhou	Northwest Plateau Institute of Biology	CN	115.07	26.73	200212	200411	2, 39	30minutely	39

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Race Bank	Race Bank	GB	0.75	53.31	200606	201304†	15, 20, 30, 50, 70, 80, 88, 89	10minutely	89
Rafsanjan	SATBA	IR	56.22	30.32	200606	200807†	2, 10, 30, 38, 40	10minutely	40
ROAM4	NBDC	US	-89.31	47.87	198310	201612*	39, 46, 47	hourly	47
Roodab	SATBA	IR	57.35	36.05	200808	201003†	2, 10, 30, 38, 40	10minutely	40
Rostamabad	SATBA	IR	49.49	36.90	200201	200307†	10, 20, 30, 40	10minutely	40
Sakaerat	National Institute of Advanced Industrial Science and Technology	TH	101.92	14.49	200012	200311*	45, 47	30minutely	47
American Samoa	ESRL	AS	-170.56	-14.25	199406	201605*	2, 21	10minutely	21
Sanar	SATBA	IR	51.31	36.50	200607	200708†	2, 10, 30, 38, 40	10minutely	40
Sarakhs	SATBA	IR	61.14	36.31	200609	200711†	2, 10, 30, 38, 40	10minutely	40
Saravan	SATBA	IR	62.26	27.42	201010	201110†	2, 30, 38, 40	10minutely	40
Saveh Site	SATBA	IR	50.40	35.08	200805	200909†	2, 10, 30, 38, 40	10minutely	40
Semnan	SATBA	IR	53.45	35.62	200907	201011†	2, 10, 30, 38, 40	10minutely	40
Seven Mile	Bonneville Power Administration	US	-121.27	45.63	200201	201804*	15, 30	10minutely	30
SGOF1	NBDC	US	-84.86	29.41	200310	201612*	20, 35	hourly	35
Shahr Abad	SATBA	IR	56.20	37.65	201104	201112†	2, 10, 30, 38, 40	10minutely	40
Shahr Babak	SATBA	IR	55.22	30.09	200609	200807†	2, 10, 30, 38, 40	10minutely	40
Shandol	SATBA	IR	61.66	31.15	201010	201201†	2, 10, 30, 38, 40	10minutely	40

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Shell Flats Mast 1	Centrica UK	GB	-3.29	53.86	201107	201311†	2, 50, 70, 80, 82	10minutely	82
Shell Flats Mast 2	Centrica UK	GB	-3.20	53.87	201107	201401†	30, 40, 50, 52	10minutely	52
Sheykh Tapeh	SATBA	IR	45.08	37.52	201207	201504†	2, 30, 38, 40	10minutely	40
Shiraz Site	SATBA	IR	52.61	29.37	200712	200906†	10, 20, 40	10minutely	40
Shooshtar	SATBA	IR	48.76	31.79	200711	200908†	2, 10, 30, 38, 40	10minutely	40
Shorjeh	SATBA	IR	49.44	36.07	200807	201001†	2, 10, 30, 38, 40	10minutely	40
skmg1	Skidaway Institute of Oceanography	US	-80.24	31.53	200409	200801	50	hourly	50
Sodankyla	Finnish Meteorological Institute	FI	26.64	67.36	200012	201412*	2, 24	30minutely	24
South Carolina	Savannah River National Laboratory	US	-81.83	33.41	200904	201712*	34, 68, 329	15minutely	329
South Pole	ESRL	US	-24.80	-89.98	200711	201812*	2, 10, 20, 30	10minutely, hourly	30
spag1	Skidaway Institute of Oceanography	US	-80.57	31.38	200401	200909	50	hourly	50
STDM4	NBDC	US	-87.22	47.18	198407	201612*	28, 35	hourly	35
Summit	ESRL	GL	-38.48	72.58	200806	201605*	2, 10, 20, 50	10minutely	50
Tafresh	SATBA	IR	50.06	34.68	201009	201302†	2, 10, 30, 38, 40	10minutely	40
Huisun	National Chung Hsing University	TW	121.13	24.08	201012	201311*	60	30minutely	60
Taleghan Site	SATBA	IR	50.57	36.12	200712	201002†	10, 20, 40	10minutely	40
Tange Hashi	SATBA	IR	52.96	29.18	201503	201509†	2, 40, 60, 78, 80	10minutely	80

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Tarom	SATBA	IR	49.03	36.66	201106	201306†	2, 10, 30, 38, 40	10minutely	40
Tiksi	Roshydromet - Finnish Meteo-rological Institue - U.S. National Oceanic and Atmospheric Administration	RU	128.89	71.60	201008	201809*	6, 9, 10, 15, 21	10minutely	21
Too Takaboon	SATBA	IR	49.52	36.91	200204	200312†	10, 20, 30	10minutely	30
Trinidad Head	ESRL	US	-124.15	41.05	200204	201605	2, 10, 20	10minutely	20
Troutdale	Bonneville Power Administration	US	-122.4	45.56	201002	201804*	30	10minutely	30
Tumbarumba	CSIRO Marine and Atmospheric Research	AU	148.15	-35.66	200101	201412*	70	hourly	70
tybg1	Skidaway Insti-tute of Oceanog-raphy	US	-79.92	31.63	200401	200801	32, 34	hourly	34
upbc1	NOAA's National Ocean Service'	US	-122.12	38.04	201302	201612	100	10minutely	100
Varzaneh	SATBA	IR	52.62	32.46	200606	200810†	2, 10, 30, 38, 40	10minutely	40
Vasf	SATBA	IR	50.93	34.19	200809	200902†	2, 10, 30, 38, 40	10minutely	40
Vielsalm	Université Catholique de Louvian	BE	6.00	50.31	199608	200904*	2, 3, 9, 12, 21, 40, 50, 51, 52	30minutely	52
Wallaby Creek	University of Western Aus-tralia	AU	145.19	-37.43	200501	200812†	2, 10, 20, 45, 90, 110	30minutely	110
Walnut Grove	ESRL/DOE	US	-121.49	38.27	200508	201611	9, 122, 244, 366, 488	hourly	488

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
Wasco	Bonneville Power Administration	US	-120.77	45.50	200509	201804*	30	10minutely	30
wdel1	Shell International EP	US	-89.55	28.66	200812	201609*	41	hourly	41
West Branch	ESRL - IOWA university	US	-91.35	41.72	200801	200807	30, 99, 379	20minutely	379
WM01	Republic of SouthAfrica - Department of Energy	ZA	16.66	-28.60	201006	201701*	6, 10, 20, 40, 60, 62	10minutely	62
WM02	Republic of SouthAfrica - Department of Energy	ZA	19.36	-31.52	201006	201701*	6, 10, 20, 40, 60, 62	10minutely	62
WM03	Republic of SouthAfrica - Department of Energy	ZA	18.42	-31.73	201006	201701*	6, 10, 20, 40, 60, 62	10minutely	62
WM04	Republic of SouthAfrica - Department of Energy	ZA	18.11	-32.85	201005	201306†	6, 10, 20, 40, 60, 62	10minutely	62
WM05	Republic of SouthAfrica - Department of Energy	ZA	19.69	-34.61	201005	201701*	6, 10, 20, 40, 60, 62	10minutely	62
WM06	Republic of SouthAfrica - Department of Energy	ZA	20.69	-32.56	201009	201612*	6, 10, 20, 40, 60, 62	10minutely	62
WM07	Republic of SouthAfrica - Department of Energy	ZA	22.56	-32.97	201005	201701*	6, 10, 20, 40, 60, 62	10minutely	62

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
WM08	Republic of SouthAfrica - Department of Energy	ZA	24.51	-34.11	201008	201701*	6, 10, 20, 40, 60, 62	10minutely	62
WM09	Republic of SouthAfrica - Department of Energy	ZA	25.03	-31.25	201009	201612*	6, 10, 20, 40, 60, 62	10minutely	62
WM10	Republic of SouthAfrica - Department of Energy	ZA	28.14	-32.09	201008	201612*	6, 10, 20, 40, 60, 62	10minutely	62
WM11	Republic of SouthAfrica - Department of Energy	ZA	28.07	-30.81	201510	201707*	6, 10, 20, 40, 60, 62	10minutely	62
WM12	Republic of SouthAfrica - Department of Energy	ZA	30.53	-29.85	201510	201707*	6, 10, 20, 40, 60, 62	10minutely	62
WM13	Republic of SouthAfrica - Department of Energy	ZA	32.17	-27.43	201510	201707*	6, 10, 20, 40, 60, 62	10minutely	62
WM14	Republic of SouthAfrica - Department of Energy	ZA	29.54	-27.88	201510	201707*	6, 10, 20, 40, 60, 62	10minutely	62
WM15	Republic of SouthAfrica - Department of Energy	ZA	27.12	-28.62	201509	201707*	6, 10, 20, 40, 60, 62	10minutely	62

* Operational

† Decommissioned

[1]TML: Top Measuring Level

Table S1: Continued

Tower name	Institution	Country	Longitude	Latitude	POR start	POR end	Measuring levels (m)	Sampling	TML[1] (m)
wslm4	Great Lakes Environmental Research Laboratory	US	-85.14	45.84	201504	201612*	36, 43	30minutely	43
Xishuangbanna	Xishuangbanna Tropical Botanical Garden	CN	101.20	21.95	200212	200511*	42, 70	30minutely	70
Zahedan	SATBA	IR	60.81	29.47	201101	201201†	2, 30, 38, 40	10minutely	40
Zarrineh2	SATBA	IR	46.93	36.06	201503	201601†	2, 40, 60, 78, 80	10minutely	80
Zartoshtabad	SATBA	IR	48.50	37.61	201408	201504†	2, 40, 60, 78, 80	10minutely	80

* Operational

† Decommissioned

[1]TML: Top Measuring Level

S2 QC main tests

S2.1 Plausible values

Wind speed and wind direction records falling outside a physically possible range of values are commonly found within the time series. They are mainly produced by gross errors in the data loggers or storage. This test detects and flags unrealistic values such as negative wind speed values or observations above a maximum allowed threshold. The absolute maximum limit has been set to the maximum wind gust measurement ever recorded on the earth surface, which is 113.3 ms^{-1} measured in Barrow Island (Australia) under the effects of Olivia cyclone in April 1996 (Courtney et al., 2012). A lower threshold can be selected from which wind speed values can be flagged as suspect. This value is set to 75 ms^{-1}, which is the one suggested by the WMO (WMO, 2007) and besides, this fixed-value also corresponds to Vaisala's sensors highest measurable value. Wind direction values falling outside the range from 0 to 360 degrees are also flagged as erroneous.

S2.2 Difference between extreme values of the wind distribution

One of the potential uses of the Tall Tower Dataset is the detection of severe weather events by looking at the extreme values of the empirical wind speed distribution. However, some of these extreme measurements might be erroneous and need to be flagged accordingly. This QC check detects and flags unrealistic extreme wind speed values of the time series by checking the difference between the maximum and the second maximum values of the distribution of wind speed values. If the difference between them exceeds the absolute value of the second maximum, the first maximum is flagged as suspect. This test runs iteratively until the previously mentioned condition is not satisfied.

S2.3 Persistence test

Wind time series are usually characterised by strong variability, alternating periods of high and low fluctuations. Nevertheless, the presence of long periods of extremely low variability can be unrealistic since they can be produced by errors in the measuring sensors or instrumental drift. The persistence test detects and flags sequences of wind speed and wind direction observations with abnormally low variability. However, it is important to take into consideration those relatively long periods with very low variability and mean wind speed values close to zero are typical of the observed natural variability (e.g., static high-pressure systems during several days in a row producing weak winds). Hence, these data cannot be considered erroneous. Thus, the persistence test does not introduce any flag to wind speeds weaker than 0.5 ms^{-1}. These measurements are then flagged as calms.

The WMO proposes that 1-minute data should vary at least 0.5 ms^{-1} over 60 consecutive wind speed values, and 10 degrees in the case of wind direction records. Otherwise data should be flagged as doubtful. These thresholds have been adapted to the resolutions reported by the towers. Thus, wind speed periods are flagged as suspect if the wind speed does not change more than 0.7 ms^{-1} in 60 consecutive values. Wind direction values will be considered suspicious when the range between the maximum and the minimum values in a sequence of 60 records is lower than 5 degrees.

The example plotted in Figure S1 shows wind speed observations measured at 18 meters at the top of the Barrow tower (Arctic Circle) during 51 consecutive days. In except of the two spikes on 14^{th} October and 3^{rd} November, wind speed values range from 4.8 ms^{-1} to 5.3 ms^{-1}. This variability is significantly low when compared with the rest of the wind series (not shown). Although the *Persistence test* flags the records as a suspect, a visual inspection reveals that they are potentially erroneous and should not be used as reliable data.

S2.4 Flat line

A sequence of numbers with null standard deviation is the extreme case of a period with low variability and indicates that several constant values are observed consecutively. The probability of recording constant values in a row decreases with the number of significant figures that a sensor can record, being almost unlikely to have more than five consecutive exact matches for wind speed (IOSS, 2017) and 40 for wind direction measurements. In this sense, data fail the *flat line* test when there exist 6 -or more- constant wind speed values in a row. This threshold is increased to 40 for the wind direction variable. Observing

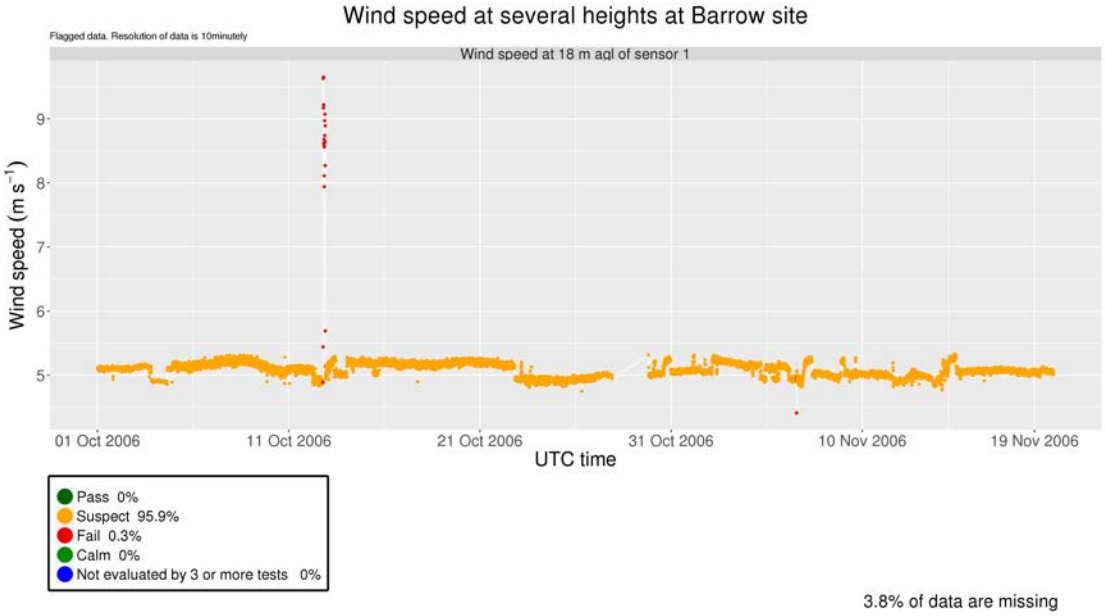


Figure S1. Wind speed time series at 18 meters above ground level at Barrow site (71.32ºN 156.61ºW, 11 m), USA

3, 4 or 5 exact consecutive matches is more likely for wind speed values, but still unlikely to happen frequently. Therefore, the tests flags as suspect those flat sequences. Analogously for wind direction data, flat sequences containing 20 to 40 wind direction records are flagged as suspect. It is also frequent to observe an alternation of no data periods with null speed values, which are usually produced by failures in the sensors or data loggers. If the period containing this alternating pattern exceeds 30 days, all the measurements within this period are flagged as erroneous.

A detection of a flat line is shown in Figure S2. Various sequences of constant values are encountered at the three different levels between September 14^{th} and September 20^{th}. Like that, flat lines are often detected simultaneously at all levels of the tower.

S2.5 Icing

Freezing rain or fog usually frosts the anemometers and vanes placed along the tall tower preventing them from measuring non-zero wind speed values and changes in the wind direction. Hence, these records should be detected by checking wind and temperature observations simultaneously. Based on Jiménez et al. (2010), data are considered wrong when the *Icing test* detects 4 or more days with 0 ms^{-1} as the maximum wind speed value and below zero temperatures during all the same period.

Wind speed series at different heights at Hegyhatsal tower are represented in Figure S3. A flat line is observed in the two uppermost levels from December 8^{th} to December 18^{th} 2002. However, the air temperature observations (Figure S4) reveal that negative Celsius temperatures occurred during all the ten days in the two top levels of the tower. Given these conditions, it is very likely that an icing event happened and frosted the two upper anemometers.

S2.6 Abnormal variations

Random and gross errors in the measurements might produce periods of abnormally high or abnormally low variability and usually, appear embedded in the wind speed time series. Various authors have proposed several different thresholds that define a period with extreme variability (see Jiménez et al. (2010)) since the threshold selection should depend on the local wind features. In an attempt of generalisation, in this work it is proposed that these limits are defined by statistical parameters

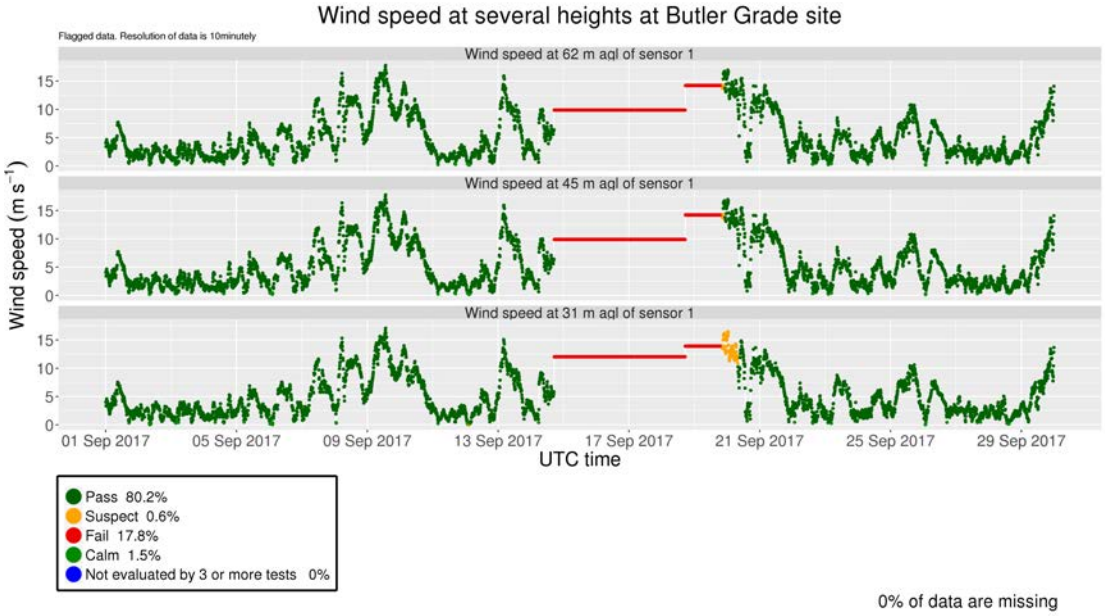


Figure S2. Wind speed time series at 31, 45 and 62 meters above ground level at Butler Grade site, USA (45.95ºN, 118.68ºW, 545 m).

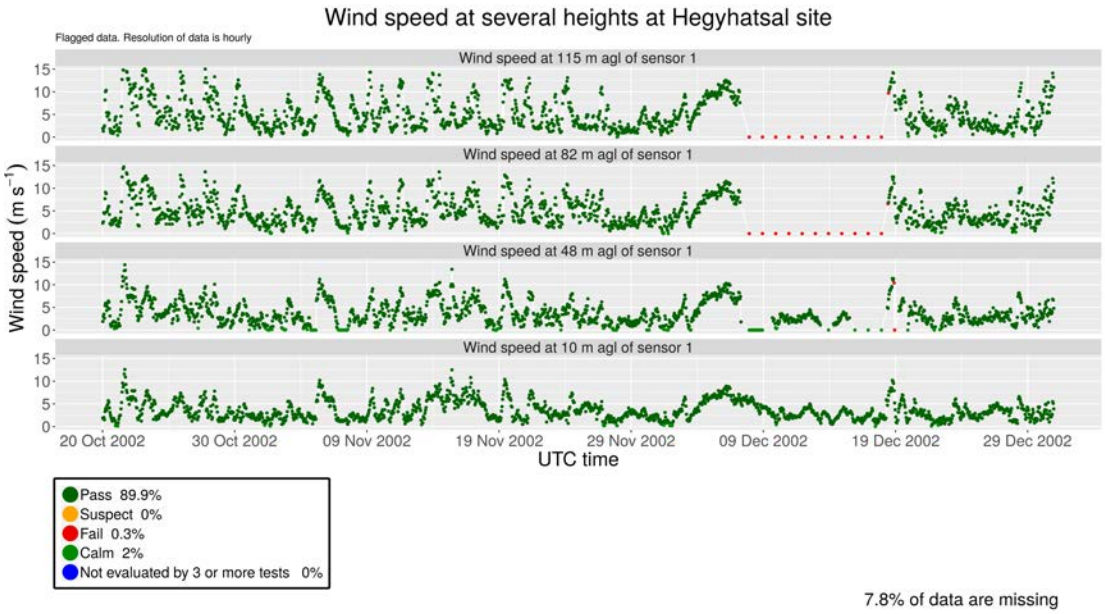


Figure S3. Wind speed time series at 10, 48, 82 and 115 meters above ground level at Hegyhatsal tall tower, Hungary (46.96ºN, 16.65ºE, 248 m).

derived from the wind distributions themselves. In this way, the abnormal variations check compares the variability (computed

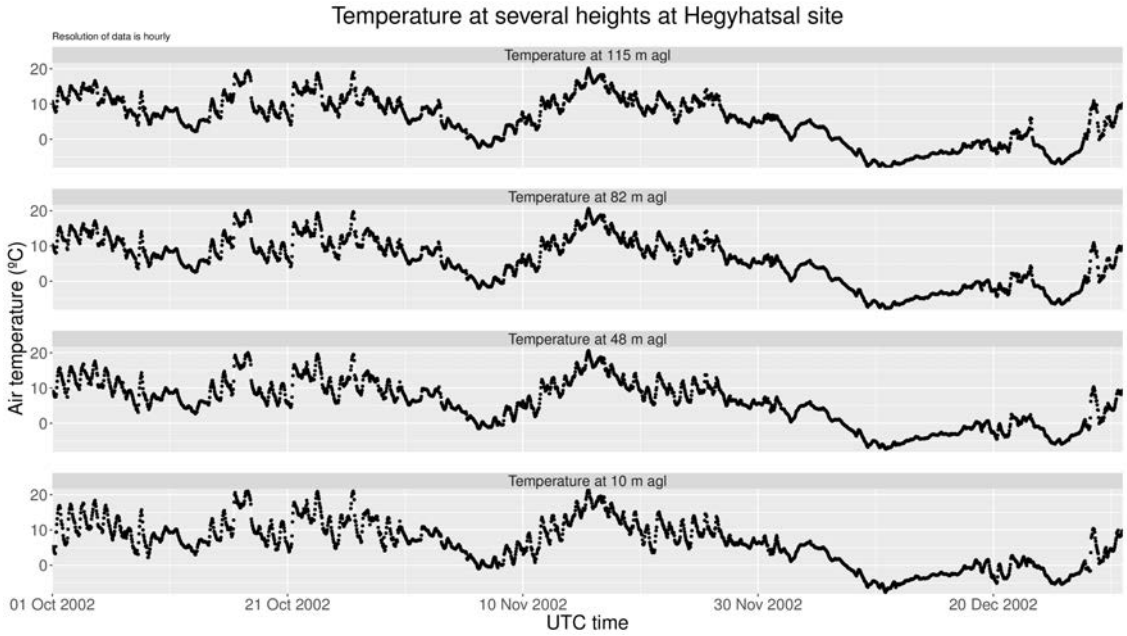


Figure S4. Temperature measurements at 10, 48, 82 and 115 meters above ground level at Hegyhatsal tall tower, Hungary (46.96ºN, 16.65ºE, 248 m).

as the variance) of 30-day periods with the mean variance of all 30-day periods of the time series using moving variances. If the standard deviation of a specific 30-day period departs more than four standard deviations from the mean standard deviation, records within these 30 days are all flagged as suspect.

S2.7 Systematic errors

Another approach to detecting random and systematic errors in the experimental measurements is based on the computation of moving averages. Similar to the abnormal variations check, this QC routine computes the mean wind speeds over a 30-day moving window. Wind speed values within a 30-day period whose average departs more than four standard deviations from the mean value of all 30-day moving means are all considered suspect.

In Figure S5, the *Systematic errors* check flags as suspect 12 consecutive days of wind speed measurements taken at the top of Hegyhatsal tower. A close inspection reveals that the minimum wind speed record is over 5 ms^{-1} during all the mentioned period, which is in disagreement with the wind speeds observed at lower levels. Indeed, the three anemometers located at 10, 48, and 82 meters report weaker winds or even calm during these 12 days. It is likely that an offset value could have been inserted in the data logger producing the inconsistency observed in the uppermost wind speed measurements. In this case, these 12 days of winds at 115 meters should not be considered reliable.

Figure S6 shows a false detection of a systematic error at WLEF tall tower. Although the test flags as suspect a period of 2 months of wind speed data at the 122-meter level, a visual inspection and comparison with winds reported at other tower heights does not reveal any inconsistency in the suspicious observations. Hence, these data should not be discarded unless a sensor failure is reported in the metadata of the site.

S2.8 Quartile occurrences

A third method to detect periods containing gross errors in the measuring process is suggested here by looking at the number of consecutive days where no value is above or below the first, second and third quartiles of the empirical wind speed distribution.

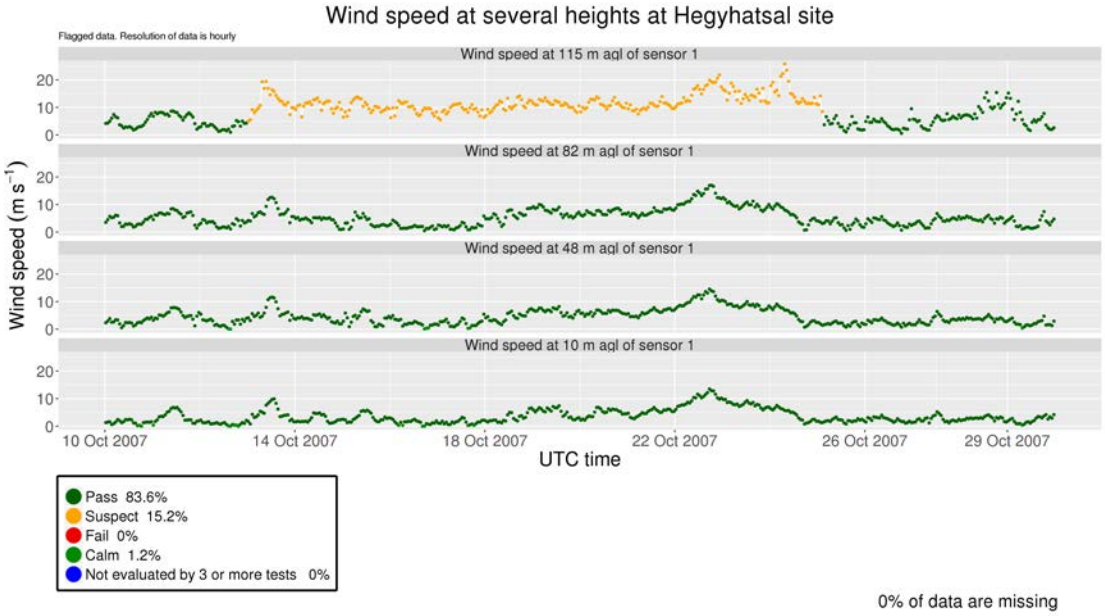


Figure S5. Wind speed time series at 10, 48, 82 and 115 meters above ground level at Hegyhatsal tall tower, Hungary (46.96ºN, 16.65ºE, 248 m).

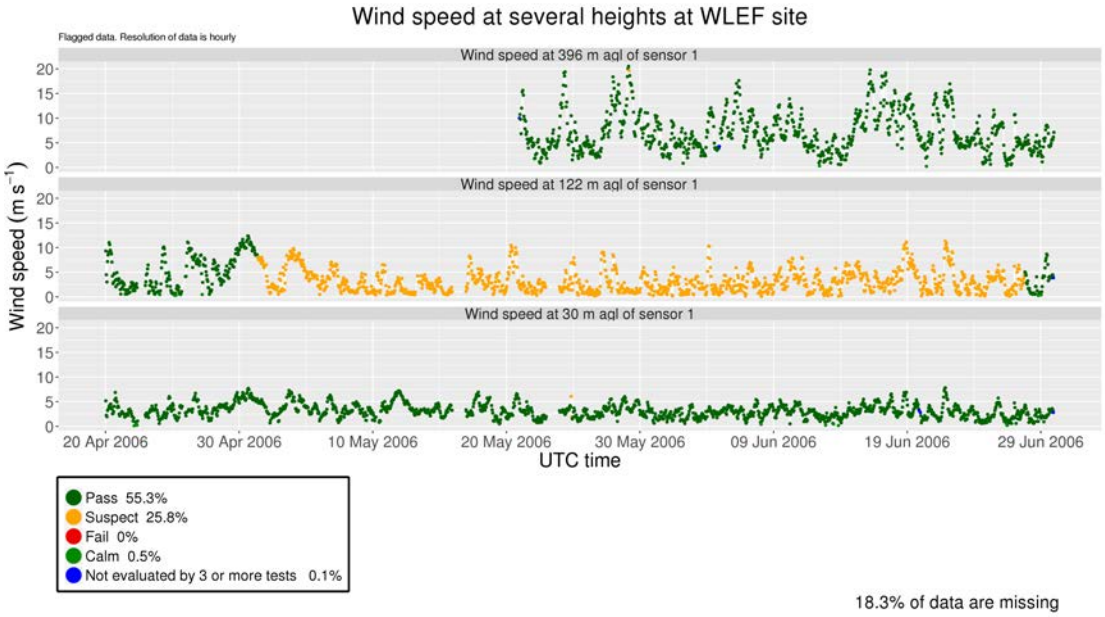


Figure S6. Wind speed time series at 30, 122, 396 meters above ground level at WLEF tall tower, USA (45.95ºN, 90.27ºW, 472 m).

Table S2 summarises the different thresholds (in days) that define the trustworthiness of an observation. As an example, the

first row indicates that if all the observations in 30 days fall above the first quartile, data within this period will be flagged as erroneous. Observations are suspicious when the period without any occurrence within the first quartile ranges between 15 and 30 days. Spans shorter than 15 days without any value falling within the first quartile are considered correct by this test.

Table S2. Threshold values (in days) that set the different levels of confidence for the *Quartile occurrences* check.

All the observations are...	Pass	Suspect	Fail
$>1^{st}$ quartile	<15	[15,30]	>30
$>2^{nd}$ quartile	<10	[10,20]	>20
$>3^{rd}$ quartile	<5	[5,10]	>10
$<1^{st}$ quartile	<5	[5,10]	>10
$<2^{nd}$ quartile	<10	[10,20]	>20
$<3^{rd}$ quartile	<15	[15,30]	>30

S2.9 Rate of change

The presence of spikes in the wind series is usually observed during extreme wind phenomena events. However, the magnitude of these peaks is constrained to a specific allowable range of values specially when the very high-frequency wind data are averaged in periods of several minutes (which is the case of the observations within the Tall Tower Dataset). This test compares pairs of adjacent observations. To pass the test successfully, differences between consecutive values must be lower than a specific threshold, that can be either dynamically established or fixed (IOSS, 2017). The *Rate of change* test uses the interquantile range (IQR) of the considered series, defined as the difference between the third and first quartiles of the empirical distribution. When the difference between two consecutive values exceeds three times the value of the IQR, both values are considered wrong. If the difference is between twofold and threefold the IQR, the pair of observations is considered as suspect.

S2.10 Step test

The *Step test* uses a similar methodology as the *Rate of change* test to detect spurious peaks of wind speed data. In the *Step test*, the maximum permissible difference between two consecutive observations is fixed to 20 ms^{-1} (WMO, 2007), instead of using a statistic derived from the wind series. Although the WMO suggests this limit specifically for 2-minutely averaged wind speed data, their usage has been deemed appropriate for the data within the Tall Tower Dataset since the time stamp samplings observed in this collection are larger. Indeed, by averaging data in longer periods, one can expect a general smoothing of the series, hence reducing the possibility of observing big data spikes.

S2.11 Repeated sequences

This check looks for sequences of observations that appear repeated within the same time series. Duplicated sequences of at least 30 wind speed values are flagged as erroneous if data do not contain any decimal places. The threshold is decreased to 20 wind speed observations if data are measured with one or more decimal digits. Wind direction series are also checked for duplicate sequences, and they are flagged when the length of the repeated sequence exceeds 30 values.

Duplicated sequences have been found in the three parallel wind time series at Abadan tall tower time series (Figure S7). A careful inspection reveals that the values within the two black rectangles in the top series match perfectly. An analogous situation is noticed for the two lower levels. This is probably due to an standard procedure to fill in no-data periods, which takes previously observed wind speed sequences of data. However, it has been deemed appropriate to the detect and consider erroneous these sequences of data.

S2.12 Tower shadow

One of the singularities of the tall tower data is that wind measurements are not taken at the top of a pole where a sensor is placed. Instead, anemometers and wind vanes are distributed along with the vertical structure of the tall tower, which usually

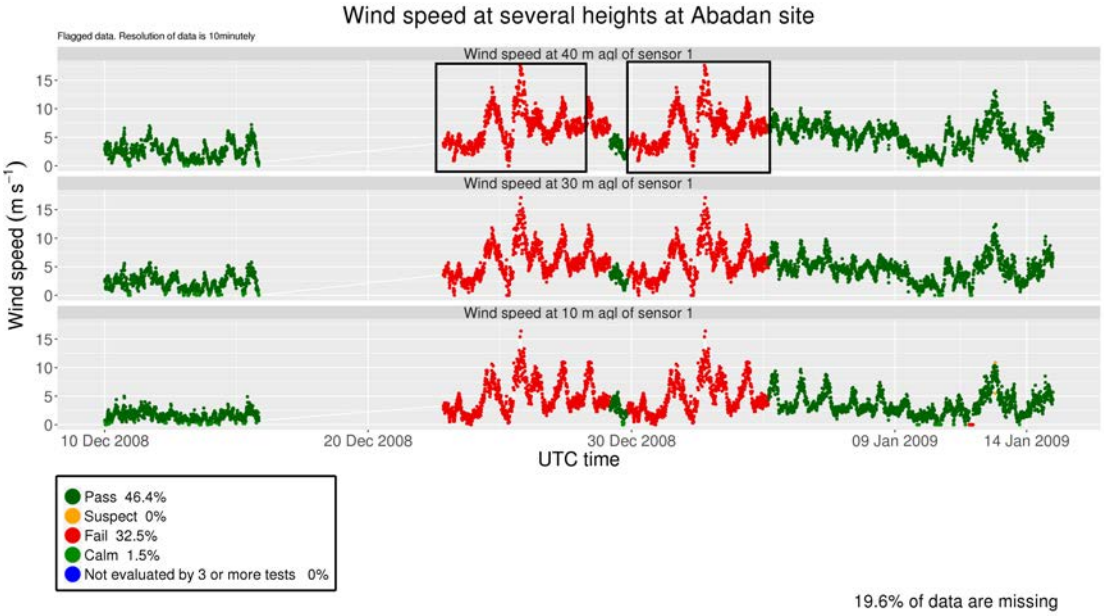


Figure S7. Wind speed time series at 10, 30 and 40 meters above ground level at Abadan met mast, Iran (30.45ºN, 48.31ºE, 4 m). The two black boxes in the upper graph represent two duplicated sequences of wind speed values within the same time series.

consists of a solid vertical cylinder or a lattice structure that inherently produces a wind shadow in the downwind area. If an anemometer is measuring in the shadow area, wind speeds are affected by this shadow and cannot be considered reliable.

To help overcome this handicap, a common practice in the instrumental installation is to place redundant sensors at the same height in booms oriented to different cardinal directions. Shadowed records can then be replaced by those from a sensor not affected. The *Tower shadow* test identifies first the shadowed directions and anemometers by dividing wind speeds from two sensors at the same level. Ideally, they should measure the same values so that the ratio is expected to be equal to the unit unless the winds from one sensor are shadowed. Then, all wind speed ratios are grouped in wind direction sectors of 1 degree. The 5^{th} and 95^{th} percentiles of the distribution generated by all the ratios are calculated next. Those directions showing ratios below the 5^{th} percentile and above the 95^{th} are considered to be in the wake of the tower. After identifying the shaded directions for each anemometer, the test marks as suspect those wind speed values affected.

Figure S8 exemplifies the previous explanation presenting the ratios between simultaneous wind speeds observations measured by redundant sensors at 60 and 100 meters at the FINO3 met mast in the North Sea. The quotient between wind speeds reported by two different sensors is approximately one for most of the wind directions. However, wind speeds coming from 50 ± 5 and 170 ± 5 degrees of direction are affected by the vertical pole at the two measuring levels. Thus, the anemometer measuring the weakened winds is identified, and those records should not be considered correct.

S2.13 Vertical ratios

QC checks that employ nearby stations are not suitable for meteorological variables with remarkably localised features such as precipitation or wind speed, because the correlation between neighbour series is considerably lower when compared to temperature or pressure time series (Dunn et al., 2012). In addition, those tests require a dense network of stations, which is not the case of the Tall Tower Dataset. However, another particularity of tall tower data is the simultaneous records taken at the same time at different heights along the mast. These series can be compared among them as they are expected to be highly correlated. The *Vertical ratios* is a particular test which considers pairs of time series measured at different heights and computes the mean ratio ($\overline{r}$) of all the pair-wise measurements ratios (r_i). To avoid duplication and save computation time, the

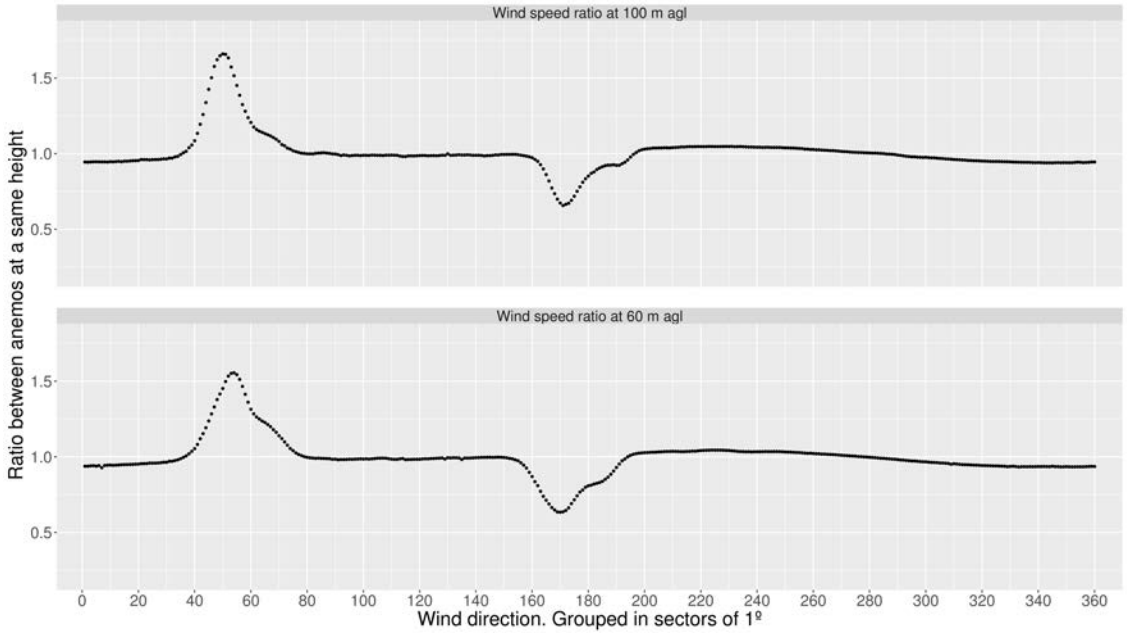


Figure S8. Ratio between simultaneously measured wind speed values at 60 and 100 meters at FINO3 met mast, Germany (55.20ºN, 7.16ºE, 0 m).

test only computes the ratio between one level an all the lower levels. In except local effects such as low-level jets, wind speeds tend to increase in height, so the computed mean ratio is expected to be greater or equal to unity. Taking this assumption into account, the *Vertical ratios* test will detect and flag as erroneous those pairs which ratio (r_i) satisfies the following condition:

$$r_i \geq \overline{r} + 30 \tag{1}$$

Values are considered dubious when the following condition is satisfied:

$$r_i \geq \overline{r} + 15 \tag{2}$$

Even though the allowable ranges of ratios was initially chosen somewhat arbitrarily, it has been tested and adjusted using the data within the Tall Tower Dataset to ensure that only gross errors are detected and flagged as erroneous. Wind speeds under 1 ms^{-1} are not considered in this test.

S2.14 Isolated pass

After running some of the QC tests, a certain amount of sequences might be flagged as wrong or dubious. These sequences can be found close in time and encircle values marked as correct by the QC checks. However, it is very likely that those presumably correct values are not be acceptable since a prolonged sensor failure may have occurred, but the previously run QC checks missed it. The *Isolated pass* check is applied after running at least one QC test and attempts to detect those apparently correct (we note that calms are also identified as good data) sequences of observations surrounded by wrong or suspect values, and change their flag into erroneous or suspect. Besides, we also force to be wrong those scattered individual records appearing randomly within long no-data periods.

A total of 12 predefined sequences (see Table S3) containing data flagged as correct ('Pass' or 'Calm') but surrounded to the left and right by, wrong ('Fail'), dubious ('Suspect') or absent ('Missing') records have been defined. Wherever these series

are found, the central ‘Pass’ or ’Calm’ values are changed from ’Pass’ to ’Fail’. Table 6 defines similar sequences, but their central records will be changed from ‘Pass’ to ‘Suspect’.

Table S3. Explicit definition of the sequences to be searched within the wind time series which central value or values flag will be changed from ‘Pass’ or ’Calm’ to ‘Fail’.

$Fail, Fail, Fail, Pass, Fail, Fail, Fail$
$\underbrace{Fail, ..., Fail}_{5}, Pass, Pass, \underbrace{Fail, ..., Fail}_{5}$
$\underbrace{Fail, ..., Fail}_{10}, Pass, Pass, Pass, \underbrace{Fail, ..., Fail}_{10}$
$\underbrace{Fail, ..., Fail}_{15}, Pass, Pass, Pass, Pass, \underbrace{Fail, ..., Fail}_{15}$
$\underbrace{Fail, ..., Fail}_{25}, Pass, Pass, Pass, Pass, Pass, \underbrace{Fail, ..., Fail}_{25}$
$\underbrace{Missing, ..., Missing}_{50}, Pass, \underbrace{Missing, ..., Missing}_{50}$
$\underbrace{Missing, ..., Missing}_{50}, Calm, \underbrace{Missing, ..., Missing}_{50}$

Table S4. Explicit definition of the sequences to be searched within the wind time series which central value or values will be changed from ‘Pass’ flag to ‘Suspect’.

$Suspect, Suspect, Suspect, Pass, Suspect, Suspect, Suspect$
$\underbrace{Suspect, ..., Suspect}_{5}, Pass, Pass, \underbrace{Suspect, ..., Suspect}_{5}$
$\underbrace{Suspect, ..., Suspect}_{10}, Pass, Pass, Pass, \underbrace{Suspect, ..., Suspect}_{10}$
$\underbrace{Suspect, ..., Suspect}_{15}, Pass, Pass, Pass, Pass, \underbrace{Suspect, ..., Suspect}_{15}$
$\underbrace{Suspect, ..., Suspect}_{25}, Pass, Pass, Pass, Pass, Pass, \underbrace{Suspect, ..., Suspect}_{25}$

S2.15 Occurrences of 0s and 360s values

The lack of coordination concerning the data storage and formatting conventions in the original data may produce some issues that must be detected. For example, in the wind speed time series, missing records are sometimes set to zero, thus leading to a spurious increase in the occurrence of the zero value. Similarly, some conventions use the value 0 degrees to refer to the northern wind direction while others identify this direction with 360 degrees. Stations with properly detailed metadata information include the convention adopted by the data managers. Regrettably, most of the stations whose data was accessed to be included in the Tall Tower Dataset did not attach such complete information. In those cases, the original basic standards such as assigning the 0 or the 360 value to the north direction need to be inferred.

This routine computes the percentage of occurrence of each of these three cases:

1. Occurrences of 0s within the wind speed time series,
2. occurrences of 0s within the wind direction series and
3. occurrences of 360s within the wind direction series.

The *Occurrences of 0s and 360s values* does not flag individual records, but provides a value for each of the series indicating the percentage of the aforementioned occurrences to the total data. The whole series is considered incorrect if any of these occurrences exceeds 30%, which has been chosen appropriately to take into account that a considerable percentage of calms may exist.

S2.16 Internal consistency

Whenever a null wind speed is recorded, the associated wind direction value is meaningless since it is very likely that the wind vane is still pointing to the direction defined by the last non-zero wind speed observation. According to the WMO guidelines, whenever a null wind speed is reported, the simultaneous wind direction measurement must be forced to be null as well. However, in the Tall Tower Dataset the zero wind direction value indicates the true North. Therefore, for null wind speed records, wind direction must be set to NA. We note that the condition must be only applied for wind measurements taken at the same height above ground level.

References

Courtney, J., Buchan, S., Cerveny, R., Bessemoulin, P., Peterson, T., Rubiera Torres, J., Beven, J., King, J., Trewin, B., and Rancourt, K.: Documentation and verification of the world extreme wind gust record: 113.3 m/s on Barrow Island, Australia, during passage of tropical cyclone Olivia, Australian Meteorological and Oceanographic Journal, 62, 1–9, https://doi.org/10.22499/2.6201.001, http://www.bom.gov.au/jshess/docs/2012/courtney.pdf, 2012.

Dunn, R. J., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., and Vose, R. S.: HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Climate of the Past, 8, 1649–1679, https://doi.org/10.5194/cp-8-1649-2012, 2012.

IOSS: Manual for Real-Time Quality Control of Wind Data, May, 2017.

Jiménez, P. A., González-Rouco, J. F., Navarro, J., Montávez, J. P., and García-Bustamante, E.: Quality assurance of surface wind observations from automated weather stations, Journal of Atmospheric and Oceanic Technology, 27, 1101–1122, https://doi.org/10.1175/2010JTECHA1404.1, 2010.

WMO: Guide to the Global Observing System. Third edition, Tech. rep., World Meteorological Organization, Geneva, http://www.wmo.int/pages/prog/www/OSY/Manual/488_Guide_2007.pdf, 2007.

Chapter 5

Reanalysis versus reality. Are they too different?

Objective

Investigate to what extent reanalyses are capable of representing the true near-surface wind speeds. These products are widely recognised as trusted sources to derive wind speed estimates. Therefore, by quantifying the error associated with each reanalysis, we ultimately help the reanalysis user choose the most accurate product.

Methodology

- Describe the main differences and agreements between five reanalyses for surface wind speeds.

- Verify the set of reanalyses against the Tall Tower Dataset.

Results

- Important discrepancies are found in terms of seasonal mean winds, interannual variability and trends.
- For example, while JRA55 displays high values of interannual variability inland, ERA-Interim and ERA5 depict systematically lower values in the same regions.
- Trends —both positive and negative— are seen over specific regions by all five reanalyses.
- Reanalyses suffer from representativeness errors, and these can be large sometimes.
- The ERA5 outperforms the other reanalyses in reproducing the tall tower winds.

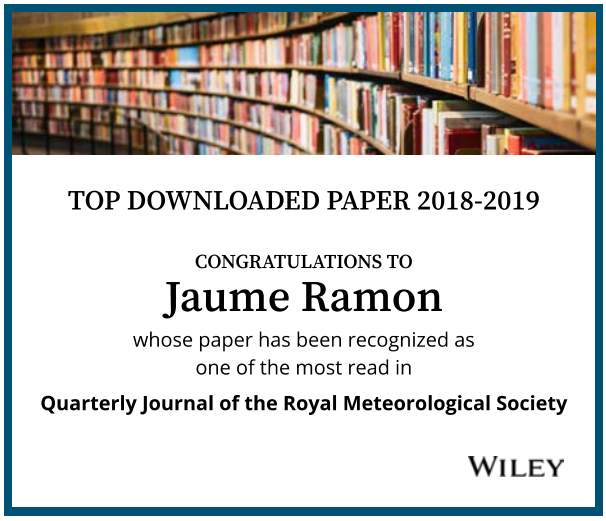
Conclusions

- We discourage using global reanalyses to obtain accurate estimates of seasonal mean winds.
- The temporal variability of near-surface wind speeds is well represented in global reanalyses, but representativeness errors make absolute values less reliable.
- The analysis of trends requires further attention, particularly in those regions where winds have been weakening during decades.

Publication

- Ramon, J., Lledo, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? *Quarterly Journal of the Royal Meteorological Society*, 145(724):3236–3251, doi: 10.1002/qj.3616

The article presented in this chapter has obtained two distinctions by the *Quarterly Journal of the Royal Meteorological Society*:



Received: 13 May 2019 | Revised: 04 July 2019 | Accepted: 10 July 2019 | Published on: 20 August 2019

DOI: 10.1002/qj.3616

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

What global reanalysis best represents near-surface winds?

Jaume Ramon[1] | Llorenç Lledó[1] | Verónica Torralba[1] | Albert Soret[1] | Francisco J. Doblas-Reyes[1,2]

[1]Barcelona Supercomputing Center (BSC), c/ Jordi Girona, Barcelona, Spain

[2]ICREA, Pg. Lluís Companys 23, Barcelona, Spain

Correspondence
Jaume Ramon, Barcelona Supercomputing Center (BSC), c/ Jordi Girona 29, Barcelona 08034, Spain.
Email: jaume.ramon@bsc.es

Abstract

Since global reanalysis datasets first appeared in the 1990s, they have become an essential tool to understand the climate of the past. The wind power industry uses those products extensively for wind resource assessment, while several climate services for energy rely on them as well. Nowadays various datasets coexist, which complicates the selection of the most suitable source for each purpose. In an effort to identify the products that best represent the wind speed features at turbine hub heights, five state-of-the-art global reanalyses have been analysed: ERA5, ERA-Interim, the Japanese 55-year Reanalysis (JRA55), the Modern Era Retrospective Analysis for Research and Applications-2 (MERRA2), and the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis 1 (R1). A multi-reanalysis ensemble approach is used to explore the main differences amongst these datasets in terms of surface wind characteristics. Then, the quality of the surface and near-surface winds is evaluated with a set of 77 instrumented tall towers. Results reveal that important discrepancies exist in terms of boreal winter seasonal means, interannual variability (IAV), and decadal linear trends. The differences in the computation of these parameters, which are mainly concentrated inland, reach up to the order of magnitude of the parameters themselves. Comparison with in situ observations shows that the ERA5 surface winds offer the best agreement, correlating and reproducing the observed variability better than a multi-reanalysis mean in 35.1% of the tall tower sites on a daily time-scale. However, none of the reanalyses stands out from the others when comparing seasonal mean winds. Regarding the IAV, near-surface winds from ERA5 offer the values closest to the observed IAV.

KEYWORDS

hub-height wind, reanalysis, surface wind, tall towers, verification, wind resource assessment

1 | INTRODUCTION

Global reanalysis datasets are an essential tool in different research disciplines to investigate past atmospheric conditions. Reanalyses are the result of combining a frozen state-of-the-art numerical model with the assimilation of past observations from several sources (Fujiwara *et al.*, 2017) to produce a consistent dataset for a long record of time (typically more than 30 years). These datasets allow us to infer features such as variability or trends for regions or

variables where in situ observations are lacking. However, these products have inherent uncertainties that derive from model simplifications, observational uncertainties, and the data assimilation procedure. Although there are several ongoing initiatives in the community to produce reanalysis datasets (e.g., Copernicus Climate Change Service (C3S), 2017), only a few efforts have been devoted to comparing their quality in a coordinated way (Fujiwara *et al.*, 2017). It is true, however, that several studies have focused on verifying reanalysis data on a regional scale (Kumar and Hu, 2012; Alvarez *et al.*, 2014; Kaiser-Weiss *et al.*, 2015; Sharp *et al.*, 2015; Olauson, 2018; Rehman Tahir *et al.*, 2018; Uotila *et al.*, 2018). Others intended to cover the whole world by using interpolated observational datasets (Donat *et al.*, 2014) or employing some stations distributed worldwide (Decker *et al.*, 2012). A more comprehensive verification would be beneficial, especially for socio-economic sectors that employ reanalysis datasets.

The energy industry is one of the biggest user groups of reanalysis datasets (Gregow *et al.*, 2015). Specifically for the wind industry, the lack of long and homogeneous records of wind speed observations has favoured the adoption of reanalyses for assessing wind resources (Cannon *et al.*, 2015). Preconstructive wind resource assessment studies need to determine long-term mean wind speed (Tammelin *et al.*, 2013) and its probability distribution accurately at each turbine location to estimate wind power generation and revenues. The characterization of year-to-year variations of the wind speed is crucial to understanding the risk of a wind farm project (Pryor *et al.*, 2006; Brower *et al.*, 2013) and also estimating the amount of debt that banks can finance. Information on long-term trends can also be used to foresee changes in the performance of wind farms during their lifetime, which typically comprises between 20 and 30 years (Jude and Leseney, 2017). All these evaluations require long records, which typically are not available at the location where the farm is planned to be built, which has led wind energy users to rely on reanalysis products.

Another frequent usage of reanalyses for already operating wind farms is to understand the causes of anomalous wind speed episodes that impacted generation and revenues on monthly or seasonal time-scales. In recent years, the climate prediction community has started to unveil the climate drivers of anomalous events such as wind droughts (e.g., Lledó *et al.*, 2018) and to produce seasonal forecasts (Doblas-Reyes *et al.*, 2013; Clark *et al.*, 2017) to anticipate those wind speed anomalies that can have an impact on wind energy activity. In this climate prediction framework, reanalyses are also used as observational reference for the adjustment of systematic errors affecting these predictions (Torralba *et al.*, 2017b) and to assess the forecast quality (Jolliffe and Stephenson, 2012).

The different applications that use—either directly or indirectly—reanalyses require a scientific evaluation of the quality of the different available reanalysis datasets. An assessment is needed especially in those cases in which one single product is employed and commonly chosen ad hoc without taking into account its related uncertainty.

In this work, we aim to shed light on the uncertainty associated with observational references to help the reanalysis user decide about the suitability of a reanalysis dataset. The assessment will be done by comparing surface wind speeds from different global reanalysis datasets and verifying them further with in situ observations to eventually select the most accurate sources. The analysis focuses on three different aspects of the datasets: mean wind speeds, variability, and trends. Section 2 describes the reanalysis and observational datasets employed. The methodology is described in section 3. An intercomparison of the reanalysis set is presented in section 4 and the verification results are presented in section 5, including an interpretation in both sections. Discussion and conclusions follow in section 6.

2 | DATASETS

2.1 | Reanalysis datasets

Five widely used state-of-the-art global reanalyses are considered in this study. They include the two European Centre for Medium-Range Weather Forecasts (ECMWF) reanalyses ERA5 (Copernicus Climate Change Service (C3S), 2017) and ERA-Interim (Dee *et al.*, 2011b), the Japanese 55-year Reanalysis (JRA55: Kobayashi *et al.*, 2015), NASA's Modern Era Retrospective Analysis for Research and Applications-2 (MERRA2: Molod *et al.*, 2015), and National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis 1 (R1: Kalnay *et al.*, 1996). Table 1 lists the main characteristics concerning the time periods, data assimilation schemes, spatiotemporal resolution, single levels employed in this study, and references.

Other reanalysis products were considered preliminarily to be included in this work, but have been disregarded for several reasons. The climate forecast system reanalysis (CFSR) dataset (Saha *et al.*, 2010), which is published by NCEP, has been produced with two different model configurations before and after March 31, 2011, including different spectral resolutions.[1] Since this discontinuity produces detectable changes in mean wind speeds, the CFSR dataset has not been considered for the present study. Previous versions of the selected datasets such as JRA-25, ERA40, and MERRA have been discarded too, as they have been discontinued and superseded by newer products released by the respective centres. The case of NCEP/NCAR R1 is an exception. After its first publication in 1996, the National Oceanic and Atmospheric Administration (NOAA)/NCEP and NCAR released a newer reanalysis

[1] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2

TABLE 1 Summary of the basic characteristics of the five reanalysis datasets used in this study (ordered alphabetically)

Name	ERA-Interim	ERA5	JRA55	MERRA2	R1
Institution	ECMWF	ECMWF	JMA	NASA GMAO	NOAA/NCEP and NCAR
Period coverage	1979 to present[a]	1980 to present[b]	1978 to present	1980 to present	1948 to present
Time resolution	6-hr	1-hr	6-hr	1-hr	6-hr
Assimilation scheme	4D-VAR	4D-VAR	4D-VAR	3D-VAR-FGAT	3D-VAR
Horizontal grid spacing	0.75° × 0.75°	0.3° × 0.3°	1.25° × 1.25°	0.5° latitude × 0.625° longitude	1.875° latitude × 2° longitude
Vertical levels	60 hybrid ($\sigma - p$)	137 hybrid ($\sigma - p$)	60 hybrid ($\sigma - p$)	72 hybrid ($\sigma - p$)	28 (σ)
Single levels available	surface	surface, 100 m	surface	surface, 50 m	surface
Top level	0.1 hPa	0.01 hPa	0.1 hPa	0.01 hPa	3 hPa
Reference	Dee *et al.* (2011b)	Copernicus Climate Change Service (C3S) (2017)	Kobayashi *et al.* (2015)	Molod *et al.* (2015)	Kalnay *et al.* (1996)

[a]Expected to be discontinued at the end of August 2019

[b]Expected to be expanded back to 1950 by the end of 2019

in 2002, namely the NCEP/DOE Reanalysis AMIP-II or R2 (Kanamitsu *et al.*, 2002). Nonetheless, some studies (e.g., Lucio-Eceiza *et al.*, 2018) have revealed its poor performance in estimating surface wind speeds, which is worse than its predecessor R1. Many regional reanalyses with higher spatial resolutions exist as well, but they do not cover the whole globe, and their usage may be limited in some applications, such as verification and bias adjustment of subseasonal to seasonal predictions made at the global scale. Moreover, all regional reanalyses are refinements of a global reanalysis produced through dynamical downscaling and data assimilation techniques (Mesinger *et al.*, 2006; Gleeson *et al.*, 2017). Therefore, their value largely depends on the quality of the driving reanalysis, so that they have not been included here.

The whole set of reanalyses used in this study assimilate both conventional data (such as surface synoptic (SYNOP) ocean stations, radiosonde profiles, or aircraft records) and satellite measurements (provided mainly by scatterometers and the Special Sensor Microwave Imager (SSMI)). All of the selected datasets include some observations of surface oceanic winds and all of them, except R1, assimilate satellite observations of ocean surface winds (Fujiwara *et al.*, 2017). It is worth mentioning that none of the reanalyses ingests surface winds from land stations. This is known to be problematic, especially over areas with inhomogeneous terrain, where stations experience strong local influences and do not represent the grid-area winds appropriately. Instead, the 10-m wind speeds, hereafter referred to as surface wind speeds, are parametrized in planetary boundary layer schemes from surface characteristics, stability indices, and the lowest native model winds. For instance, ERA5 and ERA-Interim use Monin–Obukhov theory and an open-terrain roughness to make the field more comparable with SYNOP stations. MERRA2 also uses a scheme based on the Monin–Obukhov similarity theory (Helfand and Schubert, 1995), which includes the effect of a viscous sublayer for heat and moisture transport over all surfaces except land (Molod *et al.*, 2015). The NCEP/NCAR R1 uses a bulk model (see Stull (2012), section 11.2.3) with an aerodynamic formula that considers the fluxes of sensible flux, heat, and momentum proportional to the difference between values at the surface and the adjacent atmosphere (Kanamitsu, 1989). The proportionality constants depend on the wind speed as well as the static stability of the surface layer and are based on the work of Businger *et al.* (1971), and subsequently the Monin–Obukhov similarity theory. According to Torralba *et al.* (2017a), the extrapolation of surface winds in the JRA55 reanalysis from the lowest model level is carried out assuming neutral stability. In areas where this level is located too high above the surface (such as forested zones), this assumption may lead to a significant reduction of wind speeds after extrapolation to the surface level.

Some of the most recent reanalyses (e.g., MERRA2 and ERA5) have started to provide winds at turbine hub heights, in reaction to the demands of the wind-power industry. Those winds can describe the vertical wind profile better within the first 100 m of the atmosphere. In particular, ERA5 supplies wind speeds at 100 m above the ground, while MERRA2 offers wind speeds computed at the 50-m level. In this article, they will be referred to as ERA5 and MERRA2 near-surface winds, respectively.

2.2 | Tall tower observations

The accuracy of the reanalysis in the representation of both surface and hub-height wind speed is assessed in this study through a comparison with hub-height wind-speed observations measured in tall meteorological towers (also known as met masts) distributed worldwide. This type of structure

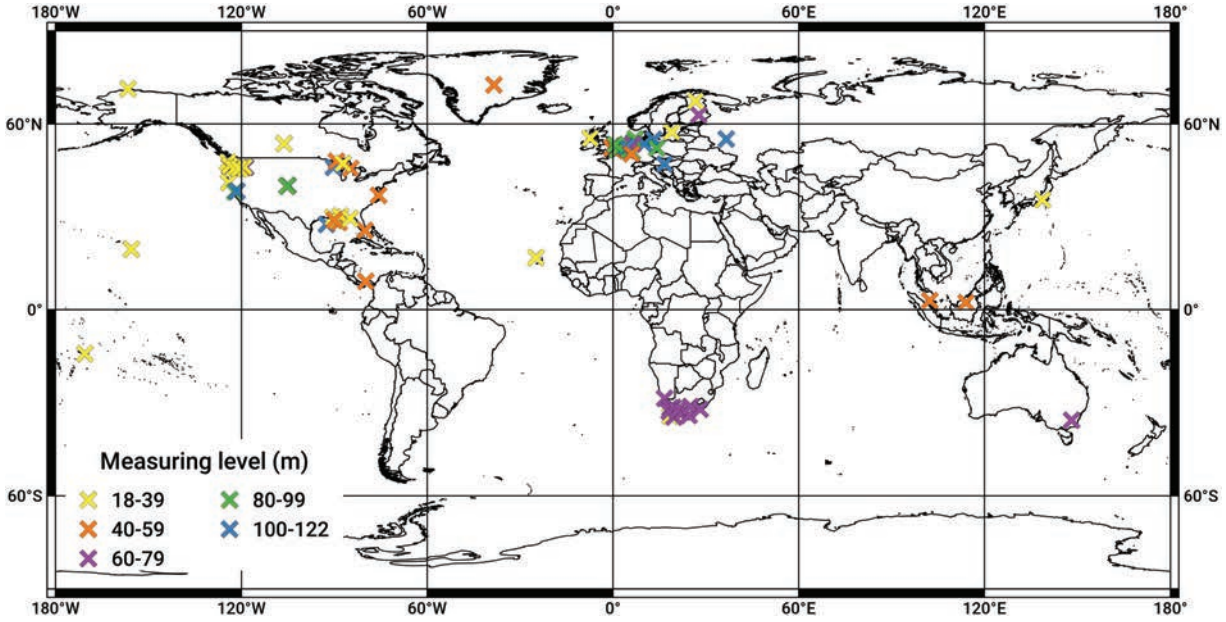


FIGURE 1 Global distribution of the 77 tall towers. Colours indicate, for each tall tower, the height of the measuring level employed in this study

allows the placement of several sensors along a high vertical tower, the altitude of which commonly varies from 10 to more than 100 m (e.g., Van Ulden and Wieringa, 1996; Kouwenhoven, 2007). Tall tower data are widely used within the wind power industry to characterize the wind-flow fine features at turbine hub heights, that is, between 80 and 120 m for modern turbines, as well as monitoring the current weather situation in wind farms. Thus, these measurements are preferred over those taken at surface level, which are usually disturbed by the surrounding elements. The maintenance of these structures, however, is much more expensive than for surface stations, and therefore these data appear sparsely and are usually difficult to find for public usage.

A new unique dataset containing quality-controlled wind observations plus other climate observations such as temperature or relative humidity from 222 different observational sites has recently been created under the name of the Tall Tower Dataset (Ramon and Lledó, 2019). The reporting time sampling of these masts ranges from 10 min to 1 hr. Due to the high tower maintenance costs mentioned above, the record length of these time series represents a major handicap in this study, as most tall towers do not have a lifespan of more than three years. Moreover, these series often contain interruptions in the measuring process, most likely produced by sensor failures or tower maintenance tasks, so that the net record length available can be shortened even more. To reach a balance between quality and amount of data, we have selected those towers containing at least three years of data. Although this threshold has been chosen somewhat arbitrarily, we considered it optimal for the purposes of the present work. Hence, the total number of towers employed in this study is eventually set to 77. Their spatial distribution is presented in Figure 1 and the main meta information is specified in the Supporting Information in Table S1. Even though the masts appear over all continents, their distribution is highly heterogeneous, and most of the locations appear clustered in Europe and North America, which are areas with a large deployment of wind energy. For each of the towers, we have selected the anemometer that is measuring at the closest level to 100 m (see Figure 1 and Table S1), which is the average of modern turbine hub heights. Note that in some towers only one anemometer is installed at the top of the mast, and it is sometimes placed far from a 100 m height.

As far as the authors know, the Tall Tower Dataset constitutes an independent set of observations that has not been assimilated previously by any reanalysis, thus assuring the statistical independence with the reanalysis products. This is important to ensure a fair comparison, since a verification with observations employed in the assimilation of a reanalysis would lead to biased scores.

3 | METHODS

3.1 | Reanalysis intercomparison

To establish the main differences and agreements amongst the five global reanalyses, the common period 1980–2017 has been selected. Firstly, the surface wind speeds from all reanalyses have been computed from the corresponding zonal and meridional components. Then, seasonal averages have been calculated from 6-hr (in the case of ERA-Interim,

JRA55, and R1 datasets) or hourly (for ERA5 and MERRA2 reanalyses) instantaneous values. Results have been produced for each season: December–January–February (DJF), March–April–May (MAM), June–July–August (JJA) and September–October–November (SON). As the five reanalyses use different grids (Table 1), an interpolation to a common grid is needed for comparison. All grids have been interpolated to the F47(T62) horizontal grid, which is R1's grid and is the coarsest out of the five reanalysis grids. A conservative interpolation method from the R package *s2dverification* (Manubens *et al.*, 2018) has been employed to this end.

The differences in terms of seasonal mean wind speeds, the interannual variability (IAV)—computed separately for each season as the standard deviation of the seasonal means within the 1980–2017 period and normalized by the seasonal climatology of all years under consideration—and the long-term trends have been assessed. A multi-reanalysis approach, which is analogous to the often applied multi-model approach in climate modelling, has been used for mean winds and IAV. Firstly, the seasonal mean and IAV are computed for each of the reanalyses. Then, the multi-reanalysis mean (MR) is computed at each grid point with a nonweighted mean. Together with the MR, the spread is computed as the range of values (i.e., the difference between the absolute maximum and minimum values) at each grid point. This parameter highlights the areas where different datasets are in disagreement. Finally, the difference between each reanalysis and the MR is computed (also referred to as the departure) in order to quantify the contribution of each of the reanalyses to the overall uncertainty.

Long-term linear trends have been computed using a simple linear regression equation, where the independent variable is time and the dependent variable is wind speed. Thus, the slope represents the rate of change in wind speed over time. This parameter has subsequently been normalized by the seasonal mean-wind speed value and unit-converted to be expressed as a percentage change per decade, which can also be referred to as a decadal trend. We also test the significance of the decadal trends being different from zero by applying a Student's *t*-test. This methodology followed in the computation of decadal trends is identical to the approach used in Torralba *et al.* (2017a) to facilitate the comparison of results.

3.2 | Verification with tall tower observations

Tall tower observations are originally available at hourly or subhourly time resolution. Daily and seasonal mean values have been computed, so they can be compared against reanalysis daily and seasonal means (which were obtained from hourly or 6-hr model outputs). Although one might argue that reanalysis products with hourly resolution have an advantage in the verification, some tests (not shown) revealed that this factor does not change any of the conclusions. Moreover, a user-oriented verification has to highlight the quality of the entire product available, including all its features, such as better spatial or time resolution.

To verify reanalysis datasets with in situ observations, the gridded data have to be interpolated horizontally at each tower location. Three horizontal interpolation methods have been tested in order to select the most appropriate one for the objectives of this work.

- Nearest-neighbour (i.e., takes the closest grid point to the location of interest). This has been considered in the literature as the most suitable method for variables with high spatiotemporal variability like precipitation (Accadia *et al.*, 2003). Additionally, this approach has also been employed in wind-speed variability studies such as that of Lucio-Eceiza *et al.* (2018).
- Nine-point average (i.e., averaging the nine nearest grid points). For variables with high spatial variability, this method produces smoother results.
- Bilinear interpolation (Press *et al.*, 2006, equation 3.6.5). Several studies (Stohl *et al.*, 2002; Mentis *et al.*, 2015) have used the bilinear method to regrid wind fields.

To compare reanalysis surface and near-surface data against each tall tower's winds, a vertical extrapolation of reanalysis wind speeds (WS) to the tower measuring height closest to 100 m (h) is performed using a power law:

$$WS(h) = WS(h_{\text{ref}}) \left(\frac{h}{h_{\text{ref}}} \right)^{\alpha}, \tag{1}$$

where h_{ref} is the reference height of the reanalysis field (i.e., 10, 50, or 100 m). Alpha (α) is a non-dimensional wind-shear exponent. A value of 0.143 has been used for land (Touma, 1977), while 0.11 has been employed over water bodies (Hsu *et al.*, 1994). We note that both exponents assume neutral stability. According to the previous two references, whereas neutral stability prevails in oceanic areas, the onshore shear exponent is more dependent on stability conditions. This approximation is fair enough when considering long-period averages (such as seasonal means), but might add some uncertainty when considering daily averages (Kelly, 2016).

Daily averages at each location have been verified in terms of spatiotemporal correlation, standard deviation, and centred root-mean-squared error (CRMSE: Jolliffe and Stephenson, 2012, equation 6.2). A Taylor diagram (Taylor, 2001) is used here to visualize these three statistical parameters from multiple reanalyses in one single graph. In addition, two tests of significance have been performed to evaluate the statistical significance of the change in correlation and standard deviation when comparing the results from two different datasets. For the correlation, the Williams test (Williams, 1959) has been employed to assess the difference between two dependent correlations that share one variable. Analogously, the

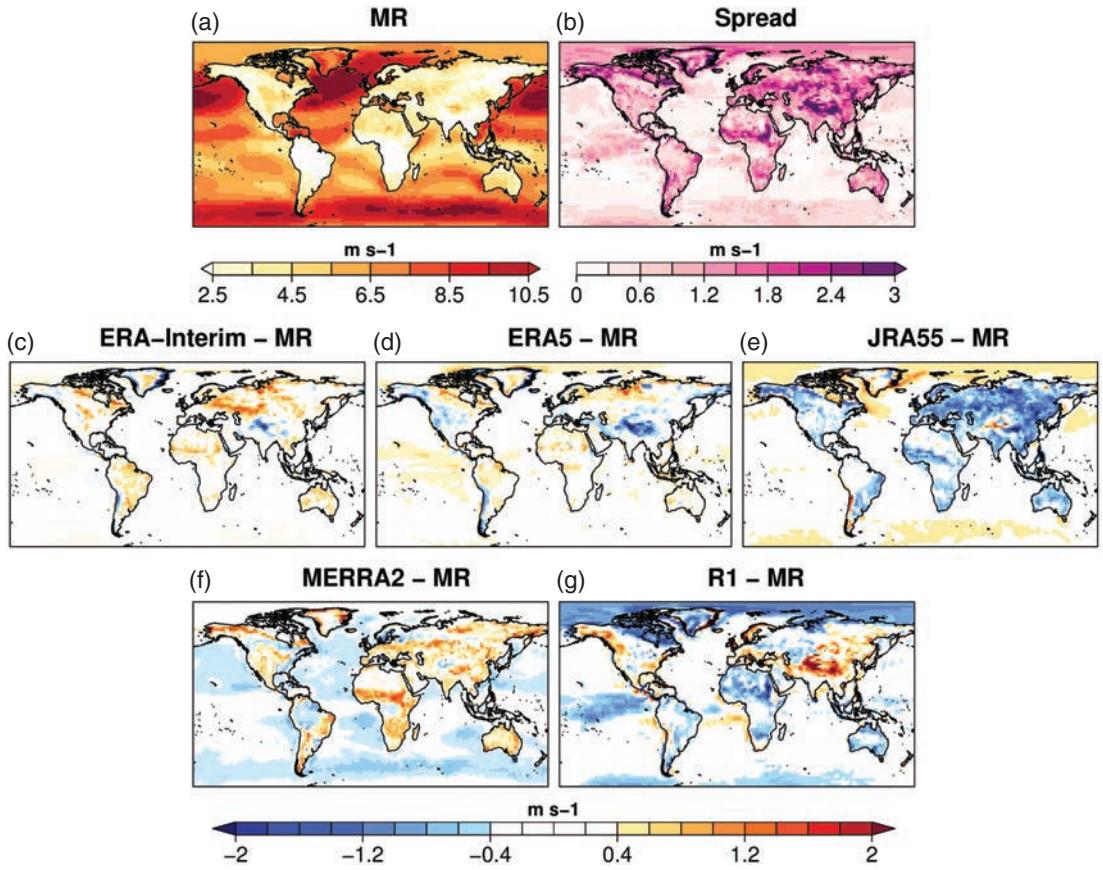


FIGURE 2 (a) Multi-reanalysis mean (MR) and (b) associated spread of seasonal surface wind speeds in DJF derived from the five reanalyses in the 1980–2017 period. Departures from MR follow for (c) ERA-Interim, (d) ERA5, (e) JRA55, (f) MERRA2, and (g) R1 reanalyses

statistical significance of the change in standard deviation has been analysed using the *F*-test (von Storch and Zwiers, 1999). Finally, reanalysis seasonal means and IAVs are compared against those computed using the observational dataset.

4 | REANALYSIS INTERCOMPARISON

A comparison of surface wind speeds from five global reanalysis datasets is presented here for DJF. Similar results are obtained for MAM, JJA, and SON, so only the comparison in the boreal summer (JJA) is included in Figures S1, S2, and S3 within the Supporting Information to avoid repetition. Agreements and differences in terms of climatology, variability, and trends at surface level are noted and described in the following subsections.

4.1 | Differences in mean wind speed

The differences between reanalyses in terms of seasonal mean wind speeds in DJF are shown in Figure 2. Although the strongest winds occur over extratropical oceans (Figure 2a) with values exceeding 10 m/s, the spread derived from the five mean wind speed fields (Figure 2b) reveals that the main differences lie inland. These discrepancies can be partly explained by differences in the land-surface roughness (which is derived from land cover and vegetation databases) and elevation representation. Indeed, the grid resolution of the reanalysis models has a strong influence on the final roughness and elevation representation. It can be noted that those products with coarser resolution (i.e., JRA55 and R1, Figure 2e and g, respectively) tend to show the highest wind speeds over high elevated mountain ranges. The JRA55 (Figure 2e) also shows a systematic negative departure pattern over continental regions, which is particularly extended over Eurasia. Over the ocean, the main differences are noticed by R1 in the eastern Pacific and at high latitudes (Figure 2g). This disparity maybe explained by the fact that R1 does not assimilate ocean surface winds from satellite observations (Fujiwara *et al.*, 2017).

4.2 | Differences in variability

The discrepancies in the year-to-year variations of the surface wind speeds in the different reanalyses have been explored through the IAV (expressed as a percentage of the mean

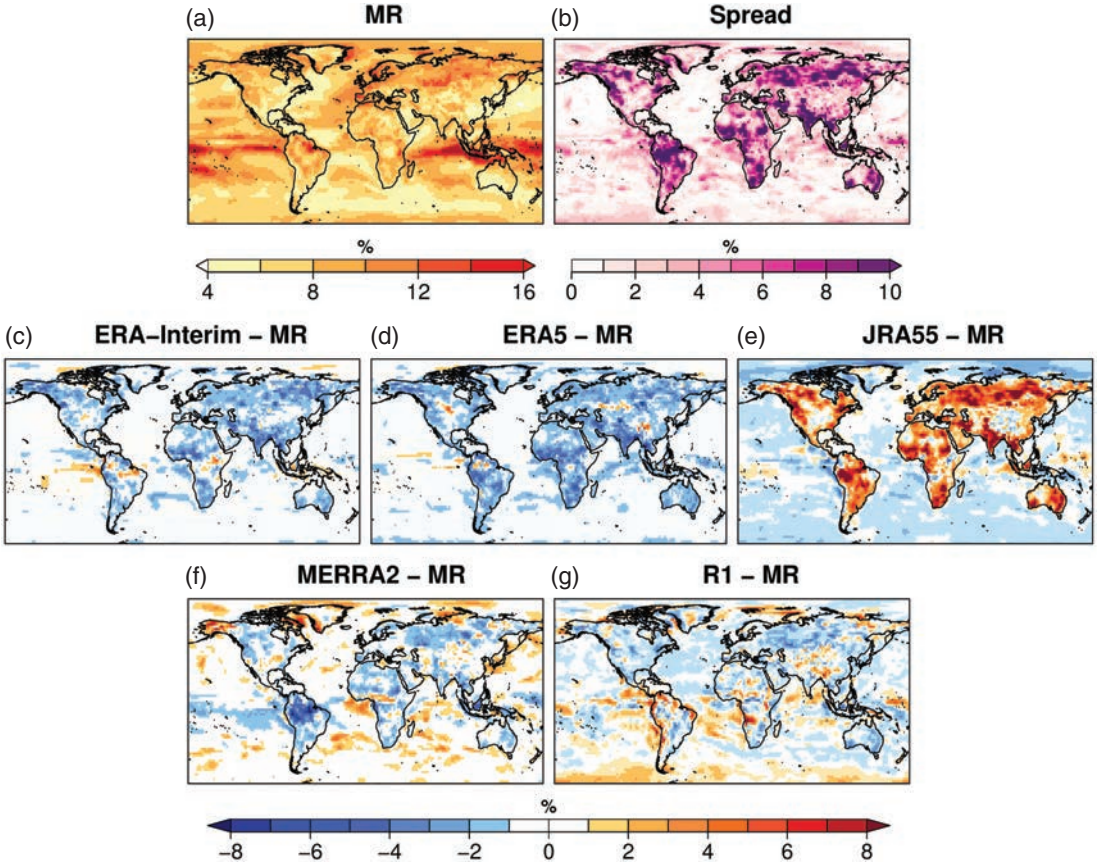


FIGURE 3 Same as in Figure 2, but for the IAV

wind speed). The MR for the IAV (Figure 3a) shows that the highest IAVs are encountered over oceanic regions along the Equator, especially in the maritime continent (Indonesia) and Pacific, where values that exceed 16% of the climatology are found. Furthermore, we also notice spots of high IAV in the Iberian and Scandinavian peninsulas, as well as northern Asia. Figure 3b informs us that the highest spread is observed inland. Indeed, the disagreements between the five reanalyses in the computation of the IAV value reach 10 percentage points, which is actually of the same order of magnitude as the IAV itself. The areas with the highest spread values often match densely vegetated regions. A detailed look at the departures of each reanalysis from the MR reveals that JRA55 (Figure 3e) concentrates the highest IAVs in the aforementioned regions and is the main contributor to the high spread. The other products offer negative departures over continental areas (Figure 3c,d,f,g), with the exception of some scattered spots of high IAV displayed by MERRA2 (Figure 3f) and R1 (Figure 3g).

4.3 | Differences in long-term trends

Linear trends are presented in Figure 4 as the rate of change of wind speed over the considered period 1980–2017 for five reanalysis products. Results for ERA-Interim (Figure 4a), JRA55 (Figure 4c), and MERRA2 (Figure 4d) are similar to those obtained by Torralba *et al.* (2017a) for the 1980–2015 period and the reader is referred there for complete information. Here, the addition of the newly released ERA5 reanalysis (Figure 4b) shows good agreement with its predecessor ERA-Interim (Figure 4a) as well as MERRA2 (Figure 4d). However, the NCEP/NCAR reanalysis (Figure 4e) shows different behaviour for the two hemispheres. For this product, decadal trends are systematically stronger in the Southern Hemisphere and the intertropical area, probably induced by the different amount of observations—which is considerably larger north of the Equator—assimilated by the R1 reanalysis in the two hemispheres (Kalnay *et al.*, 1996). In addition, since R1 does not ingest any ocean surface record from satellites, this difference between hemispheres is more appreciable.

Several causes can produce observed trends in reanalysis surface wind speeds (Thorne and Vose, 2010; Vautard *et al.*, 2010; Dee *et al.*, 2011a; Wohland *et al.*, 2019):

- decadal variations in the atmospheric circulation (e.g., changes in the location or intensity of the storm tracks),

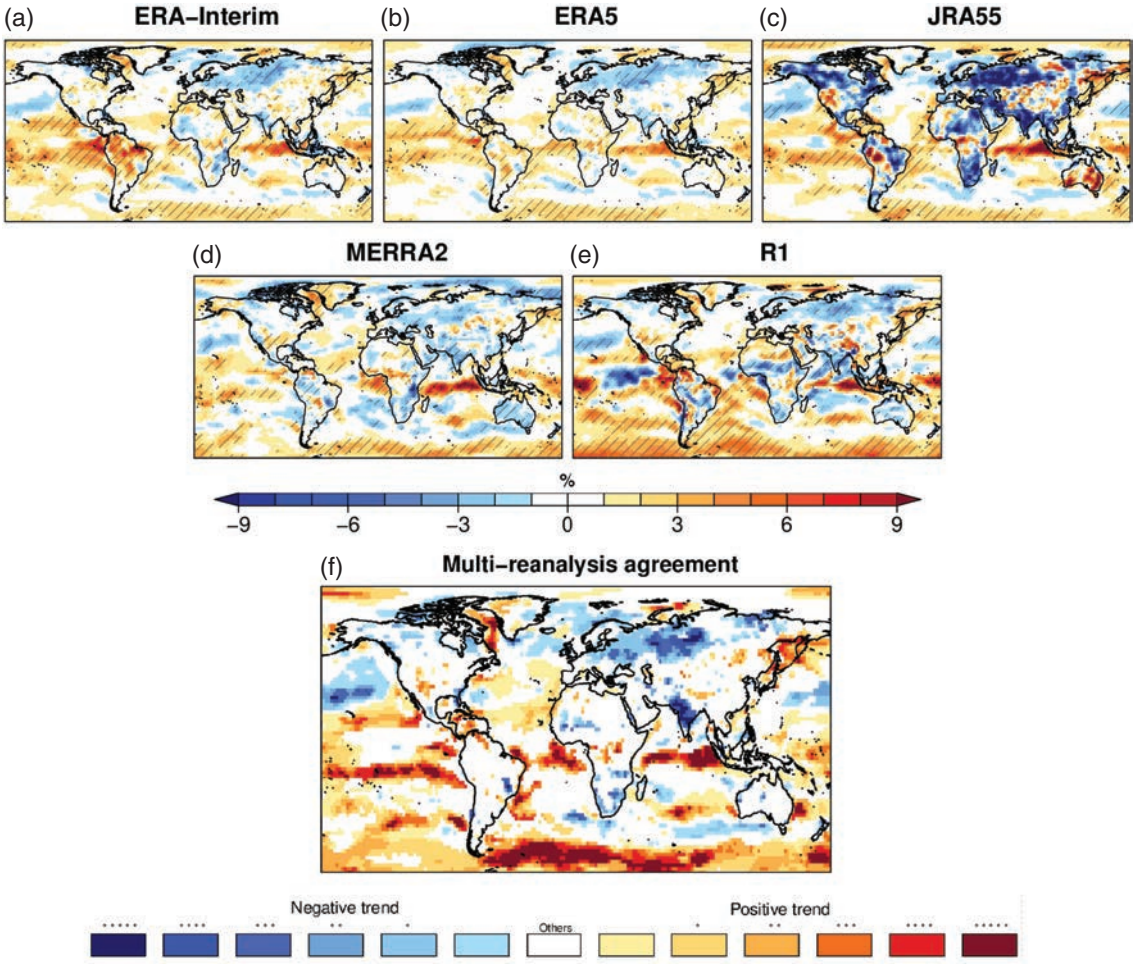


FIGURE 4 Normalized linear trend (% per decade), calculated as the linear trend of surface wind speeds divided by the seasonal mean surface wind speeds in DJF over the 1980–2017 period for (a) ERA-Interim, (b) ERA5, (c) JRA55, (d) MERRA2, and (e) R1. Hatched regions in (a)–(e) indicate where the trends are significant at the 95% confidence level. In (f) we represent an agreement map between the five reanalyses. Blues (reds) indicate agreement between the five reanalyses about the negative (positive) trends in the surface wind speed in DJF in the 1980–2017 period. An asterisk indicates that the trends are significant at the 95% confidence level: no asterisk indicates that the trends are not significant. One asterisk (∗) means that only one of the reanalyses has significant trends, two asterisks (∗∗) inform us that two reanalyses have significant trends, and so on (adapted from Torralba *et al.*, 2017a)

- changes in surface roughness due to land cover and vegetation changes,
- spurious trends in assimilated observations due to instrumental drift, measuring errors, or station relocations that produce inhomogeneities, or
- the growing number of observations available for assimilation in more recent periods.

The role of each driver, however, is still uncertain. Although separating spurious from real causes in the observed trends might be challenging, some remarks can still be made.

Firstly, changes in climate dynamics, such as the strengthening of the Walker circulation due to climate change and the subsequent reinforcement of the Hadley cell, can increase the trade winds within the intertropical area (L'Heureux *et al.*, 2013). This change in the dynamic of the tropical atmosphere is well observed in the multi-reanalysis agreement map (Figure 4f), where all five reanalyses depict a significant trend over several marine tropical areas in the Southern Hemisphere. This has been documented in previous studies, such as Wu *et al.* (2018). Similarly, a significant strengthening of wind speeds is observed in the Southern Hemisphere storm track (Figure 4f). Lee and Feldstein (2013) studied a shift of westerly winds polewards in the Southern Hemisphere and attributed this fluctuation to either an increase of greenhouse gas concentrations or a decrease of stratospheric ozone.

Secondly, surface roughness can vary substantially over vegetated regions, due to changes in the forest canopy. A statistically significant diminution of wind speeds is noticed over an extended area in Eurasia (Figure 4f), possibly mirroring the

wind-stilling effect already reported by Vautard *et al.* (2010). Vautard *et al.* (2010) attributed the stilling by alluding to an increase of the surface roughness produced partly by the growth of the forest extension in the last 30 years. A reduction of winds is observed in India too, but no particular driver has been suggested so far.

Concerning the last two items, one might expect that artefacts originated by the assimilation systems and available observations are partially compensated after considering several reanalysis products, since they employ diverging data assimilation schemes and ingest different amounts of records. For this reason, and as recommended in Wohland *et al.* (2019) for 20th century reanalyses, the use of more than one reanalysis to derive a single trends map, as in Figure 4f, can provide more robust conclusions than any single product.

5 | VERIFICATION WITH TALL TOWER OBSERVATIONS

The reanalysis intercomparison has revealed that several agreements and discrepancies do exist between these products. However, at this point it cannot be concluded which of the sources represents each surface wind speed feature better. A further comparison of the reanalysis output (hereafter referred to as *modelled* winds) against wind-speed in situ measurements (from now on referred to as *observed* winds) is needed to evaluate the uncertainty of the reanalyses in reproducing the previously described wind speed characteristics. Furthermore, this comparison can also be helpful to identify the reanalysis that represents the observed winds better within the first 100 m of the atmosphere.

In the following subsections, reanalysis surface and near-surface winds are extrapolated vertically to the selected tower measuring height (for each tower location, this is the measuring level closest to 100 m, see Table S1) and all the comparisons with tall tower data are made at that level. The set of 77 tall towers distributed worldwide is employed (see section 2.2).

5.1 | Comparison of different horizontal interpolations

The first comparison of observed and modelled daily wind speeds is presented by means of a Taylor diagram. We note that only the vertically extrapolated surface winds from the five reanalysis are utilized here. Three different interpolation methodologies (see section 3 for details) have been applied over the wind speeds from the reanalysis to produce wind speed values at the locations of the 77 tall towers. Results are presented in Figure 5. Generally, the bilinear method provides the highest correlation coefficients out of the three interpolation approaches. A maximum correlation coefficient of 0.82 is obtained for MERRA2 when

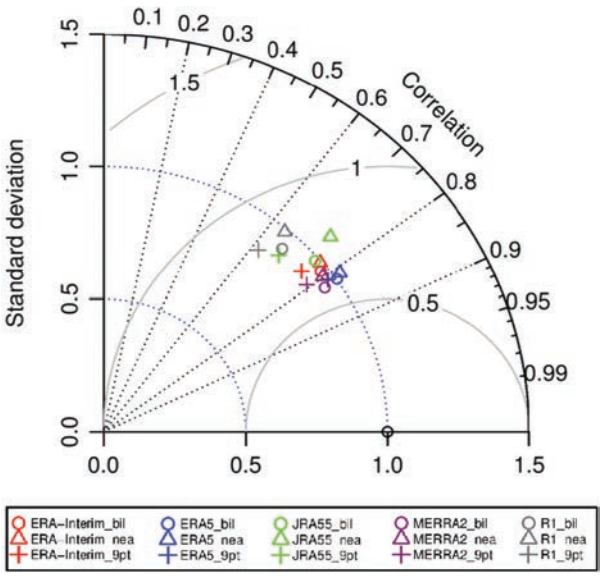


FIGURE 5 Taylor diagram of the pairwise observed and modelled daily-averaged surface wind speeds for the 77 tall towers and the five reanalyses. Three different interpolation methods are tested here: bilinear (*bil*), nearest-neighbour *(nea)* and nine-point average (*9pt*). The radial dimension represents the model standard deviations normalized by the observations. Pearson correlation coefficients are represented in angular coordinates, whereas the arcs show the CRMSEs

using this method, followed closely by ERA5 with 0.81. The lowest correlations (i.e., below 0.76) are obtained from those reanalyses with the coarsest grids, which are R1 and JRA55. The bilinear method, however, still stands as the best interpolation method in these cases. In terms of variability, both the nearest-neighbour and bilinear methods are approaches that better represent the variability of the wind speed from the tall towers. The nine-point average smooths the reanalysis winds, significantly decreasing the variability compared with the other methods. It is also worth noting that JRA55 displays a substantial overestimation of the variability for the nearest-neighbour method. This result matches the observed increase of IAV with respect to the other reanalyses in Figure 3e. The best results in terms of CRMSEs are also observed for the bilinear method; for that reason, this method has been employed for spatial interpolation of the reanalysis data to the tall tower locations for the rest of the study.

5.2 | Verification of daily-averaged surface winds

The quality of the vertically extrapolated reanalysis surface winds is assessed now in terms of correlation, standard deviation, and CRMSE. Results are presented by comparing each reanalysis product with the MR obtained from the averaging of the five reanalyses (Figure 6). Arrows within the Taylor diagrams depict the change of points from the

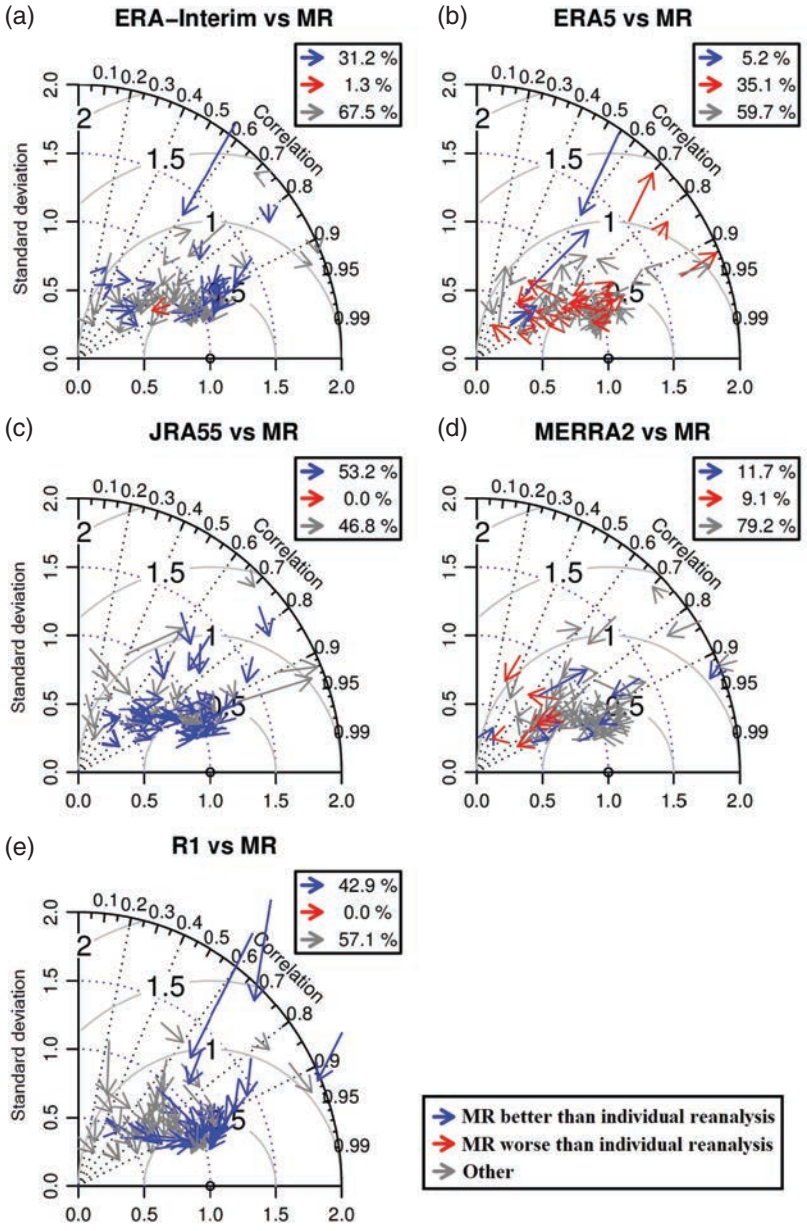


FIGURE 6 Taylor diagram of the differences in correlation, standard deviation, and CRMSE between (a) ERA-Interim, (b) ERA5, (c) JRA55, (d) MERRA2, and (e) R1, and the MR computed from all the five reanalyses. Daily-averaged modelled surface winds are vertically extrapolated and compared with the nearest 100-m measuring level winds from 77 tall towers. Blue arrows indicate that the MR provides better correlations and standard deviations closer to the observed than the reanalyses individually and that both changes are significant at the 95% confidence level. Red arrows imply that the MR improves neither the correlations nor the standard deviations, and that both changes are significant at the 95% confidence level. Grey arrows represent all other possible cases

reanalysis (beginning of the arrow) to the MR (arrowhead) produced by the difference in correlation, standard deviation, and CRMSE, compared with the in situ data set. Only ERA5 (Figure 6b), which contains 1-hr data, presents a definite improvement with respect to the MR, showing statistically significant improvements in correlation as well as standard deviation in 35.1% of the masts (i.e., a total of 27 towers). Nevertheless, the MR still improves the performance of ERA5 in 5.2% of the masts. MERRA2 (Figure 6d) also provides 1-hr data, but NASA's reanalysis does not show an overall improvement or worsening, as most of the sites (i.e., 79.2%) do not present either significant results or improvements in both correlation and standard deviation parameters. The reanalyses with 6-hr values, ERA-Interim, JRA55, and R1 (Figure 6a,c,e, respectively) show poor performance compared with the MR. Neither JRA55 nor R1 provides better

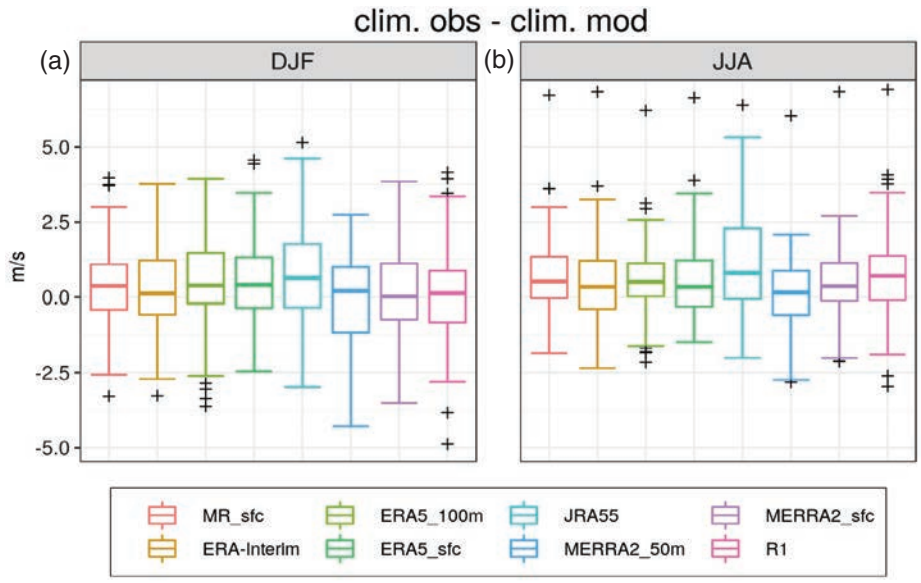


FIGURE 7 Box plots summarizing the differences between observed and modelled seasonal climatologies for 77 tall towers in (a) December–January–February and (b) June–July–August

results than the MR in any of the tall towers, and the MR presents a statistically significant improvement in half of the total tall tower set (with percentages of 53.2% and 42.9%, respectively).

5.3 | Verification of surface and near-surface winds on a seasonal time-scale

In this subsection, we assess the ERA5 and MERRA2 near-surface winds, together with surface winds from the five reanalysis products, in terms of seasonal mean winds and IAV. The MR mean, which has been computed using only surface wind fields, is also included. Both surface and near-surface winds are adjusted vertically to the closest 100-m tower heights.

Seasonal climatologies have been computed from both tall tower observations and reanalysis datasets and their differences are plotted in Figure 7 by means of box plots. In general terms, reanalysis datasets tend to show weaker seasonal mean winds than observed in both DJF and JJA, JRA55 being the dataset that provides the widest range of values, as well as the biggest underestimation, out of the five reanalyses plus the MR. This negative bias was also observed in Figure 2e extended over continental areas and is now confirmed by the observed wind data. Conversely, none of the other reanalyses stands out from the others as the best predictor of seasonal climatologies, since all of them offer quite similar results. With respect to near-surface datasets, no improvement is noticed for MERRA2 near surface winds compared with surface winds in DJF (Figure 7a), whereas, in JJA (Figure 7b), the near-surface winds tend to adjust the observed mean winds better by shifting the median of the differences to the zero value. Likewise, in the case of ERA5 it is observed that the near-surface winds reduce the spread of differences in JJA compared with the ERA5 surface winds (Figure 7b).

Analogously to Figure 7, the observed and modelled IAVs have been computed, and their differences are plotted in Figure 8. We note that the value of the IAV, as defined here, is not affected by the vertical extrapolation, as normalization by the seasonal means cancels the correction factor. As seen in Figure 8, reanalyses tend to underestimate the observed IAV, particularly in DJF (Figure 8a). This is a result somewhat expected, since grid values in a reanalysis represent an average within a cell of hundreds of square kilometres, thus smoothing the modelled variability. The wider range of values noted in DJF (Figure 8a) illustrates the complications of the reanalysis in reproducing the observed IAV in that particular season. These difficulties are related to the high variability of the wind speed in the Northern Hemisphere, where most of the towers are located. The widest spread of difference values is noticed for JRA55 in winter (Figure 8a). We note that JRA55 is the only dataset that overestimates the IAV by 5 or more percentage points in more than one tower location. These sites correspond to the Juelich, Puijo, Sodankyla, and WM09 tall towers, which are located mainly near forested areas in Europe and South Africa, which confirms the results presented in section 4.2 concerning overestimation of the variability over vegetated areas. The near-surface winds provided by ERA5 show a slight improvement in both seasons, presenting a narrower range of values oscillating around zero. In the case of NASA's reanalysis hub-height winds, only a minor enhancement of quality is found in DJF.

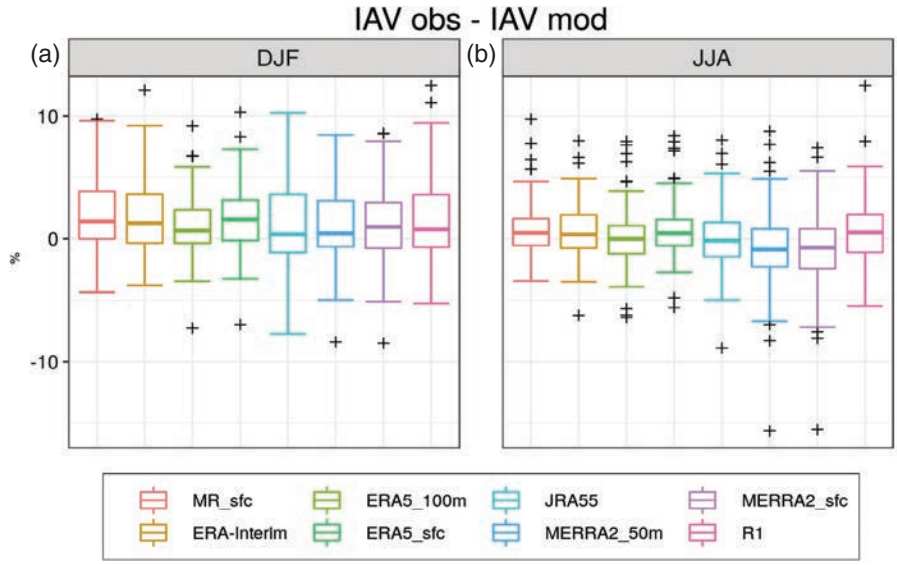


FIGURE 8 Box plots summarizing the differences between observed and modelled seasonal IAVs for 77 tall towers in (a) December–January–February and (b) June–July–August

6 | DISCUSSION AND CONCLUSIONS

Climate researchers and wind energy users often rely on reanalysis datasets for different activities such as the assessment of the wind energy resources in a particular region. However, the coexistence of several reanalysis products challenges the choice of source that characterizes the hub-height wind speed features most accurately. Indeed, disagreements in the representation of these near-surface winds have been encountered after analysing five global state-of-the-art reanalyses, demonstrating the unavoidable degree of uncertainty affecting these datasets. In this work, we describe and quantify this uncertainty in order to inform the reanalysis user about the product that describes the hub-height winds best.

Differences in DJF mean wind speed, year-to-year variability, and long-term trends have been spotted between the five global reanalyses. In particular, the most significant disagreements are encountered within continental areas. Mean wind-speed differences can be partly explained by different representations of land-surface roughness and elevation at the various grid resolutions employed in the reanalysis models. The ability of the assimilation methods to nudge fields towards observed values can also have an impact, as well as the density and quality of assimilated data or the boundary layer parametrization schemes in each model. Similarly, the main discrepancies in the IAV are seen inland. The strong IAV displayed by JRA55 is in opposition to the variability depicted by the other reanalyses. These high IAV values, as well as the systematic negative biases in the seasonal mean winds, may result from a deficiency in deriving the surface winds, as reported by Torralba *et al.* (2017a). In addition, the different amounts of available observations over time and the possible quality defects within these time series may affect the derived wind speeds. A comparison of linear trends (following Torralba *et al.*, 2017a) confirms the exceptional nature of the strong trends observed in JRA55, which are not displayed by any of the other datasets.

The multi-reanalysis agreement map reveals that, with some exceptions, significant positive and negative trends are displayed over marine areas and continental regions, respectively. All reanalyses coincide with signalling an increase of wind speeds in the Southern Hemisphere along the storm track. Taking into account the results obtained by Lee and Feldstein (2013), we point out that this strengthening of winds mirrors the shift of westerlies towards the South Pole. It is worth noting that some trends observed in only some of the reanalyses may be generated by artefacts in the assimilation process, due mainly to the abrupt evolution of the availability of observations. To overcome this handicap, we emphasize the advantages of using a multi-reanalysis approach, because most of these spurious trends can compensate each other or even cancel.

After intercomparing the set of five reanalyses, the quality of these datasets has been assessed by means of a comparison with point data measured at 77 different tall-tower sites. The interpolation of the gridded data to each tower location is carried out using the bilinear method, which has been proven to offer a slight improvement in terms of daily correlation compared with the other two approaches tested (i.e., nearest-neighbour and nine-point average). To extrapolate the surface and near-surface reanalysis data vertically to each tower measuring level, a power-law equation has been used. Even though this vertical extrapolation is very simple and

assumes neutral stability, the differences between extrapolated reanalysis winds and observed values are centred around zero. However, representativeness errors (single location versus grid area) make the distribution of errors quite large.

The verification process offered insights into which reanalyses perform better than others on daily and seasonal time-scales. Amongst all five surface wind datasets plus the MR, ERA5 shows the best results in terms of correlation and standard deviation. Indeed, the improvement in both correlation and variability with respect to the MR is statistically significant in 35% of tall tower sites, which is considerably high when compared with other reanalyses such as MERRA2 (9.1%) and ERA-Interim (1.3%). Neither JRA55 nor R1 beats the MR in any of the tower locations. Near-surface winds from ERA5 and MERRA2 have been included in some of the comparisons, particularly those made at seasonal scale. Results show a slight improvement of these variables over surface winds from the corresponding reanalyses. Notably, this enhancement is more evident for the IAV than for the seasonal means. For the latter, the high uncertainty derived from the comparison with tall tower seasonal averages leads us to discourage the use of global reanalyses to estimate mean winds. In spite of that, ERA5 outperforms all other reanalysis datasets plus the MR.

All in all, we conclude that the newly produced ERA5 near-surface wind dataset offers the best estimates of mean wind speed and variability at turbine hub heights. This is actually a crucial result, especially if we take into account that ERA5 is a reanalysis that will be updated operationally, providing useful climate information in near-real time that can be easily integrated into different decision-making processes. Moreover, many wind energy applications (e.g., the study of sudden wind changes that may lead to sudden energy ramps) can also benefit from the 1-hr data available from ERA5. Nevertheless, when it comes to the analysis of trends, we recommend a multi-reanalysis approach instead of using one single product.

Unfortunately, it has been difficult to ensure the validity of the linear trends in the five reanalyses by comparison with observations at the tall tower sites, since long and homogeneous records are not generally available. We intended to filter those tall towers within the Tall Tower Dataset, containing data spanning more than 20 years and without homogeneity issues. The experiment left only four tall towers that fulfilled this condition (Barrow, Cabauw, Mauna Loa, and National Wind Technology Center Mast 2), so that any conclusion can be derived from this exercise, due to the reduced sample of masts available for calculations. A more comprehensive description of the experiment design, as well as verification with the set of four masts, is detailed in section S3 and Figure S4 in the Supporting Information.

Although reanalyses are intended to cover the absence of high-quality and long observational records, in situ data are still needed, not only to verify climate products, but also for other applications such as model configuration. For instance, in the creation of the New European Wind Atlas (NEWA, 2019), only eight meteorological masts placed in northern Europe were used to decide the optimum configuration of the Weather Research and Forecasting (WRF) model employed in the generation of the wind atlas (Witha *et al.*, 2019), even though the domain covers the whole of Europe. In this regard, efforts should be made to facilitate the availability of hub-height wind measurements. New initiatives, such as the Tall Tower Dataset (Ramon and Lledó, 2019) within the context of the European project Integrated Approach for the development across Europe of user oriented climate indicators for GFCS high priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism, are starting to appear to promote the usage of these data. Future work will be devoted to the study of the possible improvements that a vertical extrapolation that considers the stability of the planetary boundary layer could introduce, and their implications for the impact models currently employed in the wind energy sector to perform wind power estimations.

DATA AVAILABILITY STATEMENT

Data from 181 out of the 222 tall towers, within the Tall Tower Dataset, can be publicly accessed in the data repository EUDAT at https://doi.org/10.23728/b2share.0d3a99db75df4238820ee548f35ee36b. Please refer to Ramon and Lledó (2019) for complete information on the data collection, formatting, and quality control.

ACKNOWLEDGMENTS

The authors acknowledge funding support from Integrated Approach for the development across Europe of user oriented climate indicators for GFCS high priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism (INDECIS) cofounded by the H2020 ERA-net ERA4CS (GA 690462), the S2S4E (GA 776787) project funded by the Horizon 2020 framework of the European Commission, and the MICINN grant BES-2017-082216 ("Ayudas para contratos predoctorales"). Acknowledgments are due to each of the four reanalysis production centres ECMWF, NASA, JMA, and NCEP/NCAR for facilitating access to the reanalysis data. We also thank the developers of the R package *s2dverification* (Manubens *et al.*, 2018). Very special thanks are given to Pierre-Antoine Bretonnière and Margarida Samsó for their valuable work. Authors aknowledge Dr. Karin van der Wiel, and two anonymous reviewers for their useful inputs.

We acknowledge all the principal investigators of the 77 tall towers employed in this study for sharing their valuable wind data. We thank the National Data Buoy Center

(NDBC), the Earth System Research Laboratory (ESRL), and the Physical Monitoring Program of the Smithsonian Tropical Research Institute (STRI). The Berms Aspen data were provided by NCAR/EOL under the sponsorship of the National Science Foundation. Thanks to the Bonneville Power Administration (BPA), the CESAR observatory, and Dr Fred Bosveld (KNMI) for providing Cabauw data, the South Africa Weather Service (SAWS), the Wind Atlas for South Africa project (WASA), and the United Kingdom Met Office (UKMO). Many thanks to Lena Kozlova (University of Exeter) for sharing Cape Verde tall-tower data. Acknowledgements are due to the Crown State, the BMWi (Bundesministerium fuer Wirtschaft und Energie, Federal Ministry for Economic Affairs and Energy), and the PTJ (Projekttraeger Juelich, project executing organization), the data contributors to the AsiaFlux database, Dr Ingo Lange (Hamburg University), Dr Laszlo Haszpra (Hungarian Met Service), Dr Jan Schween (University of Cologne), and Dr Frank Beyrich (Deutscher Wetterdienst) for facilitating Hamburg University, Hegyhatsal, Juelich and Lindenberg masts data, respectively. We also aknowledge Dr Rolf Neubert (University of Groningen), Met Éireann, the NREL National Wind Technology Center (NWTC: Jager and Andreas, 1996), Dr Anna Rutgersson (Uppsala University), Professor Gil Bohrer (The Ohio State University), the NoordzeeWind B.V. (NZWBV) and/or its (sub)contractors. Park Falls tower data (Davis *et al.*, 2003) were provided by AmeriFlux, for which funding was provided by the U.S. Department of Energy's Office of Science. Thanks to the FLUXNET community, who provided valuable data from different masts, and also to Christy Schultz (GMD Met–NOAA) for sharing South Pole mast data. Credit is also given to the original sources/s of the Tumbarumba Met Mast data, the Vielsalm data manager Anne de Ligne, and the data provider Tanguy Manise.

ORCID

Jaume Ramon http://orcid.org/0000-0003-2818-5206

REFERENCES

Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A. and Speranza, A. (2003) Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Weather and Forecasting*, 18, 918–932. https://doi.org/10.1175/1520-0434(2003)018<0918:sopfss>2.0.co;2.

Alvarez, I., Gomez-Gesteira, M., DeCastro, M. and Carvalho, D. (2014) Comparison of different wind products and buoy wind data with seasonality and interannual climate variability in the southern Bay of Biscay (2000–2009). *Deep-Sea Research Part II: Topical Studies in Oceanography*, 106, 38–48. https://doi.org/10.1016/j.dsr2.2013.09.028.

Brower, M.C., Barton, M.S., Lledó, L. and Dubois, J. (2013) *A study of wind speed variability using global reanalysis data*. AWS True Power. Technical Report. Available at: https://www.awstruepower.com/assets/A-Study-of-Wind-Speed-Variability-Using-Global-Reanalysis-Data.pdf [Accessed 1st August 2019].

Businger, J.A., Wyngaard, J.C., Izumi, Y. and Bradley, E.F. (1971) Flux-profile relationships in the atmospheric surface layer. *Journal of the Atmospheric Sciences*, 28, 181–189. https://doi.org/10.1175/1520-0469(1971)028<0181:fprita>2.0.co;2.

Cannon, D., Brayshaw, D., Methven, J., Coker, P. and Lenaghan, D. (2015) Using reanalysis data to quantify extreme wind-power generation statistics: a 33 year case study in Great Britain. *Renewable Energy*, 75, 767–778. https://doi.org/10.1016/j.renene.2014.10.024.

Clark, R.T., Bett, P.E., Thornton, H.E. and Scaife, A.A. (2017) Skilful seasonal predictions for the European energy industry. *Environmental Research Letters*, 12, 119602. https://doi.org/10.1088/1748-9326/aa94a7.

Copernicus Climate Change Service (C3S) (2017) *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS)*. ECMWF. Available at: https://cds.climate.copernicus.eu/cdsapp#!/home [Accessed 1st August 2019].

Davis, K., Bakwin, P.S., Yi, C., Berger, B.W., Zhaos, C., Teclaw, R.M. and Isebrands, J.G. (2003) The annual cycles of CO2 and H2O exchange over a northern mixed forest as observed from a very tall tower. *Global Change Biology*, 1278–1293. https://doi.org/10.1029/2009JD012832.

Decker, M., Brunke, M.A., Wang, Z., Sakaguchi, K., Zeng, X. and Bosilovich, M.G. (2012) Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal of Climate*, 25, 1916–1944. https://doi.org/10.1175/JCLI-D-11-00004.1.

Dee, D.P., Kallen, E., Simmons, A.J. and Haimberger, L. (2011a) Comments on "Reanalyses suitable for characterizing Long-term trends". *Bulletin of the American Meteorological Society*, 92, 65–70. https://doi.org/10.1175/2010bams3070.1.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G. and Bauer, P. (2011b) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. others. https://doi.org/10.1002/qj.828.

Doblas-Reyes, F.J., García-Serrano, J., Lienert, F., Biescas, A.P. and Rodrigues, L.R. (2013) Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4, 245–268. https://doi.org/10.1002/wcc.217.

Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T. and Zwiers, F.W. (2014) Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27, 5019–5035. https://doi.org/10.1175/JCLI-D-13-00405.1.

Fujiwara, M., Wright, J.S., Manney, G.L., Gray, L.J., Anstey, J., Birner, T., Davis, S., Gerber, E.P., Lynn Harvey, V., Hegglin, M.I., Homeyer, C.R., Knox, J.A., Krüger, K., Lambert, A., Long, C.S., Martineau, P., Molod, A., Monge-Sanz, B.M., Santee, M.L., Tegtmeier, S., Chabrillat, S., Tan, D.G., Jackson, D.R., Polavarapu, S., Compo, G.P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker,

J.S. and Zou, C.Z. (2017) Introduction to the SPARC reanalysis Intercomparison project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, 17, 1417–1452. https://doi.org/10.5194/acp-17-1417-2017.

Gleeson, E., Whelan, E. and Hanley, J. (2017) Met Éireann high resolution reanalysis for Ireland. *Advances in Science and Research*, 14, 49–61. https://doi.org/10.5194/asr-14-49-2017.

Gregow, H., Jylhä, K., Mäkelä, H.M., Aalto, J., Manninen, T., Karlsson, P., Kaiser-Weiss, A.K., Kaspar, F., Tan, D.G.H., Obregon, A. and Su, Z. (2015) Worldwide survey of awareness and needs concerning reanalyses and respondents views on climate services. *Bulletin of the American Meteorological Society*, 97, 1461–1473. https://doi.org/10.1175/bams-d-14-00271.1.

Helfand, H.M. and Schubert, S.D. (1995) Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *Journal of Climate*, 8, 784–806.

Hsu, S.A., Meindl, E.A. and Gilhousen, D.B. (1994) Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. *Journal of Applied Meteorology*, 33, 757–765. https://doi.org/10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2.

Jager, D. and Andreas, A. (1996) *NREL National Wind Technology Center (NWTC): M2 tower; Boulder, Colorado (data); NREL report no. DA-5500-56489*. National Renewable Energy Laboratory. Technical Report. Available at: https://doi.org/10.5439/1052222 [Accessed 1st August 2019]

Jolliffe, I.T. and Stephenson, D.B. (2012) *Forecast Verification*. Oxford, UK: Wiley.

Jude, M. and Leseney, H. (2017) *Influence of the choice of long term sources in France Modelled data vs measured data.* Wind Europe Resource Assessment Workshop, Edinburgh, United Kingdom. Available at: http://www.eoltech.fr/doc/Full%20paper%20-%20Workshop%20Wind%20Europe%20March17-%20Eoltech.pdf [Accessed 1st August 2019].

Kaiser-Weiss, A.K., Kaspar, F., Heene, V., Borsche, M., Tan, D.G.H., Poli, P., Obregon, A. and Gregow, H. (2015) Comparison of regional and global reanalysis near-surface winds with station observations over Germany. *Advances in Science and Research*, 12, 187–198. https://doi.org/10.5194/asr-12-187-2015.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M. (1989) Description of the NMC global data assimilation and forecast system. *Weather and Forecasting*, 4, 335–342. https://doi.org/10.1175/1520-0434(1989)004<0335:dotngd>2.0.co;2.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.K., Hnilo, J.J., Fiorino, M. and Potter, G.L. (2002) NCEP-DOE AMIP-II Renalalysys (R-2). *Bulletin of the American Meteorological Society*, 83, 1631–1643. https://doi.org/10.1175/BAMS-83-11.

Kelly, M. (2016) *Uncertainty in Vertical Extrapolation of Wind Statistics: Shear-Exponent and WAsP/EWA Methods*. DTU Wind Energy. Technical Report. https://orbit.dtu.dk/ws/files/126254660/VertExtrapUncert_Kelly2016_DTU_en_.pdf [Accessed 1st August 2019].

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan, Ser. II*, 93, 5–48. https://doi.org/10.2151/jmsj.2015-001.

Kouwenhoven, H. (2007) *User manual data files meteorological mast NoordzeeWind*. Document code: NZW-16-S-4-R03. 1, Noordzee Wind, pp. 1–8

Kumar, A. and Hu, Z.Z. (2012) Uncertainty in the ocean–atmosphere feedbacks associated with ENSO in the reanalysis products. *Climate Dynamics*, 39, 575–588. https://doi.org/10.1007/s00382-011-1104-3.

Lee, S. and Feldstein, S.B. (2013) Detecting ozone- and greenhouse gas-driven wind trends with observational data. *Science*, 339, 563–567. https://doi.org/10.1126/science.1225154.

L'Heureux, M.L., Lee, S. and Lyon, B. (2013) Recent multidecadal strengthening of the Walker circulation across the tropical Pacific. *Nature Climate Change*, 3, 571–576. https://doi.org/10.1038/nclimate1840.

Lledó, L., Bellprat, O., Doblas-Reyes, F.J. and Soret, A. (2018) Investigating the effects of Pacific sea surface temperatures on the wind drought of 2015 over the United States. *Journal of Geophysical Research: Atmospheres*, 1–13. https://doi.org/10.1029/2017JD028019.

Lucio-Eceiza, E.E., González-Rouco, J.F., García-Bustamante, E., Navarro, J. and Beltrami, H. (2018) Multidecadal to centennial surface wintertime wind variability over northeastern North America via statistical downscaling. *Climate Dynamics*, 53, 41–66. https://doi.org/10.1007/s00382-018-4569-5.

Manubens, N., Caron, L.P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N.S., Garcia-Serrano, J., Massonnet, F., Ménégoz, M., Sicardi, V., Batté, L., Prodhomme, C., Torralba, V., Cortesi, N., Mula-Valls, O., Serradell, K., Guemas, V. and Doblas-Reyes, F.J. (2018) An R package for climate forecast verification. *Environmental Modelling and Software*, 103, 29–42. https://doi.org/10.1016/j.envsoft.2018.01.018.

Mentis, D., Hermann, S., Howells, M., Welsch, M. and Siyal, S.H. (2015) Assessing the technical wind energy potential in Africa a GIS-based approach. *Renewable Energy*, 83, 110–125. https://doi.org/10.1016/j.renene.2015.03.072.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D. and Shi, W. (2006) North American regional reanalysis. *Bulletin of the American Meteorological Society*, 87, 343–360. https://doi.org/10.1175/BAMS-87-3-343.

Molod, A., Takacs, L., Suarez, M. and Bacmeister, J. (2015) Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development*, 8, 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015.

NEWA (2019) *New European Wind Atlas*. Available at: http://www.neweuropeanwindatlas.eu/ [Accessed 1st August 2019].

Olauson, J. (2018) ERA5: the new champion of wind power modelling?. *Renewable Energy*, 126, 322–331. https://doi.org/10.1016/j.renene.2018.03.056.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2006) *Numerical Recipes in C: The Art of Scientific Computing*, Vol. 62. Cambridge, UK: Cambridge University Press. Available at: https://doi.org/10.2307/2153422 [Accessed 1st August 2019].

Pryor, S.C., Barthelmie, R.J. and Schoof, J.T. (2006) Inter-annual variability of wind indices across Europe. *Wind Energy*, 9, 27–38. https://doi.org/10.1002/we.178.

Ramon, J. and Lledó, L. (2019) *The Tall Tower Dataset*. Technical Note. Barcelona: Barcelona Supercomputing Center–Centro Nacional de Supercomputación. Technical Report: 1. https://earth.bsc.es/wiki/lib/exe/fetch.php?media=library:external:technical_memoranda:technical_report_talltower_database_v2.pdf [Accessed 1st August 2019].

Rehman Tahir, Z.U., Sarfraz, M.S., Asim, M., Kamran, M.S., Imran, S. and Hayat, N. (2018) Evaluation of ERA-Interim and NCEP-CFSR reanalysis datasets against in situ measured wind speed data for Keti Bandar port, Pakistan. *Journal of Physics: Conference Series*, 1102, 012001. https://doi.org/10.1088/1742-6596/1102/1/012001.

Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.T., Chuang, H.Y., Juang, H.M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G. and Goldberg, M. (2010) The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91, 1015–1057. arXiv: 9809069v1. https://doi.org/10.1175/2010BAMS3001.1.

Sharp, E., Dodds, P., Barrett, M. and Spataru, C. (2015) Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information. *Renewable Energy*, 77, 527–538. https://doi.org/10.1016/j.renene.2014.12.025.

Stohl, A., Wotawa, G., Seibert, P. and Kromp-Kolb, H. (2002) Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories. *Journal of Applied Meteorology*, 34, 2149–2165. https://doi.org/10.1175/1520-0450(1995)034<2149:ieiwfa>2.0.co;2.

von Storch, H. and Zwiers, F.W. (1999) *Statistical Analysis in Climate Research*. Cambridge, UK: Cambridge University Press. https://doi.org/10.2307/2669798.

Stull, R.B. (2012) *An Introduction to Boundary Layer Meteorology*. Dordrecht, the Netherlands: Springer Science and Business Media.

Tammelin, B., Vihma, T., Atlaskin, E., Badger, J., Fortelius, C., Gregow, H., Horttanainen, M., Hyvönen, R., Kilpinen, J., Latikka, J., Ljungberg, K., Mortensen, N.G., Niemelä, S., Ruosteenoja, K., Salonen, K., Suomi, I. and Venäläinen, A. (2013) Production of the Finnish wind atlas. *Wind Energy*, 16, 19–35. https://doi.org/10.1002/we.517.

Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106, 7183–7192.

Thorne, P.W. and Vose, R.S. (2010) Reanalyses suitable for characterizing long-term trends. *Bulletin of the American Meteorological Society*, 91, 353–361. https://doi.org/10.1175/2009BAMS2858.1.

Torralba, V., Doblas-Reyes, F.J. and Gonzalez-Reviriego, N. (2017a) Uncertainty in recent near-surface wind speed trends: a global reanalysis intercomparison. *Environmental Research Letters*, 12, 114019. https://doi.org/10.1088/1748-9326/aa8a58.

Torralba, V., Doblas-Reyes, F.J., MacLeod, D., Christel, I. and Davis, M. (2017b) Seasonal climate prediction: A new source of information for the management of wind-energy resources. *Journal of Applied Meteorology and Climatology*, 56, 1231-1247. https://doi.org/10.1175/JAMC-D-16-0204.1.

Touma, J.S. (1977) Dependence of the wind profile power law on stability for various locations. *Journal of the Air Pollution Control Association*, 27, 863–866. https://doi.org/10.1080/00022470.1977.10470503.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthélemy, A., Bricaud, C., Carton, J., Fučkar, N., Garric, G., Iovino, D., Kauker, F., Korhonen, M., Lien, V.S., Marnela, M., Massonnet, F., Mignac, D., Peterson, K.A., Sadikni, R., Shi, L., Tietsche, S., Toyoda, T., Xie, J. and Zhang, Z. (2018) An assessment of ten ocean reanalyses in the polar regions. *Climate Dynamics*, 52, 1–38. https://doi.org/10.1007/s00382-018-4242-z.

Van Ulden, A.P. and Wieringa, J. (1996) Atmospheric boundary layer research at Cabauw. *Boundary-Layer Meteorology*, 78, 39–69. https://doi.org/10.1007/BF00122486.

Vautard, R., Cattiaux, J., Yiou, P., Thépaut, J.N. and Ciais, P. (2010) Northern hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience*, 3, 756–761. https://doi.org/10.1038/ngeo979.

Williams, E.J. (1959) *Regression Analysis*. New York, NY: Wiley.

Witha, B., Hahmann, A., Sīle, T., Barcons, J., Dörenkämper, M., Ezber, Y., Garcia, E., González-rouco, F., Leroy, G., Olsen, B.T., Sastre, M. and Söderberg, S. (2019) *The NEWA Mesoscale Wind Atlas: production and ensemble runs*. Zenodo. https://doi.org/10.5281/zenodo.2635548.

Wohland, J., Omrani, N.E., Witthaut, D. and Keenlyside, N.S. (2019) Inconsistent wind speed trends in current twentieth century reanalyses. *Journal of Geophysical Research: Atmospheres*, 124, 1931–1940. https://doi.org/10.1029/2018JD030083.

Wu, J., Zha, J., Zhao, D. and Yang, Q. (2018). Changes in Terrestrial near-Surface Wind Speed and their Possible Causes: An Overview. *Climate dynamics*, 51, 2039–2078.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ramon J, Lledó L, Torralba V, Soret A, Doblas-Reyes FJ. What global reanalysis best represents near-surface winds?. *Q J R Meteorol Soc*. 2019;145:3236–3251. https://doi.org/10.1002/qj.3616

Supporting information

S1. List of tall towers employed in this study

Table S1: List of tall towers employed in this study. A season is defined complete if each of the constituent months contain at least 50% of non-missing data.

Name	Longitude	Latitude	Offshore	Country[a]	Complete seasons (DJF/MAM/JJA/SON)	Measuring height (m)
42361	-92.490	27.550	yes	US	4/4/6/4	122
42887	-88.496	28.191	yes	US	4/4/4/5	48
BAO	-105.004	40.050	no	US	6/6/5/6	100
Barrow	-156.611	71.323	no	US	27/29/28/28	18
BCI	-79.850	9.167	no	PA	15/15/15/14	48
Berms Aspen	-106.200	53.629	no	CA	7/7/7/7	38
Braschaat	4.520	51.308	no	BE	10/16/17/16	41
burl1	-89.428	28.905	no	US	29/29/28/29	38
Butler Grade	-118.683	45.950	no	US	12/14/13/15	62
bygl1	-90.420	29.789	yes	US	9/11/11/11	31
Cabauw	4.926	51.971	no	NL	26/27/27/27	80
Cape Point	18.480	-34.350	no	ZA	5/7/7/7	30
Cardington	-0.417	52.100	no	GB	8/8/8/9	50
Chinook	-119.534	45.833	no	US	10/11/10/10	50
chlv2	-75.713	36.905	no	US	32/31/26/27	43
CVO	-24.868	16.850	no	CV	4/5/4/4	30
desw1	-124.485	47.675	no	US	30/30/32/33	31
Docking Shoal	0.648	53.158	yes	GB	3/3/3/3	90
Fino 1	6.588	54.015	yes	DE	12/14/14/12	100
Fino 2	13.154	55.007	yes	DE	10/8/10/11	102
Fino 3	7.158	55.195	yes	DE	6/8/8/7	100
fmoa1	-88.024	30.228	yes	US	8/8/8/8	36
Fuji Hokuroku	138.765	35.444	no	JP	3/3/4/4	35
fwyf1	-80.097	25.591	no	US	21/21/21/23	44
Goodnoe Hills	-120.550	45.783	no	US	14/16/14/16	59
Greater Gabbard MMZ	1.922	51.944	yes	GB	9/6/6/8	82
Gunfleet Sands	1.197	51.726	yes	GB	5/5/6/5	61
Hamburg University	10.103	53.519	no	DE	13/13/14/14	110
Hegyhatsal	16.652	46.956	no	HU	18/16/14/12	115
Ijmuiden	3.436	52.848	yes	NL	5/4/4/4	90
Inner Dowsing	0.436	53.127	yes	GB	7/7/8/9	43
Juelich	6.220	50.928	no	DE	5/5/6/6	100
Kennewick	-119.117	46.100	no	US	5/6/4/6	37

[a]Using the ISO ALPHA-2 Country codes

Table S1: Continued.

Name	Longitude	Latitude	Offshore	Country[a]	Complete seasons (DJF/MAM/JJA/SON)	Measuring height (m)
Lindenberg	14.123	52.166	no	DE	17/18/18/18	98
London Array	1.386	51.594	yes	GB	5/6/6/6	82
lopl1	-90.025	28.885	yes	US	5/4/5/6	58
Lutjewad	6.353	53.404	no	NL	14/15/12/15	60
Malin Head	-7.332	55.353	no	IE	29/30/29/30	22
Mauna Loa	-155.576	19.536	no	US	24/26/23/24	40
Megler	-123.877	46.266	no	US	7/8/8/8	53
Naselle Ridge	-123.797	46.422	no	US	7/8/8/8	30
NWTC M2	-105.235	39.911	no	US	21/20/20/21	80
NWTC M4	-105.225	39.906	no	US	3/3/3/3	100
Obninsk	36.598	55.111	no	RU	8/9/9/9	121
Oestergarnsholm	18.983	57.433	no	SE	3/4/5/6	29
OSU	-84.714	45.560	no	US	10/11/10/10	46
OWEZ	4.390	52.606	yes	NL	5/4/4/6	116
Palangkaraya	114.036	2.345	no	ID	3/4/4/4	42
Park Falls	-90.273	45.945	no	US	16/18/13/18	122
Pasoh	102.300	2.967	no	MY	6/7/7/7	53
Puijo	27.653	62.906	no	FI	9/10/8/10	75
Race Bank	0.748	53.314	yes	GB	5/5/6/6	89
roam4	-89.313	47.867	no	US	26/26/28/28	47
Samoa	-170.564	-14.247	no	AS	20/18/19/21	21
Seven Mile	-121.267	45.634	no	US	14/16/14/15	30
sgof1	-84.858	29.408	no	US	11/12/9/11	35
Sodankyla	26.638	67.362	no	FI	11/12/13/12	24
South Pole	- 24.800	-89.980	no	US	9/9/9/9	50
stdm4	-87.225	47.184	no	US	30/28/29/33	35
Summit	-38.480	72.580	no	GL	8/8/6/8	50
Trinidad	-124.151	41.054	no	US	14/14/13/14	20
Troutdale	-122.402	45.558	no	US	6/8/8/8	30
Tumbarumba	148.152	-35.657	no	AU	13/14/14/14	70
upbc1	-122.121	38.038	yes	US	3/4/4/4	100
Vielsalm	5.998	50.305	no	BE	4/4/6/5	52
Walnut Grove	-121.491	38.265	no	US	4/5/5/5	122
Wasco	-120.767	45.500	no	US	11/12/12/13	30
WM01	16.664	-28.602	no	ZA	6/6/6/7	62
WM02	19.361	-31.525	no	ZA	5/6/6/7	62
WM03	18.42	-31.731	no	ZA	6/6/6/7	62
WM04	18.109	-32.846	no	ZA	3/3/3/3	62
WM05	19.692	-34.612	no	ZA	6/5/6/6	62
WM06	20.691	-32.557	no	ZA	6/6/6/6	62
WM07	22.557	-32.967	no	ZA	6/5/7/7	62
WM08	24.514	-34.11	no	ZA	5/5/6/7	62
WM09	25.028	-31.253	no	ZA	4/6/5/6	62
WM10	28.136	-32.091	no	ZA	5/4/5/7	62

[a]Using the ISO ALPHA-2 Country codes

S2. Reanalysis intercomparison for JJA

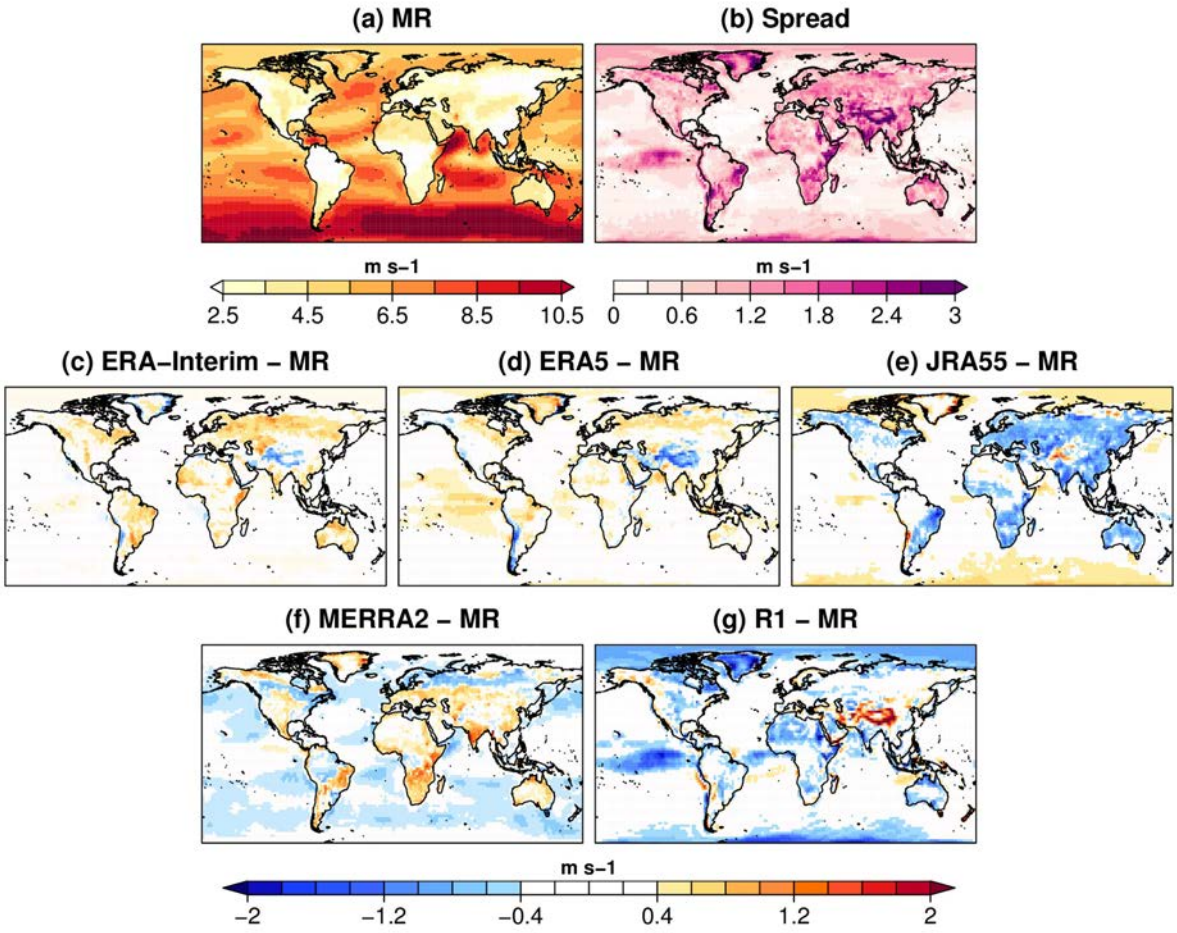


Fig. S1: (a) Multi-reanalysis mean (MR) and (b) associated spread of seasonal surface wind speeds in JJA derived from the five reanalyses in the 1980-2017 period. Departures from MR follow for (c) ERA-Interim, (d) ERA5, (e) JRA55, (f) MERRA2 and (g) R1 reanalyses.

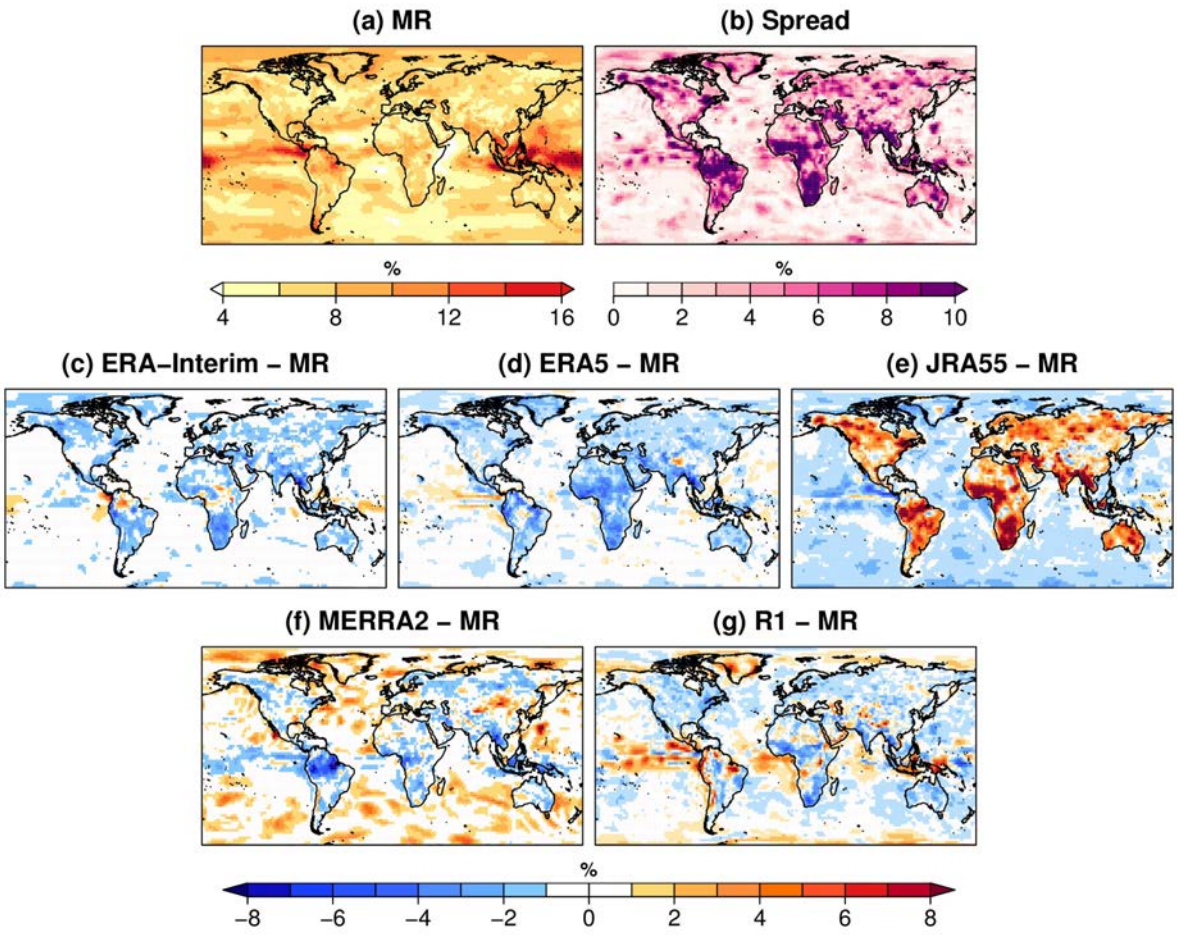


Fig. S2: Same as in Fig. S1 but for the IAV.

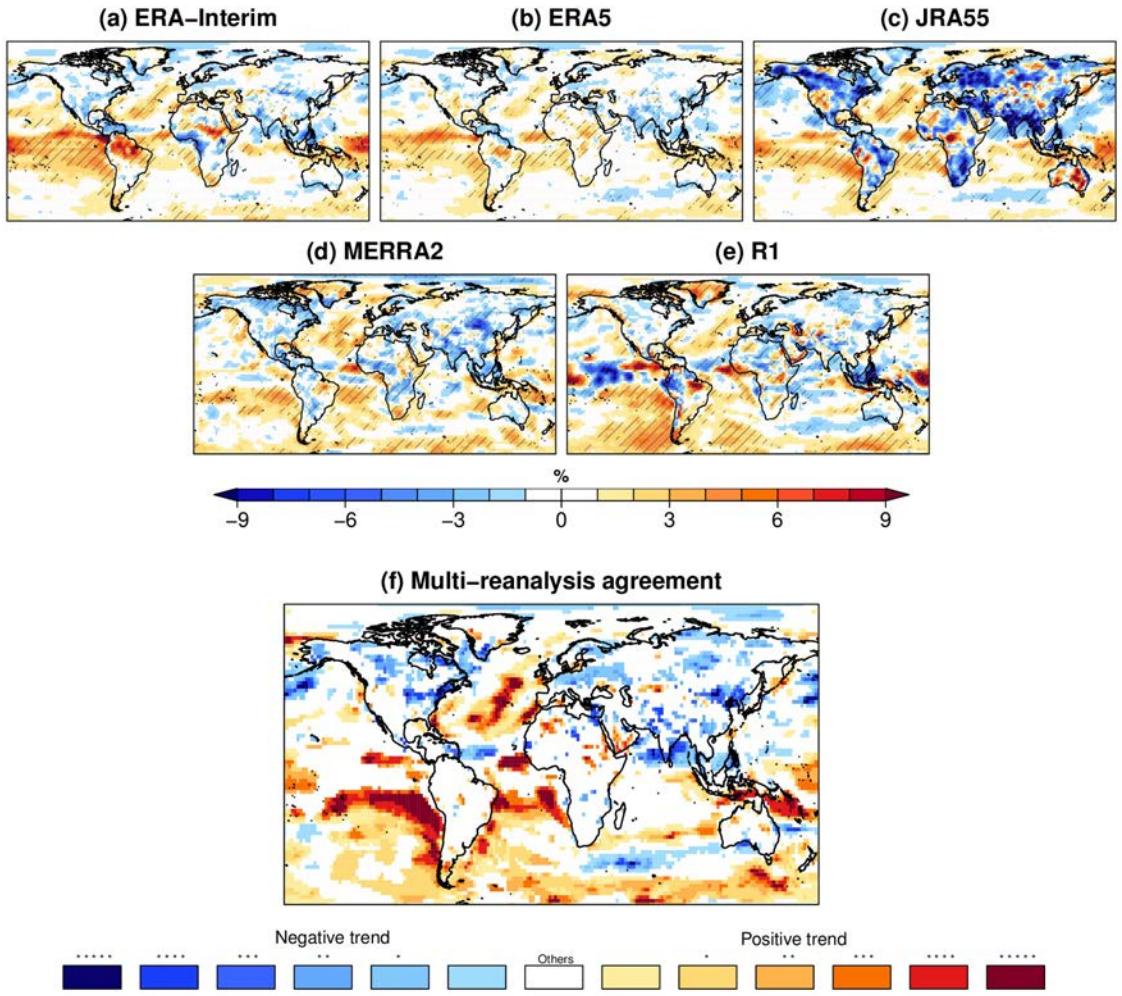


Fig. S3: Normalized linear trend (% per decade) calculated as the linear trend of surface wind speeds divided by the seasonal mean surface wind speeds in JJA over the 1980-2017 period for (a) ERA-Interim, (b) ERA5, (c) JRA55, (d) MERRA2 and (e) R1. Hatched regions in (a)-(e) indicate where the trends are significant at the 95% of confidence level. In (f) we represent an agreement map between the five reanalyses. Blues (Reds) indicate agreement between the five reanalyses about the negative (positive) trends in the 10 m wind speed in DJF in the 1980-2017 period. Asterisk indicates that the trends are significant at the 95% confidence level: no asterisk indicates that the trends are not significant. One asterisk (∗) means that only one of the reanalysis has significant trends, two asterisks (∗∗) informs that two reanalyses have significant trends, and so on (adapted from Torralba et al. (2017)).

S3. Verification of observed trends

In this subsection we aim to assess the accuracy of the wind speed trends in the reanalyses by their comparison with trends in the observations taken at the tall tower locations. Before that, some comments regarding the computation of trends must be made and taken into account.

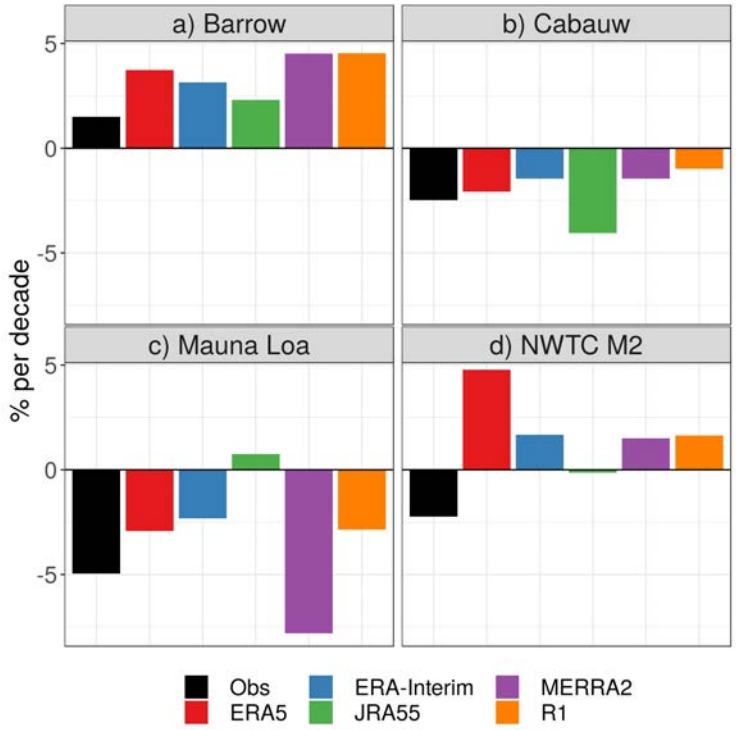


Fig. S4: Bar plot of the observed (*Obs*) and modelled linear trends expressed as a percentage per decade at (a) Barrow, (b) Cabauw, (c) Mauna Loa and (d) NWTC M2 met masts.

The computation of linear trends requires from a reference a period between 15 to 20 years of length (Liléo et al., 2013) with complete observations. A total amount of 23 tall towers within the Tall Tower Dataset have been recording for a period equal or larger than 15 years, and only 14 of these time series span more than 20 years. The number is reduced to 7 if the recording length is increased to 30 years. As mentioned in Section 2.2, these time series usually contain interruptions thus the record lengths needed to compute the long-term trends can be reduced even more.

As stated in Section 4.3, there exist several drivers such as tower maintenance activities or changes in the tower surroundings that may introduce inhomogeneities in the time series producing spurious trends hiding the real climate trend, if it really exists. Although an exhaustive quality control procedure has been applied over the Tall Tower Dataset, no attempt to homogenize these data has been performed so far. In an effort to carry out a preliminary homogeneity examination of the data from the 23 longest series (i.e., spanning at least 15 years), a homogenization test using the R Package *climatol* (Guijarro, 2011) has been performed. This evaluation reveals that there exist several breakpoints producing inhomogeneities in most of these 23 time series. To avoid these imperfections without performing exhaustive homogeneity tests that may derive in the inclusion of contaminated data within the Tall Tower Dataset, a common practice is to divide the time series into separately homogeneous periods (see Aguilar et al. (2003), Section 3.5). Unfortunately, this procedure shortens, even more, the wind series.

The result of this exercise left only four tall towers (Barrow, Cabauw, Mauna Loa and NWTC M2) with reliable measurements for the study of trends. The comparison of the observed linear trend at these sites compared with the trends obtained from the five reanalyses is shown in Fig.

S4. All the reanalysis match the sign of the trend at Barrow (Fig. S4(a)) and Cabauw (Fig. S4(b)) sites. The module, however, is not accurately reproduced, especially at Barrow where all the reanalyses overestimate the observed positive trend. All but JRA55 are able to reproduce the decrease of winds at Mauna Loa (Fig. S4(c)), but a significant disparity is noticed for their modules. None of the reanalyses with the exception of JRA55 matches the negative trend observed at the MWTC M2 mast (Fig. S4(d)). Unfortunately, these sample of four masts is too small to extract robust conclusions on the reproducibility of trends.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J., 2003. Guidelines on climate metadata and homogenization. World Meteorological Organization , 55.

Guijarro, J., 2011. User's guide to climatol. Technical Report.

Liléo, S., Berge, E., Undheim, O., Klinkert, R., Bredesen, R.E., 2013. Long-term Correction of Wind Measurements: State-of-the-art, guidelinies and future work. Technical Report 13:18. Kjeller Vindteknikk.

Torralba, V., Doblas-Reyes, F.J., Gonzalez-Reviriego, N., 2017. Uncertainty in recent near-surface wind speed trends : a global reanalysis intercomparison. Environmental Research Letters 12, 114019. doi:10.1088/1748-9326/aa8a58.

Chapter 6

Local-scale winds captured by seasonal forecasts

Objective

Generate seasonal predictions of near-surface wind speeds that eventually capture local wind effects. The local-scale is generally misrepresented in coarse-scale models because their output represents an averaged value for a grid cell of hundreds of square kilometres. In this regard, refining the output of the prediction system is a sound way forward to improving seasonal predictions' accuracy.

Methodology

- Hybrid predictions are generated through a statistical downscaling with a perfect prognosis approach that allows for transferring the coarse-scale information to the local scale.
- The four main Euro-Atlantic Teleconnections (EATC) are the predictors of the statistical model.
- The skill of hybrid predictions is compared to that of dynamical predictions of wind speed.

Results

- Overall, hybrid predictions show skill at lead times two and three, while dynamical predictions do not.
- The gain in skill by the hybrid predictions is particularly noted in central and northern Europe.
- We illustrate with an example that hybrid predictions are skilful at a site where local wind effects occur.
- Hybrid predictions can be generated using either station data or reanalysis, but specific considerations have to be taken in each case.

Conclusions

- Hybrid predictions are easily interpretable because they use the information of the large-scale atmospheric circulation.
- Hybrid predictions can be applied to anticipate wind speed anomalies at a wind farm level.

Publication

- Ramon, J., Lledó, L., Bretonnière, P.-A., Samsó, M., and Doblas-Reyes, F. J. (2021a). A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environmental Research Letters*, 16(5):054010, doi: 10.1088/1748-9326/abe491

IOP Publishing Environ. Res. Lett. 16 (2021) 054010 https://doi.org/10.1088/1748-9326/abe491

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
30 July 2020

REVISED
29 January 2021

ACCEPTED FOR PUBLICATION
9 February 2021

PUBLISHED
16 April 2021

LETTER

A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds

Jaume Ramon[1], **Llorenç Lledó**[1], **Pierre-Antoine Bretonnière**[1], **Margarida Samsó**[1] and **Francisco J Doblas-Reyes**[1,2]

[1] Barcelona Supercomputing Center (BSC), c/ Jordi Girona, 29, Barcelona 08034, Spain
[2] ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain

E-mail: jaume.ramon@bsc.es

Keywords: statistical downscaling, Euro-Atlantic Teleconnections, seasonal prediction, multi-system

Supplementary material for this article is available online

Abstract

This work provides a new methodology based on a statistical downscaling with a perfect prognosis approach to produce seasonal predictions of near-surface wind speeds at the local scale. Hybrid predictions combine a dynamical prediction of the four main Euro-Atlantic Teleconnections (EATC) and a multilinear statistical regression, which is fitted with observations and includes the EATC as predictors. Once generated, the skill of the hybrid predictions is assessed at 17 tall tower locations in Europe targeting the winter season. For comparative purposes, hybrid predictions have also been produced and assessed at a pan-European scale, using the ERA5 100 m wind speed as the observational reference. Overall, results indicate that hybrid predictions outperform the dynamical predictions of near-surface wind speeds, obtained from five prediction systems available through the Climate Data Store of the Copernicus Climate Change Service. The performance of a multi-system ensemble prediction has also been assessed. In all cases, the enhancement is particularly noted in northern Europe. By being more capable of anticipating local wind speed conditions in higher quality, hybrid predictions will boost the application of seasonal predictions outside the field of pure climate research.

1. Introduction

Recent advances in the fields of climate modelling and seasonal prediction have resulted in skilful seasonal predictions of surface variables over the extratropics (Merryfield *et al* 2020). This has, in turn, led to the development of climate services that inform weather-and-climate-vulnerable socio-economic sectors of seasonal anomalies a few months ahead (Buontempo *et al* 2018). The energy sector takes advantage of such valuable information since energy production and demand are strongly linked to climate variability. In particular, the renewable energy industry can profit from seasonal predictions of surface wind speed (Clark *et al* 2017, Torralba *et al* 2017) and wind power generation (Lledó *et al* 2019) to anticipate revenues, balance electricity supply and demand or schedule maintenance activities among others. However, those predictions still suffer from some limitations, mainly due to (1) the limited skill levels on surface variables available from current seasonal prediction systems and (2) its relatively coarse spatial scales.

Generally, seasonal anomalies of atmospheric variables arise from large-scale forcings that other components of the Earth system exert as boundary conditions, such as anomalies of sea ice extent, sea surface temperature or soil moisture. These boundary-condition forcings can be adequately represented in coarse-scale coupled models—often delivered in grids of tens of square kilometres—leading to some skill in the predictions. However, the absolute values that are experienced near the surface at the local scales can be highly affected by local effects and vary substantially even at short distances. Values of surface temperature or precipitation are affected by the local topography, particularly in complex-terrain regions (e.g. Anders *et al* 2006). Near-surface wind speeds are affected not only by topography but also by surface roughness, buildings and obstacles. For

instance, near-surface wind speed conditions can be very different at the top of a ridge, at a mountain pass or at a valley floor. These differences in magnitude are especially relevant for deriving indicators that are non-linear and therefore sensitive to absolute magnitudes, such as the capacity factor (CF) of wind power (Pickering *et al* 2020).

To transfer climate information from coarser to finer scales, many downscaling techniques have been developed and employed in weather and climate studies to refine model outputs. There are essentially two different downscaling approaches. Firstly, dynamical downscaling couples a Regional Climate Model (RCM) to a Global Circulation Model (GCM) over a limited region within the global domain, using the data from the GCM as boundary conditions. The computational costs of dynamical downscaling are rather high and the additional skill is sometimes negligible (Robertson *et al* 2012), which explains its limited use in seasonal predictions (García-Díez *et al* 2015, Schwitalla *et al* 2020).

Secondly, statistical downscaling relies on the assumption that a relationship exists between the large-scale information provided by a GCM and the fine-scale variable. Once a statistical relationship is built, local values —predictands— are inferred using large-scale information—predictors—. Then, future dynamical predictions of the large-scale variables (i.e. those generated using the physically-based equations of the dynamics of the atmosphere) can be inserted as predictors into the statistical relationship to produce local-scale predictions. Statistical downscaling techniques (see Gutierrez *et al* 2013 for a review) can be in turn subdivided depending on whether the statistical model is fitted using observational data for both predictors and predictands (known as Perfect Prognosis or PP; Klein *et al* 1959) or using data from the GCM itself (often referred to as Model Output Statistics or MOS; Glahn and Lowry 1972).

The statistical downscaling approach is relatively easy to implement with climate prediction systems containing several ensemble members and has already been employed in some studies for downscaling temperature and precipitation forecasts at seasonal timescales (e.g. Pavan and Doblas-Reyes 2013, Manzanas *et al* 2018). However, to the best of the authors' knowledge, no attempt has yet been made to downscale seasonal predictions of wind speed.

The selection of the employed predictors is vital for the success of the statistical downscaling method. Not only do the predictors need to be strongly related to the predictand, but also predictable from the dynamical model. Teleconnection indices that summarise the state of the atmospheric circulation are optimal for this purpose. In this work, four Euro-Atlantic Teleconnection (EATC) indices (namely the North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EAWR) and Scandinavian Pattern (SCA)) are employed as predictors to anticipate near-surface wind speed conditions in Europe. Those teleconnection indices are strongly related to wind speed conditions in Europe (Zubiate *et al* 2017) and wind power generation (Yang *et al* 2020), and have been recently shown to be predictable (Lledó *et al* 2020). Since the downscaled predictions combine a dynamical forecast of a circulation variable and a statistical relationship with a second variable of interest, they are referred to as hybrid predictions (see chapter 2 in WMO 2020), to differentiate them from purely statistical seasonal forecasts that employ observed values of potential forcing fields to derive the predictions (Kämäräinen *et al* 2019). Hybrid predictions take advantage of the predictability of the EATC indices from dynamical predictions, especially in winter, and thus help to overcome limitation (1). At the same time, the downscaling allows for transferring such information to a finer grid scale, circumventing limitation (2).

The objective of this work is to generate and assess the quality of a hybrid seasonal prediction of near-surface wind speeds and wind power CF by applying a statistical downscaling with a PP approach to a set of dynamical predictions of EATC indices. Sections 2 and 3 describe the data and methodology employed, respectively. Results are presented in section 4 while conclusions are drawn in section 5.

2. Datasets

The hindcasts from five different operationally-produced seasonal prediction systems have been used in this study: the System2 from Deutscher Wetterdienst (DWD2, Deutscher Wetterdienst 2019), the GloSea5-GC2 from the UK Met Office (GS5GC2, Maclachlan *et al* 2015, Williams *et al* 2015), the System 6 from Météo France (MF6, Dorel *et al* 2017), the SEAS5 (Johnson *et al* 2019) from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Seasonal Prediction System 3 from Centro Euro-Mediterraneo sui Cambiamenti Climatici (SPS3, Sanna *et al* 2017). All five prediction systems have been retrieved from the Climate Data Store data portal in a regular grid of $1° \times 1°$ of spatial resolution and covering the 1993–2016 period. Particular details of the employed seasonal prediction systems, as well as the two observational references, can be found in table 1.

The ERA5 HRES (hereafter ERA5) reanalysis dataset (Hersbach *et al* 2020) produced by the ECMWF has been used as the gridded observational reference. The dataset has been downloaded through the ECMWF retrieval system (MARS) in its native grid (i.e. 0.3° approximately), and at 1-hourly time resolution. Then, the ERA5 data has been horizontally interpolated using a conservative approach to match the spatial resolution of the predictions, allowing for bias adjustment and verification at the grid level.

Table 1. Specific details of the datasets employed.

Dataset	Type of dataset	Available period	Time resolution	Horizontal grid spacing	Ensemble members
DWD2	Seasonal prediction	1993–2016	6 h	$1^\circ \times 1^\circ$	30
GS5GC2	Seasonal prediction	1993–2016	6 h	$1^\circ \times 1^\circ$	28
MF6	Seasonal prediction	1993–2016	6 h	$1^\circ \times 1^\circ$	25
SEAS5	Seasonal prediction	1993–2016	6 h	$1^\circ \times 1^\circ$	25
SPS3	Seasonal prediction	1993–2016	6 h	$1^\circ \times 1^\circ$	40
multi-system	Seasonal prediction	1993–2016	Seasonal	$1^\circ \times 1^\circ$	148
ERA5	Reanalysis	1950–present	Hourly	$0.3^\circ \times 0.3^\circ$	—
TTD	*In-situ* observations	1984–2017	Sub-hourly	Irregular	—

Table 2. Particular details of the 17 tall towers employed in this study. *r* represents the Pearson correlation coefficient between the seasonal tall tower winds and the ERA5 100 m winds from closest grid point to each tall tower location.

ID	Name	Longitude (deg east)	Latitude (deg north)	Offshore	Measuring height (m)	Original time span[a]	*r*
T1	Braschaat	4.52	51.31	No	41	1996–2015	0.82
T2	Cabauw	4.92	51.97	No	80	1986–2017	0.91
T3	Cardington	−0.42	51.10	No	50	2004–2013	0.98
T4	Fino1	6.59	55.01	Yes	100	2004–2017	0.96
T5	Fino2	13.15	55.01	Yes	102	2007–2017	0.90
T6	Fino3	7.16	55.20	Yes	100	2009–2017	0.94
T7	Greater Gabbard MMZ	1.92	51.94	Yes	82	2005–2015	0.97
T8	Hamburg University	10.10	53.52	No	110	2004–2017	0.93
T9	Hegyhatsal	16.65	49.96	No	115	1994–2016	0.21[b]
T10	Inner Dowsing	0.44	53.13	Yes	43	1999–2008	0.83
T11	Juelich	6.22	50.93	No	100	2011–2017	0.88
T12	Lindenberg	14.12	52.17	No	98	1999–2017	0.98
T13	Lutjewad	6.35	53.40	No	60	2001–2017	0.89
T14	Malin Head	−7.33	55.35	No	22	1988–2017	0.89
T15	Obninsk	36.60	55.11	No	121	2007–2016	0.90
T16	Puijo	27.65	62.91	No	75	2005–2016	0.76
T17	Sodankyla	26.64	67.36	No	24	2000–2015	0.83

[a] May contain no-data periods.

[b] T9 will not be included in the results.

At the local scale, wind speeds measured *in-situ* at 17 tall tower locations over Europe have been considered (see details in table 2 and their spatial distribution in figure S1 (available online at stacks.iop.org/ERL/16/054010/mmedia)). Those observations have been obtained from the Tall Tower Dataset (TTD, Ramon *et al* 2020), a quality-controlled collection of wind data taken at tall meteorological masts of 20 to more than 200 m height. Since these structures measure winds simultaneously at several heights above ground, we have selected at each of the 17 locations the wind speed series which is closest to the 100-metre height. Modern wind turbines are placed at those heights since the wind flow is notably less affected by surface roughness than at surface level. The 17 time series span from 6 to 30 years within the 1984–2017 period. To unify the timespan of the series, and ensure the representativeness of the comparisons against predictions, the 17 time series have been averaged into hourly values and reconstructed to cover the entire 1981–2017 period. To this end, a Measure-Correlate-Predict approach with a simple linear regression has been employed (see Carta *et al* 2013 for further details), using as the reference series the hourly 100 m wind series of the ERA5's closest grid point to each tall tower location.

3. Methods

3.1. Hybrid predictions

Hybrid predictions for the boreal winter (December–January–February, DJF) have been produced using the PP methodology as represented in figure 1. Once the dynamical forecasts of the predictors (i.e. the EATC indices) are generated, they are used in a statistical model that accounts for variations in wind speed related to variations in the EATC indices. The statistical model has been previously built solely on observations of wind speed and EATC indices. For the purposes of our work, the PP approach represents an advantage over MOS, because (1) it uses one single statistical relationship that can be applied over various dynamical prediction systems, and (2) the amount of data available for fitting the relationship

IOP Publishing Environ. Res. Lett. 16 (2021) 054010 J Ramon et al

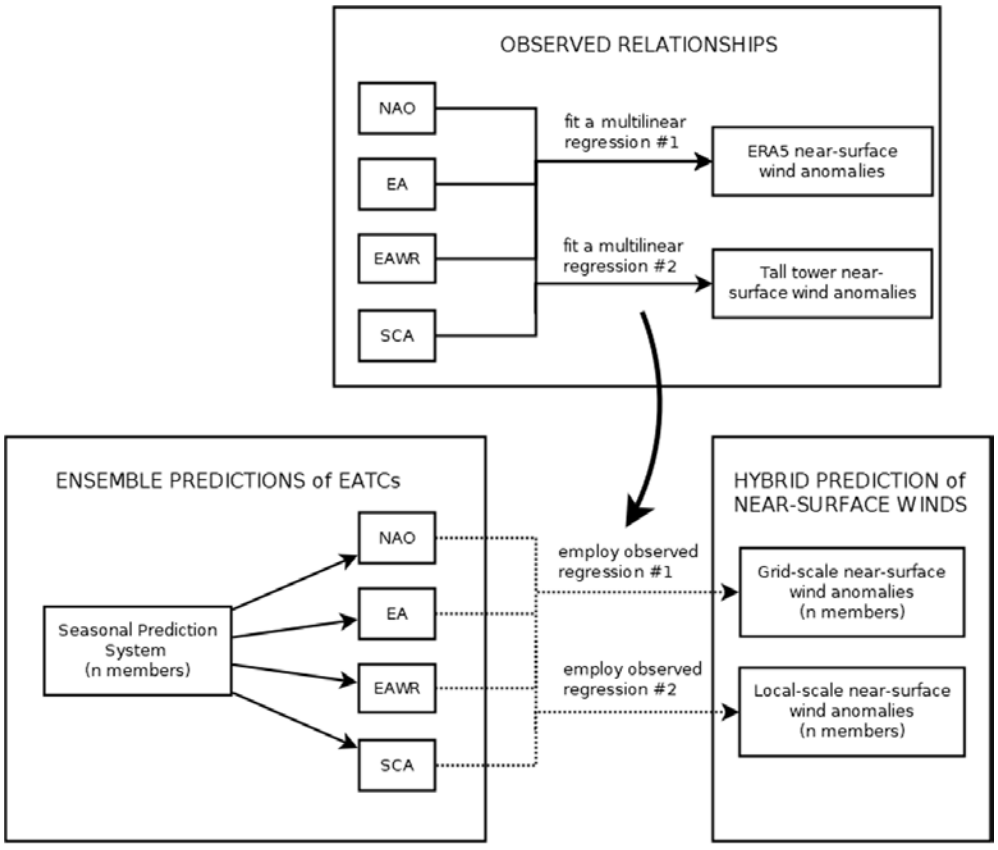


Figure 1. Schematic representation followed in the generation of the hybrid predictions.

is not limited to the length of the hindcast, but to the timespan of the observational series (Marzban *et al* 2006). A more precise description of the different steps of the PP and the generation of the hybrid predictions follows.

Firstly, four EATC indices are computed as described in Lledó *et al* (2020). The EATC patterns and indices have been derived from the 500 hPa geopotential height field anomalies employing a Rotated Empirical Orthogonal Function (REOF) analysis over the Euro-Atlantic domain [90° W–60° E; 20° N–80° N]. The four teleconnections obtained correspond to the North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EAWR) and Scandinavian pattern (SCA). This procedure has been followed to obtain both observed—using the ERA5 anomalies—and predicted—employing the anomalies from DWD2, GS5GC2, MF6, SEAS5 and SPS3—EATC indices. The observed EATC patterns are shown in figure 1 in Lledó *et al* (2020).

Then, a statistical model that relates seasonal anomalies of near-surface wind speed and the EATC indices is built from historical observations. A very simple multilinear regression model (equation (1)) has been used here, due to the rather small sample size available for fitting the model. This method has already been used in Rust *et al* (2015) to model European temperatures from several teleconnections. A model that expresses anomalies of near-surface wind speeds (predictand: w') as a linear combination of the EATC indices (predictors: *NAO, EA, EAWR, SCA*) is built separately at each grid point or tall tower location (x, y). The fit adjustment parameters a_n are obtained employing an ordinary least squares method (see their spatial distribution in figure S2). The reference period that is used in all the model fits is 1981–2017. Additionally in the generation of the multilinear models, a leave-one-out cross-validation approach has been considered. The EATC observed indices and its corresponding wind observation of the year under consideration are excluded from the sample used to estimate the fit adjustment parameters. In this way, they can be used later for verification

$$w'(x,y,t) = a_0(x,y) + a_1(x,y) * NAO(t) + a_2(x,y) * EA(t) + a_3(x,y) * EAWR(t) + a_4(x,y) * SCA(t). \quad (1)$$

To avoid overfitting in the statistical model, a selection of the best subset of predictors that retains

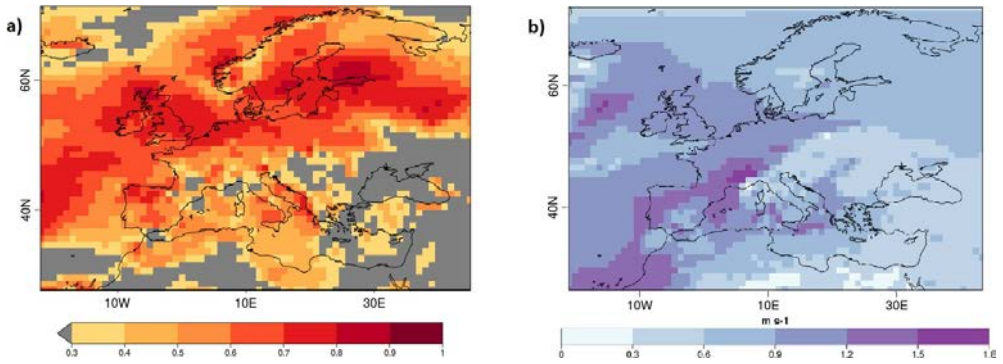


Figure 2. (a) Coefficient of determination R^2 of the linear fit between the ERA5 near-surface wind anomalies and the EATCs over the 1981–2017 period. Time resolution is seasonal and the map corresponds to DJF. The higher the R^2, the better are surface winds explained by the EATCs. Grey-masked areas show R^2 values lower than 0.3. (b) Interannual variability of ERA5 near-surface boreal-winter wind speeds over the 1981–2017 period. Interannual variability has been computed as the standard deviation of all boreal-winter seasonal means of the years under consideration.

the maximum information in the model without necessarily keeping all the predictors is made at each location by using the Akaike Information Criterion and a backward stepwise selection (James *et al* 2013). Albeit using a relatively simple statistical model, the coefficient of determination (R^2) of the fit presented in figure 2(a) shows that the EATCs explain most of the year-to-year variability (also known as interannual variability) in the near-surface winds over extended areas of Europe (figure 2(b)).

In order to obtain an ensemble of hybrid near-surface wind anomaly predictions, the individual-member predictions of EATC indices are inserted into the multi-linear regressions, both at gridded and local scales. The seasonal ensemble predictions of the EATC indices are initialised at the beginning of winter (December) and one, two and three months in advance (i.e. November, October and September, respectively). In this work, lead-zero predictions will refer to those initialised in December, lead-one predictions will be those initialised in November, and so on. Finally, all members from the five prediction system are pooled together to create a new dataset, the multi-system henceforth, with a total of 148 members. Multi-system ensemble predictions can outperform individual-system predictions (Athanasiadis *et al* 2017).

3.2. Wind capacity factor

The wind-based CF index is obtained using the 6 h wind speed data from the predictions, and the 1-hourly winds from the ERA5 and TTD. The conversion between wind speed and power output has been made employing a power curve, which takes into account the specific efficiency characteristics of the wind turbine. Specifically, a power curve for the turbine Type I defined in the IEC-61400-12-1 international standard has been considered (see IEC 2017 and Lledó *et al* 2019 for further information). Although this turbine type might not be the most suitable for all the investigated locations, it serves the purpose of investigating whether the non-linearities of its power curve affect the quality of the hybrid predictions. Once the conversion is made, CF values are obtained dividing by the nominal power capacity of the turbine. Lastly, seasonal anomalies are calculated.

Hybrid predictions of CF, which might be of particular interest at a turbine or wind farm level within the wind industry, are studied in detail at one tall tower location where local wind effects represent a huge proportion of the seasonal mean wind speed value, and subsequently the seasonal CF value. The relatively low r obtained for Puijo tall tower (T16, table 2) envisages that local wind effects are likely to occur there, and a comparison against a surface station located two kilometres away reveals so (Leskinen *et al* 2009).

3.3. Verification metrics

The quality of the hybrid predictions has been assessed employing both gridded and local-scale observations. Multiple verification scores have been considered to account for different aspects of forecast quality: association, discrimination and reliability (Jolliffe and Stephenson 2012, Mason 2018). In some of those scores, the performance of the hybrid predictions is compared to that of a benchmark prediction. Two different benchmarks have been employed: the climatological forecast (i.e. a 33% of probability for all tercile categories) and the dynamical predictions of near-surface wind speed from the considered systems (table 1). Skill scores using the climatological forecast as a reference are identified with the sub-index c while those using the dynamical prediction use d.

To prepare the dynamical prediction benchmark, seasonal anomalies of surface (10 m) wind speeds for the 1993–2016 period have been obtained at gridded and local scales (in the latter case using a

IOP Publishing Environ. Res. Lett. 16 (2021) 054010 J Ramon et al

bilinear interpolation) and then bias-adjusted using a simple bias correction approach (Torralba *et al* 2017). The method adjusts predictions to have an equivalent standard deviation and mean to that of the reference dataset, which has been the ERA5 reanalysis near-surface wind speeds. A leave-one-out cross-validation approach has been again used: the prediction to be adjusted and its corresponding observation are excluded from the sample used to estimate the adjustment parameters (see equations (1)–(4) in Torralba *et al* 2017). The multi-system of the dynamical predictions is also generated by pooling together all the bias-corrected anomalies from the five prediction systems.

The considered scores for the skill assessment are both deterministic and probabilistic, and the R packages easyVerification and SpecsVerification have been used for their computation:

- The Ensemble Mean Correlation (EMC) quantifies the association (i.e. linear dependency) between observed and predicted wind speeds. The EMC ranges from −1 to 1, with a value of 1 indicating a perfect association. A Student's *t*-test at the 95% of confidence level has been applied to emphasise statistically significant areas.
- The Relative Operating Characteristic Skill Score (ROCSS) assesses the discrimination of probabilistic single-category forecasts. Here, predictions are prepared in the form of probability of occurrence of three categories defined by the 33rd and 66th percentiles of the hindcast values. The ROCSS measures the proportion of hits (i.e. correct predictions) versus false alarms (i.e. non-occurrences that were incorrectly predicted) for each of the three categories. The ROCSS ranges from −1 to 1, with negative values indicating a weaker discrimination capacity than that of the benchmark prediction.
- The Rank Histogram (RH) tests the reliability of the probabilistic predictions, by comparing how the observations rank with respect to the ensemble members of the predictions. Reliable ensemble predictions show a flat RH, which has been statistically assessed with a decomposed Pearson's χ^2 test as in Jolliffe and Primo (2008). When the sample size is small in comparison with the number of ranks available (i.e. the ensemble size), non-flat rank histograms are likely to occur due to randomness, which is not desirable. To prevent this from happening, counts from every ten adjacent bins have been grouped so that the number of ranks has been reduced by a factor of ten.
- The Continuous Ranked Probability Skill Score (CRPSS) measures the quality of the cumulative forecast probability distribution by measuring the distance between the observed and predicted probability distributions. The CRPSS penalises both reliability and resolution—the latter is closely related to discrimination—errors. It ranges from $-Inf$ to 1, and positive values indicate an increased skill compared to the benchmark forecast. The Diebold–Mariano test (Diebold and Mariano 1995) has been applied to explore the statistical significance of the differences between the CRPSSs of hybrid and dynamical predictions.

Finally, areas where the hybrid model shows a poor performance based on the R^2 of the statistical fit being smaller than 0.3—grey areas in figure 2(a)—have been omitted in the verification. Those areas are located around the Black Sea and scattered around the northern Mediterranean, where low values of interannual variability are noted (figure 2(b)). There, winds respond mainly to mesoscale systems rather than large-scale circulation patterns, which may explain the inability of the hybrid model in reproducing the year-to-year variations of near-surface wind speeds. T9 has been omitted in the results as well since the local winds correlate very poorly with the ERA5 winds (table 2), thus not giving robustness to the Measure-Correlate-Predict reconstruction.

4. Results

In the following sections, we analyse the skill of the hybrid predictions at the local scale. We complement these results with the verification of grid-scale hybrid forecasts (i.e. adjusted to reanalysis data instead of tower observations) at a pan-European scale [27° N–72° N; 22° W–45° E]. This is important because potential users of hybrid predictions may face the limitation of the unavailability of *in-situ* local data needed to generate the predictions. The verification focuses on three key attributes of a probabilistic prediction: association, discrimination and reliability. For the sake of simplicity, results are shown only for the multi-system prediction. Remaining results for the individual systems are available from the authors upon request. We focus on the winter season, when wind speed variability is highest, and so is the importance of its anticipation.

4.1. Do hybrid predictions improve their dynamical counterparts?

The association between the observed and hybrid-predicted near-surface wind anomalies is measured by the EMC and illustrated in figures 3(a)–(d) at both local and grid scales. The EMC is a deterministic metric which is insensitive to forecast errors in the magnitudes and the spread of the ensemble, so only some association with the observations is required for a forecast to be skilful. In this regard, the negative correlation values noted across the Mediterranean basin anticipate a poor performance of the hybrid prediction over there. Conversely, positive and significant correlations above 0.6 have been obtained for lead month zero across northern Europe (figure 3(a)).

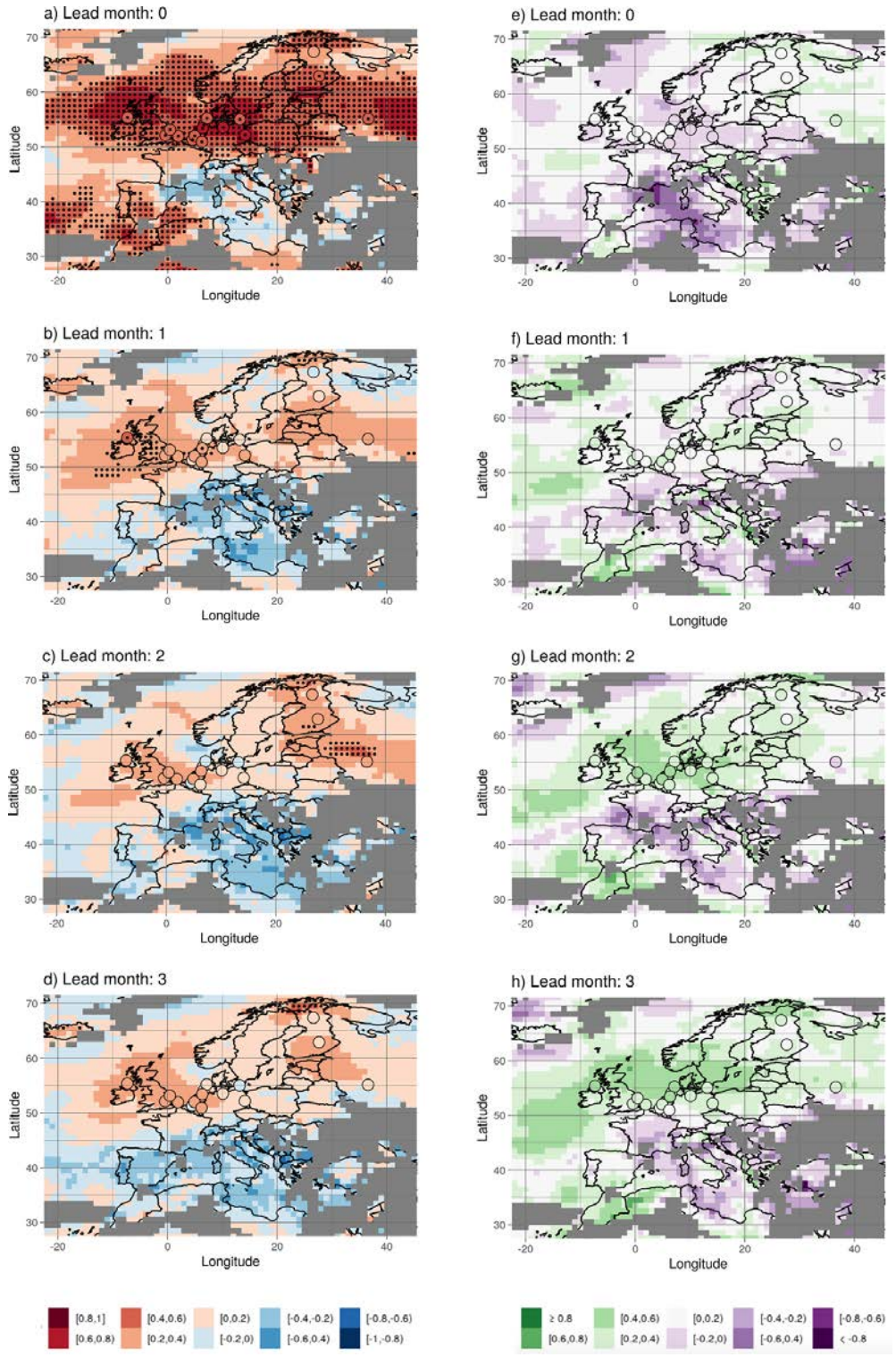


Figure 3. (a)–(d) EMC of the multi-system near-surface wind speed hybrid predictions over Europe for DJF. Black dots highlight areas statistically significant at the 95% of confidence level. (e)–(h) Differences between the EMC of the hybrid predictions and the EMC of the dynamical predictions. Filled points indicate the EMC of the hybrid predictions ((a)–(d)) and the differences in EMC ((e)–(h)) at the 16 tall tower locations. Grey-masked areas indicate regions where the R^2 of the multi-linear regression is below 0.3.

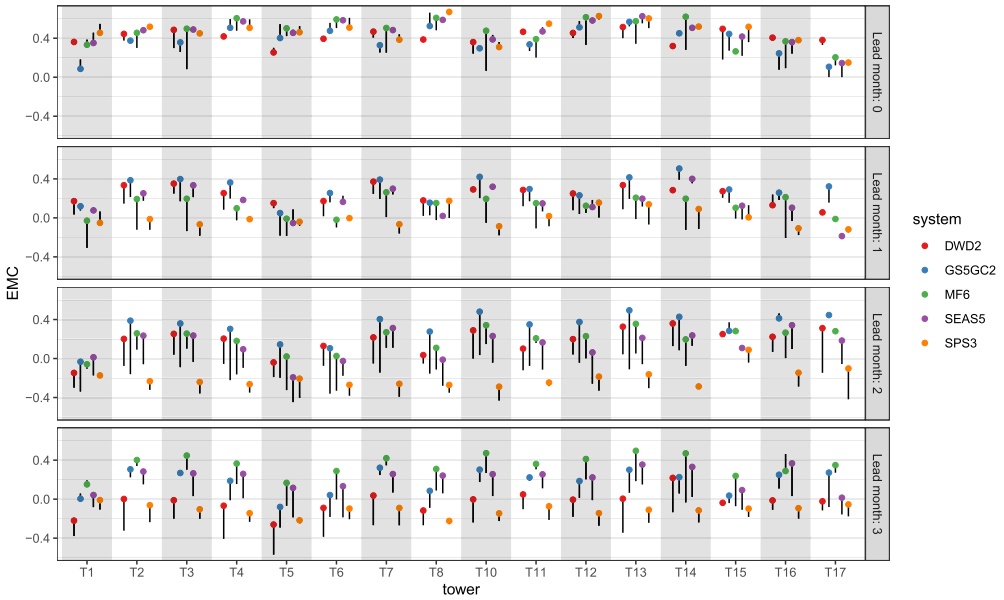


Figure 4. EMC of the DWD2, GS5GC2, MF6, SEAS5 and SPS3 near-surface wind speed predictions at the 16 tall tower locations for DJF. The circle of the lollipop shows the EMC of the hybrid prediction, whereas the opposite end represents the EMC of the dynamical prediction. Right-side-up lollipops show that the hybrid prediction outperforms its dynamical counterpart, and vice-versa. Predictions were initialised in December (lead month 0, first row), November (lead month 1, second row), October (lead month 2, third row) and September (lead month 3, fourth row).

At longer leads, correlations decrease but still depict positive values above 0.4 in the British Isles and the east of the Baltic sea. For the latter region, we observe increased and statistically significant EMC values at lead month two, which are not seen at lead months one and three. This improvement in the hybrid prediction at that particular lead month responds to an increase in the skill values of the EATC predictions. More specifically, the SCA index has the greatest weight in the hybrid model over that region (figure S2), and shows a relative maximum in correlation at lead month two (i.e. 0.42; see table S1). The differences in EMC between the hybrid and dynamical predictions (figures 3(e)–(h)) reveal that the highest gains in skill are seen at the longest leads. While the dynamical forecast offers skill only at leads zero and one (see figure S3), the hybrid prediction shows positive EMCs at all lead times. The increased scores for predictions based on the circulation patterns in the hybrid method appear to match the increase in skill seen in other recent studies (e.g. Scaife *et al* 2014, Baker *et al* 2017).

Results are similar at the local scale. The lollipop plots (figure 4) depict the most noticeable differences between hybrid and dynamical predictions at longer leads, where the improvement of the hybrid prediction is substantial for all systems but the SPS3.

The sensitivity of the predictions to discriminate between observations belonging to different categories has been explored with the ROCSS. The $ROCSS_c$ of the lower-tercile category for the multi-system hybrid and dynamical predictions is compared in figure 5. While both hybrid and dynamical predictions show similar skill score values at lead zero (panel (a); the density is centred around the $y = x$ line), it is noted that hybrid predictions enhance the discrimination ability at leads one, two and three (panels (b), (c) and (d); most of the density is found above the $y = x$ line). Furthermore, this improvement is not only restricted to a particular region but positive $ROCSS_d$ values are observed all over Europe (not shown).

Analogous results are obtained for the predictions of the upper-tercile category (figure S4). On the other hand, neither hybrid nor dynamical predictions show skill for the central-tercile category (figure S5). The lack of skill in predictions for near-normal is a recurrent issue which has already been addressed in the literature and stems from the definition of the skill scores itself, thus not requiring any physical or dynamical explanation (Van Den Dool and Toth 1991).

To gain more insight into the performance of the hybrid predictions at the local scale, we have selected four tall tower locations to evaluate the reliability of the ensemble predictions by exploring their rank histograms (figure 6). The set of four locations include T2 (Cabauw, The Netherlands), T5 (Fino2, Germany), T15 (Obninsk, Russian Federation) and T16 (Puijo, Finland), which are located in both continental—flat and complex terrain—and offshore platforms across northern Europe.

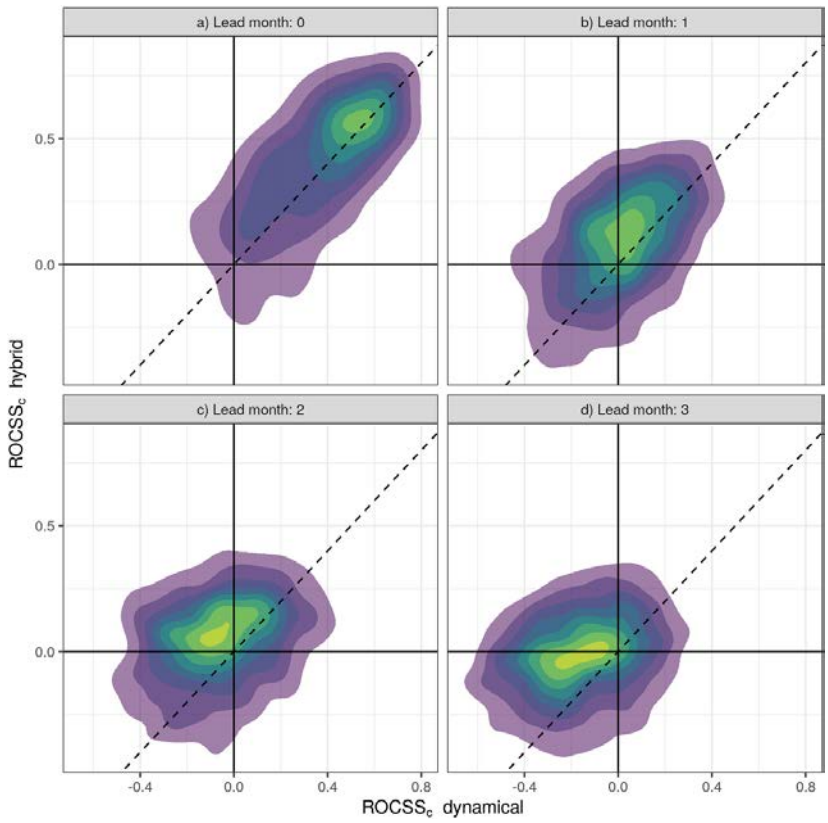


Figure 5. Two-dimensional density plots showing the ROCSS$_c$ of the lower-tercile hybrid and dynamical predictions from the multi-system at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of ROCSS$_c$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in (a) December (lead month 0), (b) November (lead month 1), (c) October (lead month 2) and (d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

Focusing on lead zero, the RHs of the hybrid predictions at T2 and T5 (figures 6(a) and (b), respectively) are both U-shaped, mirroring an overpopulation of the outermost ranks which can occur due to either a lack of ensemble mean signal or a lack of spread around the ensemble mean (Eade *et al* 2014) in the hybrid prediction. The non-flatness of the RH is statistically supported by the *p*-values of the Jolliffe–Primo statistical test—at the 95% of confidence level. Conversely, the RH at T15 (figure 6(c)) depicts an opposite convexity (i.e. overdispersion), but this outcome is not statistically significant. These results envisage a poor reliability of the multi-system hybrid predictions at these particular locations, which can also be noted for the individual systems (figures S6–S10), especially at T5. The unreliability of the multi-system hybrid predictions is observed in the RHs of 10 out of the 16 tall tower locations, while the other 6 locations show a flatter plot such as that observed at T16 (figure 6(d)). This indicates that the probability distribution of the ensemble at these six locations is in agreement with the observed values. Similar results are obtained for the other leads (not shown). The performance of the hybrid predictions could be improved further by employing calibration methods (Doblas-Reyes *et al* 2005, Manzanas *et al* 2019) or performing variance corrections to the ensemble mean and members (Eade *et al* 2014).

To complete the skill assessment we compute the CRPSS, a restrictive quality metric of the ensemble distribution that accounts for both discrimination and reliability at the same time. Figure 7 presents the CRPSS$_d$, highlighting areas where the hybrid approach improves (positive values) or degrades (negative values) the dynamical prediction. In general, the results match those discussed for the EMC and ROCSS (figures 3 and 5, respectively) with the highest gains seen for leads two and three. However, the corresponding CRPSS$_c$ values of the hybrid predictions are mostly negative (figure S11). Positive CRPSS$_c$ values are only noted for the lead-zero predictions and, in the case of MF6, the hybrid forecast is the only that offers skill (figure S12).

According to Mason (2004), some scores such as the Ranked Probability Skill Score—and thus the CRPSS—are often too harsh when the climatological

IOP Publishing Environ. Res. Lett. 16 (2021) 054010 J Ramon et al

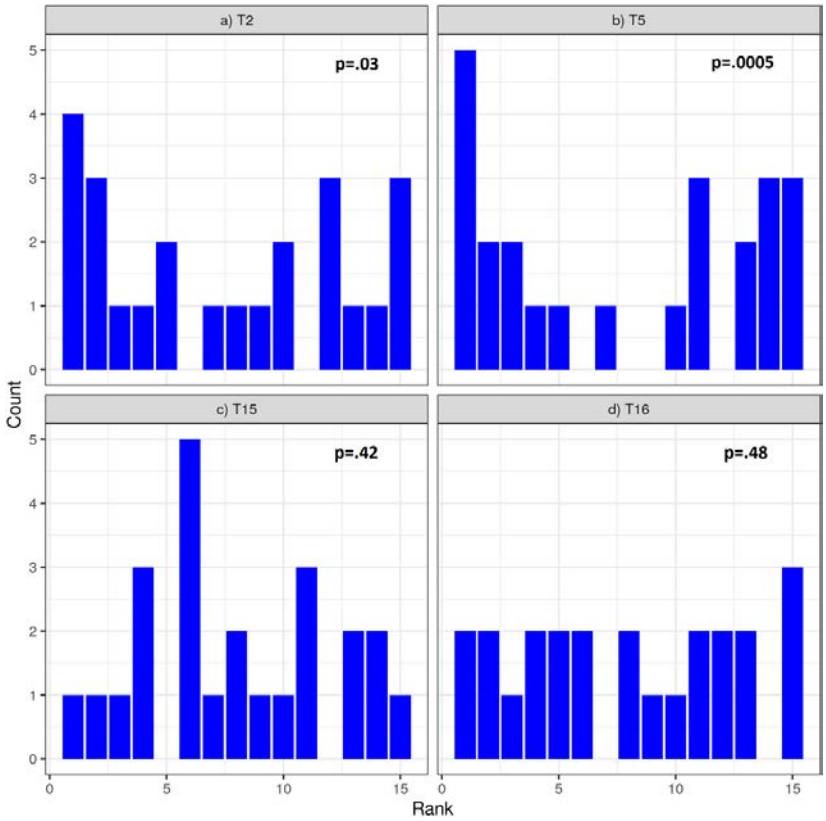


Figure 6. Rank histograms of the lead-zero multi-system near-surface wind speed hybrid predictions for DJF at (a) T2 [Cabauw; 4.92° E, 51.97° N], (b) T5 [Fino2; 13.15° E, 55.01° N], (c) T15 [Obninsk; 36.60° E, 55.11° N] and (d) T16 [Puijo; 27.65° E, 62.91° N] tall tower locations. The *p*-value of the Jolliffe–Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

forecast is considered as benchmark. This gives high chances of getting negative values even when predictions provide useful information. Such is the case observed here: although the $CRPSS_c$ is generally negative, we observe gains in association and discrimination and, in some cases, hybrid predictions are reliable. Therefore, one should not rely solely on a single skill score but take into account the whole verification.

4.2. Can hybrid predictions always be trusted at a local scale?

At this point in the results, it has been shown that hybrid predictions improve the dynamical in many aspects, primarily in northern Europe. However, little has been discussed about how hybrid predictions perform at the micro-scale level, especially when local wind effects occur. In the following, we illustrate how hybrid predictions could be applied to predict the absolute values of the wind CF at a location where local wind effects have been reported, and quantify the error made when reanalysis gridded data—which sometimes misrepresent those effects—are used to fit the hybrid model.

The ensemble predictions of CF for Puijo site are presented in figure 8 in the form of Probability Density Functions (PDF). We note that the direct output of the grid-scale hybrid predictions is considerably biased, being the seasonal mean CF systematically underestimated (figure 8(a)). A $CRPSS_c$ value of −4.215 indicates that the prediction is completely useless. A later bias adjustment of this prediction (figure 8(b)) removes the bias and adjusts the variability to that observed at Puijo—though the skill score of the prediction is still negative (−0.046), indicating a similar performance to that of a climatological forecast. Finally, the hybrid prediction fitted with *in-situ* data also adjusts well to the observed CFs, and the CRPSS increases a bit more, up to a positive value of 0.0007, indicating that the use of local observations with the hybrid method provides the most accurate prediction of seasonal CF values.

The important bias in the grid-scale predictions in figure 8(a) responds to the fact that gridded data are a representation of the average value within a grid cell of hundreds of square kilometres. Therefore, values of variables with high spatial variability such as wind speed in complex terrain regions may differ substantially from the actual values observed at

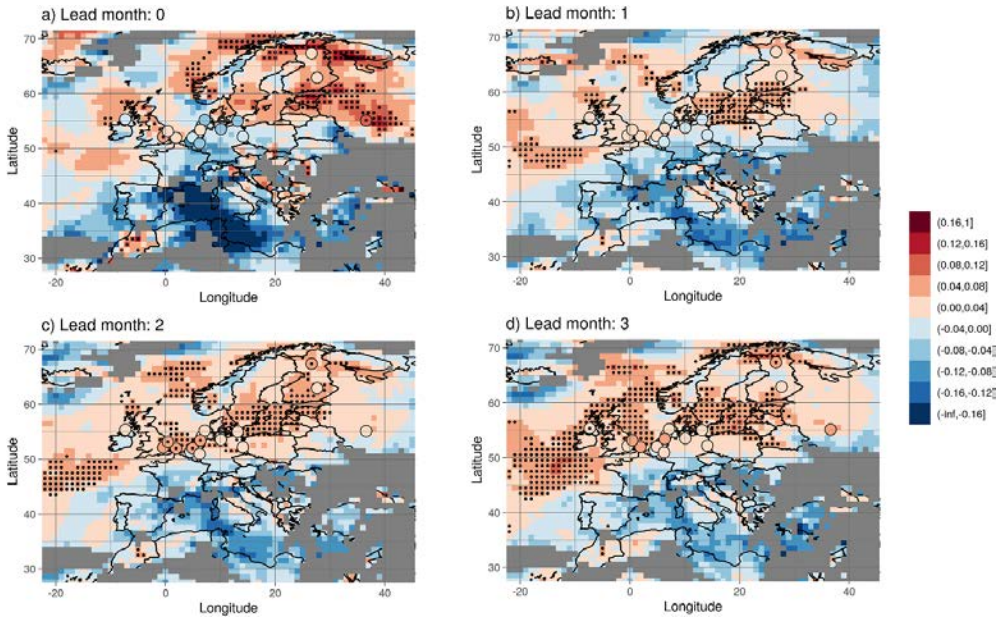


Figure 7. $CRPSS_d$ of the multi-system near-surface wind speed predictions over Europe for DJF. Predictions were initialised in (a) December (lead month 0), (b) November (lead month 1), (c) October (lead month 2) and (d) September (lead month 3). Filled points indicate the $CRPSS_d$ of the predictions at the 16 tall tower locations. Black dots highlight areas statistically significant at the 95% of confidence level. Grey-masked areas indicate regions where the R^2 of the multi-linear regression is below 0.3.

different locations within the grid cell. This misrepresentation of local values is said to produce representativeness errors. In the case of wind, local effects such as katabatic winds over complex terrain regions may account for a large proportion of the mean wind speed value, thus enlarging the representativeness error of the wind speeds in the reanalysis. These errors are propagated to the CF values, and eventually to the hybrid predictions. Hence, reanalysis gridded data are sometimes not suitable to generate hybrid predictions because these datasets are unable to represent local wind effects occurring at much finer scales, such as those observed at Puijo. A later bias-correction may enhance the grid-scale hybrid predictions, but this post-processing can only be carried out where *in-situ* measurements are available.

5. Summary and conclusions

This research proposes and applies a methodology to overcome two main restraints of seasonal predictions that jeopardises every decision based upon them. The first impediment is the limited skill levels observed in the prediction of surface variables such as wind speed, while the second is the lack of adaptation to the local scale due to the relatively coarse scales in which forecasts are delivered.

Results show that hybrid predictions of near-surface wind speed based on a PP statistical downscaling technique help reduce the effects of both issues simultaneously. Using the indices of the four main EATCs as predictors, the hybrid predictions proposed here have been shown to improve the skill of the same predictions obtained from a dynamical approach. Besides, the statistical downscaling has enabled to transfer the coarse-scale predictions to a station-scale level, and the comparison with station-based observations has revealed certain level of agreement even when local wind effects play an important role. In particular:

- Hybrid predictions enhance the skill scores of the dynamical predictions at both local and pan-European scales.
- In general, hybrid predictions are able to provide skill at leads two and three, while dynamical forecasts cannot.
- The highest gains in quality are observed in the association with the observations and the discrimination against the different observed outcomes.
- Although hybrid predictions can also be built using reanalyses, it is advisable not to use gridded data to build the statistical model over areas where local effects are considerable.
- EATC predictions—and thus hybrid predictions—provide no added value in the Mediterranean basin.

Hybrid forecasts foster the information available in the EATC predictions to anticipate near-surface wind speed or CF anomalies. The derived predictions are consistent with the main features of the atmospheric circulation, which are summarised in

IOP Publishing Environ. Res. Lett. 16 (2021) 054010 J Ramon et al

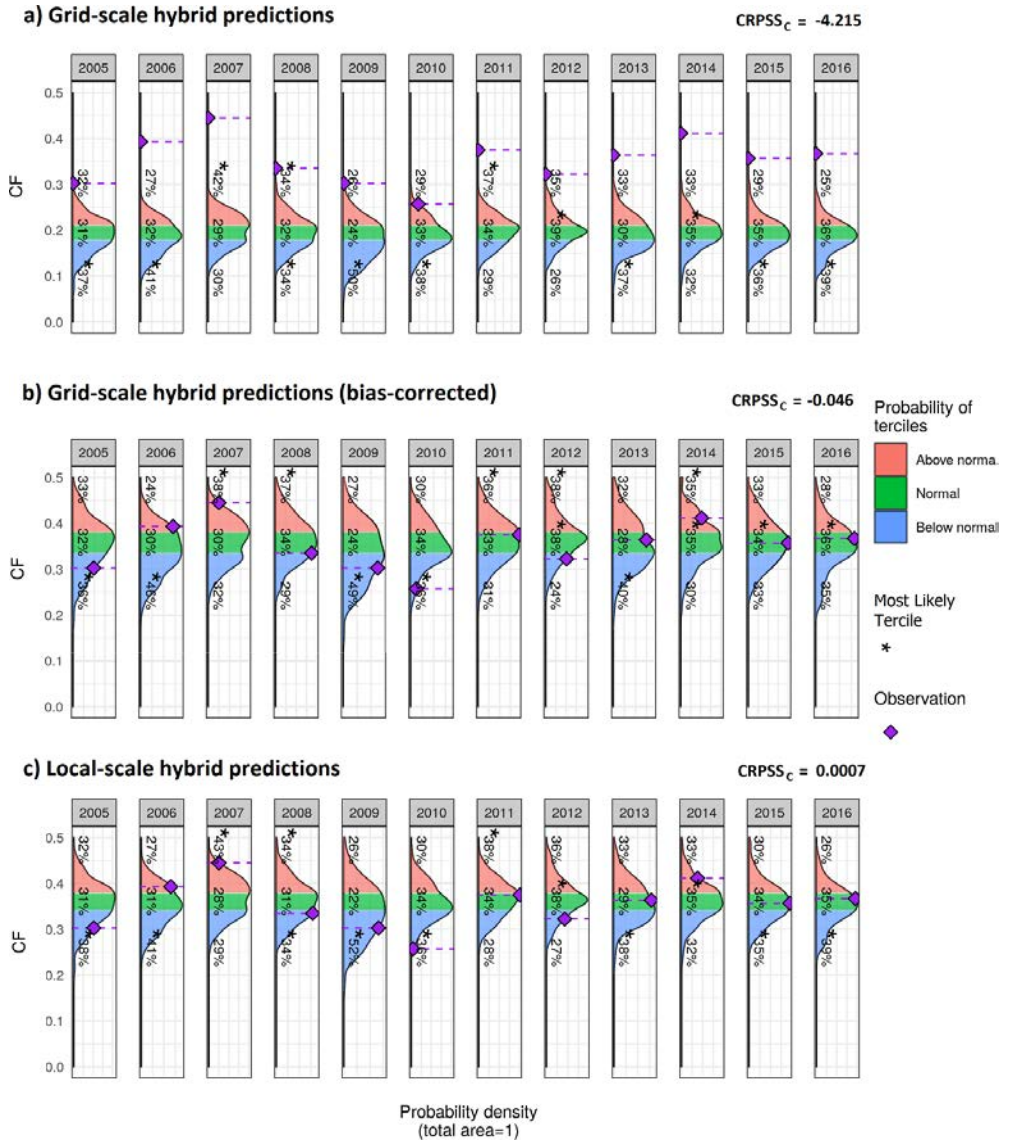


Figure 8. Lead-three hybrid predictions generated using (a) grid-scale data, (b) grid-scale data followed by a bias-correction with local observations, and (c) local-scale observations at Puijo tall tower [27.65° E, 62.91° N]. The multi-system predictions are presented in the form of probabilities for the three tercile categories for the 2005–2016 period. The $CRPSS_c$ values are calculated over the 1993–2016 hindcast period. The probability density functions have been built by dressing the ensemble members with the kernel density estimate method.

the status of the EATCs. This provides interpretability of the results, which enables users to make more informed decisions. For example, one can link higher winds across the UK and the North Sea to a positive NAO phase.

The wind power industry is one of the potential users that can profit most from hybrid predictions. Wind and CF forecasts have been proven to offer useful results at a wind farm scale, provided that site observations from a met mast are available. Moreover, the skilfulness is not restricted to the shortest leads—as it is often the case of the dynamical forecasts—but hybrid predictions issued two or three months in advance can already anticipate understanding of the conditions for the coming season.

The PP is a simple and effective approach but also suffers from some limitations. For instance, the proposed hybrid model does not account for the biases in the EATC predictions. Future work may look into existing post-processing methods like calibration techniques to bias-correct the model output. Besides, the optimal number of EATCs employed to explain the wind variability can be tuned for each region as in Bastien (2018) (chapter 3), who found varying results over France. Investigating whether these improvements lead to a marginal or substantial increase in

skill would be valuable for any potential user of the hybrid predictions.

Data availability statement

Most of the data used in this article can be accessed from publicly available sources: Climate Data Store and Tall Tower Dataset.

Acknowledgments

Authors acknowledge the funding support from project INDECIS co-funded by the H2020 ERA-net ERA4CS (GA 690462) and the MICINN Grant No. BES-2017-082216 ('Ayudas para contratos predoctorales'). The Copernicus Climate Change Service (C3S) is also acknowledged for providing seasonal predictions from several European meteorological centres, and the ERA5 reanalysis. We also thank the developers of the R packages SpecsVerification, easyVerification and CSTools. Finally, we want to express our gratitude to the 17 tall tower data providers, namely, the CESAR observatory, the Crown State, Met Éireann, the BMWi (Bundesministerium fuer Wirtschaft und Energie, Federal Ministry for Economic Affairs and Energy), the PTJ (Projekttraeger Juelich, project executing organization); and Principal Investigators Dr Fred Bosveld, Dr Ingo Lange, Dr Laszlo Haszpra, Dr Jan Schween and Dr Frank Beyrich. This article benefited from useful comments and feedback from three anonymous reviewers.

ORCID iDs

Jaume Ramon https://orcid.org/0000-0003-2818-5206
Llorenç Lledó https://orcid.org/0000-0002-8628-6876

References

Anders A M, Roe G H, Hallet B, Montgomery D R, Finnegan N J and Putkonen J 2006 Spatial patterns of precipitation and topography in the Himalaya *Geol. Soc. Am.* **398** 39–53
Athanasiadis P J, Bellucci A, Scaife A A, Hermanson L, Materia S, Sanna A, Borrelli A, MacLachlan C and Gualdi S 2017 A multisystem view of wintertime NAO seasonal predictions *J. Clim.* **30** 1461–75
Baker L H, Shaffrey L C and Scaife A A 2017 Improved seasonal prediction of UK regional precipitation using atmospheric circulation *Int. J. Climatol.* **38** e437–53
Bastien A 2018 Seasonal forecasting of wind energy resource and production in France, and associated risk PhD Thesis
Buontempo C *et al* 2018 What have we learnt from EUPORIAS climate service prototypes? *Clim. Serv.* **9** 21–32
Carta Je A, Velázquez S and Cabrera P 2013 A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site *Renew. Sustain. Energy Rev.* **27** 362–400
Clark R T, Bett P E, Thornton H E and Scaife A A 2017 Skilful seasonal predictions for the European energy industry *Environ. Res. Lett.* **12** 119602
Deutscher Wetterdienst 2019 Seasonal forecasting with the german climate forecast system (available at: https://www.dwd.de/EN/ourservices/seasonals_forecasts/project_description.html?nn=641552&lsbId=619784)
Diebold F X and Mariano R 1995 Comparing predictive accuracy *J. Bus. Econ. Stat.* **13** 253–63
Doblas-Reyes F J, Hagedorn R and Palmer T N 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting–II. Calibration and combination *Tellus* A **57** 234–52
Dorel L, Ardilouze C, Déqué M, Batté L and Guérémy J F 2017 Documentation of the METEO-FRANCE pre-operational seasonal forecasting system METEO-FRANCE *Technical Report* (available at: http://seasonal.meteo.fr/sites/data/Documentation/doc_modele/Model_MF-S6_C3S_technical_en.pdf)
Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L and Robinson N 2014 Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.* **41** 5620–8
García-Díez M, Fernández J, San-Martín D, Herrera S and Gutiérrez J M 2015 Assessing and improving the local added value of WRF for wind downscaling *J. Appl. Meteorol. Climatol.* **54** 1556–68
Glahn H R and Lowry D A 1972 The use of model output statistics (MOS) in objective weather forecasting J. Appl. Meteorol. **11** 1203–11
Gutierrez J M, Bedia J, Benestad R and Pagé C 2013 Local predictions based on statistical and dynamical downscaling *Technical Report* 308378 (SPECS)
Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1–51
IEC 2017 International Standard—IEC61400-12-1 *Technical Report* (International Electrotechnical Commission) (available at: https://webstore.iec.ch/publication/26603)
James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* (*Springer Texts in Statistics* vol 103) (New York: Springer)
Johnson S J *et al* 2019 SEAS5: the new ECMWF seasonal forecast system *Geosci. Model Dev.* **12** 1087–117
Jolliffe I T and Primo C 2008 Evaluating rank histograms using decompositions of the Chi-square test statistic *Mon. Weather Rev.* **136** 2133–9
Jolliffe I T and Stephenson D B 2012 *Forecast Verification* (Oxford: Wiley)
Kämäräinen M, Uotila P, Karpechko A Y U, Hyvärinen O, Lehtonen I and Räisänen J 2019 Statistical learning methods as a basis for skillful seasonal temperature forecasts in Europe *J. Clim.* **32** 5363–79
Klein W H, Lewis B M and Enger I 1959 Objective prediction of five-day mean temperatures during winter *J. Meteorol.* **16** 672–82
Leskinen A, Portin H, Komppula M, Miettinen P, Arola A, Lihavainen H, Hatakka J, Laaksonen A and Lehtinen K E J 2009 Overview of the research activities and results at Puijo semi-urban measurement station *Boreal Environ. Res.* **14** 576–90 (https://helda.helsinki.fi/handle/10138/233514)
Lledó L, Cionni I, Torralba Vonica, Bretonnière P-A and Samsó M 2020 Seasonal prediction of Euro-Atlantic teleconnections from multiple systems *Environ. Res. Lett.* **15** 074009
Lledó L, Torralba V, Soret A, Ramon J and Doblas-Reyes F J 2019 Seasonal forecasts of wind power generation *Renew. Energy* **143** 91–100
Maclachlan C *et al* 2015 Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system *Q. J. R. Meteorol. Soc.* **141** 1072–84
Manzanas R, Gutiérrez J M, Bhend J, Hemri S, Doblas-Reyes F J, Torralba V, Penabad E and Brookshaw A 2019 Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset *Clim. Dyn.* **53** 1287–305
Manzanas R, Lucero A, Weisheimer A and Gutiérrez J M 2018 Can bias correction and statistical downscaling methods improve

the skill of seasonal precipitation forecasts? *Clim. Dyn.* **50** 1161–76
Marzban C, Sandgathe S and Kalnay E 2006 MOS, perfect prog and reanalysis *Mon. Weather Rev.* **134** 657–63
Mason S J 2004 On using 'climatology' as a reference strategy in the Brier and the ranked probability skill scores *Mon. Weather Rev.* **132** 1891–5
Mason S J 2018 Guidance on verification of operational seasonal climate forecasts *Technical Report* (WMO) (available at: www.seevccc.rs/SEECOF/SEECOF-10/SEECOF-LRF-TRAINING/November 13th 2013/CCl verification recommendations.pdf)
Merryfield W J *et al* 2020 Current and emerging developments in subseasonal to decadal prediction *Bull. Am. Meteorol. Soc.* **101** 1–90
Pavan V and Doblas-Reyes F J 2013 Calibrated multi-model ensemble summer temperature predictions over Italy *Clim. Dyn.* **41** 2115–32
Pickering B, Grams C M and Pfenninger S 2020 Sub-national variability of wind power generation in complex terrain and its correlation with large-scale meteorology *Environ. Res. Lett.* **15** 044025
Ramon J, Lledó L, Pérez-Zañón N, Soret A and Doblas-Reyes F J 2020 The tall tower dataset: a unique initiative to boost wind energy research *Earth Syst. Sci. Data* **12** 429–39
Robertson A W, Qian J H, Tippett M K, Moron V and Lucero A 2012 Downscaling of seasonal rainfall over the philippines: dynamical versus statistical approaches *Mon. Weather Rev.* **140** 1204–18
Rust H W, Richling A, Bissolli P and Ulbrich U 2015 Linking teleconnection patterns to European temperature—a multiple linear regression model *Meteorol. Z.* **24** 411–23
Sanna A, Borrelli A, Athanasiadis P J, Materia S, Storto A, Navarra A, Tibaldi S and Gualdi S 2017 RP0285—CMCC-SPS3: the CMCC seasonal prediction system 3 *Technical Report* (Centro Euro-Mediterraneo sui Cambiamenti Climatici) (available at: www.cmcc.it/publications/rp0285-cmcc-sps3-the-cmcc-seasonal-prediction-system-3)
Scaife A A *et al* 2014 Skillful long-range prediction of European and North American winters *Geophys. Res. Lett.* **41** 2514–19
Schwitalla T, Warrach-Sagi K, Wulfmeyer V and Resch M 2020 Near-global-scale high-resolution seasonal simulations with WRF-Noah-MP v.3.8.1 *Geosci. Model Dev.* **13** 1959–74
Torralba V, Doblas-Reyes F J, MacLeod D, Christel I and Davis M 2017 Seasonal climate prediction: a new source of information for the management of wind energy resources *J. Appl. Meteorol. Climatol.* **56** 1231–47
Van Den Dool H M and Toth Z 1991 Why do forecasts for 'near normal' often fail? Weather Forecast. **6** 76–85
Williams K D *et al* 2015 The met office global coupled model 2.0 (GC2) configuration *Geosci. Model Dev.* **8** 1509–24
WMO 2020 *Guidance on Operational Practices for Objective Seasonal Forecasting* 1246 (available at: https://library.wmo.int/doc_num.php?explnum_id=10314)
Yang W, Foster K, Llorenç Lledó, Torralba V, Cortesi N, Schaller N, Cionni I, De Felice M, Brayshaw D and Bloomfield H 2020 Modes of variability in Europe and their impact on the energy indicators S2S4E project (available at: https://s2s4e.eu/sites/default/files/2020-06/s2s4e_d32.pdf)
Zubiate L, McDermott F, Sweeney C and O'Malley M 2017 Spatial variability in winter NAO—wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns *Q. J. R. Meteorol. Soc.* **143** 552–62

Supplementary Information

Table S1: EMC of the SCA index dynamical forecasts for DJF and the corresponding observed (ERA5) values for each system and lead time in the 1993–2016 hindcast period. See Lledó et al. (2020) for complete information on the skill of the remaining EATCs.

System	Lead month 0	Lead month 1	Lead month 2	Lead month 3
DWD2	0.65	0.33	0.40	0.08
GS5GC2	0.53	0.30	0.50	0.30
MF6	0.59	0.35	0.43	0.52
SEAS5	0.57	0.21	0.25	0.40
SPS3	0.61	0.20	-0.03	-0.17
multi-system	0.71	0.39	0.42	0.32

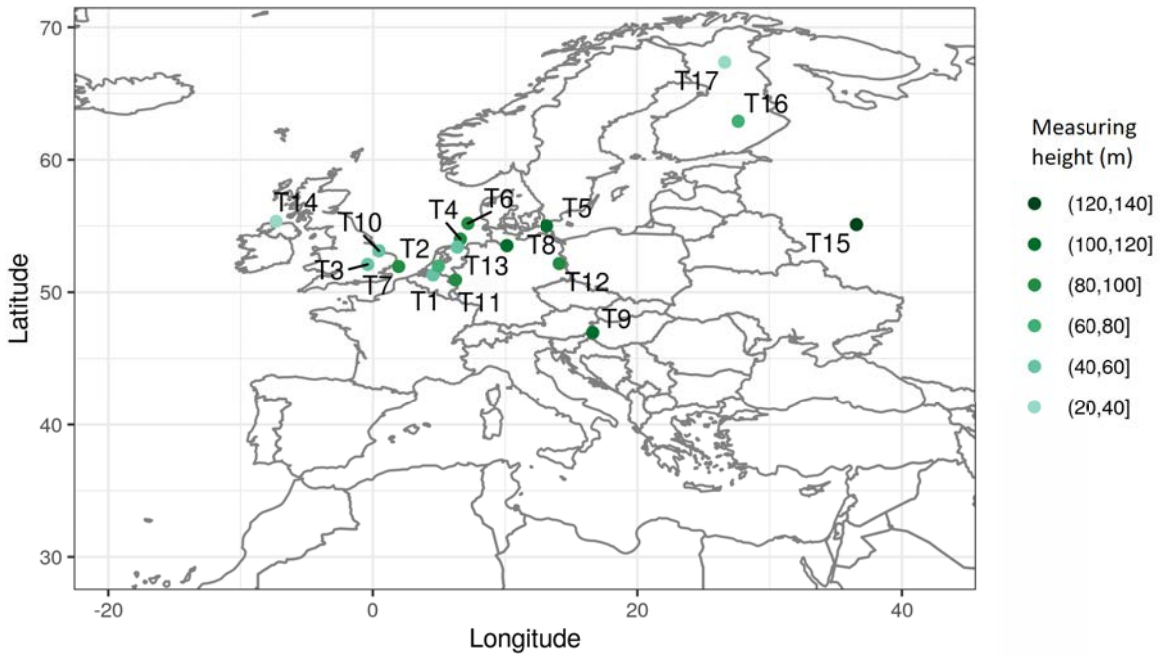


Fig. S1: Spatial distribution of the 17 tall towers considered for this study. The anemometer measuring height is also indicated.

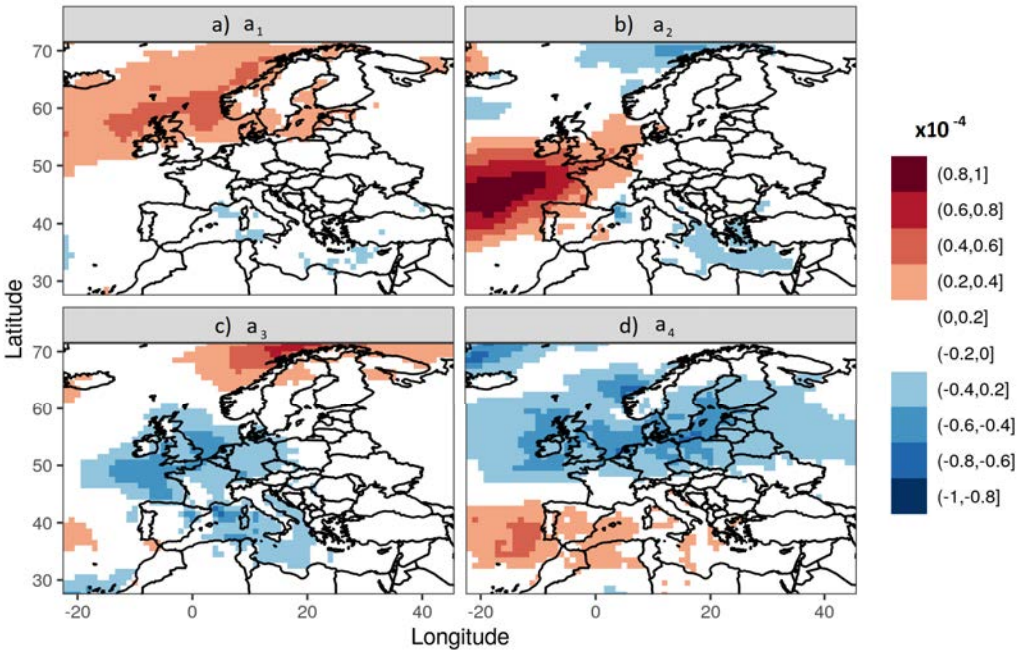


Fig. S2: Spatial distribution of the coefficients a_i; $i = 1, 2, 3, 4$ of the multilinear regression (Equation 1). The coefficients provide an estimate of the weight/importance of each EATC in the hybrid model. In particular, the four coefficients go with (a) NAO, (b) EA, (c) EAWR, (d) SCA, respectively.

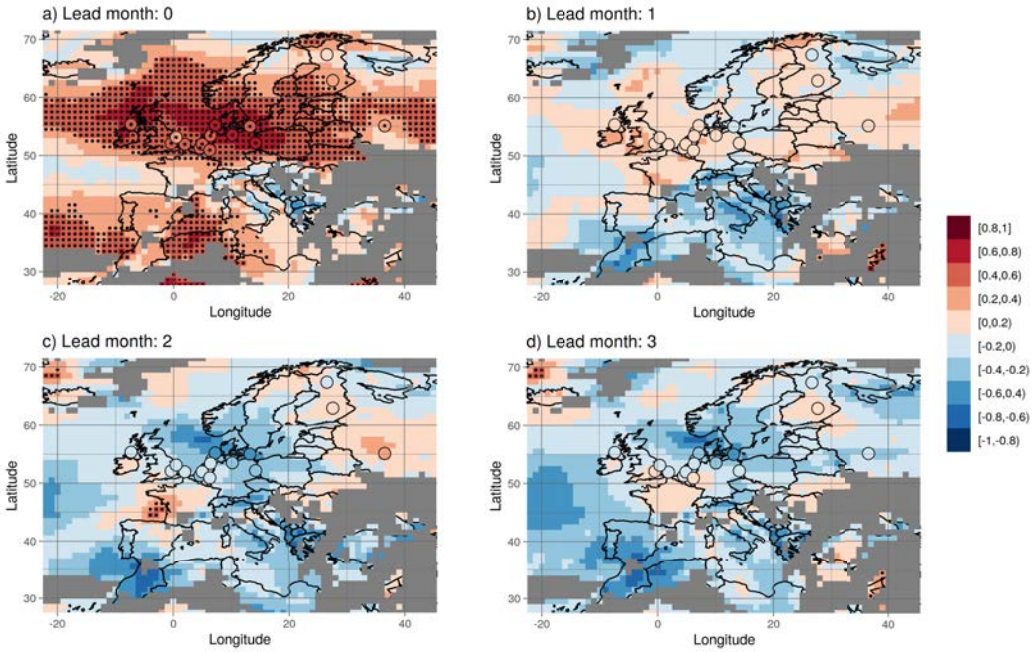


Fig. S3: EMC of the multi-system dynamical near-surface wind speed predictions over Europe for DJF. Predictions were initialised in a) December (lead month 0), b) November (lead month 1), c) October (lead month 2) and d) September (lead month 3). Filled points indicate the EMC of the predictions at the 16 tall tower locations. Black dots highlight areas statistically significant at the 95% of confidence level. Grey-masked areas indicate regions where the R^2 of the multi-linear regression is below 0.3.

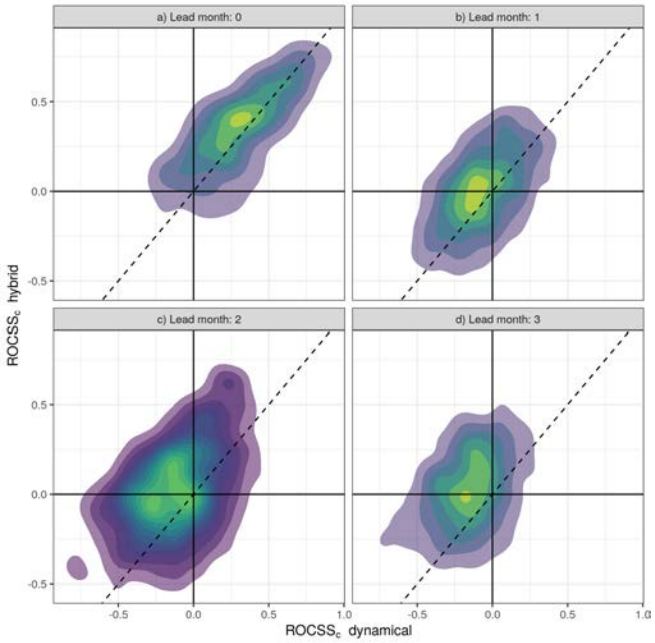


Fig. S4: Two-dimensional density plots showing the ROCSS$_c$ of the upper-tercile hybrid and dynamical predictions from the multi-system at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of ROCSS$_c$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in a) December (lead month 0), b) November (lead month 1), c) October (lead month 2) and d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

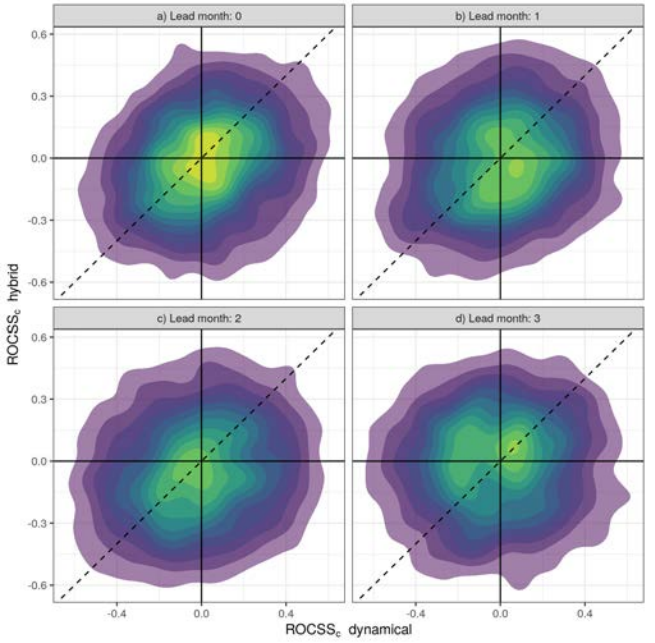


Fig. S5: Two-dimensional density plots showing the ROCSS$_c$ of the central-tercile hybrid and dynamical predictions from the multi-system at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of ROCSS$_c$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in a) December (lead month 0), b) November (lead month 1), c) October (lead month 2) and d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

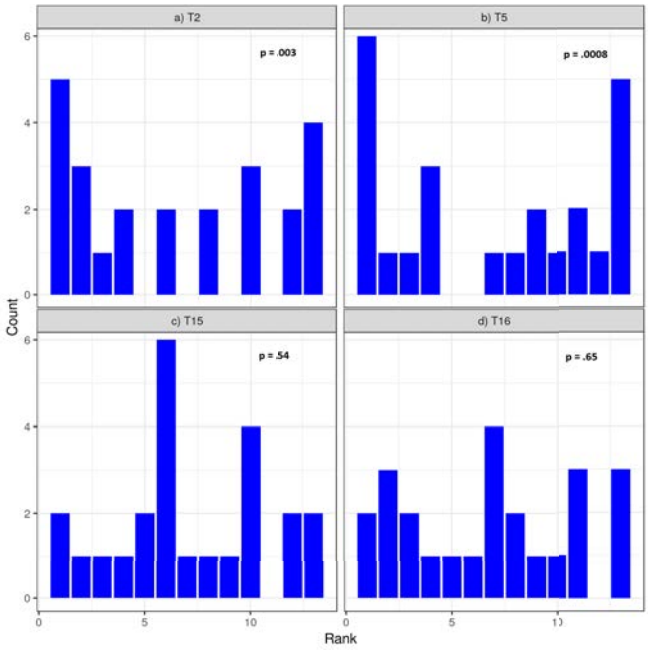


Fig. S6: Rank histograms of the lead-zero SEAS5 near-surface wind speed hybrid predictions for DJF at a) T2 [Cabauw; 4.92°E, 51.97°N], b) T5 [Fino2; 13.15°E, 55.01°N], c) T15 [Obninsk; 36.60°E, 55.11°N] and d) T16 [Puijo; 27.65°E, 62.91°N] tall tower locations. The p-value of the Jolliffe-Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

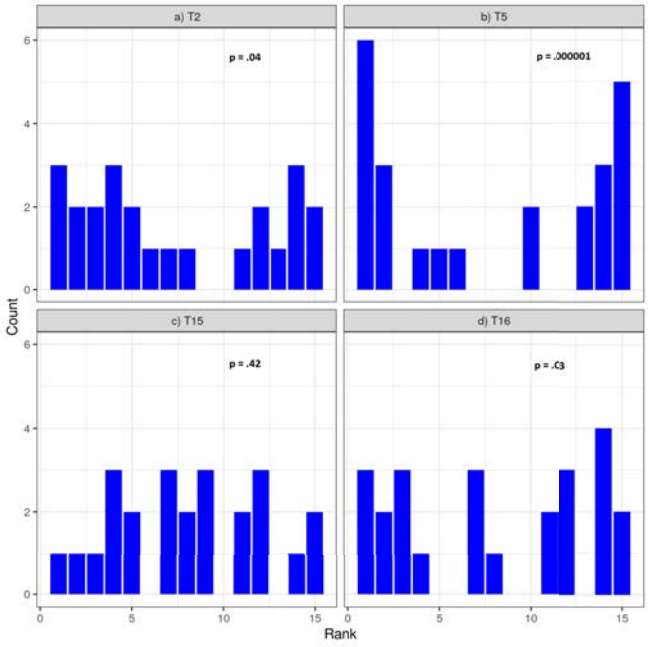


Fig. S7: Rank histograms of the lead-zero DWD2 near-surface wind speed hybrid predictions for DJF at a) T2 [Cabauw; 4.92°E, 51.97°N], b) T5 [Fino2; 13.15°E, 55.01°N], c) T15 [Obninsk; 36.60°E, 55.11°N] and d) T16 [Puijo; 27.65°E, 62.91°N] tall tower locations. The p-value of the Jolliffe-Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

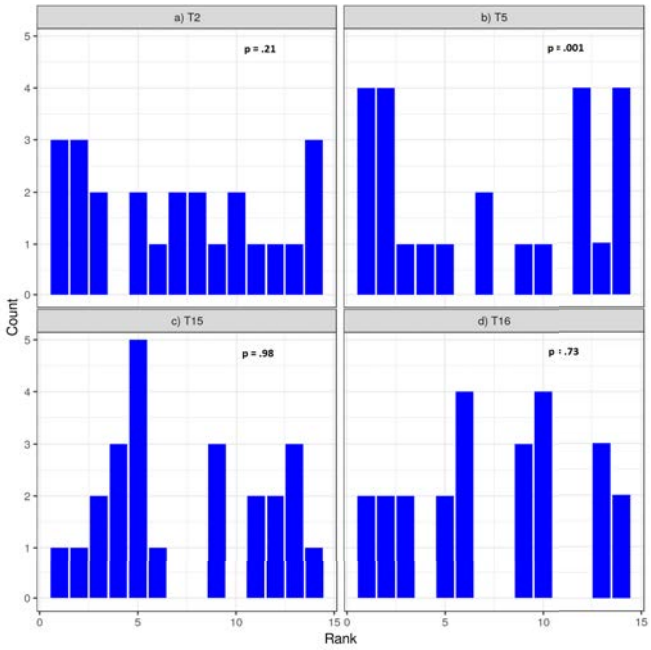


Fig. S8: Rank histograms of the lead-zero GS5GC2 near-surface wind speed hybrid predictions for DJF at a) T2 [Cabauw; 4.92°E, 51.97°N], b) T5 [Fino2; 13.15°E, 55.01°N], c) T15 [Obninsk; 36.60°E, 55.11°N] and d) T16 [Puijo; 27.65°E, 62.91°N] tall tower locations. The p-value of the Jolliffe-Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

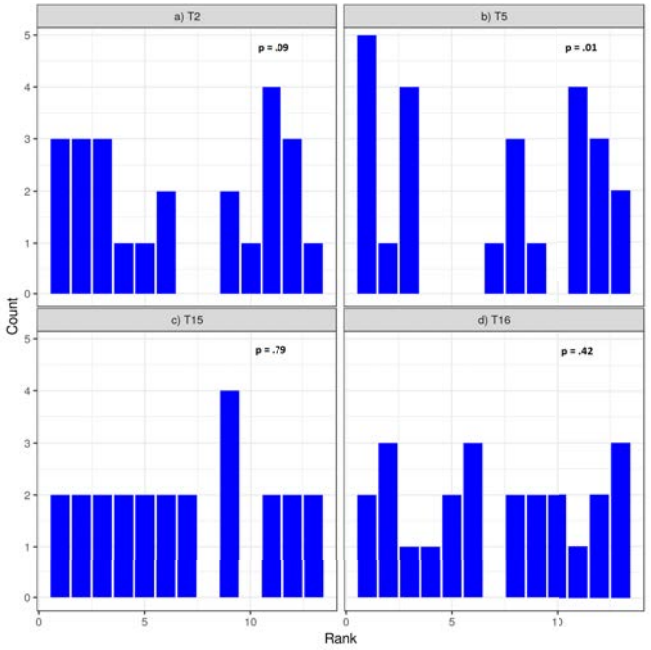


Fig. S9: Rank histograms of the lead-zero MF6 near-surface wind speed hybrid predictions for DJF at a) T2 [Cabauw; 4.92°E, 51.97°N], b) T5 [Fino2; 13.15°E, 55.01°N], c) T15 [Obninsk; 36.60°E, 55.11°N] and d) T16 [Puijo; 27.65°E, 62.91°N] tall tower locations. The p-value of the Jolliffe-Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

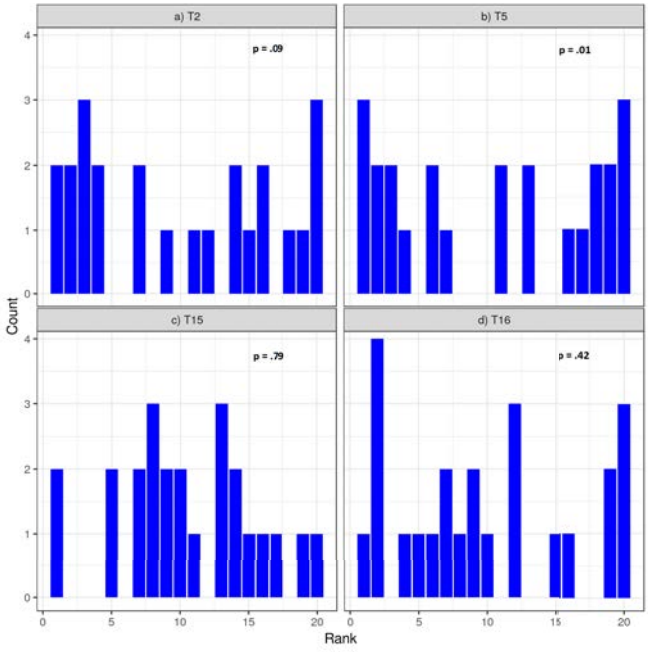


Fig. S10: Rank histograms of the lead-zero SPS3 near-surface wind speed hybrid predictions for DJF at a) T2 [Cabauw; 4.92°E, 51.97°N], b) T5 [Fino2; 13.15°E, 55.01°N], c) T15 [Obninsk; 36.60°E, 55.11°N] and d) T16 [Puijo; 27.65°E, 62.91°N] tall tower locations. The p-value of the Jolliffe-Primo test statistic for convexity under the null hypothesis of a flat rank histogram is indicated.

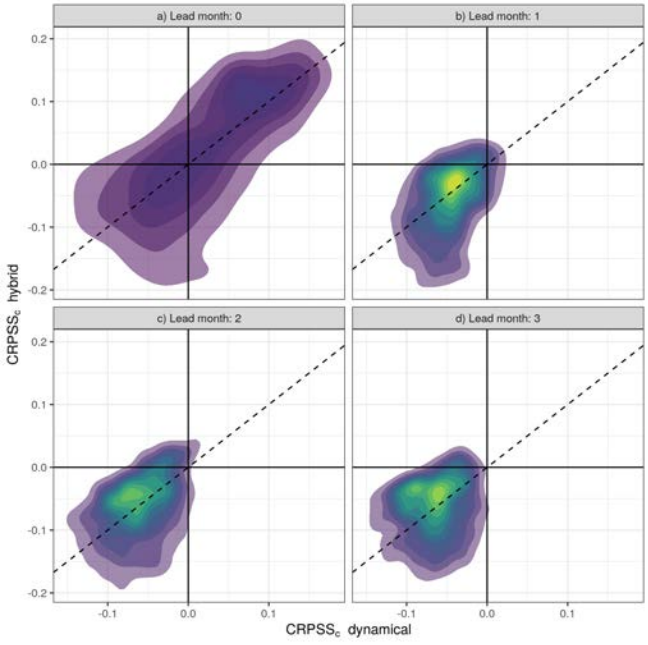


Fig. S11: Two-dimensional density plots showing the CRPSS$_c$ of the hybrid and dynamical predictions from the multi-system at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of CRPSS$_c$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in a) December (lead month 0), b) November (lead month 1), c) October (lead month 2) and d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

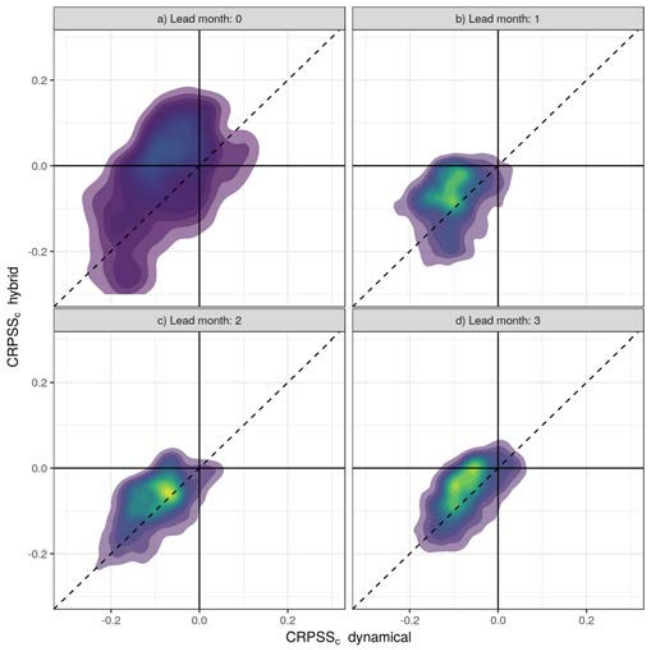


Fig. S12: Two-dimensional density plots showing the $\mathrm{CRPSS_c}$ of the hybrid and dynamical predictions from MF6 at all the grid points within the pan-European domain. On a linear scale, greens represent the highest density of points, whereas purples depict the lowest density estimates. The region above the $y = x$ dashed line indicates an improvement of $\mathrm{CRPSS_c}$ of the hybrid forecast over its dynamical counterpart. Predictions were initialised in a) December (lead month 0), b) November (lead month 1), c) October (lead month 2) and d) September (lead month 3). Grid points where the R^2 of the multi-linear regression is below 0.3 have not been included.

Chapter 7

Prediction skill inherent to observation error

Objective

Investigate the dependencies of the seasonal prediction skill on the choice of the observational reference dataset. A wide variety of skill scores are defined in a way that the forecast probabilities are contrasted against observations, being these reference observations regarded as perfect. However, recent studies such as those included in Chapters 4 and 5 of this thesis report certain levels of error affecting these observations. Is this observation error sufficient to impact the prediction skill estimates? Is it possible to achieve skill values that are independent of the choice of the reference dataset?

Methodology

- Qualitatively see the impacts of observation error through the ranking of a set of seasonal prediction systems.
- Apply a methodology to obtain unbiased estimates of the prediction skill (Brier Score).
- Compare against the traditional approach that does not account for observation error.

Results

- When observation error is not considered, the Brier Score values are highly dependent on the choice of the reference dataset.
- It may be well impossible to recognise the best prediction system among a set of candidates if observation error is not taken into account, or is not properly adjusted.
- Achieving unbiased Brier Score values is possible, as long as observation error is known.

Conclusions

- We recommend employing the error-corrected scoring rules for categorial predictands.
- Uncertainty estimates are valuable for many applications so they should be included in future observational datasets.

Publication

- The work presented in this chapter is a potential future publication to be submitted on a peer-reviewed scientific journal.

Uncertainty in the observational reference: implications in skill assessment and model ranking

Jaume Ramon[a], Llorenç Lledó[a], Christopher A.T. Ferro[b] and Francisco J. Doblas-Reyes[a,c]

[a]Barcelona Supercomputing Center (BSC), c/ Jordi Girona, 29, Barcelona 08034, Spain
[b]University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE United Kingdom
[c]ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain

Abstract. The probabilistic skill of seasonal prediction systems is often inferred using reanalysis data, assuming these benchmark observations to be error-free. However, increasing studies report non-negligible levels of uncertainty affecting reanalysis observations, especially when it comes to variables like precipitation or wind speed. This study explores how this observational uncertainty affects the skill assessment of seasonal predictions for wind speed. Two different methodologies that take into account observational uncertainty are considered: the observational probability method and the error-corrected scoring rules, the latter extended to the particular case of categorical predictands constructed by dichotomization. These two approaches are compared against the traditional procedure of not accounting for observation error. While Brier Score values depend on the choice of the observational reference if the observational probability method or the traditional approach are used, unbiased outcomes are obtained when the error-corrected Brier Score is utilised. We emphasise the perils and quantify the error committed when the observational reference, either reanalysis or point dataset, is selected arbitrarily for verifying a seasonal prediction system.

Keywords. Model scoring, observation error, Brier Score, seasonal prediction, wind speed.

7.1 Introduction

Climate predictions on seasonal time scales are becoming increasingly skilful mostly due to the recent advances in understanding of climate processes and their modelling (Merryfield et al., 2020). Still, the quality of such predictions is far from that of weather forecasts, whose higher quality allows for issuing timely early warnings for extreme events and thus reduce human and material losses. Though some of high-impact climate events can also be correctly anticipated by seasonal predictions, there is a general reluctance to make any decision based upon them.

Generally, seasonal predictions for essential climate variables (e.g., temperature, precipitation or wind speed) are provided together with a

measure of their quality, or skill, which is determined after a comparison against climate observations. Past forecasts of a prediction system are statistically assessed to judge how good that system is, or which is the "best" system among a set of candidates. Besides quality, the goodness of a forecast can also be measured by their consistency and value (Murphy, 1993), but these two aspects will not be focus of this work.

On a typical skill assessment, reference observations are considered as the 'ground truth', unquestionably exact. However, these observations might have some degree of uncertainty, especially when they come from processed datasets such as reanalyses. Reanalysis products provide gridded observational data resulting from a combination of a numerical weather prediction model and the assimilation (ingestion) of past observations from several sources. In this regard, random uncertainties in the observations (e.g. measuring errors) mix with the uncertainty coming from the numerical model (e.g. parametrizations and systematic model errors), and both propagate into the reanalysis response values.

Some authors, however, argue that the uncertainty in the observational reference (whatever type of data, not necessarily reanalysis) is significantly low when compared to the uncertainty in the seasonal prediction (Santos and Ghelli, 2012; Saetra et al., 2004). This is true, in principle, for the furthest forecast horizons (say beyond five or six months) where skill levels decay significantly and seasonal predictions are much less useful.

At shorter forecast times, current research shows that these effects are not negligible at all. What is more, different quality outcomes can be obtained depending on which observational reference has been utilised. Juricke et al. (2018) evaluated the observational uncertainty by assessing monthly and seasonal forecasts against two reanalyses with different horizontal resolutions (1° and .25°). Results revealed a strong dependence of the Brier Skill Score and reliability on the reference data. Similarly, Sunyer et al. (2013) and Gómez-Navarro et al. (2012) ranked different Regional Climate Models from best to worse using varied metrics such as the mean bias or spatial correlation. They both conclude that for most of the metrics analysed, the performance of a model is sensitive to the choice of the observational reference, and so are the scoring rankings obtained. Reichler and Kim (2008) went beyond that and showed that the error in the reanalyses can exceed that of a climate model. They refer to the fact that the driving models and assimilation schemes in some of the currently used reanalyses were designed many years ago, so that they can become deprecated nowadays. For example, the NCEP-NCAR R1 (Kalnay et al., 1996) reanalysis is still widely used, but was generated more than 20 years ago and remains unaltered ever since.

Apart from considering several data sources as a reference, the observational uncertainty can

be inferred using other approaches. For example, Massonnet et al. (2016) tackle the problem with an opposite perspective by assessing different observational datasets against a set of climate models used as a reference. Other authors (Ben-Bouallègue, 2020; Candille and Talagrand, 2008) studied and recommended the so-called *perturbed-ensemble* approach, which randomly perturbs the predicted ensemble of a probabilistic prediction with a noise meant to simulate the uncertainty in the observations. Nevertheless, it requires for making arbitrary assumptions on the magnitude and distribution of such noise. Finally, other approaches consider the observation as a probability distribution (Candille and Talagrand, 2008), or even propose a method to re-build the scoring rules to make them account for observation error (Ferro, 2017).

Most of these approaches have already been tested either with synthetic experiments or by adding arbitrary errors to the models or observations. However, newly-developed global reanalyses contain ensemble members that already allow for inferring estimates on the observation uncertainty so that the last two approaches (hereafter referred to as CT08 and F17, respectively) can be easily implemented without making further assumptions.

7.2 Scoring rules that take into account observational uncertainty: CT08 versus F17

In this section, we summarise two methodologies proposed to account for observation error in skill assessment. The methodologies are those in CT08 and F17, with a new extension for the latter approach. The objective of this work is to quantify the impact of the observational uncertainty in forecast verification by employing CT08 and F17 approaches with state-of-the-art climate data. Sections 7.3 and 7.4 list the employed datasets and present the methodology, respectively. Results are described in Section 7.5 and further discussed in Section 7.6. Finally, conclusions are drawn in Section 7.7.

The 'observational-probability' method introduced in CT08 considers the verifying observation, y, as a probability distribution, g, which is then compared against the probability distribution defined by the predicted ensemble, f. Therefore, the observed probabilities are no longer restricted to 0 or 1, but they take values within $[0, 1]$. Scoring rules such as the Brier Score (BS):

$$BS_{REF}(f, y) = (f - y)^2, \tag{7.1}$$

remain defined in this method, and are algebraically identical to their original definition:

$$BS_{CT08}(f, g) = (f - g)^2 \tag{7.2}$$

Nevertheless, some authors such as F17 consider the CT08 approach as inappropriate, since the new scoring rules are no longer proper (see definition of 'proper scoring rule'

in Jolliffe and Stephenson (2012) Sect. 2.7), which implies that better forecasts of the truth are not always rewarded. Furthermore, F17 argues that CT08's scoring rules appear biased under the observation model, indicating that the expected value of the score —the metric that determines the overall quality of a forecast— obtained using the CT08 method is not generally the same that would be obtained if the true value was known. It is therefore desirable that the expected values of the skill scores remain unaffected by observation error.

F17 proposes a general method to construct proper scoring rules that are both proper and unbiased. These error-corrected scoring rules account for observation error in a way that forecasters who issue better forecasts of the truth are favoured. In particular, for categorical predictands, the new BS is:

$$BS_{F17}(f,y) = (f-y)^2 + \\ + r_y \frac{(f-y)^2 - (f-1+y)^2}{1-r_0-r_1} \tag{7.3}$$

Where y is the observed probability which, albeit uncertain, takes the unique values 0 and 1. x refers to the true value of the observation, which is generally unknown. The observation error is included in the so-called misclassification probabilities $r_0 = Pr(y = 1 \mid x = 0)$ and $r_1 = Pr(y = 0 \mid x = 1)$. After performing the adjustment, F17 notes that BS_{F17} may happen to be negative, which is mathematically meaningless. In such cases, F17 recommends truncating the reported BS_{F17} to zero, but warns of the loss of both propriety and unbiasing conditions.

There are some particular cases for which y is not conditionally independent of f given x which, equivalently, means that both r_0 and r_1 depend of f. Such cases can arise when categorical predictands are created by dichotomizing numerical predictands, a common way of delivering the predictions at the seasonal timescale. Seasonal predictions are most commonly probabilistic forecasts derived by thresholding so that Equation 7.3 is no longer valid as it is. Instead, a slight modification has to be introduced here, i.e. the extended error-corrected BS needs to consider the dependencies $r_0(f)$ and $r_1(f)$:

$$BS_{F17-E}(f,y) = (f-y)^2 + \\ + r_y(f) \frac{(f-y)^2 - (f-1+y)^2}{1-r_0(f)-r_1(f)} \tag{7.4}$$

While the BS_{F17-E} is no longer proper (C. Ferro, personal communication), we can still achieve unbiased estimates of the score using the extension in Equation 7.4.

7.3 Datasets

The hindcasts for surface wind speeds from five different operationally-produced seasonal prediction systems have been used in this study (see details on Table 7.1). All five hindcasts have been retrieved from the Climate Data

Store[1] (CDS) data portal in a regular grid of 1°x1°, at 6-hourly time resolution and spanning the 1993–2016 period. These hindcasts will be referred to as predictions or forecasts throughout the text. Surface wind speeds from the five prediction systems have been derived from the zonal and meridional components at 6-hourly resolution. Then, monthly and seasonal averages have been prepared. Lead-zero predictions, i.e., those initialised at the beginning of the forecasted month or season, have been considered in this study.

Eight observational references providing surface wind speeds have been considered to evaluate the set of predictions at a monthly and seasonal timescales. Specific details can be found in Table 7.2. The set of datasets comprise reanalyses (i.e., ERAI, JRA55, ERA5-HRES, MERRA2 and R1, see Ramon et al. (2019) for a review), station-based datasets (i.e., HadISD), reanalyses with ensemble members (i.e., ERA5-EDA) and a multi-reanalysis (MR). All datasets cover the 1993-2016 period.

HadISD version 2.0.2.2017f (Dunn et al., 2014) is a station-based dataset consisting of 8103 stations spanning from 1931 to the end of 2017. A wide variety of quality-controlled climate variables is freely accessible at https://www.metoffice.gov.uk/hadobs/hadisd/.
Zeng et al. (2019) filtered all these 8103 stations and selected only those that have that have uninterrupted, continuous monthly records during the 1978–2017 period, thus assuring a top-level quality for the monthly mean wind speeds in the 1993-2016 period. The process left a sample of 1542 stations, which are located mainly inland, but covering the whole globe.

The newly released reanalysis ERA5-EDA (or EDA for simplicity) contains 10 members (one control plus nine perturbed members) aiming to represent estimates of analysis and short-range forecast uncertainty at each grid point (Hersbach et al., 2020). However, authors in Hersbach et al. (2020) warn of fact that this uncertainty does not represent the total observation error. Even though the generation of EDA takes into account random uncertainties in the observations, sea surface temperature and the physical parametrisations of the model, other sources of uncertainty such as systematic model errors are not considered.

Aiming to provide a more accurate representation of the total observation error, an extra observational reference has been generated in this study, i.e. the MR. Differences between different reanalyses have already been studied for surface and near-surface winds, and have been shown not to be negligible (Ramon et al., 2019). These discrepancies quantified as the spread of the sample can provide a lower bound estimate of the total observational uncertainty (Parker, 2016; Bellprat et al., 2017). The MR uses ERAI, JRA55, HRES, MERRA2 and R1 as source data and its generation is presented in the following section.

[1] cds.climate.copernicus.eu

Table 7.1: Specific details of the seasonal prediction systems employed in this study

System	Institution	Ensemble members	Reference
SEAS5	ECMWF[a]	25	Johnson et al. (2019)
SPS3	CMCC[b]	40	Sanna et al. (2017)
DWD2	Deutscher Wetterdienst	30	Deutscher Wetterdienst (2019)
GS5GC2	UK Met Office	28	Maclachlan et al. (2015)
MF6	Météo France	25	Dorel et al. (2017)

[a] European Centre for Medium-Range Weather Forecasts

[b] Centro Euro-Mediterraneo sui Cambiamenti Climatici

7.4 Methodology

7.4.1 Computation of the probabilities

Predictions are evaluated at a subset of the 1542 HadISD point locations (Figure 7.1): stations where seasonal predictions do not show skill, using HadISD as reference data, are filtered out to better visualise the effects of accounting for observation error. We note that the filtering depends on the type of the prediction (i.e., monthly or seasonal) as well as on the considered prediction system(s). Gridded data —both predictions and reanalysis observations— are interpolated bilinearly to each of the stations. Once interpolated, predicted and observed probabilities are obtained. The present work focuses on events exceeding the 90^{th} percentile, being the percentiles computed using the climatological distribution generated by each dataset separately. A probability of 1 is assigned when winds exceed the 90^{th} percentile, 0 otherwise. In case of having ensemble members, the resulting probability of occurrence is obtained by averaging all realisations. Both the predicted probabilities, f, and the observed probabilities g are computed in this way.

7.4.2 Generation of the multi-reanalysis

The MR has been produced after pooling ERAI, JRA55, HRES, MERRA2 and R1 data into one dataset. In this way, for each location and month/season, a set of five wind values representing an ensemble distribution is available. Before that pooling, however, the observed probabilities have to be computed. This is because, unlike the five prediction systems or the EDA, the ensemble of the MR has been generated from entirely independent datasets (including NWP models and data assimilation schemes). Therefore, they have differing climatologies, and the observed probabilities have to be computed for each dataset separately prior to pooling the source data into the MR.

Table 7.2: Specific details of the observational datasets employed in this study

Dataset	Institution	Ensemble members	Native time resolution	Horizontal grid spacing	Reference
ERAI	ECMWF[a]	-	6-hourly	0.75° x 0.75°	Dee et al. (2011b)
JRA55	JMA[b]	-	6-hourly	1.25° x 1.25°	Kobayashi et al. (2015)
ERA5-HRES	ECMWF[a]	-	1-hourly	0.3° x 0.3°	Hersbach et al. (2020)
MERRA2	NASA-GMAO[c]	-	1-hourly	0.5° latitude x 0.625° longitude	Molod et al. (2015)
R1	NCEP-NCAR[d]	-	6-hourly	1.875° latitude x 2° longitude	Kalnay et al. (1996)
HadISD (v.2.0.2.2017f)	UK Met Office	-	sub-daily	irregular	Dunn et al. (2014)
ERA5-EDA	ECMWF[a]	10	3-hourly	0.5625° x 0.5625°	Hersbach et al. (2020)
MR	-	5	seasonal	irregular	-

[a] European Centre for Medium-Range Weather Forecasts

[b] Japan Meteorological Agency

[c] National Aeronautics and Space Administration - Global Modeling and Assimilation Office

[d] National Centers for Environmental Prediction - National Center for Atmospheric Research

7.4.3 Preparation of the model rankings

The first part of our work investigates how observational uncertainty changes a set of rankings, in which the five prediction systems are ordered from best to worst based upon their skill. The BS of the 90^{th} percentile (BS90) of each system —lead-zero predictions for the aggregated period December-January-February, DJF— is obtained using all the set of observational references, with and without taking into account observation error. Filtering out stations where at least one of the prediction systems does not provide skill ($BSS90 < 0$) reduces the sample from 1542 to 143 stations. Moreover, since we are comparing the skill of different systems with different number of ensemble members, we consider the fair version of the BS (Ferro, 2014), fBS.

On one hand, when observation error is not accounted for (hereafter referred to as REF), all the references are used without any consideration of their uncertainty —as it would be done in a traditional forecast verification (Jolliffe and Stephenson, 2012). The BS90 is obtained as in

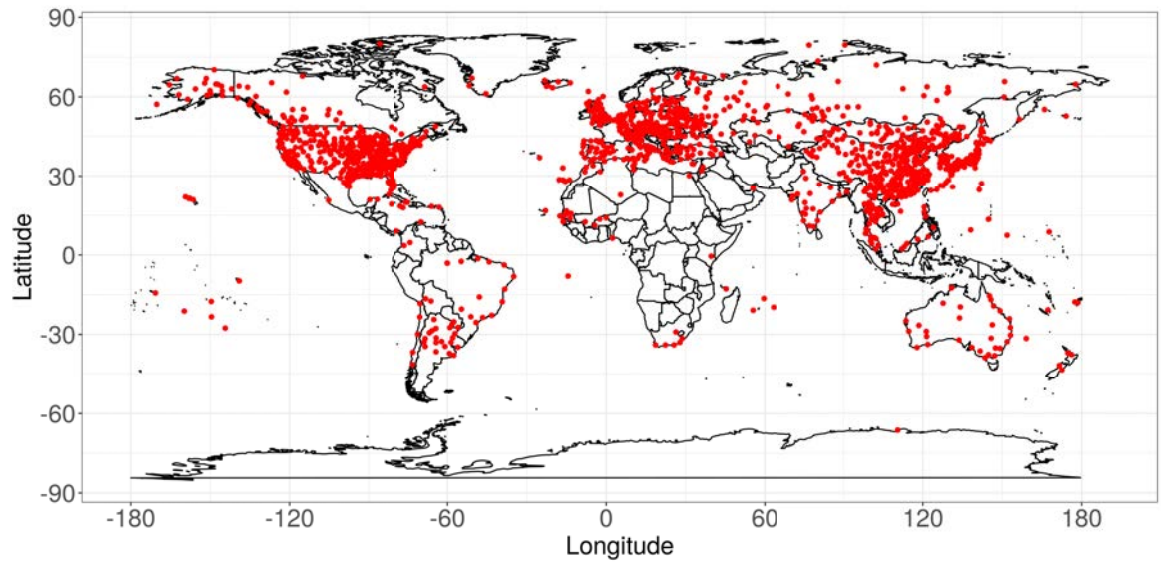


Figure 7.1: Spatial distribution of the 1542 HadISD stations.

Equation 7.1 and later transformed into its fair version, fBS90. When multiple-observations from EDA and MR, y_i, are available the BS90 is obtained for each observation separately and then aggregated into an averaged BS90. Additionally for those references, the 'observational probability' method in CT08 is used (Equation 7.2), employing the ensemble observations to build the the probability distribution of the verifying observation, g.

On the other hand, to systematically account for observation error in all observational references, F17 method is selected assuming the *true* observed value at each location, x, to be that in HadISD. These *true* observations are also utilised to compute $r_0(f)$ and $r_1(f)$, which in this case can be seen as a measure of the quality of the set of reanalyses. To this end, the forecast probabilities, f, are discretised in bins of 0.1. For each bin, the corresponding uncertain observations, y, from all the subset of stations and for each of the reanalyses and ensemble members, are contrasted against the corresponding true values, x.

Since the fair version of the BS preserves propriety, F17 method is still applicable. Firstly, the BS90 is obtained for each fixed f and y (Equation 7.1) and later transformed into its fair version, i.e. fBS90. Then, the extended error-corrected BS (BS_{F17-E}, Equation 7.4) is applied to obtain the score values that account for observation error. It is noted that when ensemble observations are available, y_i, the resulting fBS90 is the average of the fBS90 obtained for each observation separately. Besides, the fBS90 for the climatological forecast is also computed using REF, CT08 and F17. The climatological forecast is used here as a

baseline forecast, allowing the computation of the fair version of the Brier Skill Score, fBSS90.

7.4.4 Observation error in a monthly prediction

The second part of this work quantifies the error committed when verification scores are computed without assuming observation error (i.e., using REF method), or when the observational uncertainty is not properly introduced (i.e., using CT08 with ensemble reanalyses) in the BSS90 of monthly predictions. In particular, we analyse 12 monthly predictions, each one initialised at the beginning of each month of the year, and for lead time zero. To better focus on the effects of the observational uncertainty, only one prediction system, i.e. SEAS5, is analysed. The removal of the stations with no skill is carried out for each month separately, leaving a sample of 738–938 point locations, depending on the analysed month.

7.5 Results

7.5.1 The choice of the observational reference in model scoring

The fBSS90 of five prediction systems has been computed using eight different observational references, and systems have been ranked from best to worse according to their fBSS90 estimates (Figure 7.2). Scoring rankings are different depending on whether the verification takes into account the error in the observational reference.

Indeed, when observation error is not accounted for (Figure 7.2 left), the five prediction systems rank differently depending on which observational reference is used. No consensus is reached about which prediction system is best. For example, SPS3 is best when verified against JRA55, MERRA2 and R1; second best when verified against HRES, EDA, MR and HadISD, and occupies the central position when ERAI is employed. This highlights the dangers of arbitrarily selecting the reference dataset. An evaluation of the best observational reference (as in Ramon et al. (2019)) is therefore essential to obtain accurate conclusions on which product is better, i.e. more accurate, if only one reference is to be used. Comparing the results in Ramon et al. (2019), for instance we see that JRA55, which is not a good observational reference for surface wind speeds, yields systematically very low estimates of the true skill. The relationship between the quality of observational references and the forecast verification results they give has already been studied in Massonnet et al. (2016): prediction systems score better against the most recent and advanced reanalyses.

All rankings from the reanalyses-based references are generally different to those obtained with HadISD, a reference assumed here to be error-free. This mirrors the bias in the fBSS90

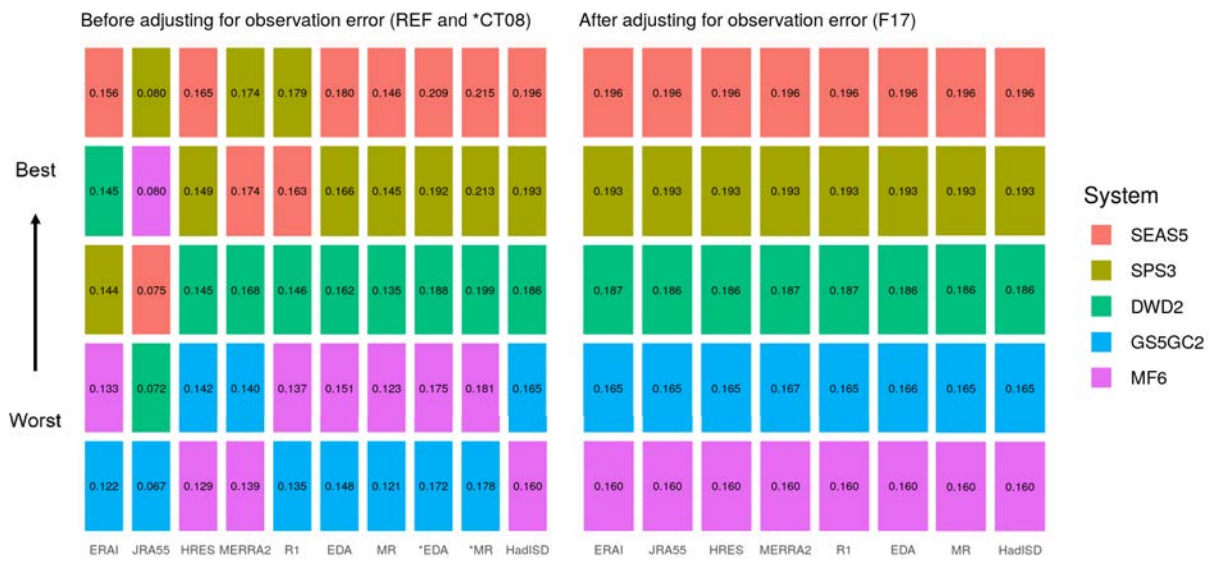


Figure 7.2: Scoring rankings for the fBSS90 of five seasonal prediction systems and eight observational references. All predictions correspond to lead time zero and for December-January-February. On the left, the observational reference values are considered perfect, except for EDA and MR whose uncertainty estimates have been accounted for as in CT08. On the right, the observational error in the BSS90 is considered as in F17, considering HadISD as ground truth. Within a panel, the lower the square, the worse the skill is.

estimates, arising from not accounting for observation error in the reanalysis-based references. When using CT08 with ensemble reanalyses, the score is less biased than the analogous verification with REF, but it is still observed a deviation (positive in this case) of the true skill score. Besides, it is noted that some reanalyses tend to favour forecasts from prediction systems that share some configuration components (e.g. atmospheric model). Such is the case of SEAS5 when assessed against ERAI, HRES and EDA, all of them produced at the ECMWF. In these evaluations, SEAS5 scores best (Figure 7.2 left). For other references like JRA55 or MERRA2, SEAS5's skill appears worse —i.e. second and third best, respectively. Whether this leads to an overestimation of the fBSS90 should be further investigated in future studies. Having said all this, having fBSS90 estimates that are independent of an arbitrary choice of the dataset would ensure fair verification outcomes and comparisons.

When observation error is considered in the error-corrected fBSS90, all scoring rankings are identical whatever the observational reference (Figure 7.2 right) —albeit the skill score values can slightly differ. Whilst SEAS5 stands out as the best prediction system, MF6 shows the lowest skill estimates.

7.5.2 Obtaining unbiased estimates of the prediction skill

The error-corrected BSS90 yields unbiased estimates for the skill of the five systems. Figure 7.3 shows the BSS90 values before and after considering the error-corrected version of the BSS90 for one of the prediction systems, namely SEAS5, for a set of monthly predictions. Firstly, when the error in the reference dataset is not taken into account, or is not properly adjusted, the estimates for the BSS90 appear generally underestimated (Figure 7.3 left). As in Figure 7.2, the highest bias is noted for JRA55, which systematically shows the worst skill estimates for all months. On the contrary, HRES is the single-value reanalyses that shows the closest skill score values to those obtained with HadISD, but still shows a negative bias. This is yet another example which shows that the more advanced products provide the best skill estimates. The discrepancies between BSS90 values obtained by the single-value reanalyses relate to their quality, being the HRES data closest to HadISD (on average) and thus of higher quality.

The skill assessment against the ensemble reanalyses shows slightly different outcomes. The verification of the monthly predictions with CT08 and the MR overestimates the true BSS90. This is also noted for EDA, but only for some specific forecast months (e.g. February and April). These results can be explained either because 1) observation error is not completely/correctly accounted for in the ensemble reanalyses or 2) CT08 methodology fails to adjust observation error properly.

Finally, when all the observational uncertainties are considered by the scoring rules, the discrepancies in the fBSS90 values vanish (Figure 7.3 right).

As stated before, the discrepancies in the BSS90 values arise from the differences in the quality of the observational references. To go into deeper detail, we explore the misclassification rates r_0 and r_1, which are also varied depending on which reference is used (Figure 7.4). These discrepancies increase as f does, since sharp forecasts (high f) are very rare for these monthly predictions. Even for some f, the corresponding values of r_1 and specially r_0 cannot be inferred. There are two reasons for this situation: i) there are no occurrences for some f values (e.g. $f = 1$ in October) and ii) for the highest values of f, it coincides that the event is almost always observed, thus impeding the computation of some values of r_0. Having said that, we can expect the estimates of the misclassification rates for the highest f to be less robust. With regard to the reanalyses individually, we do not spot any of them standing out systematically from the others. Overall, it appears EDA and HRES show the lowest values for r_0 and r_1, respectively, mirroring the good quality of the ERA5 reanalysis.

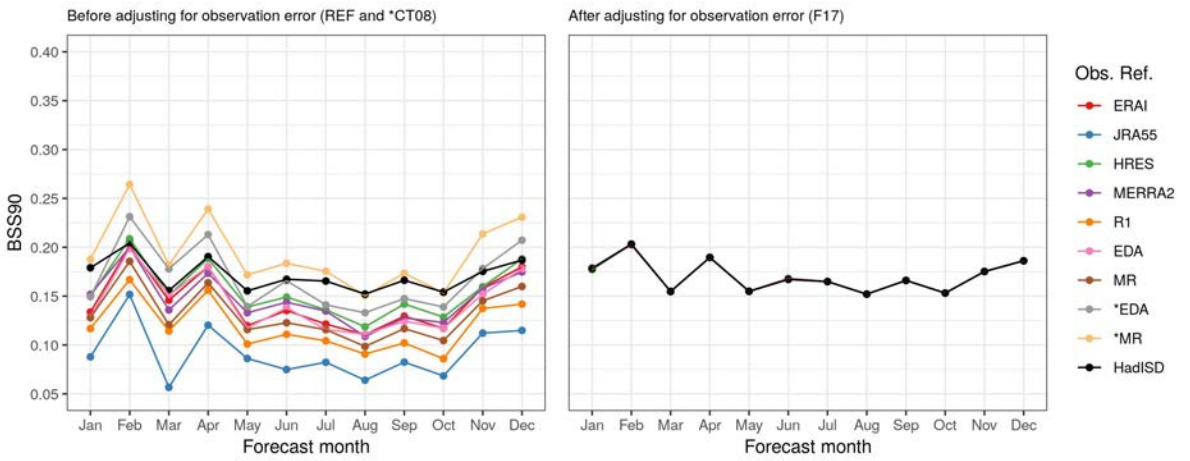


Figure 7.3: BSS90 of the lead-zero SEAS5 monthly predictions for surface wind speeds at a subset of HadISD locations before (left) and after (right) adjusting for observation error. We note that all eight curves overlap after adjusting for observation error (right).

7.6 Discussion

Observation error has been shown to play a key role in skill assessment of seasonal predictions. Different possibilities have been proposed in the literature for the uptake of such error. Here we have focused on those that consider the reference observation associated with its error, leaving the predicted ensemble elements as they are. Three different approaches have been tested: the traditional approach (REF) which considers one single verifying observation with no error, the observational probability method in CT08 with ensemble reanalyses, and F17 with all the set of observational references together with their uncertainty.

Both REF and CT08 methodologies yielded different skill score values to those that would be obtained with a perfect observational reference. In most of the cases, the real performance of the prediction systems is underestimated, in accordance with previous works (e.g. Santos and Ghelli (2012); Ferro (2017); Candille and Talagrand (2008)). The different fBSS90 and BSS90 values obtained with REF and CT08 mirror the important disagreements between reanalyses themselves and between reanalyses and point observations (Ramon et al., 2019). To catch the discrepancies between reanalyses, a MR dataset has been created in this work. Though the MR combined the data from five global reanalyses, providing a more accurate estimate of the observational uncertainty, it still offered a biased skill score. The scores that an error-free reference would achieve have not been obtained by any of the observational references. Neither is sufficient having the best single reanalysis, i.e. HRES, to have unbiased estimates of the score values. This suggests

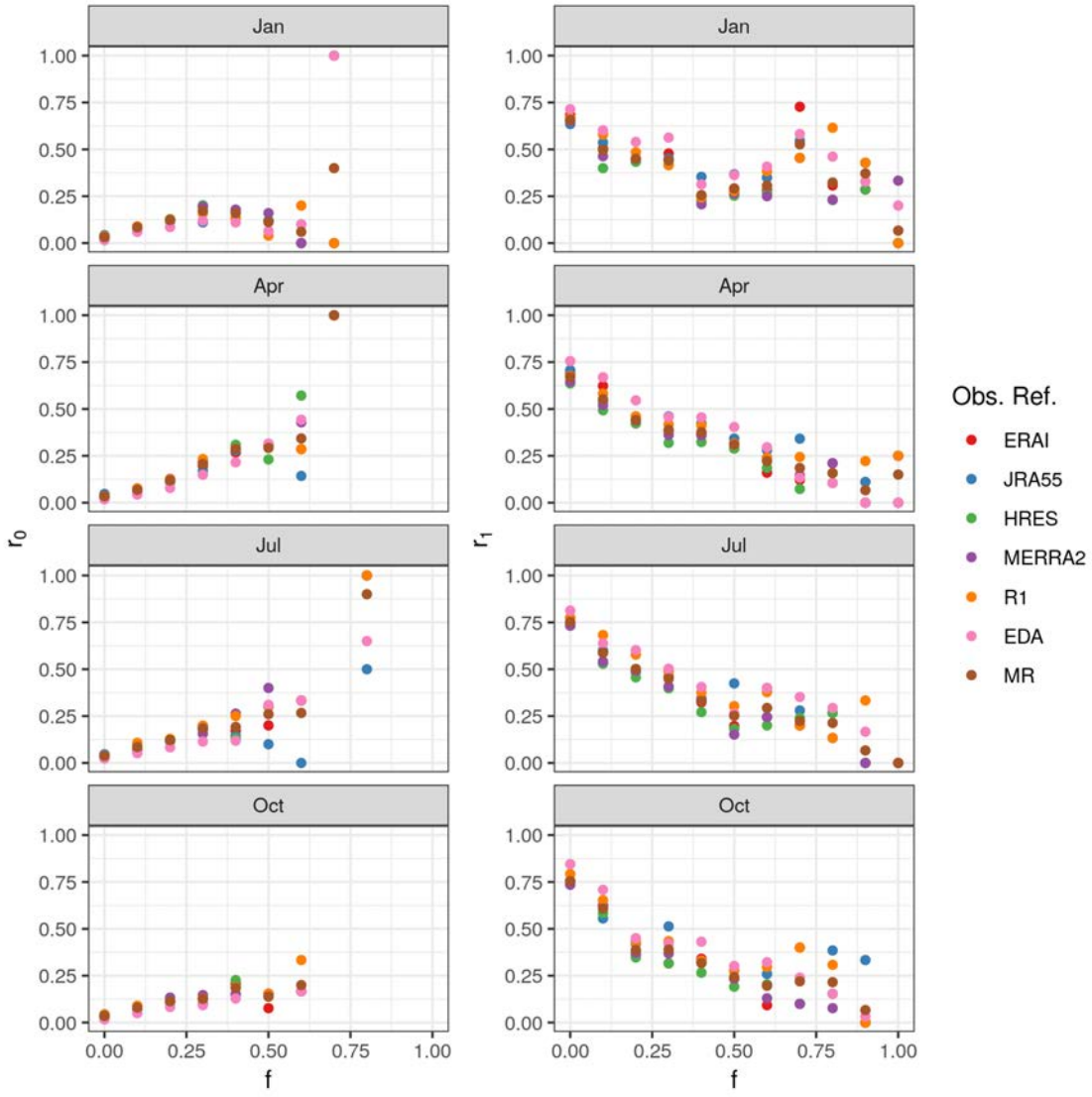


Figure 7.4: Misclassification probabilities r_0 (left column) and r_1 (right column) for the different predicted probabilities, f, and for an event exceeding the 90th percentile of the wind distribution. For the sake of simplicity, only four forecast months —i.e. January, April, July and October— have been represented.

that the differences between reanalysis values and point observations, i.e. representativeness errors, are large and are not traditionally accounted for in the verification.

Besides, an important issue has to be noted when CT08 is used with too uncertain observational references, in which the true observed outcome is almost unknown and its associated probability of occurrence is nearly the climatological value. When this happens, the uncertainty term of the BS decreases significantly, showing a misleading improvement of the overall BSS. For instance, for an event exceeding

the median value, a very uncertain observation outcome would have a probability of occurrence of approximately 50% —the observer is unsure whether the observation falls above or below the median value. In those cases, the climatological forecast will score extremely well, being it almost impossible to beat by any other forecast.

To take into account all the observation error, and ensure we are not obtaining ambiguous verification outcomes, we recommend to use F17 methodology. Although the generalised scoring rules do not preserve propriety, they still provide unbiased estimates of the true expected score. This is key in many aspects. For example, an apparent useless prediction verified against an uncertain reference may be profitable if observation error is properly considered. The downside of the method, however, is that either the misclassification probabilities r_0, r_1 and their dependence on f (parametrisation) have to be known beforehand, or true observations have to be available for generating the error distribution. In this study, we have assumed the in-situ dataset to be perfect, but other evaluation might assume a reanalyses —preferably a high-resolution or regional reanalysis— to be the ground truth.

F17 methodology has one important advantage for the particular case of the verification of seasonal predictions. Unlike weather forecasting or sub-seasonal predictions, the sample size available for the verification of seasonal forecasts (hindcast length) is often limited. Any skill score derived will be less robust, and this is particularly noticeable when considering more extreme properties such as high percentiles of the wind distribution (Sunyer et al., 2013). We have shown that the scoring rankings based on the fBSS90 are highly dependent on the verification dataset (Figure 7.2), but rankings appear more stable when considering other percentiles in the center of the distribution like the 50^{th} or 66^{th} (not shown). As long as observation error is properly accounted for, the stability of the rankings can be assured regardless of the definition of the categories or the hindcast length.

7.7 Conclusions

The main conclusions of this study are:

- Skill scores are highly sensitive to the choice of the observational reference. This poses a risk in many situations, for example, when selecting the best prediction system among a set of potential candidates.
- Observational references that provide estimates of their associated uncertainty approximate better the true skill scores.
- The Brier Score values of a seasonal prediction do not depend on the choice of the observational reference dataset if F17 methodology is used. However, the approach requires for knowledge about the error distribution of the verifying observations.

- The need for information about observational uncertainty should be taken into account in the production of future reanalysis datasets. Estimates of this uncertainty are beneficial for many applications.

Many users of seasonal predictions employ gridded forecasts to infer future anomalies of climate variables at a point scale and then verify these predictions using point data. Such is the case of the wind power industry, where wind observations taken at wind farms (point scale) are utilised in skill assessments of forecasts that represent a larger-scale quantity (grid scale, Ramon et al. (2021a)). Our work provides insights on the outcomes of verifications of gridded forecasts using point data and compare this against an analogous verification using gridded observations (reanalyses). We even go beyond this, showing how point observations can be used to adjust observation error in reanalyses, allowing reanalysis data to extend the verification to other locations where in-situ measurements are not available.

Future works could expand on F17 methodology to generalise the calculation of the estimates of the error distribution of the observations, avoiding the need of assuming a ground truth. Even though this manuscript emphasises the strong dependency of the misclassification rates on the observational reference dataset, a relationship between these and the predicted probabilities can be established in the form of, for example, a parametrisation equation. However, the extent to which it can be generalised, has to be explored in further detail.

Acknowledgements. Authors are extremely grateful to César Azorín-Molina for sharing HadISD monthly-averaged and filtered wind speed series.

Chapter 8

Conclusions and future perspectives

8.1 Conclusions

Seasonal predictions tailored to the wind energy sector represent an innovative step in using climate information to better administer the future availability of the wind resource. Nevertheless, the uptake of those predictions is not yet complete. Many barriers act as challenges for wind energy users, including the low skill levels and the difficulties in integrating such technical material into their decision-making processes. This PhD thesis provided advances in both directions.

The accomplished goals consist of the production of methodologies aimed at increasing and better characterising the quality of seasonal predictions for wind speed through appropriate climate observations. Secondarily, notable contributions have been made to provide high-quality wind observations in the form of a publicly accessible dataset as well as an exhaustive assessment of the uncertainty in reanalysis wind data. In the following, the conclusions of this PhD thesis are presented in more detail.

8.1.1 Delivering high-quality observations (Chapter 4)

Previous to the usage of wind observations in enhancing seasonal predictions, exhaustive quality control tests have to be designed and applied to ensure the high quality of the wind data. The collection of climate observations from tall towers to generate the raw version of the Tall Tower Dataset revealed the presence of erroneous data, the amount of which varied enormously from one location to another. For example, wind data outside a plausible range of values or abnormally long sequences of constant/duplicated values populate the wind series randomly.

To detect and remove all these bad data, a set of 18 quality-control tests have been run over the raw version of The Tall Tower Dataset. While most of the data, i.e. 95.2 %, were cor-

rect, some values had to be removed prior to producing the quality-controlled version of the Tall Tower Dataset. Later, these clean data have been made publicly accessible —with a few exceptions— at talltowers.bsc.es. The data portal allows for browsing through the different tall tower locations as well as a rapid visualisation of some characteristics of the wind data (e.g. timespan or extreme values). Initiatives delivering climate data freely available are undeniably beneficial for both climate scientists and wind energy users.

8.1.2 Gaining insight into the various types of error in observations (Chapters 4 and 5)

As explained in the subsection above, the errors and inconsistencies encountered within the wind series from the tall towers were diverse. This led to the creation of a quality control software suite specifically designed to detect all these errors in tall tower wind data. Among the different types of errors, random errors in wind sensors are often the most difficult to detect. Having parallel series from various instruments in the same tall tower allows for intra-station comparisons, an efficient feature to detect failures in one particular sensor, which was not implemented before in the existing quality control software. The Quality Control Software Suite for Tall Towers (QCSS4TT) is publicly available through a GitLab repository.

Reanalysis winds, on the other hand, are more susceptible to systematic errors. The comparison of surface winds from five different global reanalyses showed important differences in terms of the seasonal mean values, variability and linear trends. The differences in the seasonal mean values respond mainly to representativeness errors: the finer horizontal resolutions in the newest reanalyses such as ERA5 seem not to be enough to account for local wind effects. Nevertheless, reanalyses are still a useful source tool in downscaling studies, and the current efforts in creating high-resolution regional reanalyses appear to be fruitful in many regions like Europe (Kaiser-Weiss et al., 2019). Besides, wind speed trends have also been found in all datasets. While some of those trends may be spurious, responding to artifacts in the creation of the reanalyses, others may indicate a real change in the behaviour of wind speeds. Little attention has been traditionally paid to those fluctuations, although they can be used to foresee changes in wind power generation and the performance of wind farms during their lifetime (20-30 years typically).

8.1.3 Enhancing seasonal predictions of wind speed at the local scale (Chapter 6)

To have complete coverage of the whole globe, seasonal predictions need to be delivered on coarse grids (i.e., horizontal resolution of $\sim$100 km), thus misrepresenting atmospheric phenomena happening at much finer scales. Ac-

counting for such effects might be crucial for wind speed analyses, primarily where local-scale winds represent a high percentage of the mean value.

To take into account these fine-scale wind features, a downscaling methodology for seasonal predictions has been provided. The hybrid predictions proposed here outperform the purely dynamical predictions of near-surface wind speeds, indicating that the information of the synoptic scale provides an added value for predicting local-scale wind speeds. The strong point of these hybrid predictions is that they can be generated using a series of wind data from one single point station, provided that the timespan is sufficiently long (at least 20 years). With that, wind energy users can generate their own hybrid forecasts using their data collected at the wind power plant. Besides, seasonal predictions of wind power generation are also possible. With complete series of indicators like the wind capacity factor known beforehand, hybrid predictions can anticipate the amount of renewable production for coming months or seasons.

8.1.4 Achieving more robust estimates of the quality of seasonal predictions (Chapter 7)

The quality of seasonal predictions is assessed by comparing those forecasts against an observational reference dataset. Reanalyses most commonly serve as benchmark observations, but forecast verifications against station data are also performed. However, neither station data nor reanalyses are exempt of containing errors, which can be propagated into the skill estimates producing undesirable results. For example, one prediction system may appear skilful or not depending on which observational reference is used in the verification. Another instance may happen when selecting the best prediction system among a set of candidates, or their ranking based on their skill, being both outcomes strongly dependant on the election of the observational reference dataset.

To get rid of such dependencies, and obtain the real skill of seasonal prediction systems, skill scores need to be re-defined to account for observation error. This methodology was already proposed in the literature, but never applied to seasonal predictions for wind speed. Furthermore, an extension to the case when the predictand is categorised —as it is usually done in seasonal forecasting— is provided in this thesis. The error-corrected Brier Score yielded unbiased values of the skill of five seasonal prediction systems, allowing also a correct ranking of these. A verification against ensemble reanalyses is also introduced here, highlighting the benefits of the uncertainty estimates they provide.

8.1.5 Tailoring observations and methods to user needs (Chapters 4, 5, 6 and 7)

The scientific contributions of this work satisfy a wide variety of needs and gaps reported by users from the wind energy industry.

Firstly, the global and coarse grids of seasonal predictions are not practical for the spatial scales in which wind energy users work. At a wind farm level, for instance, local wind effects need to be characterised. That might be the case of a wind farm location over a mountain ridge, where winds tend to be stronger than in the surroundings. By using the methodology for creating the hybrid predictions with their own observation data, wind energy users can generate downscaled seasonal predictions at their locations of interest.

Another case where point wind speed observations can be used is in the verification of seasonal predictions. Our study provides an insightful comparison between a skill assessment with these local observations and reanalysis winds. We warn of the noticeable differences encountered and the perils that an arbitrary choice of an observational reference may suppose. The solution proposed is to use these point observations —which can be regarded as ground truth— to infer the error distribution of the reanalysis winds and insert it into the error-corrected scoring rules. In this way, the verification of gridded seasonal forecasts can be extended outside the measurement location, allowing a correct use of reanalyses for example to explore new areas for prospecting in the context of wind resource assessment.

One type of wind observation, which is non-standard for climate studies but massively employed within the wind industry, is taken at tall meteorological towers. These structures provide accurate estimates of the wind flow at turbine hub heights. However, those wind measurements are scarce and difficult to reach until the creation of the Tall Tower Dataset. Due to its limited spatial coverage, the Tall Tower Dataset will not discharge energy companies from installing meteorological masts before constructing wind farms. Instead, the principal usage of the Tall Tower Dataset will be to serve as a reference for inter-comparison or verification activities. Series between nearby masts can be contrasted to check for inconsistencies in the wind data or assess long-term climate variability, among other instances.

In the absence of point observations, the use of reanalyses among wind energy users has been traditionally widespread as they are handy and easy-to-use climate data sources. However, the new knowledge gained after the intercomparison and verification should warn of the pitfalls of these types of observations. Though the direct output of global reanalyses is never used to infer the mean wind speed value at a potential farm site, some applications such as wind atlases or downscaling feed on reanalysis values. Therefore, the reported problems like biases or spurious trends may eventually be transferred

to the high-resolution datasets.

8.2 Future perspectives

The work presented in this PhD thesis should serve as a starting point for coming studies aimed at improving the quality of seasonal predictions. New climate services can be produced with the most up-to-date methods and sound techniques. Those results can then be compared against those presented here to evaluate the gains of future innovations. In particular, the following ideas can be implemented.

Although this PhD thesis has focused on wind speed, the methodologies can be applied to other ECV. Being also highly variable in space and time, precipitation is also a challenging variable to be predicted at the seasonal timescale. Following an analogue methodology, hybrid seasonal forecasts for precipitation could produce accurate anticipations of the amount of water at a river catchment, which is highly beneficial for the hydropower sector. The availability of precipitation observations is not a limitation, especially in Europe, since tens of databases containing long and homogeneous records already exist, and the dense spatial distribution of stations allows for characterising fine-scale details.

Besides, a new set of techniques that are currently entering the field of climate prediction is Artificial Intelligence. Further improvements than those offered by the purely statistical approaches may come from techniques like ML. They have been shown to perform well, for example, in pattern recognition, with the consideration that substantial amounts of data are needed to train the models. This might suppose a major handicap in climate prediction, which, unlike weather forecasting, the time aggregation of the data reduces the sample size considerably (e.g. monthly, seasonal and annual means are preferred over daily averages). Therefore, working on the lengthening of the timespan of observational datasets should help increase the quality of ML models.

Only the 3% of the meteorological masts in the Tall Tower Dataset span more than 30 years. While this percentage might appear insufficient for a dataset to be helpful in climate analyses, one must bear in mind that the average lifetime of most of these structures ranges between one and five years. Fortunately, recent research projects like FINO[1] in the North Sea are installing tall towers to collect meteorological data for as many years as possible. These new records and those from newly installed masts could be added to the Tall Tower Dataset in future updates of the data collection.

Lengthening the timespan of the tall tower series may also serve for assessing climatic fluctuations and their effects on meteorological variables. The declining trend (stilling) of surface wind speeds observed since the 1980s and its recent recovery reported by Zeng et al.

[1] www.fino1.de

(2019) has been attributed to decadal climatic oscillations. Since the Tall Tower Dataset offers wind series not only at surface level but also at higher altitudes, future research may evaluate a possible decoupling of wind speed trends between these different heights. Gaining insight into the periods and drivers of such oscillations would be tremendously valuable for advancing in the modelling and initialisation of seasonal and decadal predictions.

As commented before, regional and high-resolution reanalyses (e.g. ERA5-Land, Muñoz-Sabater et al. (2021)) are fundamental when dealing with variables with high spatio-temporal variability. An analogue verification exercise to the one carried out with the five global reanalyses would be helpful to evaluate how close regional reanalyses are with the true observations.

Chapter 9

Bibliography

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J. (2003). Guidelines on climate metadata and homogenization. *World Meteorological Organization*, (1186):55.

APPLICATE (2020). How does Arctic sea ice affect energy production in mid-latitudes? https://applicate-h2020.eu/wp-content/uploads/2020/12/energy_case_study_applicate_v3.pdf.

Baars, H., Herzog, A., Heese, B., Ohneiser, K., Hanbuch, K., Hofer, J., Yin, Z., Engelmann, R., and Wandinger, U. (2020). Validation of Aeolus wind products above the Atlantic Ocean. *Atmospheric Measurement Techniques Discussions*, pages 1–27, doi: 10.5194/amt-2020-198.

Bellprat, O., Massonnet, F., Siegert, S., Prodhomme, C., Macias-Gómez, D., Guemas, V., and Doblas-Reyes, F. (2017). Uncertainty propagation in observational references to climate model scales. *Remote Sensing of Environment*, 203:101–108, doi: 10.1016/j.rse.2017.06.034.

Ben-Bouallègue, Z. (2020). Accounting for representativeness in the verification for ensemble-forecasts. Technical report, ECMWF, doi: 10.21957/5z6esc7wr, https://www.ecmwf.int/node/19544.

Bloomfield, H. C., Brayshaw, D. J., Shaffrey, L. C., Coker, P. J., and Thornton, H. E. (2016). Quantifying the increasing sensitivity of power systems to climate variability. *Environmental Research Letters*, 11(12), doi: 10.1088/1748-9326/11/12/124025.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9(10):3751–3777, doi: 10.5194/gmd-9-3751-2016.

Bonavita, M., Arcucci, R., Carrassi, A., Dueben, P., Geer, A. J., Le Saux, B., Longépé, N., Mathieu, P.-P., and Raynaud, L. (2020). Machine Learning for Earth System Observation and Prediction. *Bulletin of the American Meteorological Society*, 102(4):E710–E716, doi: 10.1175/bams-d-20-0307.1.

Brower, M. C., Bailey, B. H., Doane, J., and Eberhard, M. J. (2012). *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*. doi: 10.1002/9781118249864.

Buontempo, C., Hewitt, C. D., Doblas-Reyes, F. J., and Dessai, S. (2014). Climate service development, delivery and use in Europe at monthly to inter-annual timescales. *Climate Risk Management*, 6:1–5, doi: 10.1016/j.crm.2014.10.002.

Candille, G. and Talagrand, O. (2008). Impact of observational error on the validation of ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 134(633):959–971, doi: 10.1002/qj.268.

Cane, M. A., Zebiak, S. E., and Dolan, S. C. (1986). Experimental forecasts of El Niño. *Nature*, 321(6073):827–832, doi: 10.1038/321827a0.

Christensen, J. H., Kanikicharla, K. K., Aldrian, E., An, S. I., Albuquerque Cavalcanti, I. F., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J. K., Kitoh, A., Kossin, J., Lau, N. C., Renwick, J., Stephenson, D. B., Xie, S. P., Zhou, T., Abraham, L., Ambrizzi, T., Anderson, B., Arakawa, O., Arritt, R., Baldwin, M., Barlow, M., Barriopedro, D., Biasutti, M., Biner, S., Bromwich, D., Brown, J., Cai, W., Carvalho, L. V., Chang, P., Chen, X., Choi, J., Christensen, O. B., Deser, C., Emanuel, K., Endo, H., Enfield, D. B., Evan, A., Giannini, A., Gillett, N., Hariharasubramanian, A., Huang, P., Jones, J., Karumuri, A., Katzfey, J., Kjellström, E., Knight, J., Knutson, T., Kulkarni, A., Kundeti, K. R., Lau, W. K., Lenderink, G., Lennard, C., Leung, L. y. R., Lin, R., Losada, T., Mackellar, N. C., Magaña, V., Marshall, G., Mearns, L., Meehl, G., Menéndez, C., Murakami, H., Nath, M. J., Neelin, J. D., van Oldenborgh, G. J., Olesen, M., Polcher, J., Qian, Y., Ray, S., Reich, K. D., de Fonseca, B. R., Ruti, P., Screen, J., Sedláček, J., Solman, S., Stendel, M., Stevenson, S., Takayabu, I., Turner, J., Ummenhofer, C., Walsh, K., Wang, B., Wang, C., Watterson, I., Widlansky, M., Wittenberg, A., Woollings, T., Yeh, S. W., Zhang, C., Zhang, L., Zheng, X., and Zou, L. (2013). Climate phenomena and their relevance for future regional climate change. In *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, volume 9781107057, pages 1217–1308. doi: 10.1017/CBO9781107415324.028.

Clark, R. T., Bett, P. E., Thornton, H. E., and Scaife, A. A. (2017). Skilful seasonal predictions for the European energy industry. *Environmental Research Letters*, 12(11):119602, doi: 10.1088/1748-9326/aa94a7.

Cortesi, N., Torralba, V., Lledó, L., Manrique-Suñén, A., Gonzalez-Reviriego, N., Soret, A., and Doblas-Reyes, F. J. (2021). *Yearly evolution of Euro-Atlantic weather regimes and of their sub-seasonal predictability*, volume 56. Springer Berlin Heidelberg, doi: 10.1007/s00382-021-05679-y.

Cruz-García, R., Guemas, V., Chevallier, M., and Massonnet, F. (2019). An assessment of regional sea ice predictability in the Arctic ocean. *Climate Dynamics*, 53(1-2):427–440, doi: 10.1007/s00382-018-4592-6.

Dee, D. P., Kallen, E., Simmons, A. J., and Haimberger, L. (2011a). Comments on "Reanalyses Suitable for Characterizing Long-Term Trends". *Bulletin of the American Meteorological Society*, 92(1):65–70, doi: 10.1175/2010bams3070.1.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., and others (2011b). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, doi: 10.1002/qj.828.

Delgado-Torres, C., Caron, L.-P., Donat, M. G., Gonzalez-Reviriego, N., Athanasiadis, P. J., Bretonnière, P.-A., Ho, A.-C., Pankatz, K., Paxian, A., Pérez-Zanón, N., Samsó, M., Solaraju-Murali, B., Soret, A., and Doblas-Reyes, F. J. (2021). CMIP6/DCPP-A multi-model forecast quality assessment. *Manuscript submitted for publication*.

Deutscher Wetterdienst (2019). Seasonal forecasting with the german climate forecast system. https://www.dwd.de/EN/ourservices/seasonals_forecasts/project_description.html?nn=641552&lsbId=619784.

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4):245–268, doi: 10.1002/wcc.217.

Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting — II. Calibration and combination. *Tellus A: Dynamic Meteorology and Oceanography*, 57(3):234–252, doi: 10.3402/tellusa.v57i3.14658.

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E. P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S. W., and Taguchi, M. (2020). The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 2. Predictability Arising From Stratosphere-Troposphere Coupling. *Journal of Geophysical Research: Atmospheres*, 125(2), doi: 10.1029/2019JD030923.

Donat, M. G., Angélil, O., and Ukkola, A. M. (2019). Intensification of precipitation extremes in the world's humid and water-limited regions. *Environmental Research Letters*, 14(6), doi: 10.1088/1748-9326/ab1c8e.

Dorel, L., Ardilouze, C., Déqué, M., Batté, L., and Guérémy, J.-F. (2017). Documentation of the METEO-FRANCE Pre-Operational seasonal forecasting system. Technical report, METEO-FRANCE, http://seasonal.meteo.fr/sites/data/Documentation/doc_modele/Model_MF-S6_C3S_technical_en.pdf.

Dunn, R. J., Willett, K. M., Morice, C. P., and Parker, D. E. (2014). Pairwise homogeneity assessment of HadISD. *Climate of the Past*, 10(4):1501–1522, doi: 10.5194/cp-10-1501-2014.

European Commission (2019). The European Green Deal. Technical report, European Commision, Brussels, doi: 10.2307/j.ctvd1c6zh.7, https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf.

Falls, M., Bernardello, R., Castrillo, M., Llort, J., Acosta, M., and Galí, M. (2021). Use of Genetic Algorithms for Ocean Model Parameter Optimisation. *Manuscript submitted for publication*.

Ferro, C. A. (2014). Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923, doi: 10.1002/qj.2270.

Ferro, C. A. (2017). Measuring forecast performance in the presence of observation error. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2665–2676, doi: 10.1002/qj.3115.

Foley, A. M., Leahy, P. G., Marvuglia, A., and McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. doi: 10.1016/j.renene.2011.05.033.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Lynn Harvey, V., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C. Z. (2017). Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, 17(2):1417–1452, doi: 10.5194/acp-17-1417-2017.

García-Serrano, J., Guemas, V., and Doblas-Reyes, F. J. (2015). Added-value from initialization in predictions of Atlantic multi-decadal variability. *Climate Dynamics*, 44(9-10):2539–2555, doi: 10.1007/s00382-014-2370-7.

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., and Zorita, E. (2012). What is the role of the observational dataset in the evaluation and scoring of climate models? *Geophysical Research Letters*, 39(24):1–5, doi: 10.1029/2012GL054206.

Gregory, J. (2003). The CF metadata standard. *CLIVAR Exchanges*, 8(4):4, https://cfconventions.org/Data/cf-documents/overview/article.pdf.

Gubler, S., Sedlmeier, K., Bhend, J., Avalos, G., Coelho, C. A., Escajadillo, Y., Jacques-Coper, M., Martinez, R., Schwierz, C., de Skansi, M., and Spirig, C. (2020). Assessment of ECMWF SEAS5 seasonal forecast performance over South America. *Weather and Forecasting*, 35(2):561–584, doi: 10.1175/WAF-D-19-0106.1.

Guemas, V., Chevallier, M., Déqué, M., Bellprat, O., and Doblas-Reyes, F. (2016). Impact of sea ice initialization on sea ice and atmosphere prediction skill on seasonal timescales. *Geophysical Research Letters*, 43(8):3889–3896, doi: 10.1002/2015GL066626.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P. A., Caron, L. P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F., Fladrich, U., Franco, R. F., Garciá-Serrano, J., Von Hardenberg, J., Koenigk, T., Levine, X., Meccia, V. L., Van Noije, T., Van Den Oord, G., Palmeiro, F. M., Rodrigo, M., Ruprich-Robert, Y., Le Sager, P., Tourigny, E., Wang, S., Van Weele, M., and Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR - Description, model computational performance and basic validation. *Geoscientific Model Development*, 13(8):3507–3527, doi: 10.5194/gmd-13-3507-2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-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., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thepaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, doi: 10.1002/qj.3803.

Hewitt, C., Mason, S., and Walland, D. (2012). The global framework for climate services. *Nature Climate Change*, 2(12):831–832, doi: 10.1038/nclimate1745.

Huang, H. Y., Capps, S. B., Huang, S. C., and Hall, A. (2014). Downscaling near-surface wind over complex terrain using a physically-based statistical modeling approach. *Climate Dynamics*, 44(1-2):529–542, doi: 10.1007/s00382-014-2137-1.

Intergovernmental Panel on Climate Change (2013). Near-term Climate Change: Projections and Predictability. In Intergovernmental Panel on Climate Change, editor, *Climate Change 2013 - The Physical Science Basis*, pages 953–1028. Cambridge University Press, Cambridge, doi: 10.1017/CBO9781107415324.023.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, doi: 10.1002/qj.3130.

Jeong, J. H., Linderholm, H. W., Woo, S. H., Folland, C., Kim, B. M., Kim, S. J., and Chen, D. (2013). Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *Journal of Climate*, 26(6):1956–1972, doi: 10.1175/JCLI-D-12-00159.1.

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., Mogensen, K., Zuo, H., and Monge-Sanz, B. M. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3):1087–1117, doi: 10.5194/gmd-12-1087-2019.

Jolliffe, I. T. and Stephenson, D. B. (2012). *Forecast Verification*. doi: 10.1002/9781119960003.

Jude, M. and Leseney, H. (2017). Influence of the choice of long term sources in France Modelled data vs measured data 2 Methods 1 Background and objectives 2 Results for MERRA-2 in France. pages 2–7.

Juricke, S., MacLeod, D., Weisheimer, A., Zanna, L., and Palmer, T. N. (2018). Seasonal to annual ocean forecasting skill and the role of model and observational uncertainty. *Quarterly Journal of the Royal Meteorological Society*, 144(715):1947–1964, doi: 10.1002/qj.3394.

Kaiser-Weiss, A. K., Borsche, M., Niermann, D., Kaspar, F., Lussana, C., Isotta, F. A., van den Besselaar, E., van der Schrier, G., and Undén, P. (2019). Added value of regional reanalyses for climatological applications. *Environmental Research Communications*, 1(7), doi: 10.1088/2515-7620/ab2ec3.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437–471, doi: fg6rf9.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori,

H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1):5–48, doi: 10.2151/jmsj.2015-001, https://www.jstage.jst.go.jp/article/jmsj/93/1/93_2015-001/_article.

Koster, R. D., Mahanama, S. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J. H., Lee, W. S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., Stanelle, T., Van Den Hurk, B. J., Vitart, F., and Wood, E. F. (2011). The second phase of the global land-atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. *Journal of Hydrometeorology*, 12(5):805–822, doi: 10.1175/2011JHM1365.1.

Kushnir, Y., Robinson, W. A., Chang, P., and Robertson, A. W. (2006). The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability. doi: 10.1175/JCLI3943.1.

Lang, S. and McKeogh, E. (2011). LIDAR and SODAR measurements of wind speed and direction in upland terrain for wind energy purposes. *Remote Sensing*, 3(9):1871–1901, doi: 10.3390/rs3091871.

Leung, L. R., Hamlet, A. F., Lettenmaier, D. P., and Kumar, A. (1999). Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin. *Bulletin of the American Meteorological Society*, 80(11):2313–2329, doi: brxndr.

Lledó, L. (2020). *Climate variability predictions for the wind energy industry: a climate services perspective*. PhD thesis, Universitat de Barcelona.

Lledó, L., Bellprat, O., Doblas-Reyes, F. J., and Soret, A. (2018). Investigating the Effects of Pacific Sea Surface Temperatures on the Wind Drought of 2015 Over the United States. *Journal of Geophysical Research: Atmospheres*, pages 1–13, doi: 10.1029/2017JD028019.

Lledó, L. and Doblas-Reyes, F. J. (2020). Predicting daily mean wind speed in Europe weeks ahead from MJO status. *Monthly Weather Review*, 148(8):3413–3426, doi: 10.1175/MWR-D-19-0328.1.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2), doi: fwwt5q.

Maclachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G. (2015). Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1072–1084, doi: 10.1002/qj.2396.

Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N., and Doblas-Reyes, F. J. (2020). Choices in the verification of s2s forecasts and their implications for climate services. *Monthly Weather Review*, 148(10):3995–4008, doi: 10.1175/MWR-D-20-0067.1.

Manubens, N., Caron, L. P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., Garcia-Serrano, J., Massonnet, F., Ménégoz, M., Sicardi, V., Batté, L., Prodhomme, C., Torralba, V., Cortesi, N., Mula-Valls, O., Serradell, K., Guemas, V., and Doblas-Reyes, F. J. (2018). An R package for climate forecast verification. *Environmental Modelling and Software*, 103:29–42, doi: 10.1016/j.envsoft.2018.01.018.

Manzanas, R., Gutiérrez, J. M., Bhend, J., Hemri, S., Doblas-Reyes, F. J., Torralba, V., Penabad, E., and Brookshaw, A. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. *Climate Dynamics*, 53(3-4):1287–1305, doi: 10.1007/s00382-019-04640-4.

Marcos, R., González-Reviriego, N., Torralba, V., Soret, A., and Doblas-Reyes, F. J. (2018). Characterization of the near surface wind speed distribution at global scale: ERA-Interim reanalysis and ECMWF seasonal forecasting system 4. *Climate Dynamics*, 0(0):1–13, doi: 10.1007/s00382-018-4338-5.

Martin, A., Weissmann, M., Reitebuch, O., Rennie, M., Geiß, A., and Cress, A. (2021). Validation of Aeolus winds using radiosonde observations and numerical weather prediction model equivalents. *Atmospheric Measurement Techniques*, 14(3):2167–2183, doi: 10.5194/amt-14-2167-2021.

Massonnet, F., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J. (2016). Using climate models to estimate the quality of global observational data sets. *Science*, 354(6311):452–455, doi: 10.1126/science.aaf6369.

McNie, E. C. (2007). Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environmental Science and Policy*, 10(1):17–38, doi: 10.1016/j.envsci.2006.10.004.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910, doi: 10.1175/JTECH-D-11-00103.1.

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., Danabasoglu, G., Dirmeyer, P. A., Doblas-Reyes, F. J., Domeisen, D. I. V., Ferranti, L., Ilynia, 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., de Andrade, F. M., 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., Justin Ng, C. H., 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., and Yeager, S. (2020). Current and Emerging Developments in Subseasonal to Decadal Prediction. *Bulletin of the American Meteorological Society*, 101(6):E869–E896, doi: 10.1175/BAMS-D-19-0037.1.

Mezzina, B., García-Serrano, J., Bladé, I., Palmeiro, F. M., Batté, L., Ardilouze, C., Benassi, M., and Gualdi, S. (2020). Multi-model assessment of the late-winter extra-tropical response to El Niño and La Niña. *Climate Dynamics*, doi: 10.1007/s00382-020-05415-y.

Molina, M. O., Gutiérrez, C., and Sánchez, E. (2021). Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *International Journal of Climatology*, (March):1–15, doi: 10.1002/joc.7103.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development*, 8(5):1339–1356, doi: 10.5194/gmd-8-1339-2015.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D., Piles, M., Rodríguez-Fernández, N., Zsoter, E., Buontempo, C., and Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, (March):1–50, doi: 10.5194/essd-2021-82.

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2):281–293, doi: bbpk8s.

Ortizbeviá, M. J., Sánchezgómez, E., and Alvarez-García, F. J. (2011). North Atlantic atmospheric regimes and winter extremes in the Iberian peninsula. *Natural Hazards and Earth System Science*, 11(3):971–980, doi: 10.5194/nhess-11-971-2011.

Palmeiro, F. M., García-Serrano, J., Bellprat, O., Bretonnière, P. A., and Doblas-Reyes, F. J. (2020). Boreal winter stratospheric variability in EC-EARTH: High-Top versus Low-Top. *Climate Dynamics*, 54(5-6):3135–3150, doi: 10.1007/s00382-020-05162-0.

Parker, W. S. (2016). Reanalyses and observations: What's the Difference? *Bulletin of the American Meteorological Society*, 97(9):1565–1572, doi: 10.1175/BAMS-D-14-00226.1.

Prodhomme, C., Batté, L., Massonnet, F., Davini, P., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J. (2016a). Benefits of increasing the model resolution for the seasonal forecast quality in EC-earth. *Journal of Climate*, 29(24):9141–9162, doi: 10.1175/JCLI-D-16-0117.1.

Prodhomme, C., Doblas-Reyes, F., Bellprat, O., and Dutra, E. (2016b). Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Climate Dynamics*, 47(3-4):919–935, doi: 10.1007/s00382-015-2879-4.

Ramon, J., Lledó, L., Bretonnière, P.-A., Samsó, M., and Doblas-Reyes, F. J. (2021a). A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environmental Research Letters*, 16(5):054010, doi: 10.1088/1748-9326/abe491.

Ramon, J., Lledó, L., Ferro, C. A., and Doblas-Reyes, F. J. (2021b). Uncertainties in the observational reference: implications in skill assessment and model ranking. *Manuscript in preparation*.

Ramon, J., Lledó, L., Pérez-Zañón, N., Soret, A., and Doblas-Reyes, F. J. (2020). The Tall Tower Dataset. A unique initiative to boost wind energy research. *Earth System Science Data*, 12:429–439, doi: 10.5194/essd-12-429-2020.

Ramon, J., Lledo, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? *Quarterly Journal of the Royal Meteorological Society*, 145(724):3236–3251, doi: 10.1002/qj.3616.

Reichler, T. and Kim, J. (2008). Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. *Journal of Geophysical Research Atmospheres*, 113(5):1–13, doi: 10.1029/2007JD009278.

Rust, H. W., Richling, A., Bissolli, P., and Ulbrich, U. (2015). Linking teleconnection patterns to European temperature - A multiple linear regression model. *Meteorologische Zeitschrift*, 24(4):411–423, doi: 10.1127/metz/2015/0642.

S2S4E (2020). Benchmarking skill assessment of current sub-seasonal and seasonal forecast systems for users ' selected case studies. https://s2s4e.eu/sites/default/files/2020-06/s2s4e_d41.pdf.

Saetra, O., Hersbach, H., Bidlot, J. R., and Richardson, D. S. (2004). Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, 132(6):1487–1501, doi: c8w553.

Salinger, M. J., Renwick, J. A., and Mullan, A. B. (2001). Interdecadal Pacific Oscillation and South Pacific climate. *International Journal of Climatology*, 21(14):1705–1721, doi: 10.1002/joc.691.

Sanna, A., Borrelli, A., Athanasiadis, P. J., Materia, S., Storto, A., Navarra, A., Tibaldi,

S., and Gualdi, S. (2017). RP0285 – CMCC-SPS3: The CMCC Seasonal Prediction System 3. Technical Report RP0285, Centro Euro-Mediterraneo sui Cambiamenti Climatici, https://www.cmcc.it/publications/rp0285-cmcc-sps3-the-cmcc-seasonal-prediction-system-3.

Santos, C. and Ghelli, A. (2012). Observational probability method to assess ensemble precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138(662):209–221, doi: 10.1002/qj.895.

Shukla, J., Palmer, T. N., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., and Slingo, J. (2010). Toward a new generation of world climate research and computing facilities. *Bulletin of the American Meteorological Society*, 91(10):1407–1412, doi: 10.1175/2010BAMS2900.1.

Slingo, J. and Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4751–4767, doi: 10.1098/rsta.2011.0161.

Smith, A., Lott, N., and Vose, R. (2011). The integrated surface database: Recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92(6):704–708, doi: 10.1175/2011BAMS3015.1.

Solaraju-Murali, B., Caron, L. P., Gonzalez-Reviriego, N., and Doblas-Reyes, F. J. (2019). Multi-year prediction of European summer drought conditions for the agricultural sector. *Environmental Research Letters*, 14(12), doi: 10.1088/1748-9326/ab5043.

Sunyer, M. A., Sørup, H. J., Christensen, O. B., Madsen, H., Rosbjerg, D., Mikkelsen, P. S., and Arnbjerg-Nielsen, K. (2013). On the importance of observational data properties when assessing regional climate model performance of extreme precipitation. *Hydrology and Earth System Sciences*, 17(11):4323–4337, doi: 10.5194/hess-17-4323-2013.

Tammelin, B., Vihma, T., Atlaskin, E., Badger, J., Fortelius, C., Gregow, H., Horttanainen, M., Hyvönen, R., Kilpinen, J., Latikka, J., Ljungberg, K., Mortensen, N. G., Niemelä, S., Ruosteenoja, K., Salonen, K., Suomi, I., and Venäläinen, A. (2013). Production of the Finnish Wind Atlas. *Wind Energy*, 16(1):19–35, doi: 10.1002/we.517.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. doi: 10.1175/BAMS-D-11-00094.1.

Themeßl, M. J., Gobiet, A., and Leuprecht, A. (2011). Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31(10):1530–1544, doi: 10.1002/joc.2168.

Thorne, P. W. and Vose, R. S. (2010). Reanalyses Suitable for Characterizing Long-Term Trends. *Bulletin of the American Meteorological Society*, 91(3):353–361, doi: 10.1175/2009BAMS2858.1.

Tintó Prims, O., Castrillo, M., Acosta, M. C., Mula-Valls, O., Sanchez Lorente, A., Serradell, K., Cortés, A., and Doblas-Reyes, F. J. (2019). Finding, analysing and solving MPI communication bottlenecks in Earth System models. *Journal of Computational Science*, 36:100864, doi: 10.1016/j.jocs.2018.04.015.

Torralba, V. (2019). *Seasonal climate prediction for the wind energy sector: methods and tools for the development of a climate service*. PhD thesis, Universidad Complutense de Madrid, https://eprints.ucm.es/56841/.

Torralba, V., Doblas-Reyes, F. J., and Gonzalez-Reviriego, N. (2017a). Uncertainty in recent near-surface wind speed trends : a global reanalysis intercomparison. *Environmental Research Letters*, 12:114019, doi: 10.1088/1748-9326/aa8a58.

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., and Davis, M. (2017b). Seasonal Climate Prediction: A New Source of Information for the Management of Wind Energy Resources. *Journal of Applied Meteorology and Climatology*, 56(5):1231–1247, doi: 10.1175/JAMC-D-16-0204.1.

Torralba, V., Gonzalez-Reviriego, N., Cortesi, N., Manrique-Suñén, A., Lledó, L., Marcos, R., Soret, A., and Doblas-Reyes, F. J. (2021). Challenges in the selection of atmospheric circulation patterns for the wind energy sector. *International Journal of Climatology*, 41(3):1525–1541, doi: 10.1002/joc.6881.

Turco, M., Ceglar, A., Prodhomme, C., Soret, A., Toreti, A., and Doblas-Reyes Francisco, J. (2017). Summer drought predictability over Europe: Empirical versus dynamical forecasts. *Environmental Research Letters*, 12(8), doi: 10.1088/1748-9326/aa7859.

Van Den Dool, H. M. (1994). Searching for analogues, how long must we wait? *Tellus, Series A*, 46 A(3):314–324, doi: 10.3402/tellusa.v46i3.15481.

van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., and Hazeleger, W. (2012). Decadal prediction skill in a multi-model ensemble. *Climate Dynamics*, 38(7-8):1263–1280, doi: 10.1007/s00382-012-1313-4.

Vitart, F., Robertson, A. W., and S2S Steering Group (2015). Sub-seasonal to seasonal prediction: Linking weather and climate. Seamless Prediction of the Earth System: From Minutes to Months. Technical Report WMO No.1156, https://library.wmo.int/pmb_ged/wmo_1156_en.pdf.

Wang, L., Ting, M., and Kushner, P. J. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, 7(1), doi: 10.1038/s41598-017-00353-y.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Academic Press, 3rd editio edition.

WMO (2017). *Guide to the Global Observing System*. https://library.wmo.int/doc_num.php?explnum_id=4236.

Wu, B., Zhou, T., Li, C., Müller, W. A., and Lin, J. (2019). Improved decadal prediction of Northern-Hemisphere summer land temperature. *Climate Dynamics*, 53(3-4):1357–1369, doi: 10.1007/s00382-019-04658-8.

Zeng, Z., Ziegler, A. D., Searchinger, T., Yang, L., Chen, A., Ju, K., Piao, S., Li, L. Z., Ciais, P., Chen, D., Liu, J., Azorin-Molina, C., Chappell, A., Medvigy, D., and Wood, E. F. (2019). A reversal in global terrestrial stilling and its implications for wind energy production. *Nature Climate Change*, 9(12):979–985, doi: 10.1038/s41558-019-0622-6.

Zhang, C. (2013). Madden-julian oscillation: Bridging weather and climate. *Bulletin of the American Meteorological Society*, 94(12):1849–1870, doi: 10.1175/BAMS-D-12-00026.1.

Esta tesi doctoral s'ha imprès a Barcelona durant el setembre de 2021,
just complint-se cinc anys de la meua arribada a aquesta ciutat meravellosa.

Seasonal climate predictions appear extremely helpful in anticipating climate variations and taking timely action to manage their possible effects better. The wind power industry is a potential user of those predictions: the amount of future renewable production highly depends on wind speed anomalies. Regrettably, current seasonal predictions suffer from insufficient skill levels, and their probabilistic nature makes them harsh to understand for non-experienced users. This PhD thesis aims at improving the quality of seasonal predictions for wind speed, targeting specific needs and gaps reported by the wind energy industry. The main goal is achieved by employing high-quality wind observations. Adequate use of reanalyses or station data can enhance seasonal predictions, provided that sound methods and practical tools are defined.

Jaume Ramon Gamon (Estivella, 1993) became passionate about atmospheric sciences when, at the age of seven, he suffered the consequences of a catastrophic flooding event. In 2003 he was given his first meteorological station, which has been recording climate data since then. He has also been a frequent user of forums for weather and climate forecasting. Jaume got his degree in Physics in 2016, and his Master's in Meteorology the year after. In August 2017, when everyone was on holidays, he joined the BSC, where he had the opportunity to develop this PhD thesis. The work presented here is the climax of a life devoted to looking at the sky, so he really hopes for the reader to like it.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación