The Sailor diagram. An extension of Taylor's diagram to two-dimensional vector data.

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

[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, Parque Tecnológico de Álava, Albert Einstein 28, E-01510 Vitoria-Gasteiz, Spain
[4]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Europa Plaza 1, 20018-Donostia, Spain
[5]Oeschger Centre for Climate Change Research, University of Bern, 3010 Bern, Switzerland
[6]Climate and Environmental Physics, University of Bern, 3010 Bern, Switzerland
[7]Nuclear Engineering and Fluid Mechanics Dept., Escuela de Ingeniería de Bilbao, Plaza Ingeniero Torres Quevedo 1, 48013-Bilbao, Spain
[8]Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Otaola etorbidea 29, 20600-Eibar, Spain

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

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 idea of extending the Taylor diagram to two dimensional quantities such as currents, wind, or horizontal fluxes of water vapour, salinity, energy and other geophysical variables. 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

5 separated in a part corresponding to the bias and the relative rotation of the empirical orthogonal functions of the data. We test the performance of this new diagram to identify the differences amongst the reference dataset and the different model outputs by using examples with wind, current, vertically integrated moisture transport and wave energy flux time series. An alternative setup is shown in the last examples, where the spatial average of surface wind in the Northern and Southern Hemispheres according to different reanalyses and realizations of CMIP5 models are compared. The examples of use of the Sailor diagram

10 presented 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 also to easily retrieve the individual diagnostics of the different components of the mean squared error.

1 Introduction

It has long ago been recognized that visual tools allow an easy analysis of different properties of datasets, and 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 the scientific inquiry, see, for instance Peircean diagrammatic thinking (Dörfler, 2005). Furthermore, the visualization via diagrammatic representations does not constitute only a way of interpretation. Peircean 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) or 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 centered root mean squared error. In the Taylor diagram, every model tested is represented by a point in the diagram and visual inspection allows 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 diagnose 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) and stress the idea of flexibility associated to the Taylor diagram. Finally, since observational data also suffer from errors, an

estimation of the relevance of these observational errors in different observational datasets can also be achieved by cheking alternative observational datasets against the same reference dataset as if they were models too (Fernández et al., 2007). Thus,
50 the dispersion amongst observational datasets yields an estimate of the uncertainty of the observations.

Pearson's correlation coefficient between two scalars plays a fundamental role in the design of Taylor's diagram, and this imposes a limit to our ability to extend Taylor's diagram to vector quantities, since there does not exist a single universally accepted definition of the correlation coefficient in two dimensions.

Jupp and Mardia (1980) recognized that any multivariate definition of a correlation coefficient equivalent to Pearson's one
55 must be invariant to rotation, be close to zero for independent datasets, smaller or equal than a constant and equal to the 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) had already 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
60 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 vector sets which are perfectly correlated can be related by means of an orthogonal transformations. The vectors in this case are assumed to share a common center 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
65 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. The definitions of a two-dimensional correlation coefficient for vector datasets in the second approach are based on functions which satisfy some desirable properties, such as, for instance, the invariance of the correlation
70 to the rotation of the original datasets, 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 day, 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
75 and meteorology such as Kundu (1976) and 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 data sets by Breaker et al. (1994) and Cosoli et al. (2008), for instance.

As shown before, there is not a universally accepted way of defining two-dimensional correlation coefficient. Thus, the structure of the Taylor diagram cannot be directly extended to two dimensions following the same derivation used by Taylor
80 (2001) in his original paper.

However, a diagram similar to the one by Taylor (2001) would be an interesting addition to the set of verification tools available for modellers when comparing vectorial quantities of model output with observations. For instance, Lee et al. (2013)

presented a comparison of CMIP3, CMIP5, reanalysis and satellite based estimations of wind stress and they used the mean of the Taylor diagrams for the zonal and meridional components of the wind stress as a way to apply Taylor diagrams for vector

85   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 Figure 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 HF-Radar (Lorente et al., 2015). It also appears in the evaluation of global climate models (Martin

90   et al., 2011) using zonal and meridional components of wind speed. This result was also noticed in an analysis of moisture fluxes (Ibarra-Berastegi et al., 2011). A last example appears when wind stress components are analyzed (Chaudhuri et al., 2013), to name a few. 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.

95   In a recent paper, Xu et al. (2016) proposed a new method to overcome the defficiencies 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 Vector Field Evaluation (VFE) diagram, constructed from both components of the vectors which appear in the vector datasets that are used for the verification. In order to arrive to the same structure of the Taylor diagram commonly used for scalar variables, the

100   authors apply some approximations and normalizations to the original two-dimensional vector observations, thus leading to approximate estimations of errors.

However, in the original paper by Crosby et al. (1993), the authors show (their Figure 3) 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. Thus, we have decided to follow a new approximation which does not lead to the common Taylor

105   diagram used for scalars, but gives more information about the structure of the two-dimensional errors between vector quantities involved in the verification of a vector quantity derived from a model with its observational counterpart (reference dataset). This is the rationale which leads us to base our definition in a full use of the two-dimensional structure of the mean squared error (MSE) 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). However, we hope that our diagram will be considered a valuable

110   contribution to the set of techniques used for the evaluation of models, as it is exact (no approximations are needed) and it fully visually explores other properties of the error between the vector datasets, such as the relative rotation of the major axes of variability or the underestimation (or overestimation) along each principal axis of the covariance matrix. 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.

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

115   (properties of geophysical fluid dynamics that sailors need to know about) and because this name is very similar to the original Taylor diagram. Thus, it 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 games with words.

Section 2 presents the datasets that we have used as examples of application of our Sailor diagram. Section 3 explains the methodology that we follow to build the two-dimensional diagram. Results are included in Section 4, followed by some
120 concluding remarks in Section 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 variabless (wind or ocean currents) to additional postprocessed quantities such as vertically integrated moisture transports or wave energy fluxes.

125 ### 2.1 Wind data

The offshore wind data validation used as our first example of a Sailor diagram has been conducted from 01/01/2009 to 01/01/2015 using wind data (zonal and meridional components) from five sources (Ulazia et al., 2017):

1. A Weather Research and Forecasting Model (WRF) simulation around the Iberian Peninsula without data assimilation (experiment N).

130 2. A WRF simulation in the same area using 6-hourly 3DVAR data assimilation (experiment D). These WRF N and D simulations are thoroughly described and compared with observations in the previous article about the wind energy potential on the West Mediterranean by some of the authors (Ulazia et al., 2017).

3. The ERA-Interim reanalysis (ERAI henceforth). ERAI data (Dee et al., 2011) were downloaded from the European Centre for Medium-Range Weather Forecasts (ECMWF) with 0.75° spatial and six hourly temporal resolutions. These
135 data were also used to nest the two N and D WRF runs.

4. Wind analysis data from the second version of Cross-Calibrated Multi-Platform (CCMPv2). CCMPv2 processing combines via variational methods radiometer data, QuikSCAT and ASCAT scatterometer wind fields, moored buoy wind data, and reanalysis wind fields, to produce 6-hourly maps at a 0.25 degree grid (Hoffman et al., 2003; Atlas et al., 2011). In this paper, CCMPv2 fully assimilated Level 3 wind fields are used.

140 5. The previous four sources will be validated against in-situ observations provided by the buoy in Dragonera, near the Balearic Islands. This buoy records wind speed at a height of three meters above the sea level and other data related to the sea-state, and is operated by the Spanish State Ports Authority (Puertos del Estado) (P.P.E., 2015). This buoy is located at longitude 2.10° E, latitude 39.56° N and the distance from this location to the closest gridpoint of both WRF runs (D and N) is 8.9 km. It is located 24.7 km far away from the closest ERA Interim point.

2.2 Ocean currents

Three different data sources of ocean surface horizontal vectorial currents are also compared within the framework of the proposed methodology. They cover the Bay of Biscay area and include in-situ observations from a deep-water buoy, remotely sensed surface HF-Radar currents and an ocean modelling product. Observational products, both in-situ buoy (named DONOS-TIA buoy) and remotely sensed radar currents, belong to the Basque Meteorological Agency (EUSKALMET) and were obtained from https://www.euskoos.eus. These data follow Copernicus data standards and include data quality information. Both observational platforms provide hourly data that is punctual in the case of the buoy (approx. location 43.6º N and 2.0º W). In the case of the HF Radar dataset, it consists of a gridded dataset which covers the corner of the Bay of Biscay (approx. location 43.5-44.7º N and 3.2-1.3º W) with 5 km spatial resolution. Technical details about these observational systems and the validation of the data can be found in additional references (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`). This is a realization of the high resolution global analysis and forecasting system PSY4V3R1 that uses version 3.1 of the NEMO ocean model (Madec and the NEMO team, 2008; Lellouche et al., 2018). Zonal and meridional components of the surface level hourly mean currents ($0.083°$ spatial resolution) of this product were extracted for the Bay of Biscay area ($43° - 45°$ N and $5° - 1°$ W).

The processing chain of current data involved:

1. Filtering buoy observations using data-quality indicators and identifying latitude and longitude of buoy observations.

2. Determining the closest spatial grid-point in HF-Radar and model spatial grids using the geolocation information.

3. Extraction of the corresponding current information whenever it is concurrently available in all three data sources.

The result of this processing is a set of six hourly variables (u buoy,v buoy,u radar,v radar,u model,v model) that covers parts of years 2017 and 2018, with as many as 8319 hourly observations (47.5 % of total).

2.3 Vertically integrated water vapour transports

The vertically integrated water vapour transport is a vectorial variable that represents the overall transport of water vapour when this quantity is integrated accross the whole vertical dimension of the atmosphere. It must be calculated once the moisture and both components of the wind are determined at different vertical levels of the atmosphere. In order to test the performance of the methodology presented in this article, both components of the vertically integrated moisture flux have been calculated or downloaded from different sources. The sounding data for A Coruña (Station ID 08001, longitude $-8.41°$ E and latitude $43.36°$ N) were downloaded from the server at the University of Wyoming with a temporal resolution of 12 hours for the period 2010-2014. Both components of moisture transport were calculated from radiosonde data taking into account the pressure, temperature and mixing ratio measured at those locations. The components were obtained after the calculation of the vertical

integration of wind and specific humidity measured at several pressure levels. In this case, the specific humidity was obtained from the reported mixing ratio in the soundings using the R package *aiRthermo* (Sáenz et al., 2019).

Both components of vertically integrated moisture transport from the ERA-Interim (ERAI) were downloaded by means of the Meteorological Archival and Retrieval System (MARS) repository at ECMWF with a spatial resolution of 0.75°, while 180 vertically integrated in the original vertical levels of the ECMWF model. Then, the nearest point to A Coruña was taken from the ERA-Interim grid.

Both moisture transport components were also calculated using the moisture and wind data from two simulations created with the WRF model over the Iberian Peninsula at a spatial resolution of 15 km with 51 vertical η levels up to 20 hPa. One of them (WRF D experiment) includes the 6-hourly 3DVAR data assimilation (experiment D described before). The second 185 one (WRF N) has been run without data assimilation (experiment N). The analysis of the moisture balance over the Iberian Peninsula simulated by those experiments has been already carried out in González-Rojí et al. (2018). The components of the moisture transport were calculated by means of the vertical integration of the specific humidity and the zonal and meridional winds over the original 51 η levels of the WRF model. The nearest neighbour point in WRF's grid was compared with the sounding data.

2.4 Wave energy fluxes

Wave energy flux (WEF) represents -in general terms- the hydraulic energy held by waves. The wave energy flux is a vectorial magnitude with a direction corresponding to that of the incident waves at a given location. Having accurate estimations of WEF is a key aspect for the development of wave energy converters (WEC), intended to transform WEF into electric power. WEF is stated in kW per metre of wave front and its values are typically provided at reanalyses or buoy data repositories as a magnitude 195 plus a direction. In spite of being an oceanic variable, atmospheric notation is usually employed and the direction given usually corresponds to the direction waves come from. Some papers by the authors have focused on developing short-term forecasting models using different techniques, like ERAI physics-based forecasts, linear regression or random forests (Ibarra-Berastegi et al., 2015, 2016). From the data presented in those papers, the results corresponding to the buoy placed at Estaca de Bares ($-7.67°$ E and $44.12°$ N) are used as an application example of the diagram proposed in this paper.

2.5 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 diagram proposed in this paper can also be applied in this case, some reanalyses datasets 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 205 and ERA5 (Hersbach et al., 2018). In all those cases, we have analyzed 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 ($2.5° \times 2.5°$).

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. This case example uses the time average of surface wind obtained from ERA5 covering the period 1979-2015 over the Southern Hemisphere as the reference dataset. In order to check the behaviour of the diagram when analyzing ensembles of multimodels, we have also downloaded surface wind fields from the CMIP5 repository, historical forcing experiment. The first model ensemble (four realizations) is the IPSL model, developed at the Institute Pierre-Simon Laplace (Dufresne et al., 2013), whilst the second one (three realizations) is the MIROC model (Watanabe et al., 2010). All the models and ERA5 reanalysis gridded fields have been bilinearly interpolated to a common $1.25° \times 1°$ regular longitude-latitude grid. CMIP data also cover the period 1979-2015.

3 Methodology

In this section, we present the derivation of the 2×2 squared-error matrix that is on 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 observations and model results. Thus, these decompositions will be presented using the same notation so that the theoretical basis of the diagram is completely described and the paper is self-contained. Subsection 3.1 describes the decomposition of the matrix $\mathbf{U}$ corresponding to the reference dataset (observations) in its EOFs. Subsection 3.2 represents the decomposition of the matrix $\mathbf{V}$ corresponding to one of the models which are being compared to observations. Later, the expansion of the $\mathbf{V}$ matrix corresponding to the model is expressed as a rotation from the the EOFs derived from observations (Subsection 3.3).

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 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. The observational dataset is formed by the two-dimensional measurements $\mathbf{u}_i$, with $i = 1 \ldots N$ arranged as rows in an $N \times 2$ matrix $\mathbf{U}$. The average $\bar{\mathbf{u}}$ of the $\mathbf{u}_i$ time series will be repeated as constant rows in an $N \times 2$ matrix $\bar{\mathbf{U}}$. The $\mathbf{U}$ matrix can be expressed by means of the empirical orthogonal functions of the original vector data by using the expression

$$\mathbf{U} = \bar{\mathbf{U}} + \mathbf{P}_u^* \mathbf{\Sigma}_u \mathbf{E}_u^T = \bar{\mathbf{U}} + \mathbf{P}_u \mathbf{E}_u^T, \tag{1}$$

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. Thus, the anomalies of wind are computed as

$$\mathbf{U} - \bar{\mathbf{U}} = \mathbf{P}_u^* \mathbf{\Sigma}_u \mathbf{E}_u^T = \mathbf{P}_u \mathbf{E}_u^T \tag{2}$$

and the corresponding principal components

$$\mathbf{P}_u = \mathbf{P}_u^* \mathbf{\Sigma}_u = \left(\mathbf{U} - \bar{\mathbf{U}} \right) \mathbf{E}_u. \tag{3}$$

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

$$\mathbf{\Sigma}_u = \begin{pmatrix} \sigma_{1u} & 0 \\ 0 & \sigma_{2u} \end{pmatrix}. \tag{4}$$

Due to the fact that the rotation matrix is always full rank (in a two-dimensional space, given enough samples), the 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:

$$\mathbf{E}_u = \begin{pmatrix} \cos\theta_u & -\sin\theta_u \\ \sin\theta_u & \cos\theta_u \end{pmatrix}. \tag{5}$$

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

$$\mathbf{P}_u \mathbf{P}_u^T = \mathbf{\Sigma}_u^2, \tag{6}$$

$$\mathbf{P}_u^* \mathbf{P}_u^{*T} = \mathbb{1} \tag{7}$$

and

$$\mathbf{E}_u \mathbf{E}_u^T = \mathbf{E}_u^T \mathbf{E}_u = \mathbb{1}. \tag{8}$$

3.2 Decomposition of V in its EOFs

Following the previous notation, we also consider a time series (or time-averaged constant field over N points in a grid) of simulated wind (current, wave energy flux...) at the same observatory (or the closest grid point) formed by the two-dimensional measurements $\mathbf{v}_i$, with $i = 1 \ldots N$ arranged as rows in an $N \times 2$ matrix $\mathbf{V}$. As in Subsection 3.1, its average $\bar{\mathbf{v}}$ is arranged as constant rows in an $N \times 2$ matrix $\bar{\mathbf{V}}$. The $\mathbf{V}$ matrix can be expressed as already shown in subsection 3.1 for observations by means of the empirical orthogonal functions of the simulated wind data

$$\mathbf{V} = \bar{\mathbf{V}} + \mathbf{P}_v^* \mathbf{\Sigma}_v \mathbf{E}_v^T = \bar{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_v^T, \tag{9}$$

with equivalent interpretations and equal ranks as before for $\mathbf{P}_v^*$, $\mathbf{\Sigma}_v$, $\mathbf{E}_v$ and $\mathbf{P}_v = \mathbf{P}_v^* \mathbf{\Sigma}_v$. Similarly, the corresponding anomalies can be computed as

$$\mathbf{V} - \bar{\mathbf{V}} = \mathbf{P}_v^* \mathbf{\Sigma}_v \mathbf{E}_v^T = \mathbf{P}_v \mathbf{E}_v^T \tag{10}$$

and the corresponding principal components

$$\mathbf{P}_v = \mathbf{P}_v^* \mathbf{\Sigma}_v = \left(\mathbf{V} - \bar{\mathbf{V}} \right) \mathbf{E}_v. \tag{11}$$

Again, $\mathbf{\Sigma}_v$ is a full-rank diagonal matrix for non-degenerate cases (any realistic geophysical vector quantity with enough sampling):

$$\mathbf{\Sigma}_v = \begin{pmatrix} \sigma_{1v} & 0 \\ 0 & \sigma_{2v} \end{pmatrix}. \tag{12}$$

The $\mathbf{E}_v$ matrix can also be interpreted as a rotation matrix given by the angle θ_v formed by the leading EOF with the zonal axis as:

$$\mathbf{E}_v = \begin{pmatrix} \cos\theta_v & -\sin\theta_v \\ \sin\theta_v & \cos\theta_v \end{pmatrix}. \tag{13}$$

Finally, the same orthogonality properties as before also hold

$$\mathbf{P}_v \mathbf{P}_v^T = \mathbf{\Sigma}_v^2, \tag{14}$$

$$\mathbf{P}_v^* \mathbf{P}_v^{*T} = \mathbb{1} \tag{15}$$

and

$$\mathbf{E}_v \mathbf{E}_v^T = \mathbf{E}_v^T \mathbf{E}_v = \mathbb{1}. \tag{16}$$

Up to this point, we are just repeating the results for the matrix $\mathbf{V}$ from the ones used in section 3.1 for completeness.

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

However, since in general the EOFs derived from observations ($\mathbf{U}$) and simulations ($\mathbf{V}$) will not be the same, 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

$$\mathbf{R}_{vu} = \begin{pmatrix} \cos\theta_{vu} & -\sin\theta_{vu} \\ \sin\theta_{vu} & \cos\theta_{vu} \end{pmatrix}, \tag{17}$$

so that

$$\mathbf{E}_v = \mathbf{R}_{vu} \mathbf{E}_u, \tag{18}$$

$$\mathbf{V} = \bar{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_u^T \mathbf{R}_{vu}^T \tag{19}$$

and the corresponding principal components can be expanded as

$$\mathbf{P}_v = \left(\mathbf{V} - \bar{\mathbf{V}}\right)\mathbf{R}_{vu}\mathbf{E}_u = \tilde{\mathbf{V}}\mathbf{E}_u, \tag{20}$$

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

3.4   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

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

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

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

Substituting Eq. (1) and Eq. (9) into Eq. (21), it can be shown that

$$\boldsymbol{\Delta}_{uv}^2 = \frac{1}{N}\mathbf{B}_{uv}^2 + \frac{1}{N}\left(\mathbf{S}_{uv}^T + \mathbf{S}_{uv}\right) + \frac{1}{N}\mathbf{D}_{uv} = \frac{1}{N}\mathbf{B}_{uv}^2 + \frac{1}{N}\mathbf{C}_{uv} + \frac{1}{N}\mathbf{D}_{uv} \tag{23}$$

with

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

$$\mathbf{S}_{uv} = \left(\mathbf{E}_v\boldsymbol{\Sigma}_v\mathbf{P}_v^{*T} - \mathbf{E}_u\boldsymbol{\Sigma}_u\mathbf{P}_u^{*T}\right)\left(\bar{\mathbf{V}} - \bar{\mathbf{U}}\right) = \left(\mathbf{E}_v\mathbf{P}_v^T - \mathbf{E}_u\mathbf{P}_u^T\right)\left(\bar{\mathbf{V}} - \bar{\mathbf{U}}\right) \tag{25}$$

and

$$\mathbf{D}_{uv} = \mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + \mathbf{E}_v\boldsymbol{\Sigma}_v^2\mathbf{E}_v^T - \left(\mathbf{E}_u\boldsymbol{\Sigma}_u\mathbf{P}_u^{*T}\mathbf{P}_v^*\boldsymbol{\Sigma}_v\mathbf{E}_v^T + \mathbf{E}_v\boldsymbol{\Sigma}_v\mathbf{P}_v^{*T}\mathbf{P}_u^*\boldsymbol{\Sigma}_u\mathbf{E}_u^T\right) \tag{26}$$

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

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

$\mathbf{B}_{uv}^2$ represents the part of the squared error 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 due to the projection of the bias into the differences of vector
310   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. (2) and Eq. (10) are substituted into the definition of the matrix $\mathbf{S}_{uv}$ in Eq. (25), yielding

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

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

315   Finally, the matrix $\mathbf{D}_{uv}$ is related to the covariance matrix of anomalies, as also clearly seen if Eq. (2) and Eq. (10) 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 $\mathbf{U}$) in Eq. (13) is used. Thus, considering the expression $\mathbf{E}_v = \mathbf{R}_{vu}\mathbf{E}_u$ Eq. (17), the matrix $\mathbf{D}_{uv}$ above can be rewritten in terms of the EOFs corresponding to observations as

320 $$\mathbf{D}_{uv} = \mathbf{E}_u\boldsymbol{\Sigma}_u^2\mathbf{E}_u^T + \mathbf{R}_{vu}\mathbf{E}_u\boldsymbol{\Sigma}_v^2\mathbf{E}_u^T\mathbf{R}_{vu}^T - \left(\mathbf{E}_u\mathbf{P}_u^T\mathbf{P}_v\mathbf{E}_u^T\mathbf{R}_{vu}^T + \mathbf{R}_{vu}\mathbf{E}_u\mathbf{P}_v^T\mathbf{P}_u\mathbf{E}_u^T\right). \tag{29}$$

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:

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

The interpretation of this expression is that all the matrices involved in the mean squared error can be expressed in the axes de-

325 fined 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 the same way the Taylor diagram performs for univariate datasets. Thus, 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.

330   3.5   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 traditional terminology of principal components (Compagnucci and Richman, 2008). Thus, in the previous description, N will run along

335 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's diagram to climatology, we present an example including the comparison of multi-year averages of Northern Hemisphere surface wind vectors. For 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 applied to the data in order to avoid an excessive weight in the results of points in polar

340 latitudes which represent much a lower area in a regular longitude-latitude grid.

3.6 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 climatologies of surface wind from two models with a different number of realizations are compared with the corresponding climatology from ERA5. As described above, since this is also a comparison of data on a regular longitude-latitude grid, the covariance matrix is also built

345   over the spatial points and the external weights are also applied to avoid an overrepresentation of polar regions in the results.

3.7 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 section 4, and the code to prepare some of these plots are provided as examples in the manual of the package. Besides producing the diagrams shown as an example in this paper, the

350   package also computes all the individual terms used in the analysis of the MSE error as described in Section 3. Thus, different aspects of the main principal axes, their relative rotation, the two-dimensional correlation coefficient and the combined RMSE can be readily analyzed for different vector datasets and exported to tables which can be presented in publications.

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

4.1 Wind over a Mediterranean location

355   The first example of a Sailor diagram is shown in Figure 1 (left). 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 colored circle, except for the reference dataset, which uses a grey square. The leading EOF of every dataset is represented as the semi-major axis of the ellipse that is plotted centered at every model mean value (same colour as the one used to represent the model mean). The grey ellipse centered at each model mean represents the EOF from the reference dataset (observations).

360   Thus, the angle between the colored and grey semi-major axes represents the relative rotation 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). The comparison of these lengths between coloured and grey ellipses allows to address the question whether the model underestimates or overestimates each EOF. In this particular example, since the model vs observation biases are much lower than the variance explained by the principal axes defined by the EOFs, the interpretation of this diagram is not

365   very easy. However, it is already showing the main directions of the error matrices, their biases and the position of the reference dataset. The bottom right legend shows the total RMSE error given by Eq. (22) in subsection 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

370   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). In the scaled version of the diagram
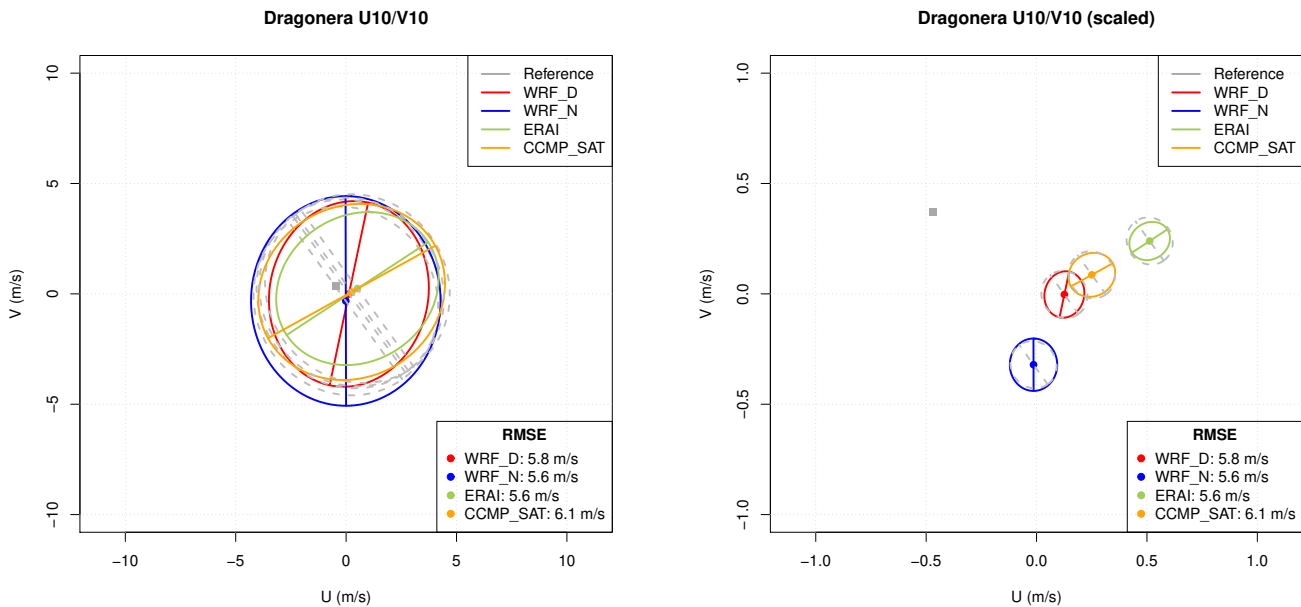
Figure 1. Sailor diagram with default parameters (left) and ellipses scaled to improve visibility of the directional error (right) for the wind observed and simulated in Dragonera (buoy in the Mediterranean).

(Figure 1, right) 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 ellipse and its major semiaxis show the main structure of the variability of the reference dataset. This grey ellipse is plotted

375   centered 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 analyzed 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 axes easily allows to compare the main directions of variability of the observed wind (grey) and modelled one, with the ones corresponding to the WRF model the closest ones to observations. It can be seen that both WRF simulations show a

380   lower rotation of their major axes with respect to the one from observations. The model EOFs are almost orthogonal from the ones in observations for the case of ERAI or CCMPv2. The bottom right legend, in any case, presents the real RMSE error without scaling its value.

4.2   Surface current in the Bay of Biscay

Figure 2 (left) shows an alternative version of the Sailor diagram. In this particular case, the bias is relatively low. Thus, in order

385   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 is only affecting the part of the RMSE derived from term
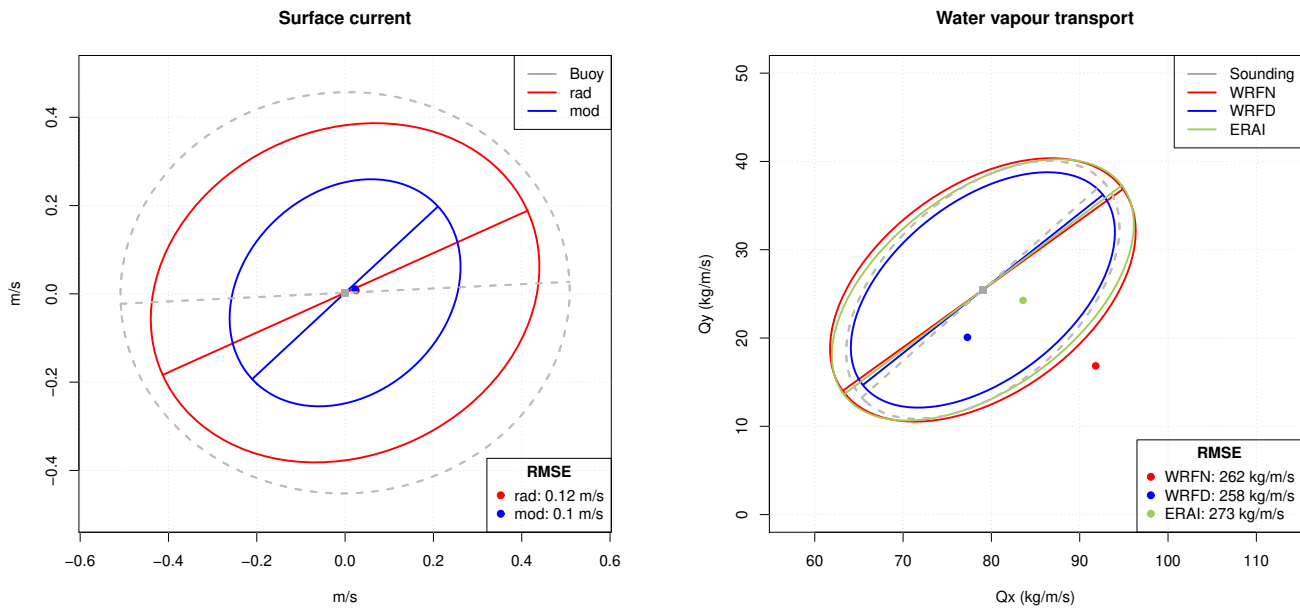
Figure 2. Sailor diagram representing the structure of errors between HF radar estimations of currents (rad) and model results (mod) with variances corresponding to EOFs scaled with scale factor 4 (left). Sailor diagram derived from vertically integrated water vapour transport (right).

$\mathbf{B}_{uv}^2$ in Eq. (23). As in the previous case, they are scaled (four 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
390   be expected, since both cases represent observational (in-situ versus remote) estimations of currents. As in the previous case, the legend at the bottom right shows unscaled total RMSE errors. 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, but especially in the case of model data.

4.3   Vertically integrated water vapour transport

395   The Sailor diagram for the vertically integrated water vapour transport can be seen in Figure 2 (right). In this case, the errors associated to the bias are smaller than the error associated to 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. This way, the errors in direction can be easily identified. For clarity, the ellipses are scaled with a scale factor of 0.1, again. It can be seen that the estimation of the EOFs is closer for the case of the simulation with data assimilation, both in direction and,

400 particularly, in the case of 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 MSE between observed and modelled vertically integrated water vapour transport are presented in Table 1. Different aspects of the main principal axes, their relative rotation, the two-dimensional correlation coefficient and the combined RMSE can be readily analyzed for the water vapour transport vectors.

	Model	σ_x	σ_y	θ_{uv}	R^2	$|\bar{\mathbf{U}} - \bar{\mathbf{V}}|$	RMSE
1	OBS	154.72	146.09	0.00	2.00	0.00	0.00
2	WRF N	173.05	149.19	-5.86	1.57	15.41	261.98
3	WRF D	149.52	133.31	-3.23	1.94	5.65	257.53
4	ERAI	171.00	148.00	-4.66	1.92	4.69	272.94

Table 1. 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 representing the standard deviation around the average values for every model (kg m^{-1} s^{-1}) and θ_{uv} (degrees) represents the relative rotation of model ellipses with respect to the observational one (Section 3.3). 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, instead of 1 for the one-dimensional counterpart). Finally, the differences between the datasets described by the bias ($|\bar{\mathbf{U}} - \bar{\mathbf{V}}|$ kg m^{-1} s^{-1}) and total root mean squared error are also shown.

405

4.4 Wave energy flux

Figure 3 (left) represents the Sailor diagram corresponding to the estimation of wave energy flux by different models against observations from a buoy (Estaca de Bares). In order to avoid many duplicate ellipses in the diagram when many models are diagnosed, it is better to center all their ellipses on top of the observational dataset's mean point, as shown in this case
410 (compare it with Figure 1, left), using an scale factor of 0.25. The statistical models involving the selection of analogs (analo), the combination of analogs and Random Forests (analorf) and the use of Random Forests alone (rf) tend to cluster in the same orientation of the covariance matrix, slightly rotated anticlockwise with respect to observations. Still, they present different levels of estimation of variance, with random forests estimating the variance the best, and models based on analogs presenting a worse estimation of the right variance. The WAM model is slightly rotated clockwise. It also presents an overestimation of
415 variance with respect to observations. On the other hand, since persistence (per) involves just a resampling of the original PDF of the observations, it leads to a perfect agreement in terms of the EOFs.

4.5 Spatial distribution of seasonally-averaged surface wind

As an example of the potential uses of the Sailor diagram, Figure 3 (right) shows the agreement of the January-averaged northern Hemisphere surface wind from different reanalyses using an scale factor of 0.2. In this case, we are assuming that
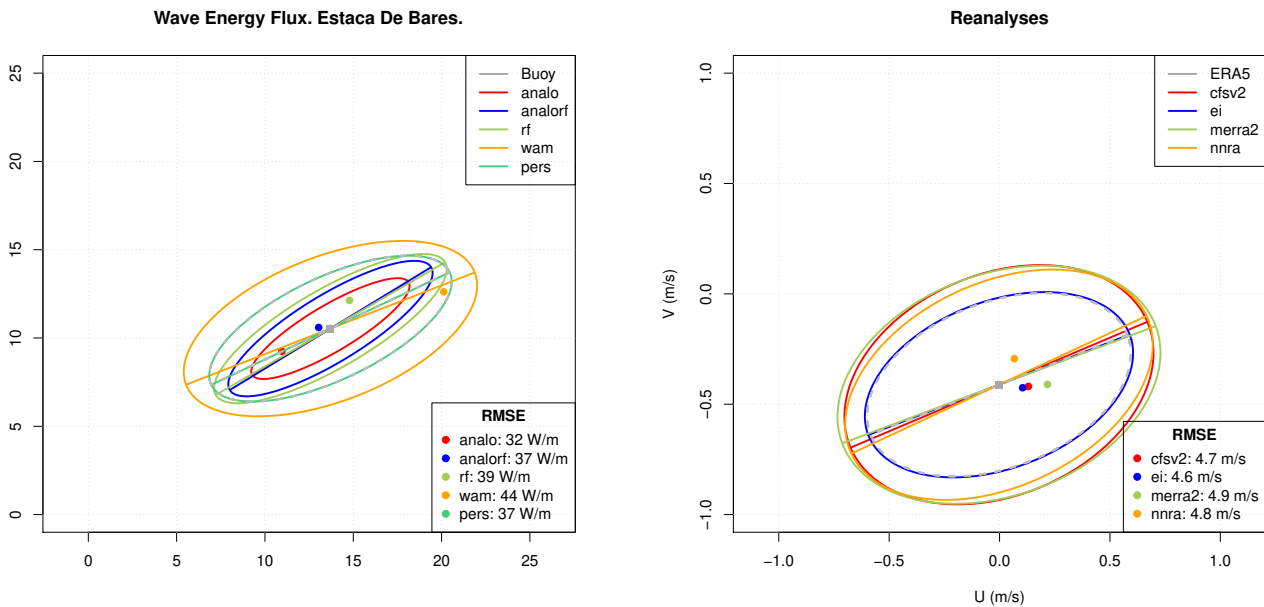
Figure 3. Sailor diagram representing the structure of errors between different physical and statistical models in the estimation of wave energy flux (left) and surface wind in January over the Northern Hemisphere (right).

ERA5 corresponds to the "perfect" dataset (observations). This is quite arbitrary, but we are performing this comparison for the sake of showing the ability of the Sailor diagram to evaluate spatial fields, too. It is clearly shown that the reanalyses produced by the ECMWF (ERA5 and ERA Interim) show the highest agreement both in terms of 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, too, it can be seen that the lowest bias is the one corresponding to ERA Interim.

4.6 Application to multimodel ensembles

In this case, we propose to define the average of all the M ensemble members of every model as the vector $\bar{\mathbf{V}}$ (Rougier, 2016). On the other side, the principal components, associated variances and eigenvectors associated with the model can be estimated from an extended data matrix $\mathbf{V}_e$ (with dimensions $NM \times 2$) built by joining all the realizations together in a single dataset. This means that the observational matrix $\mathbf{U}$ must also be extended to an $\mathbf{U}_e$ matrix sized $NM \times 2$. This can be done by repeating the observations M times to produce the $\mathbf{U}_e$ dataset and the results is shown in Figure 4 (left). 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
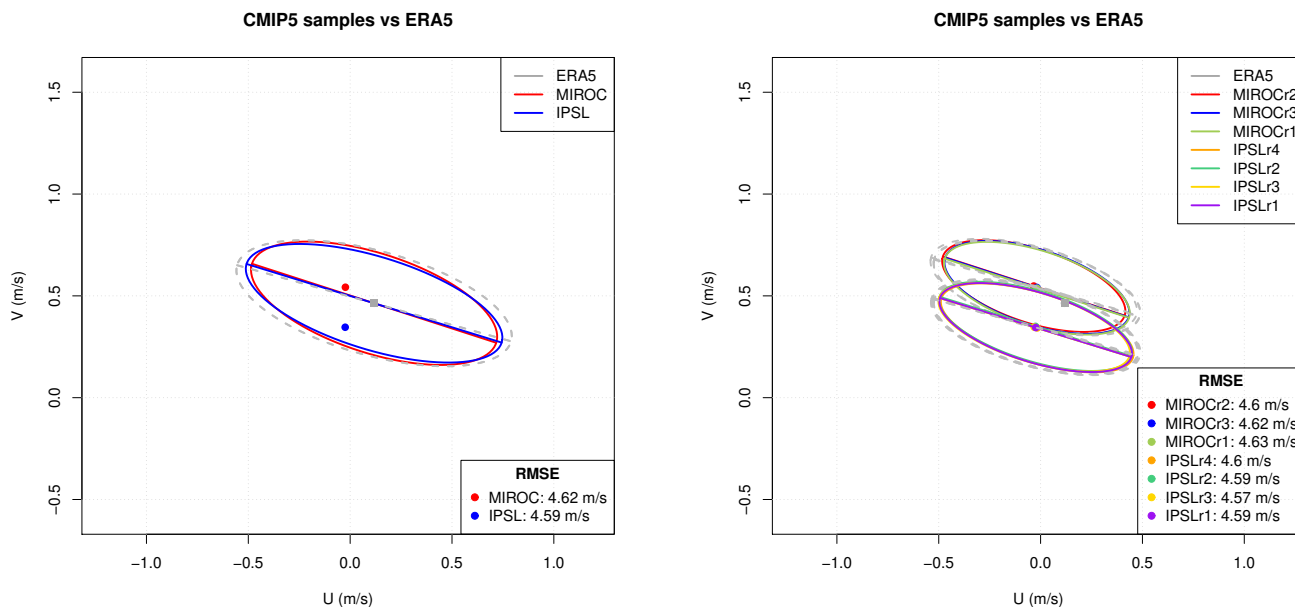
Figure 4. Sailor diagram representing the agreement between the Southern Hemisphere wind field as simulated by two models from the CMIP5 repository with ERA5 data when the reference dataset is repeated (left) or when the individual realizations of the ensemble are takes as independent datasets (right).

435   the same, such as in the CMIP set of runs, for instance. As shown in Figure 4 (left), prepared using as scale factor 0.2, the Sailor diagram shows interesting features. The two models studied agree quite well in the simulation of the spatial variability of the field (the EOFs and major/minor axes in the ellipse represent the spatial variability of the field). Both models simulate an underestimation of zonal average winds when compared with ERA5, whilst MIROC tends to overestimate the mean meridional circulation, IPSL underestimates it.

440   The second option for ensembles (same scale factor) is shown in Figure 4 (right). It consists in the use of every single realization of the ensemble as a single model. The proposed 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

445   model tend to cluster at the same position. The intra-ensemble variance of properties such as the spatial variability of the field is also quite low. Thus, in the analysis performed here, all the realizations of every model in the ensemble are very close to the reference dataset.

5   Conclusions

A new diagram for the fast evaluation of the quality of models forecasting two-dimensional vector fields or time series has been

450 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 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 MSE 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

455 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 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

460 diagram.

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, by centering all of them in the reference dataset or on top of every model being used. Thus, researchers can design a diagram that best suits their needs.

The analysis of ensembles can also be performed by means of the diagram. As shown in subsection 4.6, the diagram can

465 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. 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 analyze the impact of this in the application of the Sailor diagram to model ensembles. However, in the second case, all the members of the ensemble can also be analyzed as

470 independent realizations of the same dataset. 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. However, every member of the ensemble can also be treated individually, so that the diagram would represent the intraensemble spread.

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.

475 *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 also distributed with the package. Both elements (code and data) can also be accessed from ZENODO, DOI https://doi.org/10.5281/zenodo.3543717.

Author contributions. JS conceived the idea, 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

480 graphical representation of data. GE collaborated 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. GIB and AU provided data for the package, performed exhaustive checking of the implementation and helped in the analysis of results. All authors took active part in the writing of the paper.

Competing interests. The authors manifest they have no competing interests in the outcome of this paper-edition process.

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

485 *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). 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 Port Authorities (Puertos del Estado) and 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.

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, Bulletin of the American Meteorological Society, 92, 157–174, https://doi.org/10.1175/2010BAMS2946.1, 2011.

Boer, G. J. and Lambert, S. J.: Second order space-time climate difference statistics, Clim. Dyn., 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., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field, Journal of 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, Journal of Climate, 26, 153–170, https://doi.org/10.1175/JCLI-D-12-00090.1, 2013.

Compagnucci, R. H. and Richman, M. B.: Can principal component analysis provide atmospheric circulation or teleconnection patterns?, International Journal of Climatology, 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, Journal of Atmos. and Oceanic Technol., 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, 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.

Dörfler, W.: Diagrammatic thinking, in: Activity and sign, pp. 57–66, Springer, 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, Climate Dynamics, 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, D04 101, 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,

530   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), Journal of 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, Journal of Geophysical Research: Atmospheres, 123, 708–729,

535   https://doi.org/10.1002/2017JD027511, 2018.

Hawkins, E., Fæhn, T., and Fuglestvedt, J.: The climate spiral demonstrates the power of sharing creative ideas, Bull. of the Amer. Met. 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.,

540   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, Journal of Atmospheric and Oceanic Technology, 20,

545   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, Hydrology and Earth System Sciences, 15, 1895–1907, https://doi.org/10.5194/hess-15-1895-2011, 2011.

550   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 Engineering, 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 Journal of Oceanic Engineering, 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-

555   Roldán, A.: Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation, J. Appl. Meteorol. and 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.,

560   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, Bull. Amer. 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.

565   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., Galloudec, O. L., 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 Traon, P.-Y. L.: Recent updates to the Copernicus Marine Service global

570   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.

575   Martin, T. H. D. T. G. M., 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.,

580   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, Geoscientific Model Development, 4, 723–757, https://doi.org/10.5194/gmd-4-723-2011, 2011.

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

North, G., Bell, T., Cahalan, R., and Moeng, F.: Sampling Errors in the Estimation of Empirical Orthogonal Functions, Monthly Weather

585   Review, 110, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2, 1982.

P.P.E.: Puertos del Estado: Oceanography: Forecast, real time and climate, Spanish Government: Madrid. Updated 2015-10-11. http://www.puertos.es/en-us/oceanografia/Pages/portus.aspx, 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, 2019.

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

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

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

595   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, Geophysical Research Letters, 38, L19 607, 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, Journal of Marine Systems, 109-110, S121 –

600   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, Computers & Geosciences, 122, 113–119, https://doi.org/10.1016/j.cageo.2018.10.007, 2019.

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, Journal of Climate, 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

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, Continental Shelf Research, 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., D7, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Ulazia, A.: Multiple roles for analogies in the genesis of fluid mechanics: How analogies can cooperate with other heuristic strategies, Foundations of Science, 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, Applied 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, Journal of 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, 2 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. Mod. Dev., 9, 4365–4380, https://doi.org/10.5194/gmd-9-4365-2016, 2016.