



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

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

<sup>1</sup>Dept. Applied Physics II, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s./n., 48940 Leioa, Spain

<sup>2</sup>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

<sup>3</sup>Tecnalia, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Álava, Albert Einstein 28, 01510 Vitoria-Gasteiz, Spain

<sup>4</sup>Faculty of Pharmacy, Applied Mathematics, Statistics and O.R. Department, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain

<sup>5</sup>Nuclear Engineering and Fluid Mechanics Dept., Gipuzkoako Ingeniaritza Eskola, Europa Plaza 1, 20018 Donostia, Spain

<sup>6</sup>Oeschger Centre for Climate Change Research, University of Bern, 3010 Bern, Switzerland

<sup>7</sup>Climate and Environmental Physics, University of Bern, 3010 Bern, Switzerland

<sup>8</sup>Nuclear Engineering and Fluid Mechanics Dept., Escuela de Ingeniería de Bilbao, Plaza Ingeniero Torres Quevedo 1, 48013 Bilbao, Spain

<sup>9</sup>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 ref-

erence 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.

## 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, com-

paring 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 over-estimated 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 lim-

ited 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  $2 \times 2$  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  $38^\circ$  N and  $-124^\circ$  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  $(4.8, -6.8) \text{ m s}^{-1}$  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  $30^\circ$  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 (ERA-Interim) 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  $43.6^\circ$  N and  $2.0^\circ$  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  $43.5\text{--}44.7^\circ$  N and  $3.2\text{--}1.3^\circ$  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  $\eta$  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 **U** corresponding to the reference dataset (observations) in its EOFs. A similar notation will be used later for the decomposition of the matrix **V** corresponding to the zonal and meridional components of every model which is being compared to observations. Later, the expansion of the **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, **U** includes the zonal and meridional components of observations and so does **V** for a simulated dataset. The observational dataset is formed by the two-dimensional (zonal and meridional) components of vector measurements  $\mathbf{u}_i$ , with  $i = 1 \dots N$  arranged as rows in an  $N \times 2$  matrix **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 2 × 2 covariance matrix from the zonal and meridional components of velocity anomalies in the observations is given by

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

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

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

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

$i$ th eigenvalue so that

$$f_{ui} = \frac{\lambda_{ui}}{\sum_{i=1}^r \lambda_{ui}} \quad (3)$$

represents the fraction of variance in observations explained by the linear combination of the original variables defined by the  $i$ th 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  $e_{uj}$  with  $j = t + 1 \dots r$  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

$$\mathbf{U} = \bar{\mathbf{U}} + \mathbf{P}_u^* \boldsymbol{\Sigma}_u \mathbf{E}_u^T = \bar{\mathbf{U}} + \mathbf{P}_u \mathbf{E}_u^T, \quad (4)$$

with  $\mathbf{P}_u^*$  (an  $N \times 2$  matrix) the standardized principal components of the  $\mathbf{U}$  data,  $\boldsymbol{\Sigma}_u$  ( $2 \times 2$  matrix) the standard deviations ( $\sigma_{1u}$  and  $\sigma_{2u}$ ) of the leading and second EOFs of the  $\mathbf{U}$  field,  $\mathbf{E}_u$  ( $2 \times 2$  matrix) the matrix holding the orthogonal rotation matrix leading to the empirical orthogonal functions of the  $\mathbf{U}$  field arranged as columns, and  $\mathbf{P}_u = \mathbf{P}_u^* \boldsymbol{\Sigma}_u$  ( $N \times 2$  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

$$\mathbf{U} - \bar{\mathbf{U}} = \mathbf{P}_u^* \boldsymbol{\Sigma}_u \mathbf{E}_u^T = \mathbf{P}_u \mathbf{E}_u^T, \quad (5)$$

the corresponding principal components as

$$\mathbf{P}_u = \mathbf{P}_u^* \boldsymbol{\Sigma}_u = (\mathbf{U} - \bar{\mathbf{U}}) \mathbf{E}_u, \quad (6)$$

and their standardized counterparts as

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

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),  $\boldsymbol{\Sigma}_u$  is a full-rank diagonal matrix:

$$\boldsymbol{\Sigma}_u = \begin{pmatrix} \sigma_{1u} & 0 \\ 0 & \sigma_{2u} \end{pmatrix}, \quad (8)$$

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  $\theta_u$  formed by the leading (second) EOF with the zonal (meridional) axis as follows:

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

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

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

$$\mathbf{P}_u \mathbf{P}_u^T = \boldsymbol{\Sigma}_u^2, \quad (10)$$

as do the standardized principal components,

$$\mathbf{P}_u^* \mathbf{P}_u^{*T} = \mathbf{1}, \quad (11)$$

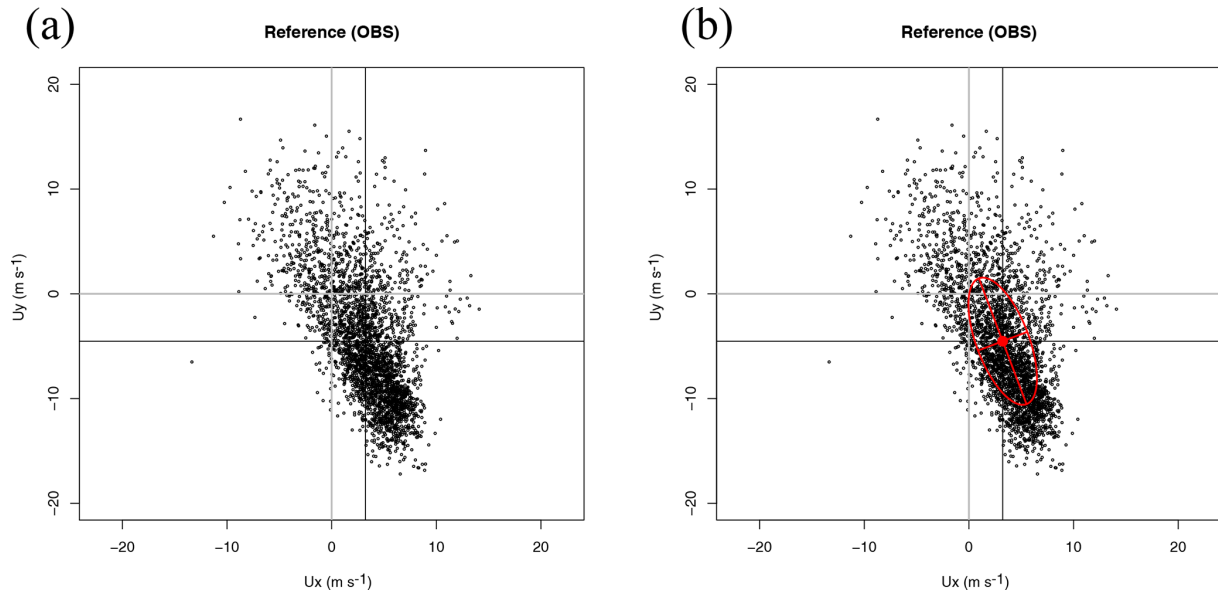
and eigenvectors (EOFs) in the horizontal plane,

$$\mathbf{E}_u \mathbf{E}_u^T = \mathbf{E}_u^T \mathbf{E}_u = \mathbf{1}. \quad (12)$$

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

$$\|\mathbf{P}_u^* \mathbf{P}_u^{*T}\|_F = \|(\mathbf{U} - \bar{\mathbf{U}}) \mathbf{E}_u \boldsymbol{\Sigma}_u^{-2} \mathbf{E}_u^T (\mathbf{U} - \bar{\mathbf{U}})^T\|_F = 1. \quad (13)$$

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,

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

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 ( $\mathbf{U}$  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  $\mathbf{v}_i$ , with  $i = 1 \dots N$ , will be arranged as rows in an  $N \times 2$  matrix  $\mathbf{V}$ . The average vector from model data  $\bar{\mathbf{v}}$  is arranged as constant rows in an  $N \times 2$  matrix  $\bar{\mathbf{V}}$ . The  $\mathbf{V}$  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:

$$\begin{aligned} \mathbf{V} &= \bar{\mathbf{V}} + \mathbf{P}_v^* \boldsymbol{\Sigma}_v \mathbf{E}_v^T = \bar{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_v^T \Leftrightarrow \\ \mathbf{V} - \bar{\mathbf{V}} &= \mathbf{P}_v^* \boldsymbol{\Sigma}_v \mathbf{E}_v^T = \mathbf{P}_v \mathbf{E}_v^T, \end{aligned} \quad (15)$$

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 $\mathbf{V}$ in the EOFs defined by observations

In general, the mean values and EOFs derived from observations ( $\mathbf{U}$ ) and simulations ( $\mathbf{V}$ ) 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

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

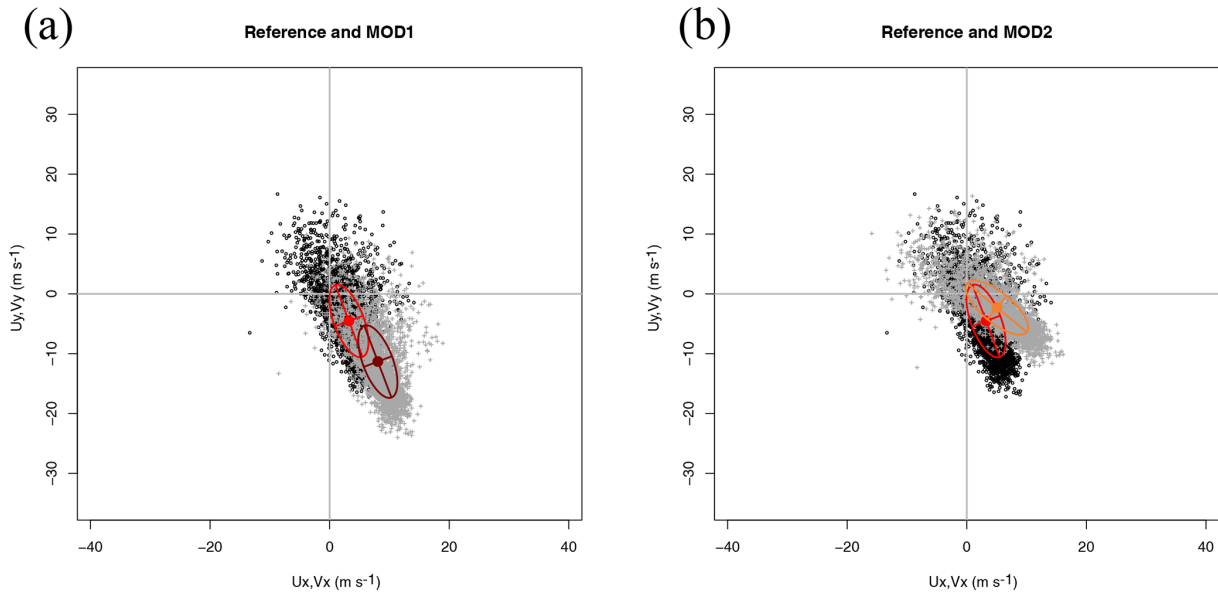
so that

$$\mathbf{E}_v = \mathbf{R}_{vu} \mathbf{E}_u, \quad (17)$$

$$\mathbf{V} = \bar{\mathbf{V}} + \mathbf{P}_v \mathbf{E}_u^T \mathbf{R}_{vu}^T, \quad (18)$$

and the corresponding principal components can be expanded as

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



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

Since both  $\mathbf{e}_{u1}$  and  $-\mathbf{e}_{u1}$  are solutions of the eigenvalue equation when the diagonalization of the two-dimensional covariance matrix is performed (the same happens with  $\mathbf{e}_{v1}$  and  $-\mathbf{e}_{v1}$  for model data),  $\theta_{vu}$  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

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

and it measures the agreement between the pairs of EOFs from observations ( $\mathbf{e}_{ui}$ ) and models ( $\mathbf{e}_{vi}$ ). 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  $\mathbf{e}_{ui}$  and  $\mathbf{e}_{vi}$ . 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  $(2 \times 2)$  matrix that represents the mean squared error between the  $\mathbf{U}$  and  $\mathbf{V}$  datasets is given by

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

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

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

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

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

with

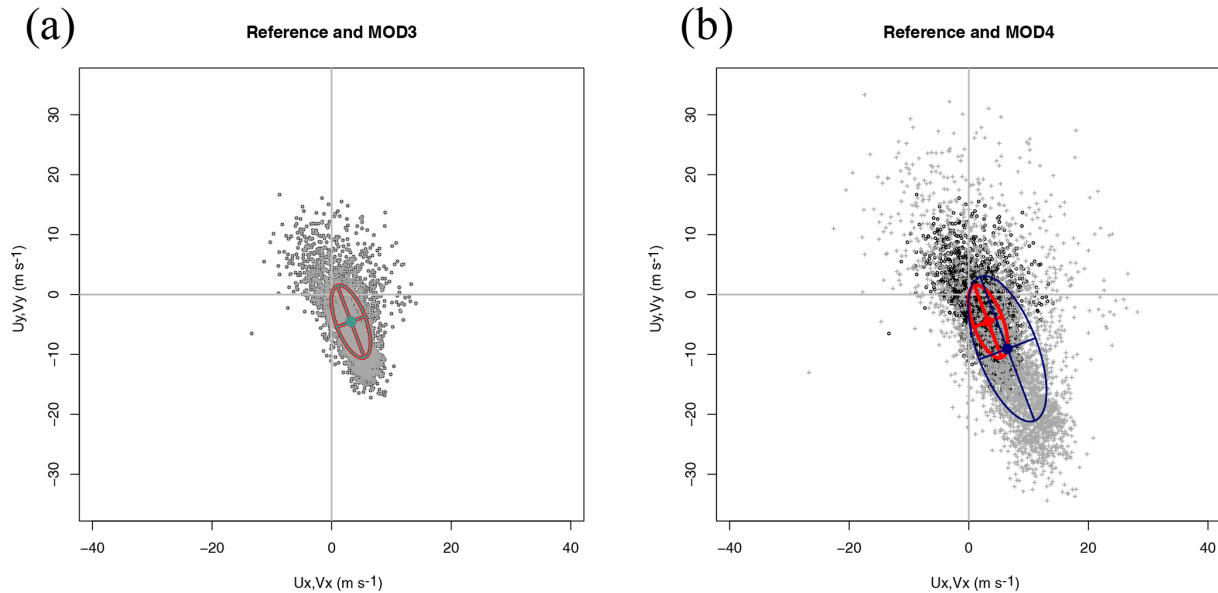
$$\mathbf{B}_{uv}^2 = (\bar{\mathbf{V}} - \bar{\mathbf{U}})^T (\bar{\mathbf{V}} - \bar{\mathbf{U}}), \tag{24}$$

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

and

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





**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

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

$\mathbf{B}_{uv}^2$  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

$$\begin{aligned} \mathbf{S}_{uv} &= \left( (\mathbf{V} - \bar{\mathbf{V}}) - (\mathbf{U} - \bar{\mathbf{U}}) \right)^T (\bar{\mathbf{V}} - \bar{\mathbf{U}}) \\ &= (\mathbf{V} - \mathbf{U})^T (\bar{\mathbf{V}} - \bar{\mathbf{U}}) - (\bar{\mathbf{V}} - \bar{\mathbf{U}})^T (\bar{\mathbf{V}} - \bar{\mathbf{U}}) = 0. \end{aligned} \quad (28)$$

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

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

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

$$\mathbf{D}_{uv} = \mathbf{E}_u \Sigma_u^2 \mathbf{E}_u^T + \mathbf{R}_{vu} \mathbf{E}_u \Sigma_v^2 \mathbf{E}_u^T \mathbf{R}_{vu}^T - \left( \mathbf{E}_u \Gamma_{vu} \mathbf{E}_u^T \mathbf{R}_{vu}^T + \mathbf{R}_{vu} \mathbf{E}_u \Gamma_{vu}^T \mathbf{E}_u^T \right). \quad (30)$$

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.  $\sigma^2$  represents the total variance ( $\text{m}^2 \text{s}^{-2}$ ) of every dataset as computed from the original zonal and meridional components.  $\sum_i \sigma_i^2$  represents the variance ( $\text{m}^2 \text{s}^{-2}$ ) of wind for every dataset (reference or model) as computed from the EOF decomposition (axes of the ellipses in the diagrams).  $\theta_u$  and  $\theta_v$  represent the angles (radians) of the semi-major axes of the ellipses calculated for reference and models.  $\theta_{vu}$  (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).  $|\text{bias}|$  represents the magnitude of the bias ( $\text{m s}^{-1}$ ). RMSE lists the root mean squared errors ( $\text{m s}^{-1}$ ). The eccentricity of the ellipses ( $\varepsilon$ ) 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	$\sigma^2$	$\sum_i \sigma_i^2$	$\theta_u$	$\theta_v$	$\theta_{vu}$	$R^2$	$ \text{bias} $	RMSE	$\varepsilon$	$g_{11}$
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

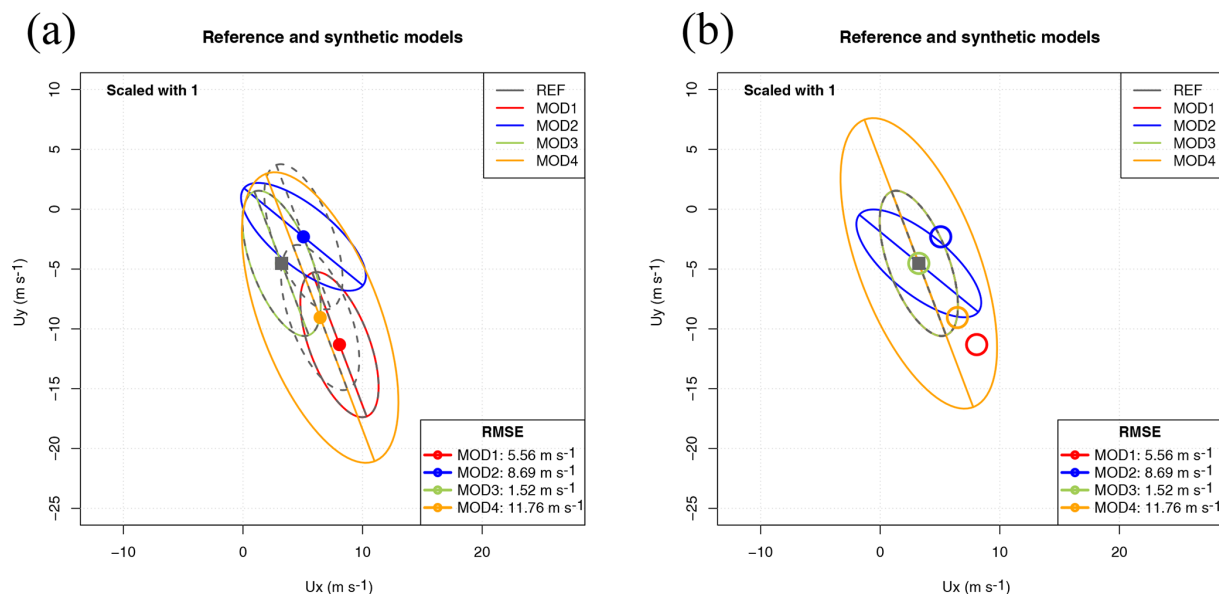
properties of the diagram. In MOD3, a simple random permutation of the original observations has been performed. Thus, there are neither biases nor rotations of the principal axes. From the point of view of the graphical example shown, it seems that the model is perfect, but it is not due to the lack 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  $\sigma^2$  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  $\theta_{vu}$  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  $\theta_u$  and  $\theta_v$  (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  $\varepsilon$ , 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  $\mathbf{e}_{ui}$  and  $-\mathbf{e}_{ui}$  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  $\theta_u$ . 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

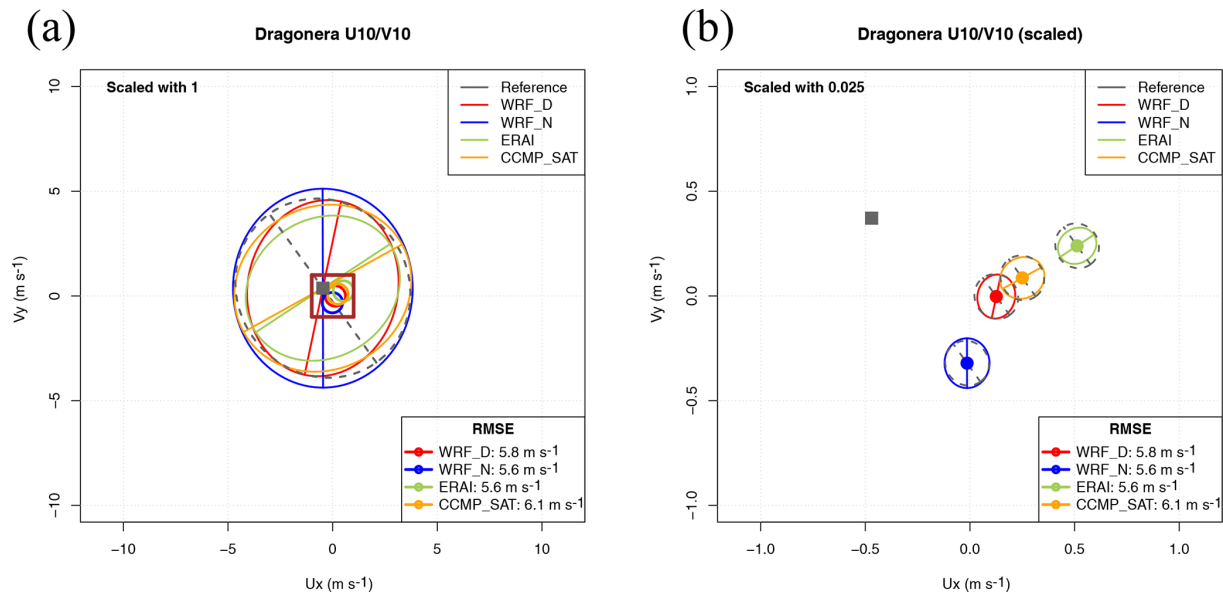
In the case of the analysis of the ability of models to represent the spatial distribution of an averaged field (a typical use of the Taylor diagram in climatology, for instance), there is no change needed to the diagram defined so far. Instead of using the  $T$  mode of principal components (covariance 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  $\phi$  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  $\phi$  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 ( $\theta_{vu}$ ) 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.

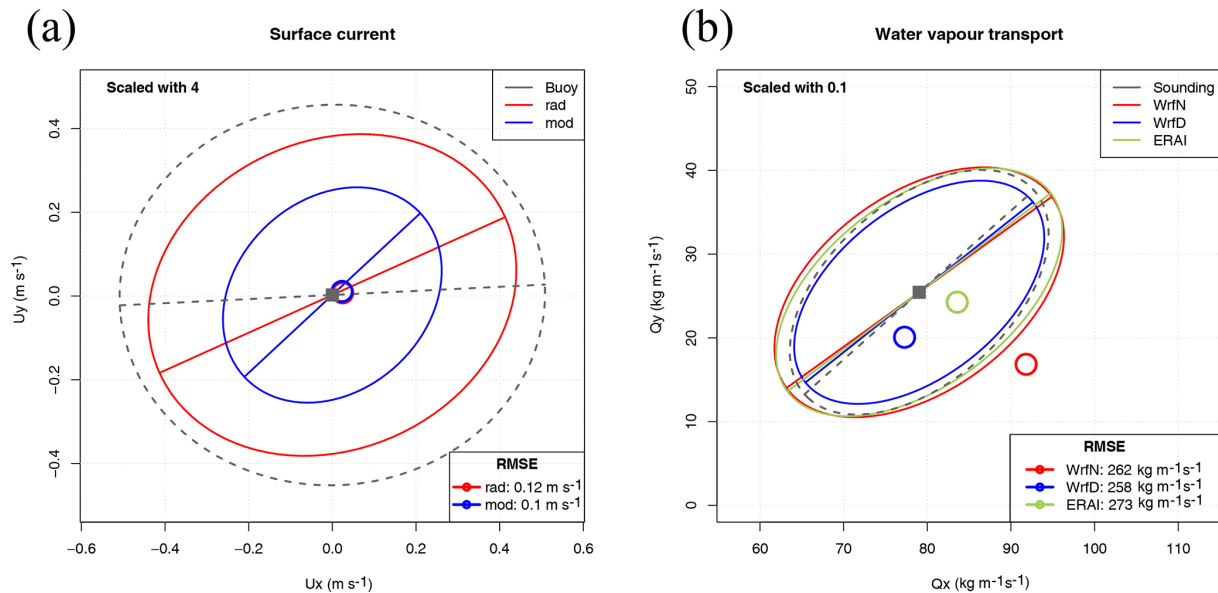
#### 4.3 Vertically integrated water vapour transport

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.  $\sigma_x$  and  $\sigma_y$  represent the semi-major and semi-minor axes of the ellipses ( $\text{kg m}^{-1} \text{s}^{-1}$ ). 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  $|\bar{\mathbf{U}} - \bar{\mathbf{V}}|$  ( $\text{kg m}^{-1} \text{s}^{-1}$ ) and total root mean squared error ( $\text{kg m}^{-1} \text{s}^{-1}$ ) are also shown. Finally, the eccentricity of the ellipses ( $\epsilon$ ) 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	$\sigma_x$	$\sigma_y$	$R^2$	$ \bar{\mathbf{U}} - \bar{\mathbf{V}} $	RMSE	$\epsilon$	$g_{11}$
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	ERA5	196.99	111.18	1.92	4.69	272.94	0.83	1.00

variances with fewer lines. It is clearly shown that the re-analyses produced by the ECMWF (ERA5 and ERA5) show the closest agreement in terms of both the smallest bias and better matching of the corresponding EOFs. The other re-analyses (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 ERA5. The easiest way to arrive at a numerically precise overall diagnostic is presented in the legend, where the aggregated RMSE is shown.

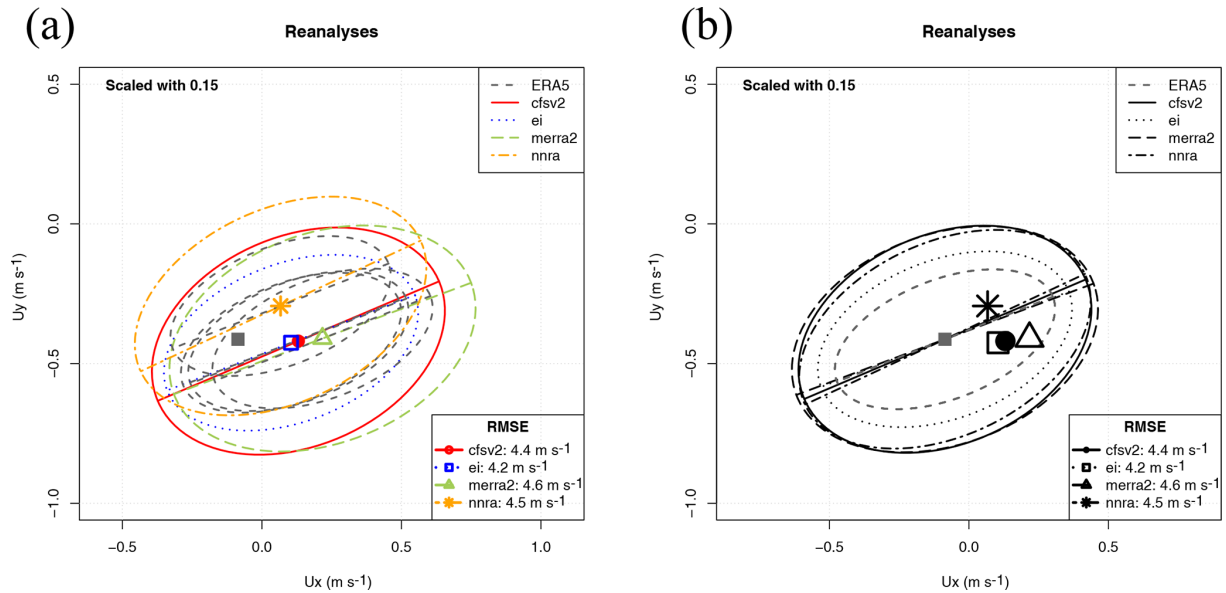
#### 4.5 Application to multimodel ensembles

In this case, we propose defining the average of all the  $M$  ensemble members of every model as the vector  $\bar{\mathbf{V}}$  (Rougier, 2016). On the other side, the principal components and the associated variances and eigenvectors can be estimated from

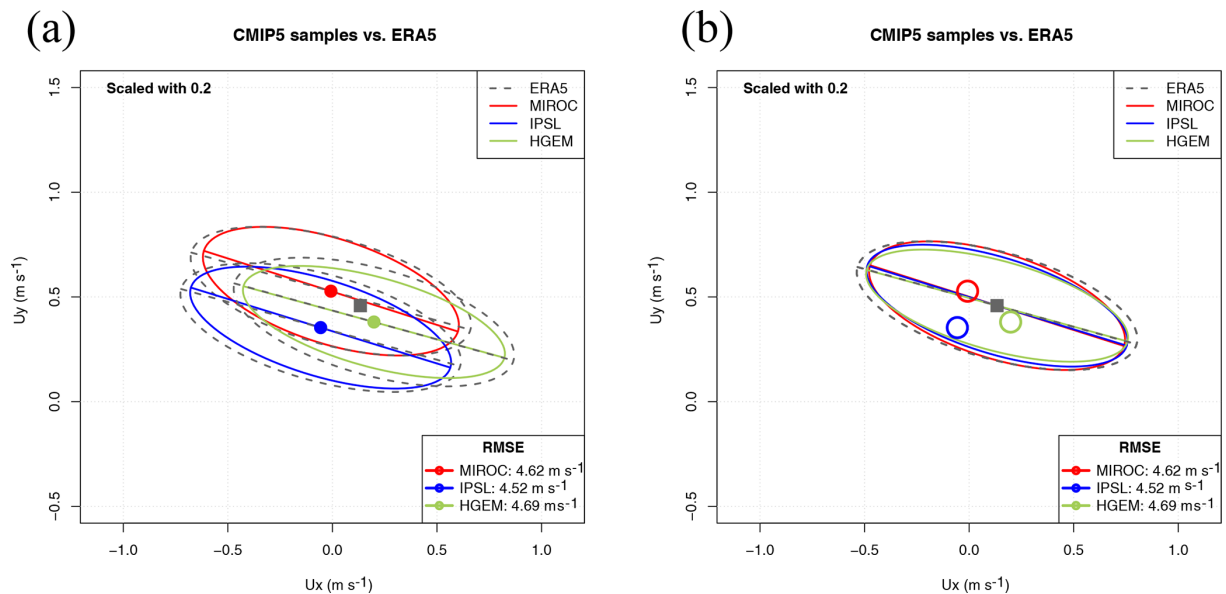
an extended data matrix  $\mathbf{V}_e$  (with dimensions  $NM \times 2$ ), which is built by joining all the realizations together in a single dataset. This means that the observational matrix  $\mathbf{U}$  must also be extended to a  $\mathbf{U}_e$  matrix (sized  $NM \times 2$ ). This can be done 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





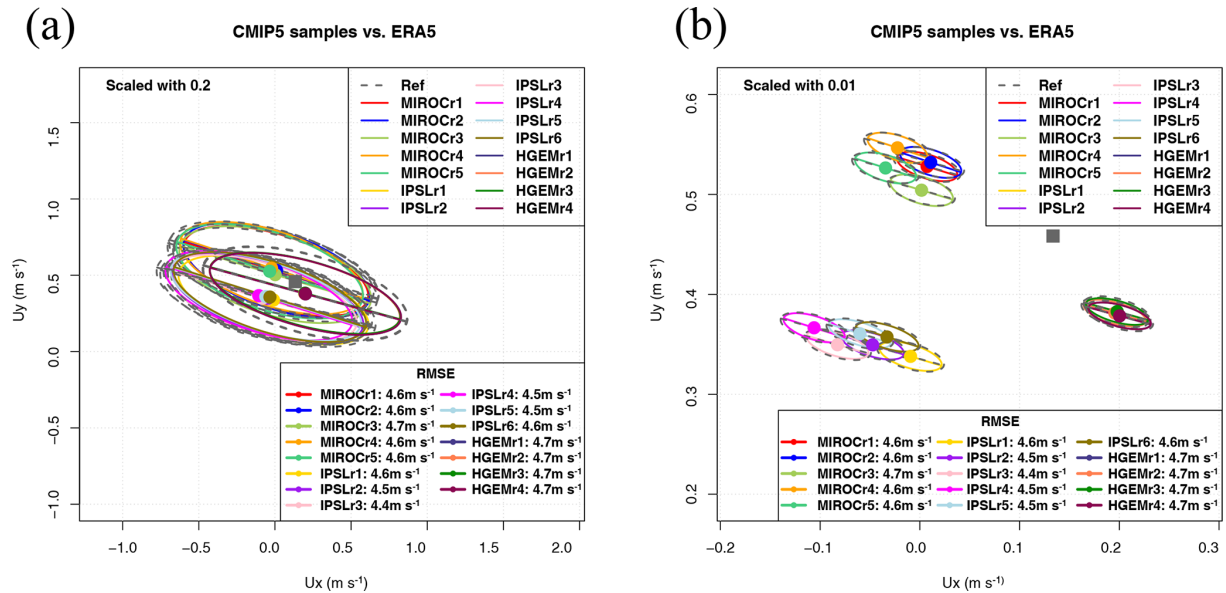
**Figure 7.** Sailor diagram representing the structure of errors in surface wind in January over the Northern Hemisphere for different reanalyses: uncentred version (a, scale factor 0.15) and centred version (b, scale factor 0.15).



**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 corre-

sponding 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, multi-platform 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](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](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 Inter-comparison 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](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](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](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](https://doi.org/10.1207/s15327906mbr0903_10), 1974.

Crosby, D. S., Beaker, L. C., and Gemmill, 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](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., Morcrette, 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., Fichetef, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Jous-saume, 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., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. o. E., Parityka, 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 Fuglestedt, 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 Boissésion, 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., Muñoz-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](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](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](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., Dreviron, 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 (north-eastern 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](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](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-Berastegi, 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.