

Article

Machine Learning as a Lens on NWP ICON Configurations Validation over Southern Italy in Winter 2022–2023—Part I: Empirical Orthogonal Functions

Davide Cinquegrana *  and Edoardo Bucchignani 

Meteorology and Climatology Lab, Centro Italiano Ricerche Aerospaziali (CIRA), Via Maiorise, 81043 Capua, CE, Italy; e.bucchignani@cira.it

* Correspondence: d.cinquegrana@cira.it

Abstract

Validation of ICON model configurations optimized over a limited domain is essential before accepting new semi-empirical parameters that influence the behavior of subgrid-scale schemes. Because such parameters can modify the dynamics of a numerical weather prediction (NWP) model in highly nonlinear ways, we analyze one season of forecasts (December 2022, January and February 2023) generated with the NWP ICON-LAM through the lens of machine learning–based diagnostics as a complement to traditional evaluation metrics. The goal is to extract physically interpretable information on the model behavior induced by the optimized parameters. This work represents the first part of a wider study exploring machine learning tools for model validation, focusing on two specific approaches: Empirical Orthogonal Functions (EOFs), which are widely used in meteorology and climate science, and autoencoders, which are increasingly adopted for their nonlinear feature extraction capability. In this first part, EOF analysis is used as the primary tool to decompose weather fields from observed reanalysis and forecast datasets. Hourly 2-m temperature forecasts for winter 2022–2023 from multiple regional ICON configurations are compared against downscaled ERA5 data and in situ observations from ground station. EOF analyses revealed that the optimized configurations demonstrate a high skill in predicting surface temperature. From the signal error decomposition, the fourth EOF mode is effective particularly during night-time hours, and contributes to enhancing the performance of ICON. Analyses based on autoencoders will be presented in a companion paper (Part II).

Keywords: numerical weather prediction; limited area mode; machine learning; empirical orthogonal function; ICON; phase space



Academic Editor: Doris Folini

Received: 3 December 2025

Revised: 20 January 2026

Accepted: 23 January 2026

Published: 26 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Meteorological phenomena in the atmosphere are not purely random; however, they are so complex and sensitive to initial conditions that from a practical point of view they behave like random fields. For this reason, probabilistic approaches (ensemble forecasting, statistical climatology, etc.) are often applied. The validation of a Numerical Weather Prediction (NWP) model is a fundamental step in ensuring the reliability of forecasts across different spatial and temporal scales. Classical validation techniques such as point-to-point error statistics and skill scores provide valuable information, but are often not sufficient to capture the spatial structures and dominant modes of variability. In this context, Empirical Orthogonal Function (EOF) analysis represents a powerful complementary approach, since

it allows the identification of patterns in atmospheric fields and helps to assess whether a model correctly reproduces the main dynamical features. Therefore, integrating EOF-based diagnostics with conventional validation methods leads to a more comprehensive assessment of NWP model performance. The EOF was introduced in meteorology by Lorenz [1] in 1956 and is still widely used in climate science. It is mathematically equivalent to the Proper Orthogonal Decomposition (POD), a statistical method suitable for dealing with many scientific problems, in applications that requires data reduction, filtering, features extractions and so on; in fluid mechanics, it was introduced to represent pseudorandom process such as coherent structures of turbulent flows [2] as a linear combination of deterministic functions (POD modes), which are modulated by uncorrelated random coefficients. Both methods are based on the eigen-decomposition of the covariance matrix, and aim to identify orthogonal modes that capture the maximum variance (or energy) of the system. Nowadays, ECMWF [3] employs PCA/EOF methods to identify Northern Hemisphere Winter Connection (the Pacific/North American pattern, the North Atlantic Oscillation, and the Eurasian pattern) and Monsoon Indexes. Those statistical methods are currently applied to other top-class NWP and climate models as well, for example to ICON model seasonal simulations at the DWD institute [4]. Even for seasonal forecasting, EOF plays a crucial role in bias correction [5–7].

An analysis of the state-of-the-art and future prospects of POD has been provided by Solari et al. [8], with special regard to its framework and applications in wind engineering. One of the authors of the present work has applied this technique in the fields of design space reduction for aerodynamic shape optimization problems [9,10] as well as embedded in a response surface tool in fluid dynamics applications [11,12]. Ramesh et al. [13] used a POD method to study the spatial and temporal variations of wind data in Singapore for the planning and execution of high-altitude balloon flights, with the main objective being the development of a reduced/low-order model of the stratospheric wind data for a faster computation of wind profiles to be used in balloon trajectory simulations. In climate and weather research field, an early work by Ferranti [14] highlighted the capacity of EOF in comparing meteorological pattern with low frequency variability as well as in contributing to the analysis of seasonal weather data and numerical forecasts. Hannachi et al. [15] investigated the possibility of using these methods to include projection into the modes of variability of observations (e.g., reanalysis). Benestad et al. [16] combined fields of standardized surface air temperature and mean sea level pressure, then used the EOFs of these fields as predictors in empirical–statistical downscaling. They also discussed how the common EOF approach enhanced capability in comparing climate models.

In a statistical manner, weather regimes represent the states for which the large-scale flow pattern resides for an assigned period (from a week to a month); in this view, the leading Empirical Orthogonal Functions (EOFs) on a large-scale flow field are computed to construct a regime [17]. In fact, in weather- and circulation-type classifications, EOF (or PCA) approaches represent the main tools used in several studies, e.g., Coast Action 733 [18], which aimed to assess and compare classifications of atmospheric circulation and typical weather situations in European regions. Pandzic [19] identified sectors with different patterns of precipitation on a regional domain over the Adriatic sea and inland Croatia by means of PCA.

This tool has also been employed in regional models to perform climate and seasonal forecasts [20]. Recently, EOF was employed to support the calibration of Numerical Weather Prediction (NWP) systems such as the ICON (ICOsahedral Non-Hydrostatic) model [21]. Via surrogate-based optimization, this technique can support an objective comparison and local enhancements between model outputs and observational datasets, supporting the identification of biases in the simulated fields [22].

The approach proposed in the present study enables a more comprehensive assessment of model performances with respect to the standard point-to-point validation metrics. This study uses the output of the NWP ICON-LAM for winter 2022–2023, considering three advanced configurations of the model applied over a domain in southern Italy encompassing northern Campania and southern Lazio while employing a spatial resolution of about 1.2 km. These ICON configurations were optimized through tuning of microphysics parameters performed in a previous work [23]. The aim of this study is to validate these datasets against observations and reanalyses in terms of 2-m temperature, not only by using classical procedures based on diurnal cycles, but also through the lens of EOF/POD in order to identify specific patterns induced by the different subgrid parameterizations. Specifically, the evaluation is performed against a collection of ground station data and against a down-scaled version of ERA5 reanalysis at 2-km resolution [24]. Analyses were performed in terms of 2-m temperature, since it represents the main response of the surface–atmosphere system to changes in model parameterizations.

One of our goals is to describe how the reduced dimensionality space can be affected by previously-tuned parameters as well as by the grid resolution, then to verify whether these key factors can be linked to the physics of the system and the way in which they improve the representation of hidden spatial and temporal patterns [25]. In addition to the comparison between reanalyses and ICON forecasts, a great emphasis is given to the representation of ICON data in the subspace defined on observed data. We emphasize the spatial pattern of the first modes extracted from the collected numerical simulation databases as well as by superimposing the time history of the scalar realization of the EOF coefficient. Consequently, EOF analysis provides an effective diagnostic framework for quantifying how well the ICON model captures the observed key modes of atmospheric variability.

Several literature studies have been based on this kind of data process, in which model outputs and observations are projected onto these modes to assess the model's ability to capture dominant variability patterns. This approach provides insights the performance of models in representing key climate modes [26]. For example, Sperber et al. [27] analyzed projections to evaluate models' ability to reproduce the inter-annual variability of the Asian summer monsoon. By comparing the projections of model ensemble members and observations onto specific EOFs, they were able to assess the models' skill in capturing key monsoon variability patterns. Another useful approach is based on the comparison between observed and numerical signals by plotting the PCs on their phase space [28–30]. In the present work, PCs from observed and numerical data are projected on a latent space and evaluated on a common observational manifold. The phase space is achieved through a subspace projection accomplished for this kind of local analysis; in this way, some comments on the forecast performance can be argued.

However, the EOF approach is limited when dealing with nonlinear signals which are difficult linearize under such hypothesis; examples include the orthonormality of the span of modes and global space domain orthogonality, where the modes may be non-physical [25]. Partly to overcome this limitation, feature extraction and dimensionality reduction by means of nonlinear methods such as encoder–decoder (autoencoder) architectures have been gaining increasing attention. These architectures are able to capture complex relationships, although they require large datasets and more effort in tuning, and are less interpretable. To explore the capabilities of this technique and compare them with the approach presented in this research, a companion 'Part II' paper based on ongoing research is currently in preparation.

The rest of this paper is organized as follows: Section 2 describes the datasets, and methodologies, and simulation domain, including the NWP model, simulation setup, and description of the test case under consideration; Section 3 describes the statistical

tools used for model evaluation; Section 4 discusses the results; and Section 5 reports the conclusions.

2. Model, Domain, and Data

2.1. The NWP ICON Model

ICON (ICOsahedral Nonhydrostatic) is a joint project between the Deutscher Wetterdienst (DWD) and the Max Planck Institute for Meteorology (MPI-M) for the development of a unified global numerical weather prediction system. It can also be run as a Limited Area Mode (ICON-LAM); in fact, since 2018 ICON-LAM has replaced COSMO-LM as the operational model of the COSMO consortium. ICON-LAM has been widely described in several works (e.g., [21]), so only the basic information is recalled here. ICON-LAM is characterized by enhanced properties, improved scalability on parallel high-performance computers, and the possibility of performing static mesh refinement. The governing equations are discretized using an unstructured mesh composed of regular icosahedra, each consisting of 20 triangular faces, ensuring exact local mass conservation and consistent tracer transport. Time integration employs a fully explicit two-time-level predictor–corrector scheme. A time-splitting approach separates the dynamical core and tracer advection from the physical parameterizations and horizontal diffusion processes. Due to the limited spatial resolution, some physical processes are not explicitly solved and need to be parameterized; in particular, these are radiation, cloud micro-physics, convection, cloud cover, non-orographic gravity waves, and sub-grid-scale orography. As a consequence, a large number of unconstrained model parameters associated with the parameterization schemes must be properly set through a model calibration procedure. In [23], the authors implemented an automatic calibration methodology to optimize some parameters based on an Efficient Global Optimization algorithm relying on a Gaussian-based metamodel. These parameters are related to turbulence and grid scale microphysics: these are the scaling factor for minimum vertical diffusion ($tkmmin$), minimum value for the turbulence coefficient ($tkhmin$), and laminar resistance ($rlam-heat$) for the former, and the snow vertical velocity ($v0snow$) for the latter. Two optimization processes were conducted, starting from a reference model configuration (identified in the following as ICON-REF) against a gridded observational dataset and a collection of station data, respectively (identified in the following as ICON-GRID and ICON-STAT), leading to two different optimized configurations for the area under investigation. Both of these configurations are characterized by larger values of $v0snow$, $tkhmin$, and $tkmmin$ and smaller values of laminar resistance ($rlam-heat$) than the reference one. Furthermore, the largest differences between ICON-GRID and ICON-STAT lie on the smallest value of $rlam-heat$ and largest value of $tkmmin$ for ICON-STAT. Complete details can be found in Table 3 of Cinquegrana et al. [23].

2.2. Domain

The domain of interest is located in southern Italy, and includes the northern part of the Campania region and southern part of the Lazio region, with the Tirrenian sea positioned on the western side of the domain (Figure 1). The study domain is ideally subdivided into two climate areas, a mild zone under the influence of the sea and a colder internal zone characterized by the presence of mountains. This domain poses relevant challenges for NWP due to its complex orography, pronounced coastal–inland contrasts, and interaction between synoptic-scale forcing and local circulations. These factors often lead to sharp spatial gradients and localized phenomena such as orographically-induced precipitation and sea–land breeze systems, which are difficult to simulate using NWP models. For this reason, the selected region provides a particularly suitable test case for model validation;

traditional point-based metrics can be complemented by EOF analysis to better assess the model's ability to capture the dominant spatial patterns and variability across the domain.

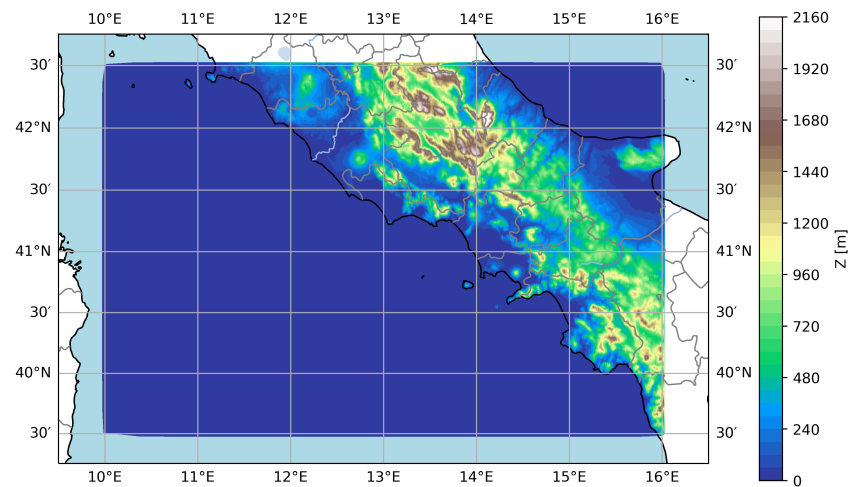


Figure 1. Illustration of the computational ICON-LAM domain in Southern Italy, with terrain elevation.

The two optimized ICON configurations along with the reference one are used in the present work to perform numerical simulations over the study domain, which is discretized using a computational grid with 110,216 cells at a horizontal resolution of about 1.2 km (grid R02B11). The grid is characterized by 65 levels in the vertical direction, and extends from the smoothed surface of the Earth up to 22 km, the first level being positioned at 20 m.

2.3. The Test Case Considered

A series of daily 24-h forecast analyses was carried out, covering the period from 1 December 2022 to 28 February 2023. Each simulation began at 00 UTC of the preceding day and concluded at 24 UTC of the target day. Boundary and initial conditions, including soil climatology, were provided by the ECMWF IFS model [31] at 3-h intervals. The 2-m temperature (referred to in the following as T2m) was extracted as the target variable.

The simulations were performed on the HPC cluster installed at CIRA, which is equipped with 40 Intel Xeon E5-2697 nodes (for a total of 1440 cores) interconnected by means of an Intel Omni-Path network at 100 Gbits.

2.4. Observed Ground Data

Validation of the ICON model in terms of T2m was performed using ground station data made available from the Meteo Italian Supercomputing Portal [32]. These stations are located within a box defined as 11.90–14.9 E, 40.5–42.0 N. These data represent true measurements, and as such provide a robust reference. However, the irregular spatial distribution of the stations does not ensure a homogeneous coverage of the study area.

Data were preprocessed by applying a spatial filtering to select stations and quality filtering to check whether each station had recorded a meaningful quantity of data for the proposed analyses. Hence, stations were filtered out when missing data was above 30%. Because the present study focuses on T2m, only 40 stations were considered; their positions are shown in Figure 2.

The considered dataset is constituted by T2m data from the whole winter season, from 1 December 2022 to 28 February 2023, at hourly frequency; hence, the array is composed of 2160 time steps (rows) and 40 stations (columns). The EOF model of the observed data was trained on this array, as explained in following sections.

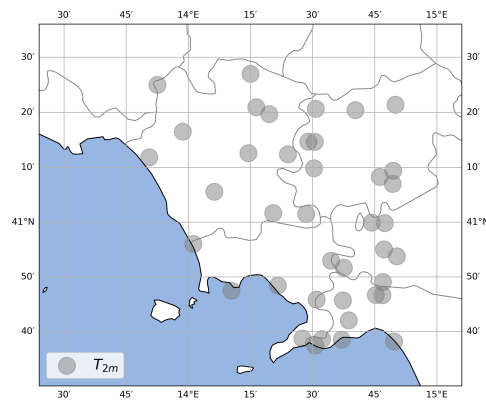


Figure 2. Location of automated weather stations for T2m, from Mistral Portal [32].

2.5. Downscaled Reanalysis

Model evaluation was also performed against VHR-REA-IT [24], a gridded hourly dataset over Italy for recent climate (1981–2023) available through the CMCC Data Delivery System (<http://dds.cmcc.it>, accessed on 17 January 2026). This dataset was obtained by means of dynamical downscaling of ERA5 Reanalysis [33] at a convection-permitting scale (0.02° , about 2.2 km) with the COSMO-CLM Regional Climate Model [34], including the TERRA-URB module for proper representation of urban areas.

It is evident that this product has its own biases and is affected by model-dependent assumptions. For this reason, it has been used as a comparative tool and is employed only to complement the observational analysis and facilitate comparison of spatially coherent and dynamical patterns.

3. Statistical Methods

Proper Orthogonal Decomposition (POD) is a technique for computing a linear basis of vectors that is optimal in some sense; in other words, it provides the best representation of a given dataset in a different reference frame originated by the POD basis vectors. A detailed mathematical formulation for its derivation can be found in [35,36]; for this reason, only the fundamental concepts and relations are recalled in Section 3.2. Section 3.3 describes the process of projecting anomaly fields derived from external datasets onto EOF subspaces. This process allows the abilities of a given trained EOF model for resolving the features of an external dataset to be described, thereby highlighting any differences in describing the same weather event.

3.1. Data Processing

Both the output of numerical models and the observational data require a preprocessing step before being analyzed using statistical methods.

For the numerical models, two-dimensional hourly T2m fields (C) from both the down-scaled ERA data and the ICON configurations share the same grid as the downscaled ERA5, namely, a grid with $ngrid = lon \times lat = 178 \times 193 = 35354$ grid points. Consequently, ICON data were upsampled to this grid using the weighted inverse-distance method available in the CDO software package [37]. The two-dimensional fields were then flattened to obtain an array $X_{nwp,models}$ of $(ntimes, ngrids)$, where the models (ERA5, ICON-REF, ICON-STAT, ICON-STAT) have dimensions (2160, 35354).

Regarding observational data, the array X_{obs} of dimension $(ntimes, nstations)$ with size (2160, 40) exhibited some data loss even after the filtering procedure described in Section 2.4. The fraction of missing records is equal to 0.4%, i.e., 362 against 86,440 T2m records). These missing values were filled using a gappy-POD algorithm [38] implemented in Python 3.9.

To define a snapshots matrix (see Section 3.2) of anomalies (S), data were centered and scaled to obtain a unit variance:

$$S = \frac{X - \bar{X}}{\sigma} \quad (1)$$

where $\bar{X}$ and σ represent the row-wise mean and standard deviation, respectively.

As regards the training of models based on the ground stations dataset (Section 4.2.2), data related to numerical $X_{nwp,models}$ were obtained sampling each of $X_{nwp,models}$ at the 40 cell center points, with a nearest-neighbor algorithm applied to the the long-lat values of the stations in X_{obs} . Thus, the snapshot matrices analyzed with EOF are $S^{fields}[ntime, ngrids]$ for each numerical model, and $S^{obs}[ntime, nstations]$ for the weather stations dataset and the sampled numerical data from the models.

3.2. EOF

The datasets are organized in a snapshot array $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$ which collects the available data in a column-wise manner, where n is the number of experiments or numerical simulations. It is assumed that each snapshot element S_k is obtained by sampling the model output in time; hence, $\mathbf{t} = \{t_1, \dots, t_m\}$, which in the present case represents the hourly T2m values. The combination of the sets $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m\}$ and $\{S_1, S_2, \dots, S_n\}$ represents the so-called training set of the database.

The computation of a POD basis consists in searching a set of orthonormal vectors $\phi_1, \phi_2, \dots, \phi_n$ such that the following relation holds:

$$S_j = \bar{S} + \sum_{i=1}^n \alpha_i(\mathbf{t}_j) \phi_i = \bar{S} + \sum_{i=1}^{\hat{n}} \alpha_i(\mathbf{t}_j) \phi_i + \epsilon_{\hat{n}}^j = \bar{S} + \sum_{i=1}^{\hat{n}} (S_j, \phi_i) \phi_i + \epsilon_{\hat{n}}^j$$

where $\bar{S}$ is a base solution [36,39] (i.e., an average field), the scalars α_i are the unknown POD coefficients, $\hat{n}$ is the number of preserved modes (with $\hat{n} \leq n$), and ϵ is the *optimal* error, i.e., the smallest possible. Indeed, for any set of orthonormal vectors $\psi_1, \psi_2, \dots, \psi_n$, the optimality condition is expressed as

$$\epsilon_{\hat{n}} = \sum_{j=1}^n \left\| S_j - \bar{S} - \sum_{i=1}^{\hat{n}} (S_j, \phi_i) \phi_i \right\| \leq \sum_{j=1}^n \left\| S_j - \bar{S} - \sum_{i=1}^{\hat{n}} (S_j, \psi_i) \psi_i \right\|.$$

In practice, the POD decomposition is obtained by taking the Singular Value Decomposition (SVD) of $\mathbf{P} = \{S_1 - \bar{S}, S_2 - \bar{S}, \dots, S_n - \bar{S}\}$:

$$\mathbf{P} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U} \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_M \\ 0 & \dots & 0 \end{pmatrix} \mathbf{V}^T \quad (2)$$

with the singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M \geq 0$. The POD basis vectors, also called POD modes or eigen-functions, are the first n column vectors of the matrix $\mathbf{U}$, while the POD coefficients $\alpha_i(\mathbf{t}_j)$ are obtained by projecting the snapshots onto the POD modes:

$$\alpha_i(\mathbf{t}_j) = (S_j - \bar{S}, \phi_i). \quad (3)$$

Usually, the singular values rapidly become small; then, a small number of basis vectors are adequate to reconstruct and approximate the snapshots, as they preserve the most significant ensemble energy contribution. In this way, POD provides an efficient mean of capturing the dominant features of a system with multiple degrees of freedom and

representing it to the desired precision by using the relevant set of modes. The reduced-order model is derived by projecting the original model onto a reduced space spanned by only some of the proper orthogonal modes or POD eigenfunctions. This process realizes a kind of lossy data compression by truncating the POD sum to $\hat{n} < n$:

$$S_j \simeq \bar{S} + \sum_{i=1}^{\hat{n}} \alpha_i(t_j) \phi_i. \quad (4)$$

A criterion for truncating the sum in Equation (4) is usually based on the total information content, which is provided by the following equation:

$$E_{\hat{n}} = \frac{\sum_{j=1}^{\hat{n}} \sigma_j^2}{\sum_{i=1}^n \sigma_i^2}. \quad (5)$$

Of course, if $\hat{n} = n$, then $E_{\hat{n}}$ is equal to unity and all modes are preserved. The amount of $E_{\hat{n}}$ can be defined as a spectral energy content. Often, the threshold value is set to 0.99, i.e., the 99% of the information content is retained, meaning that $E_{\hat{n}} \geq 0.99$ and $E_{\hat{n}-1} < 0.99$. In this way, the complex nonlinear behavior of the original model (e.g., NWP models or observational datasets) is transformed into a linear sum that is still able to represent the original dataset in an optimal way while employing a reduced number of degrees of freedom (i.e., $\hat{n}$ scalars α_k and modes $\Phi_k, k = 1, \dots, \hat{n}$).

The SVD algorithm is available in standard LAPACK solver via *scipy.linalg.svd*.

3.3. Data Projection on Subspace

After the EOF model trained with the reference dataset has been built, the external NWP anomalies can be projected on the POD basis by an inner product. The number of modes to be used for the reconstruction is $\tilde{n} < n$, as usual. Recalling the theory in Section 3.2, the reconstructed dataset G^* can be recovered as follows:

$$G^* = \bar{S} + \tilde{\Phi}(\tilde{\Phi}^T(G - \bar{S})). \quad (6)$$

The scalar POD coefficients $\tilde{A} = \{\alpha_1, \dots, \alpha_{\tilde{n}}\} = \tilde{\Phi}^T(G - \bar{S})$ are obtained by projecting the G vector, after subtraction of the database mean (i.e., the T2m anomalies), onto the preserved modes matrix $\tilde{\Phi}$. Equation (6) is basically a duplication of Equation (4); the only difference is that here the projected vector G is not included in the database, meaning that it may not belong to the subspace spanned by the POD modes.

The reconstructed field/dataset G^* can be compared to G by considering metrics based on the error $G - G^*$, and a curve is obtained by progressively increasing the number of preserved modes $\tilde{n}$. Consequently, a convergence plot can be organized showing the geometric errors versus the number of retained POD modes, as will be shown in the following section.

In this specific case, the external NWP anomalies G^* to be projected over the reference subspace were obtained with Equation (1), with the mean and standard deviation obtained from the reference dataset.

4. Results

This section is organized in the following manner: Section 4.1 evaluates ICON forecasts and ERA5 reanalysis downscaled at 2.2 km against the observational values in terms of diurnal cycles by means of classical statistical metrics; in Section 4.2, each dataset (observed, reanalysis, and forecasts) is processed with EOF and the results are compared to obtain the pattern of T2m in the full field (Section 4.2.1) and locally at the ground station points (Section 4.2.2).

The Results section ends with a comparison of the dataset over a common EOF subspace (Section 4.3) identified in the ERA5 EOF subspace for the full T2m pattern (Section 4.3.1) and observed data for local stations (Section 4.3.2). This last part can be considered as the main portion of the results, where the phase-space diagrams and error decomposition are presented.

4.1. Classical Evaluation and Validation: Diurnal Cycle

A first assessment of the capabilities of the different ICON configurations was performed considering model and observed T2m diurnal cycles obtained by averaging data over the whole winter season (from 1 December 2022 to 28 February 2023) over the 40 stations considered (shown in Figure 3). For this purpose, the nearest grid cell center to each of 40 stations was selected.

Table 1 reports different performance indicators related to the data in Figure 3 for the different ICON configurations and ERA5. It is evident that the ICON configurations overperform ERA5-2km. In particular, ICON-STAT is the best configuration in terms of mean average error (mae), normalized root mean square error (nrmse), central root mean square error (crms), correlation, and bias.

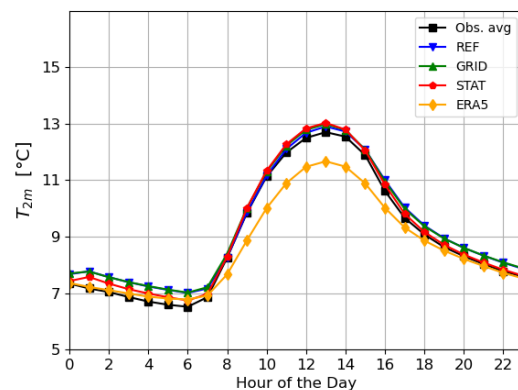


Figure 3. Diurnal cycle of T2m averaged at selected stations in winter 2022–2023 (i.e., DJF), showing ground station data vs. 2-km downscaled ERA5 reanalysis [24] and ICON NWP reference (REF) along with optimized (GRID,STAT) configurations outputs.

Table 1. Values of performance indicators related to T2m diurnal cycles in winter 2022–2023 and selected stations for different model configurations against observational data.

Model	Mae	Bias	Nrmse	Crms	Corr_Coefficient	StdRatio
ERA5	0.435	0.232	−0.353	0.0974	0.995	0.784
I-REF	0.306	0.0787	0.306	0.0563	0.998	0.950
I-GRID	0.340	0.0611	0.340	0.0587	0.999	0.960
I-STAT	0.187	0.0515	0.187	0.0349	0.999	1.014

4.2. EOF Model Comparison

In this section, EOF models trained on different datasets are compared in order to assess their capability in extracting, describing, and synthesizing the simulated and observed T2m pattern features. This model comparison was carried out by looking at the variance explained with respect to the EOF-decomposed time coefficients and the spatial eigenvectors obtained with the EOF analysis. In this way, the correlation with a reference model is evaluated both temporally (by comparing the EOF time coefficients, or PCs) and spatially (by comparing the patterns of the EOF eigenvectors).

The first comparison in Section 4.2.1 is about T2M fields from reanalysis and forecast, $S^{fields}[ntime, ngrids]$. EOF models are trained over the complete NWP forecast fields for each models' output; the reference model is represented by EOF trained on the ERA5 dataset.

The second case (in Section 4.2.2) concerns models trained on a subset of grid points corresponding to the ground station locations $S^{obs}[ntime, nstations]$ from observed data and NWP sampled data; the reference model is represented by the observed values of T2M from the ground stations.

4.2.1. Model Based on NWP Fields

After processing the datasets used to train the EOF model, the SVD supplies the eigenvalues on which eigenvectors are ordered (Equation (2)). These values are usually associated with the information content (Equation (5)) of each linear EOF eigenvector extracted; in other words, it can be associated with the variance of the processed variables. Figure 4a shows the spectrum of the variance explained as a function of the EOF model number trained with the T2m dataset. It is found that for the ERA5 dataset, the first EOF captures about 85% of the variance of temperature anomalies. It can be noted that the three datasets provided by the ICON simulations do not differ significantly one another, as they retain about 81% with the first mode. The next-most energetic modes capture less than 7% for ERA5 and about 10% for the ICON configurations.

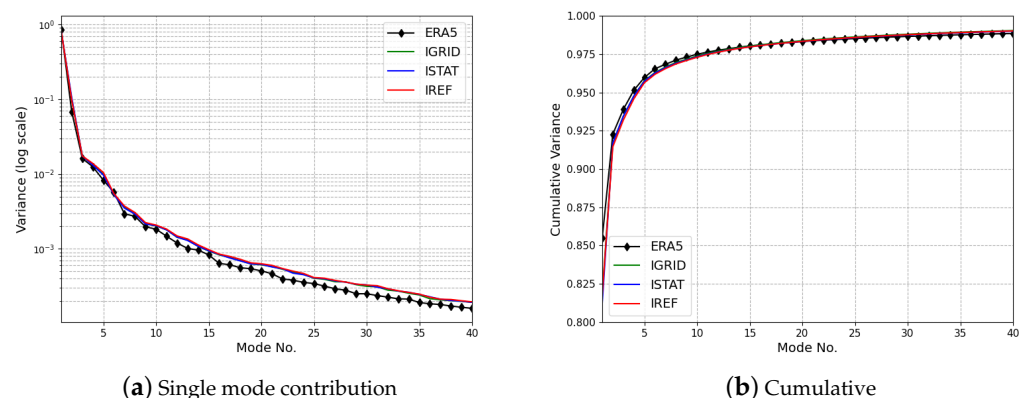


Figure 4. Normalized explained variance of the leading 40 EOFs of models trained using the T2m dataset with ERA5 and ICON configurations.

Figure 4b shows the cumulative information contents. The first three modes are able to describe 94% of the whole dataset information for ERA5. This trend is less steep for the three ICON outputs, with the features captured by the three modes achieving 93.5%. In general, each mode can be reconstructed with 98% of the variance by taking the first $\tilde{n} = 14$ modes, and with 99% by taking $\tilde{n} = 40$ modes. The reliability of the different dataset reconstructions with reduced modes, as evaluated with standard metrics, is shown in Table 2.

Figure 5 shows the hourly time series over the considered season of winter 2022–2023 for the first four T-EOF, also called time coefficients (Equation (3)) or PCs, obtained from each full dataset. These represent the realization of S-EOF eigenvectors on the related T2m field at a given time, measuring how the EOF eigenvectors influence the flow realization (i.e., the T2M anomaly) at a given timestep. The red line refers to the ERA5 reanalysis PC coefficients. It can be observed that ICON configurations show a qualitatively similar trend in time, and that differences of ICON with respect to ERA5 can be appreciated only for the fourth mode. For the higher modes, the ERA5 energy content is lower than ICON. This

behavior could be ascribed to the different grid resolution, since the smallest scales are better resolved by ICON than ERA5, distributing high energies to higher modes.

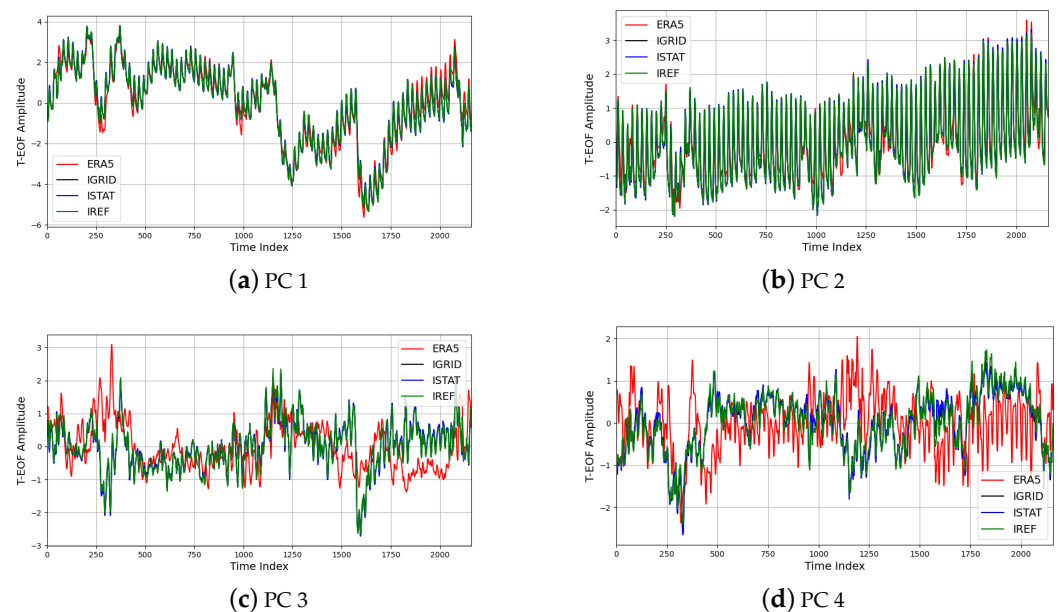


Figure 5. Time history in hourly-DJF of the first four leading PCs from EOF models trained with downscaled ERA5, ICON-REF, STAT, and GRID configurations.

In order to provide a quantitative comparison, we have calculated the correlation between ICON PC coefficients and those of ERA5. In this case, taking advantage of the spatial coherence being a reanalysis product, the downscaled ERA5 can be treated as a reference for pattern dynamics. The results for the three ICON configurations are shown in Figure 6a for $\tilde{n} = 14$, meaning that 98% of the energy is properly represented (Equation (5)). These results reveal that highly correlated modes can be found up to the eighth mode (with a Pearson coefficient higher than 0.6), while the third and fourth modes are weakly correlated. Figure 6b shows the weighted mean correlation of the PCs for ICON forecasts vs. ERA5; the weights used in this average operation are the ERA5 normalized eigenvalues (Figure 4a), which is done in order to take into account their relative contributions. ICON-REF is the most correlated configuration, i.e., it can be considered the configuration that is more similar in time to ERA5 with respect to ICON optimized configurations (the less correlated configuration is ICON-STAT).

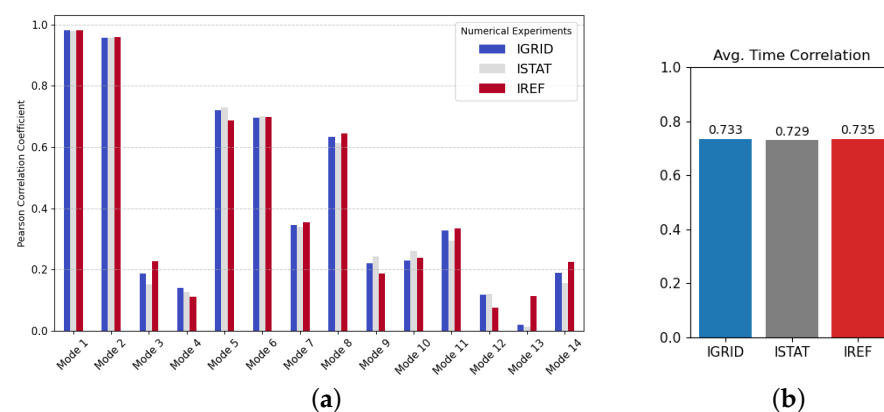


Figure 6. Time correlation between ICON configuration PCs and the corresponding ERA5 reference PCs and weighted averaged correlation: (a) correlations for the single PC and (b) weighted averaged correlation values up to $\tilde{n} = 14$.

The space patterns of the first four EOFs of T2m are shown in Figure 7, where they are related to the ERA5 (left) and ICON REF (right) configurations.

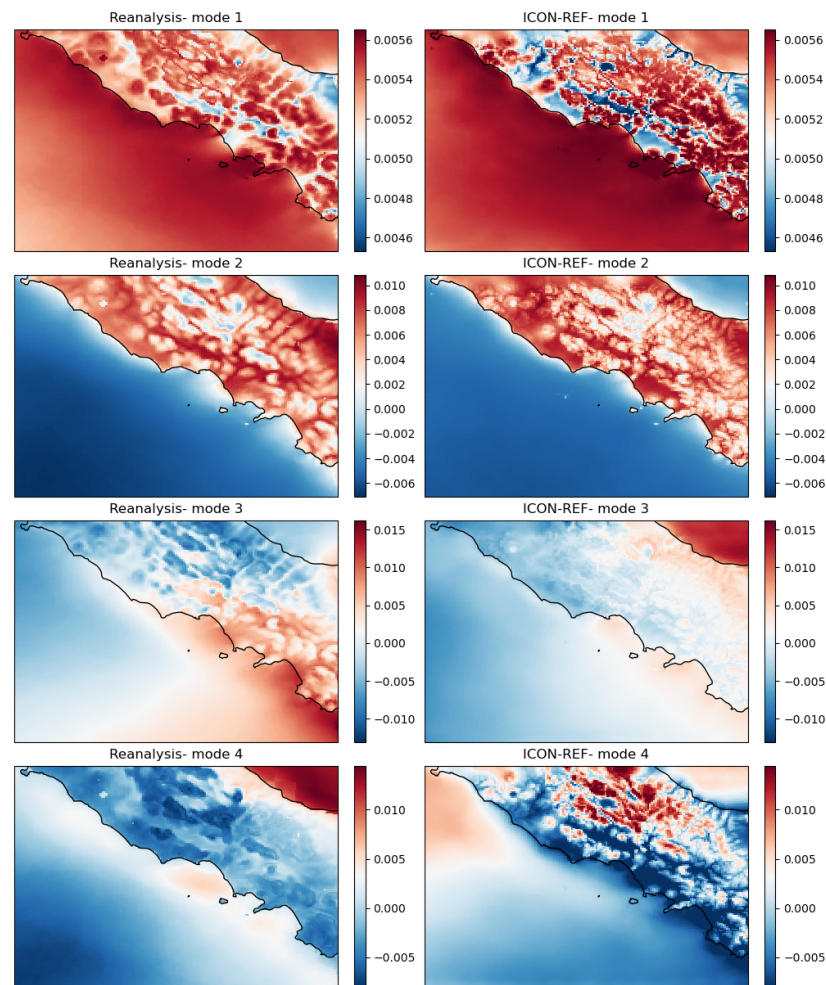


Figure 7. The first four EOF space patterns of T2m fields from the ERA5 (left) and ICON-REF (right) configurations.

It can be observed that the third and fourth modes behave differently from the re-analyses and ICON-REF; the first and second modes capture the sensitivity of T2m to the orographical features, while the ICON modes reveal the highest-resolution effects.

The correlations of the spatial patterns are shown in Figure 8, revealing that the second mode has the highest value of variance at 7%. Considering the weighted average, these are slightly less correlated than T-EOF compared to ERA5, as shown in Figure 6. The ICON-REF configuration has the highest correlation in space, while ICON-STAT has the lowest.

Figure 9 shows the spatial differences of the T2m EOF patterns related to the ICON optimized configurations (STAT and GRID) with respect to ICON-REF.

Performance indicators related to the dataset reconstruction by taking $\tilde{n} = 14$ within Equation (4) are shown in Table 2; the best data recovery according to the performance indicators of NRMSE and MAE is attained by the EOF model trained with ICON-STAT. It must be considered that when the reconstruction is performed with the whole subspace of EOF, the resulting error is null. These considerations are useful in supporting our discussion of projection errors in Section 4.3.

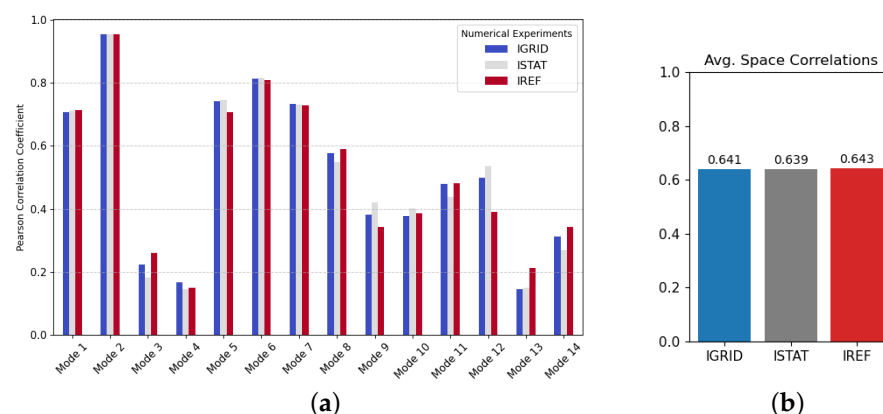


Figure 8. Space correlation of EOF eigenvectors with ERA5 from models trained with the ICON-REF, STAT, and GRID configurations: (a) correlations for single S-EOF and (b) weighted averaged correlations up to $\tilde{n} = 14$.

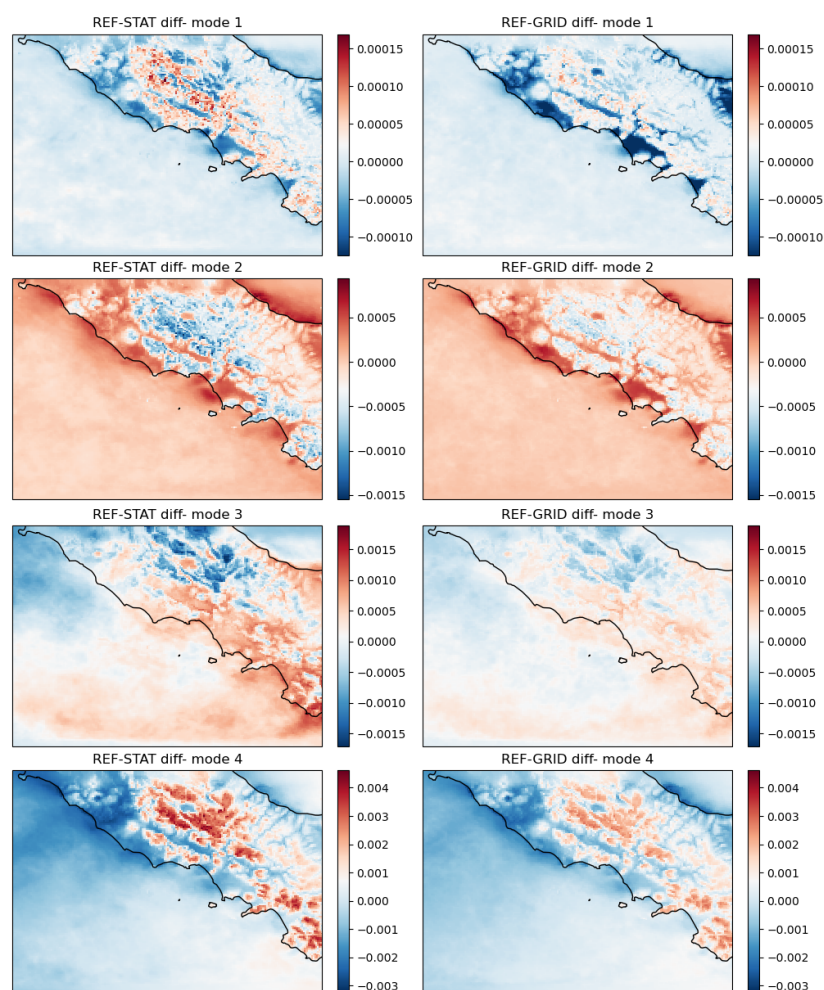


Figure 9. Spatial differences of T2m EOF patterns between ICON configurations for the first four eigenvectors: REF vs. STAT (left); REF vs. GRID (right).

Table 2. Performance indicators related to the reconstruction of T2m fields with $\tilde{n} = 14$; for each model, metrics are based on the difference between the reconstructed dataset with $\tilde{n}$ and the full dataset.

Model	RMSE	MAE	Expl. Variance
ERA5	0.143	0.106	0.98
I-REF	0.146	0.107	0.98
I-GRID	0.148	0.108	0.98
I-STAT	0.160	0.106	0.98

4.2.2. Model Based on Ground Stations Dataset

This section shows results related to five EOF models (observed, reanalysis, and three NWP ICON configurations) trained on the nearest-station set of points $S^{obs}[ntime, nstations]$, with the resulting S-EOF that can be considered as spatial eigenvectors defined on a sparse observational network. They share a common time frame with the previous analyzed T2M fields, but with the resulting spatial modes that represent patterns sampled by the observational network rather than continuous spatial fields.

Figure 10a shows the explainable variance as a function of the number of modes for the observational dataset of weather stations compared with ICON output. The first EOF mode from the observational data captures about 84% of the variance, while ICON outputs show about 94% of the energy. From the second mode onward, the observational dataset has more energy (i.e., 7% on the second and 1% on the third), while reanalysis and NWP are lower (about 3% and 0.6%, respectively).

Figure 10a shows the cumulative energy, revealing that the numerical dataset needed eleven modes to reach 99% of content, while the observational dataset needed 25.

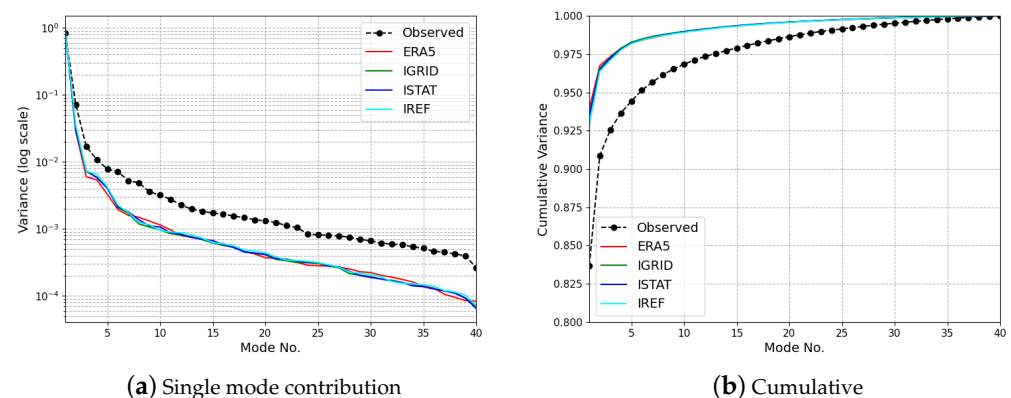


Figure 10. Spectrum and integrated variance explained as a function of the EOF model number trained with the T2m dataset from observational data, downscaled ERA5 at 2Km, and ICON output with the three configurations.

Figure 11 shows the time series of the first four scalar EOF PCs obtained from the ground stations dataset (black lines) along with the EOF model trained on the numerical dataset. It can be noted that there is a good agreement between observational and numerical PCs for the first two modes, while from the third mode onward the PCs show poor data overlap.

Figure 12a shows the correlation values of numerical PCs against the observational model for the first eleven modes (99% of variance). It can be seen that the correlation drops from the third mode, especially for ERA5. Figure 12b shows the corresponding

weighted values, revealing better performances of the ICON models, in particular the ICON-STAT configuration, compared ERA5.

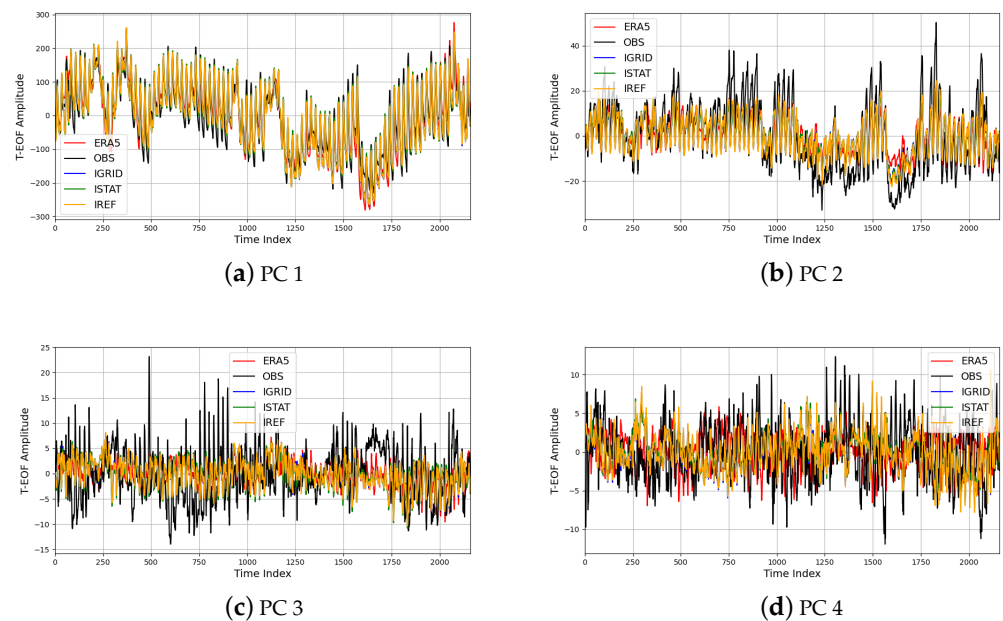


Figure 11. Time history in hourly-DJF of PCs from models trained on observational ground stations dataset and from related nearest-cell points of downscaled ERA5 and ICON (REF, STAT, and GRID) configurations.

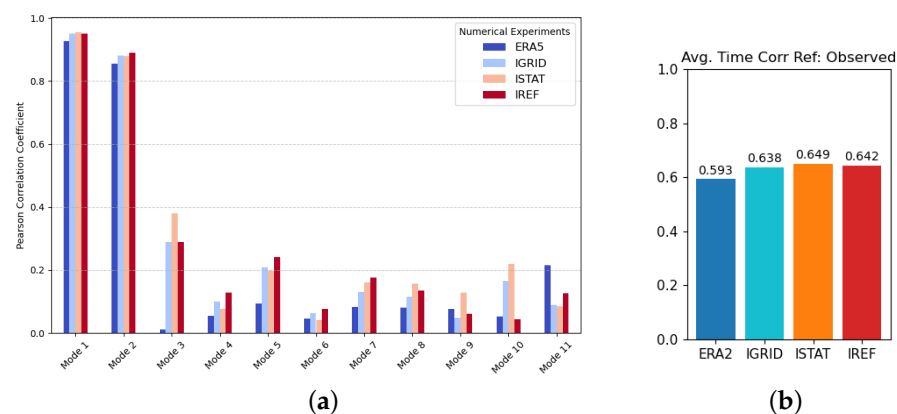


Figure 12. Time correlation of EOF coefficients from models trained with ERA5 and ICON (REF, STAT and GRID) configurations and observational data up to $\tilde{n} = 11$ (99% of variance): (a) correlations for modes and (b) weighted averaged correlations up to $\tilde{n} = 11$.

The discrete spatial patterns of the first four S-EOF obtained from the ground stations dataset are shown in Figure 13.

The correlation of numerical EOF patterns with observational EOF is worst if compared with full T2M fields, but this is well-expected because the data are strongly local and noisy. Figure 14 shows that the correlation trend with the modes is irregular and generally very poor, with the exception of the second EOF mode. Evaluating the weighted average correlations (according to normalized eigenvalues; see Figure 10, ICON-REF has the best correlation with the observed S-EOF.

Table 3 shows the values of performance indicators related to the reconstruction of each dataset relative to the ground station points by taking $\tilde{n} = 11$ in Equation (4). This metric reveals the ability of each model to represent itself given the same number of features.

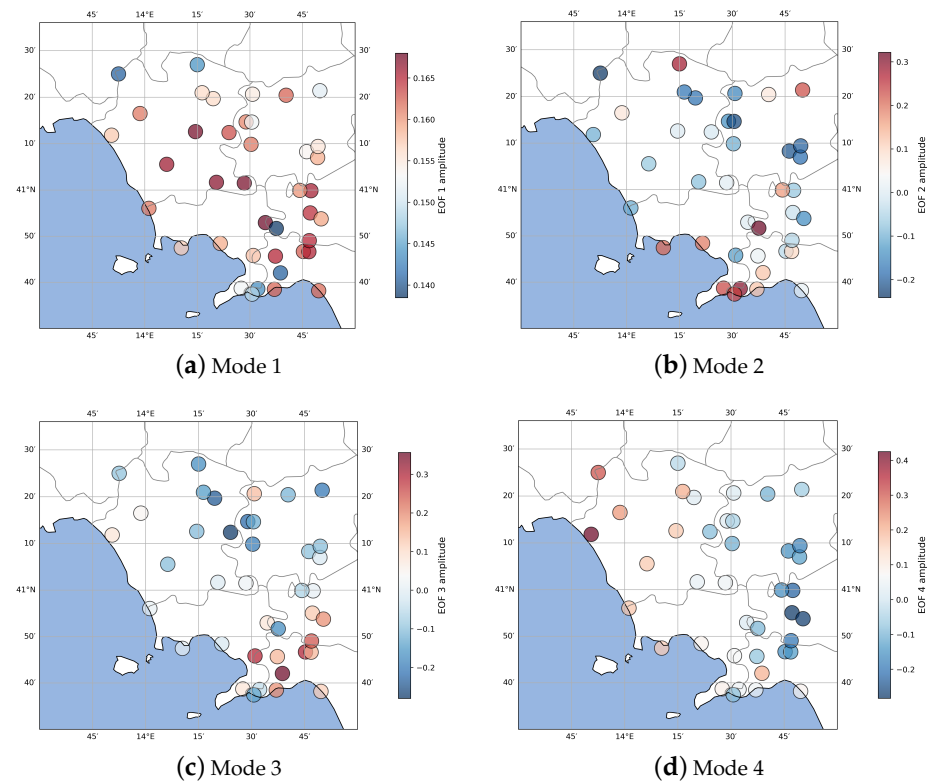


Figure 13. Spatial pattern of EOF spatial eigenvectors from models trained on the observational ground stations dataset.

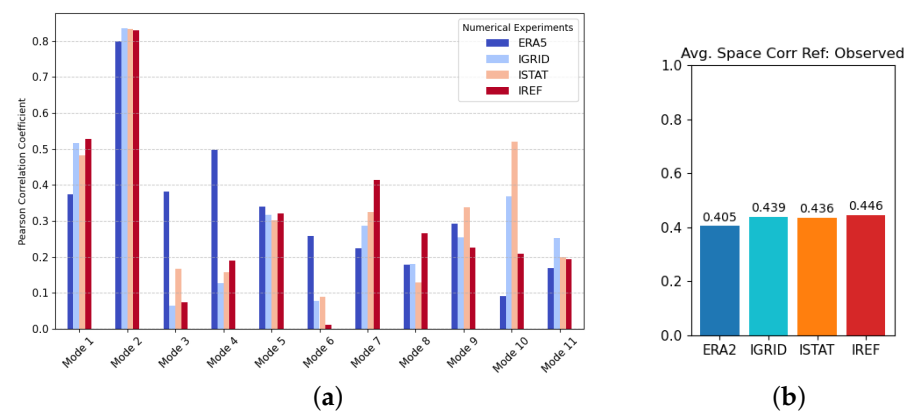


Figure 14. Space correlation in EOF eigenvector with the observation model, from models trained with ERA5 and Icon (REF, STAT and GRID) configurations: (a) correlations for modes and (b) weighted averaged correlations up to $\tilde{n} = 11$.

Table 3. Metrics of reconstructed ground stations dataset.

Model	RMSE	MAE	Expl. Variance
ERA5	0.097	0.073	0.99
I-REF	0.100	0.074	0.99
I-GRID	0.097	0.072	0.99
I-STAT	0.098	0.073	0.99
Observed	0.170	0.127	0.97

4.3. EOF Projection on Reference Subspaces

Several works in the fields of meteorology, climatology, and in fluid dynamics more generally have evaluated the performance of NWP models by means of field projections onto the EOF subspace, particularly where the EOF is trained using observational data [28–30]. This method provides a tool for assessing the ability of NWP models to capture the dominant patterns of variability observed in the atmosphere.

A fair comparison between observed and numerical signals can be achieved by plotting the PCs on their phase space, as was done for example in Molteni et al. [28], who validated the ensemble performances using EOFs. This is possible because EOFs reduce the dimensionality of the atmospheric phase space, enabling comparison of the probability density functions of observed and modeled PCs, which in turn provides an estimate of the quasi-systematic error of the model. This approach is considered a consistent diagnostic framework for seasonal comparisons. In fact, Lee [29] showed that a common basis function method allows models and observations to be compared reliably; modes of variability may be compromised or spread across multiple modes in model EOF space, since often the modes are not well-separated in either observations or models. Biggerstaff et al. [30] drew the EOF phase space related to micro-physical variables in order to more thoroughly assess the differences between observations and simulations.

In the present work, the projection onto a common basis is applied twice. The first projection uses a reference dataset consisting of ERA5 reanalysis data, projecting the full T2m model anomalies onto the related EOF subspace. The second projection uses an EOF model trained with a dataset from ground stations, against which the NWP-based models extracted at the nearest grid cell centers are compared.

In addition to the analyses presented above, particular attention is given to the reconstruction error, as it reflects the reference model's capability to reconstruct projected datasets.

4.3.1. Dataset Projection on Reanalysis Subspace

In this section, ERA5-2km is assumed as the reference model and the EOF span is the reference system on which ICON forecast datasets obtained with the three previously-described configurations have been projected on. Their representation on this new common span could support their quantitative comparison; from the inner product between EOF sub-span and DB realizations, a new time history of PCs will result, which can then be fairly compared.

The PCs' coefficient time history of projected T2m datasets were compared with the time history obtained from the reanalysis in order to highlight the different behavior of the dataset along with the similarity and correlation with the observed features for each of the numerical realizations.

The trends shown in Figure 15 are similar to those of Figure 5; in fact the time history of ERA5 are coincident but the curves related to ICON configurations are conceptually different, as explained in Equation (6). In Figure 15, the red lines represent the ERA5 reanalysis EOF coefficients.

Figure 16a shows the correlation of the PCs resulting from the dataset projections on the ERA5 EOF subspace. An almost regular decreasing trend in correlation is observed according to the EOF modes order, which can be interpreted by assuming that the highest-frequency features are less correlated with the ERA5 subspace. Figure 16b shows the weighted averaged correlation, in which weights are assumed as the fraction of the variance associated with each EOF mode. It can be seen that the best-correlated ICON configuration is ICON-REF.

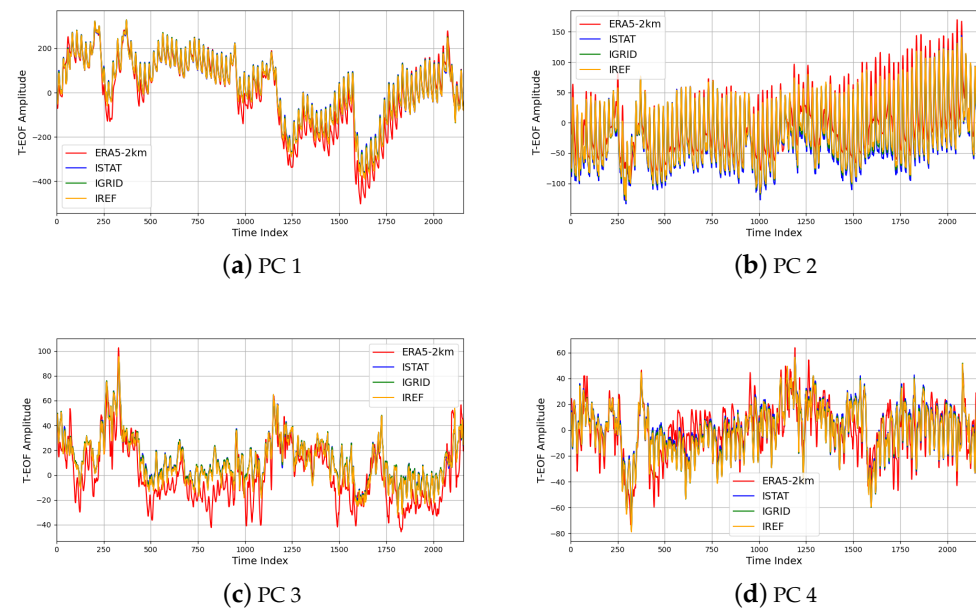


Figure 15. Time series of the first four EOF time coefficients from ICON datasets projected onto the ERA5 EOF subspace.

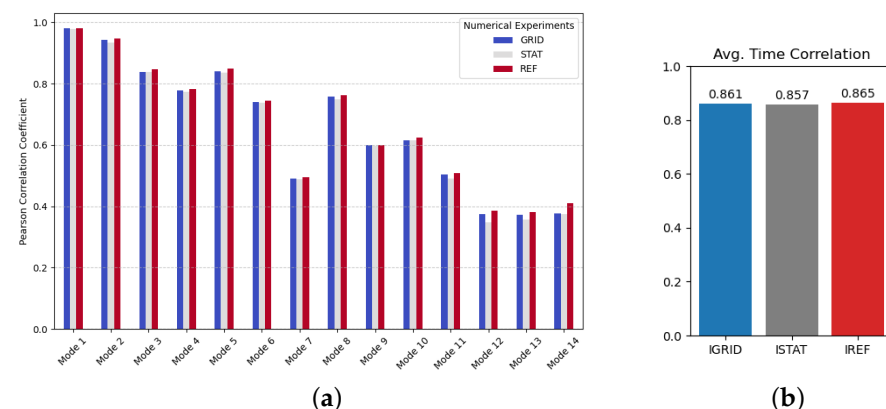


Figure 16. Time correlation as a function of modes between EOF coefficients with reanalysis model, with models trained by the ICON REF, STAT, and GRID configurations: (a) correlations for modes and (b) weighted averaged correlations.

A T2m field can be reconstructed as a linear combination of the EOF time coefficients with the EOF eigenvectors of the model trained on ERA5, as described in Equation (6). This reconstructed dataset represents a measure of the reanalysis' ability to solve the features of the ICON datasets.

The difference between the reconstructed fields and the ERA5 training dataset should help in finding the most similar models to the reanalysis. Through a convergence history of a given metric, it should be possible to individuate the features that help to pull the reconstructed fields to the reference models, and similarly the features that the model is not able to solve. These shortcomings are well expected due to the different space resolution of the reanalysis compared to ICON forecasts. Furthermore, the different contributions of the ICON configuration can be identified as converging to or diverging from the reanalysis.

Figure 17 shows the RMSE and the MAE error convergence trend as functions of the EOF modes, with $\tilde{n} = 40$. The blue lines are related to ERA5 reconstructions; they converge towards a zero error, since these are the data on which the EOF model is defined and ICON anomalies were projected. It is interesting to analyze the behavior of an externally

reconstructed ICON dataset (see Figure 15); in fact, it is evident that the errors decrease towards a minimal value achieved with $\tilde{n} \approx 12$. Additional modes cause the error to increase again, meaning that for $\tilde{n} > 12$, further terms do not add useful information that could help in reducing the gap from the ERA5 dataset. This could be ascribed to the ERA5 subspan, which from the twelfth mode onward is not able to ‘solve’ features in the ICON datasets. Regarding the ICON models, the best performance (as characterized by minimum RMSE and MAE) is achieved by the ICON-REF configuration, followed by ICON-GRID and ICON-STAT.

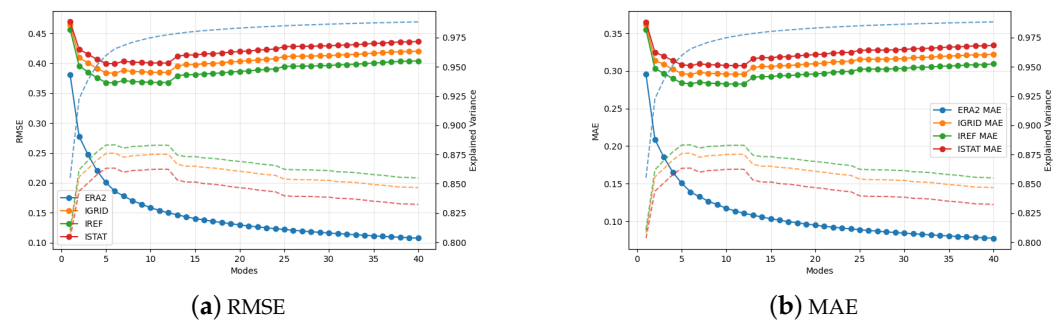


Figure 17. Convergence of metrics in dataset reconstruction with projected PCs on the ERA5 subspace from data of REF, STAT, and GRID Icon configurations, with explained variance (dashed lines).

The dotted curves in Figure 17 relate to the explained variance, the trend of which is specular to the error line. ICON anomalies solved in the ERA5 EOF subspace capture less variance than ERA5 itself (87% for ICON-REF), then it decreases after twelve modes.

4.3.2. Dataset Projection on Observed Subspace

In this section, an EOF model trained on observed ground station data is considered as reference model, while the numerical datasets are analyzed by projecting anomalies on its subspace using the operation explained in Equation (6).

As a general comment, the observed ground station data suffer from being very localized. They are more affected by noise compared to the smooth and leveraged information of ERA5 reanalyses or NWP outputs, as discussed in Section 4.2.2. Figure 18 shows the time series of the first four scalar PCs, i.e., $\tilde{\mathbf{A}}$, obtained from each numerical dataset using Equation (6). The black line represents the observed EOF time coefficients. Noticeably different behavior of the numerical projections with respect to observational values is recorded in their time evolution, starting from the third PC.

Figure 19 shows the correlation values of PCs between observational data and numerical PCs as a function of the numbers of modes. A decreasing trend is observed, which is even more evident when compared to Figure 16; this is because the weighted averaged values of the correlations are lower than those in Figure 16. The best values are achieved with the ICON-STAT configuration.

Figure 20 shows a comparison between the observed and projected PCs in their phase space; the black dots and blue dots have coordinates (α_1, α_2) and $(\tilde{\alpha}_1, \tilde{\alpha}_2)$, respectively, the latter derived from Equation (6). This is a common choice in the literature for this kind of plot to correlate, e.g., seasonal forecasts with interconnection indices [28–30], which are here inherited to analyze reanalysis and forecasts. Furthermore, each axis reports the Probability Density Function (PDF) that supports the verification of variance of each PCs; specifically, the PDF of PC1 is set on the top and the PDF of PC2 is plotted on the right side. Figure 20a compares the ERA5 and observational data; according to the top PDF, the ERA5 shows a variance almost equal to observational PC1, while on the left side a narrow curve highlights the concentration of PC2 for ERA5 in a central phase space position, with a

reduced density when $PC2 > 0$. Figure 20b is related to ICON-REF and shows an asymmetry in frequency for $PC2$, reduced for $PC2 < 0$, indicating the attitude in simulating phenomena with smaller variance described with $PC2$ (or α_2). Figure 20c,d, related to ICON-GRID and STAT, respectively, describe the variance features for $PC2$. It is worth noting that shape of the PDF is less narrowed than in the REF configuration, moving towards the shape of the PDF for the observed data.

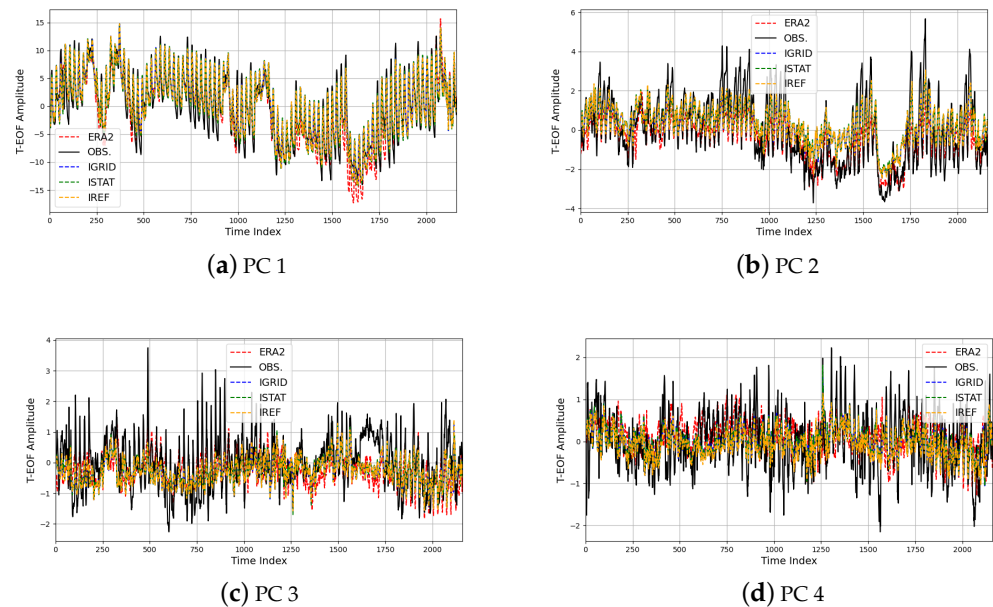


Figure 18. Time history in hourly-DJF of the first four EOF time coefficients from ICON and ERA5 T2M anomalies projected on the observed EOF subspace.

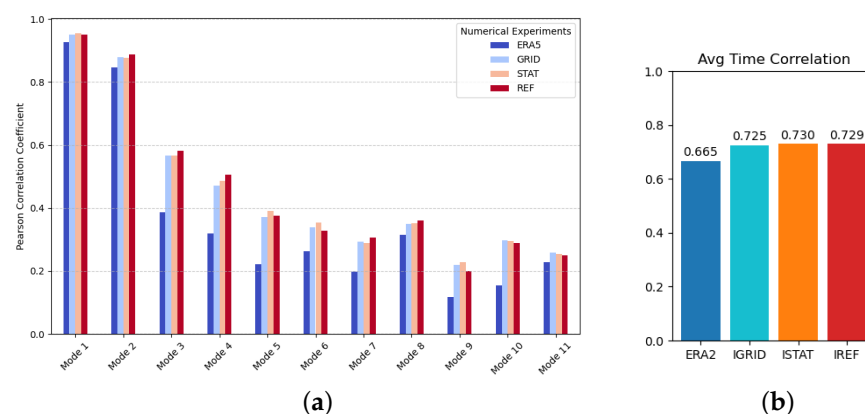


Figure 19. Time correlation as a function of modes in EOF time coefficients with the observed data model from models trained by ERA5 and ICON (REF, STAT and GRID) configurations up to eleven modes: (a) correlations for modes and (b) weighted averaged correlations.

Figure 21 compares the PCs with coordinate (α_3, α_4) (black dots) and the numerical dataset with coordinates $(\tilde{\alpha}_3, \tilde{\alpha}_4)$ (blue dots). For the fourth mode, the variance captured by numerical models is smallest compared to the observational dataset, as shown in each plot of the figure. Moreover, it is worth noting that while ERA5 and ICON-REF show a marked tendency to spread over the phase space asymmetrically, the optimized configurations appear more centered, especially for the fourth PCs.

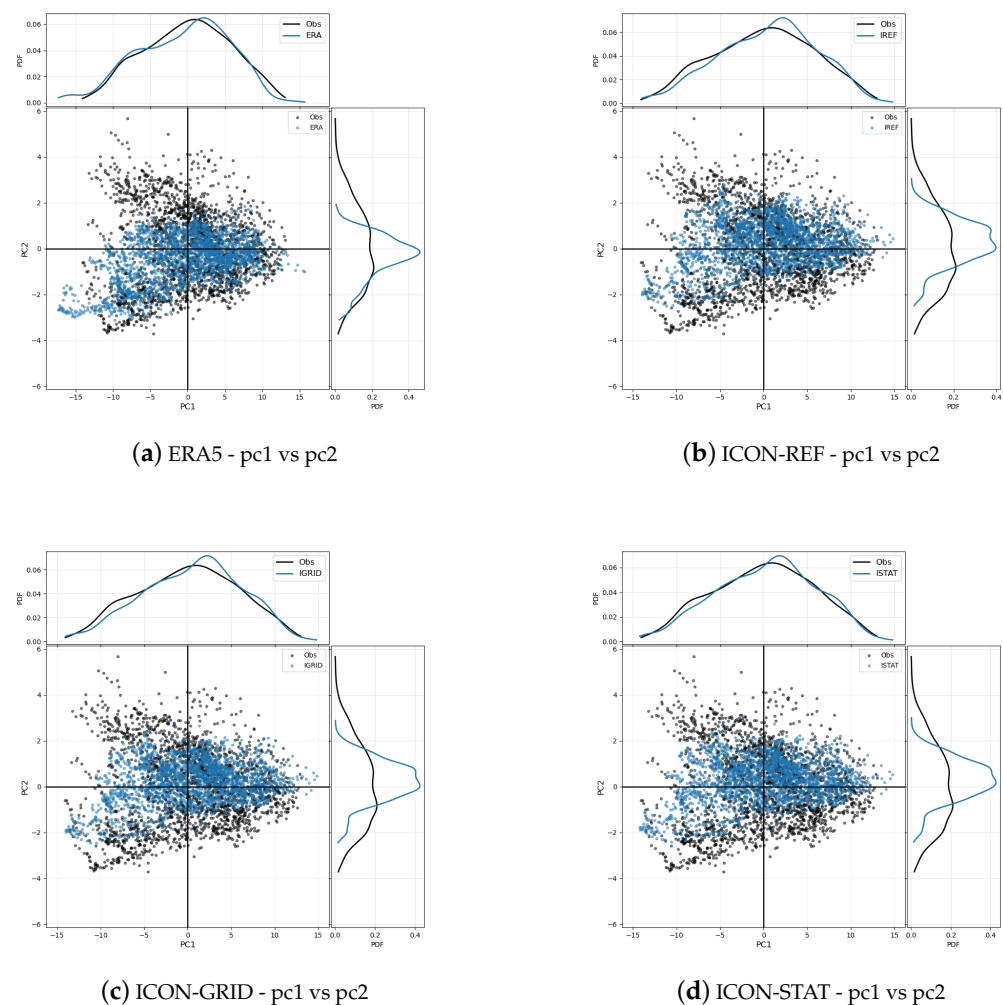


Figure 20. Phase space of PCs 1 (x-axis) and 2 (y-axis), comparing observed and projected NWP data.

As already explained, the main aim of the present work is to find a possible relationship between EOF decomposed signals and modeled phenomena, the representation of which could successively support further parameter tuning in NWP models. To this end, a decomposition of NWP output biases against ground station data was performed. The diurnal cycle of T2m was decomposed using Equation (6), and the bias of the numerical values against the observed signal was evaluated by using Equations (4) and (6) to quantify the contribution of the prevailing expansion terms to the distance, say e_i , with observed data, i.e., $S - G^*$:

$$e_i = (\alpha_i - \tilde{\alpha}_i)\phi_i. \quad (7)$$

Figure 22 shows the diurnal cycles of e_i with i varying from 1 to 6, which is a representation of the diurnal cycle of T2m bias (i.e., in this case, values of Equation (7) are expressed in °C). According to the ERA5 decomposition (black line), the bias analyzed by means of the first term e_1 shows the largest deviation from observed values (where $y = 0$ represents 'zero' error) during the first hours of the day. In ICON, it moves towards lower temperatures, with the worst result for ICON-REF (blue line). In the optimized configurations (ICON-STAT (green line) and ICON-GRID (red line)), the contribution is reduced, while the shape of the signal bias is preserved compared to ERA5. The second term e_2 again gives the largest contribution for the NWP comparison during daytime; ERA5 is the worst case, with the highest bias around noon. The third term e_3 exhibits a roughly

inverted pattern with respect to e_2 , with the optimized ICON configurations showing the smallest error. It is worth noting that the contribution of the higher e_i modes in terms of C decreases with increasing mode orders, which is as expected from the fraction of variance that they capture. While it is relatively straightforward to identify the modes that drive the deviation of ERA5 from the observed profiles, it is more difficult to discern the differences that distinguish the ICON configurations in cases such as e_2 and e_4 . A common feature is that the differences among the ICON configurations become more pronounced during night-time, with the exception of e_6 , and that they often they have opposite signs.

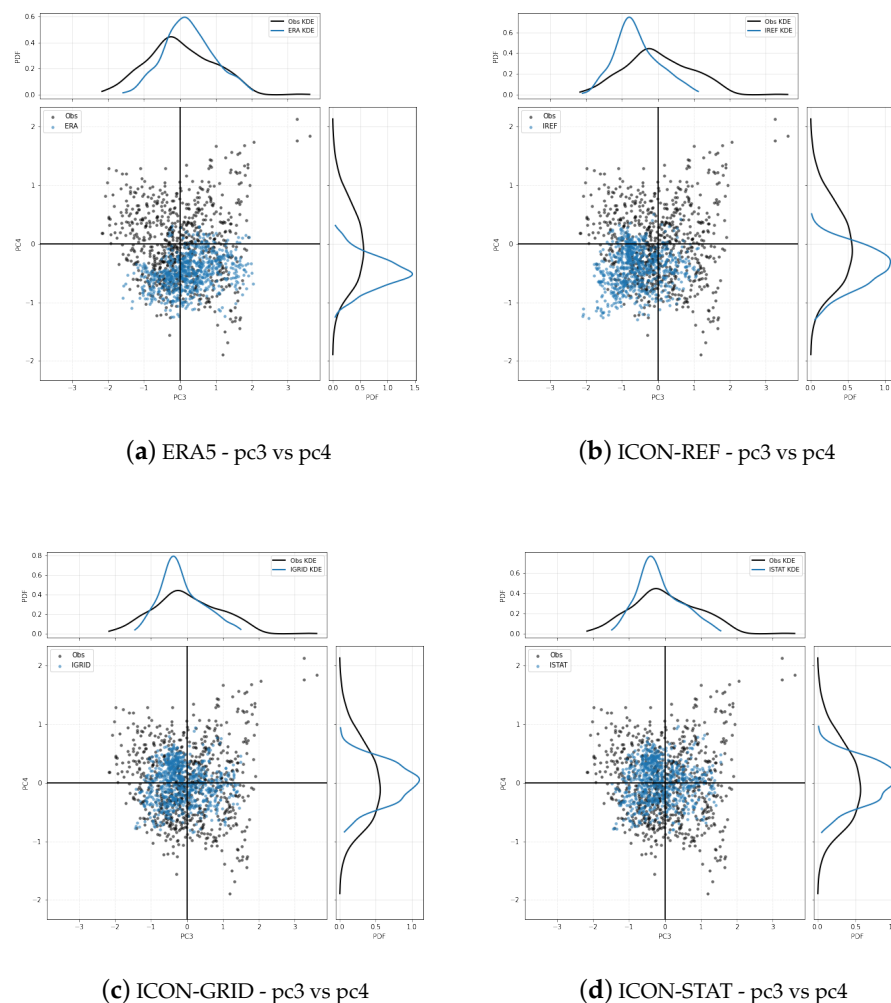


Figure 21. Phase space of PCs 3 (x-axis) and 4 (y-axis), comparing observed and projected NWP data.

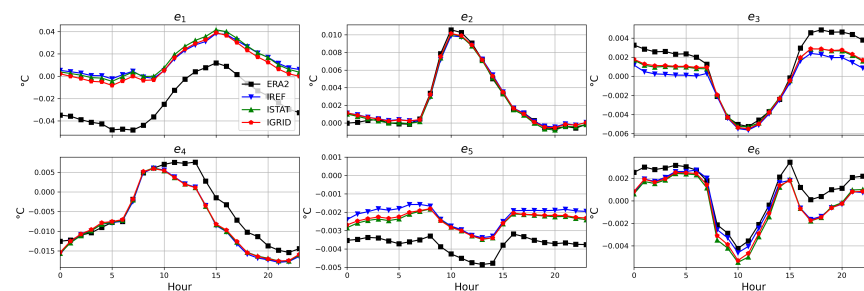


Figure 22. T2m bias valued with e_i of Equation (7) on the diurnal cycle with respect to the reconstructed observational dataset.

These components contribute to reducing error, especially during night-time, as can be seen in Figure 3. The term e_4 gives its major contribution during night-time, and the

optimized ICON configurations show a large influence of this term on the signal. Overall, e_i for $i = 1$ to 5 provide the main contributions driving the improved performances of the optimized ICON configurations, as shown in Figure 23. Figure 23 shows the RMSE and MAE error convergence trend for the external datasets as a function of the number of terms (Equation (4)) up to $\tilde{n} = 23$ (i.e., 99% of variance explained for observed dataset). Generally, the ICON configurations perform better than ERA5 (blue line), especially the optimized GRID and STAT datasets, with the latter being characterized by the minimum error level. In particular, these error indicators are characterized by a non-monotonic trend, with a point of inversion respectively added after two (ERA5 and ICON-REF) and five (GRID and STAT) modes in Equation (4). The most recognized patterns from the observed model are the ones projected from optimized configurations, up to fifth eigenvector.

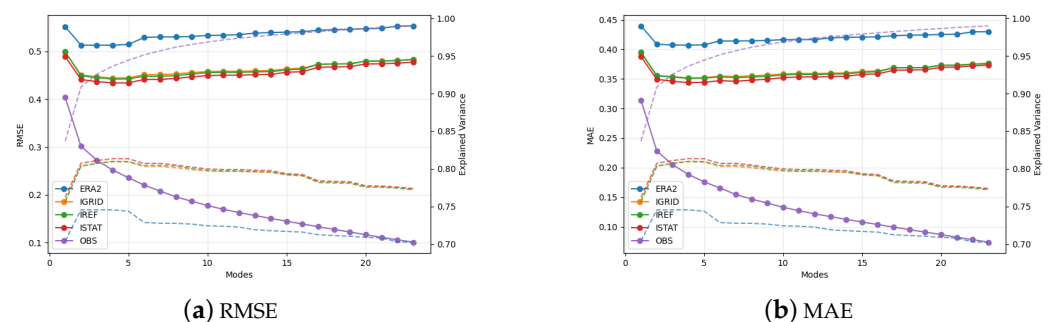


Figure 23. Convergence of metrics in dataset reconstruction with projected PCs on observed subspace, from data of ERA5 and Icon REF, STAT, and GRID configurations, with explained variance (dashed lines).

Adding further terms does not help with convergence to observed solutions, and the explained variance reaches its maximum values after the fifth EOF mode, reaching 82% of the observed model. From the signal error decomposition, almost all of the first modes act during night-time and contribute to enhancing the performance of ICON, in particular for ICON-STAT (see Figure 3); starting from sixth mode, the introduced bias is no longer recovered by the added terms.

5. Conclusions

In the present study, a detailed evaluation of previously optimized ICON-LAM configurations has been performed over a domain located in southern Italy by analyzing a series of 24 h forecasts of 2-m temperature over winter 2022–2023 through the lens of EOFs. Two reference datasets have been adopted for comparison: ERA5 reanalysis downscaled at 2.2 km, and data from a collection of ground stations. While a classical model validation using standard procedures showed that the optimized ICON configurations (particularly ICON-STAT) have strong capabilities in reproducing observational data, outperforming the reanalyses, our EOF analyses revealed a more detailed picture. In particular, the ICON-REF configuration exhibited the greatest similarity to the ERA5 reanalyses; however, the optimized configurations (especially ICON-STAT) demonstrated a higher skill in predicting the considered atmospheric variable, providing results that are a better match with observations. From the signal error decomposition, the first two modes are particularly effective during night-time hours and contribute to enhancing the performance of ICON, especially when ICON-STAT is employed. On the other side, although ICON-REF ensures good time correlation, this is penalized by poor correlation regarding spatial patterns.

ICON-GRID is characterized by a smoother response surface than ICON-STAT, reflecting reduced sensitivity to small-scale variability and localized extremes. Nevertheless, our optimization experiments indicate a coherent displacement of the parameter space

across configurations. In particular, the tuned setups were characterized by higher values of the minimum turbulent diffusion parameters ($tkhmin$ and $tkmmin$) compared to the reference configuration. This behavior enhances the persistence of turbulent kinetic energy under stable stratification, limiting the development of excessively strong near-surface temperature inversions. As a consequence, weak convective activity and associated precipitation are better represented. At the same time, the laminar heat resistance ($rlam-heat$) was reduced with respect to the reference configuration, leading to stronger surface–atmosphere heat exchange. The impact of this parameter is more pronounced under calm anticyclonic conditions with a stable nocturnal boundary layer, where local surface-layer processes dominate the thermodynamic evolution. Variations in $vosnow$ did not produce impact on the objective function within the range of investigated conditions, suggesting a secondary role of solid precipitation-related parameters compared to turbulence and heat exchange processes in controlling near-surface temperature performance.

It is worth noting that the focus of this work is on the identification of coherent large-scale patterns and systematic differences among experiments, rather than on the assessment of the statistical significance of local differences. While highly valuable, techniques for uncertainty estimation (e.g., bootstrap-based approaches) would substantially increase the methodological and computational complexity of the analysis, particularly for high-resolution EOF diagnostics, and are beyond the scope of the present study. Moreover, the inclusion of additional variables such as wind, humidity, or precipitation would allow for a more detailed analysis; however, such inclusion is beyond the scope of the present study, and would increase its complexity. In future work, it is our intention to address multivariable diagnostics.

As already explained, further analyses based on autoencoders will be presented in a companion paper (Part II, Cinquegrana and Bucchignani, in preparation). Autoencoders are a powerful tool belonging to the machine learning family and are able to overtake the linear properties of EOFs to support description of nonlinear features. Our forthcoming work will highlight the difference between these statistical tools in a practical way.

Author Contributions: Conceptualization, D.C. and E.B.; methodology, D.C. and E.B.; software, D.C.; validation, D.C.; formal analysis, E.B.; investigation, D.C.; data curation, D.C.; writing—original draft preparation, D.C.; writing—review and editing, E.B.; visualization, D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

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

References

1. Lorenz, E.N. *Empirical Orthogonal Functions and Statistical Weather Prediction*; Massachusetts Institute of Technology, Department of Meteorology: Cambridge, MA, USA, 1956; Volume 1.
2. Lumley, J.L. The Structure of Inhomogeneous Turbulent Flows. In *Atmospheric Turbulence and Radio Propagation*; Yaglom, A.M., Tatarski, V.I., Eds.; Nauka: Moscow, Russia, 1967; pp. 166–178.
3. Stockdale, T. SEAS5 User Guide. Technical Report. 2021. Available online: <https://www.ecmwf.int/en/elibrary/81237-seas5-user-guide> (accessed on 10 November 2025).
4. Dalelane, C.; Dobrynin, M.; Fröhlich, K. Seasonal Forecasts of Winter Temperature Improved by Higher-Order Modes of Mean Sea Level Pressure Variability in the North Atlantic Sector. *Geophys. Res. Lett.* **2020**, *47*, e2020GL088717. [CrossRef]
5. Golian, S.; Murphy, C. Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts. *J. Hydrometeorol.* **2022**, *23*, 1350–1363. [CrossRef]
6. Yu, Y.; Lin, Z.H.; Qin, Z.K. Improved EOF-based bias correction method for seasonal forecasts and its application in IAP AGCM4.1. *Atmos. Ocean. Sci. Lett.* **2018**, *11*, 499–508. [CrossRef]

7. Ferranti, L.; Magnusson, L.; Vitart, F.; Richardson, D.S. How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Q. J. R. Meteorol. Soc.* **2018**, *144*, 1788–1802. [\[CrossRef\]](#)
8. Solari, G.; Carassale, L.; Tubino, F. Proper orthogonal decomposition in wind engineering. Part 1: A state-of-the-art and some prospects. *Wind Struct.* **2007**, *10*, 153–176. [\[CrossRef\]](#)
9. Cinquegrana, D.; Iuliano, E. Efficient Global Optimization of a Transonic Wing with Geometric Data Reduction. In *Proceedings of the 35th AIAA Applied Aerodynamics Conference*; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 2017. [\[CrossRef\]](#)
10. Cinquegrana, D.; Iuliano, E. Investigation of adaptive design variables bounds in dimensionality reduction for aerodynamic shape optimization. *Comput. Fluids* **2018**, *174*, 89–109. [\[CrossRef\]](#)
11. Cinquegrana, D. Proper-Orthogonal-Decomposition-Based Leveraging of Reentry Vehicle's Surface Environment. *J. Spacecr. Rocket.* **2015**, *52*, 1563–1576. [\[CrossRef\]](#)
12. Cinquegrana, D.; Vitagliano, P.L. A reduced order model for boundary layer ingestion map prediction at fan inlet of rear-mounted engine nacelle. In *Proceedings of the AIAA Scitech 2021 Forum*; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 2021. [\[CrossRef\]](#)
13. Ramesh, S.S.; Lim, K.M.; Lee, H.P.; Khoo, B.C. Reduced-order modeling of stratospheric winds and its application in high-altitude balloon trajectory simulations. *J. Appl. Meteorol. Climatol.* **2017**, *56*, 1753–1766. [\[CrossRef\]](#)
14. Ferranti, L.; Molteni, F.; Brankovic, C.; Palmer, T.N. Diagnosis of Extratropical Variability in Seasonal Integrations of the ECMWF Model. *J. Clim.* **1994**, *7*, 849–868. [\[CrossRef\]](#)
15. Hannachi, A.; Finke, K.; Trendafilov, N. Common EOFs: A tool for multi-model comparison and evaluation. *Clim. Dyn.* **2023**, *60*, 1689–1703. [\[CrossRef\]](#)
16. Benestad, R.E.; Mezghani, A.; Lutz, J.; Dobler, A.; Parding, K.M.; Landgren, O.A. Various ways of using empirical orthogonal functions for climate model evaluation. *Geosci. Model Dev.* **2023**, *16*, 2899–2913. [\[CrossRef\]](#)
17. Grams, C.; Ferranti, L.; Magnusson, L. How to make use of weather regimes in extended-range predictions for Europe. *ECMWF Newsl.* **2020**, *165*, 14–19.
18. Tveito, O.E.; Huth, R.; Philipp, A.; Post, P.; Pasqui, M.; Esteban, P.; Beck, C.; Demuzere, M.; Prudhomme, C. *COST Action 733: Harmonization and Application of Weather Type Classifications for European Regions; Final Scientific Report*; Universität Augsburg: Augsburg, Germany, 2016.
19. Pandžić, K. Principal component analysis of precipitation in the Adriatic-Pannonian area of Yugoslavia. *J. Climatol.* **1988**, *8*, 357–370. [\[CrossRef\]](#)
20. Tipi di Circolazione | Consorzio LaMMA. 2025. Available online: <https://www.lamma.toscana.it/clima-e-energia/climatologia/tipi-di-circolazione-eobs> (accessed on 20 November 2025).
21. Zängl, G.; Reinert, D.; Rípodas, P.; Baldauf, M. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 563–579. [\[CrossRef\]](#)
22. Fischer, M.; Knippertz, P.; Proppe, C. Surrogate-based model parameter optimization in simulations of the West African monsoon. *Weather Clim. Dyn.* **2025**, *6*, 113–130. [\[CrossRef\]](#)
23. Cinquegrana, D.; Zollo, A.L.; Montesarchio, M.; Bucchignani, E. A Metamodel-Based Optimization of Physical Parameters of High Resolution NWP ICON-LAM over Southern Italy. *Atmosphere* **2023**, *14*, 788. [\[CrossRef\]](#)
24. Adinolfi, M.; Raffa, M.; Reder, A.; Mercogliano, P. Investigation on potential and limitations of ERA5 Reanalysis downscaled on Italy by a convection-permitting model. *J. Clim.* **2023**, *61*, 4319–4342. [\[CrossRef\]](#)
25. Hannachi, A.; Jolliffe, I.T.; Stephenson, D.B. Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.* **2007**, *27*, 1119–1152. [\[CrossRef\]](#)
26. Hodson, D.L.; Sutton, R.T.; Scaife, A.A. Signal-to-noise and predictable modes of variability in winter seasonal forecasts. *Q. J. R. Meteorol. Soc.* **2023**, *149*, 2598–2616. [\[CrossRef\]](#)
27. Sperber, K.; Brankovic, C.; Deque, M.; Frederiksen, C.; Graham, R.; Kitoh, A.; Kobayashi, C.; Palmer, T.; Puri, K.; Tennant, W.; et al. Dynamical seasonal predictability of the Asian summer monsoon. *Mon. Weather Rev.* **2001**, *129*, 2226–2248. [\[CrossRef\]](#)
28. Molteni, F.; Buizza, R. Validation of the ECMWF Ensemble Prediction System Using Empirical Orthogonal Functions. *Mon. Weather Rev.* **1999**, *127*, 2346–2358. [\[CrossRef\]](#)
29. Lee, J.; Sperber, K.R.; Gleckler, P.J.; Bonfils, C.J.W.; Taylor, K.E. Quantifying the agreement between observed and simulated extratropical modes of interannual variability. *Clim. Dyn.* **2018**, *52*, 4057–4089. [\[CrossRef\]](#)
30. Biggerstaff, M.I.; Seo, E. An EOF-based comparison and evaluation of simulated passive microwave signatures to observations over tropical oceans. *J. Geophys. Res. Atmos.* **2010**, *115*, D15209. [\[CrossRef\]](#)
31. Hortal, M. The development and testing of a new two-time-level semi-Lagrangian scheme (SETTLS) in the ECMWF forecast model. *Q. J. R. Meteorol. Soc. A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr.* **2002**, *128*, 1671–1687.
32. Bottazzi, M.; Scipione, G.; Marras, G.F.; Trotta, G.; D'Antonio, M.; Chiavarini, B.; Caroli, C.; Montanari, M.; Bassini, S.; Gascón, E.; et al. The Italian open data meteorological portal: MISTRAL. *Meteorol. Appl.* **2021**, *28*, e2004. [\[CrossRef\]](#)

33. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
34. Rockel, B.; Will, A.; Hense, A. The regional climate model COSMO-CLM (CCLM). *Meteorol. Z.* **2008**, *17*, 347–348. [[CrossRef](#)]
35. Sirovich, L. Turbulence and the dynamics of coherent structures. *Q. Appl. Math.* **1987**, *XLV*, 561–571. [[CrossRef](#)]
36. Berkooz, G.; Holmes, P.; Lumley, J. The Proper Orthogonal Decomposition in The analysis of turbulent Flows. *Annu. Rev. Fluid Mech.* **1993**, *25*, 539–575. [[CrossRef](#)]
37. Schulzweida, U. CDO User Guide. 2023. Available online: <https://zenodo.org/records/10020800> (accessed on 7 January 2026).
38. Kurucz, M.; Benczúr, A.A.; Csalogány, K. Methods for large scale SVD with missing values. In *Proceedings of the KDD Cup and Workshop*; Citeseer: San José, CA, USA, 2007; Volume 12, pp. 31–38.
39. Lucia, D.; King, P.; Beran, P. Reduced order Modeling of a two dimensional flow with moving shocks. *Comput. Fluids* **2003**, *32*, 917–938. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.