Análisis Estadístico y Predicción de Viento

Sheila Carreño-Madinabeitia

Dirigido por

Gabriel Ibarra-Berastegi

y

Jon Sáenz Agirre

Departamento de Física
Facultad de Ciencia y Tecnología
Universidad del País Vasco

2021

La versión electrónica se encuentra en https://addi.ehu.es

Vitoria-Gasteiz, Marzo 2021

Diseño de la portada de Ainhara Flores.

Maquetación y composición con la versión de LaTeX 2020.
La plantilla de LaTeX ha sido adaptada de la proporcionada por Santi González-Rojí.

A mi familia

Siempre se llega a alguna parte
si se camina lo bastante

– Lewis Carroll

Índice

Resumen

El objetivo principal de esta tesis es el análisis estadístico del viento en diferentes ámbitos relacionados, como son dos escalas temporales, extremos y una metodología de evaluación bidimensional.

El primero es la predicción de viento a muy corto o corto plazo en el cual se ha desarrollado y testado un modelo híbrido para realizar predicción horaria de las componentes zonal y meridional de viento y la racha a partir de observaciones y resultados del modelo numérico empleado, ERA-Interim, tanto en su modo análisis como predicción en tres emplazamientos del País Vasco.

En el ámbito de la validación se ha desarrollado el diagrama de Sailor para visualizar y valorar de forma conjunta las variables vectoriales bidimensionales. Además de la metodología, se ha implementado la librería SailoR del programa R, que se distribuye con licencia libre para la comunidad científica.

Aplicando el estudio a la energía eólica, se ha analizado la variabilidad a largo plazo de las variables densidad de energía eólica (WPD) y factor de capacidad (CF) offshore en el siglo XX alrededor de la PI. Se han detectado las zonas Atlántica, Golfo de León y Cabo de Gata como las de mayor potencial eólico. También se ha demostrado que las tendencias decadales tanto de WPD como de CF, son positivas y significativas en todo el dominio de estudio.

Finalmente, a partir del primer estudio se ha visto la necesidad de comenzar a trabajar en el análisis de los extremos de viento, los cuales se dan mayoritariamente en las tormentas o durante eventos de tiempo severo. Por este motivo se ha analizado si el modelo de downscaling numérico (WRF) reproduce de forma adecuada los índices de inestabilidad atmosférica CAPE, CIN y TT para la Península Ibérica (PI) para los ocho radiosondeos disponibles y para el resto de la Península a partir de los datos del modelo numérico.

Agradecimientos

A lo largo de estos años han sido muchas las personas que han compartido conmigo este viaje y quiero mostrar mi agradecimiento a todas ellas.

En primer lugar, quiero dar las gracias a mis directores Gabriel Ibarra-Berastegi y Jon Sáez por aceptarme como doctoranda y confiar en mi. Desde ese momento y a lo largo de toda la tesis el apoyo de Alain Ulazia ha sido indispensable, junto con el resto de los integrantes del grupo de investigación EOLO de la UPV/EHU.

A gradecer a Eduardo Zorita por hacer posible la estancia en Geesthacht, no solo por lo que aprendí académicamente allí, que fue mucho, si no también por darme la posibilidad de vivir en otro país y conocer a tanta gente.

Esta tesis tampoco hubiera sido posible sin el apoyo del área de Meteorología de Tecnalia, a la que agradezco al completo, a lo largo de tanto tiempo todos han contribuido de alguna manera en ella. Especialmente mi agradecimiento a Kepa Otxoa de Alda, Joseba Egaña y mis meteorólogas favoritas.

Agradecer también la última etapa a la sección al Departamento de Matemáticas de la UPV/EHU, por acogerme y animarme en la recta final desde el primer día.

Esta aventura del doctorado ha sido significativamente ; −) mucho más divertida y llevadera gracias a mis amigos Maialen Martija-Diéz y Santi González-Rojí.

Por supuesto le doy las gracias a Aritz. A estas alturas un gran experto en revisiones por pares, matriculas, acreditaciones etc. Admiración por mi familia, mi padre Carlos, mi madre Lourdes y mis hermanas Goizane y Zuriñe, por preguntarme incansablemente sobre la tesis a sabiendas que la respuesta no sería breve. Al igual que mis amigas de la cuadrilla que han escuchado durante años que aun quedaba mucho para terminar. A Lexuri, que ha confiado tanto o más que Josune en que este proceso acabaría con éxito.

Gracias a todos por apoyarme todos estos años, pero sobre todo por quererme tal y como soy.

Índice de publicaciones

Artículos publicados:

1. **Carreno-Madinabeitia S**, Ibarra-Berastegi G, Sáenz J, Zorita E, Ulazia A. *Sensitivity studies for a hybrid numerical-statistical short-term wind and gust forecast at three locations in the basque country (Spain)*. Atmosphere (Basel) 2020;11. https://doi.org/10.3390/ATMOS11010045.

2. Sáenz J, **Carreno-Madinabeitia S**, Esnaola G, González-Rojí SJ, Ibarra-Berastegi G, Ulazia A. *The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models*. Geosci Model Dev 2020;13:3221–40. https://doi.org/10.5194/gmd-13-3221-2020.

3. **Carreno-Madinabeitia S**, Ibarra-Berastegi G, Sáenz J, Ulazia A. *Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900 to 2010)*. Energy 2021, 226, 120,364, https://doi.org/10.1016/j.energy.2021.120364.

Artículos en revisión:

4. González-Rojí SJ, **Carreno-Madinabeitia S**, Sáenz J, Ibarra-Berastegi G. *Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation*. Hydrology and Earth System Sciences.

Índice de figuras

Índice de tablas

1 SÍNTESIS

1.1 Introducción

EL viento ha sido siempre un elemento relevante de la naturaleza. De hecho, para las doctrinas anteriores a lo que ahora conocemos como *ciencia moderna*, el aire junto con la tierra, el fuego y el agua, constituían los cuatro elementos básicos de la naturaleza. Ya en la *ciencia moderna* el estudio del viento se enmarca dentro de la meteorología. Desde este punto de vista, el viento es un fenómeno físico que aparece como resultado de las diferencias de presión atmosférica combinadas con otros efectos como la rotación de la Tierra.

Dentro de la meteorología las variables por excelencia y por lo tanto las más utilizadas son la precipitación y temperatura. Realizando consultas avanzadas del tipo *TITLE-ABS-KEY* (*meteorology AND variable*), en *SCOPUS*, una de las bases de datos más grandes de citas revisadas por pares, se ha comprobado que la variables precipitación (*precipitation OR rainfall*) cuentan con más de 30 mil artículos científicos y la temperatura (*temperature*) más de 28 mil. El viento (*wind*), la variable de interés de esta tesis les sigue con más de 20 mil citas.

El motivo por el que el viento es un objeto de investigación interesante es porque es un factor determinante en sectores como la propagación de los incendios, la dispersión de la contaminación atmosférica, la generación de energía eólica y el mantenimiento de sus infraestructuras, la seguridad en los puertos y aeropuertos y la navegación, por mencionar algunos ejemplos. La predicción meteorológica del viento es otro ámbito importante. Su principal finalidad es informar a la población de lo que ocurrirá en los próximos días, pero también tiene una utilidad transversal en

casi todos los sectores afectados por el viento como los mencionados anteriormente.

Horizonte temporal	**Rango**
Nowcasting	[0 s, 2 h)
Muy corto	[2 h, 12 h)
Corto	[12 h, 3 días)
Medio	[3 días, 10 días)
Extendido	[10 días, 30 días)
Largo	[30 días, 2 años)
Climático	$\geqslant$ 2 años

Tabla 1.1: Clasificación del tipo de predicción dependiendo del horizonte temporal propuesto por la OMM

La predicción meteorológica contribuye a evitar catástrofes naturales, daños personales y materiales. En concreto, según un estudio realizado por el Consorcio de Compensación de Seguros Español en el periodo 1987-2015 (*Spanish Department of Economy Industry and Competitiveness*, 2016) de todas las pérdidas económicas que se atribuyen a riesgos de la naturaleza el 15 % son causados por vientos superiores a 120 km/h y tornados. Un robusto sistema de pronóstico del viento, particularmente en el caso de eventos extremos, puede prevenir los efectos derivados de los mismos, horas antes de que ocurran, lo que permite adoptar medidas paliativas para los daños esperados.

La clasificación de las técnicas utilizadas para realizar pronóstico de viento dependiendo del horizonte de predicción no es clara, por un lado la Organización Mundial de la Meteorología (OMM) propone la clasificación de la Tabla 1.1[1]. Por otro lado, tanto *Okumus and Dinler* (2016) como *Soman et al.* (2010) coindicen en que una buena clasificación podría ser la que se muestra en la Tabla 1.2.

Horizonte temporal	**Rango**
Muy corto	[0 s, 30 min)
Corto	[30 min, 6 h)
Medio	[6 h, 1 día)
Largo	$\geqslant$ 1 día

Tabla 1.2: Clasificación del tipo de predicción dependiendo del horizonte temporal propuesto para predicción de viento

Soman et al. (2010) además clasifica los métodos de predicción en cuatro grandes grupos (persistencia, métodos físicos, métodos estadísticos y métodos híbridos). El primer grupo o método de persistencia consiste en tomar como predicción el último registro de viento disponible como proponen *Okumus and Dinler* (2016) y aplican entre otros *Torres et al.* (2005) para comparar los resultados de su modelo de predicción de viento generado con la técnica ARMA. Aunque parezca increíble, para el

[1] https://www.wmo.int/pages/prog/www/DPFS/Manual/GDPS-Supplement5-AppI-4.html

horizonte de muy corto-corto plazo, este modelo es más preciso que muchos métodos físicos y estadísticos. El segundo grupo al que hace referencia es el de los modelos físicos. Estos modelos tienen en común que utilizan parametrizaciones basadas en la descripción física de la atmosfera, que se basan a su vez en la solución numérica de las leyes de conservación del momento lineal y la energía y la ecuación de continuidad para predecir la evolución de la atmósfera, simulando los procesos de escala inferior al tamaño de celda mediante parametrizaciones. Los modelos numéricos de predicción meteorológica NWP (*Numeric Weather Prediction*) son útiles sobre todo para realizar predicciones a largo plazo, aunque dentro del horizonte de predictibilidad determinista. Dentro de los modelos NWP se encuentran tanto los modelos numéricos globales, como los modelos de escala local. Los modelos numéricos globales realizan predicciones para todo el mundo con una resolución espacial baja, el modelo *Global Forecast System* (GFS) (*Environmental Modeling Center*, 2003) es un ejemplo de ellos. El GFS [1]cuenta con una resolución horizontal de 0.5°. En cambio, los modelos de escala local se emplean en la técnica de *dowscaling* numérico. El objetivo de esta técnica es reducir la escala espacial de los modelos globales a escala local. Por la vía numérica esta reducción se realiza anidando a los resultados de un modelo global un modelo de mayor resolución, como son, entre otros, el *Weather Research and Forecasting* (WRF) (*Skamarock et al.*, 2008), *Regional Atmospheric Modeling System* (RAMS) (*Pielke et al.*, 1992) o HIRLAM-5 (*Christensen et al.*, 2007). Los modelos de *downscaling* numérico permiten realizar simulaciones de hasta 1 km de resolución horizontal (aproximadamente 0.01° de resolución espacial). El tercer grupo de técnicas de predicción cubre los modelos estadísticos, es decir, los basados en series temporales (*Torres et al.*, 2005) y redes neuronales (*Kulkarni et al.*, 2008); y los basados en regresión vectorial (*Santamaría-Bonfil et al.*, 2016) apropiados para realizar predicciones a corto plazo. Por último, se describe el grupo de las estructuras híbridas donde se encuentran los modelos de *downscaling* generados a partir de la vía estadística. Esta metodología utiliza funciones de transferencia para obtener resultados locales a partir de predictores proporcionados por modelos NWP. La función de transferencia puede generarse a partir de diferentes métodos estadísticos como, por ejemplo, el método de análogos (*Valero et al.*, 2014), redes neuronales (*Fernández-Ferrero et al.*, 2009; *Zhao et al.*, 2016), filtros de Kalman (*Cassola and Burlando*, 2012), etc. De esta forma obtienen una predicción (o conjunto de predicciones para el caso probabilístico)(*Fernández-Ferrero et al.*, 2010) adecuadas para realizar predicciones a medio-largo plazo.

En esta tesis se propone combinar las series temporales y las observaciones con el enfoque de downscaling estadístico más propio de predicciones a varios días para mejorar las predicciones a corto-medio plazo. Por lo tanto, el ***primer objetivo*** consiste en aplicar este modelo híbrido de predicción a las componentes zonal (u) y meridional (v) de viento y racha, en un horizonte que va desde una hasta 24 horas.

Para que el método sea robusto, y con el fin de identificar la técnica óptima entre un conjunto de técnicas disponibles, se han generado predicciones con tres técnicas estadísticas: análogos, regresión lineal y *random forest*.

Además, con el fin de estudiar con mayor solidez la robustez del método pro-

[1] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

puesto y con el objetivo de poder generalizarlo, las predicciones se han realizado en tres ubicaciones diferentes, mar, costa e interior. Este es el motivo por el cual se ha seleccionado inicialmente como zona de estudio el País Vasco. El País Vasco se encuentra en el Golfo de Vizcaya en el norte de la Península Ibérica (PI), incluyendo una zona montañosa que puede llegar a 1000 m a tan solo 30 km de la costa (Figura 1.1). El gradiente topográfico es elevado y el régimen del viento está sujeto a importantes influencias costeras. Por este motivo, dispone de ubicaciones con características meteorológicas muy diferentes a muy pocos kilómetros de distancia. Dado además que la doctoranda ha desarrollado parte de su carrera profesional en Euskalmet, la disponibilidad de datos de alta resolución espacial y temporal en esa zona y el conocimiento de los problemas derivados de la implementación práctica de las técnicas en predicción operativa, han ayudado a considerar la zona de estudio como óptima para llevar a cabo esta tesis doctoral.

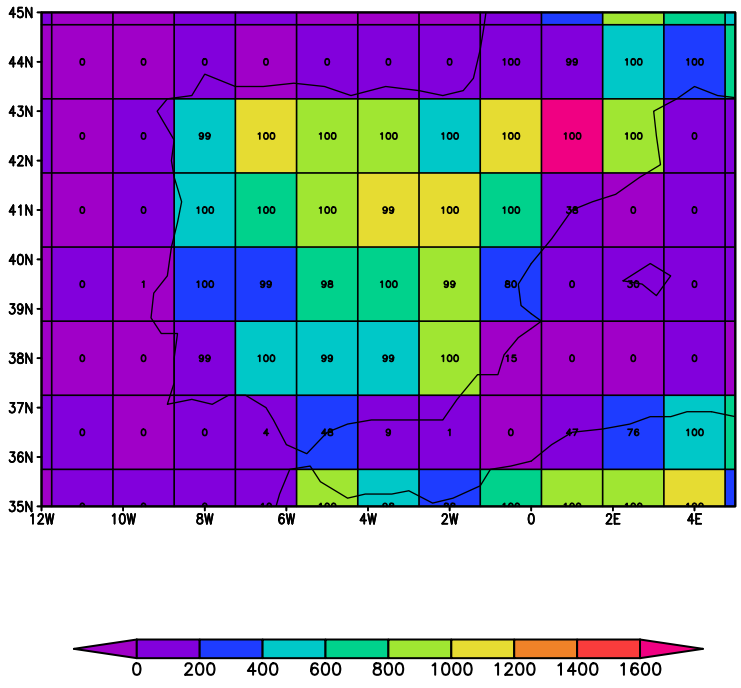

Figura 1.1: Topografía del reanálisis ERA5 con la resolución espacial de ERA20C en la PI (en números, el porcentaje de tierra de cada celda y en color, los metros de altitud)

Para comparar los resultados de los modelos implementados para las diferentes ubicaciones, técnicas estadísticas y horizontes de predicción, se han utilizado el coeficiente de determinación (R^2) y el error cuadrático medio (RMSE). A su vez, los modelos no se consideran adecuados si no son capaces de mejorar los resultados obtenidos directamente por el modelo global o, como se ha mencionado anteriormente, por el modelo de persistencia. En este caso, todas las verificaciones de las variables de estudio (u,v y racha) se han realizado de forma independiente.

De las variables empleadas en la primera parte de esta tesis (componentes zonal y meridional de viento y racha), la más difícil de predecir es la racha de viento. En

el caso de las variables instrumentales esto se debe a que la racha (en este caso, el viento máximo registrado en una hora) registra valores extremos a diferencia de u y v que reflejan valores medios. Los resultados de esta primera parte de la tesis nos han llevado a seguir investigando los valores extremos de viento y preguntarnos ¿cómo se deben analizar?

Los extremos de viento se dan mayoritariamente en las tormentas o durante eventos de tiempo severo (por ejemplo, durante el desarrollo de ciclogénesis explosivas). Existe una tradición sólidamente fundada en conceptos de meteorología dinámica y por la cual se llevan a cabo estudios de diagnóstico dinámico o termodinámico de eventos severos de viento como es el caso del ciclón extratropical Klaus (*Liberato et al.*, 2011). En este tipo de diagnósticos se analizan variables tales como la vorticidad potencial y el calentamiento diabático, las anomalías superficiales de temperatura del agua del mar y los flujos de calor latente (*Ludwig et al.*, 2014) en diversas formulaciones. Otros análisis se basan en el estudio de los términos involucrados en la tendencia de la presión (*Fink et al.*, 2012) o el análisis lagrangiano de los transportes de humedad (*Liberato et al.*, 2013). En base a este tipo de diagnósticos, se ha podido descubrir que, si bien hay factores comunes a los ciclones explosivos que afectan a la estructura de la corriente de chorro, las zonas de divergencia horizontal del viento, la estructura baroclínica de la atmósfera o el aporte de energía mediante el flujo de calor latente (*Ludwig et al.*, 2014), no todos los factores son igualmente importante en el desarrollo de todos los eventos, ya que en el caso de Xynthia (Febrero de 2010) los flujos de calor latente fueron el factor más importante (*Fink et al.*, 2012; *Ludwig et al.*, 2014), mientras que Kyrill (Enero de 2001) muestra mayor dependencia de la baroclinicidad del flujo (*Fink et al.*, 2012). Dado que esta tesis se centra en el análisis de períodos muy largos de datos (mayores de diez años por lo general), no se llevan a cabo diagnósticos detallados de eventos individuales tales como los que se realizan en las publicaciones mencionadas anteriormente. Los balances de vorticidad, energía, etc. se toman por buenos tal y como vienen representados en los modelos numéricos empleados. En lo referente al estudio de la inestabilidad convectiva, el desarrollo del paquete de código abierto *aiRthermo* nos ha permitido realizar algunos análisis de sensibilidad de los resultados obtenidos para *Convectively Avaiable Potential Energy* (CAPE) y *Convective INhibition* (CIN) (*Moncrieff*, 1981) en las diversas formulaciones del modelo WRF, pero se considera que el resto de inestabilidades dinámicas quedan fuera del ámbito de esta tesis. Igualmente, esta herramienta (*aiRthermo*), se ha utilizado para calcular la dependencia de la energía eólica disponible con la presión o la temperatura vía la consideración de la variabilidad de la densidad del aire.

En meteorología, la evaluación de las condiciones de inestabilidad convectiva en la atmósfera se basa típicamente en el cálculo de índices de inestabilidad, los más conocidos son los ya mencionados CAPE y CIN pero también *Lifted Index* (LI) (*Galway*, 1956), *K-Index* (*George*, 1960), *Total Totals index* (TT) (*Miller*, 1975) y *Showalter Index* (SI) (*Showalter*, 1953).

La convección profunda y las tormentas eléctricas son provocadas por tres factores: niveles altos de humedad en la capa límite (PBL), inestabilidad potencial y ascensos forzados (*forcing lifting*) (*Gascón et al.*, 2015; *Holley et al.*, 2014; *Johns and Doswell*, 1992; *McNulty*, 1995). CAPE y CIN proporcionan información de los

dos primeros. Es decir, CAPE y CIN ponen las condiciones para que en esa zona se puedan comenzar a desarrollar procesos convectivos atmosféricos y además proporcionan información sobre la intensidad de esos procesos (*Riemann-Campe et al.*, 2009). Sin embargo, estudios previos (*Angus et al.*, 1988; *López et al.*, 2001) sugieren que CAPE no debe usarse solo, sino que debe combinarse con otros índices. El factor final (elevación forzada) lo puede dar la orografía (*Doswell et al.*, 1998; *Siedlecki*, 2009), la convergencia de los flujos de humedad horizontales (*McNulty*, 1995) o en las regiones de costa las brisas (*van Delden*, 2001). Por tanto, la alta resolución espacial y temporal es importante para este tipo de estudios centrados en convección atmosférica. Es decir, es recomendable calcular los índices mencionados a partir de un modelo de alta resolución (*Siedlecki*, 2009). En este caso, para realizar el análisis se ha seleccionado el modelo numérico de alta resolución WRF. Como se ha expuesto anteriormente, los factores dinámicos han sido tenidos en cuenta por el modelo numérico empleado durante su integración junto con condiciones iniciales y contornos adecuados y, en nuestro caso, realizando asimilación 3DVAR cada 6 horas.

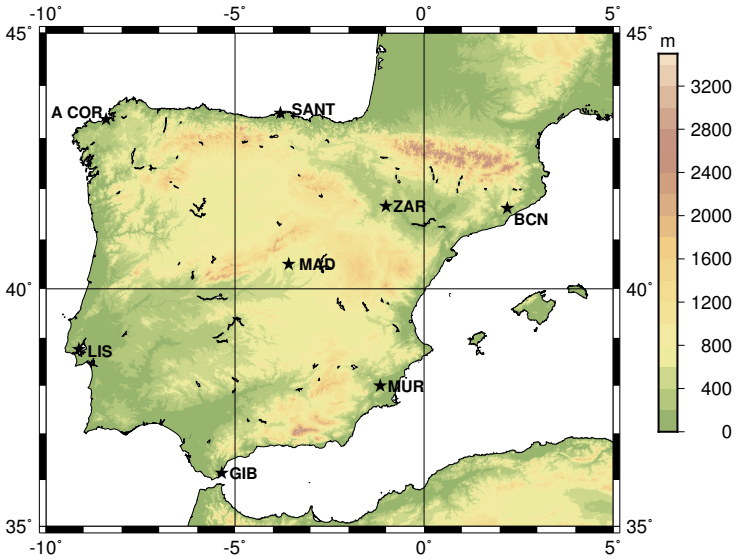

Figura 1.2: Mapa de la PI donde se muestran las ubicaciones donde habitualmente se lanzan los radiosondeos (*Sáenz et al.*, 2019)

Relacionado con el estudio de los valores extremos de viento en esta tesis se ha realizado el análisis de los índices CAPE, CIN y TT utilizando para ello las dos simulaciones diferentes del modelo WRF generadas recientemente por el grupo de investigación en el que ha trabajado la doctoranda. En este caso, una de las simulaciones (N) se ha llevado a cabo simplemente anidando el modelo WRF en el reanálisis ERA-Interim (*Dee et al.*, 2011), mientras que la segunda simulación (D) se ha llevado a cabo optimizando cada 6 horas las condiciones iniciales empleando asimilación de datos y anidada en las mismas condiciones de contorno proporcionadas por ERA-Interim. El uso de ambas simulaciones permite evaluar el impacto del uso de la asimilación de datos en la calidad de la inestabilidad atmosférica simulada. Como referencia para validar los datos de WRF, se han utilizado observaciones de datos de radiosondeo procedentes tanto de la base de datos homogeneizada IGRA (*Durre et al.*, 2006), como los proporcionados por *University of Wyoming* (2020).

Debido a que en el País Vasco los radiosondeos disponibles son muy limitados, en esta parte del análisis ha sido necesario ampliar la zona de estudio a la Península Ibérica (PI) (Figura 1.2). Esta decisión se ha visto afianzada por los artículos que destacan la región del País Vasco y el Mar Mediterráneo como lugares donde se forman tormentas severas (*Romero et al.*, 2007; *Siedlecki*, 2009; *Viceto et al.*, 2017).

En el artículo enviado para su publicación (*González-Rojí et al.*, 2021) la validación de los resultado de los índices en las diferentes simulaciones de WRF y las diferentes bases de datos observados tomadas como referencias, se ha utilizado el diagrama de Taylor (*Taylor*, 2001). Este diagrama es de gran utilidad gracias a que compara de forma visual los valores de desviación estándar (S), R^2 y RMSE de varios modelos de predicción de forma conjunta. Los dos últimos índices, como se ha mencionado anteriormente, se han utilizado también en el momento de validar de forma independiente las predicciones obtenidas de las componentes zonal y meridional (variables u y v) de viento y racha. En el momento de validar las predicciones de viento a corto-medio plazo se ha apreciado la necesidad de comparar a la vez los resultados obtenidos de las dos componentes del viento.

Combinando estas dos ideas, por un lado, la necesidad de validar las variables vectoriales bidimensionales en una sola gráfica y por el otro la idoneidad del diagrama de Taylor (*Taylor*, 2001) para comparar diferentes simulaciones o bases de datos respecto a una referencia en torno a las métricas de validación mencionadas anteriormente, se germinó el siguiente objetivo de esta tesis.

Este ***segundo objetivo*** consiste en desarrollar una nueva metodología en la que un diagrama permite de forma visual comparar diferentes bases de datos vectoriales. Este diagrama, al que se le ha llamado Sailor, se basa en la estructura matricial del RMSE para variables vectoriales.

Se ha propuesto el nombre de diagrama de Sailor (marinero en inglés) como un guiño al hecho de que es un diagrama que se puede utilizar para las variables como viento (objetivo de esta tesis) y corrientes, aunque también es válido para cualquier variable bidimensional que se desee evaluar. Y también porque este grafico está inspirado en el diagrama de Taylor. El nombre puede derivarse de la palabra Taylor simplemente cambiando dos letras de ésta (dos letras que equivalen al número de dimensiones utilizadas en el diagrama).

Por último, se ha centrado la atención en el ámbito de la energía eólica y su variabilidad a largo plazo, en escala climática. En los últimos 40 años el interés por la energía eólica se ha revivido debido a la falta de seguridad del suministro energético a partir de combustibles fósiles y los problemas medio ambientales que estos acarrean (*Kaldellis and Zafirakis*, 2011). Por este motivo, es un campo de investigación donde se espera gran desarrollo. Dentro de las energías renovables a nivel mundial, la energía eólica es la renovable con la que se genera más energía después de la energía hidroeléctrica con 591 y 1132 GW anuales respectivamente (*Murdock et al.*, 2019). En este mismo informe indican que España se encuentra dentro de los cinco países del mundo con mayor potencia instalada de generación de energía eólica. No solo eso, España y Portugal se encuentran por encima del tercer cuartil de los países de la Unión Europea que han instalado mayor potencia para generar energía eólica (*Papież et al.*, 2019). Estos datos concuerdan con los que publica *International Renewable Energy Agency* (2009) (IRENA), la Comisión

Europea de Energía *European Commission* (2019) y los planes nacionales de energía renovable *National Renewable Energy Action Plans* de España (*Gobierno de España*, 2020) y Portugal (*República Portuguesa*, 2018) que se detallan en la Tabla 1.3 junto con las previsiones de futuro hasta 2030. Los porcentajes de la Tabla 1.3 se calculan respecto a la potencia eólica instalada *onshore* por un lado y *offshore* por el otro.

	2019		**2020**	**2030**
	Onshore GW (%)	*Offshore* GW (%)	GW	GW
Mundo	594.4 (100)	20.31 (100)		
Europa	173.95 (29.26)	21.83 (77.12)	210	350
España	25.55 (4.30)	0.05 (0.02)	28.03	50.33
Portugal	5.23 (0.88)	0.08 (0.03)	5.4	9.2

Tabla 1.3: Presente y futuro de potencia eólica instalada

Además, como se muestra en la Tabla 1.3, la EU está comprometida a seguir desarrollando la energía eólica offshore. De hecho, actualmente, tiene la mayor potencia instalada de energía eólica flotante del mundo, alrededor del 77 % del total (Tabla 1.3). Además, la Comisión Europea de Energía espera que esta tecnología proporcione la mayor contribución al aumento de la energía renovable que se ha propuesto para los próximos años. Este planteamiento está alineado con la estrategia de energía eólica *offshore* dentro del acuerdo *The European Green Deal* (*European Commission*, 2019), en el que se están dando los primeros pasos hacia una UE con una economía sostenible. La energía eólica marina en Europa se centra actualmente en el Mar del Norte *The North Seas Energy Cooperation* (*European Commission*, 2020). Sin embargo, recientemente se está prestando también atención al Mar Mediterráneo y la costa atlántica.

Los indicadores más frecuentes para analizar la energía eólica son la densidad de potencia eólica (WPD) (*Manwell et al.*, 2010) y el factor de capacidad (CF) (*Masters*, 2013). El primero (WPD) se utiliza para evaluar la potencia del recurso eólico disponible en cada zona de estudio, el segundo (CF), en cambio, tiene en cuenta la carga promedio que puede producir una planta de energía de unas características concretas, teniendo en cuenta que con vientos muy flojos (o muy fuertes), los aerogeneradores no producen energía.

Debido a que el sector de la energía eólica marina se encuentra en desarrollo, en los últimos años se han publicado numerosos estudios científicos en torno a esta temática. El objetivo principal de la mayoría de estos estudios es analizar la idoneidad de las zonas para generar energía eólica en muchas y muy diferentes partes del mundo. Por ejemplo, en África (*Elsner*, 2019) han utilizado datos de satélites *Blended SeaWinds* (BSW) (*NOAA*, 2020a) en un periodo de 11 años para analizar el recurso de viento y su CF en dos tramos de mar, uno cerca de la costa y otro más alejado. En cambio, en Colombia (*Rueda-Bayona et al.*, 2019) han analizado la velocidad de viento (U) y la WPD en el periodo 1979-2015 con los datos de reanálisis NARR (*NOAA*, 2020b). El análisis de la energía eólica también es importante para el desarrollo de países o zonas más pequeñas, como es el caso de Kuwait (*Alkhalidi et al.*, 2019), el Mar Negro (*Islek et al.*, 2020) o Brunei Darassulam (*Salam et al.*,

2019) donde han calculado y analizado WPD en ubicaciones puntuales. Todos estos trabajos proporcionan una imagen fija del WPD y CF actual sin analizar la evolución pasada y/o futura.

Hasta la fecha, la mayor parte de los estudios que analizan la evolución de la energía eólica marina utilizan la variable WPD para sus proyecciones futuras (*Costoya et al.*, 2020; *Rusu and Rusu*, 2019; *Zheng et al.*, 2019). Estos estudios están orientados al análisis del impacto del cambio climático en los que subyace un interés de tipo económico ya que los cambios futuros en el recurso requerirán también cambios en la gestión de los parques eólicos actualmente funcionando (*JBA Consulting*, 2020). Pero ¿qué sabemos de la evolución de WPD de los últimos 100 años? ¿Y del CF?

A día de hoy ni la tendencia de WPD, ni el CF a lo largo del siglo XX han sido analizadas. Pero sí se han analizado otras variables de naturaleza también meteorológica como son la energía de las olas (*Penalba et al.*, 2018; *Ulazia et al.*, 2018), precipitación diaria (*Kim et al.*, 2019), velocidad de viento (U) (*Bett et al.*, 2017) y U *offshore* y altura de las olas (*Meucci et al.*, 2020; *Young and Ribal*, 2019).

Por todo lo anterior, el ***tercer objetivo*** de la tesis consiste en analizar las tendencias decadales observadas de WPD y CF alrededor de la PI a lo largo del siglo XX. Dado que la resolución espacial de los reanálisis con un período largo como el empleado en esta parte del trabajo es relativamente cruda, se ha preferido en este caso trabajar con toda la PI, ya que los resultados en esta escala espacial son más adecuados, considerando la resolución espacial de los datos disponibles. Asimismo, como se destaca en los siguientes artículos tanto la zona del Mar Mediterráneo (*Soukissian et al.*, 2017; *Ulazia et al.*, 2017), como la zona Atlántica (*Campos and Guedes Soares*, 2018; *Salvação and Guedes Soares*, 2018) son zonas con un alto WPD y por ello, es interesante desarrollar el estudio para todas sus áreas costeras.

Se deseea destacar también que en esta tesis todos los cálculos y la mayoría de los gráficos han sido realizados con el lenguaje de programación estadístico *open source* R (*R Core Team*, 2020) y el entorno de desarrollo integrado R-Studio (*RStudio Team*, 2020). A lo largo de estos años la doctoranda ha colaborado en el desarrollo de las librerías *aiRthermo*[1] y *SailoR*[2] en este entorno de programación las cuales están disponibles para uso público con licencia GPL de código abierto en el repositorio oficial de R.

1.2 Marco teórico

En este apartado se presentan las fuentes de datos y metodologías utilizadas a lo largo de la tesis.

[1] https://CRAN.R-project.org/package=aiRthermo
[2] https://CRAN.R-project.org/package=SailoR

1.2.1. Fuentes de datos

Para realizar la predicción de viento a corto-medio plazo se han utilizado tanto las observaciones de viento en tres ubicaciones concretas, como el modelo global ERA-Interim.

Observaciones de viento: las observaciones de viento del periodo de años 2007-2014 se han tomado de tres ubicaciones diferentes. La primera, la boya de Bilbao Bizkaia (43.64 °N, 3.05 °W) que depende de la agencia española de *Puertos del Estado* (2019) y dispone de datos horarios. Las otras dos, Punta Galea (43.38 °N y 3.04 °W, 61 m de altitud sobre el nivel del mar) y Alegría (42.84 °N y 2.52 °W, 545 m de elevación) son estaciones meteorológicas con observaciones cada 10 minutos. Estas dos estaciones están gestionadas por el Servicio Meteorológico Vasco (*Euskalmet - Agencia Vasca de Meteorología*, 2019).

Modelo global ERA-Interim: el modelo global seleccionado en el mismo periodo de datos (2007-2014) ha sido el reanálisis ERA-Interim del ECMWF (*Dee et al.*, 2011)[1]. El dominio seleccionado es un rectángulo que incluye a la PI (49.50 °N, 12 °W, 36 °N, 7.5 °E) con una resolución espacial de 0.75° × 0.75° (que corresponde en la latitud del estudio aproximadamente a 61 km de resolución). Las variables vinculadas con el viento seleccionadas de este modelo son: la presión media al nivel del mar (p_{msl}), la componente zonal de viento a 10 m (u_{10}); la componente meridional del viento a 10 m (v_{10}); y la temperatura a 2 m (t_2). Además, en los puntos de la rejilla más cercanos a las tres ubicaciones con datos observados (Bilbao-Bizkaia, Punta Galea y Alegria) se han recopilado los datos de prónostico de las variables u_{10} y v_{10} y racha de viento. En concreto, el modelo dispone de análisis en las ejecuciones de las t = 00 y 12 horas UTC y de predicciones cada 3 horas a partir de dichas ejecuciones. Es decir, tenemos datos de t + k, donde k = 3, 6, 9, 12, 15, 18, 21, 24 horas.

Relacionado con el primer objetivo de esta tesis y como contribución científica adicional se ha realizado la evaluación de las condiciones atmosféricas favorables para el desarrollo de evoluciones convectivas extremas utilizando dos simulaciones (N y D) del modelo regional WRF y los datos de radiosondeo de Wyoming e IGRA.

Modelo regional WRF: las simulaciones utilizadas en este estudio se han realizado con la versión 3.6.1 del modelo WRF para el período 2010-2014. En concreto se iniciaron el 1 de enero de 2009 desde un arranque en frío y en las cuales se utilizo el año 2009 como *spin-up* del modelo de tierra. En ambas simulaciones la resolución espacial es de 15 km y dispone de 51 niveles verticales hasta 20 hPa en coordenadas eta (η). En la primera, el modelo global ERA-Interim (*Dee et al.*, 2011) proporciona las condiciones iniciales y de contorno necesarias (experimento N). El segundo experimento, en cambio, presenta las mismas parametrizaciones que el anterior, pero con el paso adicional de asimilación de datos 3DVAR (*Barker et al.*, 2012, 2004) cada 6 horas (experimento D). En ambos casos disponemos de datos cada 3 horas del estado del modelo, a las t = 00, 06, 12 y 18 horas UTC de análisis (D) y a las

[1] https://www.ecmwf.int/

t = 03, 09, 15 y 21 horas UTC de pronóstico.

Datos de radiosondeos: para el periodo 2010-2014 se han obtenido datos de radiosondeos de dos fuentes, la primera directamente del servidor de la Universidad de Wyoming (*University of Wyoming*, 2020) y la segunda de la base de datos IGRA (*Durre et al.*, 2006). Solamente existen ocho estaciones de radiosondeo en la PI. En estas estaciones, los sondeos se llevan a cabo dos veces al día (t = 00 UTC y t = 12 UTC), a excepción de la estación de Lisboa que solo están disponibles una vez al día (t = 12 UTC).

Para el cálculo de WPD y CF a lo largo del siglo XX se han utilizado dos reanálisis diferentes, en ambos casos se ha seleccionado el dominio de la PI que se extiende desde (15 °W, 10 °E) longitud hasta (32.5 °N, 50 °N) latitud. Igualmente, para ambos se han seleccionado las variables u_{10}, v_{10}, componente u de viento a 100 m de altura (u_{100}), componente v de viento a 100 m (v_{10}), t_2, temperatura de rocío a 2 m (d_2) y p_{msl}.

Modelo reanálisis ERA20C: el reanálisis del siglo XX ERA20C, en adelante ERA20[1] dispone de datos entre los años 1900 y 2010 y cuenta con una resolución espacial de 1.5° x 1.5°(en las latitudes del estudio corresponde aproximadamente a 167 km de resolución). La resolución temporal de sus variables varía entre las tres horas y las seis horas. En este caso la variable t_2 es la que ha determinado que todo el estudio se realice con resolución temporal seis horaria.

Modelo reanálisis ERA5: el reanálisis ERA5 (*Hersbach*, 2016) ha sido también realizado por del ECMWF. Cuenta con datos horarios desde 1979 hasta la actualidad[2] y tiene una resolución espacial que 0.25° x 0.25° (aproximadamente 28 km de resolución). Tanto la resolución espacial como temporal de ERA5 se ha adecuado a la del reanálisis ERA20. En cambio, para definir los puntos de rejilla como tierra o mar se han seleccionado las características de los puntos de ERA5. Este reanálisis moderno está forzado por más observaciones que ERA20. De hecho, a diferencia del modelo ERA20 que cubre todo el siglo XX, en ERA5 se asimilan otro tipo de variables como por ejemplo, la radiación medida por satélite o las observaciones procedentes de radiosondeos.

1.2.2. Técnica híbrida de predicción de viento a corto-medio plazo

Para cada variable de estudio (u_{10}, v_{10} y racha), técnicas estadísticas (análogos, regresión lineal y *random forest*) y horizonte de predicción t + k (k = 1, 2, ..., 24 horas) se ha creado un modelo de predicción específico.

La técnica híbrida de predicción se basa en la generación de tres tipos de predictores que posteriormente se proponen como posibles predictores en el momento de construir los modelos matemáticos de predicción basados en las tres técnicas

[1] https://www.ecmwf.int/

[2] Durante la escritura de la tesis la cobertura temporal se ha extendido a partir de 1950

estadísticas que también se describen más adelante.

Para identificar los predictores es necesario procesar los datos observados de las ubicaciones de interés y las salidas del modelo global ERA-Interim tanto en modo análisis, como modo predicción. El origen y forma de obtención de cada tipo de predictor se detalla a continuación y se resume en la Tabla 1.4.

Última observación: se trata de obtener las observaciones de u y v, y la racha de viento a la altura de cada sensor cuando t = 0 horas. Con este predictor se pretende darle peso a la última observación registrada antes de realizar la predicción.

Predicción ERA-Interim: este tipo de predictores son los pronósticos de ERA-Interim (u, v, racha) para los puntos de rejilla más cercanos de las ubicaciones seleccionadas para t + 3, t + 6, . . . , t + 24 horas UTC. De esta forma se introduce la información del modelo global en el modelo de predicción.

Extended EOF: en el tercer tipo de predictores se pretende incluir tanto las observaciones de viento del pasado como la evolución temporal de las variables p_{msl}, u_{10}, v_{10} y t_2 del modo análisis de ERA-Interim en el dominio definido, también en el pasado, en concreto a las t - 0 h, t - 6 h, t - 12 h y t - 18 h UTC. Debido a que hay que gestionar una gran cantidad de datos y con el objetivo de reducir la dimensión del problema se han calculado funciones empíricas ortogonales extendidas (extEOF) (*Hannachi et al.*, 2007; *Weare and Nasstrom*, 1982). De esta forma, los predictores extEOF siguen capturando tanto los patrones temporales como espaciales relacionados con el viento, ya que, al calcular los componentes principales son seleccionados los que explican las mayores fracciones de varianza. Para evitar la diferencia de unidades de las variables involucradas en este cálculo se recomienda estandarizar (media = 0, varianza = 1) las variables originales, como se ha hecho en esta tesis.

Predictores	**Variables**	**Fuente**	**Horizonte (horas)**
Última obs.	u, v, racha	Obs	t+0
Pred ERA-I	u_{10}, v_{10}, racha	ERA-I pred	t+3, t+6, ..., t+24
ExtEOF	p_{msl},u_{10}, v_{10}, t_2 (dominio)	ERA-I análisis	t, t-6, t-12, t-18
	u_{10},v_{10}, racha	Obs	

Tabla 1.4: Tipos de predictores propuestos para la predición de veinto a corto-medio plazo

Una vez descritos los diferentes tipos de predictores, es momento de definir las técnicas estadísticas aplicadas. La primera de ellas consiste en la técnica de análogos (*Zorita and von Storch*, 1999). En esta técnica se pretende comparar el caso a predecir, definido por sus predictores, con los casos históricos en el archivo de entrenamiento y seleccionar la situación o situaciones más similares, convirtiéndose éstas

en la predicción de viento correspondiente a t + h horas. Es posible definir diferentes métricas de similitud (*Matulla et al.*, 2008) para realizar esta comparación. En esta tesis se han utilizado la norma euclidiana y la norma del coseno. Si definimos los predictores que forman parte del caso a predecir como un vector de dimensión p, $x = (x_1, x_2, \ldots, x_p)$, la distancia Euclidea entre los vectores x e y se define en la Ecuación 1.1.

$$d(x, y) = \|(x - y)\| = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2} \tag{1.1}$$

Mientras que la norma del coseno se define por la Ecuación 1.2.

$$d(x, y) = \|(x - y)\| = \frac{\sum_{i=1}^{p} (x_i y_i)^2}{\sqrt{\sum_{i=1}^{p} x_i^2 \sum_{i=1}^{p} y_i^2}} \tag{1.2}$$

La regresión lineal múltiple (*Weisberg*, 2005) es una técnica ampliamente conocida. En este caso los predictores que forman parte del modelo final se han incorporado a la ecuación siguiendo el criterio de información de Akaike (AIC) (*Akaike*, 1987). Aunque esta técnica parece muy simple, dentro del campo de la meteorología, existen artículos que indican que los algoritmos no lineales como las redes neuronales superan, pero no siempre de forma holgada, métodos simples como puede ser la regresión lineal y los análogos (*Eccel et al.*, 2007). Para poder comparar esta situación se ha decidido seleccionar la técnica de *random forest*.

Random forest (*Breiman*, 2001), al igual que las anteriores, es una técnica estadística de predicción. Esta técnica surge de la técnica de árboles de clasificación y la regresión (*CART – classification and regression trees*) (*Quinlan*, 1986), donde los árboles de regresión se perturban aleatoriamente para crear un bosque de árboles perturbados. Esta perturbación se realiza en dos niveles, el primero, seleccionando de forma aleatoria varias muestras de los datos originales para a continuación alimentar los diferentes árboles de regresión. Y la segunda, es en cada árbol, donde se seleccionan de forma aleatoria m predictores, dividiendo cada nodo en el mejor subconjunto de los predictores seleccionados. Una ventaja destacable de *random forest* sobre otras técnicas de regresión es que está libre de sobreajuste. Otra de las razones por la que se ha considerado esta técnica es debido a la capacidad que tiene de capturar relaciones altamente no lineales, como las que se sabe que están detrás de la física de la predicción del viento.

Los modelos generados con el procedimiento híbrido se han comparado directamente con los resultados del modelo global numérico y con el modelo de persistencia. Estos dos modelos se consideran de referencia, es decir, si no es posible realizar predicciones más precisas que estos dos modelos no merece la pena construir un modelo híbrido.

1.2.3. Índices de inestabilidad

Para obtener los índices de inestabilidad (CAPE, CIN y TT), bien del modelo numérico o bien de los radiosondeos, cuando ha sido necesario este cálculo, se ha utilizado la librería *aiRthermo* de R en su versión más reciente (versión 1.2.1).

Los índices de inestabilidad CAPE y CIN se han calculado por medio de las integrales verticales utilizando capas definidas por la resolución de la presión en los sondeos (utilizando todos los niveles disponibles). Las integrales de cada una de las capas definidas por perfiles lineales se calculan analíticamente, y la energía correspondiente a cada capa se acumula, obteniendo así el valor final de CAPE o CIN. La temperatura virtual ha sido utilizada en todas las integrales verticales empleadas para evaluar CAPE y CIN (*Doswell and Rasmussen*, 1994). Más detalles del cálculo de la evolución vertical de las parcelas de aire se pueden encontrar en *Sáenz et al.* (2019). El índice TT se ha calculado siguiendo la definición de *Miller* (1975) en la Ecuación 1.3.

$$TT = (t_{850} - t_{500}) + (d_{850} - t_{500}) \tag{1.3}$$

donde t_{850} y t_{500} son las temperaturas a 850 y 500 hPa, y d_{850} es la temperatura del punto de rocío a 850 hPa. Los índices de CAPE, CIN y TT se calculan utilizado las siguientes funciones de la librería *aiRthermo*:

CAPE_CIN(): esta función calcula los índices de inestabilidad atmosférica CAPE y CIN a partir de los datos de presión (Pa), temperatura (K) y razón de mezcla (kg/kg), obtenidos en un radiosondeo después de definir los valores iniciales de la parcela, estimando CAPE y CIN por integración numérica de los perfiles atmosféricos.

TTIndex(): el índice de inestabilidad TT (°C) se calcula a partir de la presión, temperatura (K) y razón de mezcla (kg/kg) obtenidos en un sondeo vertical.

1.2.4. Indicadores de energía eólica WPD y CF

Como se ha mencionado anteriormente en la introducción, WPD indica la energía eólica disponible en cada lugar. Es decir, la potencia media anual disponible por metro cuadrado de área del rotor de una turbina. Como se muestra en la Ecuación 1.4 este cálculo incluye tanto la información de U como la de la densidad del aire (ρ) (*Manwell et al.*, 2010).

$$WPD = \frac{1}{2}\rho(U_{90})^3 \tag{1.4}$$

Para este estudio se ha seleccionado una turbina a 90 metros de altura, para la cual se ha de calcular U a 90 m, que corresponde a la altura del buje de la turbina referencial *offshore* de 5 MV de *National Renewable Energy Laboratory* (NREL) (*Jonkman et al.*, 2009). Como los valores de la U cambian verticalmente con la altura de forma logarítmica, es necesario calcular la rugosidad de la superficie del mar z_0 a partir de la U a 10 y 100 m. Para ello utilizamos las ecuaciones 1.5 y 1.6.

En un segundo paso, con la velocidad a 90 m y la Ecuación 1.4 se ha obtenido WPD.

$$\frac{U_{100}}{U_{10}} = \frac{log(100/z_0)}{log(10/z_0)} \tag{1.5}$$

$$U_{90} = U_{10}\frac{log(90/z_0)}{log(10/z_0)} \tag{1.6}$$

CF, en cambio mide la cantidad de energía producida respecto a la que se produciría funcionando siempre a potencia nominal (P_R). El CF es ampliamente utilizado para medir la carga promedio de cualquier planta de energía como son los parques eólicos (*Barthelmie and Jensen*, 2010; *Boccard*, 2009; *Komusanac et al.*, 2019) o las plantas fotovoltaicas (*Izquierdo et al.*, 2010; *Wang et al.*, 2017) entre otros. Se puede utilizar para comparar diferentes instalaciones entre sí o incluso diferentes tecnologías (*Cheng and Yu*, 2013). Para calcular CF de una turbina comercial, primero se ha normalizado la velocidad de viento (U_n) con la densidad de aire (Ecuación 1.7) y a continuación se ha obtenido el CF con la Ecuación 1.8 (*Masters*, 2013). En esta estimación se utilizaron distintas turbinas de referencia, con potencia nominal entre 1 y 2 MW (*Masters*, 2013). Además, el término restante de esa ecuación que introduce las características de la turbina (P_R y D) es irrelevante al calcular la diferencia absoluta de CF en porcentaje, es decir, sólo interviene la altura del buje.

$$U_n = U(\rho/\rho_0)^{1/3} \tag{1.7}$$

$$CF = 0.087U_n - P_R/D^2 \tag{1.8}$$

donde P_R es la potencia nominal en kW de la turbina y D el diámetro en metros.

En este caso hemos querido tener en cuenta la densidad de aire correspondiente en cada punto de rejilla en cada momento (*Ulazia et al.*, 2017, 2019a,b,c). Para ello, se ha seguido la metodología detallada por *Ibarra-Berastegi et al.* (2019). En este artículo se explica cómo la densidad del aire depende de las variables p_{msl} en Pa (presión a nivel del mar), temperatura virtual (t_v) en K y $R_d \approx 287\ JK^{-1}kg^{-1}$ que corresponde a la constante de la mezcla de gases que forma aire seco y están relacionadas como se observa en la Ecuación 1.9.

$$\rho = \frac{p_{msl}}{R_d t_v} \tag{1.9}$$

Para obtener la t_v se ha utilizado la Ecuación 1.10. La temperatura virtual depende de las variables t_2 (temperatura a 2 m), p_{msl} y la relación ε (Ecuación 1.11) entre las constantes de gas que corresponden a aire seco R_d y vapor de agua R_w $\approx$487 $JK^{-1}kg^{-1}$.

$$t_v = \frac{t_2}{1-(1-\varepsilon)\frac{e}{p_{msl}}} \tag{1.10}$$

$$\varepsilon = \frac{R_d}{R_w} \approx 0.622 \tag{1.11}$$

Por lo tanto, para obtener la densidad del aire es necesario tener en cuenta es-

tas ecuaciones y aproximaciones. Para ello en esta tesis se ha utilizado la librería *aiRthermo* (*Sáenz et al.*, 2019). Esta librería, permite obtener los resultados de una forma directa, ya que tiene funciones implementadas para aplicar las ecuaciones detalladas anteriormente. En este caso, se han utilizado las siguientes funciones de *aiRthermo*:

TTdP2rh(): con esta función se calcula la humedad relativa (rh) a partir de la presión (Pa), t_2 (K) y la temperatura del punto de rocío (d_2) (K). La temperatura de rocío es la temperatura en la que el vapor de agua que contiene el aire empieza a condensarse. Las unidades de rh son %.

rh2w(): se utiliza para calcular la razón de mezcla (w) a partir de rh (%), presión (Pa) y t_2 (K). La razón de mezcla se define como la masa de vapor de agua por la masa de aire seco de una parcela. Las unidades de w vienen dadas por kg/kg.

DensityMoistAir(): por último, esta función permite calcular la densidad de aire en kg m^{-3} a partir de presión (Pa), t_2 (K) y w (kg/kg).

1.2.5. Ajuste de los datos de reanálisis con la técnica *quantile matching*

El reanálisis del siglo XX ERA20, es un reanálisis que asimila menos observaciones que los reanálisis del periodo instrumental moderno como es ERA5. Por tanto, el objetivo del ajuste de los datos de este reanálisis con ERA5 es trasladar la información adicional que contiene ERA5 a ERA20. Para realizar esta calibración se ha utilizado la técnica de *quantile matching* que consiste en calcular por un lado los percentiles a corregir de ERA20 y por otro lado los percentiles de ERA5 durante los años en los que se dispone datos de ambos reanálisis (1979-2010). En la Figura 1.3 se ha realizado el ajuste para los datos de WPD del punto de rejilla más cercano (44°N, 3.5°W) a la boya de Bilbao Bizkaia (43.64°N, 3.05°W). El resultado de la función de densidad acumulada muestra cómo los datos en el periodo 1979-2010 de ERA20 se corrigen y pasan a ser más parecidos a los datos de ERA5. De esta forma el reanálisis con información más detallada sirve de base para corregir el más extenso obteniendo así el reanálisis resultante ERA20 ajustado (ERA20adj). La técnica de *quantile matching* ha sido ampliamente utilizada para analizar las tendencias en el siglo XX de otras variables meteorológicas. Por ejemplo, *Penalba et al.* (2018) y *Ulazia et al.* (2018) han analizado la variación del recurso energético de las olas en la costa oeste de Irlanda y en la costa de Chile a lo largo del siglo XX. En ambos casos han utilizado ERA20 corregido por ERA-Interim para realizar este estudio. *Kim et al.* (2019) han analizado la precipitación diaria en Corea del Sur durante el siglo XX. Para ello han utilizado ERA20 al que han eliminado el sesgo con la técnica de *quantile matching*. En este caso en vez de utilizar datos de un reanálisis más actual han utilizado datos observados. En cambio, para la variable U en superficie (*offshore* y *onshore*) en Europa, *Bett et al.* (2017) han aplicado esta misma técnica con el reanálisis 20CR (*Compo et al.*, 2011) del siglo XX ajustado con datos de ERA-Interim.

Se ha seleccionado la técnica de *quantile matching* debido a que como indican

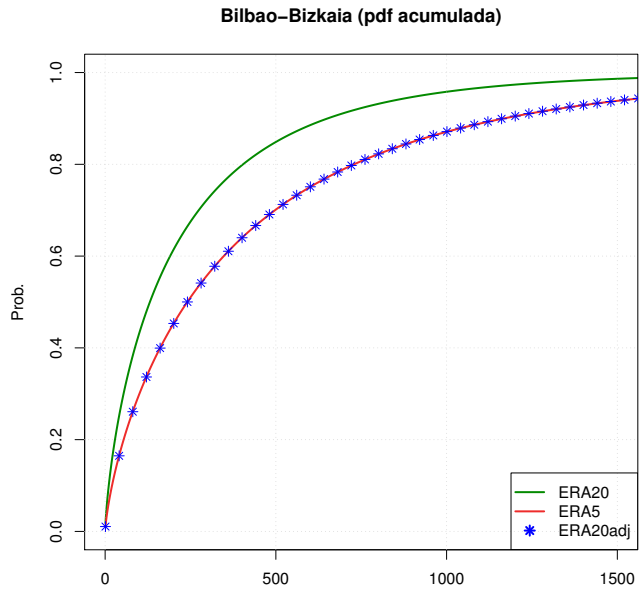

Figura 1.3: Función de distribución de Weibull acumulada de la variable WPD en el periodo 1979-2010 para los reanálisis ERA20, ERA5 y ERA20adj en el punto de rejilla más cercano a la boya Bilbao-Bizkaia

Gudmundsson et al. (2012) las transformaciones no paramétricas son mejores que las paramétricas o las que se basan directamente en la distribución. Dentro de las no paramétricas consideran que *quantile matching* es mejor que SSPLINE (*Hastie et al.*, 2001). En el mencionado artículo también señalan que el éxito de estas trasformaciones es probablemente su flexibilidad y la independencia de una función predeterminada. En este mismo artículo mencionan que los métodos no paramétricos tienen en promedio mejor habilidad para reducir errores sistemáticos, también en el caso de percentiles muy altos (extremos). En este estudio se ha utilizado la librería *qmap* del repositorio de R[1].

Esta técnica de ajuste de datos se ha aplicado directamente a las variables de estudio WPD y CF en vez de las variables originales. Esto es debido a que las relaciones entre la ρ y la U y WPD y CF son no lineales, por lo que se ha considerado más interesante calibrar directamente la variable objetivo para reducir los errores finales.

1.2.6. Cálculo de las tendencias decadales: Theil-Sen

Las tendencias decadales se han calculado utilizando la técnica de Theil-Sen (*Sen*, 1968; *Theil*, 1950). Este método es mucho más robusto que la regresión lineal simple, ya que calcula la mediana de las pendientes entre todos los posibles pares de puntos. Por ello, Theil-Sen logra evitar las distorsiones que puedan darse debido a los valores extremos, y permite obtener intervalos de confianza más robustos al 95 %.

[1] https://CRAN.R-project.org/package=qmap

Para el cálculo de las tendencias decadales, tanto temporales como espaciales, se ha partido de los datos mensuales de las variables WPD y CF. Con el objetivo de que el ciclo estacional no enmascare las tendencias, se han calculado las anomalías de todo el periodo (1900-2010) respecto a un periodo de referencia de 30 años (1981-2010) que se ha empleado para definir el ciclo estacional mencionado. Para representar estas tendencias, en el caso temporal se han generado diagramas de cajas, donde cada caja corresponde a los datos mensuales de cada década en todos los puntos de la rejilla. Junto con estos diagramas se ha calculado la tendencia de las medianas decadales con Theil-Sen.

Para poder realizar un análisis más profundo de las tendencias por zonas, se han calculado mapas que muestran en cada punto la pendiente de las tendencias decadales también calculados mediante el método Theil-Sen.

1.2.7. Validaciones

Para validar el primer objetivo de esta tesis se han utilizado los índices de validación RMSE y R^2. Estos mismos indicadores se muestran de manera visual en el diagrama de Taylor para comparar los índices de inestabilidad relacionados con eventos extremos de viento. Además, con la finalidad de avanzar hacia una herramienta de validación de variables bidimensionales se ha desarrollado un método nuevo, el diagrama de Sailor.

Es habitual utilizar RMSE y R^2 (*Müller*, 2011; *Rozas-Larraondo et al.*, 2014; *Valero et al.*, 2014) para validar las predicciones de viento a corto-medio plazo. El RMSE (*Stanski et al.*, 1989) (Ecuación 1.12) es una métrica simple que indica la magnitud de los errores medios ponderado según el cuadrado del error sin indicar la dirección de la desviación. Es decir, el RMSE de media da más peso a los errores grandes que a los errores pequeños. Por tanto, se utiliza cuando los errores grandes son particularmente indeseables. Solamente se puede utilizar con muestras suficientemente grandes debido a la sensibilidad a estos grandes errores. $RMSE = 0$ es el error correspondiente a un modelo perfecto.

$$RMSE = \left(\frac{1}{p} \sum_{i=1}^{p} (F_i - O_i)^2 \right)^{1/2}, \tag{1.12}$$

donde F_i es el valor de la predicción i-ésima y O_i es el valor de la observación i-ésima de una serie de longitud p.

El coeficiente de correlación de Pearson (R) (*WWRP/WGNE Joint Working Group on Forecast Verification Research*, 2015) en cambio (Ecuación 1.13), mide la correspondencia entre los valores observados y predichos. Es una buena medida de asociación lineal. En concreto, mide la dispersión de los pares de valores (F_i, O_i) respecto a la recta que mejor ajusta todo el conjunto de los datos. Este estadístico no tiene en cuenta el sesgo y es sensible a los valores atípicos.

$$R = \frac{\sum_{i=1}^{p} (F_i - \bar{F})(O_i - \bar{O})]}{\sqrt{\sum_{i=1}^{p} (F_i - \bar{F})^2} \sqrt{\sum_{i=1}^{p} (O_i - \bar{O})^2}}, \tag{1.13}$$

donde, de nuevo, F_i es el valor de la predicción i-ésima y O_i es el valor de la

observación i-ésima de una serie de longitud p y donde $\bar{F}$ y $\bar{O}$ son las medias correspondientes a los valores de predicción y de observación respectivamente. Además, R $= \pm 1$ implica correlación lineal perfecta y $R = 0$ que no existe esa correlación lineal entre ambos grupos de datos. En esta tesis se ha utilizado el coeficiente de determinación R^2 que corresponde al cuadrado del coeficiente de correlación de Pearson. Este estadístico también se puede interpretar como la proporción de varianza total explicada de la variable que pretende explicar.

A lo largo de esta tesis, para la mayor parte de los estadísticos utilizados se ha considerado un nivel de confianza del 95 %. El nivel de confianza se ha calculado utilizando la técnica de *bootstrap* (*Efron*, 1979), en concreto, remuestreando en 1000 ocasiones.

En este apartado también vamos a definir la desviación estándar muestral (S) debido a que este estadístico se utiliza en el diagrama de Taylor. En este caso la S calcula la dispersión de los datos en su mismo grupo de datos. En la Ecuación 1.14 se muestra la S correspondiente a los datos de pronóstico, pero de la misma manera es posible calcularla para los datos observados.

$$S_F = \left(\frac{1}{p} \sum_{i=1}^{p} (F_i - \bar{F})^2 \right)^{1/2} \tag{1.14}$$

El diagrama de Taylor relaciona a través del teorema del coseno los índices RMSE centrado (RMSE*) o lo que es lo mismo el RMSE calculado a partir de anomalías, R^2 y las desviaciones estándar. La relación geométrica entre los índices estadísticos que se trazan en el diagrama posibilita la comparación visual rápida entre diferentes realizaciones.

Con la ayuda de la Figura 1.4 se va a describir la forma en que se debe interpretar el diagrama de Taylor. Puesto que la referencia respecto a si misma siempre tiene un $R^2 = 1$ y un RMSE*$= 0$, se coloca en el eje x, en el punto donde corresponde su S.

En el diagrama en sus ejes X e Y se indica la S, los semicírculos que unen ambos ejes, por tanto, también indican la S. Si se consideran coordenadas polares, la distancia radial viene dada por la S de cada serie. El arco con la línea discontinua indica la desviación estándar de la referencia y los modelos que se encuentran en esa línea muestran la misma desviación que la referencia.

El eje de correlación se mide de forma angular respecto al eje Y, cuanto más cerca del eje Y se encuentre el punto del modelo menor correlación muestra respecto a la referencia, en cambio, cuanto más cerca esté del eje X más alta es su correlación respecto a la referencia. Por último, el RMSE* se presenta mediante los semicírculos centrados en el punto de referencia (*observed*). Cuanto más alejado esté el semicírculo en el que se encuentra el punto del modelo de la referencia mayor será el RMSE* entre la referencia y modelo a validar.

Por lo tanto, el punto del modelo que más cerca se encuentre del punto de la referencia es el que corresponde a la mejor realización. Como se ha dicho anteriormente el diagrama de Sailor se basa en la estructura bidimensional del RMSE de dimensiones 2 x 2. De igual forma que para el diagrama de Taylor, vamos a describir cómo se interpretaría un diagrama de Sailor.

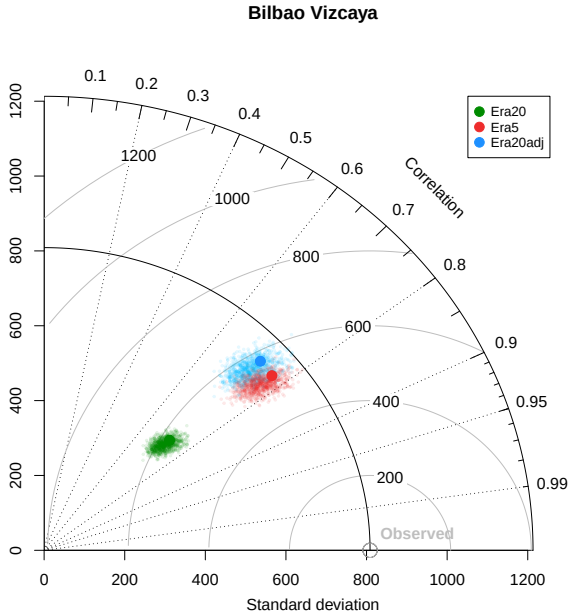

Figura 1.4: Ejemplo de diagrama de Taylor

Para interpretar el diagrama de Sailor y su tabla de índices de diagnóstico asociada vamos a fijarnos en el caso en el que las elipses están centradas en el punto medio de los datos observados o de referencia (Figura 1.5 y tabla 1.5). La elipse correspondiente a los datos de referencia u observaciones bidimensionales (**U**) se traza con una línea discontinua y su punto medio con un cuadrado, todo ello de color gris. Por lo tanto, para ver si los datos de los modelos bidimensionales (**V**) muestran sesgo con respecto a la referencia es necesario comparar la distancia de los círculos de diferentes colores desde el cuadrado gris. En este caso concreto, el MOD1 muestra sesgo respecto a las observaciones. De hecho, si nos fijamos en *biasMag* de la tabla 1.5 observamos un sesgo de 8.34 m/s. Si los ejes de la elipse del modelo son más grandes o pequeños que los ejes de referencia esto quiere decir que los valores del modelo están escalados frentes a los de la referencia. Por ejemplo, en el MOD4 observamos cómo las desviaciones típicas (σ_x y σ_y) del modelo son mayores que los correspondientes a la referencia. Además, si los semi-ejes están rotados, esto indica que la estructura de la covarianza del modelo es diferente a la de la referencia (MOD2). El índice que mejor indica la rotación de los semi-ejes respecto a la referencia es el valor absoluto del coeficiente de congruencia (*Cheng et al.*, 01 Jun. 1995) del primer EOF g_{11} (Ecuación 1.15). Este estadístico nos muestra la orientación relativa del primer vector propio entre la referencia y el modelo teniendo en cuenta el primer EOF del las observaciones y el modelo respectivamente.

$$g_{11} = |EOF_{u_1} \cdot EOF_{v_1}| \tag{1.15}$$

Por lo tanto, el MOD2 de la Figura 3 muestra esa rotación con un g_{11} de 0.87. Podría ser, que una elipse de modelo se superponga a la elipse de la referencia. En ese caso podríamos pensar que el modelo la simula perfectamente, pero existe una excepción, es necesario comprobar que entre ambos casos el RMSE es nulo

y la correlación es total. Cabe la posibilidad de que los datos del modelo estén desordenados y que exista esta contradicción. Por lo tanto, es necesario fijarnos en el índice RMSE que se muestra en el diagrama o en el RMSE y correlación que aparecen en la tabla de diagnóstico. Por lo tanto, el MOD3 de la Figura 1.5 no se trata de un modelo que simula perfectamente las observaciones, en realidad es un modelo que tiene los resultados permutados, ya que, su RMSE es de 1.52 m/s. La Tabla 1.5 para MOD3 a pesar de mostrar un coeficiente de congruencia de uno la R^2 es nula, así como un RMSE diferente de cero.

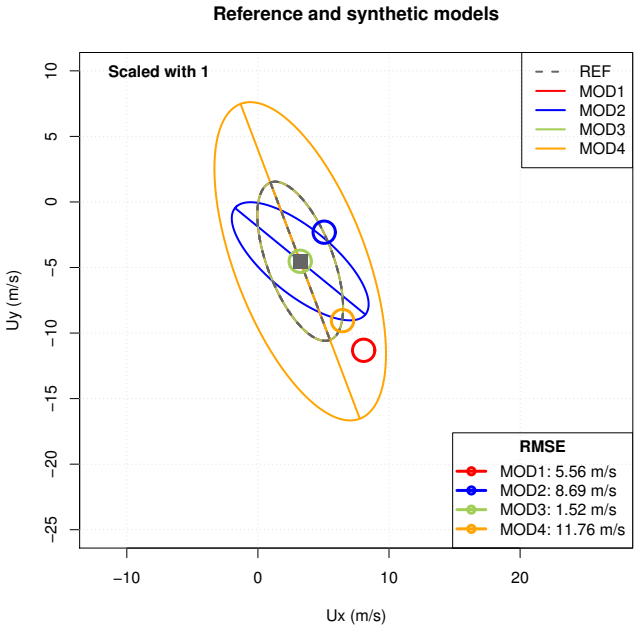

Figura 1.5: Ejemplo de diagrama de Sailor. Correponde a la Figura 4 b del artículo (*Sáenz et al.*, 2020)

	Name	σ_x	σ_y	g_{11}	R^2vec	biasMag	RMSE
1	Ref	6.42	2.52	1.00	2.00	0.00	0.00
2	MOD1	6.42	2.52	1.00	2.00	8.34	5.56
3	MOD2	6.42	2.52	0.87	2.00	2.88	8.69
4	MOD3	6.42	2.52	1.00	0.00	0.00	1.52
5	MOD4	12.84	5.03	1.00	2.00	5.56	11.76

Tabla 1.5: Ejemplo de tabla de índices de diagnóstico de Sailor. Corresponde a la Tabla 1 del artículo (*Sáenz et al.*, 2020)

Otro caso singular del diagrama de Sailor es en el que los semi-ejes son de la misma longitud y por lo tanto forman un círculo. Es decir, la varianza explicada por el EOF_{u1} y el EOF_{u2} es la misma. En este caso, los vectores propios podrían ir en cualquier dirección y dificultar la comparación de la rotación. Por este motivo, se ha añadido la excentricidad (ε) como índices de diagnóstico, el cual nos permite conocer la forma de la elipse (Ecuación 1.16). En este caso es recomendable cerciorarse de que la excentricidad no es nula antes de comparar las rotaciones de las elipses.

$$\varepsilon_u = \sqrt{1 - \frac{S_{u_1}^2}{S_{u_2}^2}} \tag{1.16}$$

Para generar el diagrama y los índices de diagnóstico de Sailor se ha creado la librería de código abierto bajo licencia GPL *SailoR* de R[1]. En esta librería las dos funciones principales son *SailoR.Indices()* y *SailoR.Plot()*.

SailoR.Indices(): esta función devuelve un objeto *data.frame* con todos los estadísticos necesarios para construir el diagrama de Sailor. Los argumentos necesarios para poder realizar los cálculos son:

- **ref**: un objeto de clase *data.frame* con las observaciones de referencia. Este objeto debe tener tres columnas, con el nombre, el componente zonal y el componente meridional.
- **mod**: un objeto de clase *data.frame* con los datos de los modelos. Este objeto debe tener tres columnas, con el nombre del modelo, el componente zonal y el componente meridional.
- **Ensembles**: si se establece este argumento como *TRUE*, los m modelos diferentes incluidos en el argumento **mod** se tomarán como un solo modelo y los datos de la referencia se repetirán m veces.

El resultado que nos proporciona esta función es un objeto *data.frame* que contiene los siguientes estadísticos, donde **U** se refiere al modelo de referencia y **V** a los modelos a validar frente a **U**.

- **meanU**: la media de los datos de **U** (referencia).
- **meanV**: la media de los datos de **V** (modelos).
- **RMSE**: RMSE entre **U** y **V**.
- **sdUx**: S de la componente zonal de **U**.
- **sdUy**: S de la componente meridional de **U**.
- **sdVx**: S de la componente zonal de **V**.
- **sdVy**: S de la componente zonal de **V**.
- **Eu**: matriz de EOF de los datos de **U**.
- **Ev**: matriz de EOF de los datos de **V**.
- **Sigmau**: matriz diagonalizada con las S de **U** (de dimensiones 2x2).
- **Sigmav**: matriz diagonalizada con las S de **V** (de dimensiones 2x2).

[1] https://CRAN.R-project.org/package=SailoR

- **TotVarU**: varianza total de los datos de **U**.
- **TorVarV**: varianza total de los datos de **V**.
- **thetau**: ángulo de EOF_1 (**U**) respecto al eje zonal.
- **thetav**: ángulo de EOF_1 (**V**) respecto al eje meridional.
- **thetavu**: ángulo de rotación desde el EOF de **U** hasta el EOF de **V**.
- **Rvu**: correlación vectorial cuadrática en dos dimensiones definido por *Breaker et al.* (1994).
- **EccentricityU**: excentricidad de las elipses de **U**.
- **EccentricityV**: excentricidad de las elipses de **V**.
- **congruenceEOF**: coeficiente de congruencia (valor absoluto) de los EOF_1 de **V** respecto de EOF_1 de **U**.

SailoR.Plot(): esta función genera el diagrama de Sailor llamando internamente a la función *SailoR.Indices()*. Los argumentos que permiten que el diagrama sea flexible son los siguientes:

- **ref**: igual que para *SailoR.Indices()*.
- **mod**: igual que para *SailoR.Indices()*.
- **referenceName**: en este argumento se puede proporcionar un nombre externo para el conjunto de datos de referencia, el cual se mostrará en las leyendas.
- **ColourList**: en este parámetro se puede introducir un vector con los colores que se desee utilizar para trazar los modelos. El color gris se reserva para los datos de referencia. Si es *NULL*, se utilizan los colores predeterminados: *gray40*, *red*, *blue*, *darkolivegreen3*, *orange*, *seagreen3*, *gold*, *purple* y *pink*.
- **linestype**: se puede indicar un vector con los tipos de líneas seleccionados para trazar las elipses de los modelos. La opción *dashed* está reservada para la elipse de referencia. Las otras opciones posibles son *solid*, *dotted*, *dotdash*, *longdash* o *twodash*.
- **bias_pch**: en este vector se pueden incluir la forma de los símbolos seleccionados para los puntos medios. La opción predeterminada proporciona un círculo cuando *docenter* es igual *TRUE* y un punto cuando es *FALSE*.
- **sfactor**: se trata del factor común de escala que se aplicará a las elipses de cada modelo.
- **docenter**: este argumento debe de ser *TRUE*, si se desea centrar todas las elipses sobre la de referencia.
- **xlim**: límite del eje X del gráfico.

- **ylim**: límite del eje Y del gráfico.
- **xlab**: título del eje X.
- **ylab**: título del eje Y.
- **plotmain**: título general del diagrama.
- **plotlegend**: este argumento debe ser *TRUE*, si se desea agregar al gráfico una leyenda predeterminada.
- **plotRMSElegend**: este argumento debe ser *TRUE*, si se desea agregar al gráfico los índices RMSE de cada modelo.
- **RMES_legend_units**: las unidades que se quieran incluir en la leyenda RMSE.
- **RMSE_legend_Rounding**: número de decimales de RMSE que se quiera mostrar en su leyenda.
- **plotscalelegend**: este argumento debe ser *TRUE*, si se desea agregar al gráfico la leyenda donde se muestra el factor de escala que se ha aplicado a las elipses.
- **Ensembles**: igual que para SailoR.Indices().

La técnica matemática detrás de estos cálculos se presenta en el artículo *The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models* (*Sáenz et al.*, 2020) que se adjunta como segunda contribución en esta tesis.

1.3 Objetivos

En esta tesis se ha realizado un trabajo de *downscaling* estadístico en modo pronóstico, así como dinámico en modo análisis. También se han calculado índices de inestabilidad, se han desarrollado herramientas de evaluación de modelos y se han calculado tendencias. Las variables objeto han sido el viento y el ciclo de agua en la atmósfera y las escalas temporales van desde horas hasta décadas.

Por ello, en este apartado se describen tres objetivos principales junto con los hitos necesarios para alcanzarlos.

1.3.1. Primer objetivo

Este objetivo trata de desarrollar y testar una nueva metodología para realizar predicción de viento a corto-medio plazo:

- Acreditar el rendimiento de diferentes técnicas estadísticas (en este caso regresión lineal, análogos y *random forest*) y analizar si la bondad de una técnica prevalece frente al resto. Además, para obtener resultados más robustos se han comparado los resultados obtenidos con los modelos híbridos y con las predicciones de referencia que son la persistencia y los resultados del modelo global.
- Analizar el comportamiento de los tres tipos de predictores proporcionados (última observación, predicción del modelo global y extEOF). Es decir, se ha estudiado cuál es su relevancia y si han sido seleccionados para el modelo final.
- Generalizar el método para diferentes emplazamientos, analizando las similitudes y disimilitudes entre los resultados de las ubicaciones seleccionadas definidas como océano, costa e interior.

Este objetivo se alcanzó en la publicación *Sensitivity studies for a hybrid numerical-statistical short-term wind and gust forecast at three locations in the Basque Country (Spain)* (*Carreno-Madinabeitia et al.*, 2020). Dentro de este primer objetivo, relacionado con los extremos de viento y en proceso de publicación, se analiza si el modelo de *downscaling* numérico reproduce de forma adecuada los índices de inestabilidad CAPE, CIN y TT para la PI.

- Se desarrolla una comparativa entre las dos simulaciones del modelo WRF de *downscaling* numérico, los índices de las base de datos IGRA y los registros de Wyoming. Así como, con los índices obtenidos a partir de los registros de Wyoming calculados con la librería *aiRthermo*. Esta comparativa se va a realizar en los ocho puntos de radiosondeo disponibles en la PI.
- Realizar un análisis estacional de los resultados. Debido a que, las tormentas y los eventos de tiempo severo se dan más en unas épocas del año que en otras.
- Generalizar y analizar los mapas de los tres índices de inestabilidad analizados en este trabajo. Esto es posible a partir de los datos espaciales para toda la PI del modelo de *downscaling* numérico.

Este estudio se encuentra en revisión en *Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation* (*González-Rojí et al.*, 2021).

1.3.2. Segundo objetivo

En este objetivo se desarrolla un nuevo diagrama (Sailor) para validar de forma conjunta y visual las variables vectoriales bidimensionales.

- A partir del RMSE bidimensional se ha desarrollado la base algebraica para que el diagrama tenga consistencia matemática.
- Además, puesto que los datos obtenidos a través de un sistema de predicción por conjuntos son muy comunes en el campo de la meteorología, se ha adaptado la metodología del diagrama para que también sea viable para este tipo de datos.

- Extensión de la metodología para que también sea posible trabajar con datos espaciales.
- Una vez desarrollada la metodología, implementación de una herramienta accesible y con licencia de código abierto para su uso por la comunidad científica.

Este objetivo es objeto de la publicación *The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models* (*Sáenz et al.*, 2020).

1.3.3. Tercer objetivo

El último objetivo analiza la variabilidad de las variables WPD y CF *offshore* en el siglo XX alrededor de la PI.

- Para ello es necesario obtener datos de calidad del siglo XX.
- Identificar desde el punto de vista de WPD y CF las zonas con mayor incidencia de energía eólica *offshore* y por lo tanto adecuadas para la generación de este tipo de energía renovable.
- Caracterizar las tendencias decadales tanto de manera temporal como espacial de WPD y CF a lo largo del siglo XX.
- Comparar los resultados obtenidos con las proyecciones futuras ya desarrolladas en otras investigaciones.
- Estimar las implicaciones de estas tendencias en la industria de la energía eólica.

Este objetivo se cubre en la publicación de titulo *Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900 to 2010)* (*Carreno-Madinabeitia et al.*, 2021). En la Tabla 1.6 se relaciona cada objetivo con la metodología y las fuentes de datos utilizadas en cada artículo.

1.4 Resumen y discusión de los resultados

En este punto se ha realizado un resumen de los resultados obtenidos a lo largo de esta tesis, divididos en tres grandes estudios. Los resultados obtenidos al implementar una metodología para realizar la predicción de viento a corto-medio plazo (*Carreno-Madinabeitia et al.*, 2020) se describen a continuación.

- Se ha conseguido establecer un método general para realizar la predicción de viento a corto-medio plazo. Los resultados obtenidos son robustos, ya que son consistentes para las diferentes ubicaciones seleccionadas (océano, costa e interior) con un nivel de confianza del 95 %.

Objetivo	**Metodología**	**Fuentes datos**
Predicción viento	*Downscaling* estadístico: - Análogos - Regresión lineal - *Random forest* - Persistencia	- Obs. viento - Modelo ERA-I
Índices inestabilidad	- CAPE - CIN - TT	- Modelo WRF - Radiosondeos
Diagrama Sailor	Validación	- Variables vectoriales
Evolución WPD y CF	- Ajuste *quantile matching* - Tendencias Theil-Sen	- Reanálisis ERA20 - Reanálisis ERA5

Tabla 1.6: Resumen de objetivos, metodología y fuentes de datos utilizados en esta tesis

- Si se pone el foco en identificar la mejor técnica estadística, en general *random forest* es un poco mejor que el resto (regresión lineal y análogos), pero no siempre ni de forma significativa. De este resultado se concluye que, en este caso, la técnica estadística utilizada no es un factor que marque claramente la diferencia, es más importante la información que se introduce a través de los predictores.

- La validación entre la mejor técnica estadística (*random forest*) y el uso simple de persistencia o pronósticos del modelo global ERA-Interim indica que random forest supera a estos más allá del horizonte de tiempo de $1-4$ horas al 95 % de confianza.

- Por otro lado, la persistencia es la mejor opción para el horizonte de predicción $1-4$ horas. Este estudio es un claro ejemplo de la importancia de comparar los resultados de los modelos de predicción para un horizonte corto-medio, con los obtenidos por la persistencia.

- Al analizar los predictores que forman parte de los modelos finales, se ha ha detectado que los tres tipos de predictores (última observación, extEOF y predicciones del modelo global) son relevantes para el pronóstico. En este estudio, para todos los pronósticos y predicciones hasta $2-5$ h en el futuro, la entrada más influyente es la última observación registrada. Así, la memoria del sistema juega un papel clave en la construcción de la predicción y explica por qué los resultados de persistencia son mejores en las primeras horas de predicción. Más allá de este horizonte, las predicciones del modelo global ERA-Interim son el predictor con más peso dentro de los predictores seleccionados. El grupo de extEOF (aunque no los mismos en todos los casos) también juega un papel importante, como la segunda o tercera entrada más influyente, lo que

confirma la necesidad de utilizar los tres grupos de predictores para construir modelos híbridos.

Además, dentro del marco de investigación de este primer objetivo y dada la importancia de los eventos extremos de viento y como parte de contribución científica futura (*González-Rojí et al.*, 2021) se han analizado los índices de inestabilidad (CAPE, CIN y TT) necesarios para identificar las condiciones bajo las que se pueden desarrollar tormentas convectivas y los extremos de viento asociadas a ellos. Al comparar las simulaciones del modelo numérico de *downscaling* WRF se ha obtenido, como se preveía, que la simulación con asimilación de datos simula mejor los índices de inestabilidad que la simulación sin asimilación.

Tomando los datos de la base de datos IGRA como referencia, se observa que tanto los datos adquiridos directamente por la universidad de Wyoming, como los calculados con la librería *aiRthermo* son muy similares. Por este motivo, se dan por buenas las metodologías utilizadas por la universidad de Wyoming y la librería *aiRthermo*. Aunque, y como era de esperar debido a que las metodologías utilizadas para obtener los índices de CAPE y CIN tienen sus diferencias en aspectos tales como la forma de integración o el desarrollo de la parcela inicial desde la que se inicia la evolución vertical, las bases de datos muestran más similitudes entre sí en el momento de calcular TT que CAPE y CIN. Esto se debe a que CAPE y CIN son sensibles al estado de la parcela inicial desde la que se calcula la evolución inicial y TT no depende de este factor.

Siguiendo los objetivos marcados se han analizado los diferentes índices de inestabilidad en cada estación del año. En general, la diferencia estacional es pequeña en todas las ubicaciones. En invierno se ha observado que las masas de aire inestables se sitúan sobre todo en las ubicaciones de la costa Atlántica de la PI.

Una vez verificado que la simulación con asimilación de datos reproduce de forma adecuada los índices de inestabilidad de interés, se han generado los mapas para toda la PI. En estos mapas se observan patrones heterogéneos, pero también se ve cómo, del mismo modo que se intuía al analizar las ubicaciones de los radiosondeos, en invierno las masas de aire inestables se colocan en toda la costa del Atlántico, pero sobre todo en la esquina noroeste de la PI.

El resultado más importante al desarrollar el diagrama de Sailor (*Sáenz et al.*, 2020) ha sido haber logrado relacionar algebraicamente el RMSE de dos dimensiones con los ejes de las elipses. Basándose en esta metodología, este diagrama es capaz de comparar diferentes modelos frente a una observación o modelo de referencia. En el diagrama se puede observar directamente tanto el sesgo bidimensional de los diferentes modelos frente a la observación (comparando los puntos centrales de las elipses), su covarianza (en los semi-ejes) o el aumento o disminución del orden de los datos (tamaño de los semi-ejes).

Además, junto con el diagrama de Sailor se proponen diferentes índices de diagnóstico que lo acompañan para facilitar la interpretación de los resultados. En concreto, hay una situación que puede llevar a error en la que dos elipses son iguales pero los datos que representan no. Este sería el ejemplo de una simulación climática casi perfecta en la que no hay correspondencia temporal en el *tiempo* de la simulación, pero sí en su *clima*. Es decir, las PDFs de $\mathbf{U}$ y de $\mathbf{V}$ son iguales. En esta

situación los ejes principales están rotados y es necesario revisar el RMSE y R^2 para darse cuenta de que no se trata de modelos iguales. También es necesario analizar si la excentricidad (ε) es no nula, debido a que la interpretación de la rotación puede ser errónea en el caso en el que la elipse se degenera en una circunferencia.

Se ha desarrollado y puesto a disponible para la comunidad investigadora la librería *SailoR* de R CRAN. Con esta librería es posible generar el diagrama y los índices de validación. Además, *SailoR* permite, entre otras opciones, personalizar los diagramas aplicándoles un factor de escala o generar los gráficos centrando o no las elipses en la media de los datos de referencia. De esta forma independientemente de cuál sea la variable o los modelos a analizar es posible representar de forma clara los resultados en un diagrama. Incluso para datos espaciales y datos obtenidos a través de un sistema de predicción por conjuntos.

Para finalizar, quedan por mencionar los resultados obtenidos al analizar la variabilidad de WPD y CF en el siglo XX (*Carreno-Madinabeitia et al.*, 2021). En primer lugar, se ha demostrado que la corrección de los datos de reanálisis del siglo XX (ERA20) ajustados mediante la técnica *quantile matching* y un reanálisis actual (ERA5), son adecuados para llevar a cabo este estudio.

Utilizando los datos corregidos, el resultado medio obtenido en la PI para ambas variables de estudio en el periodo 1900-2010 a una altura de 90 m, es de WPD 582 Wm^{-2} y CF de 0.38. Las zonas con mayor potencial eólico son la zona Atlántica y el Golfo de León con WPD y CF mayores de 600 Wm^{-2} y 0.4 respectivamente. Además, destaca la zona de Cabo de Gata con valores de 590 Wm^{-2} y 0.37.

En esta investigación se ha demostrado también que las tendencias decadales tanto de WPD como de CF, son positivas y significativas en todo el dominio de estudio. Para WPD la pendiente media es de 15.86 $Wm^{-2}decada^{-1}$ y de 0.008 $decada^{-1}$ para el caso de CF. Esto implica que en las 11 décadas podría llegar a ser de 174 Wm^{-2} (0.088) para WPD (CF). Los máximos, al igual que para las intensidades medias, se dan en la zona Atlántica y el Golfo de León. En cambio, la pendiente media de Cabo de Gata no destaca como sí lo hacía como zona de mayor potencial eólico.

Puesto que estos resultados se han aplicado a una turbina de referencia de 5 MW (*Jonkman et al.*, 2009) y teniendo en cuenta esta evolución, para los parques eólicos offshore en las últimas 11 décadas habría supuesto un incremento de un 20 % en sus ganancias. Este resultado se podría alargar en el tiempo hasta finales del siglo XXI, si tenemos en cuenta que según *Zheng et al.* (2019) es posible que aumenten las zonas con WPD mayor a 400 Wm^{-2}.

Esta tendencia histórica es de gran interés dada la penetración esperada de la energía eólica marina en Europa y las posibles reducciones en el precio de la electricidad y las emisiones de CO_2 que pueden seguir en el futuro. La UE está comprometida con la expansión de la industria de las energías renovables, y la viabilidad económica de cualquier parque eólico depende directamente de la disponibilidad de WPD. Un estudio de viabilidad de larga duración requiere un conocimiento dinámico de la evolución de la WPD en una ubicación determinada, no es suficiente con la información que proporciona una imagen estática. Los Planes de Acción Nacionales de Energía Renovable de España y Portugal (*Gobierno de España*, 2020; *República Portuguesa*, 2018) incorporan estimaciones para la energía eólica instalada en el mar

que se espera que alcance los 59.53 GW en la PI en 2030 (Tabla 1.3). Además, en los últimos años se ha reducido a 4.5 €/MWh el precio de mercado de la electricidad por cada bloque de 12.8 TWh instalado debido precisamente a la penetración de la energía eólica (*Pereira and Saraiva*, 2013). También se ha demostrado que la implementación de decenas de GW en una región puede reducir las emisiones de CO_2 en casi un 70 % (*Kempton et al.*, 2007).

Finalmente, cabe señalar que el aumento de CF observado en función de la disponibilidad de recursos en las últimas décadas podría traducirse en incrementos aún mayores en la producción de energía debido a que se esperan avances técnicos en el diseño aerodinámico e hidrodinámico de los aerogeneradores flotantes. A medida que los avances técnicos en el campo de los sistemas de anclaje en alta mar sean posibles, en un futuro próximo será posible también construir más instalaciones de energía eólica marina. Esto incluye la generación de hidrógeno o metanol a miles de kilómetros de la costa que abre la posibilidad de producirlos en parques eólicos ubicados en lugares lejanos con una WPD óptima/ máxima (*Babarit et al.*, 2018).

Referencias

Akaike, H. (1987), Factor analysis and AIC, *Psychometrik*, *52*(3), 317–332.

Alkhalidi, M. A., S. K. Al-Dabbous, S. Neelamani, and H. A. Aldashti (2019), Wind energy potential at coastal and offshore locations in the state of Kuwait, *Renewable Energy*, *135*(2019), 529–539, doi:10.1016/j.renene.2018.12.039.

Angus, P., S. Rasmussen, and K. Seiter (1988), Short-term prediction of thunderstorm probability and intensity by screening observational and derived predictors, *Reprints AMS 15th Conference on Severe Local Storms, Baltimore, MA*, pp. 368–371.

Babarit, A., J. C. Gilloteaux, G. Clodic, M. Duchet, A. Simoneau, and M. F. Platzer (2018), Techno-economic feasibility of fleets of far offshore hydrogen-producing wind energy converters, *International Journal of Hydrogen Energy*, *43*(15), 7266–7289, doi:10.1016/j.ijhydene.2018.02.144.

Barker, D., X.-Y. Huang, Z. Liu, T. Auligné, X. Zhang, S. Rugg, R. Ajjaji, A. Bourgeois, J. Bray, Y. Chen, M. Demirtas, Y.-R. Guo, T. Henderson, W. Huang, H.-C. Lin, J. Michalakes, S. Rizvi, and X. Zhang (2012), The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA, *Bulletin of the American Meteorological Society*, *93*(6), 831–843, doi:10.1175/BAMS-D-11-00167.1.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao (2004), A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results, *Monthly Weather Review*, *132*(4), 897–914, doi:10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.

Barthelmie, R. J., and L. E. Jensen (2010), Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm, *Wind Energy*, *13*(June), 573–586, doi:DOI:10.1002/we.408.

Bett, P. E., H. E. Thornton, and R. T. Clark (2017), Using the Twentieth Century Reanalysis to assess climate variability for the European wind industry, *Theoretical and Applied Climatology*, *127*(1-2), 61–80, doi:10.1007/s00704-015-1591-y.

Boccard, N. (2009), Capacity factor of wind power realized values vs. estimates, *Energy Policy*, *37*(7), 2679–2688, doi:10.1016/j.enpol.2009.02.046.

Breaker, L. C., W. H. Gemmill, and D. S. Crosby (1994), The Application of a Technique for Vector Correlation to Problems in Meteorology and Oceanography, *Journal of Applied Meteorology*, *33*(11), 1354–1365, doi:10.1175/1520-0450(1994)033<1354:TAOATF>2.0.CO;2.

Breiman, L. (2001), Random Forests, *Machine learning*, *45*, 5–32, doi:10.1023/A:1010933404324.

Campos, R. M., and C. Guedes Soares (2018), Spatial distribution of offshore wind statistics on the coast of Portugal using Regional Frequency Analysis, *Renewable Energy*, *123*, 806–816, doi:10.1016/j.renene.2018.02.051.

Carreno-Madinabeitia, S., G. Ibarra-Berastegi, J. Sáenz, E. Zorita, and A. Ulazia (2020), Sensitivity studies for a hybrid numerical-statistical short-term wind and gust forecast at three locations in the basque country (Spain), *Atmosphere*, *11*(1), doi:10.3390/ATMOS11010045.

Carreno-Madinabeitia, S., G. Ibarra-Berastegi, J. Sáenz, and A. Ulazia (2021), Long-term changes in offshore wind power density and wind turbine capacity factor in the iberian peninsula (1900–2010), *Energy*, *226*, 120,364, doi:https://doi.org/10.1016/j.energy.2021.120364.

Cassola, F., and M. Burlando (2012), Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, *Applied Energy*, *99*, 154–166, doi:10.1016/j.apenergy.2012.03.054.

Cheng, H. P., and M. T. Yu (2013), Effect of the transmission configuration of wind farms on their capacity factors, *Energy Conversion and Management*, *66*, 326–335, doi:10.1016/j.enconman.2012.12.011.

Cheng, X., G. Nitsche, and J. M. Wallace (01 Jun. 1995), Robustness of low-frequency circulation patterns derived from eof and rotated eof analyses, *Journal of Climate*, *8*(6), 1709 – 1713, doi:10.1175/1520-0442(1995)008<1709:ROLFCP>2.0.CO;2.

Christensen, O. B., M. Drews, J. H. Christensen, K. Dethloff, K. Ketelsen, I. Hebestadt, and A. Rinke (2007), The HIRHAM Regional Climate Model Version 5 (beta), *Technical Report 06-17;*, *5*, 1–22.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. BroNnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, O. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley (2011), The Twentieth Century Reanalysis Project, *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28, doi:10.1002/qj.776.

Costoya, X., A. Rocha, and D. Carvalho (2020), Using bias-correction to improve future projections of offshore wind energy resource: A case study on the Iberian Peninsula, *Applied Energy*, *262*(January), 114,562, doi:10.1016/j.apenergy.2020.114562.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes,

A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. Mcnally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, and F. Vitart (2011), 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.

Doswell, C. A., and E. N. Rasmussen (1994), The Effect of Neglecting the Virtual Temperature Correction on CAPE Calculations, *Weather and Forecasting*, *9*(4), 625–629, doi:10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2.

Doswell, C. A., C. Ramis, R. Romero, and S. Alonso (1998), A Diagnostic Study of Three Heavy Precipitation Episodes in the Western Mediterranean Region, *Weather and Forecasting*, *13*(1), 102–124, doi:10.1175/1520-0434(1998)013<0102:ADSOTH>2.0.CO;2.

Durre, I., R. S. Vose, and D. B. Wuertz (2006), Overview of the Integrated Global Radiosonde Archive, *Journal of Climate*, *19*(1), 53–68, doi:10.1175/JCLI3594.1.

Eccel, E., L. Ghielmi, P. Granitto, R. Barbiero, F. Grazzini, and D. Cesari (2007), Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models, *Nonlinear Processes in Geophysics*, *14*, 211–222, doi:10.5194/npg-14-211-2007.

Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, *7*(1), 1–26, doi:10.1214/aos/1176344552.

Elsner, P. (2019), Continental-scale assessment of the African offshore wind energy potential: Spatial analysis of an under-appreciated renewable energy resource, *Renewable and Sustainable Energy Reviews*, *104*(July 2018), 394–407, doi:10.1016/j.rser.2019.01.034.

Environmental Modeling Center (2003), The GFS Atmospheric Model, Note 442., *Tech. Rep. November*.

European Commission (2019), The European Green Deal.

European Commission (2020), The North Seas Energy Cooperation.

Euskalmet - Agencia Vasca de Meteorología (2019), Euskalmet - Agencia Vasca de Meteorología.

Fernández-Ferrero, A., J. Sáenz, G. Ibarra-Berastegi, and J. Fernández (2009), Evaluation of statistical downscaling in short range precipitation forecasting, *Atmospheric Research*, *94*(3), 448–461, doi:10.1016/j.atmosres.2009.07.007.

Fernández-Ferrero, A., J. Sáenz, and G. Ibarra-Berastegi (2010), Comparison of the performance of different analog-based bayesian probabilistic precipitation forecasts over Bilbao, Spain, *Monthly Weather Review*, *138*(8), 3107–3119, doi:10.1175/2010MWR3284.1.

Fink, A. H., S. Pohle, J. G. Pinto, and P. Knippertz (2012), Diagnosing the influence of diabatic processes on the explosive deepening of extratropical cyclones, *Geophys. Res. Lett.*, *39*, L07,803, doi:10.1029/2012GL051025.

Galway, J. G. (1956), The Lifted Index as a Predictor of Latent Instability, *Bulletin of the American Meteorological Society*, *37*(10), 528–529, doi:10.1175/1520-0477-37.10.528.

Gascón, E., A. Merino, J. Sánchez, S. Fernández-González, E. García-Ortega, L. López, and L. Hermida (2015), Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe, *Atmospheric Research*, *164-165*, 194–209, doi:10.1016/j.atmosres.2015.05.012.

George, J. J. (1960), *Weather Forecasting for Aeronautics*, 411 pp., Academic Press, San Diego.

Gobierno de España (2020), Borrador Actualizado del Plan Nacional Integrado de Energía y Clima 2021-2030, pp. 1–8.

González-Rojí, S. J., S. Carreno-Madinabeitia, J. Sáenz, and Ibarra-Berastegi (2021), Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation, *Hydrology and Earth System Sciences*.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen (2012), Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - A comparison of methods, *Hydrology and Earth System Sciences*, *16*(9), 3383–3390, doi:10.5194/hess-16-3383-2012.

Hannachi, A., I. T. Jolliffe, Stephenson, and D. B. (2007), Review Empirical orthogonal functions and related techniques in atmospheric science: A review, *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, *27*(May), 1119–1152, doi:10.1002/joc.1499.

Hastie, T., R. Tibsharani, and J. Friedman (2001), *The Elements of Statistical Learning*, vol. 27, 83–85 pp., doi:10.1007/b94608.

Hersbach, H. (2016), The ERA5 Atmospheric Reanalysis, *AGU fall meeting abstracts*.

Holley, D. M., S. R. Dorling, C. J. Steele, and N. Earl (2014), A climatology of convective available potential energy in Great Britain, *International Journal of Climatology*, *34*(14), 3811–3824, doi:10.1002/joc.3976.

Ibarra-Berastegi, G., A. Ulazia, J. Saénz, and S. J. González-Rojí (2019), Evaluation of lebanon's offshore-wind-energy potential, *Journal of Marine Science and Engineering*, *7*(10), 1–13, doi:10.3390/jmse7100361.

International Renewable Energy Agency (2009), IRENA.

Islek, F., Y. Yuksel, and C. Sahin (2020), Spatiotemporal long-term trends of extreme wind characteristics over the Black Sea, *Dynamics of Atmospheres and Oceans*, *90*(September 2019), 101,132, doi:10.1016/j.dynatmoce.2020.101132.

Izquierdo, S., C. Montañs, C. Dopazo, and N. Fueyo (2010), Analysis of CSP plants for the definition of energy policies: The influence on electricity cost of solar multiples, capacity factors and energy storage, *Energy Policy*, *38*(10), 6215–6221, doi:10.1016/j.enpol.2010.06.009.

JBA Consulting (2020), Performance indicators for offshore wind farms in Europe from 1977 to 2100 derived from climate projections, doi:10.24381/cds.42c6e4dd.

Johns, R. H., and C. A. Doswell (1992), Severe Local Storms Forecasting, *Weather and Forecasting*, *7*(4), 588–612, doi:10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2.

Jonkman, J., S. Butterfield, W. Musial, and G. Scott (2009), Definition of a 5-MW Reference Wind Turbine for Offshore System Development, *Tech. rep.*

Kaldellis, J. K., and D. Zafirakis (2011), The wind energy (r)evolution: A short review of a long history, *Renewable Energy*, *36*(7), 1887–1901, doi:10.1016/j.renene.2011.01.002.

Kempton, W., C. L. Archer, A. Dhanju, R. W. Garvine, and M. Z. Jacobson (2007), Large CO2 reductions via offshore wind power matched to inherent storage in energy end-uses, *Geophysical Research Letters*, *34*(2), doi:10.1029/2006GL028016.

Kim, D. I., H. H. Kwon, and D. Han (2019), Bias correction of daily precipitation over South Korea from the long-term reanalysis using a composite gamma-pareto distribution approach, *Hydrology Research*, *50*(4), 1138–1161, doi:10.2166/nh.2019.127.

Komusanac, I., D. Fraile, G. Brindley, C. Walsh, and I. Pineda (2019), Wind energy in Europe in 2018, *Trends and statistics*, p. 32.

Kulkarni, M. a., S. Patil, G. V. Rama, and P. N. Sen (2008), Wind speed prediction using statistical regression and neutral network, *Journal of Earth System Science*, *117*(4), 457–463, doi:10.1007/s12040-008-0045-7.

Liberato, M. L. R., J. G. Pinto, I. F. Trigo, and R. M. Trigo (2011), Klaus - an exceptional winter storm over northern Iberia and southern France, *Weather*, *66*, 330–334, doi:10.1002/wea.755.

Liberato, M. L. R., J. G. Pinto, R. M. Trigo, P. Ludwig, P. Ordóñez, D. Yuen, and I. F. Trigo (2013), Explosive development of winter storm Xynthia over the subtropical North Atlantic ocean, *Nat. Hazards Earth Sys. Sci.*, *13*, 2239–2251, doi:10.5194/nhess-13-2239-2013.

López, L., J. L. Marcos, J. L. Sánchez, A. Castro, and R. Fraile (2001), CAPE values and hailstorms on northwestern Spain, *Atmospheric Research*, *56*(1-4), 147–160, doi:10.1016/S0169-8095(00)00095-8.

Ludwig, P., J. G. Pinto, M. Reyers, and S. L. Gray (2014), The role of anomalous SST and surface fluxes over the southeastern North Atlantic in the explosive

development of windstorm Xynthia, *Quart. J. Roy. Meteorol. Soc.*, *140*, 1729–1741, doi:10.1002/qj.2253.

Manwell, J. F., J. G. McGowan, and A. L. Rogers (2010), *Wind energy explained: theory, design and application*, Wiley.

Masters, G. M. (2013), *Renewable and efficient electric power systems*, John Wiley & Sons.

Matulla, C., X. Zang, X. L. Wang, J. Wang, E. Zorita, S. Wagner, and H. von Storch (2008), Influence of similarity measures on the performance of the analog method for downscaling daily precipitation, *Climate Dynamics*, *30*, 133–144, doi:10.1007/s00382-007-0277-2.

McNulty, R. P. (1995), Severe and Convective Weather: A Central Region Forecasting Challenge, *Weather and Forecasting*, *10*(2), 187–202, doi:10.1175/1520-0434(1995)010<0187:SACWAC>2.0.CO;2.

Meucci, A., I. R. Young, O. J. Aarnes, and Ø. Breivik (2020), Comparison of wind speed and wave height trends from twentieth-century models and satellite altimeters, *Journal of Climate*, *33*(2), 611–624, doi:10.1175/JCLI-D-19-0540.1.

Miller, R. C. (1975), Notes on Analysis and Severe-Storm Forecasting Procedures of the Air Force Global Weather Central, *AWS Technical Report*, *200*.

Moncrieff, M. W. (1981), A theory of organized steady convection and its transport properties, *Quarterly Journal of the Royal Meteorological Society*, *107*(451), 29–50, doi:10.1002/qj.49710745103.

Müller, M. D. (2011), Effects of model resolution and statistical postprocessing on shelter temperature and wind forecasts, *Journal of Applied Meteorology and Climatology*, *50*(8), 1627–1636, doi:10.1175/2011JAMC2615.1.

Murdock, H. E., D. Gibb, T. André, F. Appavou, A. Brown, and B. Epp (2019), Renewables 2019 Global Status Report, *Tech. Rep. 3*, doi:10.3390/resources8030139.

NOAA (2020a), NCEP North American Regional Reanalysis: NARR.

NOAA (2020b), Blended Sea Winds.

Okumus, I., and A. Dinler (2016), Current status of wind energy forecasting and a hybrid method for hourly predictions, *Energy Conversion and Management*, *123*, 362–371, doi:10.1016/j.enconman.2016.06.053.

Papież, M., S. Śmiech, and K. Frodyma (2019), Factors affecting the efficiency of wind power in the European Union countries, *Energy Policy*, *132*(May), 965–977, doi:10.1016/j.enpol.2019.06.036.

Penalba, M., A. Ulazia, G. Ibarra-Berastegui, J. Ringwood, and J. Sáenz (2018), Wave energy resource variation off the west coast of Ireland and its impact on realistic wave energy converters' power absorption, *Applied Energy*, *224*(November 2017), 205–219, doi:10.1016/j.apenergy.2018.04.121.

Pereira, A. J., and J. T. Saraiva (2013), Long term impact of wind power generation in the Iberian day-ahead electricity market price, *Energy*, *55*, 1159–1171, doi: 10.1016/j.energy.2013.04.018.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland (1992), A comprehensive meteorological modeling system-RAMS, *Meteorology and Atmospheric Physics*, *49*(1-4), 69–91, doi:10.1007/BF01025401.

Puertos del Estado (2019), Miniterio de transporte movilidad y agencia urbana.

Quinlan, J. R. (1986), Induction of decision trees, *Machine learning*, *1*, 81–106.

R Core Team (2020), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

República Portuguesa (2018), Plano Nacional Integrado Energia e Clima 2021-2030, pp. 1–106.

Riemann-Campe, K., K. Fraedrich, and F. Lunkeit (2009), Global climatology of Convective Available Potential Energy (CAPE) and Convective Inhibition (CIN) in ERA-40 reanalysis, *Atmospheric Research*, *93*(1-3), 534–545, doi:10.1016/j.atmosres.2008.09.037.

Romero, R., M. Gayà, and C. A. Doswell (2007), European climatology of severe convective storm environmental parameters: A test for significant tornado events, *Atmospheric Research*, *83*(2-4), 389–404, doi:10.1016/j.atmosres.2005.06.011.

Rozas-Larraondo, P., I. Inza, and J. A. Lozano (2014), A Method for Wind Speed Forecasting in Airports Based on Nonparametric Regression, *Weather and Forecasting*, *29*(6), 1332–1342, doi:10.1175/WAF-D-14-00006.1.

RStudio Team (2020), RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rueda-Bayona, J. G., A. Guzmán, J. J. C. Eras, R. Silva-Casarín, E. Bastidas-Arteaga, and J. Horrillo-Caraballo (2019), Renewables energies in Colombia and the opportunity for the offshore wind technology, *Journal of Cleaner Production*, *220*, 529–543, doi:10.1016/j.jclepro.2019.02.174.

Rusu, E., and L. Rusu (2019), Evaluation of the wind power potential in the European nearshore of the Mediterranean Sea, *E3S Web of Conferences*, *103*(2019), 1–6, doi:10.1051/e3sconf/201910301003.

Sáenz, J., S. J. González-Rojí, S. Carreno-Madinabeitia, and G. Ibarra-Berastegi (2019), Analysis of atmospheric thermodynamics using the R package aiRthermo, *Computers and Geosciences*, *122*, 113–119, doi:10.1016/j.cageo.2018.10.007.

Sáenz, J., S. Carreno-Madinabeitia, G. Esnaola, S. J. González-Rojí, G. Ibarra-Berastegi, and A. Ulazia (2020), The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models, *Geoscientific Model Development*, *13*(7), 3221–3240, doi:10.5194/gmd-13-3221-2020.

Salam, M. A., M. G. Yazdani, Q. M. Rahman, D. Nurul, S. F. Mei, and S. Hasan (2019), Investigation of wind energy potentials in Brunei Darussalam, *Frontiers in Energy*, *13*(4), 731–741, doi:10.1007/s11708-018-0528-4.

Salvação, N., and C. Guedes Soares (2018), Wind resource assessment offshore the Atlantic Iberian coast with the WRF model, *Energy*, *145*, 276–287, doi:10.1016/j.energy.2017.12.101.

Santamaría-Bonfil, G., A. Reyes-Ballesteros, and C. Gershenson (2016), Wind speed forecasting for wind farms: A method based on support vector regression, *Renewable Energy*, *85*, 790–809, doi:10.1016/j.renene.2015.07.004.

Sen, P. K. (1968), Estimates of the regression coefficient based on Kendall's tau, *Journal of the American Statistical Association*, *63*, 1379–1389, doi:10.1080/01621459.1968.10480934.

Showalter, A. K. (1953), A Stability Index for Thunderstorm Forecasting, *Bulletin of the American Meteorological Society*, *34*, 250–252.

Siedlecki, M. (2009), Selected instability indices in Europe, *Theoretical and Applied Climatology*, *96*(1-2), 85–94, doi:10.1007/s00704-008-0034-4.

Skamarock, W., J. Klemp, J. Dudhi, D. Gill, D. Barker, M. Duda, X.-Y. Huang, W. Wang, and J. Powers (2008), A Description of the Advanced Research WRF Version 3, *NCAR Technical Note NCAR/TN-475+STR*, (June), 113, doi:doi:10.5065/D68S4MVH.

Soman, S. S., H. Zareipour, O. Malik, and P. Mandal (2010), A review of wind power and wind speed forecasting methods with different time horizons, *North American Power Symposium 2010, NAPS 2010*, (October), doi:10.1109/NAPS.2010.5619586.

Soukissian, T., F. Karathanasi, and P. Axaopoulos (2017), Satellite-Based Offshore Wind Resource Assessment in the Mediterranean Sea, *IEEE Journal of Oceanic Engineering*, *42*(1), 73–86, doi:10.1109/JOE.2016.2565018.

Spanish Department of Economy Industry and Competitiveness (2016), No Title.

Stanski, H. R., L. J. Wilson, and W. R. Burrows (1989), Survey of Common Verfication Methods in Meteorology, *Atmospheric Research*, pp. 9–42, doi:10.1017/CBO9781107415324.004.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.

Theil, H. (1950), *A Rank-Invariant Method of Linear and Polynomial Regression Analysis*, 386–392 pp., In: Proceedings of Koninalijke Nederlandse Akademie van Weinenschatpen, doi:10.1007/978-94-011-2546-820.

Torres, J. L., A. García, M. De Blas, and A. De Francisco (2005), Forecast of hourly average wind speed with ARMA models in Navarre (Spain), *Solar Energy*, *79*(1), 65–77, doi:10.1016/j.solener.2004.09.013.

Ulazia, A., J. Sáenz, G. Ibarra-Berastegi, S. J. González-Rojí, and S. Carreno-Madinabeitia (2017), Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean, *Applied Energy*, *208*(September), 1232–1245, doi:10.1016/j.apenergy.2017.09.030.

Ulazia, A., M. Penalba, A. Rabanal, G. Ibarra-Berastegi, J. Ringwood, and J. Sáenz (2018), Historical evolution of the wave resource and energy production off the chilean coast over the 20th Century, *Energies*, *11*(9), doi:10.3390/en11092289.

Ulazia, A., A. Nafarrate, G. Ibarra-Berastegi, J. Sáenz, and S. Carreno-Madinabeitia (2019a), The consequences of air density variations over northeastern Scotland for offshore wind energy potential, *Energies*, *12*(13), doi:10.3390/en12132635.

Ulazia, A., J. Sáenz, G. Ibarra-Berastegi, S. J. González-Rojí, and S. Carreno-Madinabeitia (2019b), Global estimations of wind energy potential considering seasonal air density changes, *Energy*, *187*(August), 115,938, doi:10.1016/j.energy.2019.115938.

Ulazia, A., G. Ibarra-Berastegi, J. Sáenz, S. Carreno-Madinabeitia, and S. J. González-Rojí (2019c), Seasonal correction of offshore wind energy potential due to air density: Case of the Iberian Peninsula, *Sustainability (Switzerland)*, *11*(13), 2014–2016, doi:10.3390/su11133648.

University of Wyoming (2020), Sounding data.

Valero, F., A. Pascual, and M. L. Martín (2014), An approach for the forecasting of wind strength tailored to routine observational daily wind gust data, *Atmospheric Research*, *137*, 58–65, doi:10.1016/j.atmosres.2013.09.019.

van Delden, A. (2001), The synoptic setting of thunderstorms in western Europe, *Atmospheric Research*, *56*(1-4), 89–110, doi:10.1016/S0169-8095(00)00092-2.

Viceto, C., M. Marta-Almeida, and A. Rocha (2017), Future climate change of stability indices for the Iberian Peninsula, *International Journal of Climatology*, *37*(12), 4390–4408, doi:10.1002/joc.5094.

Wang, Z., Y. Li, K. Wang, and Z. Huang (2017), Environment-adjusted operational performance evaluation of solar photovoltaic power plants: A three stage efficiency analysis, *Renewable and Sustainable Energy Reviews*, *76*(December 2016), 1153–1162, doi:10.1016/j.rser.2017.03.119.

Weare, B. C., and J. S. Nasstrom (1982), Examples of Extended Empirical Orthogonal Function Analyses, doi:10.1175/1520-0493(1982)110<0481:EOEEOF>2.0.CO;2.

Weisberg, S. (2005), Applied Linear Regression, *Wiley, New Jersey (2005)*.

WWRP/WGNE Joint Working Group on Forecast Verification Research (2015), Forecast Verification methods.

Young, I. R., and A. Ribal (2019), Multiplatform evaluation of global trends in wind speed and wave height, *Science*, *364*(6440), 548–552, doi:10.1126/science.aav9527.

Zhao, W., Y. M. Wei, and Z. Su (2016), One day ahead wind speed forecasting: A resampling-based approach, *Applied Energy*, *178*, 886–901, doi:10.1016/j.apenergy.2016.06.098.

Zheng, C. W., X. yan Li, X. Luo, X. Chen, Y. hao Qian, Z. hua Zhang, Z. sheng Gao, Z. bo Du, Y. bo Gao, and Y. ge Chen (2019), Projection of Future Global Offshore Wind Energy Resources using CMIP Data, *Atmosphere - Ocean*, *57*(2), 134–148, doi:10.1080/07055900.2019.1624497.

Zorita, E., and H. von Storch (1999), The analog method as a simple statistical downscaling technique:comparison with more complicated methods, *Journal of Climate*, *12*, 2474–2489, doi:10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.

2 CONCLUSIONES

Las conclusiones generales de desarrollar el primer objetivo (*Carreno-Madinabeitia et al.*, 2020), es decir, desarrollar un modelo híbrido de predicción de viento a corto-medio plazo han sido las siguientes.

- Se ha desarrollado un método completo de predicción de viento a corto plazo con potencial para aplicar en cualquier lugar.
- Se ha reforzado la idea de que, en los modelos de predicción con un horizonte de predicción menor a un día, es necesario comparar sus resultados con los obtenidos a través de la persistencia.
- Los tres tipos de predictores propuestos en esta tesis (última observación, predicción del modelo global y extEOF) son adecuados y necesarios para realizar predicciones de este tipo.
- El avance fundamental es que una forma de mejorar los resultados del modelo de predicción de corto-medio plazo es introduciendo la información de un modelo global como predictores.

Dentro del primer objetivo se ha continuado estudiando los extremos de viento, en concreto se ha analizado si el modelo de WRF reproduce de forma adecuada los índices de inestabilidad CAPE, CIN y TT (González-Rojí et al. 2021) (en revisión):

- La simulación WRF con datos asimilados (D) reproduce mejor los índices de inestabilidad que la simulación sin asimilación (N).

- La mayoría de los estudios que se centran en este campo no utilizan la simulación del modelo a escala local con asimilación de datos 3DVAR. La introducción de la asimilación de datos mejora los resultados si la comparamos con la misma ejecución pero sin la mencionada asimilación.

- El modelo WRF reproduce adecuadamente estos índices, sobre todo para el caso del índice TT.

- Respecto a los resultados estacionales, el índice TT muestra errores muy pequeños en todas las estaciones.

- En invierno las masas de aire inestable se encuentran sobre todo en la costa Atlántica de la PI.

- Además, en este caso se abre la puerta a realizar la predicción de los extremos de viento a partir de la información de los índices de inestabilidad.

Al desarrollar el diagrama de Sailor (*Sáenz et al.*, 2020) las conclusiones obtenidas son las siguientes:

- Se ha desarrollado un nuevo diagrama para comparar diferentes modelos de variables bidimensionales frente a los datos de referencia.

- Se ha relacionado algebraicamente el índice RMSE bidimensional con el sesgo y la varianza. Además, todas las matrices involucradas en el cálculo del RMSE bidimensional se pueden expresar en función del primer y segundo EOF de los datos de referencia. Es decir, sin tener en cuenta los EOFs de los datos del modelo, solamente con la matriz de rotación que los relaciona.

- Este diagrama se ha generalizado para poder utilizarlo con datos obtenidos a través de un sistema de predicción por conjuntos o datos espaciales.

- Se ha generado el software *SailoR* para distribuirlo de forma libre y que su uso sea fácil por otros investigadores.

En el último objetivo, en el cual se han analizado los cambios a largo plazo (1900-2010) de las variables WPD y CF en el mar en el siglo XX (Carreno-Madinabeitia et al. 2021):

- Se ha observado que la calibración *quantile matching* es adecuada para realizar este tipo de estudios.

- La calibración *quantile matching*, con el objetivo de reducir errores finales debido a la relación de no linealidad entre la ρ y U, se ha realizado directamente a las variables WPD y CF.

- Se han identificado las zonas del Golfo de León y el Cabo de Gata, así como la zona Atlántica del dominio de estudio como zonas de un alto potencial para generar energía eólica.

- También se ha comprobado que durante el siglo XX tanto WPD como CF han aumentado de forma significativa.
- Es el primer estudio que analiza las tendencias de las variables WPD y CF en un periodo tan largo del pasado.

3 ANEXO: Trabajos publicados o aceptados

Predicción de viento

Carreno-Madinabeitia S, Ibarra-Berastegi G, Sáenz J, Zorita E, Ulazia A. *Sensitivity studies for a hybrid numerical-statistical short-term wind and gust forecast at three locations in the basque country (Spain)*. Atmosphere (Basel) 2020;11. https://doi.org/10.3390/ATMOS11010045.

Atmosphere 2019	
Total Cites	3218
Journal Impact Factor (IF)	2.397
5 Year IF	2.437
Immediacy Index	0.535
IF witout Journal Self Cites	1.923
Meteorology & Atmospheric Sciences	
Rank	48/93
Quantile	Q_3
Journal IF Percentile	48.925

Tabla 3.1: Índices de calidad de la revista *Atmosphere* del año 2019

atmosphere

MDPI

Article

Sensitivity Studies for a Hybrid Numerical–Statistical Short-Term Wind and Gust Forecast at Three Locations in the Basque Country (Spain)

Sheila Carreno-Madinabeitia [1,2,*], Gabriel Ibarra-Berastegi [3,4], Jon Sáenz [2,4], Eduardo Zorita [5] and Alain Ulazia [6]

1 TECNALIA, Parque Tecnológico de Álava, Albert Einstein 28, E-01510 Vitoria-Gasteiz (Araba/Álava), Spain
2 Applied Physics II Department, Faculty of Science and Technology, University of the Basque Country, E-48940 Leioa, Spain; jon.saenz@ehu.eus
3 NE and Fluid Mechanics Department, Faculty of Engineering, University of the Basque Country, E-48013 Bilbao, Spain; gabriel.ibarra@ehu.eus
4 Joint Research Unit, BEGIK, Spanish Institute of Oceanography-University of the Basque Country, Plentzia Itsas Estazioa (PIE), E-48620 Plentzia, Spain
5 Institute of Coastal Research, Helmholtz-Zentrum-Geesthacht, 21502 Geesthacht, Germany; Eduardo.Zorita@hzg.de
6 NE and Fluid Mechanics Department, Faculty of Engineering, University of the Basque Country, E-20600 Eibar, Spain; alain.ulazia@ehu.eus
* Correspondence: Sheila.Carreno@Tecnalia.com

Received: 28 November 2019; Accepted: 27 December 2019; Published: 29 December 2019

check for updates

Abstract: This study evaluates the performance of statistical models applied to the output of numerical models for short-term (1–24 h) hourly wind forecasts at three locations in the Basque Country. The target variables are horizontal wind components and the maximum wind gust at 3 h intervals. Statistical approaches such as persistence, analogues, linear regression, and random forest (RF) are used. The verification statistics used are coefficient of determination (R^2) and root mean square error (RMSE). Statistical models use three inputs: (1) Local wind observations; (2) extended EOFs (empirical orthogonal functions) derived from past local observations and ERA-Interim variables in a previous 24-h period covering a domain around the area of study; and (3) wind forecasts provided by ERA-Interim. Results indicate that, for horizons less than 1–4 h, persistence is the best model. For longer predictions, RF provides the best forecasts. For horizontal components at 4–24 h horizons, RF slightly outperformed ERA-Interim wind forecasts. For gust, RF performs better than ERA-Interim for all the horizons. Persistence is the most influential factor for 2–5 h. Beyond this horizon, predictors from the ERA-Interim wind forecasts led the contribution. Hybrid numerical–statistical methods can be used to improve short-term wind forecasts.

Keywords: short-term forecast; wind; statistical forecast; random forest; ERA-Interim; persistence

1. Introduction

Providing accurate short-term wind prediction is an important goal for any meteorological service [1]. Wind storms in Spain produce important economic losses, representing 15% of all natural hazards damage compensations reimbursed by insurance companies from 1987 to 2015 [2]. A robust wind forecasting system must issue reliable warnings hours before potentially dangerous events occur, such as the windstorm Klaus [3]. Better short-term wind forecasts contribute to reducing economic damage and help save lives [4]. Reliable wind forecasts find many applications, including air pollution [5], wind power [6–13], forest fires [14], sea navigation [15,16], and airports [17,18].

Wind predictions are issued based on meteorological numerical weather forecast (NWF) systems, in which data assimilation (DA) plays an important role in order to properly correct errors in the forecasting model by the use of observations in the data assimilation stages. In these NWF systems, the data assimilation methods optimally estimate the state of the atmosphere [19,20] and a numerical weather forecast model is used to predict the future states of the atmosphere. Two examples of these advanced NWP systems are the ERA-Interim [21] global atmospheric reanalysis model, from 1979 until 2019 [21], and NCEP re-analyses [22]. They are not real-time operational NWF systems, since the DA and the forecasting models are frozen in time, but they are re-run for the whole period using all the available observations commonly used in numerical weather forecasts in order to avoid, as much as possible, errors in the record due to changes in the structure (architecture, resolution, kind of observations processed) of the archives derived from operational NWF systems. In the case of the ERA-Interim re-analyses, besides analyses, forecasts are also provided starting at the analysis times. These forecasts are the ones used in this paper as a surrogate of an operational model. However, re-analyses are quite coarse-resolution models when compared with operational models of the same generation. In the particular case of the ERA-Interim, the spatial resolution is 0.75°, too coarse for local wind forecasts.

To fill this resolution gap, numerical and statistical downscaling techniques are customarily used. Numerical downscaling increases the resolution by nesting a higher-resolution model, such as WRF [23], RAMS [24], or Hirlam [25], into the global NWF. Statistical downscaling, as an alternative downscaling method, uses transfer functions from predictors provided by low-resolution models to local predictands, as in Valero et al. [26], who used ERA-Interim data followed by an analogue downscaling method. Interesting results were obtained by Pascual et al. [27] using ECMWF's ERA 40. Wind forecasts for horizons between a few hours to some days often involve the application of statistical post-processing to NWF model outputs. Global forecast system (GFS) [28], North American mesoscale forecast system (NAM) [29], and rapid update cycle (RUC) [30] NWF outputs were used, followed by statistical analogues post-processing in Nagarajan et al. [31]. In particular, in that paper, the objective was to forecast wind speeds at 6 h intervals up to 72 h horizons. Another example is Cassola and Burlando [8], where the BOLAM numerical model [32] was used, along with Kalman filter-based post-processing for forecasting wind speed for a horizon of up to 36 h at 6 h steps. A combined approach (GFS plus statistical downscaling) can also be found in other studies, such as Zhao et al. [33], where hourly averaged wind speed was predicted for 24 h horizons. A meteorological model like WRF [23], followed by analogues and Bayesian inference, was also used in complex terrain by Manor and Berkovic [34]. In this context, the Greek National meteorological service model (GNMS) applies locally recurrent neural networks [9] for hourly averaged wind speed with a 72 h horizon every hour. A study involving the same geographical area as this study [18] combined the forecasts by a global GFS model with a post-processing stage based on machine-learning techniques. The second stage, which used random forest (RF) amongst other methods, improved the quality of the raw forecast provided by the global NWF model. These results will be extended to different areas (ocean, coast, interior) in this contribution. Gust predictions are also of interest. Valero et al. [26] proposed a daily forecast derived from ERA-Interim by means of the analogue technique. Patlakas et al. [35] used the SKIRON/Dust [36] model, followed by Kalman Filter-based post-processing. Finally, Zjavka [37] predicted hourly gusts with the ALADIN model [38], using the polynomial neural networks technique. Statistical downscaling models in this context are calibrated with past observations and are later used to forecast future meteorological events. This is usually highly site-dependent, because the model structure and parameters are specifically developed for the location [8,34,39–42].

For very short horizons (a few hours in advance) time-series based statistical models are commonly used. For example, hourly averaged wind forecasts have been issued by statistical models such as support vector regression (SVR) and artificial neural networks (ANN) [10,43,44]. Also, ARIMA models [41] were used to predict average wind speed. The authors of [45,46] developed a method that makes use of regime-switching for forecast wind speed hourly 12 h ahead. This way, they deal

with non-stationarity in wind dynamics. A comparison of machine learning models for forecasting wind during the next hour was presented by Feng et al. [11]. Along these lines, ARIMA-ANN and ARIMA-Kalman Filter models have also been used [47] for 3 h forecasting. Hourly observations were used as target variables with ANN and genetic expression programming (GEP) [48]. A method similar to analogues [49] was also applied to forecast wind speed in the next hour [50]. Even [51] monthly mean wind speed has been forecasted combining time-series with ANN.

In summary, long-term (beyond a few days) forecasts are commonly issued by means of a global NWF followed by statistical post-processing [7]. While for time horizons of a few hours, forecast time-series based statistical models are usually applied, we propose to use both approaches in a hybrid wind forecast system.

Thus, we follow the well-stablished technique of testing the hybrid numerical–statistical methods by using data from a frozen assimilation system. This allows us to perform a sensitivity study of the relative merits of different statistical techniques under controlled conditions without introducing spurious modifications into the results due to changes in the configuration of the operational NWF system [8,18,26,27,31,33,39,52].

A comparison of more than 14 studies based on statistical methods for wind forecasting can be found in Okumus and Dinler [12] and references therein. As mentioned in that review, these kinds of wind-forecasting studies are sometimes difficult to compare. The root mean square error (RMSE) and the coefficient of determination (R^2) are widely used in this field of research [8,18,26,27,31,33,42]. The main advantage of using R^2 over the plain correlation is that the coefficient of determination is an indicator that represents the fraction of the overall variability explained by the model. A complete evaluation requires additional information on the error, either expressed in absolute terms (RMSE) or in percentage terms.

For this study, hybrid models were constructed using local observations and ERA-Interim data for short-term forecasting of wind speed and wind gust up to 24 h in the future at three selected locations in the Basque Country during 2007–2014 (Figure 1). The statistical algorithms used were analogues, linear regression, and random forest (RF). They are here applied to locations over the ocean, the coast, and the interior (see Figure 1).

Taking into account the results in aforementioned studies, our objectives were: First, a comparison of the performance of different hybrid numerical–statistical techniques in the forecasting of u and v components for three observatories located in the ocean, at the coast, and in the interior of the Basque Country in the Bay of Biscay. Secondly, a comparison of the performance of the best statistical model with two reference forecasts, the nearest grid cell direct model output from the coarse-scale NWF model, which provided the predictors and persistence. The rationale behind the use of these reference models (ERA-Interim and persistence) is that any forecasting effort for a certain number of hours in the future is justified only if the new model outperforms existing ones. The comparison of performances is presented both for hourly wind data and wind gusts. The robustness of the method is evaluated against the selection of the domain and by applying it to three different sites which present different wind regimes. Finally, the most significant predictors are also identified.

Section 2 presents the datasets and method used in this study. Section 3 describes the results, Section 4 discusses the conclusions of this study, and Section 5 presents a discussion along with a future outlook.

Figure 1. Locations of interest, Bilbao Bizkaia buoy (blue circle), Punta Galea station (orange circle), Alegria station (red circle). Distances in km to the nearest ERA-Interim node (magenta star) and from the coast to the Bizkaia buoy.

2. Material and Methods

2.1. Data

The Basque country is located in the Bay of Biscay, an area with elevations that may reach 1000 m just 30 km away from the coast. The topographic gradient is moderately steep, and the wind regime is subject to important coastal influences. Thus, three locations are chosen for this study (Figure 1). They represent ocean, coastal, and interior wind regimes, as can be seen in their wind roses (Figure 2). One of them is a buoy (Bilbao Bizkaia), representing an oceanic regime. The other two sites were selected to compare the results of the sea with the coast and an inland zone in the area of interest.

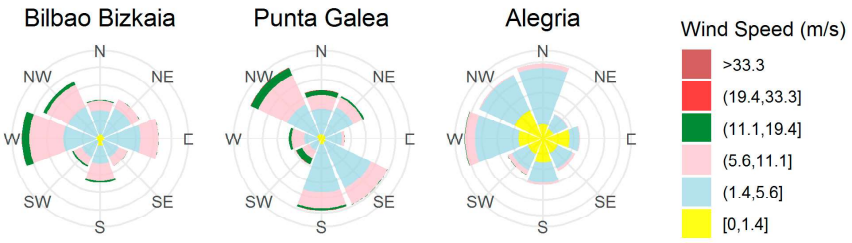

Figure 2. Wind roses corresponding to the period (2007–2014).

The Bilbao Bizkaia buoy (43.64° N, 3.05° W) is run by the Spanish Puertos del Estado agency [53] and is located offshore, approximately 30 km from the coast. As can be seen from the wind roses, it is not affected by sea breezes and, since there are no natural obstacles around it, the wind tends to flow freely. The second location is Punta Galea, at the coastline, at the top of a cliff (43.38° N and 3.04° W,

61 m elevation above sea level). Punta Galea is exposed to sea–land interactions affecting the wind regime, such as sea breezes. The third wind sensor is located in Alegria, on a plateau in the south of the Basque Country (42.84° N and 2.52° W, 545 m elevation), 60 km inland. The dominant wind pattern is west–east but the highest wind corresponds to a south–west direction. Both inland stations are run by the Basque Meteorological Service (Euskalmet).

At each location, u, v, and wind gust at the sensor height above ground level (Bilbao Bizkaia: 3 m, Punta Galea: 12 m, and Alegria: 10 m) are measured. u and v were calculated using only the last measurements (10 min) prior to the hourly time step at the three locations. These are the operational averages regularly implemented and subsequently made public by the institutions running the sensors. The zonal (u) and meridional (v) projections of the wind vector were obtained from magnitude and direction recorded at the three locations.

A second group of target variables was used at the three locations: Hourly maximum wind gusts. The wind gusts were computed from the maximum value of the gust recorded in the last 10 min at Punta Galea and Alegria. The maximum value for the Bilbao Bizkaia buoy was calculated using the measurements taken over the last 10 min of each hourly step because the buoy records data every hour while the meteorological stations do this every 10 min. Only the magnitude of wind gust was used. To compare the data with the ERA-Interim forecast, the wind gust since previous post-processing (3 h period) was calculated; this will hereinafter be referred to simply as wind gust. All the data span the period 2007–2014.

Data from ECMWF's ERA-Interim reanalysis [21,54] were used. The variables selected as likely to impact upcoming wind speed were mean sea level pressure (*Msl*); the zonal component of wind at 10 m (m/s) (u_{10}); the meridional component of wind at 10 m (v_{10}); and temperature at 2 m (T_2). The domain selected (Figure 3) is a rectangle spanning (49.50° N, 12° W) to (36° N, 7.5° E) with a spatial resolution of 0.75° × 0.75°. Analyses at 6 h time steps for the 2007–2014 period were downloaded from ECMWF. Additionally, u_{10} and v_{10} and wind gusts in forecast mode from ERA-Interim for the nearest grid points from the three chosen locations were also used. From the initial conditions at t = 0000 UTC and t = 1200 UTC, ERA-Interim forecasts of these variables are available for $t + k$, where k = (3, 6, 9, 12, 15, 18, 21, 24) h in the future in the meteorological archival and retrieval system (MARS) server of the ECMWF. All these variables were used to feed the different wind forecasting models as if the statistical models were using forecasts from an operational model.

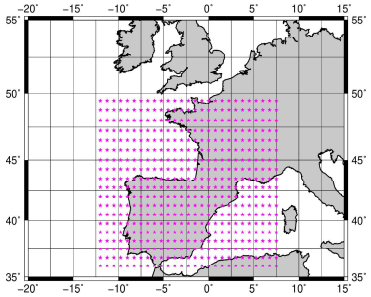

Figure 3. ERA-Interim domain used for downloading analysis data.

2.2. Method

In order to achieve a rigorous evaluation of the performance of statistical models, the original database was split into two datasets, one to fit and train the models (2007–2010) and the other to test the models on independent data (2011–2014). RF uses train and test datasets [18,55]; this method does not require validation datasets, such as the neural network method. The number of cases is shown in Table 1.

Table 1. Number of cases in each location and period.

	Data	Train	Test
Bilbao Bizkaia Buoy	3792	1896	1896
Punta Galea Station	4461	2231	2230
Alegria Station	4084	2042	2042

The raw input data were pre-processed as follows. At a given t time, each model at the different lead times was fed with:

1. Observations of u and v, and the wind gust at each sensor height when $t = 0$;
2. ERA-Interim forecasts (u, v, wind gust) for the nearest grid points from the selected locations (Figure 1). This NWF is based on a four-dimensional variational assimilation analysis, with a time window of 12 h, and produces forecasts with time steps that range from 3 to 24 h in the future, as explained in Section 2.1;
3. Time-lagged wind observation at $t - 0$, $t - 6$, $t - 12$, and $t - 18$ h. Time-lagged observations and ERA-Interim analyses (see domain in Figure 3) were used. In order to reduce the dimensionality of the time-lagged ERA-Interim variables (Msl, u_{10}, v_{10}, and T_2), extended empirical orthogonal functions (extEOFs) of the original variables were calculated [56,57]. In this way, both spatial and temporal patterns can be captured in the space corresponding to the principal components, and the leading extEOFs hold the highest fractions of the total variance. The resulting extended principal components have not been rotated, since they were just used for compressing the information and they are, therefore, orthogonal. Extended EOFs have been applied, dealing with different geophysical variables, such as waves [52,58], ENSO events [59], and surface moisture flux and precipitation [55]. In the case of this study, important associations (correlations) were detected between variables, in addition to non-negligible autocorrelations and spatial correlations throughout the selected spatial and time domain. Due to the different physical quantities (K, m/s and so on) of the variables involved, at an initial stage they were standardized ($mean$ = 0, $variance$ = 1), and the final number of leading extEOFs used for this study was 26, selected under the condition of jointly retaining at least 90% of the overall variance, following the final criteria developed by authors for similar geophysical studies after careful evaluation of different alternatives [52,55].

A summary of the raw input data is shown below (Table 2).

Table 2. Summary description of the statistical model raw inputs Input.

Name	Variables	Source	Time
Last obs	u, v, wind gust	Observations	t h
ERA-I Forecast	u_{10}, v_{10}, wind gust	ERA-I forecast	$t + 3$, $t + 6$, … $t + 24$ h
ExtEOF	Msl, u_{10}, v_{10}, T_2 (domain) u, v, wind gust	ERA-I analisis Observations	t, $t - 6$, $t - 12$, $t - 18$ h

Three statistical techniques were used: Analogues, linear regression, and RF. Linear regression is a common method in statistical prediction. Second, the analogues method is a non-linear technique that has been often used since the early days of meteorological forecasting [49]. The third technique was random forests (RF), a machine learning technique that we wished to compare with the two other, simpler methods.

For each target variable and statistical technique, a specific model was built for each of the forecasting horizons at $t + k$ ($k = 1, 2, \ldots, 24$) h. This means that, for u and v, a total of 432 models were constructed (two target variables times three locations times three techniques times 24 hourly steps). In the case of the wind gust, only the best technique identified for u and v was used RF. For wind gust,

72 models (one target variable times three locations times one technique times 24 hourly steps) were built. As a result, for this study, a total number of 504 models were developed and evaluated.

In the analogue technique [49], for each case in the test period, the most similar situation is sought in the training period and this becomes the wind prediction at $t + h$ hours. Patterns can be selected using various similarity metrics [60], and, for this study, both Euclidean and cosine norms were applied. Given the state of the predictors to be used by the vector $x = (x_1, \ldots, x_p)$ with p the dimension of the state-space, the Euclidean distance between vectors x and y is defined by:

$$d(x, y) = \left\| x - y \right\| = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2} \quad (1)$$

whilst the cosine distance is defined by:

$$d(x, y) = \left\| x - y \right\| = \frac{\sum_{i=1}^{p} x_i y_i}{\sqrt{\sum_{i=1}^{p} {x_i}^2} \sqrt{\sum_{i=1}^{p} {y_i}^2}} \quad (2)$$

Random forest is a predictive algorithm that uses bootstrap technique to combine different decision trees, where each tree is built with observation data and random variables. RF [61] was considered due to its ability to mathematically capture highly non-linear relationships, like those known to be involved in the physics of wind forecasting. RF is a development of classification and regression trees (CART [62]), where the regression trees are randomly perturbed and a forest of perturbed trees is created. Trees are perturbed at two levels:

1. A number of bootstrap samples from original data are drawn and used to feed the different regression trees;
2. At each tree, each node is split using the best among a subset of m predictors randomly chosen at that node.

A key advantage of RF over other regression techniques is that it is free from overfitting, due to the use of the strong law of large numbers [61]. In RF, it is possible to estimate how important a given input is by calculating its percentage increase in the mean square error of the regression, if that particular predictor, is removed. This ranks the most influential inputs [63]. For this work, this importance ranking was estimated from the training dataset and, following the general procedure, the RF outputs were estimated as the average of the outputs of the trees [64,65].

In the case of the classical multiple linear regression, the candidate input variables are incorporated into the equation following the Akaike information criterion (AIC) [66]. Performance comparison for other atmospheric fields indicates that non-linear algorithms like neural networks outperform, but not overwhelmingly so, more simple methods like linear regression [67] and analogues [49,68]. RF is perhaps more difficult to implement than analogues or linear regression, but it has been successfully employed in other studies [55,69,70]. In particular, for the variable studied here—wind—[18] RF was used, as well as other machine learning techniques, to forecast wind speed. This latter study is the most similar to ours, because both forecast wind for same locations and used NWF model predictors for the statistical models. Details on most mathematical aspects of RF can be found in the literature [61,71]. For comparison purposes, in addition to the abovementioned models, the plain forecasts provided by ERA-Interim for time $t + k$ (k = 3 h, 6 h, 9 h, ... , 24 h), as well as persistence were also considered as additional models for each location and forecasting horizon.

Forecast accuracy was evaluated against observations exclusively using data belonging to the test period. The number of cases ranges from 1896 to 2231 (Table 1). The performance assessment was carried out by means of the usual indicators typically employed in the literature: R^2 and RMSE [18,26,42,72]. In this study, a given model will be considered superior to another if both indicators, at a 95% confidence level, are better. All the comparisons using these indicators were assessed at a 95% confidence level

using boundaries computed by bootstrap resampling [73]. In the process of selecting bootstrap samples, the time order of the series was taken into account.

3. Results

3.1. Model Performance for u and v

First, in this section, the results of three statistical models (linear regression, analogues, and RF) have been shown and compared for u and v targets. Then, the best result has been compared against persistence and ERA-Interim forecasts. The sensitivity to the domain size is also verified and finally the relevant inputs are identified.

3.1.1. Statistical Models

The forecasts by the three statistical models (linear regression, analogues, and RF) were compared in order to clarify which technique performed better. The goal was to identify the best general technique.

The results of this analysis are summarized in Tables 3 and 4 and Figure 4. These tables and figure show a performance comparison for the statistical models built using the three techniques mentioned above. It can be seen that models based on RF, analogues, and linear regression tend to perform similarly for all the forecasting horizons (1 h, 2 h, . . . , 24 h) and only marginal—though statistically significant—improvements can be detected, depending on the location and forecasting horizons. In the case of one location (Bilbao Bizkaia), linear regression-based models performed slightly better for more forecasting horizons, while RF-based models exhibited a marginal improvement in the case of Punta Galea and Alegria. So, in general terms, RF-based models appear to perform best at the three locations.

Table 3. R^2 verification index of u and v wind components for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and three statistical methods (random forest, analogs, and linear regression). The best verification values are typed using bold font.

			1 h	6 h	12 h	18 h	24 h
u	BB	RF	0.89–0.91	0.69–0.74	0.65–0.7	0.64–0.7	0.6–0.65
		AN	0.85–0.88	0.64–0.7	0.59–0.65	0.58–0.65	0.55–0.61
		LR	**0.9–0.92**	**0.7–0.75**	**0.67–0.72**	**0.65–0.71**	**0.62–0.67**
	PG	RF	**0.78–0.82**	**0.57–0.63**	**0.51–0.58**	**0.52–0.59**	**0.49–0.55**
		AN	0.75–0.79	0.53–0.59	0.48–0.55	0.49–0.56	0.47–0.54
		LR	**0.78–0.82**	0.54–0.61	0.46–0.53	0.47–0.54	0.45–0.51
	Al	RF	**0.65–0.71**	**0.42–0.5**	**0.42–0.51**	**0.37–0.45**	**0.39–0.48**
		AN	0.6–0.67	0.37–0.45	0.37–0.46	0.33–0.42	0.34–0.43
		LR	0.64–0.71	0.38–0.46	0.39–0.47	0.33–0.4	0.37–0.46
v	BB	RF	0.78–0.83	0.57–0.65	**0.52–0.6**	**0.51–0.6**	**0.48–0.57**
		AN	0.78–0.83	0.58–0.65	0.5–0.58	0.51–0.59	0.45–0.53
		LR	**0.82–0.86**	**0.59–0.66**	0.49–0.57	0.51–0.59	0.46–0.54
	PG	RF	0.82–0.86	**0.71–0.76**	**0.66–0.71**	**0.67–0.72**	**0.63–0.68**
		AN	0.82–0.86	0.66–0.71	0.61–0.66	0.62–0.67	0.57–0.63
		LR	**0.83–0.87**	0.67–0.72	0.62–0.67	0.64–0.68	0.59–0.65
	Al	RF	**0.75–0.79**	**0.62–0.68**	**0.64–0.69**	**0.61–0.66**	**0.61–0.67**
		AN	0.74–0.78	0.57–0.63	0.61–0.66	0.56–0.62	0.56–0.62
		LR	**0.75–0.79**	0.58–0.64	0.59–0.64	0.56–0.62	0.58–0.63

Table 4. RMSE verification index of u and v wind components for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and three statistical methods (random forest, analogs, and linear regression). Lowest values of RMSE are written using bold font.

			1 h	6 h	12 h	18 h	24 h
u	BB	RF	1.41–1.55	2.33–2.55	2.46–2.66	2.49–2.73	2.65–2.84
		AN	1.58–1.75	2.53–2.78	2.72–2.94	2.73–2.96	2.83–3.04
		LR	**1.24–1.42**	**2.27–2.49**	**2.4–2.58**	**2.43–2.65**	**2.57–2.77**
	PG	RF	**1.92–2.09**	**2.74–2.94**	**2.81–3**	**2.9–3.12**	**2.94–3.13**
		AN	2–2.18	2.79–2.98	2.85–3.05	2.91–3.1	2.94–3.14
		LR	1.89–2.08	2.81–3	2.94–3.13	3.04–3.23	3.03–3.22
	Al	RF	**1.15–1.27**	**1.43–1.58**	**1.54–1.69**	**1.5–1.65**	**1.59–1.74**
		AN	1.26–1.39	1.5–1.66	1.63–1.76	1.53–1.69	1.67–1.82
		LR	1.17–1.29	1.48–1.64	1.61–1.73	1.54–1.7	1.63–1.77
v	BB	RF	1.53–1.74	2.24–2.5	**2.36–2.61**	**2.4–2.69**	**2.45–2.69**
		AN	1.56–1.73	2.23–2.47	2.43–2.67	2.43–2.7	2.55–2.79
		LR	**1.38–1.58**	**2.21–2.44**	2.45–2.69	2.45–2.71	2.52–2.75
	PG	RF	1.92–2.18	**2.47–2.68**	**2.7–2.92**	**2.66–2.86**	**2.83–3.06**
		AN	1.91–2.14	2.67–2.9	2.9–3.14	2.84–3.07	3.06–3.3
		LR	**1.84–2.07**	2.65–2.85	2.86–3.07	2.8–3.01	2.95–3.17
	Al	RF	1.07–1.16	**1.2–1.3**	**1.21–1.3**	**1.21–1.31**	**1.23–1.34**
		AN	1.08–1.17	1.27–1.39	1.27–1.37	1.28–1.39	1.31–1.43
		LR	**1.07–1.15**	1.27–1.37	1.29–1.38	1.28–1.38	1.3–1.39

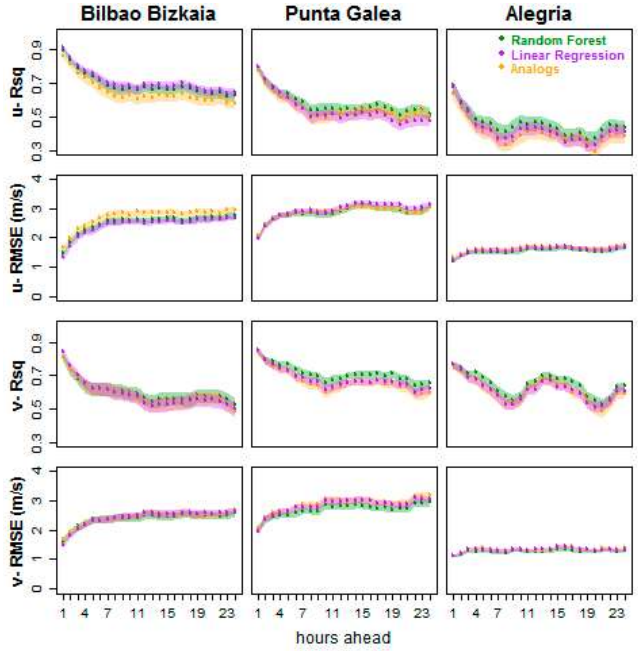

Figure 4. R^2 and root mean square error (RMSE) verification indices of u and v wind components for the coming 24 h, for three locations, and three statistical methods.

3.1.2. Model Evaluation

The next step was to compare the results of the statistical models based on RF (the best technique identified in the previous step) against persistence and ERA-Interim forecasts for the nearest grid point. The results can be seen in Figure 5 and Tables 5 and 6 for the Bilbao Bizkaia buoy, Punta Galea, and Alegria locations. The errors of the physics-based ERA-Interim forecasts remain almost constant up to the 24 h lag forecast [74].

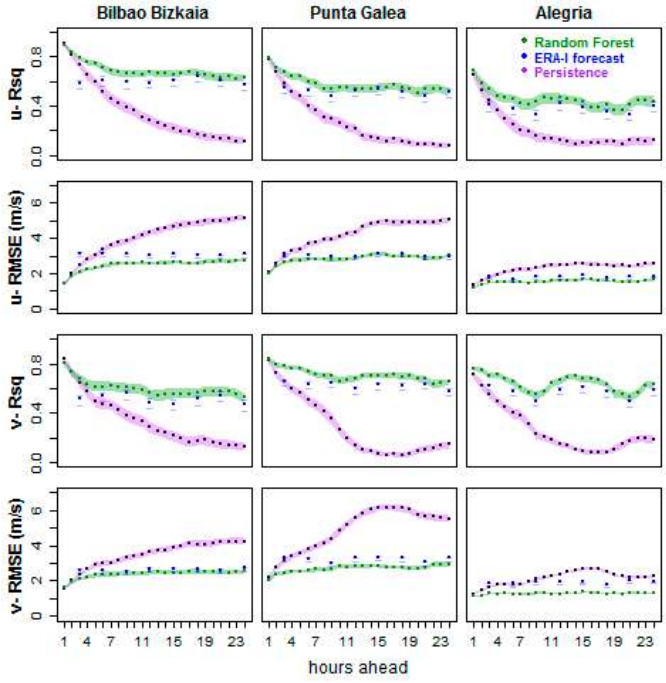

Figure 5. R^2 and RMSE verification indices of u and v wind components for the coming 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and comparing three methods in the test period.

If u and v are jointly analyzed, the results indicate that for horizons shorter than 2–4 h, neither the RF-based models nor ERA-Interim forecasts outperform persistence. Beyond this horizon, up to 24 h in the future, for the Bilbao Bizkaia location, the RF-based models exhibit higher R^2 values and lower RMSE than persistence or ERA-Interim models for u forecasts. In the case of v, both RF and ERA-Interim clearly outperform persistence. Considering that wind prediction involves a joint forecast of u and v, the results for Bilbao Bizkaia indicate that, below a forecasting horizon of 4 h, persistence is the best option, and, from 4 h to 24 h in the future, RF-based models outperform the rest, but that both are not always significantly better than ERA-Interim forecasts.

A similar pattern can be observed for the other two locations, Punta Galea and Alegria, although persistence outperformed the other techniques for a shorter forecast horizon (the first 2–4 h). For longer horizons, the ERA-Interim and RF-based models performed better, although, for v, RF models exhibited somewhat better results than ERA-Interim forecasts. For u forecasts, RF and ERA-Interim performed similarly, and a joint evaluation of wind forecasts once again indicates a slightly better overall performance of RF-based models. In all cases, it must be noted that the RMSE were lower for

the different techniques and forecasting horizons for Alegria, but this was due to the lower wind speed values recorded at this location.

A seasonally stratified analysis (not shown) confirmed the abovementioned ranking, although in summer, since wind is generally weaker, the error values described by the RMSE are also smaller.

Table 5. R^2 index (95% confidence interval) of u and v wind components for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and RF, ERA-Interim forecasts, and persistence models. Highest values of R^2 are highlighted by means of bold font.

			1 h	6 h	12 h	18 h	24 h
u	**BB**	**RF**	0.89–0.91	**0.69–0.74**	**0.65–0.7**	**0.64–0.7**	**0.6–0.65**
		ERA-I F		0.57–0.65	0.56–0.62	0.61–0.67	0.54–0.6
		Pers	**0.89–0.92**	0.48–0.56	0.25–0.33	0.14–0.21	0.09–0.14
	PG	**RF**	**0.78–0.82**	**0.57–0.63**	**0.51–0.58**	**0.52–0.59**	**0.49–0.55**
		ERA-I F		0.49–0.56	0.49–0.55	0.48–0.55	0.48–0.55
		Pers	0.75–0.8	0.35–0.44	0.18–0.25	0.09–0.15	0.06–0.11
	Al	**RF**	**0.65–0.71**	**0.42–0.5**	**0.42–0.51**	**0.37–0.45**	**0.39–0.48**
		ERA-I F		0.34–0.42	0.38–0.46	0.31–0.39	0.36–0.44
		Pers	0.62–0.69	0.2–0.29	0.09–0.16	0.08–0.15	0.08–0.16
v	**BB**	**RF**	0.78–0.83	**0.57–0.65**	**0.52–0.6**	**0.51–0.6**	**0.48–0.57**
		ERA-I F		0.5–0.58	0.45–0.53	0.48–0.56	0.43–0.51
		Pers	**0.81–0.86**	0.43–0.52	0.25–0.33	0.14–0.22	0.09–0.16
	PG	**RF**	**0.82–0.86**	**0.71–0.76**	**0.66–0.71**	**0.67–0.72**	**0.63–0.68**
		ERA-I F		0.61–0.66	0.57–0.63	0.6–0.65	0.55–0.61
		Pers	0.8–0.85	0.46–0.54	0.11–0.17	0.04–0.08	0.12–0.18
	Al	**RF**	**0.75–0.79**	**0.62–0.68**	**0.64–0.69**	**0.61–0.66**	**0.61–0.67**
		ERA-I F		0.55–0.61	0.56–0.62	0.55–0.6	0.56–0.61
		Pers	0.68–0.74	0.36–0.44	0.12–0.18	0.06–0.11	0.15–0.22

Table 6. RMSE index (95% confidence interval) of u and v wind components for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and RF, ERA-Interim forecasts, and persistence models. Lowest values of RMSE are highlighted by means of bold font.

			1 h	6 h	12 h	18 h	24 h
u	**BB**	**RF**	1.41–1.55	**2.33–2.55**	**2.46–2.66**	**2.49–2.73**	**2.65–2.84**
		ERA-I F		3.05–3.27	2.97–3.17	2.97–3.18	3–3.21
		Pers	**1.31–1.51**	3.21–3.56	4.15–4.5	4.7–5.07	5–5.37
	PG	**RF**	**1.92–2.09**	**2.74–2.94**	**2.81–3**	**2.9–3.12**	**2.94–3.13**
		ERA-I F		3–3.18	2.87–3.05	3.02–3.2	2.92–3.11
		Pers	1.95–2.2	3.51–3.84	4.21–4.52	4.74–5.09	4.9–5.25
	Al	**RF**	**1.15–1.27**	**1.43–1.58**	**1.54–1.69**	**1.5–1.65**	**1.59–1.74**
		ERA-I F		1.63–1.78	1.77–1.89	1.69–1.85	1.82–1.95
		Pers	1.28–1.44	2.06–2.29	2.42–2.64	2.37–2.58	2.42–2.65
v	**BB**	**RF**	1.53–1.74	**2.24–2.5**	**2.36–2.61**	**2.4–2.69**	**2.45–2.69**
		ERA-I F		2.46–2.69	2.58–2.83	2.55–2.79	2.62–2.86
		Pers	**1.44–1.66**	2.82–3.15	3.45–3.78	3.9–4.25	4.05–4.44
	PG	**RF**	**1.92–2.18**	**2.47–2.68**	**2.7–2.92**	**2.66–2.86**	**2.83–3.06**
		ERA-I F		3.15–3.37	3.17–3.4	3.17–3.38	3.19–3.42
		Pers	2–2.31	3.63–3.98	5.39–5.76	5.92–6.31	5.32–5.74
	Al	**RF**	**1.07–1.16**	**1.2–1.3**	**1.21–1.3**	**1.21–1.31**	**1.23–1.34**
		ERA-I F		1.69–1.81	1.86–1.99	1.72–1.84	1.91–2.05
		Pers	1.2–1.32	1.76–1.92	2.31–2.45	2.45–2.59	2.2–2.36

3.1.3. Sensitivity to the Domain Size

One of the input groups was the extEOF derived from the ERA-Interim values of different variables, plus local observations, as explained in Section 2.2. As mentioned above, the geographical boundaries of those variables used to calculate the extEOF were (49.50° N, 12° W) to (36° N, 7.5° E) (Figure 3). The reason that these boundaries were selected is that analysis of the variables inside this domain could allow us to capture the large-scale synoptic fields influencing the wind field at the three locations.

However, it was necessary to evaluate how the size of the selected domain could affect the results. To that end, a new set of extEOF was derived from the same variables, but defined for a smaller domain, from (45° N, 12° W) to (40° N, 0° E) (Figure 6). These new extEOFs were used to build the RF models again, and their performance was evaluated following the same method as before.

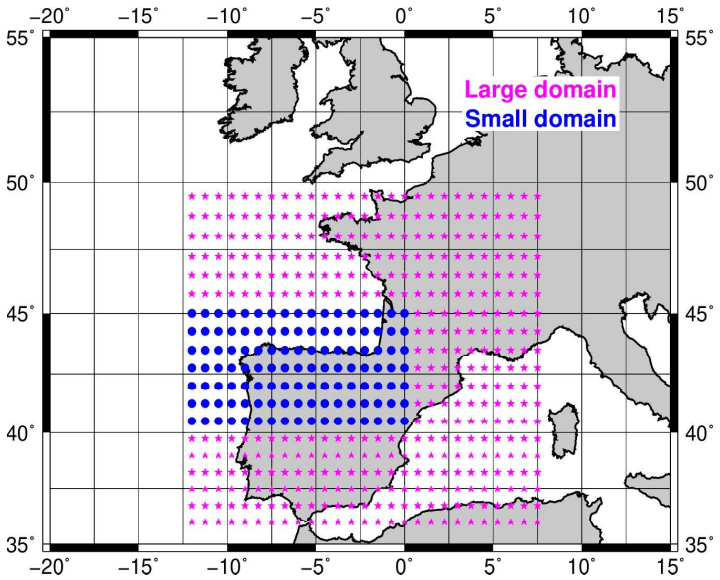

Figure 6. ERA-Interim domains used for downloading analyses.

The rationale behind this testing of the domain is that the observed wind field in the Basque Country is, as is common in extratropical areas, very likely close to geostrophy. The ability of a smaller domain to provide information at this level must be checked. The use of a smaller domain would make calculations more simple, quicker, and straightforward. For this reason, a much smaller domain has also been considered.

The results of the RF-based models obtained using the two sets of extEOFs derived from the two domains can be seen in Figure 7. This figure clearly shows that the statistical indicators of performance for R^2 and RMSE (95% confidence boundaries) are entirely overlapping, thus indicating a similar performance. For this reason, it can be concluded that there is negligible sensitivity of results to a domain size in the range analyzed in this paper.

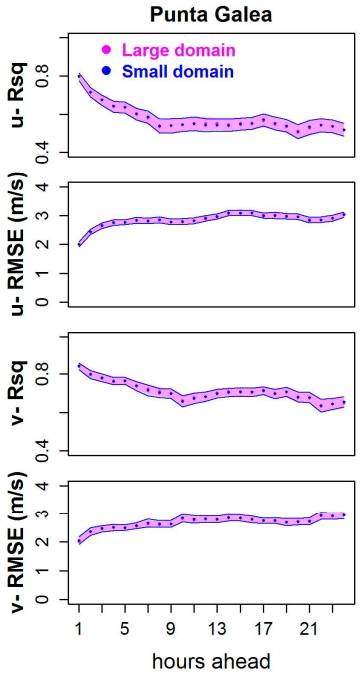

Figure 7. R^2 and RMSE verification indices for the u and v wind components for the coming 24 h at Punta Galea, using the RF method to compare the results in two different domains.

3.1.4. Identification of Relevant Inputs

The RF-based models were identified as those that, in general terms, performed best. Thus, these models were compared with the ERA-Interim and persistence forecasts. However, it should be noted that, for the three locations, the statistical models built with RF roughly identify the same three groups of inputs as being the most influential. These predictors are current local wind observation (u, v); the first, second, fourth, eighth, and ninth extEOFs; and the ERA-Interim forecasts for the nearest grid points. Figure 8 shows the most influential input for the different horizons for RF-based models.

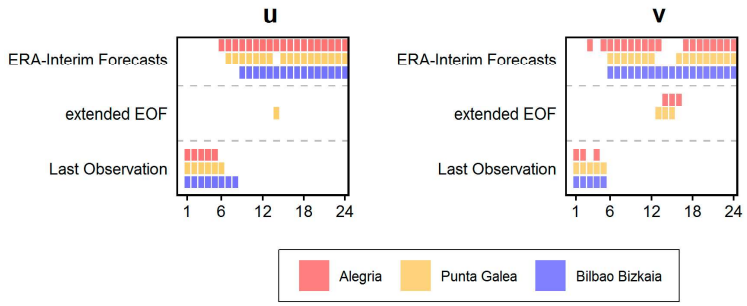

Figure 8. Most important input (last observation, extEOF, or ERA-I Forecast) of u and v wind components for the coming 24 h, at three locations using the RF method.

For predictions for shorter horizons than 2–5 h, the most influential variable tends to be the current observation. This indicates that for short horizons, the system's memory accounts for the highest fraction of overall predictability achieved by the models, which do not outperform persistence in the first 2–4 h. For medium-range predictions, persistence, ERA-Interim forecasts, and extEOFs tend to be either the most, or second most, influential inputs. Finally, for forecasting horizons longer than 7–9 h, the wind forecasts of ERA-Interim for the nearest grid points are the most influential predictors. All these conclusions are for the most important variable in each model, but if the second one (or even the third) is considered (results not shown), extEOFs become more influential.

3.2. Model Performance for Wind Gusts

Considering the results obtained under Section 3.1, the model evaluation against persistence and ERA-Interim forecasts and the identification of important inputs for the downscaling model presented in the previous subsections. The best-performing one has been used for wind gust variable.

3.2.1. Model Evaluation

A similar general approach to that described above was also applied to construct the forecasting models at the same three locations, with wind gust as the target variable. The forecasting horizons were also the same (from 1 to 24 h in the future), but, since in most cases RF seemed to perform somewhat better for u and v, this was the only technique analyzed in a more detailed way. Similar predictors were also used in the models (Table 2). For this case, the inputs were:

1. All the information up to $t = 0$, last 3 h of observed wind gusts, and the same set of extEOFs corresponding to the initial domain and variables;
2. Instead of using the u and v values observed at time $t = 0$, the selected input was the maximum wind gust since previous post-processing;
3. The ERA-Interim forecasts were those corresponding to the upcoming maximum wind gusts. These forecasts involved 3 h time steps from $t + 3$ to $t + 24$ h.

Therefore, for this target variable a total of 72 RF-based models were evaluated, the results of which are shown in Figure 9 and Tables 7 and 8. It can be seen that the best choice for modelling the first 1–2 h was persistence. From this time onwards, the results indicate that, for the three locations and all forecasting horizons, the RF-based models exhibited a significantly smaller error than either persistence or ERA-Interim forecasts. In general terms, wind gust results are better than u and v's previously obtained results.

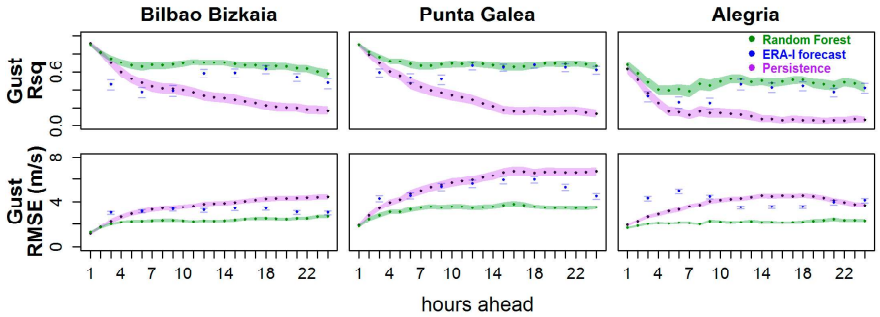

Figure 9. R^2 and RMSE verification indices for wind gust components over the coming 24 h, at three locations, and comparing three methods for the test period.

Table 7. R^2 index (95% confidence interval) of wind gust for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and RF, ERA-Interim forecasts, and persistence models. The results corresponding to the best model are highlighted using bold font.

			1 h	6 h	12 h	18 h	24 h
Gust	BB	RF	**0.9–0.93**	**0.67–0.72**	**0.64–0.69**	**0.61–0.67**	**0.58–0.64**
		ERA-I F		0.38–0.46	0.45–0.53	0.39–0.47	0.43–0.51
		Pers	**0.9–0.93**	0.49–0.57	0.31–0.4	0.19–0.26	0.16–0.22
	PG	RF	0.9–0.92	**0.68–0.73**	**0.65–0.71**	**0.62–0.68**	**0.63–0.69**
		ERA-I F		0.5–0.57	0.58–0.64	0.49–0.55	0.56–0.62
		Pers	**0.9–0.93**	0.45–0.53	0.24–0.31	0.13–0.2	0.12–0.18
	Al	RF	**0.73–0.78**	**0.51–0.58**	**0.54–0.61**	**0.47–0.53**	**0.5–0.57**
		ERA-I F		0.37–0.45	0.48–0.55	0.36–0.44	0.47–0.54
		Pers	0.69–0.75	0.28–0.36	0.02–0.06	0–0.03	0.12–0.19

Table 8. RMSE index (95% confidence interval) of wind gust for the steps 1, 6, 12, 18, and 24 h, for three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and RF, ERA-Interim forecasts, and persistence models. Lowest RMSE values are highlighted by using bold font.

			1 h	6 h	12 h	18 h	24 h
Gust	BB	RF	1.14–1.28	**2.19–2.35**	**2.34–2.53**	**2.4–2.57**	**2.55–2.74**
		ERA-I F		3.22–3.42	2.98–3.2	3.18–3.38	3.05–3.27
		Pers	**1.09–1.27**	2.88–3.15	3.57–3.86	4.02–4.34	4.26–4.59
	PG	RF	**1.67–1.92**	**3.11–3.42**	**3.31–3.63**	**3.42–3.72**	**3.44–3.79**
		ERA-I F		5.16–5.53	4.91–5.32	5.15–5.55	4.86–5.29
		Pers	1.67–1.96	4.49–4.94	5.72–6.2	6.36–6.89	6.48–7.03
	Al	RF	**1.59–1.77**	**1.98–2.17**	**2.07–2.27**	**2.07–2.3**	**2.17–2.36**
		ERA-I F		4.11–4.39	3.59–3.85	4.16–4.44	3.76–4.01
		Pers	1.75–1.94	2.86–3.17	4.08–4.35	4.12–4.39	3.46–3.79

3.2.2. Identification of Important Inputs

The most important input for maximum wind gust forecasts over the last 3 h period is shown in Figure 10 for different lead times. The previous pattern was confirmed; in the first 4 h, the selected input was the last observation, and in the last 10 h, it was the ERA-Interim forecast for the nearest grid points. The results for the u wind component and wind gust were very similar. The horizons for persistence and ERA-Interim forecast were almost the same. They only differed in the selection of extEOFs for medium-range predictions. In the same way as previously, all these conclusions are for the most influential inputs in each model, but, if the second one (or even the third) is considered (results not shown), extEOFs gain relevance (results not shown).

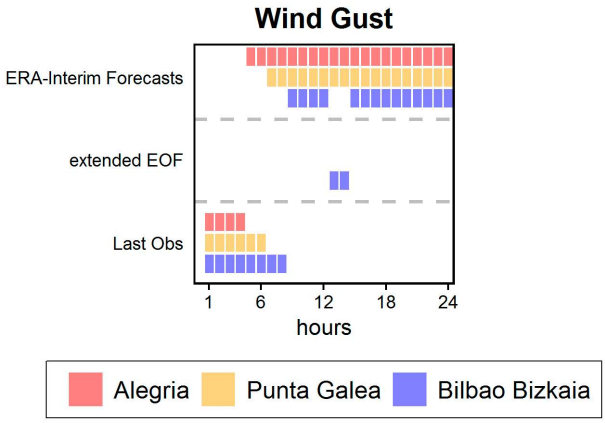

Figure 10. Most important input (last observation, extended empirical orthogonal functions (extEOFs), or ERA-Interim forecast) for the wind gust for the next 24 h, at three locations (Bilbao Bizkaia, Punta Galea, and Alegria), and using the RF method.

4. Discussion

The approach used in this paper follows the general pattern of using forecasts provided by various types of numerical models (such as the prognostic model used in the ERA-Interim reanalysis) followed by a statistical modelling stage based on different mathematical algorithms [33]. In our case, as shown above, the models include additional candidate inputs, like extEOFs plus local observations. The comparison between the performance of analogues, linear regression, and RF indicates that RF outperforms linear regression and analogues, but not in all cases, and never overwhelmingly.

Since the persistence of wind is not null, any modelling effort must also outperform persistence (Okumus and Dinler, 2016). Persistence, and ERA-Interim forecasts (playing the role of the surrogate of an operational NWF system) represent readily available forecasting models. The performance comparison between the best statistical technique (RF) and the plain use of persistence or ERA-Interim forecasts indicates that RF outperforms the others beyond the time horizon of 1–4 h. On the other hand, persistence is always a better option than the methods tested here for up to 1–4 h into the future. Preliminary studies with a low number of cases (10 days) indicate that a modified general regression neural network could beat persistence, even for short-term horizons (below 5 h). Other studies [9,41,48], using different methods, have found that RMSE values are better than for plain persistence, even for the first hour. However, in these studies, the number of cases is not very high, and the authors do not assess the statistical significance of the performance differences. All this indicates that the specific conditions in which persistence is outperformed for very short-term forecasting horizons are yet to be fully understood, and further research in this area may be needed. In our study, for all predictands and forecasts up to 2–5 h in the future, the most influential input is the current local observation. Thus, the system's memory plays a key role in building the prediction and explains why persistence is not beaten in these cases in the immediate short-term. Beyond this horizon, ERA-Interim forecasts are the most important predictor. The group of extEOFs (but not the same ones in all cases) also plays an important role, as the second/third most influential input, thus confirming the need to use all three general groups of inputs.

These results are in general agreement with earlier studies [7], even for the same area [18], where different statistical approaches applied to outputs from numerical models were used. Soman et al. [7] describe that, for horizons longer than 6 h, hybrid structures, such as the numerical model with post-processing, or the numerical model plus time series techniques, also exhibit better results.

Santamaría-Bonfil et al. [10] and Hu et al. [50] reached the same conclusion for wind speed forecasting. More specifically, their models are based on time series with SVM and analogue methods, respectively. The authors indicate that persistence is the best model up to the 5 h horizons, signifying that they reached broadly similar conclusions to those in this study.

RF is one of the algorithms considered by Rozas-Larraondo et al. [18], but they issued forecasts only up to a 3 h lead time. In general terms, these authors combined the outputs of an NWF—GFS in this case—with METAR data as additional inputs to build a statistical model. In our study, we have extended the method to another variable (wind gust), statistical prediction techniques (analogues and linear regression), longer forecast lead times, and different data compression techniques, with the use of predictors such as extEOFs. Our results for longer lead times support the suggestion that RF-based statistical downscaling models constitute an effective tool for providing short-time wind forecasts, particularly when an NWF predictor is included in lead times longer than 3 h. Since we found that these results hold for three different locations (ocean, coast, and interior), with different wind regimes and an assessment of errors to a 95% confidence level by means of bootstrapping, we think the results are quite robust.

It must be noted that the models shown in this paper were built for three locations, representative of different areas in the region: Sea, coast, and inland. The surroundings of the three locations and the wind regimes are all very different. It must, however, be highlighted that the results show a common pattern with regard to the best statistical technique (RF) to be used, for the forecasting horizons and also for the most influential input groups. Despite this general common pattern, the particular aspects of the RF models built for the individual sites, such as the specific inputs selected at each location for each forecasting horizon (Figures 8 and 10), are different. Moreover, similarly to other statistical forecasting models [8,34,39], especially for sites where complex topography is an important factor, the RF models developed here are not exchangeable, but rather site-dependent and valid only for their specific locations. Okumus and Dinler [12] also present the idea that results using similar approaches to that in this study depend on target location and data.

The distance from the locations and the nearest ERA-Interim grid points, around 14–25 km, could impact (Figure 1) the results. By applying these techniques to a higher-resolution operational model, better results from the model can be expected, because the model output would be available at a smaller distance from the wind station. As mentioned above, the wind forecasts provided by the model used in the reanalysis play an important role in building the predictions [31]. The impact of the distance between grids and points needs to be assessed with further studies, and with new reanalyses, like ERA5, since operational models are often run at much higher resolutions. In any case, if higher-resolution models were used, better inputs would also be introduced in the statistical model. Thus, it can be expected that the RF-based models' results presented in this paper would be relatively robust to the spatial resolution of the NWF model, which provides the inputs. However, inhomogeneities in the archives derived from operational models (not the ones used in reanalyses) make this somewhat problematic when evaluating a relatively long time period (eight years), such as the period in this study.

Since meteorological services already have access to the numerical outputs from global and regional numerical models, the availability of advanced statistical methods coupled to numerical outputs, such as the one presented in this study, is important, because it allows a local wind prediction system to be developed with a lead time of several hours (1–24 h). From the perspective of meteorological agency, forecasts are very important, but especially so in emergency cases: To more accurately forecast wind speed and direction for the next 24 h when strong winds are expected, and when wind gusts are expected to surpass the threshold levels corresponding to a warning that should be issued, because, on those occasions, the wind could result in dangerous situations for the population [1,4].

Regarding other applications, it should be highlighted that the wind energy sector already uses [13] deep techniques to improve the renewable energy forecasts. Having access to wind forecasts that are as accurate as possible could help better integrate wind energy into the electric grid [6,75]. In this field,

short ranges of between 10 min and 2 h are customarily used for controlling and regulating turbines. According to our results, further study is necessary at this point, because the proposed models perform more poorly than persistence for this horizon. On the other hand, electricity markets require longer prediction ranges, usually between 13 and 37 h in the future, with hourly temporal resolution. In this case, we propose that from 13 to 24 h RF forecasting should be used, while, from 24 to 37 h forecasts, an NWF model is probably the best tool (valid for ERA-Interim, at least). We hope that the method proposed in this paper can be used to optimize the position in the electricity auction market of the stakeholders in the wind energy sector, although future studies should evaluate this in detail, considering the drawbacks of higher resolution (operational models) versus longer homogeneous training length (reanalyses).

5. Conclusions

Although applied to three particular locations, the present study contains a complete method for short-term wind prediction that can easily be transferred and potentially applied elsewhere. This method involves a machine-learning technique like RF and, in addition to numerical forecasts (provided here by ERA-Interim as hindcasts) and local wind observations, it incorporates extended EOFs as additional inputs. The result is wind forecasts at hourly intervals up to 24 h in the future.

Although previously discussed in general terms, the forecasting errors of the RF-based models presented in this paper cannot be directly compared to results from previous studies. Nevertheless, it can be stated that, in absolute terms, the performance is, in general, similar to the one reported by other studies, and that errors are lower than the plain use of either persistence or direct model output from ERA-Interim. For this reason, in our study, RF-based models provide clear added value and represent a significant improvement over the use of the above-mentioned readily available forecasts for forecast horizons of 4–24 h, and sometimes 1–24 h forecasting horizons.

Below this forecasting range, the memory of the system (persistence) is the best predictor. The comparison between three different types of techniques (RF, linear regression, and analogues) evaluated in this study indicates that, in general, RF performs slightly better than the others, but not always and not for all forecasting horizons.

The results are common to the three analysed variables (u, v, and wind gust), forecasting horizons (24), and sites (three). Since the analysis covered many input cases (over 1800), many years, different locations (ocean, coastal, and inland places), and statistical techniques, we can consider this a robust outcome.

The next step in this research path, once the u and v wind components and wind gust have been analyzed, will be to apply the same method to extreme events. This requires a longer training database, due to a very low probability of this type of event, but a similar approach to that presented in this paper, based on RF and similar types of inputs, looks highly promising. ERA5 and UERRA reanalyses [54] have recently been made available with 0.3° and 1 h (ERA5) and 11 km × 11 km and 6 h (UERRA) spatial and temporal resolutions for a long time span: (1979–2018) for ERA5; and (1962–2018) for UERRA. As in the ERA-Interim model, forecasts are also available, so ERA5 or UERRA, along with RF, will be used in future studies. The prediction of extreme events and the appropriate management of this information represent important challenges for meteorological services and insurance companies. Furthermore, RF can also yield probabilistic predictions, something that is useful when different alert levels have to be launched to protect lives and properties. In this paper, different statistical models were compared for the short-term forecasting of wind and wind forecasts over the ocean, the coast, and an interior site in the Basque Country. The design of this study follows the reasoning presented in many references already discussed in the introduction to this paper [8,12,26,27,31,33]. This has also been the authors' experience in previous papers [52,76]. It is a common result in the literature that the use of a statistical model, which is fed with data provided by a numerical weather forecast model, yields better results than the use of a statistical model alone or a numerical model alone. There are interesting new developments in the field of statistical wind forecast [45,46]. Since the regime-switch

modelled in these new developments is already described by the numerical forecast that is being used to feed the postprocessing statistical model, the regime-switch is already being taken into account by the random forest model. A definition of a regime-switch technique, which blends numerical inputs to previous states of the observed wind, would be an interesting future new development.

Author Contributions: Conceptualization, S.C.-M., G.I.-B., J.S., and E.Z.; methodology, S.C.-M., G.I.-B., J.S., E.Z., and A.U.; software, S.C.-M. and G.I.-B.; validation, S.C.-M. and G.I.-B.; writing—original draft preparation, S.C.-M., G.I.-B., and J.S.; writing—review and editing, G.I.-B., J.S., E.Z., and A.U.; visualization, S.C.-M.; supervision, G.I.-B., J.S., E.Z., and A.U.; funding acquisition, G.I.-B. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Spanish Government, MINECO project CGL2016-76561-R (MINECO/EU ERDF), and the University of the Basque Country (project GIU17/02).

Acknowledgments: The ECMWF ERA-Interim data used in this study are freely available and were downloaded from the ECMWF-MARS Data Server at no cost. The authors would like to express their gratitude to Puertos del Estado for kindly providing the data for this study. More particularly, the authors want to thank the Emergencies and Meteorology Directorate (Basque Regional Government) for public provision of data and operational service financial support. Most of the calculations and plots were carried out within the framework of R [77] and GMT [78].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Freebairn, J.W.; Zillman, J.W. Economic benefits of meteorological services. *Meteorol. Appl.* **2002**, *9*, 33–44. [CrossRef]
2. Spanish Deparment of Economy Industry and Competitiveness. *The Spanish Extraordinary Risk Coverage System*; Spanish Deparment of Economy Industry and Competitiveness: Madrid, Spian, 2016; pp. 1987–2015.
3. Liberato, M.L.R.; Pinto, J.G.; Trigo, I.F.; Trigo, R.M. Klaus-An exceptional winter storm over northern Iberia and southern France. *Weather* **2011**, *66*, 330–334. [CrossRef]
4. Solari, G.; Repetto, M.P.; Burlando, M.; de Gaetano, P.; Pizzo, M.; Tizzi, M.; Parodi, M. The wind forecast for safety management of port areas. *J. Wind Eng. Ind. Aerodyn.* **2012**, *104*, 266–277. [CrossRef]
5. Ibarra-Berastegi, G.; Elias, A.; Barona, A.; Saenz, J.; Ezcurra, A.; Diaz de Argandoña, J. From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao. *Environ. Model. Softw.* **2008**, *23*, 622–637. [CrossRef]
6. Pinson, P. Wind Energy: Forecasting Challenges for Its Operational Management. *Stat. Sci.* **2013**, *28*, 564–585. [CrossRef]
7. Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different time horizons. In Proceedings of the North American Power Symposium 2010, Arlington, TX, USA, 26–28 September 2010. [CrossRef]
8. Cassola, F.; Burlando, M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Appl. Energy* **2012**, *99*, 154–166. [CrossRef]
9. Barbounis, T.G.; Theocharis, J.B.Ă. Locally recurrent neural networks for long-term wind speed and power prediction. *Neurocomputing* **2006**, *69*, 466–496. [CrossRef]
10. Santamaría-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2016**, *85*, 790–809. [CrossRef]
11. Feng, C.; Cui, M.; Hodge, B.M.; Zhang, J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* **2017**, *190*, 1245–1257. [CrossRef]
12. Okumus, I.; Dinler, A. Current status of wind energy forecasting and a hybrid method for hourly predictions. *Energy Convers. Manag.* **2016**, *123*, 362–371. [CrossRef]
13. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]
14. Cheney, N.P.; Gould, J.S.; Catchpole, W.R. The Influence Of Fuel, Weather And Fire Shape Variables On Fire-Spread In Grasslands. *Int. J. Wildl. Fire* **1993**, *3*, 31–44. [CrossRef]
15. Del Prete, R.; Pezzoli, A.; Pezzoli, G. Current methods for meteorological and marine forecasting for the assistance of navigation and shipping operations. *J. Navig.* **1999**, *52*, 104–118. [CrossRef]

16. Tagliaferri, F.; Viola, I.M.; Flay, R.G.J. Wind direction forecasting with artificial neural networks and support vector machines. *Ocean Eng.* **2015**, *97*, 65–73. [CrossRef]
17. Traveria, M.; Escribano, A.; Palomo, P. Statistical wind forecast for Reus airport. *Meteorol. Appl.* **2010**, *17*, 485–495. [CrossRef]
18. Rozas-Larraondo, P.; Inza, I.; Lozano, J.A. A Method for Wind Speed Forecasting in Airports Based on Nonparametric Regression. *Weather Forecast.* **2014**, *29*, 1332–1342. [CrossRef]
19. Lewis, J.M.; Lakshmivarahan, S.; Dhall, S. *Dynamic Data Assimilation*; Cambridge University Press: Cambridge, UK, 2009.
20. Cushman-Roisin, B.; Beckers, J.-M. Introduction to Geophysical Fluid Dynamics - Physical and Numerical Aspects. *Int. Geophys.* **2011**, *101*, 701–724.
21. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]
22. Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **1996**, *77*, 437–471. [CrossRef]
23. Skamarock, W.C.; Klemp, J.B.; Dudhi, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. *A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR*; University Corporation for Atmospheric Research: Boulder, CO, USA, 2008; p. 113.
24. Pielke, R.A.; Cotton, W.R.; Walko, R.L.; Tremback, C.J.; Lyons, W.A.; Grasso, L.D.; Nicholls, M.E.; Moran, M.D.; Wesley, D.A.; Lee, T.J.; et al. A comprehensive meteorological modeling system-RAMS. *Meteorol. Atmos. Phys.* **1992**, *49*, 69–91. [CrossRef]
25. Unden, P.; Rontu, L.; Jarvinen, H.; Lynch, P.; Calvo, J.; Cats, G.; Cuxart, J.; Eerola, K.; Fortelius, C.; Garcia-moya, J.A.; et al. HIRLAM-5 Scientific Documentation. 2002. Available online: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.3794 (accessed on 28 November 2019).
26. Valero, F.; Pascual, A.; Martín, M.L. An approach for the forecasting of wind strength tailored to routine observational daily wind gust data. *Atmos. Res.* **2014**, *137*, 58–65. [CrossRef]
27. Pascual, A.; Valero, F.; Martín, M.L.; Morata, A.; Luna, M.Y. Probabilistic and deterministic results of the ANPAF analog model for Spanish wind field estimations. *Atmos. Res.* **2012**, *108*, 39–56. [CrossRef]
28. Environmental Modeling Center. *The GFS Atmospheric Model, Note 442*; Environmental Modeling Center: College Park, MA, USA, 2003.
29. Janjic, Z.; Gall, R.; Pyle, M.E. Scientific Doucmentation for NMM Solver. 2010. Available online: https://opensky.ucar.edu/islandora/object/technotes%3A490 (accessed on 28 November 2019).
30. Benjamin, S.G.; Grell, G.A.; Brown, J.M.; Smirnova, T.G.; Bleck, R. Mesoscale Weather Prediction with the RUC Hybrid Isentropic–Terrain-Following Coordinate Model. *Mon. Weather. Rev.* **2004**, *132*, 473–494. [CrossRef]
31. Nagarajan, B.; Delle Monache, L.; Hacker, J.P.; Rife, D.L.; Searight, K.; Knievel, J.C.; Nipen, T.N. An Evaluation of Analog-Based Postprocessing Methods across Several Variables and Forecast Models. *Weather Forecast.* **2015**, *30*, 1623–1643. [CrossRef]
32. Buzzi, A.; Fantini, M.; Malguzzi, P.; Nerozzi, F. Validation of a limited area model in cases of mediterranean cyclogenesis: Surface fields and precipitation scores. *Meteorol. Atmos. Phys.* **1994**, *53*, 137–153. [CrossRef]
33. Zhao, W.; Wei, Y.M.; Su, Z. One day ahead wind speed forecasting: A resampling-based approach. *Appl. Energy* **2016**, *178*, 886–901. [CrossRef]
34. Manor, A.; Berkovic, S. Bayesian Inference aided analog downscaling for near-surface winds in complex terrain. *Atmos. Res.* **2015**, *164–165*, 27–36. [CrossRef]
35. Patlakas, P.; Drakaki, E.; Galanis, G.; Drakaki, E. Wind gust estimation by combining a numerical weather prediction model and statistical post-processing. *Energy Procedia* **2017**, *125*, 190–198. [CrossRef]
36. Spyrou, C.; Mitsakou, C.; Kallos, G.; Louka, P.; Vlastou, G. An improved limited area model for describing the dust cycle in the atmosphere. *J. Geophys. Res. Atmos.* **2010**, *115*, 1–19. [CrossRef]
37. Zjavka, L. Wind speed forecast correction models using polynomial neural networks. *Renew. Energy* **2015**, *83*, 998–1006. [CrossRef]
38. Bouttier, F. The Météo-France NWP System: Description, Recent Changes and Plans. 2010. Available online: http://www.umr-cnrm.fr/gmap/nwp/nwpreport.pdf (accessed on 28 November 2019).

Diagrama de Sailor

Sáenz J, **Carreno-Madinabeitia S**, Esnaola G, González-Rojí SJ, Ibarra-Berastegi G, Ulazia A. *The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models*. Geosci Model Dev 2020;13:3221–40. https://doi.org/10.5194/gmd-13-3221-2020.

Geoscientific Model Development (2019)	
Total Cites	9092
Journal Impact Factor (IF)	5.240
5 Year IF	5.768
Immediacy Index	0.962
IF witout Journal Self Cites	4.715
MGeosciences, multidisciplinary	
Rank	9/200
Quantile	Q_1, D_1
Journal IF Percentile	95.750

Tabla 3.2: Índices de calidad de la revista *Geoscientific Model Development* del año 2019

Geosci. Model Dev., 13, 3221–3240, 2020
https://doi.org/10.5194/gmd-13-3221-2020

Geoscientific Model Development
Open Access
EGU

The Sailor diagram – A new diagram for the verification of two-dimensional vector data from multiple models

Jon Sáenz[1,2], **Sheila Carreno-Madinabeitia**[3,4], **Ganix Esnaola**[5,2], **Santos J. González-Rojí**[6,7], **Gabriel Ibarra-Berastegi**[8,2], **and Alain Ulazia**[9]

[1]Dept. Applied Physics II, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s./n., 48940 Leioa, Spain
[2]Joint Research Unit BEGIK, Instituto Español de Oceanografía (IEO) Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Plentziako Itsas Estazioa, Areatza Pasealekua, 48620 Plentzia, Spain
[3]Tecnalia, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Álava, Albert Einstein 28, 01510 Vitoria-Gasteiz, Spain
[4]Faculty of Pharmacy, Applied Mathematics, Statistics and O.R. Department, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain
[5]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Europa Plaza 1, 20018 Donostia, Spain
[6]Oeschger Centre for Climate Change Research, University of Bern, 3010 Bern, Switzerland
[7]Climate and Environmental Physics, University of Bern, 3010 Bern, Switzerland
[8]Nuclear Engineering and Fluid Mechanics Dept., Escuela de Ingeniería de Bilbao, Plaza Ingeniero Torres Quevedo 1, 48013 Bilbao, Spain
[9]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Otaola etorbidea 29, 20600 Eibar, Spain

Correspondence: Jon Sáenz (jon.saenz@ehu.eus)

Received: 11 October 2019 – Discussion started: 19 November 2019
Revised: 26 May 2020 – Accepted: 8 June 2020 – Published: 15 July 2020

Abstract. A new diagram is proposed for the verification of vector quantities generated by multiple models against a set of observations. It has been designed with the objective, as in the Taylor diagram, of providing a visual diagnostic tool which allows an easy comparison of simulations by multiple models against a reference dataset. However, the Sailor diagram extends this ability to two-dimensional quantities such as currents, wind, horizontal fluxes of water vapour and other geophysical variables by adding features which allow us to evaluate directional properties of the data as well. The diagram is based on the analysis of the two-dimensional structure of the mean squared error matrix between model and observations. This matrix is separated in a part corresponding to the bias and the relative rotation of the two orthogonal directions (empirical orthogonal functions; EOFs) which best describe the vector data. Since there is no truncation of the retained EOFs, these orthogonal directions explain the total variability of the original dataset. We test the performance of this new diagram to identify the differences amongst the reference dataset and a series of model outputs by using some synthetic datasets and real-world examples with time series of variables such as wind, current and vertically integrated moisture transport. An alternative setup for spatially varying time-fixed fields is shown in the last examples, in which the spatial average of surface wind in the Northern and Southern Hemisphere according to different reanalyses and realizations from ensembles of CMIP5 models are compared. The Sailor diagrams presented here show that it is a tool which helps in identifying errors due to the bias or the orientation of the simulated vector time series or fields. The R implementation of the diagram presented together with this paper allows us also to easily retrieve the individual diagnostics of the different components of the mean squared error and additional diagnostics which can be presented in tabular form.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

It has been a long time since visual tools were recognized as an easy way to analyse different properties of datasets. This appreciation is at the root of simple and effective visualizations for exploratory data analysis such as the well-known Hovmöller diagram (Hovmöller, 1949) and the box plot (McGill et al., 1978). A visual tool for presenting temperature anomalies has also been recently recognized as a very effective way of presenting information regarding the evolution of climate to general audiences (Hawkins et al., 2019). Visual tools are very helpful in scientific inquiry; see, for instance, Peirce's diagrammatic thinking (Dörfler, 2005). Furthermore, the visualization via diagrammatic representations does not only constitute a way of interpretation. Peirce's theory of signs and other studies on scientific creative thinking show that diagrams, together with analogy or extreme thinking, also constitute a way of reasoning and knowledge generation (Dörfler, 2005; Ulazia, 2016).

Visual representation of data allows a fast and intuitive interpretation of many of their characteristics. This has led to the development of many special types of diagrams particularly in the field of model verification. These diagrams present different measures of forecast quality as in the case of the well-known relative operating characteristic curve (Wilks, 2006) and a combination of success ratio and probability of detection (Roebber, 2009) to name a few.

Boer and Lambert (2001) designed a diagram based on second-order space–time differences between model simulations and observations as a tool to diagnose the performance of climate models. Their diagram was based on simple quantities such as mean square differences, variances and Pearson's correlation coefficient between observations and model runs. They used the analytical relationship between the standard deviation of the datasets, their common correlation coefficient and the squared difference between the datasets. They also showed that the diagram could be used for the evaluation of model ensembles.

Following a similar reasoning, Taylor (2001) presented a diagram which has become a well-known and popular tool for the evaluation of model simulations against observed data (in general, a *reference* dataset). In the so-called Taylor diagram, the horizontal axis represents the standard deviation of the reference dataset, the radial distance represents the standard deviation ratio of the forecast against the reference and the angular distance from the x axis is related to the correlation coefficient between the reference dataset (also referred to as observations) and every model run. The distance from the point assigned to a model in the diagram to the point representing the reference dataset is related to the centred root mean squared error (RMSE). In the Taylor diagram, every model tested is represented by a point in the diagram, and visual inspection allows us to easily determine which points are closer (i.e. present lower error) to observations. This approach works for any number of models, and, therefore, comparing models using the Taylor diagram is in general faster and easier than using an equivalent table listing the different error measures. This explains the success of the diagram, as shown by the fact that the paper describing it has been cited more than 2300 times at the time of writing this contribution. This diagram is a tool that helps in the fast diagnosis of the relative merits of the models. Aspects such as under- or overestimated variance, incorrect phasing of the seasonal cycle and many others are reflected in the relative position of the points characterizing a model in the diagram. The diagram is flexible enough so that it can be extended to ensembles of models. More specific developments such as incorporating bootstrap techniques for the estimation of confidence intervals can be easily done (González-Rojí et al., 2018; Ulazia et al., 2017), which stresses the idea of flexibility associated with the Taylor diagram. Finally, since observed data also suffer from errors, an estimation of the relevance of these observational errors in different datasets can also be achieved by checking alternative measured datasets against the same reference as if they were models too (Fernández et al., 2007). Thus, the dispersion amongst observational datasets yields an estimate of the uncertainty of the observations (González-Rojí et al., 2019).

Pearson's correlation coefficient between two scalars plays a fundamental role in the design of Taylor diagrams, but a single universally accepted definition of the correlation coefficient in two dimensions does not exist. Jupp and Mardia (1980) recognized that any multivariate definition of a correlation coefficient equivalent to Pearson's one must be invariant to rotation, close to zero for independent datasets, smaller than or equal to a constant, and equal to that constant only if the datasets are related to each other by means of a function. Since they based their definition on these properties, they found that the sum of the squared canonical correlations was a potential definition of the squared correlation coefficient that met the previous requisites. In a previous paper, Cramer (1974) defined the two-dimensional correlation coefficient by means of the product of the canonical correlations. In this case, a low canonical correlation yields a low correlation coefficient because of the use of the product.

Stephens (1979) defined two versions of correlation between vectors by means of functions which satisfy the requirement that two perfectly correlated vector sets can be related by means of an orthogonal transformation. In this case, the vectors are assumed to share a common centre and to be unit vectors so that this measure cannot be used to identify biases between datasets or different standard deviations. In any case, the author correctly asserted that invariance to rotation does not lead to a unique definition of correlation coefficient for multivariate datasets.

Robert et al. (1985) presented an interesting review of different alternatives to compute the correlation coefficient for vector quantities. They recognized that two approaches to the problem exist. The first one is based on the use of canonical correlations between multivariate datasets. In the second ap-

proach, the definition of a two-dimensional correlation coefficient for vector datasets is based on functions which satisfy some desirable properties, such as the invariance of the correlation to the rotation of the original datasets or the existence of a limit constant for linearly related vectors as earlier suggested by Jupp and Mardia (1980).

Despite these many previous studies, it is a fact that up to today, several alternative versions of correlation coefficients between vectors exist. The fact that the definition of a two-dimensional correlation coefficient must satisfy the properties mentioned before was also followed by Crosby et al. (1993), who presented an in-depth review of previous definitions in oceanography and meteorology such as by Kundu (1976). Crosby et al. (1993) also stated different possible definitions of the correlation coefficient. Amongst them, they proposed a definition similar to the one used by Jupp and Mardia (1980). This definition was later applied to real marine and atmospheric datasets by Breaker et al. (1994) and Cosoli et al. (2008), for instance. A similar result is obtained in the case of complex correlation coefficients (Schreier, 2008). In this case, too, the literature (Hanson et al., 1992; Schreier, 2008) shows that there is not a unique definition of the complex correlation coefficient. One of the potential definitions is the one based on canonical correlation analysis, which is connected to the minimum squared error and highest mutual information in the signals being compared. This result is consistent with the definition of R^2 by Jupp and Mardia (1980) and Crosby et al. (1993).

However, the diagram designed by Taylor (2001) for scalar variables is being used by modellers when comparing vectorial quantities of model output with observations. For example, Lee et al. (2013) presented a comparison of CMIP3, CMIP5, reanalysis and satellite-based estimations of wind stress by means of the average of the Taylor diagrams for the zonal and meridional components of the wind stress as a way to apply Taylor diagrams to vector quantities. A different strategy is followed, for instance, in Jiménez et al. (2010). In this case, the behaviour of several models for the zonal and meridional components is not the same in terms of the identification of the model rankings. The best model for the zonal component in terms of its Taylor diagram is not the best one for the meridional component (see their Fig. 6). This is a typical problem which arises when using the Taylor diagram with vector data as also shown in a study about currents measured by means of an high-frequency (HF) radar (Lorente et al., 2015). It also appears in the evaluation of global climate models using zonal and meridional components of wind speed (The HadGEM2 Development Team, 2011) and in an analysis of moisture fluxes (Ibarra-Berastegi et al., 2011). A last example appears when wind stress components are analysed (Chaudhuri et al., 2013). A different alternative which allows the use of the Taylor diagram for verification of wind estimations against observations is to use it as a tool to verify the magnitude of the wind (Ulazia et al., 2016, 2017; Rabanal et al., 2019). However, even in this case, the results are limited since the information regarding errors in the direction of the vectors is lost.

In a recent paper, Xu et al. (2016) proposed a new method to overcome the deficiencies of the Taylor diagram for vector datasets and produced a new type of diagram visually equal to the original Taylor diagram but which can be used for vector quantities. It is constructed on the basis of pattern similarities of vector observations and model runs, and they call it a vector field evaluation (VFE) diagram. It is constructed from both components of the vectors which appear in the vector datasets that are used for the verification. In order to arrive at the same structure as the Taylor diagram, the authors apply some normalization to the original two-dimensional vector quantities.

However, in the original paper by Crosby et al. (1993), the authors demonstrate that two-dimensional fields showing a perfect correlation according to their definition do not have to be simple two-dimensional counterparts of what we expect in the one-dimensional case (see their Fig. 3). Thus, instead of trying to simply replicate the structure of the original Taylor diagram, we have decided to follow a new approximation which gives more information about the structure of the two-dimensional errors between vector quantities derived from models and their observational counterpart (reference dataset). In order to achieve this goal, we have based our definition on the analysis of the two-dimensional structure of the mean squared error between both vectorial datasets. This does not allow us to reduce our diagram to the well-known Taylor diagram used for scalars as the one produced by Xu et al. (2016) does. However, we hope that our diagram will be considered a valuable contribution to the set of techniques used for the evaluation of models as it visually explores other properties of the error between the vector datasets, such as the relative rotation of the major axes of variability and the underestimation (or overestimation) of variance along each principal axis of the covariance matrix. As will be shown in this contribution, this is an important diagnostic error which would otherwise be lost.

Empirical orthogonal functions (EOFs) are commonly used in the literature for the decomposition of geophysical fields in their temporal and spatial variability (Hannachi et al., 2007). The use of an EOF-based decomposition of a geophysical field is particularly relevant because it produces linear combinations of the original variables (principal components) which are uncorrelated, thus leading to better basis for subsequent stages of the analysis. These uncorrelated principal components are important bricks in the development of statistical analyses based on canonical correlation or multiple regression models, for instance (Barnett and Preisendorfer, 1987; Bretherton et al., 1992). Besides that, these linear combinations are also able to explain decreasing fractions of variance so that the EOFs form an interesting orthogonal basis for data compression and dimensionality reduction (Monahan et al., 2009). However, the reduction in variance is achieved by truncating the amount of EOFs that

are kept for the analysis to a number of EOFs lower than the rank of the corresponding covariance matrix. In the case of our paper, as will be discussed later, the original 2×2 covariance matrix is expressed by two EOFs so that there is no truncation in the process, as discussed in detail in Sect. 3.1.

It is the authors' need to find a solution to problems found in the past when using the Taylor diagram for vector quantities that inspired this proposal. The Sailor diagram provides a full analysis of the two-dimensional covariance matrix of the observed and simulated vector fields, and, at the same time, it yields exact numerical estimations of the RMSE between those vector fields. Additional diagnostics presented in this contribution such as the relative rotation of the principal axes can be obtained following our methodology. Thus, this contribution provides a useful tool for the verification of simulated vector fields.

We propose the name Sailor diagram as a joke due to the fact that it is a diagram which can be used for winds and currents (properties of geophysical fluid dynamics that sailors need to know about) and because this name is very similar to the original Taylor diagram. Thus, the name can be derived from the original Taylor just by changing two letters in the word (two letters equal the number of dimensions used in the diagram) following the idea behind Lewis Carroll's word ladder puzzles.

Section 2 presents the datasets that we have used as examples of the application of our Sailor diagram. Section 3 explains the methodology that we follow to build the two-dimensional diagram. Results are included in Sect. 4, followed by some concluding remarks in Sect. 5.

2 Data

In order to show that the diagram that we propose is of general interest and can be applied in different studies involving vector magnitudes, we have selected some examples ranging from evident variables (wind or ocean currents) to additional post-processed quantities such as vertically integrated moisture transports.

2.1 Wind data

The first wind dataset that will be used in this paper corresponds to a 1-year-long dataset of hourly wind (zonal and meridional components) from ERA5 reanalysis at the point 38° N and −124° W, by Los Angeles, and we will refer to it as reference (Ref) onwards. In order to produce synthetic models which are affected by individual sources of error, we have prepared a perturbed version of this dataset which we refer to as MOD1 and for which we have just added a constant bias of $(4.8, -6.8)\,\mathrm{m\,s^{-1}}$ to the zonal and meridional components of the wind, respectively. In order to address a second source of error (a change in the simulated direction), we have applied a counterclockwise rotation of 30° to the original dataset in order to produce MOD2. The rotation produces a change in the principal axes of the distribution of zonal and meridional wind and a new bias too since it rotates the original averaged wind. A third source of error (lack of temporal correlation) is addressed by resampling (without repetition) the original Ref dataset to produce MOD3, which is characterized by perfect mean wind (no bias) and direction of major and minor axes of the distribution of wind but no correlation of wind events. A final synthetic dataset (MOD4) is produced by scaling the wind distribution with a constant factor (2) so that both the mean and the standard deviations of wind are affected.

Next, offshore wind data are also used as our first example of a Sailor diagram constructed with realistic data. The wind dataset (zonal and meridional components) extends from 1 January 2009 to 1 January 2015 and includes five sources (Ulazia et al., 2017). Two Weather Research and Forecasting (WRF) model simulations around the Iberian Peninsula are used, one with 3DVAR data assimilation every 6 h (experiment D) and the second one without data assimilation (experiment N). ERA-Interim (ERAI) data (Dee et al., 2011) were also used to nest the two (N and D) WRF runs, and these data are also compared with observations. Fully assimilated level 3 wind analysis data from the second version of Cross-Calibrated Multi-Platform (CCMPv2) are also used (Hoffman et al., 2003; Atlas et al., 2011) for the evaluation. The previous sources will be validated against in situ observations provided by the buoy in Dragonera, near the Balearic Islands, a buoy which is managed by Spanish National Ports Authority Puertos del Estado (2015).

2.2 Ocean currents

Three different data sources of ocean surface horizontal vectorial currents are also compared with in situ data. They cover the Bay of Biscay area and include in situ observations from a deepwater buoy, remotely sensed surface HF radar currents and an ocean modelling product. Observational products, both an in situ buoy (named DONOSTIA buoy) and remotely sensed radar currents, belong to the Basque Meteorological Agency (EUSKALMET) and were obtained from https://www.euskoos.eus (last access: 3 July 2020). They provide hourly data that are punctual in the case of the buoy (approx. location 43.6° N and 2.0° W). In the case of the HF radar dataset, the data consist of a gridded dataset which covers the corner of the Bay of Biscay (approx. location 43.5–44.7° N and 3.2–1.3° W) with 5 km spatial resolution (Rubio et al., 2011, 2013; Solabarrieta et al., 2014). The ocean modelling product used in this example is the global analysis and forecast product of the Copernicus Marine Environment Monitoring Service (CMEMS), available through their data portal (identifier `GLOBAL_ANALYSIS_FORECAST_PHY_001_024`) (Madec and the NEMO team, 2008; Lellouche et al., 2018).

2.3 Vertically integrated water vapour transports

Zonal and meridional components of vertically integrated water vapour transports were calculated or downloaded from different sources. First, observations were obtained from the sounding data for A Coruña (station ID 08001; 43.36° N and −8.41° E) with a temporal resolution of 12 h for the period 2010–2014. Both components of vertically integrated moisture transport from ERAI in the original vertical levels of the European Centre for Medium-Range Weather Forecasts (ECMWF) model were downloaded by means of the Meteorological Archival and Retrieval System (MARS) repository at ECMWF for the nearest point to A Coruña.

Both moisture transport components were also calculated using the moisture and wind data from the previously mentioned N and D simulations created with the WRF model over the Iberian Peninsula as described by González-Rojí et al. (2018). The components of the moisture transport were calculated at the nearest point in WRF's grid by means of the vertical integration of the specific humidity (Sáenz et al., 2019a) and the zonal and meridional winds over the original 51 η levels of the WRF model.

2.4 Verification of spatial vector fields

An important application of the Taylor diagram is the verification of climate models, and, as such, it is often used to verify the spatial structure of climate model outputs. In order to show that the Sailor diagram proposed in this paper can also be applied for this purpose, some reanalyses are compared. The original NCEP/NCAR first generation reanalysis (Kalnay et al., 1996) is compared to more modern reanalyses such as MERRA2 (Gelaro et al., 2017), CFSv2 (Saha et al., 2014), ERAI and ERA5 (Hersbach et al., 2018). In all those cases, we have analysed the January average of the monthly values covering a common period (2011–2018), regridded by means of bilinear interpolation to the grid corresponding to the NCEP/NCAR reanalysis case (2.5° × 2.5°).

Finally, in terms of the application of the diagram to a typical case in the analysis of climate models, we use time-averaged wind speed over the Southern Hemisphere (1979–2005). This case example uses the time average of surface wind obtained from ERA5 as the reference dataset. In order to check the behaviour of the diagram when analysing ensembles of multimodels, we have also downloaded surface wind fields of the historical forcing experiment contributed by three models from the CMIP5 repository for the same period. The first set includes six realizations by the IPSL model developed at the Institute Pierre-Simon Laplace (Dufresne et al., 2013). The second one (including five realizations) derives from the MIROC model (Watanabe et al., 2010). The third case includes four realizations integrated using the HadGEM2-ES model (Collins et al., 2011) from the Hadley Centre. All the models and ERA5 reanalysis gridded fields have been bilinearly interpolated to a common 1.25° × 1° regular longitude–latitude grid. This example is selected to illustrate the way the diagram can be applied to the analysis of ensemble data even if the number of realizations for each model is different.

3 Methodology

In this section, we present the derivation of the 2×2 squared-error matrix that is the basis of the definition of the diagram that is proposed later. The two-dimensional squared error matrix is decomposed in the empirical orthogonal functions (EOFs) corresponding to the covariance matrix defined by the zonal and meridional components of observations (and similarly for the covariance matrix defined by each model). Section 3.1 describes the decomposition of the matrix $\mathbf{U}$ corresponding to the reference dataset (observations) in its EOFs. A similar notation will be used later for the decomposition of the matrix $\mathbf{V}$ corresponding to the zonal and meridional components of every model which is being compared to observations. Later, the expansion of the $\mathbf{V}$ matrix corresponding to the model is expressed as a rotation from the EOFs derived from observations (Sect. 3.2).

3.1 Decomposition of U in its EOFs

We consider a time series or spatial field of a two-dimensional vectorial variable such as horizontal wind, vertically integrated moisture transport or horizontal currents, for instance. It has been measured at an observatory or buoy (time series), or it is a time average over a grid (the case of the evaluation of a climatology derived from climate models). By now, we will consider that we are evaluating a time series of N samples, but later we will present results where the N represents the number of grid points where a time-averaged field is defined. Note that in the following presentation, $\mathbf{U}$ includes the zonal and meridional components of observations and so does $\mathbf{V}$ for a simulated dataset. The observational dataset is formed by the two-dimensional (zonal and meridional) components of vector measurements $\boldsymbol{u}_i$, with $i = 1 \ldots N$ arranged as rows in an $N \times 2$ matrix $\mathbf{U}$. The average $\overline{\boldsymbol{u}}$ of the $\boldsymbol{u}_i$ time series will be repeated as constant rows in an $N \times 2$ matrix $\overline{\mathbf{U}}$. The 2×2 covariance matrix from the zonal and meridional components of velocity anomalies in the observations is given by

$$\mathbf{S}_u = \frac{1}{N}\left(\mathbf{U} - \overline{\mathbf{U}}\right)^T \left(\mathbf{U} - \overline{\mathbf{U}}\right) = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix}. \quad (1)$$

According to the traditional use of the EOF decomposition of geophysical fields, the eigenvalues and eigenvectors of the covariance matrix from observations $\mathbf{U}$ can be computed by means of the expression

$$\mathbf{S}_u \boldsymbol{e}_{ui} = \lambda_{ui} \boldsymbol{e}_{ui}, \quad (2)$$

with $\mathbf{S}_u$ the covariance matrix in Eq. (1), $\boldsymbol{e}_{ui}$ the ith eigenvector of the observational vector field and λ_{ui} the corresponding

ith eigenvalue so that

$$f_{ui} = \frac{\lambda_{ui}}{\sum_{i=1}^{r} \lambda_{ui}} \tag{3}$$

represents the fraction of variance in observations explained by the linear combination of the original variables defined by the ith eigenvector of the covariance matrix (Monahan et al., 2009). In the general case of the EOF analysis in climatological analyses, the rank of the covariance matrix r in Eq. (3) extends to (at most) the minimum between the number of grid points (N_g) and the number of samples in the dataset (N). In the general case, in order to achieve a truncation of the original dataset, a number t of EOFs lower than the rank of the covariance matrix ($t < r$) is selected so that the signal in the subspace that can not be spanned by eigenvectors $\boldsymbol{e}_{uj}$ with $j = t+1 \ldots r$ becomes the part of the original dataset which is truncated. However, in our use of EOFs below, the original covariance matrix as defined in Eq. (1) is of rank 2 or full rank for any realistic non-linear flow. Since two EOFs ($t = r = 2$) will be used in the expansion of the datasets, no truncation is applied and the full variance in the original dataset will be analysed in the equations that follow.

Thus, the $\mathbf{U}$ matrix can be expressed by means of the two empirical orthogonal functions of the original vector data (which constitute a complete basis of the horizontal plane) by using the expression

$$\mathbf{U} = \overline{\mathbf{U}} + \mathbf{P}_u^* \mathbf{\Sigma}_u \mathbf{E}_u^T = \overline{\mathbf{U}} + \mathbf{P}_u \mathbf{E}_u^T, \tag{4}$$

with $\mathbf{P}_u^*$ (an $N \times 2$ matrix) the standardized principal components of the $\mathbf{U}$ data, $\mathbf{\Sigma}_u$ (2×2 matrix) the standard deviations (σ_{1u} and σ_{2u}) of the leading and second EOFs of the $\mathbf{U}$ field, $\mathbf{E}_u$ (2×2 matrix) the matrix holding the orthogonal rotation matrix leading to the empirical orthogonal functions of the $\mathbf{U}$ field arranged as columns, and $\mathbf{P}_u = \mathbf{P}_u^* \mathbf{\Sigma}_u$ ($N \times 2$ matrix) the variance-holding principal components. Please note that when the standardized principal components $\mathbf{P}_u^*$ are used, this matrix is always multiplied by the corresponding standard deviations so that no variance is lost in the process. Thus, the anomalies of wind are computed without any loss of variance as

$$\mathbf{U} - \overline{\mathbf{U}} = \mathbf{P}_u^* \mathbf{\Sigma}_u \mathbf{E}_u^T = \mathbf{P}_u \mathbf{E}_u^T, \tag{5}$$

the corresponding principal components as

$$\mathbf{P}_u = \mathbf{P}_u^* \mathbf{\Sigma}_u = \left(\mathbf{U} - \overline{\mathbf{U}}\right) \mathbf{E}_u, \tag{6}$$

and their standardized counterparts as

$$\mathbf{P}_u^* = \left(\mathbf{U} - \overline{\mathbf{U}}\right) \mathbf{E}_u \mathbf{\Sigma}_u^{-1}. \tag{7}$$

Unless the wind (current) time series is completely arranged across a straight line (something which is very unlikely in observed vector variables unless the flow is stationary and laminar), $\mathbf{\Sigma}_u$ is a full-rank diagonal matrix:

$$\mathbf{\Sigma}_u = \begin{pmatrix} \sigma_{1u} & 0 \\ 0 & \sigma_{2u} \end{pmatrix}, \tag{8}$$

with $\sigma_{1u} > \sigma_{2u}$. Due to the fact that the rotation matrix is always full rank (in the two-dimensional space spanned by the zonal and meridional components, given enough samples), the $\mathbf{E}_u$ matrix can also be interpreted geometrically as a rotation matrix expressed as a function of the angle θ_u formed by the leading (second) EOF with the zonal (meridional) axis as follows:

$$\mathbf{E}_u = \begin{pmatrix} \cos\theta_u & -\sin\theta_u \\ \sin\theta_u & \cos\theta_u \end{pmatrix}. \tag{9}$$

The first column of the $\mathbf{E}_u$ matrix is the first eigenvector of observations in the horizontal plane, $\boldsymbol{e}_{u1}$. Similarly, the second column of $\mathbf{E}_u$ corresponds to $\boldsymbol{e}_{u2}$, the second eigenvector of the observational covariance matrix.

The principal components and EOF rotation matrices fulfil the well-known orthogonality properties

$$\mathbf{P}_u \mathbf{P}_u^T = \mathbf{\Sigma}_u^2, \tag{10}$$

as do the standardized principal components,

$$\mathbf{P}_u^* \mathbf{P}_u^{*T} = \mathbb{1}, \tag{11}$$

and eigenvectors (EOFs) in the horizontal plane,

$$\mathbf{E}_u \mathbf{E}_u^T = \mathbf{E}_u^T \mathbf{E}_u = \mathbb{1}. \tag{12}$$

Figure 1a illustrates in a scatterplot the distribution of measurements of zonal and meridional wind components in the Ref dataset and is presented to make the next step in the derivation of the Sailor diagram easily understandable. Figure 1b shows on top of the previous scatterplot the ellipses centred on the mean of the reference dataset applied in the Sailor diagrams by using the semi-major and semi-minor axes as defined by the EOF decomposition of the two-dimensional covariance matrix of the zonal and meridional components of the original vector field, the directions of the principal axes (matrix $\mathbf{E}_u$), and the standard deviations corresponding to the principal components $\mathbf{P}_u$. From Eqs. (7) and (11), the quadratic form leading to the ellipses in the diagram can be obtained by applying the Frobenius norm to Eq. (11) as

$$||\mathbf{P}_u^* \mathbf{P}_u^{*T}||_F = ||\left(\mathbf{U} - \overline{\mathbf{U}}\right) \mathbf{E}_u \mathbf{\Sigma}_u^{-2} \mathbf{E}_u^T \left(\mathbf{U} - \overline{\mathbf{U}}\right)^T||_F = 1. \tag{13}$$

The principal components are combined according to the quadratic form in Eq. (13). This shows that the ellipse produced from the EOF decomposition of the two-dimensional covariance matrix is a good way to make a simple and clear

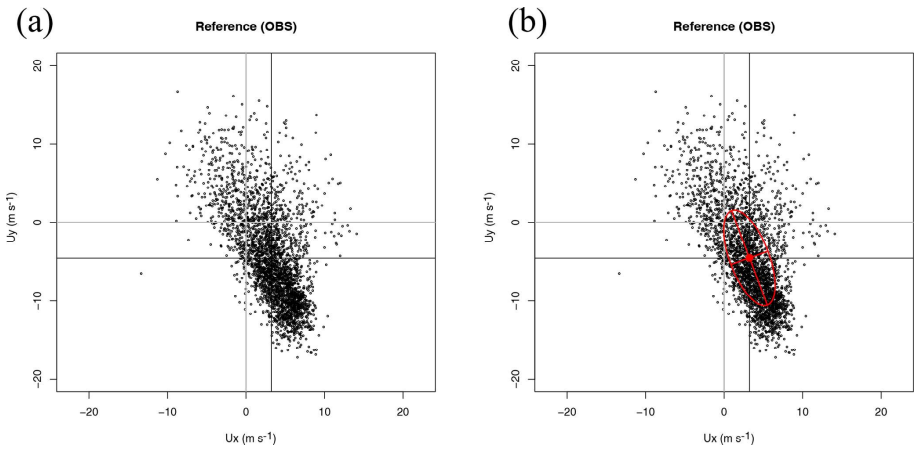

Figure 1. Scatterplot of wind in dataset Ref **(a)** and its decomposition in terms of the principal axes corresponding to the covariance matrix of the zonal and meridional components **(b)** as defined in Eq. (4).

representation of the original scatterplot. The eccentricity of the ellipse,

$$\varepsilon_u = \sqrt{1 - \frac{\sigma_{2u}^2}{\sigma_{1u}^2}}, \tag{14}$$

is an interesting indicator for additional diagnostics designed for testing the reliability of rotation angles due to the degeneracy of the eigenvalues.

Following similar notation to the one used for the observations (**U** matrix), the time series (or time-averaged constant field over N points in a grid) of simulated wind (current, wave energy flux, vertically integrated moisture transport, etc.) at the same observatory (or the closest grid point) formed by the two-dimensional (zonal and meridional components) simulations $\boldsymbol{v}_i$, with $i = 1 \ldots N$, will be arranged as rows in an $N \times 2$ matrix **V**. The average vector from model data $\overline{\boldsymbol{v}}$ is arranged as constant rows in an $N \times 2$ matrix $\overline{\mathbf{V}}$. The **V** matrix (and its anomalies) can be expressed, as was done for observations in Eqs. (4) and (5) above, by means of the empirical orthogonal functions of the two-dimensional covariance matrix from simulated zonal and meridional components of wind (current, moisture transport, etc.) data:

$$\begin{aligned} &\mathbf{V} = \overline{\mathbf{V}} + \mathbf{P}_v^* \boldsymbol{\Sigma}_v \mathbf{E}_v^T = \overline{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_v^T \;\Leftrightarrow \\ &\mathbf{V} - \overline{\mathbf{V}} = \mathbf{P}_v^* \boldsymbol{\Sigma}_v \mathbf{E}_v^T = \mathbf{P}_v \mathbf{E}_v^T, \end{aligned} \tag{15}$$

with equivalent interpretations and equal ranks for $\mathbf{P}_v^*$, $\boldsymbol{\Sigma}_v$, $\mathbf{E}_v$ and $\mathbf{P}_v = \mathbf{P}_v^* \boldsymbol{\Sigma}_v$ as presented before for observations.

3.2 Expansion of the matrix V in the EOFs defined by observations

In general, the mean values and EOFs derived from observations (**U**) and simulations (**V**) will not be the same. This is shown in Fig. 2, with panel (a) clearly showing a change in the bias between both datasets and a counterclockwise rotation in the case of panel (b), as is expected from the way these synthetic datasets were produced. It is clearly seen that in the case of MOD1 the structure of the covariance matrix has not changed, whilst a different orientation (but no scaling of the semi-major and semi-minor axes) appears in the case of MOD2.

In order to identify these kinds of errors (derived from rotations of the axes), the orthonormal EOFs in the $\mathbf{E}_v$ matrix can be expressed as the result of a rotation applied to the EOFs derived from the observations (accepting these as *true* EOFs). Thus, the rotation matrix $\mathbf{R}_{vu}$ is defined by an angle $\theta_{vu} = \theta_v - \theta_u$ as

$$\mathbf{R}_{vu} = \begin{pmatrix} \cos\theta_{vu} & -\sin\theta_{vu} \\ \sin\theta_{vu} & \cos\theta_{vu} \end{pmatrix}, \tag{16}$$

so that

$$\mathbf{E}_v = \mathbf{R}_{vu} \mathbf{E}_u, \tag{17}$$

$$\mathbf{V} = \overline{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_u^T \mathbf{R}_{vu}^T, \tag{18}$$

and the corresponding principal components can be expanded as

$$\mathbf{P}_v = \left(\mathbf{V} - \overline{\mathbf{V}}\right) \mathbf{R}_{vu} \mathbf{E}_u = \tilde{\mathbf{V}} \mathbf{E}_u, \tag{19}$$

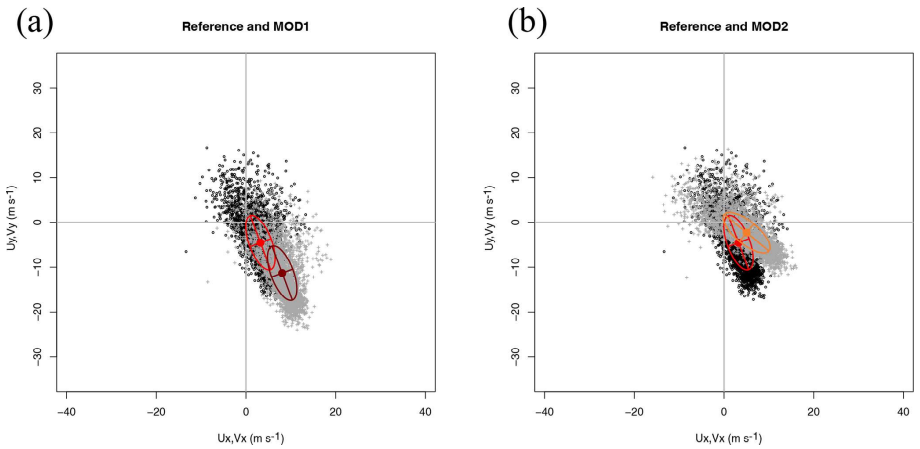

Figure 2. Scatterplot of wind in datasets Ref (panel **a**, black circles) and MOD1 (panel **a**, grey crosses) and their decomposition in terms of the principal axes corresponding to the covariance matrix of the zonal and meridional components of each dataset as defined in Eqs. (12) and (15). The comparison of the reference dataset (black circles) with model MOD2 (grey crosses) is shown in panel **(b)**.

with $\tilde{\mathbf{V}} = \left(\mathbf{V} - \overline{\mathbf{V}}\right)\mathbf{R}_{vu}$ representing the model-based $\mathbf{V}$ anomalies *rotated* to the basis given by the EOFs corresponding to observations.

Since both $\boldsymbol{e}_{u1}$ and $-\boldsymbol{e}_{u1}$ are solutions of the eigenvalue equation when the diagonalization of the two-dimensional covariance matrix is performed (the same happens with $\boldsymbol{e}_{v1}$ and $-\boldsymbol{e}_{v1}$ for model data), θ_{vu} may make it difficult to understand values even for eigenvectors which span similar subspaces. This is due to the fact that both $\theta_{vu} = 0$ and $\theta_{vu} = \pi$ refer to eigenvectors that point in perfect directions. In order to provide an easier to interpret diagnostic of the adequacy of the EOFs from observations and the model, the absolute value of the congruence coefficient (Cheng et al., 1995) can also be used. It is defined as

$$g_{ii} = |\boldsymbol{e}_{ui} \cdot \boldsymbol{e}_{vi}|, \tag{20}$$

and it measures the agreement between the pairs of EOFs from observations ($\boldsymbol{e}_{ui}$) and models ($\boldsymbol{e}_{vi}$). Since this coefficient equals the cosine of the angle between both directions and since the absolute value is used, the closer its value is to 1, the better the agreement will be between $\boldsymbol{e}_{ui}$ and $\boldsymbol{e}_{vi}$. Due to the orthogonality relationship between the EOFs, only the congruence coefficient for EOF1 is computed since it is equal to the one computed using EOF2 (matrices $\mathbf{E}_u$ and $\mathbf{E}_v$ are orthonormal).

3.3 Expansion of the mean squared error

The (2×2) matrix that represents the mean squared error between the $\mathbf{U}$ and $\mathbf{V}$ datasets is given by

$$\boldsymbol{\Delta}^2_{uv} = \frac{1}{N}(\mathbf{V} - \mathbf{U})^T (\mathbf{V} - \mathbf{U}), \tag{21}$$

and the aggregated scalar mean squared error of both components of the vector dataset is given by its Frobenius norm

$$\varepsilon^2 = ||\boldsymbol{\Delta}^2_{uv}||_F. \tag{22}$$

Substituting Eqs. (4) and (15) into Eq. (21), it can be shown that

$$\begin{aligned}\boldsymbol{\Delta}^2_{uv} &= \frac{1}{N}\mathbf{B}^2_{uv} + \frac{1}{N}\left(\mathbf{S}^T_{uv} + \mathbf{S}_{uv}\right) + \frac{1}{N}\mathbf{D}_{uv} \\ &= \frac{1}{N}\mathbf{B}^2_{uv} + \frac{1}{N}\mathbf{C}_{uv} + \frac{1}{N}\mathbf{D}_{uv},\end{aligned} \tag{23}$$

with

$$\mathbf{B}^2_{uv} = \left(\overline{\mathbf{V}} - \overline{\mathbf{U}}\right)^T \left(\overline{\mathbf{V}} - \overline{\mathbf{U}}\right), \tag{24}$$

$$\begin{aligned}\mathbf{S}_{uv} &= \left(\mathbf{E}_v \boldsymbol{\Sigma}_v \mathbf{P}^{*T}_v - \mathbf{E}_u \boldsymbol{\Sigma}_u \mathbf{P}^{*T}_u\right)\left(\overline{\mathbf{V}} - \overline{\mathbf{U}}\right) \\ &= \left(\mathbf{E}_v \mathbf{P}^T_v - \mathbf{E}_u \mathbf{P}^T_u\right)\left(\overline{\mathbf{V}} - \overline{\mathbf{U}}\right),\end{aligned} \tag{25}$$

and

$$\begin{aligned}\mathbf{D}_{uv} &= \mathbf{E}_u \boldsymbol{\Sigma}^2_u \mathbf{E}^T_u + \mathbf{E}_v \boldsymbol{\Sigma}^2_v \mathbf{E}^T_v \\ &\quad - \left(\mathbf{E}_u \boldsymbol{\Sigma}_u \mathbf{P}^{*T}_u \mathbf{P}^*_v \boldsymbol{\Sigma}_v \mathbf{E}^T_v + \mathbf{E}_v \boldsymbol{\Sigma}_v \mathbf{P}^{*T}_v \mathbf{P}^*_u \boldsymbol{\Sigma}_u \mathbf{E}^T_u\right),\end{aligned} \tag{26}$$

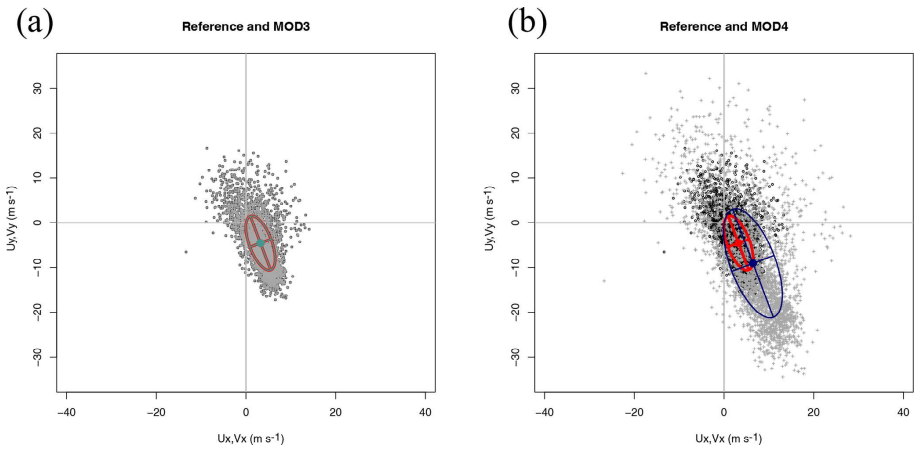

Figure 3. Scatterplot of wind in datasets Ref (panel **a**, black circles) and MOD3 (panel **a**, grey crosses) and their decomposition in terms of the principal axes corresponding to the covariance matrix of the zonal and meridional components of each dataset. The comparison of the reference dataset (black circles) with model MOD4 (grey crosses) is shown in panel **(b)**.

which can also be written using non-standardized $\mathbf{P}_u$ and $\mathbf{P}_v$ principal components as

$$\mathbf{D}_{uv} = \mathbf{E}_u \boldsymbol{\Sigma}_u^2 \mathbf{E}_u^T + \mathbf{E}_v \boldsymbol{\Sigma}_v^2 \mathbf{E}_v^T - \left(\mathbf{E}_u \mathbf{P}_u^T \mathbf{P}_v \mathbf{E}_v^T + \mathbf{E}_v \mathbf{P}_v^T \mathbf{P}_u \mathbf{E}_u^T \right). \quad (27)$$

$\mathbf{B}_{uv}^2$ represents the part of the squared error which is due to the magnitude of the bias vector (difference of both means) between both vector datasets.

The (symmetric) matrix $\mathbf{C}_{uv} = \mathbf{S}_{uv}^T + \mathbf{S}_{uv}$ reflects the error which is due to the projection of the bias onto the differences of vector anomalies. Since the bias matrices are constant, the sum of the projections become the sum of anomalies, and, as such, they become zero. This interpretation is clear if Eq. (5) and the corresponding one for the model anomalies are substituted into the definition of the matrix $\mathbf{S}_{uv}$ in Eq. (25), yielding

$$\mathbf{S}_{uv} = \left(\left(\mathbf{V} - \overline{\mathbf{V}} \right) - \left(\mathbf{U} - \overline{\mathbf{U}} \right) \right)^T \left(\overline{\mathbf{V}} - \overline{\mathbf{U}} \right) = (\mathbf{V} - \mathbf{U})^T \left(\overline{\mathbf{V}} - \overline{\mathbf{U}} \right) - \left(\overline{\mathbf{V}} - \overline{\mathbf{U}} \right)^T \left(\overline{\mathbf{V}} - \overline{\mathbf{U}} \right) = 0. \quad (28)$$

Since this matrix is zero, $\mathbf{C}_{uv}$ will also be zero.

Finally, the matrix $\mathbf{D}_{uv}$ is related to the covariance matrix of anomalies which is also clearly seen if Eq. (5) and the corresponding one for simulated data are substituted into Eq. (27).

In order to improve the graphical interpretation of the components of the error, the expression of the empirical orthogonal functions of $\mathbf{V}$ as a rotation of the *true* ones (derived from observations $\mathbf{U}$) is used. Thus, considering Eq. (17), the matrix $\mathbf{D}_{uv}$ above can be rewritten in terms of the EOFs corresponding to observations as

$$\mathbf{D}_{uv} = \mathbf{E}_u \boldsymbol{\Sigma}_u^2 \mathbf{E}_u^T + \mathbf{R}_{vu} \mathbf{E}_u \boldsymbol{\Sigma}_v^2 \mathbf{E}_u^T \mathbf{R}_{vu}^T - \left(\mathbf{E}_u \mathbf{P}_u^T \mathbf{P}_v \mathbf{E}_u^T \mathbf{R}_{vu}^T + \mathbf{R}_{vu} \mathbf{E}_u \mathbf{P}_v^T \mathbf{P}_u \mathbf{E}_u^T \right). \quad (29)$$

If $\boldsymbol{\Gamma}_{vu} = \mathbf{P}_u^T \mathbf{P}_v$ is proportional to the covariance between both datasets' principal components, the above expression can be written as

$$\mathbf{D}_{uv} = \mathbf{E}_u \boldsymbol{\Sigma}_u^2 \mathbf{E}_u^T + \mathbf{R}_{vu} \mathbf{E}_u \boldsymbol{\Sigma}_v^2 \mathbf{E}_u^T \mathbf{R}_{vu}^T - \left(\mathbf{E}_u \boldsymbol{\Gamma}_{vu} \mathbf{E}_u^T \mathbf{R}_{vu}^T + \mathbf{R}_{vu} \mathbf{E}_u \boldsymbol{\Gamma}_{vu}^T \mathbf{E}_u^T \right). \quad (30)$$

The interpretation of this expression is that all the matrices involved in the mean squared error can be expressed in the axes defined by the leading and second EOFs of the $\mathbf{U}$ (observational) dataset. Thus, using the axes corresponding to the observational dataset $\mathbf{U}$, we can produce a diagram which gives us a fast visual impression of the structure of the error in two-dimensional variables in the same way the Taylor diagram performs for univariate datasets. Therefore, the diagram presented in this contribution includes not only scalar information in the estimation of the error but also information regarding the main directions of variability of the vectors and their differences by means of the characteristics of the ellipses defined by Eq. (19) from the different datasets.

Figure 3 presents two interesting cases. The first case, MOD3, is implausible from the point of view of a real model, but it constitutes an interesting case study to analyse the

Table 1. Individual components of the error for the synthetic datasets used for the illustration of the methodology. σ^2 represents the total variance ($m^2\,s^{-2}$) of every dataset as computed from the original zonal and meridional components. $\sum_i \sigma_i^2$ represents the variance ($m^2\,s^{-2}$) of wind for every dataset (reference or model) as computed from the EOF decomposition (axes of the ellipses in the diagrams). θ_u and θ_v represent the angles (radians) of the semi-major axes of the ellipses calculated for reference and models. θ_{vu} (radians) represents the relative rotation of the semi-major axis of the model data with respect to the observations. R^2 represents the two-dimensional squared correlation coefficient (sum of the squared canonical correlations). |bias| represents the magnitude of the bias ($m\,s^{-1}$). RMSE lists the root mean squared errors ($m\,s^{-1}$). The eccentricity of the ellipses (ε) is the same for all the synthetic datasets because of the way they have been built. Finally, g_{11} represents the congruence coefficient (Eq. 20) for EOF1 of all models with respect to EOF1 as derived from observations.

Model	σ^2	$\sum_i \sigma_i^2$	θ_u	θ_v	θ_{vu}	R^2	|bias|	RMSE	ε	g_{11}
Ref	47.56	47.56	1.93						0.92	
MOD1	47.56	47.56		1.93	0.00	2.00	8.34	5.56	0.92	1.00
MOD2	47.56	47.56		2.46	0.52	2.00	2.88	8.69	0.92	0.87
MOD3	47.56	47.56		−1.21	0.72	0.00	0.00	1.52	0.92	1.00
MOD4	190.24	190.24		1.93	0.00	2.00	5.56	11.76	0.92	1.00

properties of the diagram. In MOD3, a simple random permutation of the original observations has been performed. Thus, there are neither biases nor rotations of the principal axes. From the point of view of the graphical example shown, it seems that the model is perfect, but it is not due to the lack of temporal correlation between model and observations. This is only apparent if the full RMSE is taken into account, as shown in Table 1. Thus, a legend with the RMSEs as defined in Eq. (22) must be added to the plot in order to arrive at a precise comparison of datasets. The comparison of columns σ^2 and $\sum_i \sigma_i^2$ in Table 1 shows that the full variance of the datasets is taken into account in the EOF decomposition as both columns present the same values.

On the other side, panel (b) in Fig. 3 shows that for the scaled dataset (MOD4), the sizes of the major and minor axes of the ellipses allow a fast visual detection of the scaling present in the dataset. The individual components of the error for all the synthetic datasets used in the description of the methodology are also presented in Table 1. The eighth column shows in full the RMSEs between vector fields. It is apparent from this aggregated estimation of error that it properly evaluates the differences due to the lack of correlation that have been mentioned in the case of MOD3 (no bias and perfect orientation and axes of the ellipses) too. The rotation angle (column θ_{vu} in radians) correctly identifies the way the errors have been introduced in the different synthetic models. Despite the rotation of the ellipses apparent in columns θ_u and θ_v (the case of MOD2), the fact that the semi-axes are of the same relative length is clearly seen by the value of the eccentricity ε, which also supports the way the ellipses are presented in Figs. 1–3. On the other side, the interpretation of the angles is complicated by the fact that both $\boldsymbol{e}_{ui}$ and $-\boldsymbol{e}_{ui}$ are a correct solution to the eigenvalue problem in Eq. (2). This is apparent in the case of MOD3, in which the eigenvalue problem yields eigenvectors pointing in the same direction with a different sign so that $\theta_v = -1.21 + \pi$ yields the same value as θ_u. The orientation of both eigenvectors is the same for all models except MOD2, as effectively shown by column g_{11} in Table 1, which holds the absolute value of the congruence coefficient.

The different properties of the synthetic datasets presented so far can be abbreviated in Fig. 4, which presents in panels (a) and (b) uncentred and centred (respectively) versions of the Sailor diagram. In the uncentred version of the Sailor diagram, each ellipse, as defined by Eq. (13), is centred on its own average. This allows an easy interpretation of the bias term. In order to improve the interpretability of the rotation and scaling parameters of the ellipses (semi-major and semi-minor axes and standard deviations), the ellipse corresponding to observations is also drawn in grey centred on the same average of every model. In this way, the rotations and scalings of the vectors produced by models can easily be compared against the ones drawn from observations. However, in some cases (depending on the relative values of the bias and the standard deviations), it might be more interesting to plot all the ellipses centred on the mean corresponding to the observations and identify the bias using coloured dots, as shown in the centred version of the diagram (panel b in Fig. 4).

An additional reason which supports the Sailor diagram introducing powerful diagnostics for vector data is properly shown in Table 1. According to the column which shows the squared correlation coefficient, all models show a perfect match ($R^2 = 2$) for the two-dimensional correlation coefficient except the one built by randomly resampling the data (MOD3). However, Figs. 2 and 4 clearly show that the wind data in MOD2 is rotated with respect to the reference dataset. This is not detected by R^2 because it yields perfect results by design when there is a linear relationship between both vector datasets (Crosby et al., 1993). However, an analysis based on the full components of the RMSE, as the one performed in the Sailor diagram does (Fig. 4 and Table 1), clearly highlights these directional problems. The squared two-dimensional correlation $R^2 = 2$ reflects that there is a perfectly linear relationship (rotation in this case). However, a full diagnostic of the differences between observations and

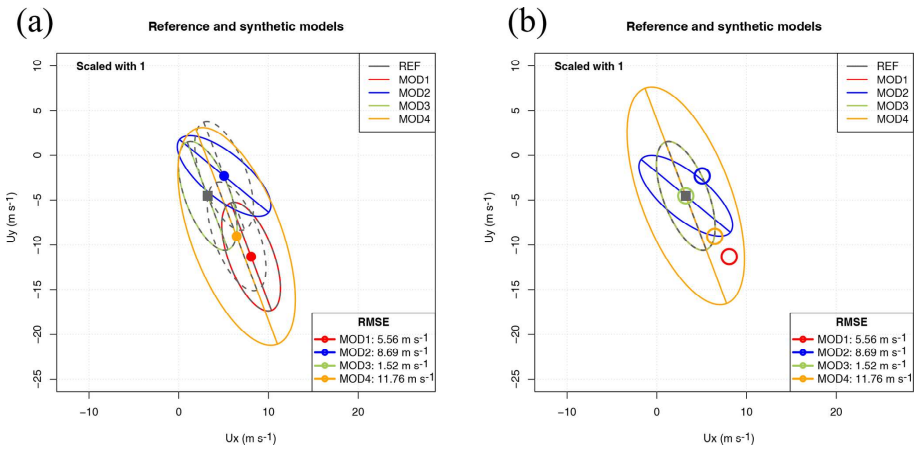

Figure 4. Uncentred **(a)** and centred **(b)** versions of the Sailor diagram after placing the ellipses from all the synthetic datasets on the same plot.

model data (such as finding the rotation angle previously mentioned) must involve a full analysis of the directional error.

3.4 Extension of the methodology to spatial fields

In the case of the analysis of the ability of models to represent the spatial distribution of an averaged field (a typical use of the Taylor diagram in climatology, for instance), there is no change needed to the diagram defined so far. Instead of using the T mode of principal components (covariance matrix defined by temporal covariances), we can just use the S mode in the traditional terminology of principal components (Compagnucci and Richman, 2008). Thus, in the previous description, N will run along the grid points, and the two-dimensional biases and covariances are computed in the spatial domain, but the error analysis is still being performed onto two-dimensional vectors. As an example of this very common case in the application of Taylor diagrams to climatology, we present an example including the comparison of multi-year averages of Northern Hemisphere surface wind vectors. In the case of spatial grids, an external standard area weighting by means of factors given by $\sqrt{\cos\phi}$ with ϕ latitude (North et al., 1982) is commonly applied to the data in order to avoid an excessive weight in the results of points in polar latitudes which represent a much lower area in a regular longitude–latitude grid.

3.5 Use of the diagram with ensembles of models

As a final example, the use of the diagram with a multimodel ensemble is shown. In this case, the long-term (27 years) climatologies of surface wind over the Southern Hemisphere from three models with a different number of realizations are compared with the corresponding climatology from ERA5. As described above, since this also involves a comparison of data on a regular longitude–latitude grid, the covariance matrix is built over the spatial points and the external weights given by $\sqrt{\cos\phi}$ with ϕ latitude (North et al., 1982) are applied to avoid an overrepresentation of polar regions in the results.

3.6 R package implementing the methodology

The authors have created an R package called `SailoR` which is freely available in the Comprehensive R Archive Network (CRAN). The package has been used to produce the plots presented in Sect. 4, and the code to prepare some of these plots and tables is provided as examples in the manual of the package. Besides producing the diagrams shown as examples in this paper, the package also computes all the individual terms used in the analysis of the RMSE as described in Sect. 3. Thus, different aspects of the main principal axes, their relative rotation, the two-dimensional correlation coefficient and the combined RMSEs can be readily analysed for different vector datasets and exported to tables which can be presented in publications.

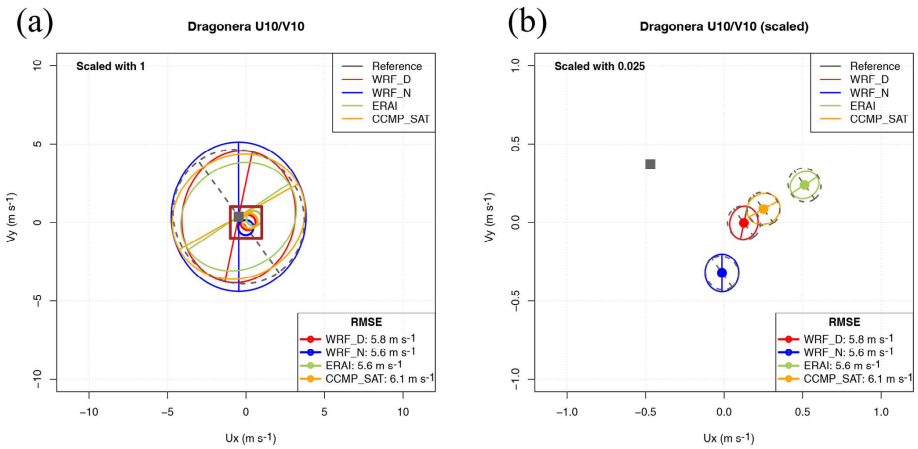

Figure 5. Sailor diagram with default parameters **(a)** and ellipses with a scale factor of 0.025 to improve visibility of the directional error **(b)** for the wind observed and simulated in Dragonera (buoy in the Mediterranean).

4 Use of the elements in the error matrix in the diagram

4.1 Wind over a Mediterranean location

The first example of a Sailor diagram built using real data is shown in Fig. 5 (panel a). In it, the x axis represents the zonal component of wind and the y axis its meridional component. The mean 2D vector corresponding to each of the datasets is represented by a coloured circle except for the reference dataset, which uses a grey square. The leading EOF of the two-dimensional covariance matrix of zonal and meridional components of every dataset is represented by the direction corresponding to the semi-major axis of the ellipse that is plotted centred on every model's mean value (same colour as the one used to represent the model's mean). The second EOF of each model is perpendicular to the previous direction by design due to the orthonormality constraint in Eq. (12). The grey ellipse centred on each model's mean represents the EOF from the reference dataset (observations). Thus, the angle between the coloured and grey semi-major axes represents the relative rotation (θ_{vu}) between EOFs from observations and simulations. The lengths of the semi-major and semi-minor axes (colour and grey) show the variances explained by each EOF (model and reference) at their principal axes. The comparison of these lengths between coloured and grey ellipses allows us to address the question of whether the model underestimates or overestimates the variances at each of the principal axes. In this particular example, since the model versus observation biases are much lower than the variance explained by the principal axes defined by the EOFs, the interpretation of this diagram is not very easy. However, it is already showing the main directions of the error matrices, their biases and the position of the reference dataset. The legend in the lower-right corner shows the total RMSEs given by Eq. (22) in Sect. 3, which takes into account both the contribution from the bias (distance of the points to the reference dataset's mean) and the different orientation and lengths of the major and minor axes (EOFs).

In order to show that different designs optimize the information transmitted by the diagram, in the second diagram prepared using the data from the same example, the ranges of both axes are limited and the ellipses corresponding to the main directions of the error matrix are accordingly scaled by means of a small scale factor (0.025). The brown square in panel (a) shows the area which is amplified in panel (b), and it illustrates the role played by the scale factor, which reduces (or amplifies) the size of the axes of the ellipses, thus making it easier to appreciate the relative differences in biases while still making it possible to get access to the information relative to the rotation of the principal axes. In the scaled version of the diagram (Fig. 5, panel b), it can be seen that the distance between every coloured point corresponding to a given model to the grey square represents the bias amongst the datasets, and they can effectively be visually compared. On the other hand, the grey ellipses and their semi-major and semi-minor axes show the main structure of the variability of the reference dataset. This grey ellipse is plotted centred on the point representing the mean of every model, where the EOFs corresponding to that model are also shown for comparison. Both ellipses (the one corresponding to the model being analysed and the one corresponding to the reference dataset) are scaled by the same scale factor so that they are not deformed during the scaling process. The

use of ellipses and their major and minor axes allows us to easily compare the main directions of variability of the observed (grey) and modelled (coloured) winds. It shows that the ones corresponding to the WRF model are the closest ones to observations. It can be seen that both WRF simulations show a smaller rotation of their major axes with respect to the one from observations. The model EOFs are almost orthogonal from the ones in observations in the cases of ERAI and CCMPv2 (CCMP_SAT in legend). The legend in the lower-right corner, in any case, presents the real RMSEs without scaling their values.

In this particular case, it might seem sensible to think that the fact that the variances of the major and minor axes are very close points to a weakness in the diagram since, in that case, the determination of the angle of the axes will be arbitrary. However, it has to be considered that the final index of agreement would still be the RMSE, which does not depend on the eigenvectors of the covariance matrix. Thus, the results in terms of direction might not be very reliable in a case when the eigenvalues are degenerated, but the RMSE is not affected by this problem. Thus, the use of the eccentricity of the ellipses (provided as an output in our R package) can be useful in diagnosing those cases (in which eccentricity is very low) that make estimations of relative rotations difficult. For a more precise determination of the reliability of the rotation angle, a bootstrap analysis of the rotation angles can also be conducted, if needed, since the evaluation of the angles is independent of the production of the diagram.

4.2 Surface current in the Bay of Biscay

Figure 6 (panel a) shows an alternative version of the Sailor diagram. In this particular case, the bias is relatively low. Thus, in order to ease the interpretation of the structure of the errors, the ellipses representing the first and second EOFs are drawn on top of the point corresponding to observations. The fact that the bias is small only affects the part of the RMSE derived from the term $\mathbf{B}_{uv}^2$ in Eq. (23). As in the previous case, they are scaled (4 times larger) in order to improve their visibility. It is clear that the relevant part in terms of the errors of models versus observations is not the bias but the way the variability is represented, instead. The HF radar data's leading EOF (observational data, actually) is closer to the one from in situ observations, as could be expected since both cases represent observational (in situ versus remote) estimations of currents. In this case, the ellipses clearly show not only the difference in the orientation of the EOFs but also the underestimation of the variability present both in radar data and especially in the case of model data. As in the previous case, the legend in the lower-right corner shows unscaled total RMSEs.

4.3 Vertically integrated water vapour transport

The Sailor diagram for the vertically integrated water vapour transport can be seen in Fig. 6 (panel b). In this case, the errors associated with the bias are smaller than the error associated with the covariance. However, since the errors in the anomalies are not very large, the visibility of the diagram has been increased by plotting all of the ellipses on top of the observational point (centred diagram). In this way, the errors in direction can be easily identified. For clarity, the ellipses are again scaled with a scale factor of 0.1. It can be seen that the estimation of the EOFs is closer in the case of the simulation with data assimilation, both in direction and, particularly, in the amount of variance represented since WRF N and ERAI slightly overestimate the water vapour fluxes.

A selection of the tabular results corresponding to the RMSE between observed and modelled vertically integrated water vapour transport is presented in Table 2. Different aspects of the main principal axes such as their semi-major and semi-minor axes, their relative rotation, the two-dimensional correlation coefficient, and the combined RMSE can be readily analysed for the water vapour transport vectors. The two-dimensional correlation coefficient R^2 and the RMSE are better for WRF D than for the other models. There is good agreement in the overall orientation of the leading EOF for all datasets, with the bias being smallest for ERAI.

4.4 Spatial distribution of seasonally averaged surface wind

As an example of the potential uses of the Sailor diagram, Fig. 7 (panel a) represents in an uncentred version of the Sailor diagram the agreement of the January-averaged Northern Hemisphere surface wind from different reanalyses using a scale factor of 0.15. In order to show that the use of different linestyles and colours can lead to diagrams which can be better interpreted, panel (a) is presented with different linestyles. On the other side, Fig. 7 (panel b) shows the agreement of the January-averaged Northern Hemisphere surface wind from different reanalyses using a scale factor of 0.15 in a centred version of the Sailor diagram. In these cases, we are assuming that ERA5 corresponds to the "perfect" dataset (observations). The selection of a reanalysis as a perfect model is quite arbitrary, but we are performing this comparison for the sake of showing the ability of the Sailor diagram to evaluate spatial fields as was done in the initial design of the Taylor diagram. In panel (b), a black-and-white version of the diagram is used, to show that it can also be used without different colours if the linestyle and character used for the reference points are changed. In the black-and-white version, a centred version of the diagram is used. Since all the ellipses corresponding to the different models are plotted on top of the observational average point, the number of ellipses to be used is smaller and the diagram better reflects the directionality problems and the under- or overestimations of

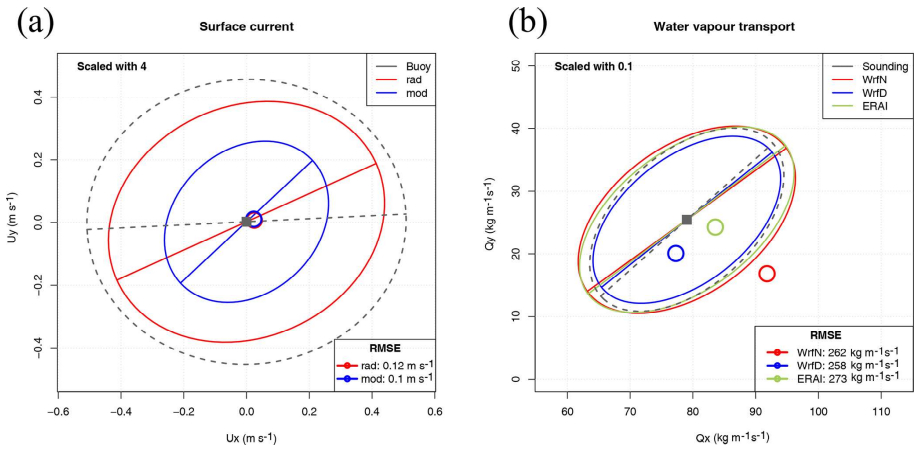

Figure 6. Sailor diagram representing the structure of errors between HF radar estimations of currents (rad) and model results (mod) with variances corresponding to EOFs scaled with a scale factor of 4 **(a)**. Sailor diagram derived from vertically integrated water vapour transport **(b)** with a scale factor of 0.1.

Table 2. Agreement of simulations by different models with observed vertically integrated water vapour transport from soundings. σ_x and σ_y represent the semi-major and semi-minor axes of the ellipses (kg m^{-1} s^{-1}). The R^2 column represents the value of the two-dimensional correlation coefficient following Crosby et al. (1993) ($R^2 = 2$ for a perfect model). The differences between the datasets described by the bias $|\overline{\mathbf{U}} - \overline{\mathbf{V}}|$ (kg m^{-1} s^{-1}) and total root mean squared error (kg m^{-1} s^{-1}) are also shown. Finally, the eccentricity of the ellipses (ε) and the congruence coefficient g_{11} of the EOF1 of every model with the one derived from observations are also shown. The congruence coefficient g_{11} represents the absolute value of the cosine of the relative rotation of model ellipses with respect to the observational one (Sect. 3.2).

	Model	σ_x	σ_y	R^2	$\lvert\overline{\mathbf{U}} - \overline{\mathbf{V}}\rvert$	RMSE	ε	g_{11}
1	OBS	183.45	107.83				0.81	
2	WRF N	195.53	118.21	1.57	15.41	261.98	0.80	0.99
3	WRF D	173.47	100.19	1.94	5.65	257.53	0.82	1.00
4	ERAI	196.99	111.18	1.92	4.69	272.94	0.83	1.00

variances with fewer lines. It is clearly shown that the reanalyses produced by the ECMWF (ERA5 and ERAI) show the closest agreement in terms of both the smallest bias and better matching of the corresponding EOFs. The other reanalyses (CFSRv2, MERRA2 and NNRA) group along the same semi-major axis, but they overestimate the variability when compared with ERA5. In terms of the bias as well, it can be seen that the lowest bias is the one corresponding to ERAI. The easiest way to arrive at a numerically precise overall diagnostic is presented in the legend, where the aggregated RMSE is shown.

4.5 Application to multimodel ensembles

In this case, we propose defining the average of all the M ensemble members of every model as the vector $\overline{\mathbf{V}}$ (Rougier, 2016). On the other side, the principal components and the associated variances and eigenvectors can be estimated from an extended data matrix $\mathbf{V}_e$ (with dimensions $NM \times 2$), which is built by joining all the realizations together in a single dataset. This means that the observational matrix $\mathbf{U}$ must also be extended to a $\mathbf{U}_e$ matrix (sized $NM \times 2$). This can be done by repeating the observations M times to produce the $\mathbf{U}_e$ dataset. This ensures that the algorithm will work because the covariance matrices involved will still be of full rank. However, it has to be considered that, in this case, the number of effective degrees of freedom (Bretherton et al., 1999) in both $\mathbf{U}_e$ and $\mathbf{V}$ datasets will not be the same. This would also be a problem for different models $\mathbf{V}_i$ and $\mathbf{V}_j$ if the number of members in their ensembles are not the same, such as in the CMIP set of runs, for instance.

As shown in Fig. 8 (panel a), prepared using a scale factor of 0.2, the Sailor diagram shows interesting features. The three models studied agree quite well in the simulation of the spatial variability of the field (the EOFs and major/minor

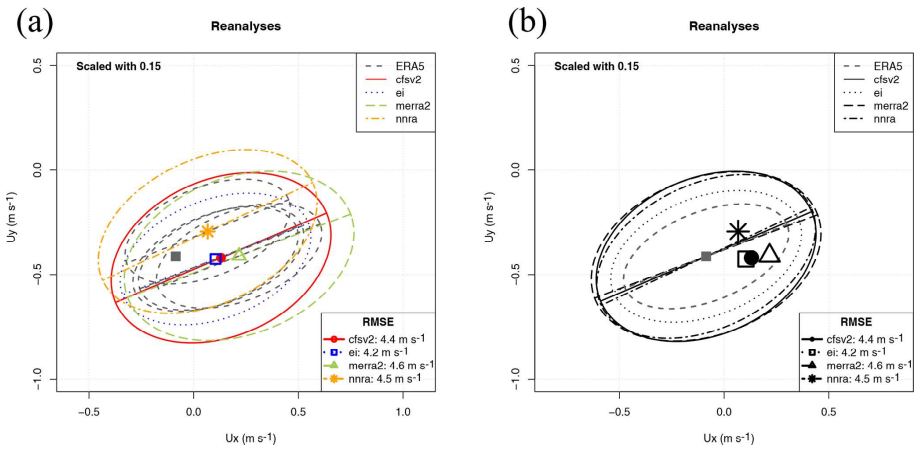

Figure 7. Sailor diagram representing the structure of errors in surface wind in January over the Northern Hemisphere for different reanalyses: uncentred version (**a**, scale factor 0.15) and centred version (**b**, scale factor 0.15).

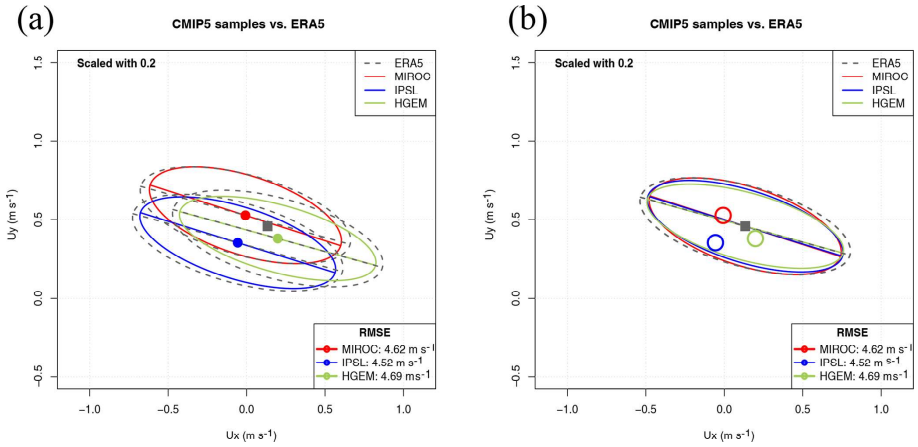

Figure 8. Sailor diagram representing the agreement between the Southern Hemisphere wind field as simulated by three models from the CMIP5 repository with ERA5 data when the reference dataset is repeated in an extended matrix. An uncentred version of the diagram **(a)** is compared with a centred version of the diagram **(b)**.

axes in the ellipse represent the spatial variability of the field). The direction of the EOFs in this case do not represent the physical direction of wind in the hemisphere but the orientation of the leading EOFs. That is, the main axis of spatial variability in the zonal and meridional directions over the hemisphere (in this case, the diagram represents a time-mean-averaged field in a T mode EOF decomposition). The analysis of the biases shows that both MIROC and IPSL models underestimate zonal average winds when compared with ERA5, whilst HadGEM2-ES shows a slightly higher zonal mean wind. This information can be obtained from the structure of the biases alone. The zonal component of the mean winds, as represented by points (square for the reference), is close to zero for MIROC and IPSL, but it is positive for ERA5 and HadGEM2-ES. Conversely, MIROC tends to overestimate the mean meridional circulation (red point

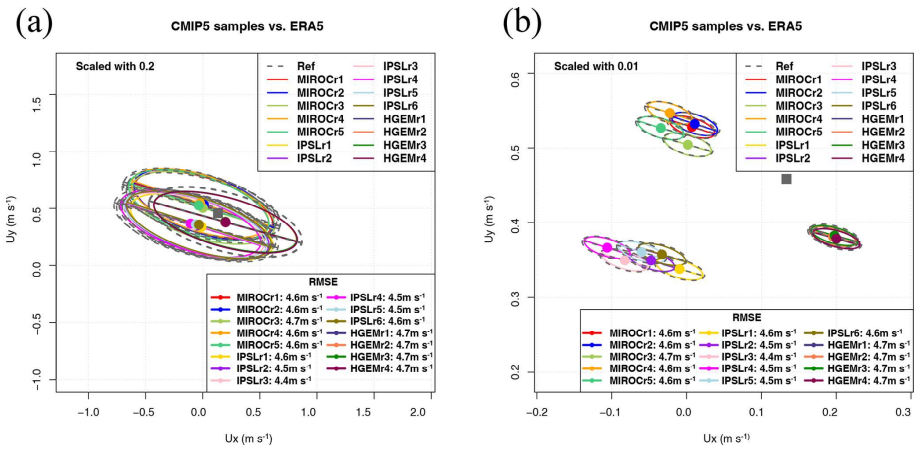

Figure 9. Sailor diagram representing the agreement between the Southern Hemisphere wind field as simulated by three models from the CMIP5 repository with ERA5 data when the reference dataset is repeated in an extended matrix **(a)** or when the individual realizations of the ensemble are taken as independent datasets **(b)**.

placed higher than ERA5), and HadGEM2-ES (green point) and IPSL (blue point) underestimate it. In order to show a clearer picture, Fig. 8 (panel b) presents a centred version of the Sailor diagram. The use of centring adds an interesting degree of freedom for the user to enhance the visibility of different aspects of the diagram such as the rotation of the EOFs. For centred diagrams, the ellipses are drawn on top of the mean hemispheric wind. Thus, only one instance of the observational ellipse is plotted. The analysis of the position of the average points leads to the same conclusions regarding the biases as before. Panel (b) in Fig. 8 shows that climate models slightly underestimate the spatial variations of the Southern Hemisphere winds (their semi-major and semi-minor axes are shorter). However, the leading EOF of the spatial variability is very close in all models, as should be expected from the horizontal structure of long-term winds (trades in tropical regions, westerlies in the extratropics). These features are properly simulated by climate models for the long-term average fields.

The second option for ensembles (same scale factor) is shown in Fig. 9 (panel a). It consists of the use of every single realization of the ensemble as a single model. This case is not of a great scientific interest, but we are presenting it in order to show the behaviour of the diagram with a high number of models (15 realizations in this case). The diagram leads to a neat comparison of the relative performance of the different members of the ensemble. This information might be interesting because of scientific reasons such as that the initialization of the members of the ensemble uses different techniques which need testing, for instance. In the case shown, the conclusion is quite clear: the averaged bias is relatively independent of the realization, and the averages corresponding to every model tend to cluster at the same position with very low biases. The inter-model variability is very low, as could be expected from long-term (27 years) time-averaged fields from climate models. Besides that, the intra-ensemble variance of properties such as the spatial variability of the field is also quite low so that the ellipses derived from different realizations in the same model almost completely overlap. Thus, in the analysis performed here, all the realizations of every model in the ensemble are very close to the reference dataset. In order to illustrate the possibility of playing with an additional degree of freedom (scale parameter), panel (b) of Fig. 9 represents the same ellipse with a very small scale factor. It can be seen that all the realizations by HadGEM2-ES are still more or less on top of each other, whilst the ellipses drawn for the realizations from other models start to separate. However, even in this extreme case, the analysis of the directionality of the leading EOF for 15 realizations is still possible since every ellipse can be compared with the reference one corresponding to observations. The legend showing the RMSE supports the conclusion that both the inter-model and intra-model variabilities are very low, as can be expected from a long-term (27 year) averaged hemispheric wind.

The final decision on the use of one approach (Fig. 8) or the other (Fig. 9) for the analysis of ensemble integrations is open to the reader since one or the other will be used by experts to answer different questions, such as whether the internal variability of the ensemble (in terms of bias and principal directions) is high or low. This might be important in some cases such as operational forecasting but not in others, such as long-term averaged spatial fields.

5 Conclusions

A new diagram for the fast evaluation of the quality of models forecasting two-dimensional vector fields or time series has been presented. As Taylor (2001) properly stated in his seminal paper, a new diagram will only be accepted by users if it helps in the fast and efficient intercomparison of model results against observational datasets. The authors of this paper developed the Sailor diagram in order to fill a gap that we detected when comparing two or more vector fields in our own work. In our previous papers in which we worked with vector fields (Ibarra-Berastegi et al., 2015, 2016), we solved this problem by duplicating the Taylor diagram once for each component. The Sailor diagram merges the same information and allows a straightforward visual comparison while rigorously providing the numeric values of the RMSE. It provides additional diagnostics which allow a complete analysis of the errors in the simulated directions too.

The authors hope that the results presented so far demonstrate that the Sailor diagram achieves that goal. First, the diagram relies on the partition of the two-dimensional RMSE in its bias and covariance parts. Those two terms are presented in the diagram separately. Thus, those two components of the error can be easily identified for the different datasets. Second, the covariance part is decomposed in terms of the corresponding principal components (empirical orthogonal functions). The structure of the covariance matrix of models and observations can also be effectively compared in the presented diagram, both in terms of the length of their semi-axes (fraction of variance) and in the relative rotation of every model against the reference dataset. This allows us to easily identify in the diagram if the models under- or overestimate the variance along any of the main axes and whether the main directions of variability in models and observations are relatively rotated or not. Thus, both two-dimensional bias and covariance can be visually identified from the diagram. Since the decomposition of the horizontal vector field is performed by means of two EOFs, there is not a loss in the variance of the observed or simulated datasets which are being compared.

The diagram might provide inaccurate estimations of the relative rotations of the principal axes of the distribution of vector components in cases in which both eigenvalues were degenerated and the eigenvectors were affected by substantial sampling uncertainty. In any case, through a diagnostic produced by the package we provide, the eccentricity of the ellipses, Eq. (13) can be used to detect this risk. Nevertheless, even if the eigenvalues were degenerated, the final classification of models is performed in terms of the RMSE, which is a measure of error that is not affected by this degeneracy.

The diagram is easily customizable in order to increase the ability to identify features of the datasets being verified by means of the use of scale factors for the ellipses (compare both panels in Fig. 5). The diagram can also be customizable by centring all of the ellipses on top of the average corresponding to the reference dataset instead of plotting all of them on top of every model being used. Thus, researchers can design a diagram that best suits their needs. In any case, if the number of models being tested is very high, many lines will appear, which will make it difficult to interpret uncentred diagrams. Thus, the option to separately use centred or uncentred diagrams and different scale factors allows us to customize the diagram to increase the ability to discriminate between similar biases (use smaller scale factors) or rotation angles (use centred diagrams). In any case, the error scores provided by our implementation (total RMSE, rotation angle, fractions of variance, R^2 and many others), as described in Sect. 3, can also be used in tabular form for a pre-screening of the multimodel dataset. Then, as a final step, only the most interesting models might be presented. Thus, the combination of centring and scaling strategies and tabular indices as described in Sect. 4 will lead to an effective verification of vector fields.

The analysis of ensembles can also be performed by means of the diagram. As shown in Sect. 4.5, the diagram can accommodate this case by using two different policies. In the first case, all the M members of the ensemble belonging to a single model can be mixed in a unique dataset, but this involves repeating the block of observations M times (Fig. 8). This implies that the analysis of the results presented in the diagram in this case must consider the different number of effective degrees of freedom very carefully, and further research should be performed to analyse the impact of this in the application of the Sailor diagram to model ensembles. However, in the second case, all the ensemble members are analysed as independent realizations of the same dataset (Fig. 9). This tends to clutter the diagram, but these results are not affected by problems related to the number of effective degrees of freedom in the different datasets used to build the diagram. The decision on the use of one or the other depends on the application intended by the user.

As a conclusion, we hope that the diagram presented here, together with an R implementation of it freely available in CRAN, will ease the verification of vector fields derived from geoscientific models in the future.

Code and data availability. The code used to prepare the figures in this paper is described in the examples of the manual of the R package `SailoR` available from CRAN https://cran.r-project.org/package=SailoR (last access: 5 July 2020). The data used to produce these figures are also distributed with the package. The version of the package used to prepare the figures in this paper can be found at https://doi.org/10.5281/zenodo.3543716 (Sáenz et al., 2019b).

Author contributions. JS conceived the idea, performed most of the mathematical analysis, and wrote some parts of the linear algebra code and most of the paper. SCM collaborated in the analysis of the matricial structure of the error and wrote substantial parts of the code, particularly the graphical representation of data. GE collab-

orated in the preparation and testing of the linear algebra part and provided data for the tests. SJGR prepared the R package distributed with the paper and its documentation. SCM, GIB and AU provided data for the package, performed exhaustive checking of the implementation and helped in the analysis of the results. All authors took active part in the writing of the paper.

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

Disclaimer. The code is made publicly available without any warranty.

Acknowledgements. The ECMWF ERA-Interim data used in this study have been obtained from the ECMWF MARS data server. The authors wish to express their gratitude to the Spanish National Ports Authority (Puertos del Estado) and the Basque Meteorological Agency (Euskalmet) for being kind enough to provide data for this study and for allowing us to make the data publicly available in the `SailoR` package. CMIP5 model output data provided by ESGF have been used for this paper. Constructive comments by three anonymous reviewers have led to a better version of the paper.

Financial support. This research has been supported by the Spanish Government's MINECO grant and ERDF (grant no. CGL2016-76561-R) and the UPV/EHU (grant no. GIU17/02).

Review statement. This paper was edited by Olivier Marti and reviewed by three anonymous referees.

References

Atlas, R., Hoffman, R. N., Ardizzone, J., Leidner, S. M., Jusem, J. C., Smith, D. K., and Gombos, D.: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications, B. Am. Meteorol. Soc., 92, 157–174, https://doi.org/10.1175/2010BAMS2946.1, 2011.

Barnett, T. P. and Preisendorfer, R.: Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis, Mon. Weather Rev., 115, 1825–1850, https://doi.org/10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2, 1987.

Boer, G. J. and Lambert, S. J.: Second order space-time climate difference statistics, Clim. Dynam., 17, 213–218, https://doi.org/10.1007/PL00013735, 2001.

Breaker, L. C., Gemmill, W. H., and Crosby, D. S.: The application of a technique for vector correlation to problems in Meteorology and Oceanography, J. Appl. Meteor., 33, 1354–1365, https://doi.org/10.1175/1520-0450(1994)033<1354:TAOATF>2.0.CO;2, 1994.

Bretherton, C. S., Smith, C., and Wallace, J. M.: An Intercomparison of Methods for Finding Coupled Patterns in Climate Data, J. Climate, 5, 541–560, https://doi.org/10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2, 1992.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field, J. Climate, 12, 1990–2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2, 1999.

Chaudhuri, A. H., Ponte, R. M., Forget, G., and Heimbach, P.: A Comparison of Atmospheric Reanalysis Surface Products over the Ocean and Implications for Uncertainties in Air–Sea Boundary Forcing, J. Climate, 26, 153–170, https://doi.org/10.1175/JCLI-D-12-00090.1, 2013.

Cheng, X., Nitsche, G., and Wallace, J. M.: Robustness of Low-Frequency Circulation Patterns Derived from EOF and Rotated EOF Analyses, J. Climate, 8, 1709–1713, https://doi.org/10.1175/1520-0442(1995)008<1709:ROLFCP>2.0.CO;2, 1995.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051–1075, https://doi.org/10.5194/gmd-4-1051-2011, 2011.

Compagnucci, R. H. and Richman, M. B.: Can principal component analysis provide atmospheric circulation or teleconnection patterns?, Int. J. Climatol., 28, 703–726, https://doi.org/10.1002/joc.1574, 2008.

Cosoli, S., Gačić, M., and Mazzoldi, A.: Variability of currents in front of the Venice Lagoon, Northern Adriatic Sea, Ann. Geophys., 26, 731–746, https://doi.org/10.5194/angeo-26-731-2008, 2008.

Cramer, E. M.: A generalization of vector correlation and its relation to canonical correlation, Multivariate Behavioural Research, 9, 347–351, https://doi.org/10.1207/s15327906mbr0903_10, 1974.

Crosby, D. S., Beaker, L. C., and Gemill, W. H.: A proposed definition for vector correlation in Geophysics: Theory and Application, J. Atmos. Ocean. Tech., 10, 355–367, https://doi.org/10.1175/1520-0426(1993)010<0355:APDFVC>2.0.CO;2, 1993.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, I., Biblot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Greer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Mong-Sanz, B. M., Morcette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dörfler, W.: Diagrammatic thinking, in: Activity and sign, Springer, 57–66, 2005.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet,

N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Clim. Dynam., 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Fernández, J., Montávez, J. P., Sáenz, J., González-Rouco, J. F., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, J. Geophys. Res., 112, D04101, https://doi.org/10.1029/2005JD006649, 2007.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R. d., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Bucha rd, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. o. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., and Díaz de Argandoña, J.: Moisture Balance Over the Iberian Peninsula According to a Regional Climate Model: The Impact of 3DVAR Data Assimilation, J. Geophys. Res.-Atmos., 123, 708–729, https://doi.org/10.1002/2017JD027511, 2018.

González-Rojí, S. J., Wilby, R. L., Sáenz, J., and Ibarra-Berastegi, G.: Harmonized evaluation of daily precipitation downscaled using SDSM and WRF+WRFDA models over the Iberian Peninsula, Clim. Dynam., 53, 1413–1433, https://doi.org/10.1007/s00382-019-04673-9, 2019.

Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, Int. J. Climatol., 27, 119–1152, https://doi.org/10.1002/joc.1499, 2007.

Hanson, B., Klink, K., Matsuura, K., Robeson, S. M., and Willmott, C. J.: Vector correlation: Review, Exposition and Geographic Application, Ann. Assoc. Am. Geogr., 82, 103–116, 1992.

Hawkins, E., Fæhn, T., and Fuglestvedt, J.: The climate spiral demonstrates the power of sharing creative ideas, B. Am. Meteorol. Soc., 100, 753–756, https://doi.org/10.1175/BAMS-D-18-0228.1, 2019.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A. J., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., and Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWP, Tech. Rep. 27, ECMWF, https://doi.org/10.21957/tkic6g3wm, 2018.

Hoffman, R. N., Leidner, S. M., Henderson, J. M., Atlas, R., Ardizzone, J. V., and Bloom, S. C.: A Two-Dimensional Variational Analysis Method for NSCAT Ambiguity Removal: Methodology, Sensitivity, and Tuning, J. Atmos. Ocean. Tech., 20, 585–605, https://doi.org/10.1175/1520-0426(2003)20<585:ATDVAM>2.0.CO;2, 2003.

Hovmöller, E.: The trough-and-ridge diagram, Tellus, 1, 62–66, https://doi.org/10.1111/j.2153-3490.1949.tb01260.x, 1949.

Ibarra-Berastegi, G., Saénz, J., Ezcurra, A., Elías, A., Diaz Argandoña, J., and Errasti, I.: Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression, Hydrol. Earth Syst. Sci., 15, 1895–1907, https://doi.org/10.5194/hess-15-1895-2011, 2011.

Ibarra-Berastegi, G., Sáenz, J., Esnaola, G., Ezcurra, A., and Ulazia, A.: Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models, Ocean Eng., 104, 530–539, https://doi.org/10.1016/j.oceaneng.2015.05.038, 2015.

Ibarra-Berastegi, G., Sáenz, J., Esnaola, G., Ezcurra, A., Ulazia, A., Rojo, N., and Gallastegui, G.: Wave Energy Forecasting at Three Coastal Buoys in the Bay of Biscay, IEEE J. Oceanic Eng., 41, 923–929, https://doi.org/10.1109/JOE.2016.2529400, 2016.

Jiménez, P. A., González-Rouco, J. F., García-Bustamante, E., Navarro, J., Montávez, J. P., de Arellano, J. V.-G., Dudhia, J., and Muñoz-Roldán, A.: Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation, J. Appl. Meteorol. Climatol., 49, 268–287, https://doi.org/10.1175/2009JAMC2175.1, 2010.

Jupp, P. E. and Mardia, K. V.: A general correlation coefficient for directional data and related regression problems, Biometrika, 67, 163–173, https://doi.org/10.2307/2335329, 1980.

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.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–470, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kundu, P. K.: Ekman veering observed near the ocean bottom, J. Phys. Ocean., 6, 238–242, https://doi.org/10.1175/1520-0485(1976)006<0238:EVONTO>2.0.CO;2, 1976.

Lee, T., Waliser, D. E., Li, J.-L. F., Landerer, F. W., and Gierach, M. M.: Evaluation of CMIP3 and CMIP5 wind stress climatology using satellite measurements and atmospheric reanalysis products, J. Climate, 26, 5810–5826, https://doi.org/10.1175/JCLI-D-12-00591.1, 2013.

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C.-E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., and Le Traon, P.-Y.: Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system, Ocean Sci., 14, 1093–1126, https://doi.org/10.5194/os-14-1093-2018, 2018.

Lorente, P., Piedracoba, S., Soto-Navarro, J., and Alvarez-Fanjul, E.: Evaluating the surface circulation in the Ebro delta (northeastern Spain) with quality-controlled high-frequency radar measurements, Ocean Sci., 11, 921–935, https://doi.org/10.5194/os-11-921-2015, 2015.

Madec, G. and the NEMO team: NEMO ocean engine, Tech. rep., Institut Pierre-Simon Laplace (IPSL), France, 2008.

McGill, R., Tukey, J. W., and Larsen, W. A.: Variations of Box Plots, Am. Stat., 32, 12–16, https://doi.org/10.2307/2683468, 1978.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H. P., Stephenson, D. B., and North, G. R.: Empirical Orthogonal Functions: The medium is the Message, J. Climate, 22, 6501–6514, https://doi.org/10.1175/2009JCLI3062.1, 2009.

North, G., Bell, T., Cahalan, R., and Moeng, F.: Sampling Errors in the Estimation of Empirical Orthogonal Functions, Mon. Weather Rev., 110, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2, 1982.

Puertos del Estado: Oceanography: Forecast, real time and climate, Spanish Government: Madrid, available at: http://www.puertos.es/en-us/oceanografia/Pages/portus.aspx (last access: 3 July 2020), 2015.

Rabanal, A., Ulazia, A., Ibarra-Berastegi, G., Sáenz, J., and Elosegui, U.: MIDAS: A Benchmarking Multi-Criteria Method for the Identification of Defective Anemometers in Wind Farms, Energies, 12, 28, https://doi.org/10.3390/en12010028, 2019.

Robert, P., Cléroux, R., and Ranger, N.: Some results on vector correlation, Comput. Stat. Data Anal., 3, 25–32, https://doi.org/10.1016/0167-9473(85)90055-6, 1985.

Roebber, P. J.: Visualizing multiple measures of Forecast Quality, Wea. Forecast., 24, 601–608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Rougier, J.: Ensemble averaging and mean squared error, J. Climate, 29, 8865–8870, https://doi.org/10.1175/JCLI-D-16-0012.1, 2016.

Rubio, A., Reverdin, G., Fontán, A., González, M., and Mader, J.: Mapping near-inertial variability in the SE Bay of Biscay from HF radar data and two offshore moored buoys, Geophys. Res. Lett., 38, L19607, https://doi.org/10.1029/2011GL048783, 2011.

Rubio, A., Fontán, A., Lazure, P., González, M., Valencia, V., Ferrer, L., Mader, J., and Hernández, C.: Seasonal to tidal variability of currents and temperature in waters of the continental slope, southeastern Bay of Biscay, J. Mar. Syst., 109–110, S121–S133, https://doi.org/10.1016/j.jmarsys.2012.01.004, 2013.

Sáenz, J., González-Rojí, S. J., Carreno-Madinabeitia, S., and Ibarra-Berastegi, G.: Analysis of atmospheric thermodynamics using the R package aiRthermo, Comput. Geosci., 122, 113–119, https://doi.org/10.1016/j.cageo.2018.10.007, 2019a.

Sáenz, J., Carreno-Madinabeitia, S., González-Rojí, S. J., Esnaola, G., Ibarra-Berastegi, G., and Ulazia, A.: SailoR: An Extension of the Taylor Diagram to Two-Dimensional Vector Data, Zenodo, https://doi.org/10.5281/zenodo.3543716, 2019b.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, J. Climate, 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Schreier, P. J.: A Unifying Discussion of Correlation Analysis for Complex Random Vectors, IEEE T. Signal Proces., 56, 1327–1336, https://doi.org/10.1109/TSP.2007.909054, 2008.

Solabarrieta, L., Rubio, A., Castanedo, S., Medina, R., Charria, G., and Hernández, C.: Surface water circulation patterns in the southeastern Bay of Biscay: New evidences from HF radar data, Cont. Shelf Res., 74, 60–76, https://doi.org/10.1016/j.csr.2013.11.022, 2014.

Stephens, M. A.: Vector correlation, Biometrika, 66, 41–48, https://doi.org/10.2307/2335240, 1979.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

The HadGEM2 Development Team: G. M. Martin, Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., Jones, C. D., McDonald, R. E., McLaren, A. J., O'Connor, F. M., Roberts, M. J., Rodriguez, J. M., Woodward, S., Best, M. J., Brooks, M. E., Brown, A. R., Butchart, N., Dearden, C., Derbyshire, S. H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J. M., Falloon, P. D., Gedney, N., Gray, L. J., Hewitt, H. T., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W. J., James, P. M., Johns, T. C., Johnson, C. E., Jones, A., Jones, C. P., Joshi, M. M., Keen, A. B., Liddicoat, S., Lock, A. P., Maidens, A. V., Manners, J. C., Milton, S. F., Rae, J. G. L., Ridley, J. K., Sellar, A., Senior, C. A., Totterdell, I. J., Verhoef, A., Vidale, P. L., and Wiltshire, A.: The HadGEM2 family of Met Office Unified Model climate configurations, Geosci. Model Dev., 4, 723–757, https://doi.org/10.5194/gmd-4-723-2011, 2011.

Ulazia, A.: Multiple roles for analogies in the genesis of fluid mechanics: How analogies can cooperate with other heuristic strategies, Found. Sci., 21, 543–565, 2016.

Ulazia, A., Sáenz, J., and Ibarra-Berastegi, G.: Sensitivity to the use of 3DVAR data assimilation in a mesoscale model for estimating offshore wind energy potential. A case study of the Iberian northern coastline, Appl. Energy, 180, 617–627, https://doi.org/10.1016/j.apenergy.2016.08.033, 2016.

Ulazia, A., Sáenz, J., Ibarra-Berastegui, G., González-Rojí, S. J., and Carreno-Madinabeitia, S.: Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean, Appl. Energy, 208, 1232–1245, https://doi.org/10.1016/j.apenergy.2017.09.030, 2017.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H., and Kimoto, M.: Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity, J. Climate, 23, 6312–6335, https://doi.org/10.1175/2010JCLI3679.1, 2010.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, Burlington, MA, 2nd edn., 2006.

Xu, Z., Hou, Z., Han, Y., and Guo, W.: A diagram for evaluating multiple aspects of model performance in simulating vector fields, Geosci. Model Dev., 9, 4365–4380, https://doi.org/10.5194/gmd-9-4365-2016, 2016.

Corrigendum to "The Sailor diagram. A new diagram for the verification of two-dimensional vector data from multiple models" published in Geoscientific Model Development, 13(7), 3221–3240, 2020.

Jon Sáenz[1, 2], **Sheila Carreno-Madinabeitia**[3, 4], **Ganix Esnaola**[5, 2], **Santos J. González-Rojí**[6, 7], **Gabriel Ibarra-Berastegi**[8, 2], **and Alain Ulazia**[9]

[1]Dept. Applied Physics II, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s./n., 48940-Leioa, Spain
[2]Joint Research Unit BEGIK, Instituto Español de Oceanografía (IEO)-Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Plentziako Itsas Estazioa, Areatza Pasealekua, 48620-Plentzia, Spain
[3]TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Álava, Albert Einstein 28, E-01510 Vitoria-Gasteiz, Spain
[4]Faculty of Pharmacy, Applied Mathematics, Statistics and O.R. Departme nt, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain
[5]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Europa Plaza 1, 20018-Donostia, Spain
[6]Oeschger Centre for Climate Change Research, University of Bern, 3010 Bern, Switzerland
[7]Climate and Environmental Physics, University of Bern, 3010 Bern, Switzerland
[8]Nuclear Engineering and Fluid Mechanics Dept., Escuela de Ingeniería de Bilbao, Plaza Ingeniero Torres Quevedo 1, 48013-Bilbao, Spain
[9]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Otaola etorbidea 29, 20600-Eibar, Spain

Correspondence: Jon Sáenz (jon.saenz@ehu.eus)

Abstract. A minor correction is needed in some of the equations of our original paper. The error in the paper affects exclusively the numerical RMSE values. The rest of the findings in the paper (ellipses, orientation of axes and structure of the diagram are not affected). The practical implementation of the software has also been corrected in the CRAN and Zenodo servers.

We made publicly available the R package SailoR together with our recently published paper (doi: 10.5194/gmd-13-3221-2020). We have been informed by one of its users that he was getting some suspicious RMSE values. We have made a thorough revision of all the equations involved both in the paper and the package and we have identified a typo in the paper and a missing N factor that propagated to the computation of the RMSE values.

There is a missing N factor in equations (10) and (11), which should be written (corrections in red):

$$\mathbf{P}_u^T \mathbf{P}_u = N\boldsymbol{\Sigma}_u^2, \tag{1}$$

$$\mathbf{P}_u^{*T} \mathbf{P}_u^* = N\mathbb{1}. \tag{2}$$

This missing N factor does not affect the majority of the results of the paper, and its main finding (the structure of the diagram) is not affected at all. However, it propagated through equations (26) to (30) into the computations of the numerical RMSE values in the paper and the SailoR package. Equation (26) must be rewritten:

$$\mathbf{D}_{uv} = N\mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + N\mathbf{E}_v\boldsymbol{\Sigma}_v^2\mathbf{E}_v^T - \left(\mathbf{E}_u\boldsymbol{\Sigma}_u\mathbf{P}_u^{*T}\mathbf{P}_v^*\boldsymbol{\Sigma}_v\mathbf{E}_v^T + \mathbf{E}_v\boldsymbol{\Sigma}_v\mathbf{P}_v^{*T}\mathbf{P}_u^*\boldsymbol{\Sigma}_u\mathbf{E}_u^T\right). \tag{3}$$

Equation (27) must be rewritten as

$$\mathbf{D}_{uv} = N\mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + N\mathbf{E}_v\boldsymbol{\Sigma}_v^2\mathbf{E}_v^T - \left(\mathbf{E}_u\mathbf{P}_u^T\mathbf{P}_v\mathbf{E}_v^T + \mathbf{E}_v\mathbf{P}_v^T\mathbf{P}_u\mathbf{E}_u^T\right). \quad (4)$$

And the same change must be applied to Eqs. (29) and (30):

$$\mathbf{D}_{uv} = N\mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + N\mathbf{R}_{vu}\mathbf{E}_u\boldsymbol{\Sigma}_v^2\mathbf{E}_u^T\mathbf{R}_{vu}^T - \left(\mathbf{E}_u\mathbf{P}_u^T\mathbf{P}_v\mathbf{E}_u^T\mathbf{R}_{vu}^T + \mathbf{R}_{vu}\mathbf{E}_u\mathbf{P}_v^T\mathbf{P}_u\mathbf{E}_u^T\right), \quad (5)$$

$$\mathbf{D}_{uv} = N\mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + N\mathbf{R}_{vu}\mathbf{E}_u\boldsymbol{\Sigma}_v^2\mathbf{E}_u^T\mathbf{R}_{vu}^T - \left(\mathbf{E}_u\boldsymbol{\Gamma}_{vu}\mathbf{E}_u^T\mathbf{R}_{vu}^T + \mathbf{R}_{vu}\mathbf{E}_u\boldsymbol{\Gamma}_{vu}^T\mathbf{E}_u^T\right). \quad (6)$$

As we stated before, this error only affects the RMSE values, and the rest of the diagnostics shown in Tables 1 and 2 are correct. As an example, we reproduce correct Tables 1 and 2 below with the values which change from the ones in the paper marked in red.

There is a typo which affected the transposes in equations (10) and (11) in the paper, but it has no consecuences, since we used the transposes correctly in all the equations (27), (28), (29) and (30). The transposes affect equation (13) in the paper, which should read (considering the N factor as well):

$$|\mathbf{P}_u^{*T}\mathbf{P}_u^*| = |\boldsymbol{\Sigma}_u^{-1}\mathbf{E}_u^T\left(\mathbf{U}-\bar{\mathbf{U}}\right)^T\left(\mathbf{U}-\bar{\mathbf{U}}\right)\mathbf{E}_u\boldsymbol{\Sigma}_u^{-1}| = N. \quad (7)$$

However, this equation was not used to plot the ellipses. Thus, the ellipses in the Figures of the paper are correct in shape, orientation and size and need no change at all. In order to show that the structure of the plots does not change with the corrections described above, we have selected Figs. 6b and 8a of the original manuscript. Figs. 1 and 2 in this corrigendum show the original version of the figures included in the published paper (left), but also their corrected counterparts (right). The only difference between them are the values of RMSE, which are marked in blue in the original version and in red in the new, corrected one. The same happens for all the figures in the paper. Thus, the main objective of the paper (the design of the diagram) is not affected by the numerical error in the RMSE values of the legends.

We have already uploaded a corrected version (1.2) of the SailoR package to the CRAN server so that future users of the software should use this version 1.2 or later. We are sorry for the inconveniences.

Author contributions. JS, SCM, GE, GIB, AU and SJGR collaborated in identifying the error and checking the new version of the package. JS, SCM and SJGR prepared the new (corrected) version of the R package. All authors took active part in the writing of this correction to the paper.

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

Disclaimer. The code is made publicly available without any warranty.

Code and data availability. The code used to prepare the figures in this paper is described as examples in the manual of the R package `SailoR`, available from CRAN https://cran.r-project.org/package=SailoR. The data used to produce these figures are distributed with the package. The package can alternatively be downloaded from Zenodo in https://doi.org/10.5281/zenodo.4046409.

Acknowledgements. This work has been funded by the Spanish Government's MINECO project CGL2016-76561-R (AEI/FEDER EU) and the University of the Basque Country (UPV/EHU funded project GIU17/02). We appreciate the feedback by SailoR package user Kwun Yip Fung, graduate student (PhD candidate) at the Department of Geological Sciences, Jackson School of Geosciences, the University of Texas at Austin.

Table 1. Original Table 1 from original manuscript with corrected values of RMSE marked in red font.

Model	σ^2	$\sum_i \sigma_i^2$	θ_u	θ_v	θ_{vu}	R^2	$\|\texttt{bias}\|$	RMSE	ε	g_{11}
Ref	47.56	47.56	1.93						0.92	
MOD1	47.56	47.56		1.93	0.00	2.00	8.34	8.34	0.92	1.00
MOD2	47.56	47.56		2.46	0.52	2.00	2.88	4.37	0.92	0.87
MOD3	47.56	47.56		-1.21	0.72	0.00	0.00	9.01	0.92	1.00
MOD4	190.24	190.24		1.93	0.00	2.00	5.56	8.45	0.92	1.00

Table 2. Original Table 2 from original manuscript with corrected values of RMSE marked in red font.

	Model	σ_x	σ_y	R^2	$\|\bar{\mathbf{U}} - \bar{\mathbf{V}}\|$	RMSE	ε	g_{11}
1	OBS	183.45	107.83				0.81	
2	WRF N	195.53	118.21	1.57	15.41	88.21	0.80	0 .99
3	WRF D	173.47	100.19	1.94	5.65	31.30	0.82	1 .00
4	ERAI	196.99	111.08	1.92	4.69	42.14	0.83	1 .00

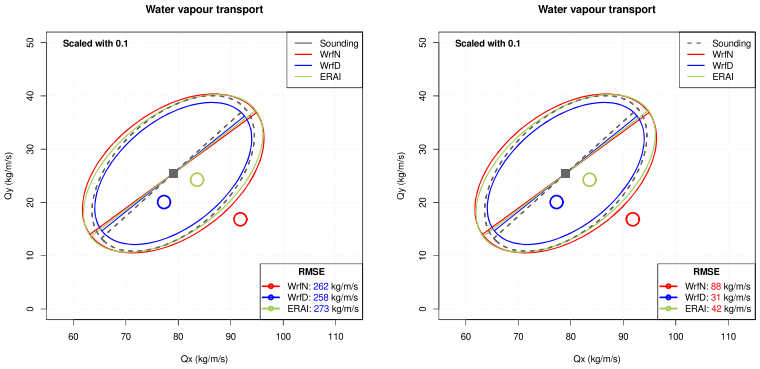

Figure 1. Fig. 6b of the published manuscript with the original RMSE values in blue font (left), and the new version of the figure with the corrected RMSE values in red font (right).

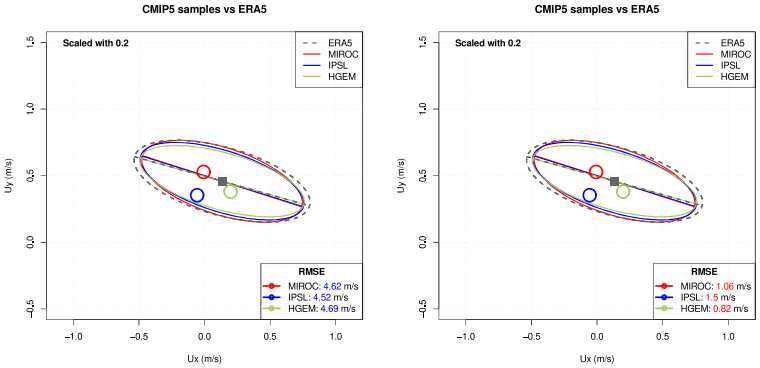

Figure 2. Fig. 8a of the published manuscript with the original RMSE values in blue font (left), and the new version of the figure with the corrected RMSE values in red font (right).

Energía eólica

Carreno-Madinabeitia S, Ibarra-Berastegi G, Sáenz J, Ulazia A. *Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900 to 2010)*. Energy 2021, 226, 120,364, https://doi.org/10.1016/j.energy.2021.120364.

Energy (2019)	
Total Cites	79807
Journal Impact Factor (IF)	6.082
5 Year IF	6.046
Immediacy Index	1.561
IF witout Journal Self Cites	4.924
Energy & Fuels	
Rank	20/112
Quantile	Q_1
Journal IF Percentile	82.589

Tabla 3.3: Índices de calidad de la revista *Energy* del año 2019

Energy 226 (2021) 120364

Contents lists available at ScienceDirect

Energy

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

Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900–2010)

Sheila Carreno-Madinabeitia [a, *], Gabriel Ibarra-Berastegi [b, e], Jon Sáenz [c, e], Alain Ulazia [d]

[a] Faculty of Pharmacy, Department of Mathematics, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain
[b] Faculty of Engineering, Energy Engineering Department, University of the Basque Country (UPV-EHU), Bilbao, Spain
[c] Faculty of Science and Technology, Department of Physics, University of the Basque Country (UPV-EHU), Leioa, Spain
[d] Faculty of Engineering, Energy Engineering Department, University of the Basque Country (UPV-EHU), Eibar, Spain
[e] Plentziako Itsas Estazioa (BEGIK), University of the Basque Country (UPV-EHU), Plentzia, Spain

ARTICLE INFO

Article history:
Received 2 September 2020
Received in revised form
20 January 2021
Accepted 9 March 2021
Available online 15 March 2021

Keywords:
Offshore wind power density
Capacity factor
Trends
ERA5
ERA-20C
20th century
Fluid mechanics

ABSTRACT

This study analysed temporal and spatial changes in offshore wind power density (WPD) and capacity factor (CF) around the Iberian Peninsula during the 20th century by analysing data from ERA20 and ERA5. Both WPD and CF were calculated using reanalysis data considering a wind turbine with a hub height of 90 m and incorporating the effect of air density changes. Since ERA5 assimilates more observations, the data from ERA20 was bias-corrected using quantile matching, with ERA5 reanalysis data as the reference. As both variables are based on wind speed (WS), this variable was also corrected and analysed. The results show that the mean values for WPD, CF, and WS during the 20th century were highest in the Atlantic zone and the Gulf of Lyon and lowest around the Balearic Islands. The results of the assessment of decadal trends using the Theil–Sen estimator show that all indicators increased significantly in the waters of the Iberian Peninsula during the study period (1900–2010). Considering the mean slope over this period, the change over the entire period could amount to 174 Wm^{-2} for WPD, 8.8% for CF, and 1.1 ms^{-1} for WS. Based on these changes, offshore wind turbines would have increased their returns by approximately 20% over the 11 decades.

1. Introduction

In recent years, there has been growing interest in renewable energy as a means to mitigate the negative effects of climate change. Liu et al. [1] demonstrated a direct association between renewable energy consumption and reductions in CO_2 emissions in multiple countries (Brazil, India, China, and South Africa).

Worldwide, wind power is the second leading source of renewable energy after hydroelectric power. These sources generate 591 and 1132 GW per year, respectively, according to the Renewables Global Status Report in 2019 [2]. This report indicates that Spain is among the five countries with the greatest installed wind power capacity worldwide. Moreover, Spain and Portugal are above the third quantile in the European Union (EU) in terms of installed wind power capacity [3]. These data are in line with those published by the International Renewable Energy Agency [4], which are summarised in Table 1. It also shows how the installed wind power capacity is expected to increase by 2030 according to the EU Directorate-General for Energy [5] and National Renewable Energy Action Plans of Spain [6] and Portugal [7].

The EU is committed to the development of offshore wind power in the long term. It is the region with the highest floating wind energy capacity, approximately 77% (Table 1) of the global total. The EU Directorate-General for Energy forecasts that this technology will contribute the most to the increase in renewable power proposed for the near future. This is in line with the offshore wind strategy in the European Green Deal [8], marking the first steps towards a more sustainable economy in the EU. In Europe, offshore wind power is focused in the North Sea, supported through the North Seas Energy Cooperation [9]. Recently, however, attention has been focused on the Mediterranean Sea and Atlantic Coast.

The most widely used indicators for measuring wind energy potential are wind power density (WPD) [10] and capacity factor (CF) [11], which, in the literature, are customarily derived from wind speed (WS). The former is used for wind resource assessments of an area under study, while the CF for a given wind turbine

* Corresponding author.
E-mail address: Sheila.Carreno@ehu.eus (S. Carreno-Madinabeitia).

https://doi.org/10.1016/j.energy.2021.120364
0360-5442/

Table 1
Present and future installed wind power capacity (GW).

	2019		2020	2030
	Onshore GW (%)	Offshore GW (%)	GW	GW
World	594.4 (100)	20.31 (100)		
Europe	173.95 (29.26)	21.83 (77.12)	210	350
Spain	25.55 (4.30)	0.05 (0.02)	28.03	50.33
Portugal	5.23 (0.88)	0.08 (0.03)	5.4	9.2

at a specific location is an indicator of actual power generation with respect to the nominal power. According to a European study [12], the mean CF for European wind turbines assessed in 2018 was 24%, rising to 37% when considering only offshore wind power. Notably, the Hywind Scotland pilot park, an offshore floating wind farm, has achieved a CF of 53.8% [13]. Both WPD and CF should be calculated by combining site-specific air density and wind speed data. However, few studies have considered the impact of changes in air density [14,15] on WPD. Considering the 1950–2010 period in the Northern Hemisphere [16], oscillations of approximately 6% in WPD can be attributed to air density changes alone. Specifically, for the Iberian Peninsula (IP), during the 2009–2014 period, the variation may reach 8% in the case of WPD and 1% for CF [17].

In recent years, numerous studies have been published on offshore wind power, and their main objective has been to analyse the wind power potential of many sites across the world that are markedly different from each other. For example, for Africa [18], Blended Sea Winds (BSW) [19] datasets from satellite observations over 11 years were used to analyse wind resources and CFs in two scenarios, one near the coast and the other in deep waters. In Colombia, WS and WPD were investigated for the 1979–2015 period [20] using North American Regional Reanalysis data [21]. WPD analysis has also been carried out at specific sites in Kuwait [22]. In the northeast of Scotland [15], estimates of WPD and CF have been obtained using ERA5 [23] reanalysis data for the 2008–2017 period. Several studies have also been conducted in the waters around the IP. The WPD in the Mediterranean Sea was analysed [24] based on BSW datasets from the National Oceanic and Atmospheric Administration (NOAA) for a 20-year period (1995–2014). Salvação et al. [25] used the 6-hourly output of the Weather Research and Forecasting (WRF) model [26] to calculate WPD and CF offshore the IP during 2004–2013. All of these studies provide an average or static picture of the current WPD and CF, using at most 37 [20] years of data. In these works, long-term changes and the effects of air density have not been considered.

To date, most studies analysing changes in offshore wind power have used WPD based on WS data to make future projections. These studies typically focus on analysing the impact of climate change, with an underlying interest in economic analysis, as future changes in the available resources will also require changes in the management of wind farms already in operation. Specifically, Zheng et al. [27] used observations obtained between 1980 and 1999 as a reference to analyse the future projections of worldwide WPD from the Coupled Model Intercomparison Project 5 (CMIP5) [28], considering the low and medium climate change scenarios (Representative Concentration Pathways [RCPs] 2.6 and 4.5). Costoya et al. [29] performed a similar analysis for the western part of the IP, considering the years 2025–2100, with the worst-case scenario (RCP 8.5). Another study carried out in the Mediterranean Sea [30] analysed an intermediate scenario (RCP 4.5) considering both 1976–2005 and 2021–2050 periods.

To demonstrate the capability of climate projections combined with operational tools for offshore wind farm planning, the European Centre for Medium-Range Weather Forecasts (ECMWF) undertook a project that consisted of simulations of seven wind farms in Northern Europe using the projections of the HIRHAM5/EC-EARTH model with RCPs 4.5 and 8.5 as input [31]. This was the first study to explore the effect of climate change on the operation and maintenance of floating wind farms. The results obtained [32] concerning the energy generated suggested a 3% reduction (RCP 8.5) and a 2% reduction in mean WS (RCP 4.5 and 8.5). As this study only considered one model and two climate scenarios, there is a need for further research in this field.

To the best of the author's knowledge, no studies have investigated long-term changes in WPD or CF during the 20th century. However, analyses were carried out for other variables, such as wave energy [33,34], daily precipitation [35], WS [36] and ocean WS and wave height [37,38].

Previous studies have generally estimated WPD and CF using reanalysis data. ERA5 reanalysis has been widely used to analyse wind in specific regions. Jourdier [39], using ERA5 and other reanalyses such as NASA's MERRA2 and COSMO-REA6 from the German Weather Service (DWD), concluded that ERA5 provides good results with a low bias for WS in the northern half of France. In Brazil, De Assis Tavares et al. [40] validated WS from the CFSv2, ERA5, and MERRA2 reanalyses at five buoys, concluding that ERA5 yielded the best results. With the same objective and results, Olauson [41] compared ERA5 with MERRA2 for the specific case of wind power for 1051 wind turbines in five countries: Germany, Denmark, France, Sweden, and the USA (Bonneville). One of the conclusions of this work was that ERA5 data can be used to accurately estimate the wind power potential. This database was also used to calculate WPD and CF in another region close to the IP, namely, the northeast of Scotland [15].

According to Wohland et al. [42], in long-term assessments, it is necessary to complement the current reanalysis ERA5, with long-term reanalysis datasets such as ERA-20C [37] or 20CR [36,43]. These long-term reanalyses (ERA-20C and 20CR) provide estimations of the state of the atmosphere through longer periods, albeit with a coarser spatial resolution and less assimilated observations than reanalyses such as ERA5. ERA20-C has been selected because it is widely used in studies that analyse trends over the 20th century [33–35].

Nonetheless, these long-term reanalyses require careful recalibration [44–47] with current reanalysis datasets (e.g. ERA5), which cover a shorter timespan (approximately 40 years). These modern reanalyses are better constrained by observations as they assimilate other types of data that are not used by the assimilation model in ERA-20C, such as satellite radiance data, vertical soundings, and surface observations. For this reason, to extend the study back to 1900, it is necessary to calibrate 20th century data by combining information from the two above-mentioned reanalyses in their overlapping period.

The objectives of this study are as follows:

1. To analyse long-term changes in offshore WPD and CF over 111 years (1900–2010) in the waters around the IP.
2. To characterise both the spatial structure and the low-frequency time variations of three variables (WPD, CF, and WS).
3. To classify the area of study into subareas according to its wind power potential as defined by the average WPD.

For all three objectives, the effect of air density oscillations needed to be incorporated into the calculations. As a reference, the National Renewable Energy Laboratory's (NREL) 5-MW turbine [48], FAST 5 MW, with 126 m diameter and 90 m hub height was selected.

The paper is divided as follows: Section 2 presents the datasets and the methods used in this study. Section 3 describes the results,

Section 4 discusses the results, and Section 5 concludes this study and outlines areas for future work.

2. Data and methods

2.1. Data

In this study, offshore WPD, CF, and WS around the IP were analysed, specifically for the region from 15° W to 10° E and 32.5°—to 50 °N, as shown in Fig. 1. To calculate and validate the results of WPD, CF, and WS over the 20th century, two sources of data were used: the ECMWF 20th century reanalysis, ERA-20C (ERA20 hereafter), and ERA5 reanalysis.

ERA20 is based on data from 1900 to 2010 and has a spatial resolution of 1.5° × 1.5°. The variables used for calculating both WPD and CF are as follows:

- 10 m zonal wind component (U_{10})
- 10 m meridional wind component (V_{10})
- 100 m zonal wind component (U_{100})
- 100 m meridional wind component (V_{100})
- 2 m temperature (T_2)
- 2 m dewpoint temperature (d_2)
- Atmospheric sea level pressure (mslp)

The temporal resolution of these variables ranges between 3 and 6 h, and for this study, the resolution was taken as 6 h in all cases, as only 6-hourly data are available for temperature.

To combine ERA20 data with ERA5, ERA5 data corresponding to the same set of variables were downloaded at the same spatial and temporal resolution. For this study, an ERA5 land—sea mask was used to identify the grid points over the sea (Fig. 1).

2.2. Methods

2.2.1. Calculation of WS, WPD, and CF

In this study, a 90-m high turbine was selected, and hence, the first step was to calculate the WS at that hub height. This is the height of the hub of the National Renewable Energy Laboratory's 5-MW offshore turbine [48] which was used as a reference. WS at 10 and 100 m were calculated using zonal and meridional components at those heights according to Equations (1) and (2).

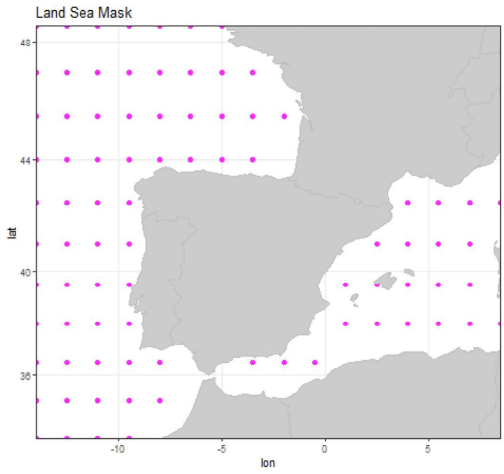

Fig. 1. Masks used by ERA5 and ERA20, the magenta dots indicating points identified as water.

$$WS_{10} = \left(U_{10}^2 + V_{10}^2\right)^{1/2} \tag{1}$$

$$WS_{100} = \left(U_{100}^2 + V_{100}^2\right)^{1/2} \tag{2}$$

As WS changes logarithmically with height, sea surface roughness z_0 was calculated from WS at 10 and 100 m, for each grid point at 6 h intervals according to Equation (3). Then, WS at 90 m was calculated by applying Equation (4).

$$\frac{WS_{100}}{WS_{10}(10)} = \frac{\log\left(\frac{100}{z_0}\right)}{\log\left(\frac{10}{z_0}\right)} \tag{3}$$

$$WS_{90}(90) = WS_{10}(10)\frac{\log\left(\frac{90}{z_0}\right)}{\log\left(\frac{10}{z_0}\right)} \tag{4}$$

The second variable of interest in this study was WPD, an indicator of the wind energy available at a given site and height. Specifically, it is the power available per square meter of area normal to the incoming wind (Equation (5)). This calculation includes both WS at 90 m and air density (ρ) [10].

$$WPD = \frac{1}{2}\rho(WS_{90})^3 \tag{5}$$

The other variable of interest, CF, indicates the energy produced by turbines with respect to what would be produced if turbines were operating at their rated power (P_R). CF is widely used to measure the average load of power plants such as wind farms [12,49,50] and solar photovoltaic arrays [51,52]. It can be used to compare different plants and technologies [53]. To calculate the CF of a commercial turbine, first, WS was normalised (WS_n). Using air density (Equation (6)), the CF was calculated using Equation (7) [11]. In this estimation, many reference turbines with rated power of approximately 1 and 2 MW [11] have been used. The authors have already demonstrated the accuracy of Equation (7) for turbines that are larger (with longer and wider blades) and taller [15,54,55], such as the aforementioned NREL 90-m 5 MW turbine. In the previous articles, the CF relative error for a 5 MW turbine was less than 5% if the CF was calculated using the power curve procedure instead of calculating it with Equation (7). Additionally, the remainder terms in the equation that accounts for the turbine characteristics (P_R and D) are not relevant when the absolute difference in CF as a percentage has been calculated. Only the hub height makes a difference.

$$WS_n = WS\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{3}} \tag{6}$$

$$CF = 0.087\ WS_n - \frac{P_R}{D^2} \tag{7}$$

where P_R is the rated power of the turbine in kW and D is the diameter in meters.

The accurate calculation of the WPD and normalised WS involves the previously calculated air density. As mentioned above,

although air density is customarily considered to be constant in most studies, $\rho_0 = 1.225$ kg/m^3, it can be accurately calculated if pressure, temperature, and humidity information are known. As ERA5 and ERA20 data include atmospheric pressure at sea level (mslp) and 2 m and dewpoint temperatures (T_2 and d_2), the air density at each grid point was calculated to obtain better results for WPD and CF. Details on performing these calculations using T_2, d_2, and mslp with the R package *aiRthermo* [56] can be found in the work of Ibarra-Berastegi et al. [14].

2.2.2. Adjustment of reanalysis data

The 20th century ERA20 reanalysis assimilates fewer observations than more recent reanalyses, such as ERA5. Therefore, the adjustment of ERA20C data using ERA5 aims to transfer the additional information available in ERA5 to ERA20. For the calibration process, the quantile matching technique was used, which involves calculating the percentiles of ERA20 data to be corrected, and the percentiles of ERA5 data for the common years in which data are available from both reanalyses (1979–2010). In this way, the relatively more recent ERA5 reanalysis that incorporates more detailed information can be used to correct biases in the ERA20 reanalysis, which covers a longer period, thus obtaining an adjusted ERA20 reanalysis (ERA20adj).

This approach is a widely used method for bias correction with the aim of achieving various goals, such as the adjustment and homogenisation of climate data series [57,58], or as in this study, the correction of biases in 20th century reanalysis using more recent reanalysis for a shorter period (in this case, 32 years). This methodology has been previously applied for different target variables such as WS [36] and wave energy resources [33,34]. For this purpose, Gudmundsson et al. [59] indicate that non-parametric transformations perform better than parametric transformations or those based directly on the distribution. Among possible non-parametric techniques, quantile matching seems to outperform statistical smoothing spline methods. The success of the non-parametric approach is likely due to its flexibility and as it does not depend on a preset function. The aforementioned authors also indicate that non-parametric methods are, on average, better able to reduce systematic errors, even in the case of the highest percentiles (the extremes).

In this study, the variables of interest are WPD, CF, and WS, which were calculated from the information available in the reanalysis. Considering that the relationships of WPD and CF with air density and WS are nonlinear, the best approach to reduce the final errors was to directly calibrate WPD and CF.

2.2.3. Calculation of mean data and decadal changes

First, mean values at each grid point based on 6-hourly WPD, CF, and WS data from the whole period were computed. Maps with three variables were produced for different seasons, considering December to February to be winter; March to May, spring; June to August, summer; and September to November, autumn. The spatial mean of the variables in the map were also shown on all maps over land.

To calculate long-term changes, WPD, CF, and WS monthly data were used. To reduce the interference of seasonal variations in the variables, anomalies were identified for the whole period (1900–2010). A 30 year reference period (1981–2010) was used to compute the seasonal cycle, and the anomalies were identified by subtracting the seasonal cycle from the original data at each grid point.

Box plots were used to graphically represent the variability of WPD, CF, and WS over time. To build the box plot for each decade, the corresponding monthly anomalies of each grid point were used.

Table 2
WPD classifications categories at 90 m.

Categories	$WPD_{90\ m}$ (Wm^{-2})	$WS_{90\ m}$ (ms^{-1})
1. High	>1050	>9.8
2. Medium-high	(640, 1050]	(8.4, 9.8]
3. Medium	(440, 640]	(7.4, 8.4]
4. Medium-low	(240, 440]	(5.9, 7.4]
5. Low	≤240	≤5.9

Together with the box plots, the trends from the decadal median were calculated using the Theil–Sen estimator [60,61] from monthly anomalies. This method is much more robust than a simple linear regression, as it calculates the mean of the slopes between all possible pairs of points. For this reason, it can reduce the influence of extreme values and allows the calculation of 95% confidence intervals.

2.2.4. Classification of areas by WPD values

According to classical ranges used in wind energy, WPD classification maps were generated according to the usual division into five major wind energy potential categories [62]. To that end, the WPD at 90 m height was calculated (Table 2) for all grid points in the studied area. The procedure to perform the classification was as follows: first, the annual mean values of WPD and WS at each grid point were calculated; next, considering Table 2, the categories were selected; and finally, the mean category for the entire period was calculated.

3. Results

3.1. Calibration of ERA20adj data

The panels in Fig. 2 show the results for each reanalysis for the overlapping period (1979–2010) in the three datasets (ERA20, ERA5, and ERA20adj).

The differences in WPD, CF, and WS between ERA5 and ERA20 were as large as 648 Wm^{-2}, 0.24, and 2.8 ms^{-1}, respectively (not shown). The greatest differences between these datasets are found at the points closest to the coast and in the Mediterranean Sea. As expected, because of adjusting ERA20 data using ERA5, the resulting dataset (ERA20adj) shows the same patterns of high WPD, CFs, and WSs as in ERA5, particularly in the Mediterranean Sea and in the northwest of the study area.

3.2. Spatial means

Considering the WPD and CF mean values in ERA20adj data for the entire period 1900–2010 (Fig. 3), interesting zones with high WPD values (above 600 Wm^{-2}) can be observed. In particular, in the Atlantic, including around A Coruña and along the French coast, as well as the Gulf of Lyon in the Mediterranean Sea. The region around Cabo de Gata should also be highlighted, as the mean WPD there is as high as 590 Wm^{-2}. Note that the areas with the highest WPDs also have the highest CFs (WSs), with values above 0.4 (8 ms^{-1}) being found in the northwest of the study area. Around the Gulf of Lyon, values between 0.35 and 0.48 (between 7.1 and 9.1 ms^{-1}) have been estimated, and 0.37 (7.8 ms^{-1}) was estimated around Cabo de Gata (Fig. 3). Based on the overall mean values of WPD and WS for the whole period, a classification of grid points was carried out by wind energy potential (Table 2), thus identifying five major regions in the study area. The most energetic areas are located NW of the Iberian Peninsula and in the Gulf of Lyon (Figs. 3 and 4), while the WPD tends to decrease near the coast of the IP.

The behaviour of the study variables WPD, CF, and WS differed

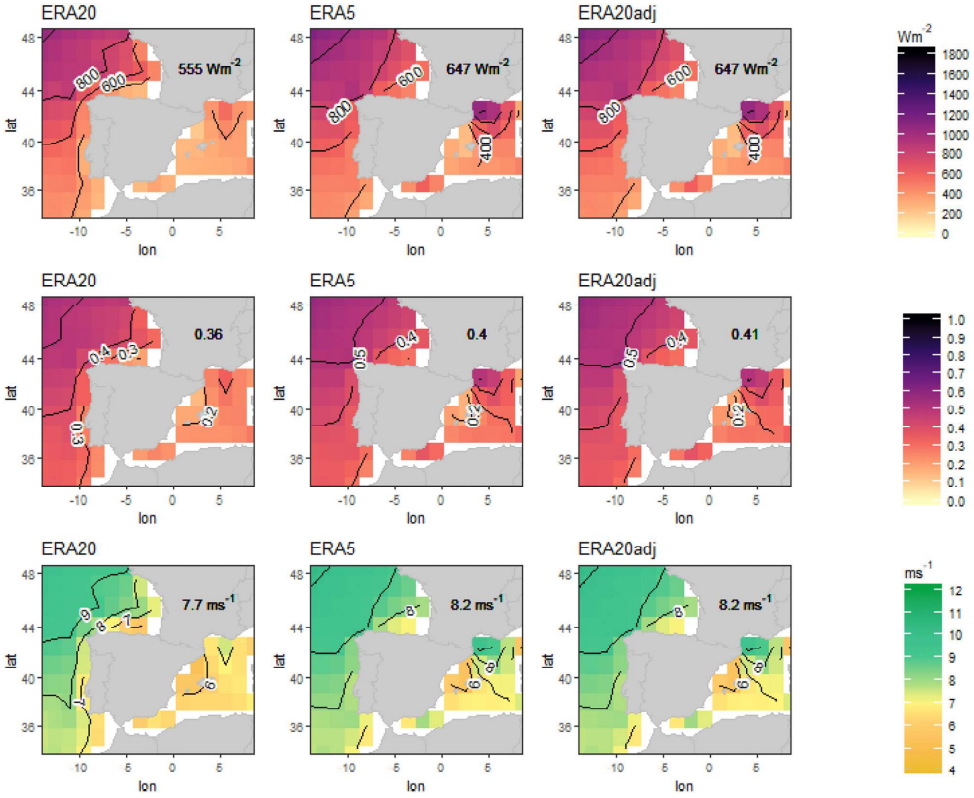

Fig. 2. Mean values of WPD (top), CF (middle), and WS (bottom) for ERA20, ERA5, and ERA20adj in the period 1979–2010, the numerical values written over land correspond to the areal mean WPD and CF.

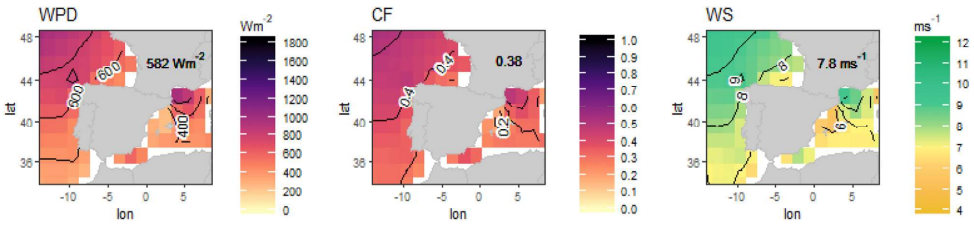

Fig. 3. Mean WPD (left), CF (middle), and WS (right) from ERA20adj in the period 1900–2010, 90 m above the sea surface.

depending on the season (Fig. 5). The values of both WPD and CF are above the mean in winter and below it in summer across the entire study area. In the regions with the highest mean values across all periods, that is, in the NW of the IP and the Gulf of Lyon, differences between winter and summer are ± 400 Wm^{-2}, ± 0.15, and ± 1.7 ms^{-1} for WPD, CF, and WS, respectively.

3.3. Temporal and spatial trends

The decadal trends over the period 1900–2009 were also calculated and derived from the monthly data corresponding to ERA20Adj at each grid point. For graphical representation purposes, the decades have been defined as follows: 1900–1909, 1910–1919, … and 2000–2009. Thus, a total of 11 decades were considered.

The box plot in Fig. 6 was generated using the monthly

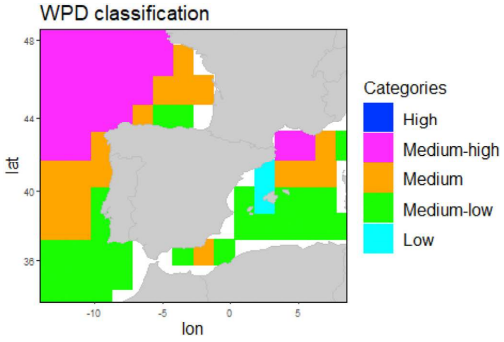

Fig. 4. Classic wind energy classification at a height of 90 m (Table 2).

anomalies. In the figure, a positive trend over the decades in the medians of WPD, CF, and WS can be observed. Specifically, WPD has a positive trend of 13.32 $Wm^{-2}decade^{-1}$ (95% confidence interval from 12.24 to 14.57 $Wm^{-2}decade^{-1}$), calculated using the Theil–Sen estimator. Moreover, as shown in Table 3, the median in each decade is higher than that in the previous one, except in 1920–1929. In the case of CF, there is a trend of 0.72% per decade (95% confidence interval from 0.65% to 0.8%). Additionally, the WS has a positive trend of 0.091 $ms^{-1}decade^{-1}$ (95% confidence interval from 0.082 to 0.1 $ms^{-1}decade^{-1}$).

To check the temporal and spatial evolution, decadal mean values of WPD, CF, and WS were calculated. Fig. 7 shows the maps of the first (1900–1909), sixth (1950–1959), and the last decade (2000–2009) of the study period. It should be noted (Fig. 7) that areas with greater intensity increase over the decades. To find out which areas have the greatest increase, the maps in Fig. 8 show the slopes of the regression lines calculated using the robust Theil–Sein estimator. Grid cells are shaded if the decadal trends are different from zero at the confidence level of 0.95.

To allow a more detailed analysis of the trends across the study area, on one hand, the slopes of the decadal trends at each grid point obtained using the Theil-Sen method [36,37] have been plotted. On the other hand, the mean values of each variable from the first (1900–1909), middle (1950–1959), and last (2000–2009) decades have been plotted.

The slope of the WPD trend is positive and significant across the

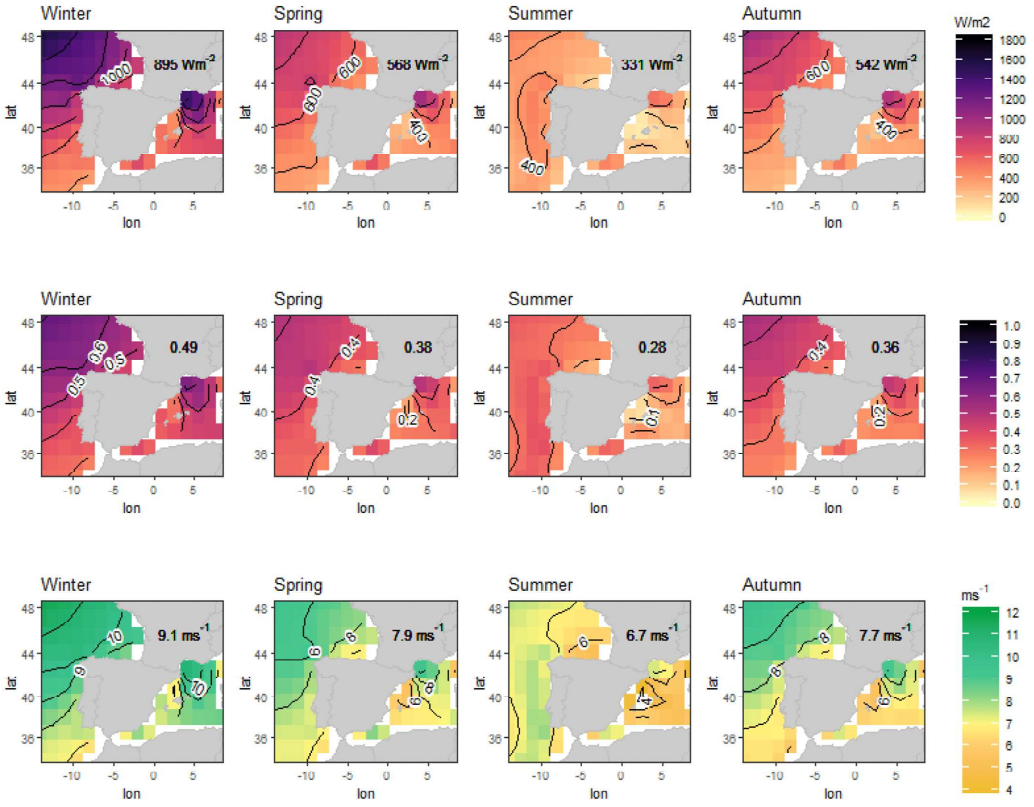

Fig. 5. Seasonal mean values of WPD, CF, and WS by season, in the top, middle, and bottom rows, respectively, from ERA20adj in the period 1900–2010, 90 m above the sea surface.

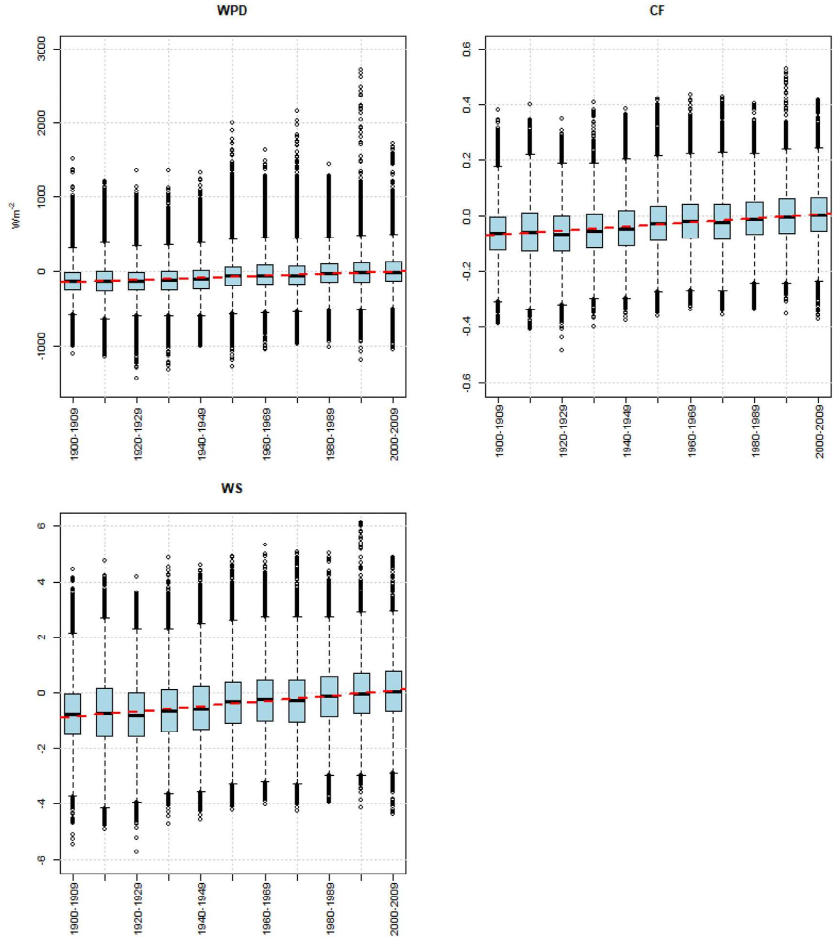

Fig. 6. Decadal box plot of monthly anomalies in WPD, CF, and WS in the period 1900–2009 from ERA20adj. The red dashed line shows the Theil–Sen estimator of the median values. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Decadal statistics for the monthly anomalies in WPD in the period 1900–2009.

(Wm^{-2})	Min	P_{25}	P_{50}	P_{75}	Max
1900–1909	−1100.12	−248.11	−132.65	−20.3	1521.68
1910–1919	−1160.29	−259.33	−121.74	−1.17	1217
1920–1929	−1437.48	−246.58	−128.05	−13.25	1353.85
1930–1939	−1327.37	−240.27	−118.38	−2.45	1355.39
1910–1949	−995.4	−231.92	−104.76	15.99	1328.39
1950–1959	−1289.49	−191.61	−62.18	59.89	1995.25
1960–1969	−1047.57	−174.57	−55.43	79.3	1632.15
1970–1979	−968.11	−176.45	−53.15	70.75	2153.63
1980–1989	−1021.41	−146.86	−33.66	93.7	1438.84
1990–1999	−1197.34	−136.85	−19.61	111.63	2718.13
2000–2009	−1042.76	−127.79	−12.17	122.75	1727.58

Statistical values indicated as: Min = minimum; P_{25} = 25th percentile; P_{50} = 50th percentile; P_{75} = 75th percentile; and max = maximum.

entire study area (Fig. 8), and is stronger in the Atlantic zone and the Gulf of Lyon, and much weaker around the Balearic Islands. The point with the steepest slope is at the north-western point of the study area (14 °W, 48.5 °N), with a slope of 27.03 $Wm^{-2}decade^{-1}$. The grid point in the Gulf of Lyon (5.5 °E, 42.5 °N) also shows a notably steep slope of 23.81 $Wm^{-2}decade^{-1}$. Further, the mean slope for the entire study area is 15.86 $Wm^{-2}decade^{-1}$, which over the 11 decades, could amount to 174 Wm^{-2}. A similar pattern is observed in the case of CF and WS, namely, a trend that is positive and significant across the study area, which is strongest in the Atlantic and weakest near the Balearic Islands. The mean slopes for CF and WS in the study area are 0.008 $decade^{-1}$ and 0.091 ms^{-1} $decade^{-1}$, which across the 11 decades would be 0.088, and 1 ms^{-1}, respectively.

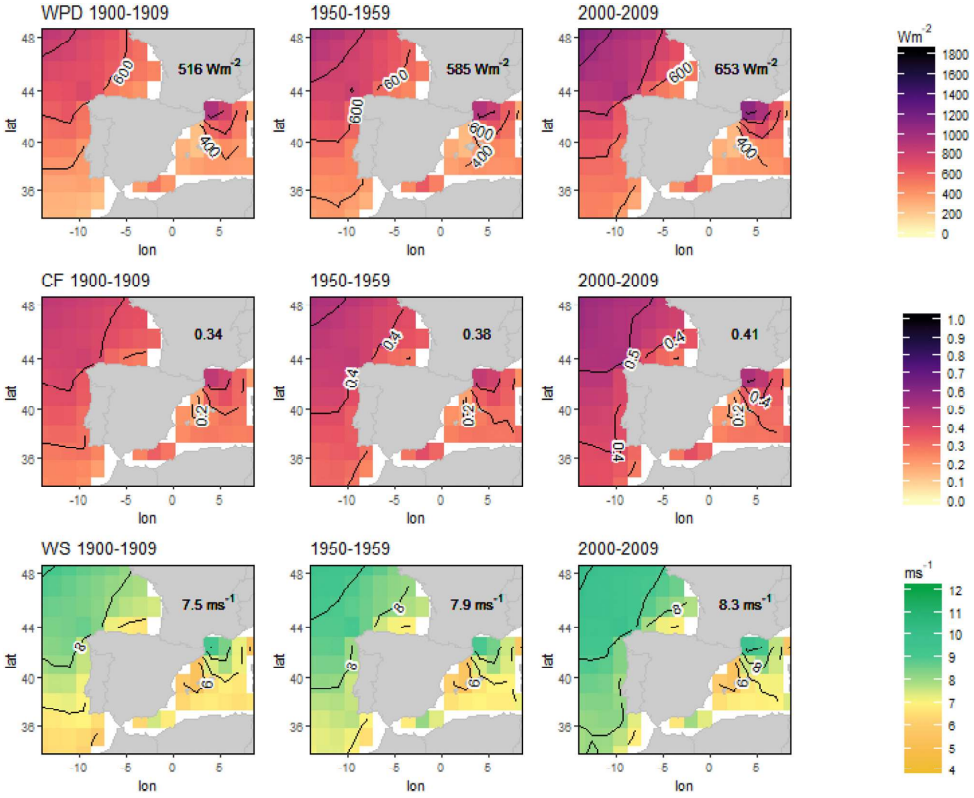

Fig. 7. Mean WPD (left), CF (middle), and WS (right) from ERA20adj for the decades 1900–1909, 1950–1959, and 2000–2009.

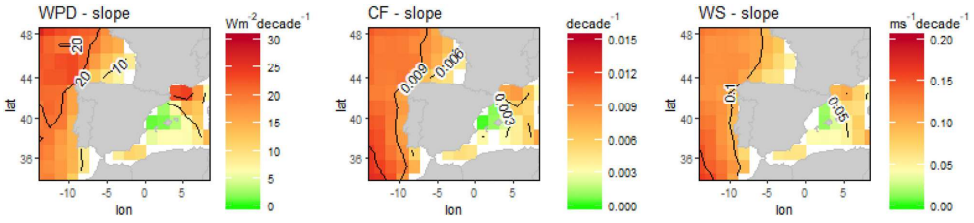

Fig. 8. Map of the decadal trends in WPD, CF, and WS for 1900–2010 obtained using monthly anomalies calculated using ERA20adj data.

4. Discussion

During the period of overlap between ERA20 and ERA5 (1979–2010), it can be seen that the ERA20 reanalysis data underestimates ERA5 for the three target variables WPD, CF, and WS (Fig. 2). Furthermore, as shown by the bias between reanalyses in Fig. 2, the greatest differences are found mainly at points close to the coast, with the largest error in the Gulf of Lyon, a highly energetic area. This is to be expected because the closest grid points to the coast show the greatest variability in topography and present small-scale atmospheric phenomena such as breezes that coarse-resolution models are not able to properly characterise [18].

The areas with the greatest wind potential are located in the northwest of the study area and the Gulf of Lyon with WPD and CF values above 600 Wm^{-2} and 0.4, respectively (Fig. 3). In addition, the area around Cabo de Gata exhibits a high potential, with values of 590 Wm^{-2} and 0.37, respectively. As shown below, these results are in line with previous studies that highlight the same areas as

having the greatest wind energy potential. Furthermore, the results of two studies [63,64] that used ERA5 data to validate the methodology correlated well with those in the present study. The first one [63] focused on Portuguese waters as simulated with the WRF model. It highlighted the north of Portugal, near the coast, as the region with the highest WPD values (909 Wm^{-2}). Using the methodology explained in subsection '2.2.1 Calculation of WS, WPD, and CF' but with the conditions of Campos et al. [63] (using ERA5 without adjusting air density and at a height of 10 m), a WPD of 884 Wm^{-2} was obtained for the same area (point 10 °W, 41 °N). This value is somewhat lower in the current study, probably due to the finer spatial resolution of WRF (0.081° × 0.097°) better representing the characteristics of areas close to the shore. Using the WRF model, Salvador et al. [64] analysed the area around Galicia and found the highest WPD values in the northwestern zone from Cape Finisterre to Cape Ortegal. Specifically, in A Coruña, the WPD reached 1000 Wm^{-2} at a height of 120 m. ERA5 yielded almost identical results, namely, 1001 Wm^{-2} under the same conditions. Another study based on WRF highlighted the potential of A Coruña in the Galician coasts [25] (from 14° to 5 °W and 35° −45° N) with a mean CF of 29–48%. Despite the great differences between the studies, a CF of 35–54% was found in this study in the same domain. Again, the results are notably similar, thus confirming the validity of this study.

Similar studies have also been carried out extensively in the Mediterranean Sea. Ulazia et al. [65] indicated that the Gulf of Lyon had notable potential, with a high CF of 0.46, compared to Menorca, with a considerably lower mean CF (0.3) and a decreasing trend towards the east Mediterranean [14]. Along similar lines, despite differences in the conditions, CFs of 0.46 for the Gulf of Lyon (5.5 °E, 42.5 °N) and 0.27 for the area around Menorca (4 °E, 39.5 °N) were obtained in this study. Based on satellite data, Sukissian et al. [24] also concluded that the Gulf of Lyon (5.5 °E, 42.5 °N) is a candidate for wind farms and highlighted the Cabo de Gata for its relatively high WPD.

To the best of the authors' knowledge, this study is the first to analyse, in depth, the trends and behaviour of WPD, CF, and WS throughout the 20th century around the IP. Analysis of the decadal median trends indicates significant positive trends during 1900–2010. This result is present for all grid points across the study area. Regarding the study variables, the mean WPD and CF values obtained were 15.86 $Wm^{-2}decade^{-1}$ and 0.008 $decade^{-1}$, respectively. Likewise, for peak values, the strongest trends were found in the Atlantic and Gulf of Lyon. In contrast, Cabo de Gata does not stand out, compared to the results for surrounding areas, as has been found previously. The results of this study also show positive trends for WS (0.091 $ms^{-1}decade^{-1}$), in line with recently published studies analysing offshore WS during the 20th century [37,38]. Meucci et al. [37] proved that ERA20 reanalyses present positive trends of WS at a height of 10 m in the Northern Hemisphere, exceeding 0.18 $ms^{-1}decade^{-1}$. However, as per Young et al. [38] 10 m WS satellite data (1985–2018) indicated a significant growth in the Atlantic North zone of 0.1 $ms^{-1}decade^{-1}$, and no clear trend was present in the Gulf of Lyon and Bay of Biscay. Compared with the results obtained in Fig. 6, growth during 1990–1999 and 2000–2009 tended to stabilize. There is no available information for 2010–2019.

Understanding the past evolution of WPD and CF during the last century can be helpful to put into context the estimations for climate change-driven future trends described in other studies. Rusu et al. [30], analysed 30 years of historical data (1976–2005) for the Mediterranean Sea, and found a slight increase in the yearly mean of approximately 8 $Wm^{-2}decade^{-1}$. In contrast, the projections for the period 2021–2050, under a moderate scenario (RCP 4.5), indicated a reduction in wind energy (approximately 12 $Wm^{-2}decade^{-1}$). The Gulf of Lyon (43.23 °N, 4.77 °E) was also identified as a suitable area for wind energy generation. Instead, Costoya et al. [29] investigated the waters off the west of the IP considering RCP 8.5 and analysed the years 2025–2100 and pointed to a reduction in WPD, except in the northwest corner of the IP where hardly any changes or a slight increase can be expected. Zheng et al. [27] concluded that as per RCPs 2.6 and 4.5, although the number of regions with high offshore wind potential (WPD > 800 Wm^{-2}) will slightly decrease worldwide, the number of zones with WPD > 400 Wm^{-2} will increase by the end of the 21st century, and the coasts of the IP lie in this range. Therefore, although previous studies have forecasted reductions in WPD, the aforementioned authors suggest that there may be more zones with wind energy levels above 400 Wm^{-2}, implying an overall increase in WPD. In relation to this, there is a need for further detailed analysis of future projections seeking to identify when and where decreases or increases in WPD can be expected.

Over the entire study period (1900–2010), the mean increase in CF, in the coastal areas around the IP, was 8.8%. In other words, the CF of a reference NREL 5 MW turbine (P_R = 5 MW and D = 126 m) would have increased from 34% (the CF mean of the first decade 1900–1909, Fig. 8) to 42.8%. Since annual energy production is proportional to CF [10], this would mean an increase in production of more than 900 MWh. For a floating turbine, such as the NREL 5 MW turbine, the increase in annual energy produced over 11 decades with this magnitude of change would be more than 4000 MWh, although the final production of a wind farm cannot be directly derived from the electricity generated by a single turbine. In addition, the studies available on this topic looking to the future are not conclusive, and it is believed that, due to climate change, WPDs could continue to increase in areas that currently have the highest values around the IP coast [43], although this should be analysed in more detail. Considering trends in the last century, the increase in WPD represents an increase in economic returns of at least 20% for each wind turbine compared with early 20th century production, since the rise from 34% to 42.8% represents in relative terms an increase of more than 20%.

This historical trend is of great interest given the expected increase in offshore wind energy in Europe, and the potential reductions in the price of electricity and CO_2 emissions that may follow in the future. The EU is committed to expanding the renewable energy industry, and the economic feasibility of any wind farm is directly dependent on the WPD. It seems clear that a long-life feasibility study requires dynamic knowledge of the WPD evolution at a given location instead of only a static value. The Spanish and Portuguese National Renewable Energy Action Plans incorporates estimations for offshore installed wind power that are expected to reach 59.53 GW in IP [6,7] by 2030. Furthermore, in recent years, there has been a 4.5 €/MWh reduction in the market price of electricity for every 12.8 TWh block installed, due to the penetration of wind energy into the energy market [66]. It has also been demonstrated that the implementation of tens of GW in a region can reduce CO_2 emissions by almost 70% [67].

Finally, it should be pointed out that the increase in CF observed based on resource availability in recent decades could translate to even greater increases in energy production with technical advancements in the aerodynamic and hydrodynamic design of floating turbines. As more technical advances in the field of deep-sea anchoring systems become feasible, more offshore wind power facilities could be built. This includes the generation of hydrogen or methanol at thousands of kilometres from the coast at far-offshore wind farms with an optimal/maximum WPD [68].

5. Conclusions and future outlook

This study shows that the calibration of ERA20 data by quantile-matching with ERA5 data is appropriate to carry out the analysis of the long-term variability of WPD, CF, and WS in the waters off IP. It is shown that during the overlapping years (1979–2010), ERA5 and ERA20adj are very similar.

Considering the values of three variables (WPD, CF, and WS) over the IP, the Atlantic zone, the Gulf of Lyon, and Cabo de Gata were identified as suitable areas to generate offshore wind energy. It has been also proved that during the 20th century and the first decade of 21st, WPD, CF, and WS increased at a rate of 15.86 $Wm^{-2}decade^{-1}$, 0.008 $decade^{-1}$, and 0.091 $ms^{-1}decade^{-1}$, respectively, which corresponds to a total increase of 174 Wm^{-2}, 0.088, and 1 ms^{-1}, respectively. In this period, the areas that have the greatest increase in WPD are the Atlantic zone and the Gulf of Lyon, while for CF, only the Atlantic zone has a significant increase.

Finally, applying these results to a reference turbine throughout the 20th century, an increase in profits of 20% was estimated. This increase could extend over time, especially considering the result of Zheng et al. [27] who considered that the number of zones with WPD > 400 Wm^{-2} will increase by the end of the 21st century.

In the near future, an ERA5 dataset from 1950 to the present is expected to become available. With the inclusion of 30 more years of current reanalysis data, more robust results could be obtained. Additionally, the complete projections of CMIP6 will soon be available, making it possible to assess potential future wind power. Further projections are needed to improve the estimates of what will happen from now until 2100.

In this study, ERA20 reanalysis data from the 20th century (herein ERA20) were used, but this analysis could also be carried out using other atmospheric reanalyses covering the 20th century, such as the NOAA's 20CR. This could be interesting, as previous studies comparing the 20th century reanalyses from NOAA with those from the ECMWF have demonstrated differences [42,69].

Credit author statement

SCM: Methodology, Software, Visualisation, Writing – original draft. GIB: Methodology, Supervision, Writing – review & editing, Project administration JS: Methodology, Supervision, Writing – review & editing, Funding acquisition AU: Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

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

Acknowledgments

This work was financially supported by the Spanish Government through the MINECO project CGL2016-76561-R (MINECO/ERDF, UE) and the University of the Basque Country (UPV/EHU, GIU 17/002). Reanalysis data were downloaded at no cost from the ECMWF. All calculations and plots were carried out in the R framework [70].

References

[1] Liu J-L, Ma C-Q, Ren Y-S, Zhao X-W. Do real output and renewable energy consumption BRICS countries. Energies 2020;13.
[2] Murdock HE, Gibb D, André T, Appavou F, Brown A, Epp B. Renewables 2019 global Status report, 8; 2019. https://doi.org/10.3390/resources8030139.
[3] Papież M, Śmiech S, Frodyma K. Factors affecting the efficiency of wind power in the European Union countries. Energy Pol 2019;132:965–77. https://doi.org/10.1016/j.enpol.2019.06.036.
[4] International renewable energy Agency. IRENA; 1981. https://www.irena.org/. [Accessed 5 June 2020].
[5] European Commission. Renowable energy Onshore and offshore wind. 2019. https://ec.europa.eu/energy/topics/renewable-energy/onshore-and-offshore-wind_en. [Accessed 5 June 2020].
[6] Gobierno de España. Borrador Actualizado del Plan Nacional Integrado de Energía y Clima 2021-2030. 2020. p. 1–8.
[7] República Portuguesa. Plano Nacional Integrado Energia e Clima 2021-2030. 2018. p. 1–106.
[8] European Commission. The European green deal. 2019. https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en. [Accessed 5 June 2020].
[9] European Commission. the North seas energy Cooperation. 2020. https://ec.europa.eu/energy/topics/infrastructure/high-level-groups/north-seas-energy-cooperation_en. [Accessed 5 June 2020].
[10] Manwell JF, McGowan JG, Rogers AL. Wind energy explained: theory, design and application. Wiley; 2010.
[11] Masters GM. Renewable and efficient electric power systems. John Wiley & Sons; 2013.
[12] Komusanac I, Fraile D, Brindley G, Walsh C, Pineda I. Wind energy in Europe in 2018. Trends Stat 2019:32.
[13] Smith AZ. Energy numbers 2020. https://energynumbers.info/uk-offshore-wind-capacity-factors. [Accessed 5 June 2020].
[14] Ibarra-Berastegi G, Ulazia A, Saénz J, González-Rojí SJ. Evaluation of Lebanon's offshore-wind-energy potential. J Mar Sci Eng 2019;7:1–13. https://doi.org/10.3390/jmse7100361.
[15] Ulazia A, Nafarrate A, Ibarra-Berastegi G, Sáenz J, Carreno-Madinabeitia S. The consequences of air density variations over northeastern Scotland for offshore wind energy potential. Energies 2019;12. https://doi.org/10.3390/en12132635.
[16] Ulazia A, Sáenz J, Ibarra-Berastegi G, González-Rojí SJ, Carreno-Madinabeitia S. Global estimations of wind energy potential considering seasonal air density changes. Energy 2019;187:115938. https://doi.org/10.1016/j.energy.2019.115938.
[17] Ulazia A, Ibarra-Berastegi G, Sáenz J, Carreno-Madinabeitia S, González-Rojí SJ. Seasonal correction of offshore wind energy potential due to air density: case of the Iberian Peninsula. Sustain Times 2019;11. https://doi.org/10.3390/su11133648.
[18] Elsner P. Continental-scale assessment of the African offshore wind energy potential: spatial analysis of an under-appreciated renewable energy resource. Renew Sustain Energy Rev 2019;104:394–407. https://doi.org/10.1016/j.rser.2019.01.034.
[19] NOAA.. Blended Sea winds 2020. https://www.ncdc.noaa.gov/data-access/marineocean-data/blended-global/blended-sea-winds. [Accessed 18 March 2020].
[20] Rueda-Bayona JG, Guzmán A, Eras JJC, Silva-Casarín R, Bastidas-Arteaga E, Horrillo-Caraballo J. Renewables energies in Colombia and the opportunity for the offshore wind technology. J Clean Prod 2019;220:529–43. https://doi.org/10.1016/j.jclepro.2019.02.174.
[21] NOAA. NCEP north American regional reanalysis: NARR 2020. https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html.
[22] Alkhalidi MA, Al-Dabbous SK, Neelamani S, Aldashti HA. Wind energy potential at coastal and offshore locations in the state of Kuwait. Renew Energy 2019;135:529–39. https://doi.org/10.1016/j.renene.2018.12.039.
[23] Hersbach H. The ERA5 atmospheric reanalysis. AGU Fall Meet Abstr 2016;2016. NG33D–01.
[24] Soukissian T, Karathanasi F, Axaopoulos P. Satellite-based offshore wind resource assessment in the Mediterranean Sea. IEEE J Ocean Eng 2017;42:73–86. https://doi.org/10.1109/JOE.2016.2565018.
[25] Salvação N, Guedes Soares C. Wind resource assessment offshore the Atlantic Iberian coast with the WRF model. Energy 2018;145:276–87. https://doi.org/10.1016/j.energy.2017.12.101.
[26] Skamarock WC, Klemp JB, Dudhi J, Gill DO, Barker DM, Duda MG, Powers JG. A Description of the Advanced Research WRF Version 3 (No. NCAR/TN-475+STR). University Corporation for Atmospheric Research 2008. https://doi.org/10.5065/D68S4MVH.
[27] Zheng C, wei, yan Li X, Luo X, Chen X, Qian Y hao, Zhang Z hua, et al. Projection of future global offshore wind energy resources using CMIP data. Atmos-Ocean 2019;57:134–48. https://doi.org/10.1080/07055900.2019.1624497.
[28] Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, et al. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. J Adv Model Earth Syst 2013;5:572–97. https://doi.org/10.1002/jame.20038.
[29] Costoya X, Rocha A, Carvalho D. Using bias-correction to improve future projections of offshore wind energy resource: a case study on the Iberian Peninsula. Appl Energy 2020;262:114562. https://doi.org/10.1016/j.apenergy.2020.114562.
[30] Rusu E, Rusu L. Evaluation of the wind power potential in the European nearshore of the Mediterranean Sea. E3S Web Conf 2019;103:1–6. https://doi.org/10.1051/e3sconf/201910301003.
[31] Consulting JBA. Performance indicators for offshore wind farms in Europe from 1977 to 2100 derived from climate projections. Copernicus Clim Chang Serv 2020. https://doi.org/10.24381/cds.42c6e4dd.
[32] Consulting JBA. Technical Guide : offshore wind farm operations and

maintenance C3S _ 422 _ Lot2 _ Deltares - European Services. 2019.
[33] Penalba M, Ulazia A, Ibarra-Berastegui G, Ringwood J, Sáenz J. Wave energy resource variation off the west coast of Ireland and its impact on realistic wave energy converters' power absorption. Appl Energy 2018;224:205–19. https://doi.org/10.1016/j.apenergy.2018.04.121.
[34] Ulazia A, Penalba M, Rabanal A, Ibarra-Berastegi G, Ringwood J, Sáenz J. Historical evolution of the wave resource and energy production off the chilean coast over the 20th Century. Energies 2018;11. https://doi.org/10.3390/en11092289.
[35] Kim DI, Kwon HH, Han D. Bias correction of daily precipitation over South Korea from the long-term reanalysis using a composite gamma-pareto distribution approach. Nord Hydrol 2019;50:1138–61. https://doi.org/10.2166/nh.2019.127.
[36] Bett PE, Thornton HE, Clark RT. Using the Twentieth Century Reanalysis to assess climate variability for the European wind industry. Theor Appl Climatol 2017;127:61–80. https://doi.org/10.1007/s00704-015-1591-y.
[37] Meucci A, Young IR, Aarnes OJ, Ø Breivik. Comparison of wind speed and wave height trends from twentieth-century models and satellite altimeters. J Clim 2020;33:611–24. https://doi.org/10.1175/JCLI-D-19-0540.1.
[38] Young IR, Ribal A. Multiplatform evaluation of global trends in wind speed and wave height. Science 2019;364:548–52. https://doi.org/10.1126/science.aav9527. 80-.
[39] Jourdier B. Evaluation of ERA5 and other reanalyses to simulate wind power production over France. EMS Annu Meet Abstr 2019;16:2019.
[40] de Assis Tavares LF, Shadman M, de Freitas Assad LP, Silva C, Landau L, Estefen SF. Assessment of the offshore wind technical potential for the Brazilian Southeast and South regions. Energy 2020;196:117097. https://doi.org/10.1016/j.energy.2020.117097.
[41] Olauson J. ERA5: the new champion of wind power modelling? Renew Energy 2018;126:322–31. https://doi.org/10.1016/j.renene.2018.03.056.
[42] Wohland J, Omrani NE, Witthaut D, Keenlyside NS. Inconsistent wind speed trends in current twentieth century reanalyses. J Geophys Res Atmos 2019;124:1931–40. https://doi.org/10.1029/2018JD030083.
[43] Wohland J, Eddine Omrani N, Keenlyside N, Witthaut D. Significant multi-decadal variability in German wind energy generation. Wind Energy Sci 2019;4:515–26. https://doi.org/10.5194/wes-4-515-2019.
[44] Trenberth KE, Olson JG. An evaluation and Intercomparison of global analyses from the national meteorological center and the European Centre for medium range weather forecasts. Bull Am Meteorol Soc 1988;69:1047–57. https://doi.org/10.1175/1520-0477(1988)069<1047:AEAIOG>2.0.CO;2.
[45] Simmons AJ, Jones PD, da Costa Bechtold V, Beljaars ACM, Källberg PW, Saarinen S, et al. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. J Geophys Res Atmos 2004;109:1–18. https://doi.org/10.1029/2004JD005306.
[46] Sterl A. On the (in)homogeneity of reanalysis products. J Clim 2004;17:3866–73. https://doi.org/10.1175/1520-0442(2004)017<3866:OTIORP>2.0.CO. 2.
[47] Bloomfield HC, Shaffrey LC, Hodges KI, Vidale PL. A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern Hemisphere storminess in the ERA20C reanalysis. Environ Res Lett 2018;13. https://doi.org/10.1088/1748-9326/aad5c5.
[48] Jonkman J, Butterfield S, Musial W, Scott G. Definition of a 5-MW reference wind turbine for offshore system development. 2009.
[49] Barthelmie RJ, Jensen LE. Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm. Wind Energy 2010;13:573–86. https://doi.org/10.1002/we.408.
[50] Boccard N. Capacity factor of wind power realized values vs. estimates. Energy Pol 2009;37:2679–88. https://doi.org/10.1016/j.enpol.2009.02.046.
[51] Izquierdo S, Montañs C, Dopazo C, Fueyo N. Analysis of CSP plants for the definition of energy policies: the influence on electricity cost of solar multiples, capacity factors and energy storage. Energy Pol 2010;38:6215–21. https://doi.org/10.1016/j.enpol.2010.06.009.
[52] Wang Z, Li Y, Wang K, Huang Z. Environment-adjusted operational performance evaluation of solar photovoltaic power plants: a three stage efficiency analysis. Renew Sustain Energy Rev 2017;76:1153–62. https://doi.org/10.1016/j.rser.2017.03.119.
[53] Cheng HP, Yu MT. Effect of the transmission configuration of wind farms on their capacity factors. Energy Convers Manag 2013;66:326–35. https://doi.org/10.1016/j.enconman.2012.12.011.
[54] Ulazia A, Ibarra-Berastegi G, Sáenz J, Carreno-Madinabeitia S, González-Rojí SJ. Seasonal correction of offshore wind energy potential due to air density: case of the Iberian Peninsula. Sustain Times 2019;11. https://doi.org/10.3390/su11133648. 2014–6.
[55] Ulazia A, Saenz J, Ibarra-Berastegui G. Sensitivity to the use of 3DVAR data assimilation in a mesoscale model for estimating offshore wind energy potential. A case study of the Iberian northern coastline. Appl Energy 2016;180:617–27. https://doi.org/10.1016/j.apenergy.2016.08.033.
[56] Sáenz J, González-Rojí SJ, Carreno-Madinabeitia S, Ibarra-Berastegi G. Analysis of atmospheric thermodynamics using the R package aiRthermo. Comput Geosci 2019;122. https://doi.org/10.1016/j.cageo.2018.10.007.
[57] Yang S, Wang XL, Wild M. Homogenization and trend analysis of the 1958-2016 in situ surface solar radiation records in China. J Clim 2018;31:4529–41. https://doi.org/10.1175/JCLI-D-17-0891.1.
[58] Squintu AA, van der Schrier G, Brugnara Y, Klein Tank A. Homogenization of daily temperature series in the European climate assessment & dataset. Int J Climatol 2019;39:1243–61. https://doi.org/10.1002/joc.5874.
[59] Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T. Technical Note: downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. Hydrol Earth Syst Sci 2012;16:3383–90. https://doi.org/10.5194/hess-16-3383-2012.
[60] Theil H. A rank-invariant method of linear and polynomial regression analysis. In: Proceedings of koninalijke nederlandse akademie van Weinenschatpen; 1950. https://doi.org/10.1007/978-94-011-2546-8 20.
[61] Sen PK. Estimates of the regression coefficient based on Kendall's tau. J Am Stat Assoc 1968;63:1379–89. https://doi.org/10.1080/01621459.1968.10480934.
[62] Troen I, Lundtang Petersen E. European wind atlas - Errata. 1989.
[63] Campos RM, Guedes Soares C. Spatial distribution of offshore wind statistics on the coast of Portugal using Regional Frequency Analysis. Renew Energy 2018;123:806–16. https://doi.org/10.1016/j.renene.2018.02.051.
[64] Salvador S, Costoya X, Sanz-Larruga FJ, Gimeno L. Development of offshore wind power: contrasting optimal wind sites with legal restrictions in Galicia, Spain. Energies 2018;11:2011–20. https://doi.org/10.3390/en11040731.
[65] Ulazia A, Sáenz J, Ibarra-Berastegi G, González-Rojí SJ, Carreno-Madinabeitia S. Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean. Appl Energy 2017;208:1232–45. https://doi.org/10.1016/j.apenergy.2017.09.030.
[66] Pereira AJC, Saraiva JT. Long term impact of wind power generation in the Iberian day-ahead electricity market price. Energy 2013;55:1159–71. https://doi.org/10.1016/j.energy.2013.04.018.
[67] Kempton W, Archer CL, Dhanju A, Garvine RW, Jacobson MZ. Large CO2 reductions via offshore wind power matched to inherent storage in energy end-uses. Geophys Res Lett 2007;34. https://doi.org/10.1029/2006GL028016.
[68] Babarit A, Clodic G, Delvoye S, Gilloteaux JC. Exploitation of the far-offshore wind energy resource by fleets of energy ships - Part 1: energy ship design and performance. Wind Energy Sci 2020;5:839–53. https://doi.org/10.5194/wes-5-839-2020.
[69] Befort DJ, Wild S, Kruschke T, Ulbrich U, Leckebusch GC. Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. Atmos Sci Lett 2016;17:586–95. https://doi.org/10.1002/asl.694.
[70] R Core Team. R: A Language and Environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/2020.

4 SUPLEMENTO: Trabajo en revisión

Análisis de tormentas

González-Rojí SJ, **Carreno-Madinabeitia S**, Sáenz J, Ibarra-Berastegi G. *Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation*. Hydrology and Earth System Sciences.

Hydrology and Earth System Sciences (2019)	
Total Cites	19937
Journal Impact Factor (IF)	5153
5 Year IF	5460
Immediacy Index	1.004
IF witout Journal Self Cites	4.695
Geosciences, multidisciplinary	
Rank	10/200
Quantile	Q_1
Journal IF Percentile	95.250

Tabla 4.1: Índices de calidad de la revista *Hydrology and Earth System Sciences* del año 2019

Changes in the simulation of atmospheric instability over the Iberian Peninsula due to the use of 3DVAR data assimilation

Santos J. González-Rojí[1,2], Sheila Carreno-Madinabeitia[3,4], Jon Sáenz[5,6], and Gabriel Ibarra-Berastegi[7,6]

[1]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.
[2]Climate and Environmental Physics, University of Bern, Bern, Switzerland.
[3]Department of Applied Mathematics, Statistics and O.R., University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain.
[4]TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Álava, Vitoria-Gasteiz, Spain.
[5]Department of Applied Physics II, University of the Basque Country (UPV/EHU), Leioa, Spain.
[6]Plentzia Itsas Estazioa, PIE, University of the Basque Country (UPV/EHU), Plentzia, Spain.
[7]Department of NE and Fluid Mechanics, University of the Basque Country (UPV/EHU), Bilbao Engineering School, Bilbao, Spain.

Correspondence: Santos J. González-Rojí (santos.gonzalez@climate.unibe.ch)

Abstract. The ability of two downscaling experiments to correctly simulate thermodynamic conditions over the Iberian Peninsula (IP) is compared in this paper. To do so, three parameters used to evaluate the unstable conditions in the atmosphere are evaluated: TT index, CAPE and CIN. The WRF model is used for the simulations. The N experiment is driven by ERA-Interim's initial and boundary conditions; The D experiment has the same configuration as N, but the 3DVAR data assimilation step is additionally run at 00, 06, 12 and 18 UTC. Eight radiosondes are available over the IP, and the values for these indices directly retrieved from the Integrated Global Radiosonde Archive (IGRA) were chosen as reference in the validation of both simulations. Additionally, vertical temperature and moisture profiles from the radiosondes provided by the University of Wyoming were used to calculate three parameters commonly used to represent atmospheric instability by our own methodology using the R package *aiRthermo*. According to the validation, the correlation, Standard Deviation (SD) and Root Mean Squared Error (RMSE) obtained by the experiment D for all the variables at most of the stations are better than those for N. The different methods produce small discrepancies between the values for TT, but these are larger for CAPE and CIN due to the dependency of these quantities on the initial conditions assumed for the calculation of a lifted air parcel. Similar results arise from the seasonal analysis concerning both WRF experiments: N tends to overestimate or underestimate (depending on the parameter) the variability of the reference values of the parameters, but D is able to capture it in most of the seasons. In general, D is able to produce more reliable results due to the more realistic values of dew point temperature and virtual temperature profiles over the IP. The heterogeneity of the studied variables is highlighted in the mean maps over the IP. According to those for D, the unstable air masses are found along the entire Atlantic coast during winter, but in summer they are located particularly over the Mediterranean coast. The convective inhibition is more extended towards inland at 00 UTC in those areas. However, high values are observed near the southeastern corner of the IP (near Murcia) also at 12 UTC. Finally, no linear relationship between TT, CAPE or CIN was found, and consequently, CAPE and CIN should be preferred for the study of the instability of the atmosphere as more atmospheric layers are employed during their calculation than for TT index.

1 Introduction

Precipitation is one of the most important variables involved in the water balance, and its variability determines the water resources of the planet. Following the definitions of regional models, precipitation can be separated in two categories: large-scale and convective precipitation. In general, convective precipitation is frequently associated with precipitation extreme events due to high intensity over a short duration. However, the simulation of these events is a well-known problem in the modelling community (Sillmann et al., 2013) due to restrictions in the resolution, poor representation of complex topography, insufficient assimilated observations, forecast errors or deficiencies in the microphysics schemes in the numerical models. In order to avoid these problems, as previously done in the literature (Viceto et al., 2017), this paper focuses on the evaluation of the atmospheric conditions favourable for the development of convective precipitation rather than the validation of the simulation of extreme events.

The evaluation of the atmospheric conditions is typically based on the calculation of some instability indices such as Lifted Index (LI) (Galway, 1956), K-Index (George, 1960), Total Totals index (TT) (Miller, 1975) or Showalter Index (S) (Showalter, 1953). These conditions can be also evaluated by Convective Available Potential Energy (CAPE) (Moncrieff, 1981) or Convective INhibition (CIN) (Moncrieff, 1981). All of these variables are commonly used in the literature for this kind of studies (e.g., Ye et al., 1998; DeRubertis, 2006; Viceto et al., 2017). CAPE and CIN are based on the adiabatic lifting of a parcel, while most of the others are based on differences in the values of several variables at different pressure levels. The deep convection is caused by three ingredients: high levels of moisture in the planetary boundary layer (PBL), potential instability and forced lifting (Johns and Doswell, 1992; McNulty, 1995; Holley et al., 2014; Gascón et al., 2015). CAPE and CIN provide information about the first two ingredients (Holley et al., 2014), and both can give details about the genesis and intensity of the atmospheric convection (Riemann-Campe et al., 2009). However, previous studies (Angus et al., 1988; López et al., 2001) suggest that CAPE should not be used alone, but should be combined with other indices. The final ingredient, which is the forced lifting, is usually caused by the orography (Doswell et al., 1998; Siedlecki, 2009), the convergence of horizontal moisture fluxes (McNulty, 1995) or the breezes in coastal regions (van Delden, 2001). Thus, the high spatial and temporal resolution is important for this kind of studies focusing on the atmospheric convection, and that is why regional simulations are needed (Siedlecki, 2009).

The probability of occurrence of convective precipitation is not the same along the day, and previous studies support that the maximum convection takes place in the afternoon and evening (Siedlecki, 2009; Virts et al., 2013; Piper and Kunz, 2017; Enno et al., 2020). According to van Delden (2001), the preferred time in most of Western Europe is between 18 and 24 UTC, with the exception of the island of Corsica where the sea breeze causes convection usually between 6-12 UTC. In open sea areas, the lightning activity peaks in the morning (Enno et al., 2020), associated to thunderstorms caused by land breezes at night (Virts et al., 2013). A regional study focusing over the UK (Holley et al., 2014) suggests that the reduction overnight of CAPE is over 500 J/kg.

On the global scale, CAPE follows the spatial pattern of surface specific humidity and air temperature, which means that it increases from pole to Equator (Riemann-Campe et al., 2009). The minimums are obtained in arid regions and over areas with cold water up-welling. Focusing on Europe, convective storms develop for lower values than the U.S. (Graf et al., 2011), and several studies tried to determine the most active regions. Amongst them, Romero et al. (2007) found that the region with highest instability is located along a zonal belt over the south-central Europe, particularly over the west Mediterranean sea and the surrounding areas. This agrees with Brooks et al. (2003), who found that the favourable environment for thunderstorms is developed in southern Europe, and that the highest number of days in such a regime are located over the Iberian Peninsula (hereafter, IP), south of the Alps, and Northern Balkans. However, van Delden (2001) found that the southwestern France and the Basque Country seem to be a preferred region for the formation of severe storms that drift towards the northeast. More recent studies based on lightning data (Enno et al., 2020) and regional climate models using higher resolution (Mohr et al., 2015; Rädler et al., 2018) highlighted the same areas with favourable environments for thunderstorms in Europe, which are located in particular over northern Italy (Po Valley), east of the Adriatic Sea (Albania, Bosnia and Serbia) and in the northeastern IP and Southern France (near the Gulf of Lyon).

Over the IP, the seasonality of precipitation is determined by different sources of moisture due to seasonal variations of the global atmospheric circulation and contrasting climatic regions (influenced by the strong topography). Northern and western IP are mainly affected by stratiform precipitation during winter, while eastern and southern IP receive great amounts of precipitation during autumn due to convective activity (Rodríguez-Puebla et al., 1998; Esteban-Parra et al., 1998; Romero et al., 1999; Iturrioz et al., 2007). Maximum precipitation amounts over central IP are measured in early spring (Tullot, 2000).

Previous studies over the IP (Viceto et al., 2017) suggest that CAPE shows a high spatiotemporal variability: the values in winter and spring over land are small due to the reduced surface temperature, and the differences between Atlantic and Mediterranean regions are remarkable during summer. According to Siedlecki (2009), the mean values range from below 50 J/kg in the north to between 100 and

200 J/kg at the Mediterranean coast (some events can even reach 1000 J/kg). As Romero et al. (2007), Viceto et al. (2017) also stated that CAPE is low during autumn in the Atlantic and continental regions, but high in the areas surrounding the Mediterranean sea. This seasonality was also observed for other indices such as K-index or TT, which show maximum values during summer (Siedlecki, 2009). Observations proved that annual precipitation over eastern stations is mostly accumulated during autumn, as a result of the cumulative warming of the Mediterranean sea due to summer insolation (Romero et al., 2007; Iturrioz et al., 2007), and later entry of very hot and humid air into the IP while cold air is present at higher levels (Dai, 1999; Eshel and Farrell, 2001; Correoso et al., 2006). Additionally, September and October are the months with highest frequency of waterspouts and tornadoes near the Balearic Islands (Gayà et al., 2001). Over the northwestern IP, the mean values of the CAPE when hailstorms occur is 360 J/kg, while for thunderstorms it is only 259 J/kg (López et al., 2001). The dispersion of these values is really high (almost 350 J/kg over the whole sample), which is similar to that found in previous studies (Alexander and Young, 1992; Lucas et al., 1994). The values are similar to those observed in other regions of Europe, but lower than those values obtained in studies based on synoptic or lightning data for severe hailstorms (around 500 J/kg) (Kunz, 2007; Púčik et al., 2015; Taszarek et al., 2017). Due to global warming, the conditions necessary for the development of extreme precipitation events will be enhanced (Brooks, 2013; Rädler et al., 2019). The frequency and intensity of climate extremes will be magnified (Diffenbaugh et al., 2013), projecting larger values of CAPE at the Mediterranean coast during summer and autumn (Marsh et al., 2009; Viceto et al., 2017).

The main objective of this paper is to evaluate the performance of two simulations created by using the Weather and Research Forecasting (WRF) model (Skamarock et al., 2008) (including or not the extra 3DVAR data assimilation step) at reproducing the atmospheric conditions that can cause convective precipitation over the IP if the third ingredient (e.g. lifting) is fulfilled. We are not restricting our analysis only to convective situations, and the entire period from 2010-2014 will be considered. For the evaluation, the comparison of pseudo-soundings extracted from the model against real observations will be carried out. Additionally, the seasonal patterns of different variables commonly used to represent atmospheric instability will be studied. Moreover, this study will also help us to accurately determine the regions of the IP more prone to develop unstable thermodynamic conditions. If the condition of the forced lifting is also fulfilled, convective precipitation can be developed in those areas. As shown before, a highly demanding feature in model simulations, a topic with great importance nowadays due to the large damage that extreme convective events can cause to society, and which frequency will be increased in the future.

The novelty of this study lies in the inclusion of data assimilation step in the downscaling experiment used for the analysis of some instability indices, as most of the previous studies are mainly based on simulations driven by boundary conditions after its initialization (e.g., Marsh et al., 2009; Holley et al., 2014; Mohr et al., 2015). To those not familiar with the data assimilation process, its main objective is to produce more reliable and accurate initial conditions for regional models. This is achieved once the effect of the assimilated observations is used to modify the fields of temperature, wind and pressure in order to make them closer to the observations. The impact of the data assimilation is not restricted only to the location of the observations being assimilated. First, the improvements due to the analysis are propagated zonally, meridionally and vertically to the nearby grid points of the domain by means of the background error covariance matrix (Barker et al., 2004, 2012). Second, after the simulation in the new cycle is performed from the initial conditions achieved through assimilation, they propagate in the next six hours by means of advection, thus affecting areas distant from the original observations.

This paper is organised as follows: The details of the configuration of the WRF model used in both experiments are presented in section 2, along with a brief outline of the methodologies used in the study. The main results are presented in section 3, while they are compared against previous studies presented in the introduction. Finally, we conclude with some remarks about our research in section 4.

2 Data and Methodology

2.1 WRF model configuration

Two experiments were carried out using version 3.6.1 of the WRF model for the period 2010-2014. In both simulations, ERA-Interim provides the initial and boundary conditions (Dee et al., 2011). Six-hourly data at 0.75 degrees were downloaded from the Meteorological Archival and Retrieval System (MARS) repository at European Centre for Medium-Range Weather Forecasts (ECMWF). Analyses of temperature, relative humidity, both horizontal wind components and geopotential height at 20 pressure levels (5, 10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 850, 900, 925, 950, 1000 hPa) were used to feed WRF. Both simulations were started on the 1st of January, 2009 from a cold start. Following similar methodologies to previous studies (Argüeso et al., 2011; Zheng et al., 2017), the entire year 2009 was selected as the spin-up for the land surface model included in WRF, and consequently, it was omitted in the study presented here.

One of the experiments (hereafter, N) was nested inside ERA-Interim as usual in numerical downscaling experiments, which means that the model is driven by the boundary conditions after its initialization. It is generated running 6-hour long segments that are restarted from the restart file

produced at the end of previous segment, which is similar to a continuous WRF run where the boundary conditions are provided to the model every 6-hours after the initialization of the model. The other experiment (D) relies on the same setup, but with the additional 3DVAR data assimilation step (Barker et al., 2004, 2012) that is run every 6 hours (at 00, 06, 12 and 18 UTC). In this case, 12-hour long segments starting at every analysis time (00, 06, 12 and 18 UTC) are used. The analyses are generated from the outputs of the model at a 6-hour forecast step from the previous segment as first guess in a 3DVAR data assimilation scheme. In both experiments, the outputs are saved every 3 hours, which means that analyses (00, 06, 12 and 18 UTC) and 3-hour forecasts (at 03, 09, 15 and 21 UTC) are included in our results. In the data assimilation step, quality controlled temperature, moisture, pressure and wind observations in PREPBUFR format from the NCEP ADP Global Upper Air and Surface Weather Observations dataset (referenced as $ds337.0$ at NCAR's Research Data Archive) were included. Only those observations included in a time-window of two-hours centered in the analysis times were assimilated.

As Figure 1 shows, the domain focuses over the IP, but it also includes parts of Europe, Africa and the Atlantic ocean. As stated by previous studies (Jones et al., 1995; Rummukainen, 2010), the set-up of the domain used in this study prevents border-effects affecting our results as mesoscale systems can develop freely. The spatial resolution of both experiments is 15 km, and they include 51 vertical levels up to 20 hPa in eta (η) coordinates.

Apart from the ERA-Interim data, sea surface temperature (SST) of the model was updated on a daily basis using the high-resolution dataset *NOAA OI SST v2* (Reynolds et al., 2007) developed by the National Oceanic and Atmospheric Administration (NOAA). Additionally, the following parameterizations for the physics of the model were included in both WRF simulations: five-class microphysics scheme (WSM5) (Hong et al., 2004), MYNN2 planetary boundary layer scheme (Nakanishi and Niino, 2006), Tiedtke cumulus convection scheme (Tiedtke, 1989; Zhang et al., 2011), RRTMG scheme for both long and shortwave radiation (Iacono et al., 2008), and NOAH land surface model (Tewari et al., 2004).

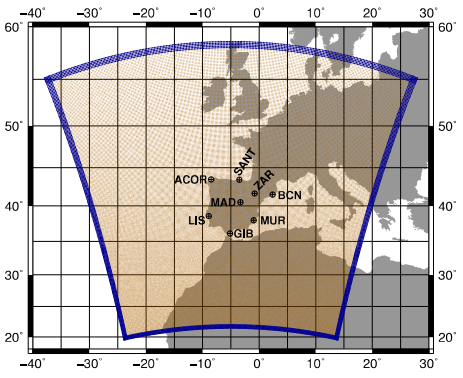

Figure 1. The domain used in both WRF simulations is presented with dark orange dots, while the dark blue region highlights the relaxation zone. The location of all the radiosondes available over the IP is also presented with quartered circles.

The background error covariance matrices were created before running the simulation with 3DVAR data assimilation. To do so, the CV5 method included in WRFDA (Parrish and Derber, 1992) was used. A separate simulation initialized at 00 and 12 UTC and spanning 13 months (from January 2007 to February 2008) was necessary for the calculation of these matrices. Independent matrices were created for each month, and each of them was calculated taking into account a 90 days period centered on each month.

Both simulations were already presented and validated in previous studies by the authors. Integrated water vapor, precipitation and evaporation over the IP were validated against station measurements and gridded datasets including independent satellite data in González-Rojí et al. (2018), and the outputs produced by D were always superior to N and the driving reanalysis ERA-Interim (for the latter, at least comparable for some variables). The closure of the water balance was also better for D. Additionally, the precipitation from D exhibited similar capabilities to the one downscaled with statistical methods (González-Rojí et al., 2019). Furthermore, the wind field from D showed also improvements compared to ERA-Interim, and consequently, that data were used for the calculation of the offshore wind energy potential in the west Mediterranean (Ulazia et al., 2017). Afterwards, that study was extended to every coast of the IP (Ulazia et al., 2019). The moisture recycling over the IP was also evaluated in González-Rojí et al. (2020), highlighting the reliable results produced by the experiment including data assimilation and the importance of the moisture recycling at the Mediterranean coast during spring and summer.

The effect of data assimilation on moisture and temperature was measured by the analysis increments (analysis minus background) in González-Rojí et al. (2018). The effect of the data assimilation is more intense at 12 UTC compared to the other times, and particularly for summer (see their Figure 13). The spatial analysis of these values highlight that the effect of data assimilation is not homogeneous over the IP, and it concentrates mainly in the southeastern IP and both Guadalquivir and Ebro basins. Southern IP has been already highlighted by previous studies as a region where cold biases are observed in WRF simulations during summer (Fernández et al., 2007; Argüeso et al., 2011; Jerez et al., 2012). The fact that the effect of data assimilation is concentrated in that region in our WRF simulations is not a coincidence, and thus, the data assimilation helps to reduce that bias to some extent.

2.2 Radiosonde data

Atmospheric radiosonde data were downloaded from the server of the University of Wyoming (freely accessible at http://weather.uwyo.edu/upperair/sounding.html). Even if the University of Wyoming does not apply any quality control to the data, this dataset was already used in previous studies by the authors and none of the values were taken as erroneous. Moreover, data from the Integrated Global Radiosonde Archive (IGRA) created by NOAA was also included in this study. This dataset is available online after several quality control procedures.

Only eight radiosondes are available over the IP: A Coruna (ACOR), Santander (SANT), Zaragoza (ZAR), Barcelona (BCN), Madrid (MAD), Lisbon (LIS), Gibraltar (GIB) and Murcia (MUR). The location of each station is presented in Figure 1. Measurements are carried out every day at midday and midnight (00 and 12 UTC, corresponding to 02 and 14 LT summer time, respectively), with the exception of Lisbon where they are only available at 12 UTC (13 LT summer time). Additionally, the amount of data available for Gibraltar is extremely scarce since August 2012.

Temperature and mixing ratio were retrieved at all the available pressure levels at each location from the University of Wyoming database and from the IGRA dataset. Moreover, the values of the TT, CAPE and CIN as calculated directly by the creators were also retrieved. However, only the values computed from the IGRA dataset were assumed as the reference in our analysis. Additionally, vertical profiles of temperature and mixing ratio downloaded from the University of Wyoming were also used to calculate TT, CAPE and CIN following our own methodology using the *aiRthermo* R package (further details can be found in the next subsection). The comparison between the original values of the indices retrieved and our results can give us information about whether their discrepancies are only due to differences in the calculation procedure.

It must be said that all the radiosondes presented here were assimilated during the 3DVAR data assimilation step in WRF. However, we do not assimilate directly any of the evaluated parameters or precipitation, as we mainly assimilate pressure, temperature, humidity and wind. Additionally, as already stated, only eight radiosondes are available over the IP. The validation of the results against the assimilated radiosondes (even if we don't assimilate directly the studied variables) can be seen as biased, but we cannot exclude some of these measurements from the data assimilation process only to be able to validate the simulation afterwards with such a reduced amount of data available (e.g. assimilate only four radiosondes, and validate the simulation with the remaining four radiosondes). Thus, in order to get the most accurate results as possible from the model, all the available measurements are used.

Moreover, we also assume that these radiosondes were very likely assimilated in ERA-Interim. Nevertheless, the impact of this is insignificant in our simulations as we only used ERA-Interim data as boundary conditions for our regional model after the initial run (January 1st, 2009). We only analysed the outputs from the model after one year of spin-up, so the results are only taking into account the variability corresponding to the regional climate model.

2.3 Methodology

2.3.1 Calculation of parameters representing atmospheric instability

For both simulations, the nearest grid point to the real latitude and longitude of each radiosonde was determined, and the corresponding pseudo-sounding (pressure, temperature and mixing ratio) at 00 and 12 UTC was obtained at WRF's original η levels. We did not consider the averaged value of several grid points as we would be considering a extremely large area to be compared against radiosonde data. For example, if we consider an array of 3×3 grid points, we would be taking into account an area of 2025 km^2 (45 km $\times$ 45 km, as the horizontal resolution of our domain is 15 km), which is not suitable to be compared against in-situ data. Additionally, according to Xu et al. (2015), for every sounding balloon, the vertical profile of the atmosphere up to 6 km is already measured for a drifting distance of 7.5 km (half the spatial resolution we use) even if the samples are taken during a clear or cloudy day (see their Figure 6). That means that our spatial resolution is suitable for the direct comparison of the nearest points against radiosonde data, as we do not neglect the horizontal drift of the sounding balloons. Thus, averaging the neighbour grid points would be suitable for the validation of results when convection-permitting scales are used, but not in our case as the spatial resolution of our model run is 15 km.

Extracting pseudo-sounding from reanalysis or model data is nothing new, and Lee (2002) or Molina et al. (2020) amongst others showed that these pseudo-soundings are able to reproduce reasonably well the atmospheric conditions measured by real soundings. However, as highlighted by Holley et al. (2014), this procedure takes into account a stationary column at a fixed time, which can influence the comparison to real radiosonde data as these measurements are not instantaneous and not in a straight vertical line as the balloons used deviate because of wind.

In order to calculate TT, CAPE and CIN using the pseudo-soundings from the model, the R package *aiRthermo* was employed (Sáenz et al., 2019). The most recent version was selected (version 1.2.1), which is publicly available in the CRAN repository (https://cran.r-project.org/package=aiRthermo). Both CAPE and CIN are calculated by means of the vertical integrals using discrete slabs defined by the resolution of pressure in the soundings (using all the available levels). The integrals for each of the slabs enclosed by linear profiles are computed analytically, and the energy corresponding to each slab is accumulated, producing the final

value of CAPE or CIN. The virtual temperature was used in every integral (Doswell and Rasmussen, 1994). Further details about the functions used for the calculation of the vertical evolution of the air parcels can be found in Sáenz et al. (2019) and also in the manual of the R package *aiRthermo* associated to that publication.

Additionally, in order to calculate CAPE and CIN in the most similar way to the University of Wyoming with the aim of reducing the differences between the values due to different calculation procedures, the average of the lower vertical levels was set as the initial representative parcel (Craven et al., 2002; Siedlecki, 2009; Letkewicz and Parker, 2010). As in Siedlecki (2009), the averaged values from the lowest 500 m were used in this study. Furthermore, in order to avoid that the averaged initial parcel state is still too hot compared to the ambient conditions (in that case, CIN will never be computed as the parcel is already artificially buoyant), an isobaric precooling was applied if needed. To do that, the parcel is cooled along an isobar until it crosses the sounding so that it is not buoyant at the initial state.

The TT index was calculated following the definition from Miller (1975). It is defined as:

$$TT = (T_{850} - T_{500}) + (D_{850} - T_{500}) \quad (1)$$

where T_{850} and T_{500} are the temperatures at 850 and 500 hPa, and D_{850} is the dew point temperature at 850 hPa. According to the ECMWF (Owens and Hewson, 2018), thunderstorms are likely when the values for this index are above 44°C. However, other values can be found in the literature: 48.1°C for southern Germany (Kunz, 2007), 46.7°C for the Netherlands (Haklander and Van Delden, 2003) or 46°C for Switzerland (Huntrieser et al., 1997).

It can be seen that this index is not highly dependant on the initial conditions for its calculation as it only depends on temperature at two discrete pressure levels, while CAPE and CIN are very sensitive to the initial conditions used for the simulated ascent. TT avoids this problem, but the results can suffer from errors due to inversion layers (Siedlecki, 2009). It must be pointed out that the dew point temperature is needed for TT, and that it is highly important for the calculation of the Lifting Condensation Level (LCL) while calculating CAPE and CIN. In the case of the radiosonde data, the indices are calculated using the measured dew point temperature at 850 hPa when is needed, while in our method, this variable is calculated from the temperature and mixing ratio at that pressure level. This can cause small differences in the results even if the same original radiosonde data are used.

Further indices could be calculated from the pseudo-soundings obtained from the outputs of the model or real observations. However, keeping in mind that the main objective of this study is to evaluate the difference in the performance of two simulations at reproducing the unstable atmospheric conditions that can cause convection, we needed to restrict our study to a small set of the indices calculated and provided directly by the radiosonde data holders used in this study: IGRA and University of Wyoming. By doing that, we can compare our results to those obtained by them, and infer which simulation performs better (that including data assimilation or the one without). In both cases, CAPE, CIN, TT index, LI, S index or K index are provided. In the case of the University of Wyoming, SWEAT is also included, but not in IGRA. Then, only six indices were available for us for the validation of our data. The R-package aiRthermo also allows the calculation of these six indices, so it was not a restricting feature in our analysis. However, previous studies reported a strong correlation between CAPE and LI (Blanchard, 1998; López et al., 2001), and K-index is also based on temperature at different pressure levels, so it suffers from the same problems as TT. Consequently, in order to avoid these connections between indices, we restricted this study to TT, CAPE and CIN.

2.3.2 Analysis

Once TT, CAPE and CIN are calculated at the nearest grid points to radiosonde locations of both simulations (N and D), and also those using the original sounding data (labelled as 'aiRthermo' in the results), we obtain a time series with a 12-hourly temporal resolution for each index. These values can be compared against the reference values of the indices retrieved directly from IGRA and the University of Wyoming (labelled as 'Reference' and 'Wyoming' in the next figures). The comparison between 'Wyoming' and 'aiRthermo' aims to achieve an estimation of the error/differences due to the different methods applied by both sources of results. This comparison was based on independent locations over the IP (separated by several kilometers), so a Taylor diagram was chosen as the best option to show Pearson's correlation (r), Root Mean Squared Error (RMSE) and Standard Deviation (SD) of each experiment in the same plot. In order to determine which experiment is doing the best job at simulating the reference values of the variables, the procedure explained by Taylor (2001) was followed: the dots that lie nearest to the reference on the X axis represent variables that agree well with observations (high correlations and low RMSEs), and those lying near the highlighted arc will present comparable standard deviations to the observations.

Additionally, the bootstrap technique with resampling was applied to the results in order to represent an estimation of the sampling errors from each experiment (Efron and Gong, 1983; Wilks, 2011). In our case, the original time series used in the Taylor diagrams consist of 60 values, each of them for the corresponding month along the period 2010-2014 (12 months $\times$ 5 years). For the bootstrap, we created 1000 perturbed time series taking into account different samples of the data. 67 % of the new time series (2/3 of the length of the original time series – 40 values in our case) is made from the original data, and the remaining 33 % (1/3 – 20 values) is chosen from those values already taken from the original data. For each correlation calculated, the same samples

are taken from all datasets and experiments. The variability of the Pearson's and Spearman's correlations obtained with these synthetic time series were shown by means of Box-Whiskers plots.

Then, the seasonal analysis of each parameter at each location was carried out. In this case, the variability of the results is showed by different Box-Whiskers plots. Each season was defined as follows: winter is defined from December to February (DJF), spring from March to May (MAM), summer from June to August (JJA) and autumn from September to November (SON).

Finally, the calculation of TT, CAPE and CIN was extended to every grid point included in a mask defined for the land points of the IP over model's domain. The spatial distribution of the mean values of them at 00 and 12 UTC during winter and summer was calculated. These maps show the spatial distribution of TT, CAPE and CIN over the land grid points in the IP which are more prevalent in each season.

3 Results

Taylor diagrams for the TT index calculated for each radiosonde of the IP are shown in Figure 2. The Box-Whiskers associated to the correlations (both Pearson's and Spearman's) obtained for each of the 1000 time-series created with the bootstrap technique are also included. According to the Taylor diagrams, as expected, the best experiment reproducing the reference values is Wyoming, followed by aiRthermo (the real measurements of temperature, mixing ratio and pressure from the sounding were used to calculate TT with our methodology), D and later by N. Wyoming obtains the closest values to the observations at all the stations. The results for aiRthermo are quite similar to Wyoming, except for Murcia where D is better reproducing the reference data. The correlations are always above 0.99 for Wyoming, 0.98 for aiRthermo, 0.97 for D and 0.75 for N. The observed SD is really well simulated by Wyoming, aiRthermo and D, but N underestimates it at most of the stations as it is only able to reproduce the one in Santander and A Coruna. The RMSE is below 0.6 °C for aiRthermo, below 1 °C for D and below 2.5 °C for N.

The bootstrap analysis is consistent with the results obtained in the Taylor diagrams, and it shows that the Pearson's correlations are always above 0.99 for Wyoming and 0.98 for aiRthermo (again, with the exception of Murcia where they are above 0.9). The correlations are always above 0.95 for D. In the case of N, the spread of the values is much larger than for aiRthermo and D, and their median values are obtained between 0.8 and 0.9. If we change to Spearman's correlations, we can see that values are similar but with a small decrease of the values (particularly in Gibraltar, Murcia and Madrid).

Thus, as expected, we obtain the most similar results to those calculated from IGRA (Reference) with the values from Wyoming and those calculated with the real measurements from the soundings (that is, aiRthermo). However, we can still detect small differences between the values of the datasets due to the use of measured dew point temperature in Wyoming, whilst it is computed from temperature and mixing ratio in IGRA and aiRthermo. These differences are more remarkable in Murcia. Between both WRF experiments, it is clear that the experiment including the 3DVAR data assimilation is able to outperform the standard simulation only driven by the reanalysis data at the boundaries of the domain (N). The differences between both WRF simulations are highlighted, particularly at those stations located at the Mediterranean coast (Barcelona, Murcia and Gibraltar) and in Lisbon.

In the case of CAPE, the validation results are presented in Figure 3. The best experiment reproducing the results is aiRthermo, followed by Wyoming, D and finally by N. The correlations are at all the stations above 0.99 in aiRthermo and 0.95 for Wyoming, while for D they are above 0.9 and above 0.7 for N. A similar behaviour is observed for SD and RMSE. The largest RMSEs are obtained for Barcelona and Murcia (both in the Mediterranean region).

The bootstrap analysis shows that the highest Pearson's correlations are obtained by aiRthermo and Wyoming, but followed really closely by D. As for TT index, N presents the worst performance and the largest spread. If we consider, instead, the use of Spearman's correlations, we can see that the values are similar at most of the stations and only in A Coruna and Santander there is a strong decrease of the values.

As stated before in Section 2.3.1, the calculation of CAPE is more sensitive to subtle differences in the methodology than that of TT, and this is highlighted in the validation of these results. Even if the same data are used for the calculation of CAPE (Wyoming and aiRthermo used the same measurements as input), it is clear that small differences in the initial conditions can result in serious discrepancies between both methods as stated by Siedlecki (2009). The largest RMSEs for aiRthermo can be found at Barcelona and Murcia. As for the TT index, two stations at the Mediterranean coast present the largest differences between the experiments. However, while the computation of TT from both WRF simulations produces standard deviations similar to the observed ones, the results for CAPE substantially overestimate the variance of Atlantic sites (A Coruna, Santander and Lisbon) and Madrid, or overestimate it at the Mediterranean coast (Barcelona, Murcia and Gibraltar). Anyway, it can be seen that data assimilation improves the simulation of CAPE over the IP.

Finally, the validation of CIN is presented in Figure 4. As for CAPE, the best results are obtained again by aiRthermo, followed by Wyoming, D and N (with the exception in Gibraltar where D and N are really similar). aiRthermo obtains in every station correlations above 0.97, followed by Wyoming, D and N with correlations above 0.93, 0.85 and

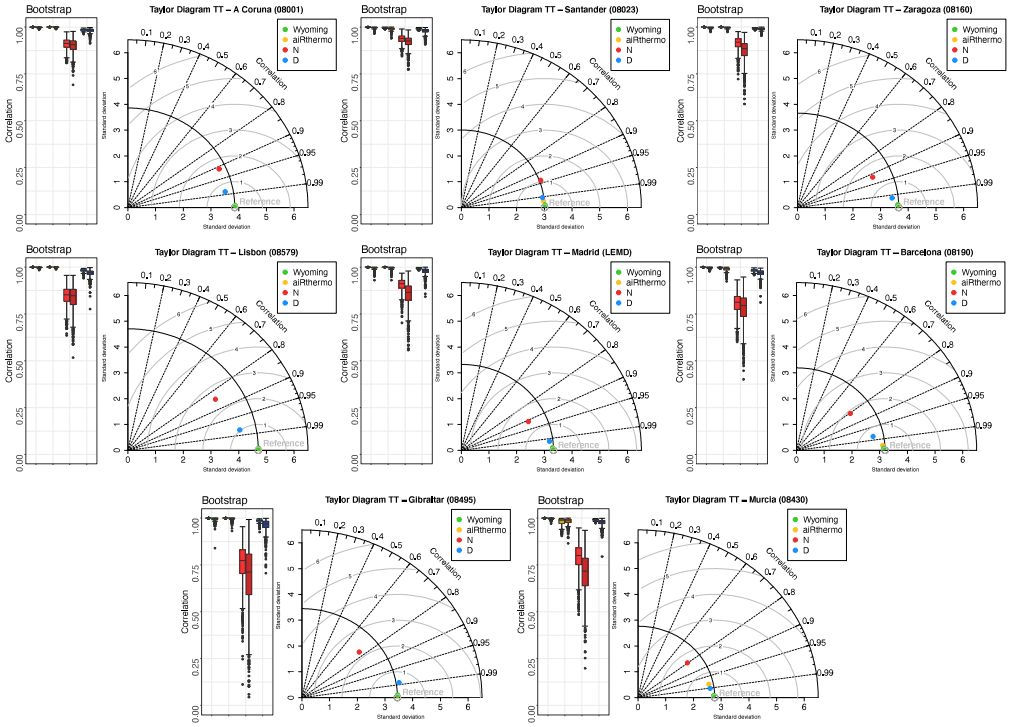

Figure 2. Taylor diagrams showing the r, RMSE and SD values for Wyoming, aiRthermo, N and D compared to TT values retrieved directly from IGRA (Reference). On the left side of each Taylor diagram, a Box-Whiskers plot is added in order to show Pearson's and Spearman's correlations between each experiment and the reference data (lighter and darker colours, or first and second columns of the Box-Whiskers respectively). The bootstrap technique with resampling was used to create 1000 synthetic time-series. Wyoming, aiRthermo, N and D are plotted in green, orange, red and blue respectively.

0.65 respectively. Both WRF experiments overestimate or underestimate it depending on the station (particularly N in Lisbon, Madrid, Murcia and Zaragoza). The RMSE is always larger for N, and particularly in Murcia and Gibraltar where the values exceed the 40 J/kg.

The bootstrap analysis presents the same results as for CAPE (Figure 3). However, for Gibraltar, as shown in the Taylor diagram, both WRF experiments produce similar Pearson's correlation values during the bootstrap. If we consider, instead, Spearman's correlations, the worse performance of N is perceptible in A Coruna, Santander, Murcia and Gibraltar. However, in Gibraltar, differences between both WRF experiments arise: WRF D obtained better correlations than WRF N as in the other stations. In contrast to previous results, the poorest correlations for CIN are obtained at stations located at the Atlantic coast as Lisbon and A Coruna.

As for CAPE, the differences between aiRthermo and Wyoming are highlighted here. This result supports the idea that small differences in the initial conditions of the lifted air parcel and the determination of the LCL due to differences in the dew point temperature can cause large differences in the values of CIN even if the same vertical profile of temperature and mixing ratio are used for its calculation. Again, the differences between both WRF experiments are important and the experiment including data assimilation (D) presents generally closer results to the observed ones.

The seasonal analysis of the five datasets (Reference, Wyoming, aiRthermo, N and D) for TT index is presented in Figure 5. In this case, Wyoming, aiRthermo and D are able to correctly simulate the reference seasonal variability of TT index at all the stations and all the seasons. However, N tends to overestimate the values of TT in every season and for most of the stations over the IP. The difference in TT is

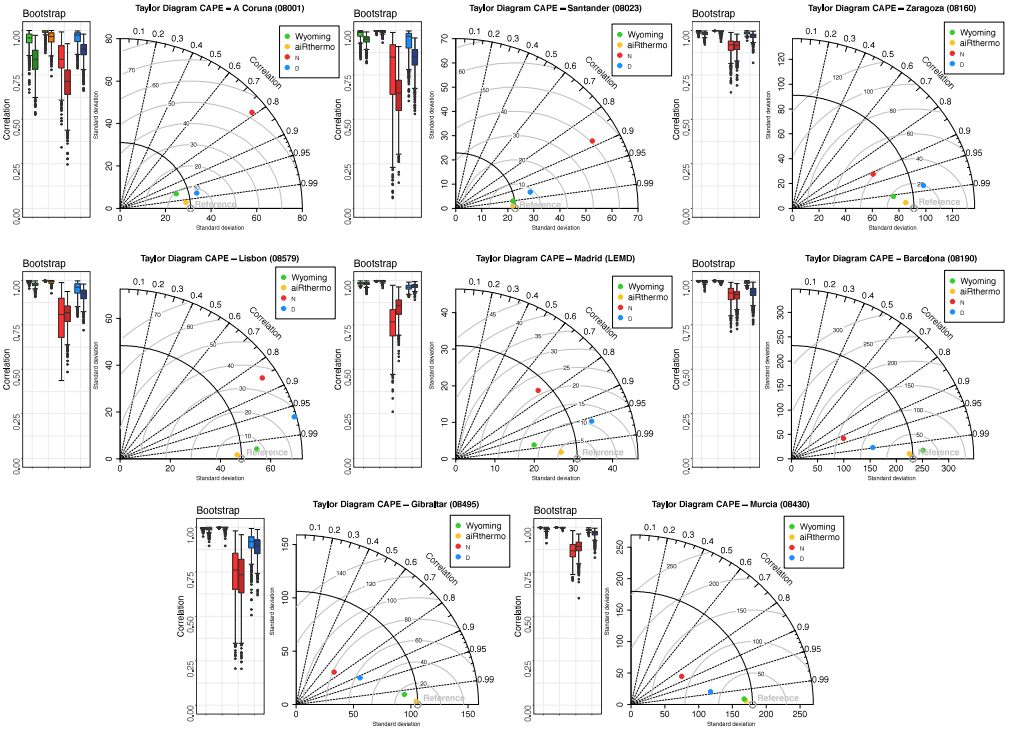

Figure 3. Same as Figure 2 but for CAPE.

most important during winter months as a severe overestimation by the simulation N without data assimilation can be seen. It can be tracked on the basis of Figures A1, A2 and A3 to an improved representation of temperature at 500 hPa. As shown by Figure A.1, the winter temperature at 850 hPa is higher for D than for N, but this would lead to higher values of TT, so that it does not explain the observed discrepancies for the N simulation. For the case of dew point temperature at 850 hPa (Figure A2), it is higher for the N simulation than for D, which leads to higher values of TT for N. Additionally, the temperature at 500 hPa is higher for D than for N, and this also leads to higher values of TT for N than for D. It is thus, clear, that the improvement in the simulation of TT during winter for the D simulation is due to an improved simulation of moisture and temperature at the low (850 hPa) and mid-troposphere (500 hPa) derived from the assimilation of soundings. The same diagnostic can be done for spring, another season during which N overestimates TT in many of the soundings (A Coruna, Santander, Lisbon and Barcelona). They are located in areas where the difference between the dew point temperature from simulation D minus the one from N is negative (see Figure A2).

A Coruna and Santander present the largest values during winter. Higher values than in winter are observed during spring, and the maximum is recognizable in Madrid, which is the only station located over central IP. Even if the maximum is located there, the other stations also present values above 38°C. During summer, central, eastern and southern stations (Madrid, Barcelona, Zaragoza and Murcia) are the ones presenting higher values. In that season, the Atlantic stations (A Coruna and Santander) and Gibraltar present values below 40°C. The values of TT in summer at those stations are smaller than the ones in winter, which occurs mainly due to the combined effect of the high increasing values of temperature at 850 and 500 hPa (about 15 and 10 degrees respectively) and the smaller increase of dew point temperature (only a few degrees) in those regions from winter to summer (see Figures A1, A2 and A3 from the Appendix). Finally, all the stations show similar values in autumn, with the exception of Gibraltar where the values are smaller.

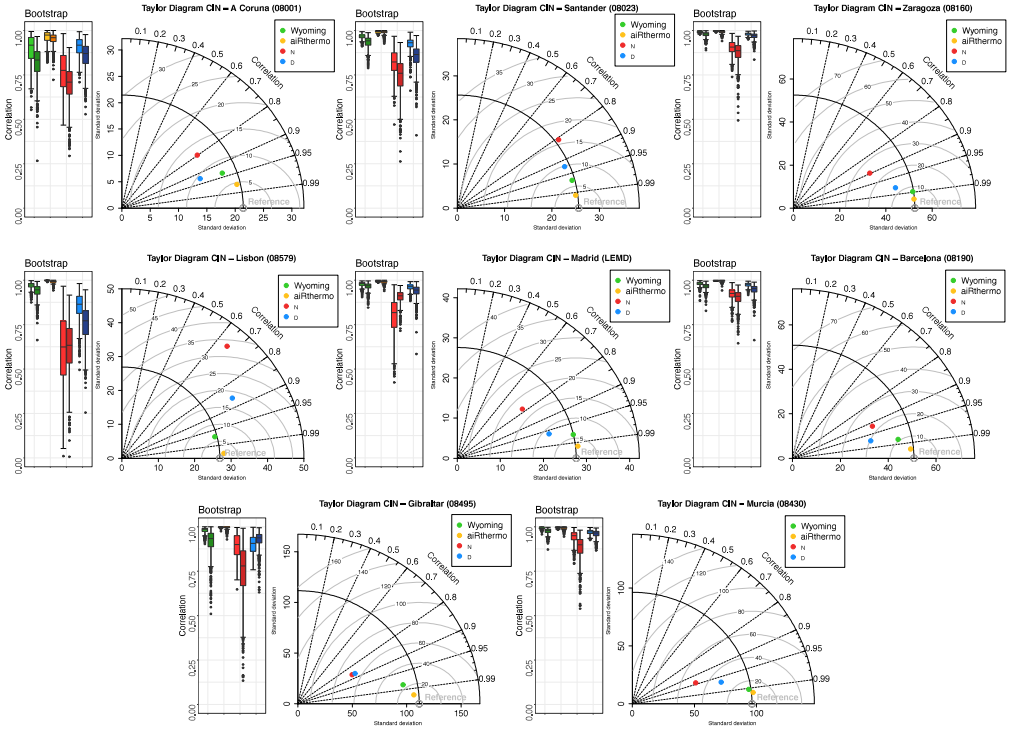

Figure 4. Same as Figures 2 and 3 but for CIN.

The seasonal analysis for CAPE is presented in Figure 6, and it highlights the spatial and temporal heterogeneity of the areas where unstable air masses can be observed over the IP, as also shown by Holley et al. (2014). Wyoming and aiRthermo are able to reproduce (as expected) the variability of the reference values, and D is able to capture the spread of the values at most of the stations during winter, summer and autumn. However, both WRF experiments (particularly D) overestimate CAPE at most of the stations in spring due to the differences in the virtual temperature in lower levels compared to reference data (colder near surface and warmer near 800 hPa) and with lifted parcels for D slightly warmer than the reference ones and those for N (see Figure A4).

The experiment without data assimilation (N) tends to overestimate CAPE in winter and to underestimate it in summer. In winter, this overestimation is caused mainly by colder conditions in the 850-750 hPa pressure levels and warmer lifted air parcels (particularly in Lisbon, A Coruna and Santander - See Figure A5). A detailed analysis of the vertical structure of the differences of both simulations against IGRA for virtual temperature and mixing ratio (Figure A6) shows that the vertical structure of moisture is improved in the D simulation, thus leading to a better estimation of CAPE through the troposphere due to improved estimations of virtual temperature.

On the contrary, in summer, the underestimations of CAPE by the N simulation are caused by warmer conditions in the lower pressure levels compared to the reference, which produces that the lifted parcel crosses the sounding in a lower pressure level than D, and consequently, underestimating CAPE (particularly for Barcelona, Murcia and Gibraltar - See Figure A7). A detailed analysis for Barcelona (Figure A8) shows that there exists a substantial underestimation of moisture at lowest levels by the N simulation, something which is consistent with findings in González-Rojí et al. (2018) in a verification with independent non-assimilated MODIS integrated water vapour. This is also observed in Murcia (Figure not shown). During spring and autumn, the underestimations or overestimations of N depend on the station and a clear pattern is not observed.

The lowest values of CAPE are obtained during winter (below 50 J/kg at all the stations), and the largest ones are

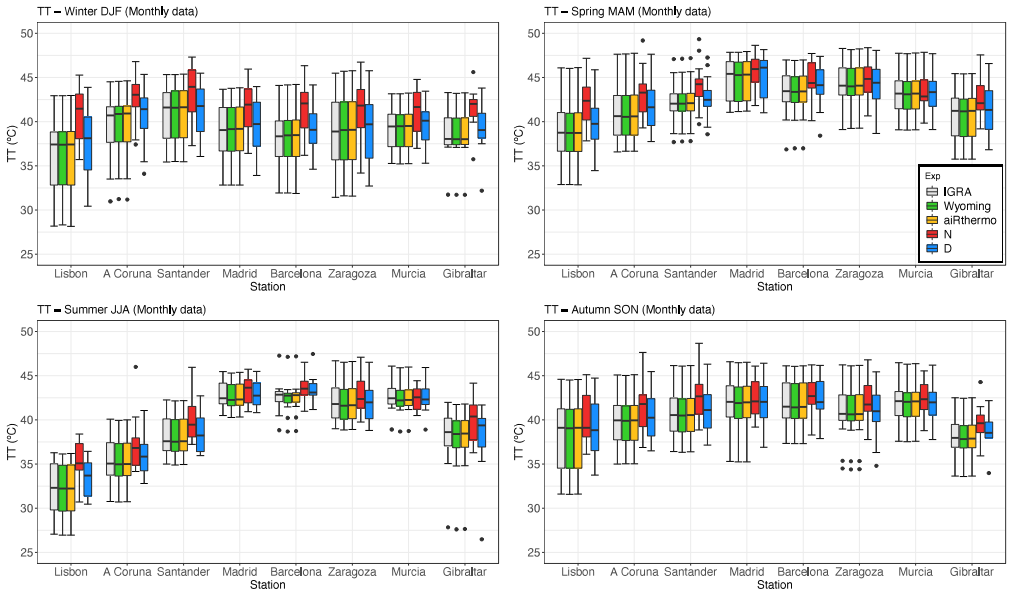

Figure 5. TT index for the reference data (grey), Wyoming (green), aiRthermo (orange), N (red) and D (blue) computed at each station for every season: winter (top panel, left), spring (top panel, right), summer (bottom panel, left) and autumn (bottom panel, right).

observed in summer (reaching 500 J/kg at some stations). However, as stated before, the distribution of CAPE is not homogeneous and different regions are prone to higher values during each season. During winter, the three Atlantic stations (A Coruna, Santander and Lisbon) and Gibraltar present the highest values of CAPE over the IP. In general, the values are below 50 J/kg, but some events can exceed 100 J/kg. During spring, the distribution of CAPE is quite homogeneous over the IP and only stations such as Lisbon, Madrid or Gibraltar present slightly higher values of CAPE than the other stations. In summer, only the stations located in the eastern and southern parts of the IP present remarkable values of CAPE. Particularly, the highest CAPE values are located at the Mediterranean coast (Barcelona and Murcia). Finally, during autumn, the regions with high CAPE are extended towards the inland of the IP, such as Madrid and Zaragoza. During this season, some extreme events can reach values over 1000 J/kg over the Mediterranean coast. This feature was already observed by Siedlecki (2009). All these seasonal changes in CAPE also agree with previous studies based on CAPE (Romero et al., 2007; Viceto et al., 2017).

Finally, the seasonal analysis for CIN is presented in Figure 7, and it highlights the stations where the inhibition is important. In general, Wyoming tends to underestimate the values of CIN at most of the stations and in every season, while aiRthermo is able to capture it. Both WRF simulations (but particularly the experiment without data assimilation) tend to underestimate the observed variability.

The values of CIN are smaller in winter and spring, and the maximum is observed in summer. During winter, CIN is higher in Gibraltar and at the Atlantic stations (Lisbon, A Coruna and Santander) than at the other stations from the IP. However, these values are small compared to those for other seasons. During spring, the values are higher than in winter, and similar values are observed at most of the stations (around 10 J/kg), with the exception of Barcelona where the CIN reaches values of 20 J/kg. In summer, the values are higher at all the stations, but particularly in those for the eastern and southern IP (Barcelona, Zaragoza, Murcia and Gibraltar). The same regime is observed for autumn, but the values are smaller than in summer. These values during summer and autumn are in agreement with Siedlecki (2009), who found CIN means above 100 J/kg in the west Mediterranean sea and surrounding countries.

As stated before, in the final phase of this study, the same procedure for the calculation of TT, CAPE and CIN at each station was extended to each grid point included in the IP. The mean winter and summer spatial patterns at 00 and 12 UTC were calculated for both WRF experiments. The spatial distribution for TT is shown in Figure 8, which highlights

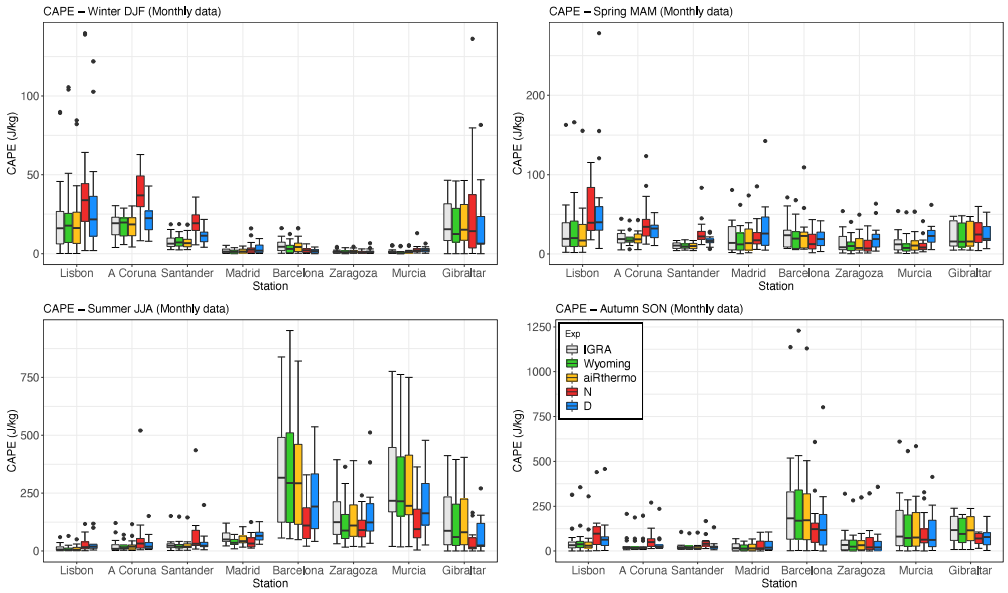

Figure 6. Same as Figure 5 but for CAPE.

the heterogeneity of the results. The differences between both simulations are observable, but also those between day and night. Additionally, it can be seen that TT cannot be calculated in most of the mountain regions of the IP because the 850 hPa layer is near the surface or below ground.

During winter, the maps of TT show that N yields higher values than D, which is in agreement with the overestimation observed in Figure 5. At 00 UTC, according to D, the regions where unstable air masses are observed are those at the Cantabrian coast and in the southeastern IP. Both regions are surrounded by remarkable mountainous systems such as the Cantabrian Range and the Baetic system, which can trigger convection by orographically induced lifting. For N, these areas are also extended to the rest of the IP, with the exception of the southwestern corner where the values are small. At 12 UTC, after solar irradiance has started heating up the land, the regions extend towards inland areas. In the experiment with data assimilation (D), most of the northern plateau presents high values of TT, and the lowest values are observed near the coastal valleys of the southwestern corner and the Mediterranean coast (like the Ebro basin or Murcia). In the case of N, the lowest values are observed mainly in the southwestern IP near the Guadalquivir valley. According to Figures A1, A2 and A3, the lowest values of TT observed near the coastal valleys of both WRF experiments are a consequence of low dew point temperatures there mainly due to dry air.

As in Figure 5, much higher TT values are obtained during summer over the IP, particularly at 12 UTC. At 00 UTC, a west-east gradient is observed in both WRF simulations. However, the values depicted at the Mediterranean coast are higher for the experiment including data assimilation (D). At 12 UTC, the regions with unstable air masses extend towards the central area. In this case, they are located near the Pyrenees, and in the proximity of the Iberian, Central and Baetic systems. The minimum TTs are observed in the western part of the IP, but particularly near Lisbon. The intensity of the most extreme values of TT is higher in D (with data assimilation).

The maps of CAPE at 00 and 12 UTC during winter and summer are presented in Figure 9. During winter, as shown in Figure 6, the N experiment presents higher values than D. At 00 UTC, the patterns are really similar for both WRF experiments. The main difference between them is observed at the western Atlantic coast of the IP, where higher CAPE values are obtained for D. At 12 UTC, the unstable air masses are found at the western coast of the IP in both simulations, and they extend further inland than at 00 UTC, particularly near the Tagus and Guadalquivir rivers. Again, the values are higher for N, but the pattern is similar in both experiments.

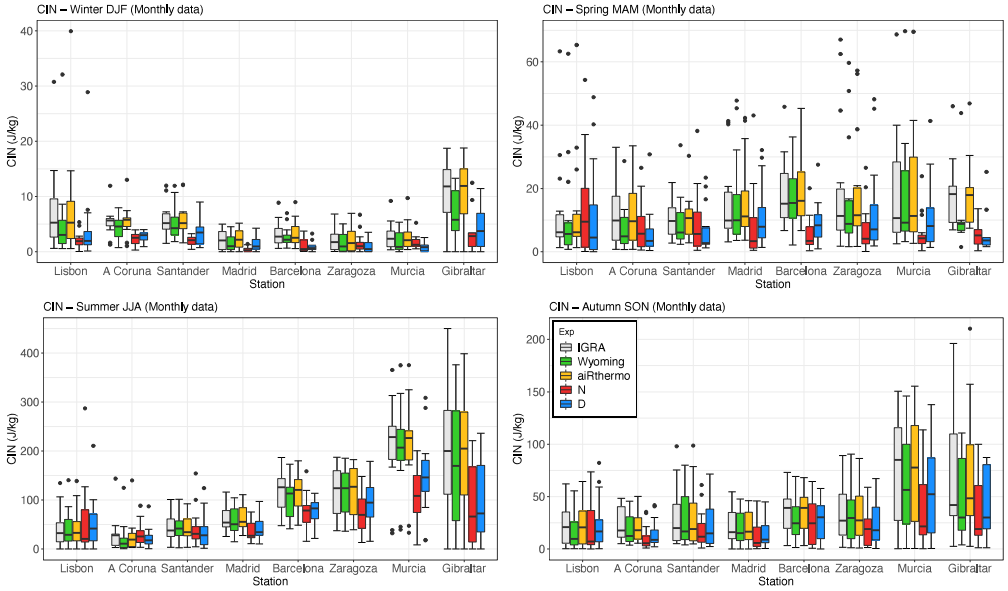

Figure 7. Same as Figures 5 and 6 but for CIN.

Compared to what is observed during winter, CAPE is higher during summer for the experiment including data assimilation. This is in agreement with the station analysis from previous Figure 6. At 00 UTC, the area with higher CAPE is observed in the northern and eastern IP, but particularly near the Mediterranean coast. However, at 12 UTC, this area with high values (over 250 J/kg) extends towards the interior and in the experiment including data assimilation it also covers the southern part of the Pyrenees. Additionally, high values are observed in most of the IP (except the southwestern corner for N), but particularly in the simulation including data assimilation. The patterns of CAPE obtained for both winter and summer are in agreement with those from Viceto et al. (2017), even if we differentiate between 00 and 12 UTC and we studied different periods (1986-2005 in their study, 2010-2014 in ours).

Finally, regarding CIN, the maps for the mean values at 00 and 12 UTC during winter and summer are presented in Figure 10. In reverse to what we found for CAPE, CIN is usually higher at 00 UTC than at 12 UTC (with the exception of Murcia in summer at 12 UTC), as it could be expected due to the stabilizing effect of nocturnal radiation. During winter, at 00 UTC, both simulations show small values over the IP, and only some high values are observed in the western and southwestern corners of the IP (and particularly for N) because of higher atmospheric stability due to surface cooling. At 12 UTC, the areas are confined to those coastal regions, but they also extend to the Mediterranean coast in the D experiment.

During summer, at 00 UTC, the most remarkable values are obtained in both simulations along the Ebro basin and near the Mediterranean coast. However, the CIN inland is higher for D. At 12 UTC, less inhibition is observed in the eastern valleys of the IP (with the exception of Murcia, which presents extremely high values of CIN). At the same time, an increase in the convective inhibition over the Guadalquivir basin is shown. The extension towards the interior is again higher for D (including data assimilation).

Comparing the results shown in Figures 5 and 6 (or in Figures 8 and 9), it seems that there is a discrepancy between the results for TT and CAPE since maximal values of these indices are not observed in the same regions. However, it must be taken into account that these results for CAPE (and CIN) are obtained from the entire series of 12-hourly values obtained during 2010-2014. Thus, Figures 6 and 7 (also Figures 9 and 10) must be compared in combination to the values of TT.

Additionally, atmospheric conditions can be analysed using TT, CAPE or CIN, but the relationship between them is not linear: the R^2 between TT and CAPE is below 0.2 and the R^2 between CAPE and CIN is below 0.1 for all the stations and seasons, particularly for stable or neutral atmospheres

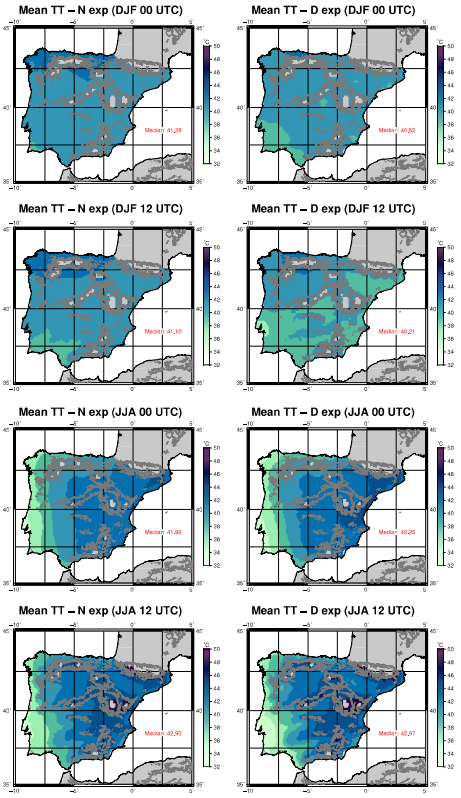

Figure 8. Spatial distribution of mean TT for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter (rows 1 and 2) and summer (rows 3 and 4) at 00 (rows 1 and 3) and 12 UTC (rows 2 and 4). The median value (°C) of each map is presented in the bottom right corner of the plots.

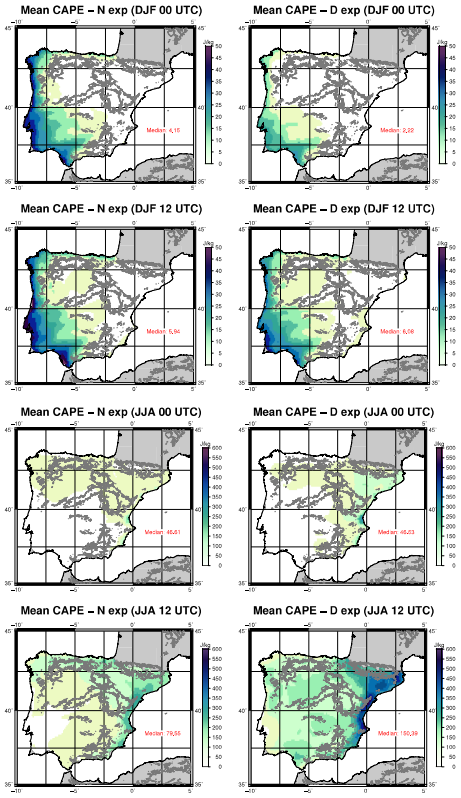

Figure 9. Same as Figure 8 but for CAPE.

(not shown here). Thus, since the calculation of CAPE and CIN takes into consideration the vertical profile of the atmosphere until the Level of Neutral Buoyancy, they should be considered more reliable than TT, which only takes into account two pressure levels.

Taking in consideration the information presented above, some clear patterns arise from these results. During winter, the areas with unstable air masses are located at the Atlantic coast of the IP, and the instability is larger during the afternoon as CIN is really high in those region until 12 UTC. However, during summer, the unstable areas are located to the north of the Mediterranean coast. CAPE is even larger during the afternoon in those regions, but some unstable areas can also appear before 12 UTC even if the inhibition is high during that period (CAPE is also high).

4 Conclusions

The main purpose of this paper is to evaluate the ability to simulate unstable conditions over the IP with two high-resolution simulations created with the state-of-the-art model WRF. One of these simulations is driven by the boundary conditions provided by ERA-Interim reanalysis (N experiment), while the second one presents the same configuration but including the additional 3DVAR data assimilation step every 6 hours (D experiment). Three parameters were evaluated: TT index, CAPE and CIN. All of them were calculated from the outputs of the model using the publicly available R package *aiRthermo*, also developed by the authors.

In order to validate these parameters, their values were downloaded from the University of Wyoming server for the

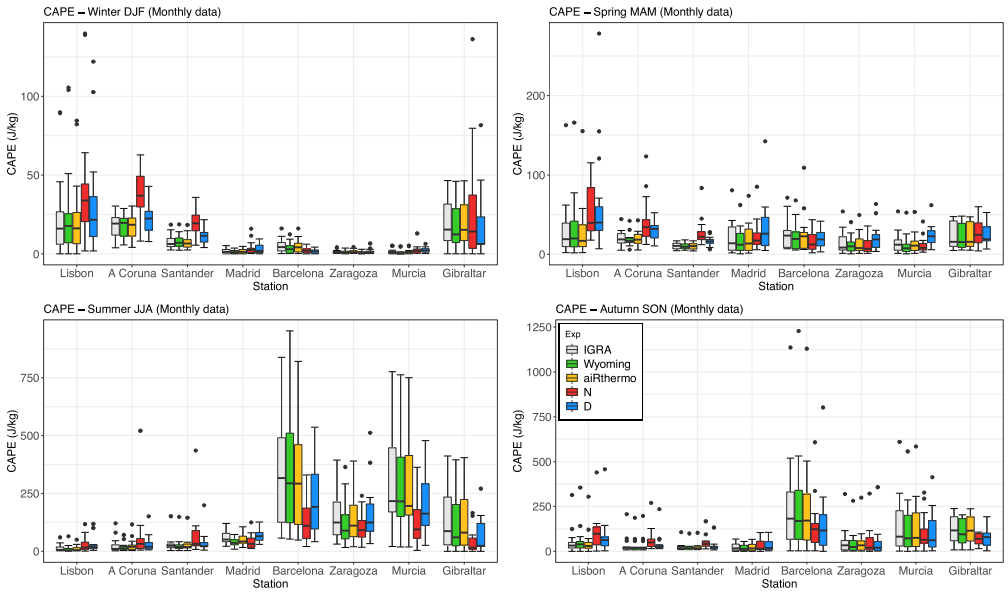

Figure 6. Same as Figure 5 but for CAPE.

only stations such as Lisbon, Madrid or Gibraltar present slightly higher values of CAPE than the other stations. In summer, only the stations located in the eastern and southern parts of the IP present remarkable values of CAPE. Particularly, the most active ones are those located in the Mediterranean coast (Barcelona and Murcia). Finally, during autumn, the regions with high CAPEs are extended towards the inland of the IP, such as Madrid and Zaragoza. During this season, some extreme events can reach values over 1000 J/kg over the Mediterranean coast. This feature was already observed by Siedlecki (2009).

Finally, the seasonal analysis for CIN is presented in Figure 7, and it highlights the stations where the inhibition is important. In general, Wyoming tends to underestimate the values of CIN in most of the stations and in every season, while aiRthermo is able to capture it. Both WRF simulations (but particularly the experiment without data assimilation) tend to underestimate the observed variability.

The values of CIN are smaller in winter and spring, and the maximum is observed in summer. During winter, CIN is higher in Gibraltar and in the Atlantic stations (Lisbon, A Coruna and Santander) than in the other stations from the IP. However, these values are small compared to those from other seasons. During spring, the values are higher than in winter, and similar values are observed in most of the stations (around 10 J/kg), with the exception of Barcelona where the CIN reaches values of 20 J/kg. In summer, the values are higher in all the stations, but particularly in those from the eastern and southern IP (Barcelona,

ing winter and summer are similar to those obtained by the regional analysis performed by Viceto et al. (2017). However, their values are comparable to those obtained by our experiment without data assimilation (N). In this case, the data assimilation (D) produces higher values, but much more realistic than the ones from those simulations (N) without it (according to Figure 6).

Finally, no linear relationships were found between the studied parameters as the R^2 is always below 0.2, independently of the station and the season. Thus, since the calculation of CAPE and CIN takes into consideration the vertical profile of the atmosphere until the Level of Neutral Buoyancy, they should be considered more reliable than TT, which only takes into account two pressure levels.

Data availability. These results can be reproduced using the post-processed outputs from the model available in https://doi.org/10.5281/zenodo.3611343.

Author contributions. The methodology and the software was developed by S.J.G.-R., S.C.-M. and J.S.; The conceptualization, preparation of datasets and analysis was carried out by all the authors; The original draft of the paper was written by S.J.G.-R., but all the authors took part in the edition and revision of it.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. S.J.G.R. is now funded by the Oeschger Centre for Climate Change Research (OCCR), but during his PhD he was supported by a FPI Predoctoral Research grant (MINECO, BES-2014-069977). This study was also supported through MINECO project CGL2016-76561-R from the Spanish Government (MINECO/ERDF, UE) and the grant GIU 17/002 from the University of the Basque Country. The computational resources were provided by I2BASQUE, and the authors thank the creators of WRF/ARW and WRFDA systems. Authors thank the anonymous reviewers for their comments, which have helped to improve the paper. Finally, most of the calculations were carried out with R (R Core Team, 2018), and the authors want to thank all the authors of the packages used for it.

References

Alexander, G. D. and Young, G. S.: The Relationship between EMEX Mesoscale Precipitation Feature Properties and Their Environmental Characteristics, Monthly Weather Review, 120, 554–564, https://doi.org/10.1175/1520-0493(1992)120<0554:TRBEMP>2.0.CO;2, 1992.

Angus, P., Rasmussen, S., and Seiter, K.: Short-term prediction of thunderstorm probability and intensity by screening observational and derived predictors, in: Reprints AMS 15th Conference on Severe Local Storms, Baltimore, MA, pp. 368–371, 1988.

Argüeso, D., Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Esteban-Parra, M. J., Dudhia, J., and Castro-Díez, Y.: Evaluation of WRF Parameterizations for Climate Studies over Southern Spain Using a Multistep Regionalization, Journal of Climate, 24, 5633–5651, https://doi.org/10.1175/JCLI-D-11-00073.1, 2011.

Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X.: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA, Bulletin of the American Meteorological Society, 93, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1, 2012.

Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A. J., and Xiao, Q. N.: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results, Monthly Weather Review, 132, 897–914, https://doi.org/10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2, 2004.

Blanchard, D. O.: Assessing the Vertical Distribution of Convective Available Potential Energy, Weather and Forecasting, 13, 870–877, https://doi.org/10.1175/1520-0434(1998)013<0870:ATVDOC>2.0.CO;2, 1998.

Brooks, H.: Severe thunderstorms and climate change, Atmospheric Research, 123, 129 – 138, https://doi.org/https://doi.org/10.1016/j.atmosres.2012.04.002, 2013.

Brooks, H. E., Lee, J. W., and Craven, J. P.: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data, Atmospheric Research, 67-68, 73 – 94, https://doi.org/https://doi.org/10.1016/S0169-8095(03)00045-0, 2003.

Correoso, J. F., Hernández, E., García-Herrera, R., Barriopedro, D., and Paredes, D.: A 3-year study of cloud-to-ground lightning flash characteristics of Mesoscale convective systems over the Western Mediterranean Sea, Atmospheric Research, 79, 89 – 107, https://doi.org/https://doi.org/10.1016/j.atmosres.2005.05.002, 2006.

Craven, J. P., Jewell, R. E., and Brooks, H. E.: Comparison between Observed Convective Cloud-Base Heights and Lifting Condensation Level for Two Different Lifted Parcels, Weather and Forecasting, 17, 885–890, https://doi.org/10.1175/1520-0434(2002)017<0885:CBOCCB>2.0.CO;2, 2002.

Dai, A.: Recent changes in the diurnal cycle of precipitation over the United States, Geophysical Research Letters, 26, 341–344, https://doi.org/10.1029/1998GL900318, 1999.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

DeRubertis, D.: Recent Trends in Four Common Stability Indices Derived from U.S. Radiosonde Observations, Journal of Climate, 19, 309–323, https://doi.org/10.1175/JCLI3626.1, 2006.

Diffenbaugh, N. S., Scherer, M., and Trapp, R. J.: Robust increases in severe thunderstorm environments in response to greenhouse forcing, Proceedings of the National Academy of Sciences, 110, 16 361–16 366, https://doi.org/10.1073/pnas.1307758110, 2013.

Doswell, C. A. and Rasmussen, E. N.: The Effect of Neglecting the Virtual Temperature Correction on CAPE Calculations, Weather and Forecasting, 9, 625–629, https://doi.org/10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2, 1994.

Doswell, C. A., Ramis, C., Romero, R., and Alonso, S.: A Diagnostic Study of Three Heavy Precipitation Episodes in the Western Mediterranean Region, Weather and Forecasting, 13, 102–124, https://doi.org/doi:10.1175/1520-0434(1998)013<0102:ADSOTH>2.0.CO;2, 1998.

Efron, B. and Gong, G.: A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, The American Statistician, 37, 36–48, https://doi.org/10.1080/00031305.1983.10483087, 1983.

Enno, S.-E., Sugier, J., Alber, R., and Seltzer, M.: Lightning flash density in Europe based on 10 years of ATDnet data, Atmospheric Research, 235, 104 769, https://doi.org/https://doi.org/10.1016/j.atmosres.2019.104769, 2020.

Eshel, G. and Farrell, B. F.: Thermodynamics of Eastern Mediterranean Rainfall Variability, Journal of the Atmospheric Sciences, 58, 87–92, https://doi.org/10.1175/1520-0469(2001)058<0087:TOEMRV>2.0.CO;2, 2001.

Esteban-Parra, M. J., Rodrigo, F. S., and Castro-Diez, Y.: Spatial and temporal patterns of precipitation in Spain for the period 1880–1992, International Journal of Climatology, 18, 1557–1574, https://doi.org/10.1002/(SICI)1097-0088(19981130)18:14<1557::AID-JOC328>3.0.CO;2-J, 1998.

Fernández, J., Montávez, J. P., Sáenz, J., González-Rouco, J. F., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2005JD006649, 2007.

Galway, J. G.: The lifted index as a predictor of latent instability, Bulletin of the American Meteorological Society, 37, 528–529, 1956.

Gascón, E., Merino, A., Sánchez, J., Fernández-González, S., García-Ortega, E., López, L., and Hermida, L.: Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe, Atmospheric Research, 164-165, 194 – 209, https://doi.org/https://doi.org/10.1016/j.atmosres.2015.05.012, 2015.

Gayà, M., Homar, V., Romero, R., and Ramis, C.: Tornadoes and waterspouts in the Balearic Islands: phenomena and environment characterization, Atmospheric Research, 56, 253 – 267, https://doi.org/https://doi.org/10.1016/S0169-8095(00)00076-4, 2001.

George, J. J.: Weather forecasting for aeronautics, Academic Press, San Diego, page 411, 1960.

González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., and Díaz de Argandoña, J.: Moisture balance over the Iberian Peninsula according to a regional climate model: The impact of 3DVAR data assimilation., Journal of Geophysical Research: Atmospheres, 123, 708–729, https://doi.org/10.1002/2017JD027511, 2018.

González-Rojí, S. J., Wilby, R. L., Sáenz, J., and Ibarra-Berastegi, G.: Harmonized evaluation of daily precipitation downscaled using SDSM and WRF+WRFDA models over the Iberian Peninsula, Climate Dynamics, 53, 1413–1433, https://doi.org/10.1007/s00382-019-04673-9, 2019.

González-Rojí, S. J., Sáenz, J., Díaz de Argandoña, J., and Ibarra-Berastegi, G.: Moisture Recycling over the Iberian Peninsula: The Impact of 3DVAR Data Assimilation, Atmosphere, 11, 19, https://doi.org/10.3390/atmos11010019, 2020.

Graf, M. A., Sprenger, M., and Moore, R. W.: Central European tornado environments as viewed from a potential vorticity and Lagrangian perspective, Atmospheric Research, 101, 31 – 45, https://doi.org/https://doi.org/10.1016/j.atmosres.2011.01.007, 2011.

Haklander, A. J. and Van Delden, A.: Thunderstorm predictors and their forecast skill for the Netherlands, Atmospheric Research, 67-68, 273 – 299, https://doi.org/https://doi.org/10.1016/S0169-8095(03)00056-5, 2003.

Holley, D. M., Dorling, S. R., Steele, C. J., and Earl, N.: A climatology of convective available potential energy in Great Britain, International Journal of Climatology, 34, 3811–3824, https://doi.org/10.1002/joc.3976, 2014.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather Review, 132, 103–120, https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

Huntrieser, H., Schiesser, H. H., Schmid, W., and Waldvogel, A.: Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland, Weather and Forecasting, 12, 108–125, https://doi.org/10.1175/1520-0434(1997)012<0108:COTAND>2.0.CO;2, 1997.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD009944, 2008.

Iturrioz, I., Hernández, E., Ribera, P., and Queralt, S.: Instability and its relation to precipitation over the Eastern Iberian Peninsula, Advances in Geosciences, 10, 45–50, https://doi.org/10.5194/adgeo-10-45-2007, 2007.

Jerez, S., Montavez, J. P., Gomez-Navarro, J. J., Jimenez, P. A., Jimenez-Guerrero, P., Lorente, R., and Gonzalez-Rouco, J. F.: The role of the land-surface model for climate change projections over the Iberian Peninsula, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD016576, 2012.

Johns, R. H. and Doswell, C. A.: Severe Local Storms Forecasting, Weather and Forecasting, 7, 588–612, https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2, 1992.

Jones, R. G., Murphy, J. M., and Noguer, M.: Simulation of climate change over europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries, Quarterly Journal of the Royal Meteorological Society, 121, 1413–1449, https://doi.org/10.1002/qj.49712152610, 1995.

Kunz, M.: The skill of convective parameters and indices to predict isolated and severe thunderstorms, Natural Hazards and Earth

System Sciences, 7, 327–342, https://doi.org/10.5194/nhess-7-327-2007, 2007.

Lee, J.: Tornado Proximity Soundings from the NCEP/NCAR Reanalysis Data, Ph.D. thesis, University of Oklahoma, 2002.

Letkewicz, C. E. and Parker, M. D.: Forecasting the Maintenance of Mesoscale Convective Systems Crossing the Appalachian Mountains, Weather and Forecasting, 25, 1179–1195, https://doi.org/10.1175/2010WAF2222379.1, 2010.

López, L., Marcos, J. L., Sánchez, J. L., Castro, A., and Fraile, R.: CAPE values and hailstorms on northwestern Spain, Atmospheric Research, 56, 147 – 160, https://doi.org/https://doi.org/10.1016/S0169-8095(00)00095-8, 2001.

Lucas, C., Zipser, E. J., and LeMone, M. A.: Convective Available Potential Energy in the Environment of Oceanic and Continental Clouds: Correction and Comments, Journal of the Atmospheric Sciences, 51, 3829–3830, https://doi.org/10.1175/1520-0469(1994)051<3829:CAPEIT>2.0.CO;2, 1994.

Marsh, P. T., Brooks, H. E., and Karoly, D. J.: Preliminary investigation into the severe thunderstorm environment of Europe simulated by the Community Climate System Model 3, Atmospheric Research, 93, 607 – 618, https://doi.org/https://doi.org/10.1016/j.atmosres.2008.09.014, 2009.

McNulty, R. P.: Severe and Convective Weather: A Central Region Forecasting Challenge, Weather and Forecasting, 10, 187–202, https://doi.org/10.1175/1520-0434(1995)010<0187:SACWAC>2.0.CO;2, 1995.

Miller, R. C.: Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central, vol. 200, AWS Technical Report, 1975.

Mohr, S., Kunz, M., and Geyer, B.: Hail potential in Europe based on a regional climate model hindcast, Geophysical Research Letters, 42, 10,904–10,912, https://doi.org/10.1002/2015GL067118, 2015.

Molina, D. S., Fernández-González, S., González, J. C. S., and Oliver, A.: Analysis of sounding derived parameters and application to severe weather events in the Canary Islands, Atmospheric Research, 237, 104 865, https://doi.org/https://doi.org/10.1016/j.atmosres.2020.104865, 2020.

Moncrieff, M. W.: A theory of organized steady convection and its transport properties, Quarterly Journal of the Royal Meteorological Society, 107, 29–50, https://doi.org/10.1002/qj.49710745103, 1981.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, Boundary-Layer Meteorology, 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

Owens, R. G. and Hewson, T.: ECMWF Forecast User Guide, ECMWF, https://doi.org/10.21957/m1cs7h, 2018.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, Monthly Weather Review, 120, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, 1992.

Piper, D. and Kunz, M.: Spatiotemporal variability of lightning activity in Europe and the relation to the North Atlantic Oscillation teleconnection pattern, Natural Hazards and Earth System Sciences, 17, 1319–1336, https://doi.org/10.5194/nhess-17-1319-2017, 2017.

Púčik, T., Groenemeijer, P., Rýva, D., and Kolář, M.: Proximity Soundings of Severe and Nonsevere Thunderstorms in Central Europe, Monthly Weather Review, 143, 4805–4821, https://doi.org/10.1175/MWR-D-15-0104.1, 2015.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

Rädler, A. T., Groenemeijer, P., Faust, E., and Sausen, R.: Detecting Severe Weather Trends Using an Additive Regressive Convective Hazard Model (AR-CHaMo), Journal of Applied Meteorology and Climatology, 57, 569–587, https://doi.org/10.1175/JAMC-D-17-0132.1, 2018.

Rädler, A. T., Groenemeijer, P. H., Faust, E., Sausen, R., and Púčik, T.: Frequency of severe thunderstorms across Europe expected to increase in the 21st century due to rising instability, Npj Climate and Atmospheric Science, 2, 30, https://doi.org/10.1038/s41612-019-0083-7, 2019.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, Journal of Climate, 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Riemann-Campe, K., Fraedrich, K., and Lunkeit, F.: Global climatology of Convective Available Potential Energy (CAPE) and Convective Inhibition (CIN) in ERA-40 reanalysis, Atmospheric Research, 93, 534 – 545, https://doi.org/https://doi.org/10.1016/j.atmosres.2008.09.037, 2009.

Rodríguez-Puebla, C., Encinas, A. H., Nieto, S., and Garmendia, J.: Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula, International Journal of Climatology, 18, 299–316, https://doi.org/10.1002/(SICI)1097-0088(19980315)18:3<299::AID-JOC247>3.0.CO;2-L, 1998.

Romero, R., Ramis, C., and Guijarro, J.: Daily rainfall patterns in the Spanish Mediterranean area: an objective classification, International Journal of Climatology, 19, 95–112, https://doi.org/10.1002/(SICI)1097-0088(199901)19:1<95::AID-JOC344>3.0.CO;2-S, 1999.

Romero, R., Gayà, M., and Doswell, C. A.: European climatology of severe convective storm environmental parameters: A test for significant tornado events, Atmospheric Research, 83, 389 – 404, https://doi.org/https://doi.org/10.1016/j.atmosres.2005.06.011, 2007.

Rummukainen, M.: State-of-the-art with regional climate models, Wiley Interdisciplinary Reviews: Climate Change, 1, 82–96, https://doi.org/10.1002/wcc.8, https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.8, 2010.

Sáenz, J., González-Rojí, S. J., Carreno-Madinabeitia, S., and Ibarra-Berastegi, G.: Analysis of atmospheric thermodynamics using the R package aiRthermo, Computers & Geosciences, 122, 113 – 119, https://doi.org/https://doi.org/10.1016/j.cageo.2018.10.007, 2019.

Showalter, A. K.: A Stability Index for Thunderstorm Forecasting, Bulletin of the American Meteorological Society, 34, 250–252, 1953.

Siedlecki, M.: Selected instability indices in Europe, Theoretical and Applied Climatology, 96, 85–94, https://doi.org/10.1007/s00704-008-0034-4, 2009.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, Journal of Geophysical Research: Atmospheres, 118, 1716–1733, https://doi.org/10.1002/jgrd.50203, 2013.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note NCAR/TN-475+STR, https://doi.org/10.5065/D68S4MVH, 2008.

Taszarek, M., Brooks, H. E., and Czernecki, B.: Sounding-Derived Parameters Associated with Convective Hazards in Europe, Monthly Weather Review, 145, 1511–1528, https://doi.org/10.1175/MWR-D-16-0384.1, 2017.

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

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, vol. 1115, 2004.

Tiedtke, M.: Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Tullot, I. F.: Climatología de España y Portugal, vol. 76, Universidad de Salamanca, 2000.

Ulazia, A., Sáenz, J., Ibarra-Berastegui, G., González-Rojí, S. J., and Carreno-Madinabeitia, S.: Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean, Applied Energy, 208, 1232 – 1245, https://doi.org/https://doi.org/10.1016/j.apenergy.2017.09.030, 2017.

Ulazia, A., Ibarra-Berastegi, G., Sáenz, J., Carreno-Madinabeitia, S., and González-Rojí, S. J.: Seasonal Correction of Offshore Wind Energy Potential due to Air Density: Case of the Iberian Peninsula, Sustainability, 11, https://doi.org/10.3390/su11133648, 2019.

van Delden, A.: The synoptic setting of thunderstorms in western Europe, Atmospheric Research, 56, 89 – 110, https://doi.org/https://doi.org/10.1016/S0169-8095(00)00092-2, 2001.

Viceto, C., Marta-Almeida, M., and Rocha, A.: Future climate change of stability indices for the Iberian Peninsula, International Journal of Climatology, 37, 4390–4408, https://doi.org/10.1002/joc.5094, 2017.

Virts, K. S., Wallace, J. M., Hutchins, M. L., and Holzworth, R. H.: Highlights of a New Ground-Based, Hourly Global Lightning Climatology, Bulletin of the American Meteorological Society, 94, 1381–1391, https://doi.org/10.1175/BAMS-D-12-00082.1, 2013.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, vol. 100, Academic Press, 2011.

Xu, G., Xi, B., Zhang, W., Cui, C., Dong, X., Liu, Y., and Yan, G.: Comparison of atmospheric profiles between microwave radiometer retrievals and radiosonde soundings, Journal of Geophysical Research: Atmospheres, 120, 10,313–10,323, https://doi.org/10.1002/2015JD023438, 2015.

Ye, B., Del Genio, A. D., and Lo, K. K.-W.: CAPE Variations in the Current Climate and in a Climate Change, Journal of Climate, 11, 1997–2015, https://doi.org/10.1175/1520-0442-11.8.1997, 1998.

Zhang, C., Wang, Y., and Hamilton, K.: Improved Representation of Boundary Layer Clouds over the Southeast Pacific in ARW-WRF Using a Modified Tiedtke Cumulus Parameterization Scheme, Monthly Weather Review, 139, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1, 2011.

Zheng, D., Van Der Velde, R., Su, Z., Wen, J., and Wang, X.: Assessment of Noah land surface model with various runoff parameterizations over a Tibetan river, Journal of Geophysical Research: Atmospheres, 122, 1488–1504, https://doi.org/10.1002/2016JD025572, 2017.

Appendix A: Supplementary figures

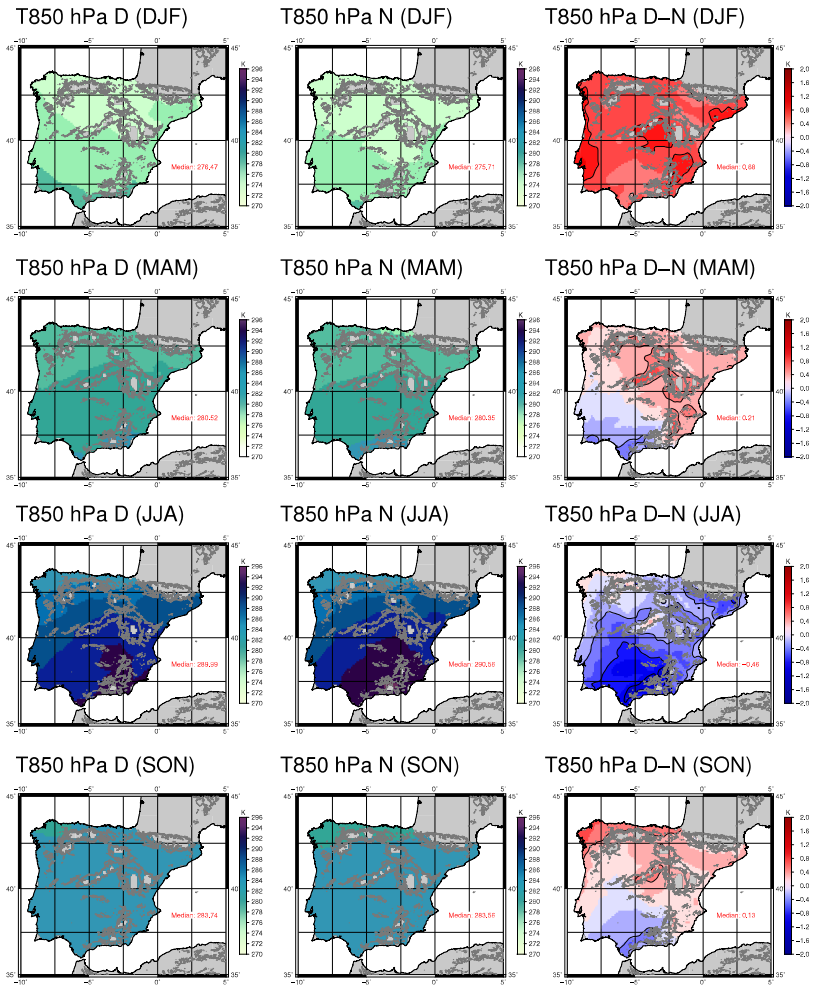

Figure A1. Spatial distribution of mean temperature at 850 hPa for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter and summer. The third columns presents the difference between both WRF experiments (D-N). The median value (K) is in the bottom right corner of the plots.

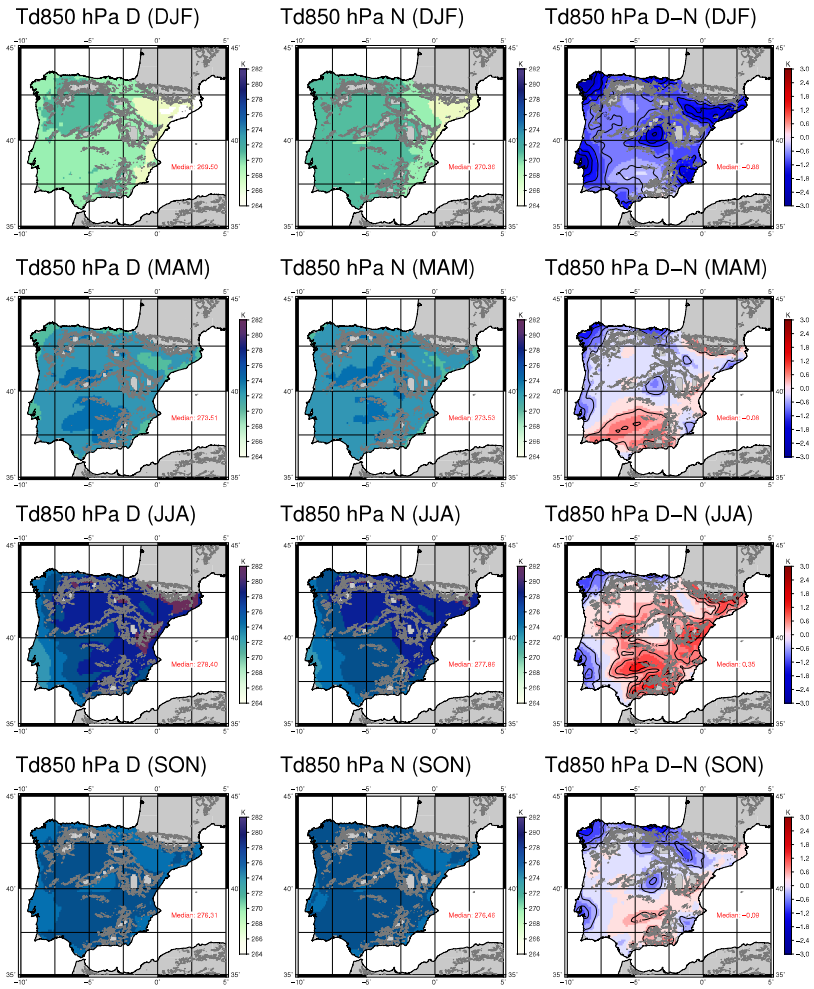

Figure A2. Same as Figure A1 but for dew point temperature at 850 hPa.

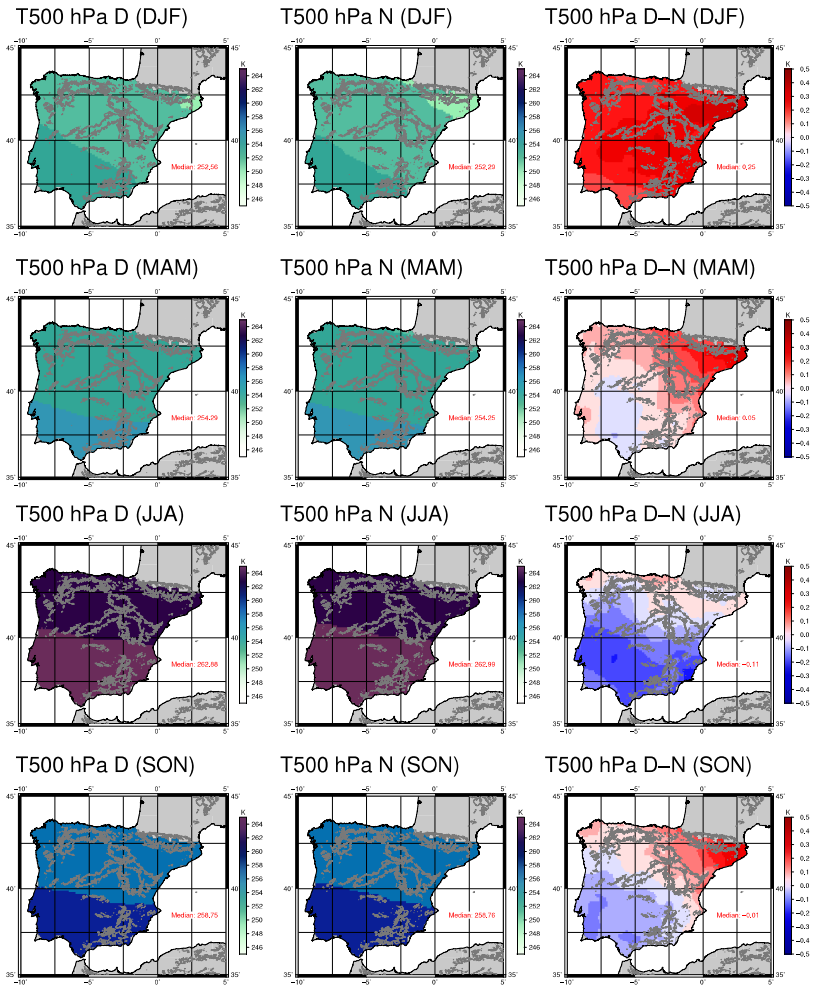

Figure A3. Same as Figures A1 and A2 but for temperature at 500 hPa.

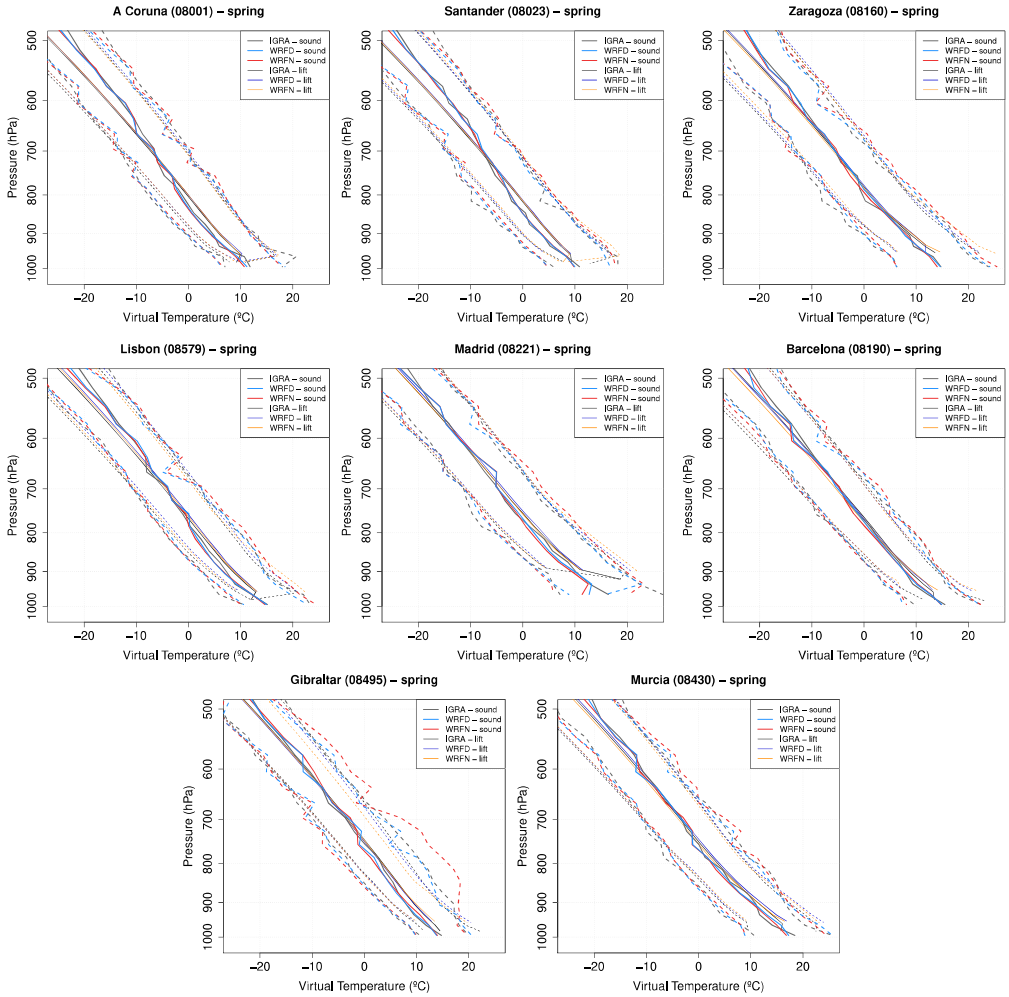

Figure A4. Vertical profiles of virtual temperature for the sounding levels and for the lifted parcel during spring. The dashed lines represent the 5 and 95 percentiles, and the solid lines the median.

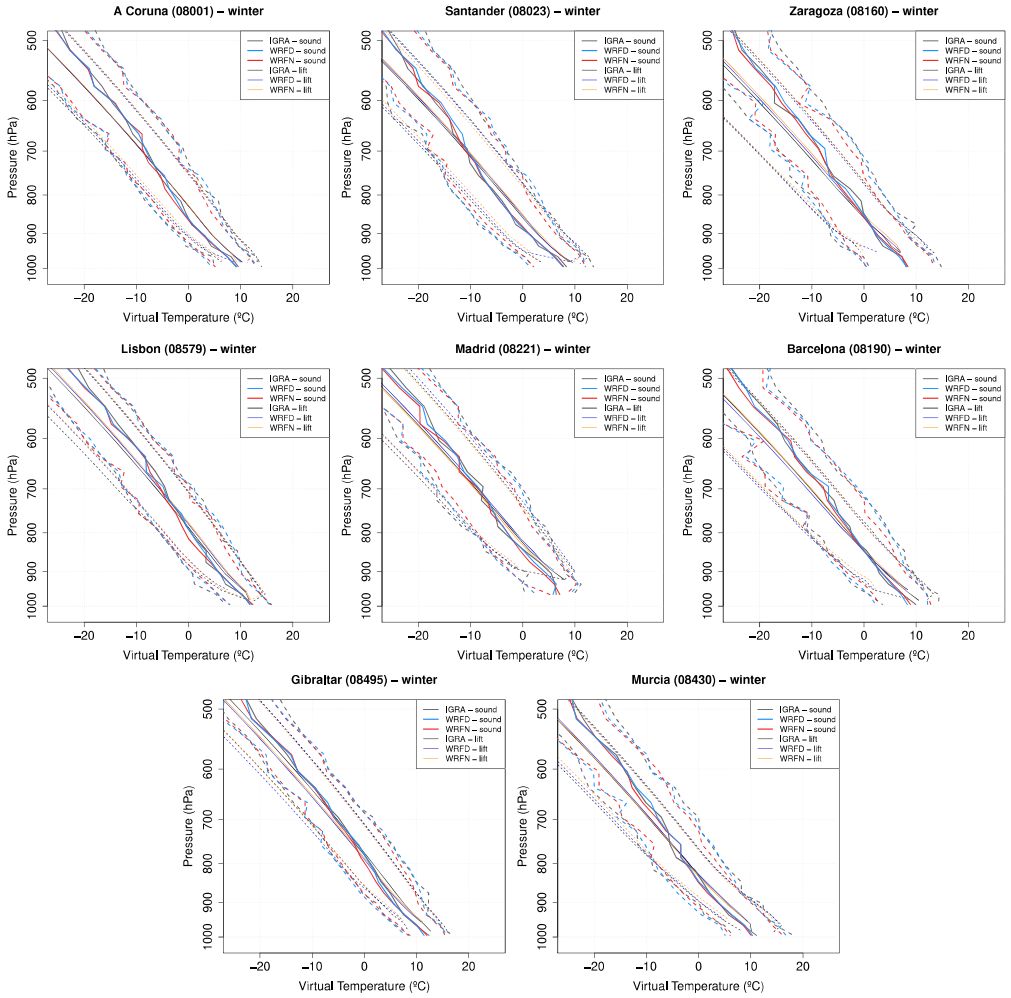

Figure A5. Same as Figure A4 but for winter.

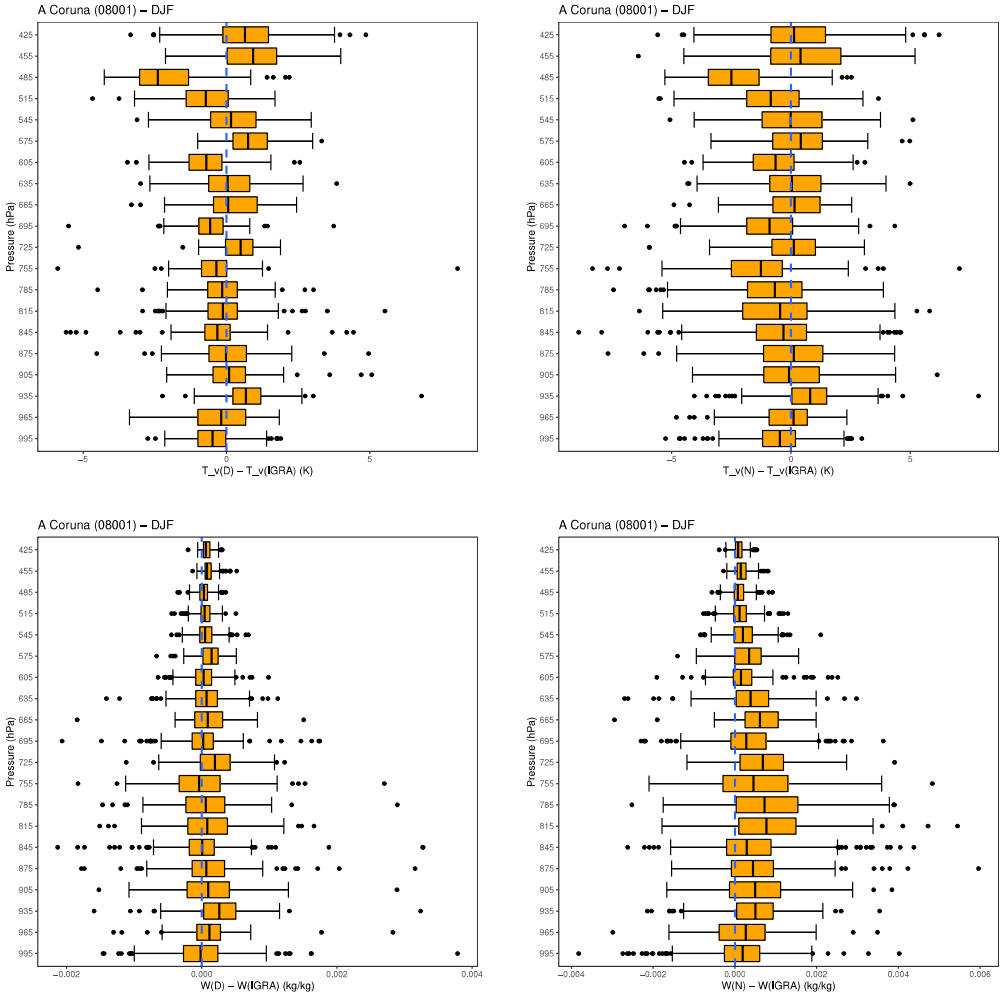

Figure A6. Vertical profile of the differences of virtual temperature (K, upper panels) and mixing ratio (kg/kg, lower panels) in A Coruna during winter from the D simulation minus IGRA (left column) and the N simulation minus IGRA (right column).

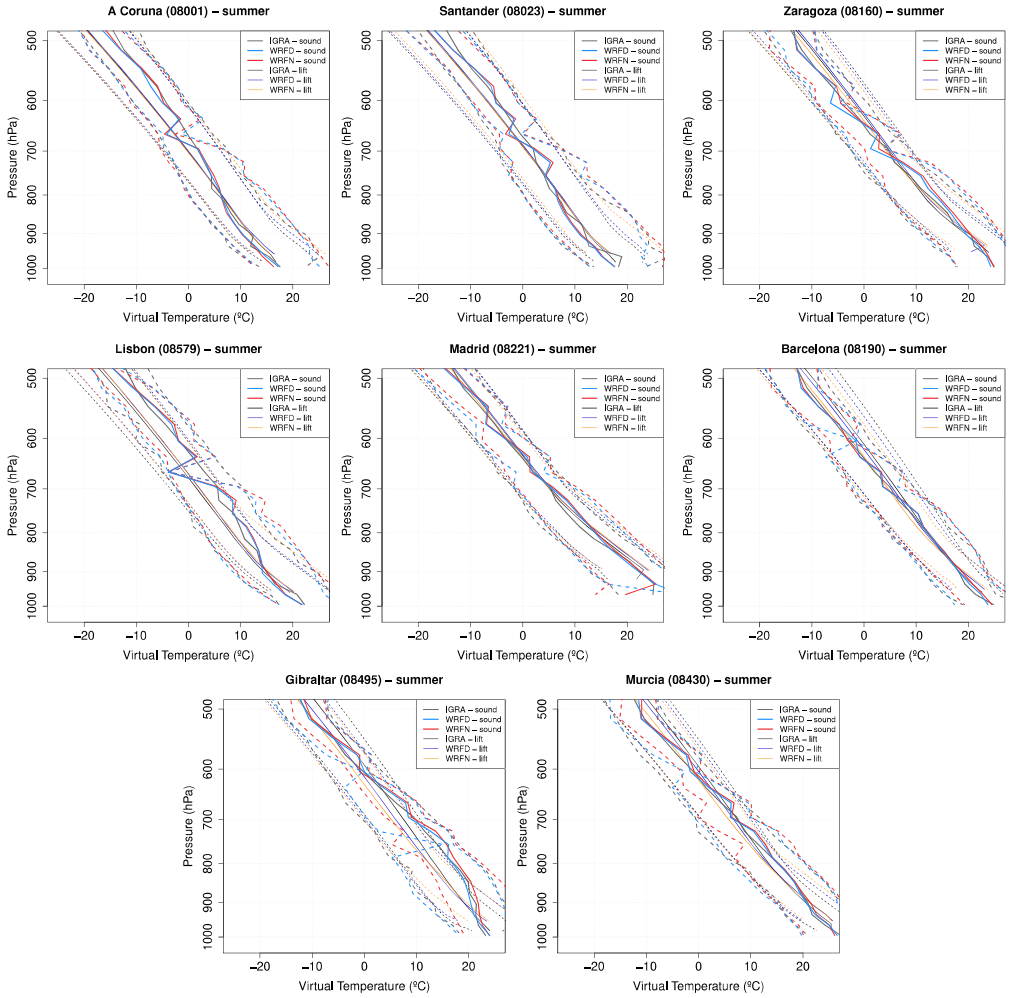

Figure A7. Same as Figures A4 and A5 but for summer.

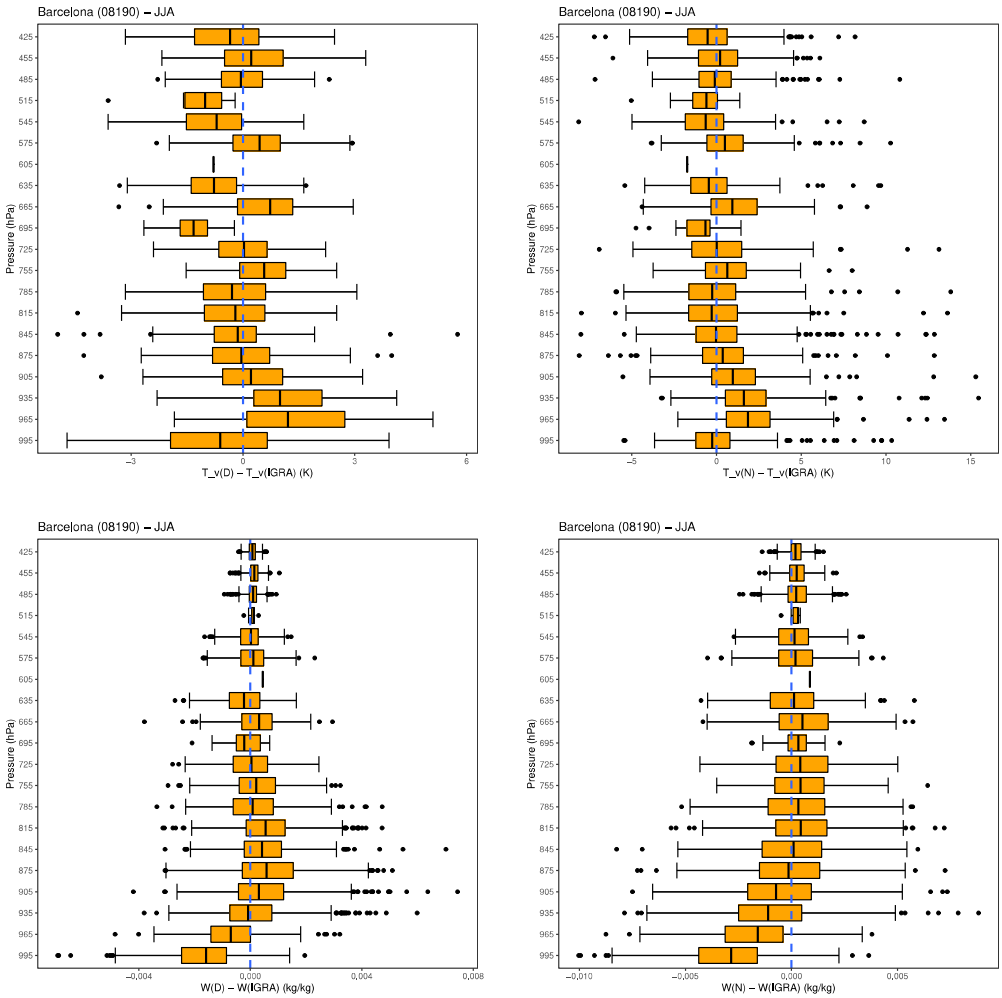

Figure A8. Same as Fig A6, but for Barcelona during summer.