Extracting latent variables from forecast ensembles and advancements in similarity metric utilizing optimal transport

Seiya Nishizawa[1]

[1]RIKEN Center for Computational Science

February 16, 2024

Abstract

This study presents a novel methodology for extracting latent variables from high-dimensional sparse data, particularly emphasizing spatial distributions such as precipitation distribution. This approach utilizes multidimensional scaling with a distance matrix derived from a new similarity metric, the Unbalanced Optimal Transport Score (UOTS). UOTS effectively captures discrepancies in spatial distributions while preserving physical units. This is similar to mean absolute error, however it considers location errors, providing a more robust measure crucial for understanding differences between observations, forecasts, and ensembles. Probability distribution estimation of these latent variables enhances the analytical utility, quantifying ensemble characteristics. The adaptability of the method to spatiotemporal data and its ability to handle errors suggest its potential as a promising tool for diverse research applications.

Extracting latent variables from forecast ensembles and advancements in similarity metric utilizing optimal transport

S. Nishizawa[1]

[1]RIKEN Center for Computational Science

Key Points:

- Novel method reveals hidden information from spatial ensemble data for understanding probability distributions.
- Technique extracts essential similarities and differences in sparse distributions, aiding interpretation for improved analysis.
- Approach is adaptable to different data types, making it promising for diverse scientific fields.

Corresponding author: Seiya Nishizawa, `s-nishizawa@riken.jp`

Abstract

This study presents a novel methodology for extracting latent variables from high-dimensional sparse data, particularly emphasizing spatial distributions such as precipitation distribution. This approach utilizes multidimensional scaling with a distance matrix derived from a new similarity metric, the Unbalanced Optimal Transport Score (UOTS). UOTS effectively captures discrepancies in spatial distributions while preserving physical units. This is similar to mean absolute error, however it considers location errors, providing a more robust measure crucial for understanding differences between observations, forecasts, and ensembles. Probability distribution estimation of these latent variables enhances the analytical utility, quantifying ensemble characteristics. The adaptability of the method to spatiotemporal data and its ability to handle errors suggest its potential as a promising tool for diverse research applications.

Plain Language Summary

This study introduces a new method to understand weather patterns by simplifying complex data. A mathematical technique was developed to efficiently identify hidden information from patterns. This assists meteorologists in understanding the weather with greater accuracy. This method simplifies weather data by highlighting the essential similarities and differences between weather patterns, making it easier for scientists to interpret and use the resultant data effectively. This study offers a new and efficient way to make sense of vast weather data, benefiting meteorological research, and potentially improving weather forecasting. The technique contributes to the meteorological field, in addition it also contributes to various fields with sparse distribution data.

1 Introduction

Probabilistic forecasts play a pivotal role in systems characterized by chaotic or stochastic behavior, such as weather forecasting (Gneiting & Katzfuss, 2014). Ensemble simulations are commonly employed to estimate the probability distributions of future states (Wilks, 2006). However, evaluating the predictive distribution in such multivariate, high-dimensional systems poses challenges, for instance in considering spatially distributed phenomena (Murphy, 1991).

While univariate cases allow straightforward distribution definitions based on ensemble member results, multivariate cases, particularly in high-dimensional systems such as weather forecasting, face the "curse of dimensionality" (Scott, 1992). Representing joint distribution that matches the state vector's dimensionality becomes infeasible owing to this issue, which influences accurate probability estimations.

Current discussions often focus on one-dimensional distributions, considering points individually (e.g., grid points) or single statistical quantities, such as spatial averages (Gneiting & Katzfuss, 2014). However, this point-wise approach could overlook crucial spatial patterns, especially in sparse quantities such as precipitation, leading to an overestimation of discrepancies between states, particularly in high-resolution simulation results (Gilleland et al., 2009).

This study tackles these limitations by leveraging the power of latent variables to capture the underlying structure and reduce complexity within high-dimensional ensemble data. Latent variables are hidden factors that influence observable data and represent underlying structure and relationships in the data (Loehlin, 2004; Lee, 2007). The dimensionality of these latent variables can be significantly smaller than that of real variables (for example, Turk & Pentland, 1991). This is supported by the fact that some systems operate within small, embedded manifolds of lower dimensions, known as ranks (Foias et al., 1988; Constantin, 1989). By extracting these latent variables from ensemble simulations, I aim to create a lower-dimensional space that faithfully represents the essential features of the original data.

This study proposed a new novel methodological approach for extracting meaningful latent variables from high-dimensional ensembles. Dimensionality reduction is a key technique in this method to capture essential features and reduce the high-dimensional ensemble data to more manageable data. The state vector obtained within the low-dimensional space through dimension reduction serves as an estimate of the underlying latent variables. In the consideration of the probability distribution of ensemble data, the distance between the ensemble members is important as it is intrinsic to the spread of the distribution. Therefore, to capture the essential characteristics of the probability distribution of ensemble data by reconstructing the probability distribution in a low-dimensional latent variable space, a reduction method that preserves the distance is appropriate.

Several dimensionality reduction techniques exist. Principal component analysis (PCA) is widely used for dimensionality reduction. However, it has limitations in nonlinear systems (for example, Nishizawa & Yoden, 2004) and is based on point-wise calculation. The variational autoencoder (VAE; Kingma & Welling, 2013) can extract nonlinear relationships, and has received much attention in recent years, however, distance information is lost due to normalization. It also faces challenges in predictive forecast problems owing to limited ensemble sizes for training. In most practical cases, the ensemble size is less than a hundred, and the size is insufficient to obtain latent variables from the ensemble data by the VAE. Moreover, this approach requires previous knowledge of the effective dimensions, which is the minimum number of dimensions of the latent vectors necessary for a sufficiently accurate representation of the underlying physically meaningful structure of the original high-dimensional data. With some other dimension reduction methods, such as locally linear embedding (LLE; Roweis & Saul, 2000), t-distributed stochastic neighbor embedding (t-SEN; Van der Maaten & Hinton, 2008), uniform manifold approximation and projection (UMAP; McInnes et al., 2018), and densMAP (Narayan et al., 2021), the distance is not maintained because low-dimensional variables are reconstructed based on weights or probabilities corresponding to neighboring points.

In this study, multidimensional scaling (MDS; for example, Cox & Cox, 2000), specifically classical MDS, a.k.a., principal coordinate analysis, is utilized to construct a Euclidean low-dimensional space, where the distances between samples correspond to the distance in the original high-dimensional state space. This method preserves the distance in a dimension-reduced latent variable space and does not require prior knowledge of the effective dimension of the system. MDS operates as a linear procedure. A nonlinear dimension reduction technique, Isomap (Tenenbaum et al., 2000), was also examined. Isomap extends MDS by capturing nonlinear manifolds embedded within the original space. By employing geodesic distance with a neighborhood graph, Isomap can be applied to complex data structures beyond linear representations. The influence of the linear limitation of MDS on the extracted state vectors was examined by employing Isomap.

The dimension reduction with MDS and Isomap is performed using a similarity metric for all pairs of input samples. The units and magnitude of the similarity metric are retained as a distance in the low-dimensional space. Therefore, the validity of the extracted latent variables significantly depends on the definition of the similarity metric used, and the choice of an appropriate similarity metric is a critical aspect of this method.

To measure the similarity of two different states, several metrics exist. Existing metrics often fall short of capturing overall differences, leading to potentially misleading interpretations. These metrics include traditional metrics, such as the mean absolute error (MAE), root mean squared error (RMSE), and Pearson correlation coefficient (CORR). In addition, these include scores considering event-based dichotomous variables, such as the frequency bias (FB, also called as bias ratio; for example, Wilks, 2006), equitable threat score (ETS; Gilbert, 1884), and fractions skill score (FSS; N. M. Roberts & Lean, 2008;

N. Roberts, 2008). These were calculated point-wise, with the exception of FB and FSS. The point-wise comparison is known to double-penalize small-scale discrepancies (Gilleland et al., 2009). Among them, FSS is a score that allows some spatial displacement and is widely used for high-resolution simulations. However, as it is based on a categorized or thresholded quantity, it does not consider amplitude differences. When considering scores with different thresholds simultaneously to determine the event, the amplitude difference may be implicitly interpreted. In cases with a large number of samples, the interpretation of multiple scores may require complex and difficult considerations. Recently, structure, amplitude and location (SAL; Wernli et al., 2008) and its extension for ensemble forecast (eSAL; Radanovics et al., 2018) have been used to evaluate validity of forecasts. However, they are not a single score but a combination of three independent scores corresponding to structure, amplitude and location errors. A comprehensive single score is preferred for several purposes, for example, estimating or characterizing a probability distribution from ensemble. Therefore, FSS, SAL and eSAL are not suitable for the purposes like probability distribution estimation. The displacement and amplitude score (DAS; Keil & Craig, 2009) is a single score of combination of displacement and amplitude differences, containing more information than the traditional scores. However, there are several arbitrary in definitions and computational procedures. Keil and Craig (2007) showed that $D_{\max}$, which is the maximum search distance, has a great decisive impact on the result. It can only take discontinuous values: $D_{\max}$ is proportional to grid spacing times a power of two. Therefore, it may be difficult to choose an appropriate value based on physical considerations owing to its discontinuous constraint. They suggested that other parameters had a minor impact, however, non-negligible arbitrariness which they did not discuss exists. The score was defined such that the amplitude difference between one distribution and the morphed distribution of the other becomes the lowest; however, no condition was provided for the morphing flow, and in general, many possible flows can achieve the smallest amplitude difference. Thus, there are many possibilities for displacement, and the total score depends on the displacement. Another arbitrary factor is the difference in weight between the displacement and amplitude. This score is a combination of these two differences. As they have different units, the differences are normalized or nondimensionalized. On the original definition of DAS, the normalization factors are determined such that the two terms have equivalent weights. However, there are other possibilities to choose the weights than the equivalent weights. In this sense, the weight parameter is inherently arbitrary. Indeed, there are arbitrary in the definition of the normalization factor of the amplitude error term I_0. In addition, there is considerable arbitrariness in its computational procedure, resulting in a variation in the score. In fact, this study's implementation of computing the DAS results in a non-negligible difference in the obtained score compared to Keil and Craig (2009) for the same distributions owing to the undocumented details in the procedure. Another critical issue is that the procedure does not consider mass conservation during morphing.

To address these issues in the existing metrics, this study introduces the Unbalanced Optimal Transport Score (UOTS) as a novel similarity metric specifically designed for evaluating spatial distribution discrepancies. UOTS considers both amplitude and location differences in a unified manner, as does the DAS. However, the two terms of the displacement and amplitude differences have the same units and can be compared directly. Therefore, nondimensionalization is not needed to combine them into a single score. UOTS is a more straightforward score that considers both displacement and amplitude differences than DAS. UOTS also has the same units as the original quantity, which facilitates physical interpretations. UOTS offers significant advantages over existing metrics by minimizing arbitrariness in its mathematical definition and providing clearer physical interpretations, particularly regarding its hyperparameters.

The effectiveness of this approach and the suitability of UOTS in extracting meaningful latent vectors are demonstrated through experiments with synthetic and real-world meteorological data. This method is expected to provide valuable insight into high-dimensional

Table 1. Metrics for similarity used in this study. The rightmost column shows the conversion equation from the metric to the corresponding distance.

Abbreviation	Name	Distance
UOTS	Unbalanced optimal transport score	$UOTS$
DAS	Displacement and amplitude score	DAS
FSS	Fractions skill score	$1 - FSS$
MAE	Mean absolute error	MAE
RMSE	Root mean squared error	$RMSE$
CORR	Pearson correlation coefficient	$\sqrt{2(1 - CORR)}$
ETS	Equitable threat score	$1 - ETS$
FB	Frequency bias	$\lvert \log(FB) \rvert$

ensemble data, leading to improved probability distribution estimation and ultimately, more accurate and informative forecasts.

2 Methods

2.1 Extracting Latent Variables

In this subsection, the proposed approach to extract latent variables from high-dimensional ensembles is described. The methodological approach was divided into two steps.

1. Calculation of a similarity metric for all pairs of ensemble members and observations.
2. Extraction of latent variables in a low-dimensional space from the distance matrix based on the similarities.

2.1.1 Metric for Similarity

Assessing the similarity between spatial distributions requires a robust metric that capture various discrepancies, including amplitude, location, area, and shape differences. In this study, various metrics were employed to measure the similarity between spatial distributions. Table 1 summarizes the metrics used in this study.

2.1.1.1 Unbalanced Optimal Transport Score The UOTS proposed in this study serves as a novel similarity metric tailored to assess spatial distribution discrepancies. The UOTS is defined as follows:

$$UOTS = \frac{1}{N} \min_{\gamma \in \mathbb{R}^{N^2}_{\geq 0}} \left[\left\{ 2 \sum_{i_1, i_2} \gamma_{i_1 i_2} \left(\frac{\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|_2}{L} \right)^q \right\} + \|\gamma \mathbf{1} - \phi_1\|_1 + \left\|\gamma^T \mathbf{1} - \phi_2\right\|_1 \right], \quad (1)$$

where $\mathbf{x}_i$ represents the location of the point i, $\phi_1(\mathbf{x}_i)$ and $\phi_2(\mathbf{x}_i)$ are mass distribution in the two distributions which are to be compared. γ is the transport matrix, which is a $N \times N$ matrix whose element $\gamma_{i_1 i_2}$ is a non-negative real number representing the mass transported from $\mathbf{x}_{i_1}$ to $\mathbf{x}_{i_2}$. $\mathbf{1}$ is a vector whose elements are all unity, and $\|\bullet\|_p$ represents the L^p norm. The superscript T represents transposition. N is the vector length, i.e., $i = 1, \cdots, N$. In this study, the score is defined as divided by N, however the number of nonzero elements can be used instead of N, depending on the purpose.

The UOTS is defined based on optimal transport (OT), which is a mathematical problem introduced by Monge (1781). OT is an optimization problem of determining mass transport plan to minimize the overall cost of moving one mass distribution onto another one with respected to given costs of moving a unit of mass between all pairs of spatial points. It has been widely used in various fields, especially in the machine learning field, as a measure of similarity of non-negative distribution, such as probability distributions. For OT, two distributions to be compared must have the identical mass; probability density distributions have unit mass, however, spatial distributions, such as precipitation distribution, can have various mass. Therefore, the OT cannot be used to measure similarity for these distributions.

Unbalanced OT (UOT) is an extension of the OT to enable to apply to distributions with different total mass. UOT has some variants in the form representing the mass difference. The most popular form is using Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) (Frogner et al., 2015). Another form is using L^1 norm, which is called the partial optimal transport (Caffarelli & McCann, 2010; Chizat et al., 2018; Figalli, 2010), flat metric (Peyré & Cuturi, 2019), or Kantrovidge–Rubinshutain distance (Hanin, 1992; Lellmann et al., 2014). UOTS is based on the UOT with L^1 form, normalized by the q-square of the length scale L. Thanks to the normalization, UOTS has the units of mass (or the units of the original quantity). In OT, the total mass is unity, or its value is usually normalized by the total mass, and thus the cost value has units of distance. Therefore, the optimal value in OT is often referred to as Wasserstein “distance”.

The OT and UOT have advantages over conventional metrics, such as point-wise norms and relative entropy, such as KL divergence, (Séjourné et al., 2023; De Plaen et al., 2023). One advantage is their ability to capture global structure, considering the overall distribution and global relationship. They are sensitive to geometry and shapes, which an important feature as similarity metric. Another advantage is robustness to noise and outliers, as since they have information across the entire distribution, the impacts of individual anomalies are reduced. Therefore, OT and especially UOT are less affected by noise and outliers, which are often contained in practical dataset.

UOTS inherits the advantages of OT and UOT. The UOTS captures both amplitude and location differences and is robust to noise and outliers. The UOTS employs the L^1 norm to express the mass difference as in the partial optimal transport, and can be interpreted as the mean absolute error when spatial displacements are considered. The first term in the brackets on the right-hand side of Eq. 1 penalizes mass transport or displacement of the distribution. $\gamma\mathbf{1}$ and $\gamma^T\mathbf{1}$ denote the transported source and target distributions, respectively. Therefore, the second and third terms represent the mean absolute error after the transport. Through mass transport, the absolute error can be decreased. On the other hand, larger transport costs more. The UOTS is to be determined to minimize the sum of the transport cost and the resulting absolute error, therefore, UOTS can be considered the mean absolute error with location error correction.

Its formulation involves the hyperparameters L and q. The parameter L determines the distance threshold for identifying same phenomena. Patterns exceeding this threshold are considered different. For the i_1 and i_2 index pairs, where $\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|_2 > L$, the optimized value of $\gamma_{i_1 i_2}$ must be zero; otherwise, the first term representing the transport cost outweighs the second and third terms representing the amplitude difference. To understand this, a simplest case of two points x_1 and x_2 can be considered, in which $\phi_1(x_1) > 0, \phi_2(x_2) > 0$, and $\phi_1(x_2) = \phi_2(x_1) = 0$. Then, it is evident that $\gamma_{12} \leq \min(\phi_1(x_1), \phi_2(x_2))$ and $\gamma_{11} = \gamma_{22} = \gamma_{21} = 0$. UOTS can be written as:

$$
\begin{aligned}
UOTS &= \frac{1}{2} \min_{\gamma_{12} \in \mathbb{R}_{\geq 0}} \left\{ 2\gamma_{12} \left(\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2}{L} \right)^q + (\gamma_{12} - \phi_1(x_1)) + (\gamma_{12} - \phi_2(x_2)) \right\} \\
&= \frac{1}{2} \min_{\gamma_{12} \in \mathbb{R}_{\geq 0}} \left[2\gamma_{12} \left\{ \left(\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2}{L} \right)^q - 1 \right\} + \phi_1(x_1) + \phi_2(x_2) \right]. \qquad (2)
\end{aligned}
$$

Therefore, $\gamma_{12} = 0$ since $\left(\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2}{L} \right)^q - 1 > 0$.

Next, the range of possible values for UOTS is considered. In general, the lower bound is zero but there is no upper bound. It is instructive to consider the limit values for L, as the limit values give a range of the possible values of UOTS. As L decreases, transportation costs increase and the UOTS approaches MAE:

$$\lim_{L \to 0} UOTS = \frac{1}{N} \left(\|\phi_{\text{int}} - \phi_1\|_1 + \|\phi_{\text{int}} - \phi_2\|_1 \right) = \frac{1}{N} \|\phi_1 - \phi_2\|_1 , \tag{3}$$

where ϕ_{int} is the intersection of ϕ_1 and ϕ_2. This is because $\gamma_{i_1 i_2}$ must be zero for $i_1 \neq i_2$, for which $\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|_2 > 0$, otherwise the term representing transport cost becomes infinity. On the other hand, as L increases, transportation costs decrease and the UOTS approaches the mean mass difference, or bias:

$$\lim_{L \to \infty} UOTS = \frac{1}{N} \left(\|\gamma \mathbf{1} - \phi_1\|_1 + \left\|\gamma^T \mathbf{1} - \phi_2\right\|_1 \right) = \frac{1}{N} \left| \|\phi_1\|_1 - \|\phi_2\|_1 \right| , \tag{4}$$

because $\gamma^T \mathbf{1} = \phi_2$ and $\|\gamma \mathbf{1}\|_1 = \|\phi_2\|_1$, when $\|\phi_1\|_1 \geq \|\phi_2\|_1$; and vise versa. For intermediate L, the UOTS have the value between the two limits (MAE and the mean mass difference) as $\frac{1}{N} \left| \|\phi_1\|_1 - \|\phi_2\|_1 \right| \leq UOTS \leq \frac{1}{N} \|\phi_1 - \phi_2\|_1$.

The other hyperparameter q affects the transportation cost per mass. The larger the value of q, the more the difference in position is disregarded and the more tolerant the score is for small displacement errors. This is because $\frac{\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|_2}{L} \leq 1$ where $\gamma_{i_1 i_2} > 0$.

In the actual computation in this study, the minimization problem for this optimization was solved by using the Sinkhorm algorithm (Cuturi, 2013) with a reservoir of dustbin points by incorporating a regularization term $\lambda \Omega(\gamma)$ as in ordinary partial optimal transport. Here, Ω and λ represent the entropy regularization function and its coefficient, respectively, and $\Omega(\gamma) = \sum_{i_1, i_2} \gamma_{i_1 i_2} \log(\gamma_{i_1 i_2})$. In this study, the parameter λ was fine-tuned to the smallest possible value without causing computational divergence.

The UOTS introduces a novel approach that comprehensively evaluates the similarity between spatial distribution patterns, while having a clear physical interpretation of its hyperparameters.

2.1.2 Extraction of Latent Variables in Dimension-Reduced Space

Before extracting the latent variables in a reduced space, a distance matrix was constructed from the similarity metric between all pairs of the ensemble members and the observational data. In the process of constructing the distance matrix, it is crucial to transform metrics into values resembling distances that signify zero for identical distributions, nonnegatives, and symmetry, as detailed in Table 1.

2.1.2.1 Multidimensional Scaling MDS allows for the extraction of state vectors in Euclidean space while preserving the given distance. It also reveals the relative importance of each coordinate and the number of effective dimensions based on stress functions. The state vector in the space can be obtained by solving an eigendecomposition problem of the matrix $K = -\frac{1}{2} HDH$, where D is the distance matrix and H is the centering matrix $H = I - \frac{1}{N} \mathbf{1}\mathbf{1}^T$, and I is the identity matrix. The vector $\mathbf{v}_m$ corresponding the state of the m-th ensemble member is obtained such that its k-th element is $v_{km} = \sqrt{\lambda_k} u_{mk}$, where λ_k and u_{mk} are the k-th eigenvalue and the m-th element of the k-th eigenvector, respectively. Coordinates corresponding to larger eigenvalues are more principal. By considering only a small number of principal eigenvalues/eigenvectors (i.e., principal coordinates), a vector of smaller dimension can be obtained. It is noted that the result of the MDS is identical to PCA when the Euclidean distance, i.e., RMSE, in the original high-dimensional space is used as the similarity metric (Cox & Cox, 2000).

The stress function S is computed as follows:

$$S = \sqrt{\frac{\sum_{m_1<m_2}(\hat{d}_{m_1m_2} - d_{m_1m_2})^2}{\sum_{m_1<m_2}\hat{d}_{m_1m_2}}}, \tag{5}$$

where $\hat{d}_{m_1m_2}$ represents the distance in the reduced space ($\hat{d}_{m_1m_2} = \|\mathbf{v}_{m_1} - \mathbf{v}_{m_2}\|_2$), and $d_{m_1m_2}$ is the distance derived from the similarity metric in the original space between members m_1 and m_2. The stress function changes value depending on how many dimensions of the $\mathbf{v}_m$ are considered when calculating the distance $\hat{d}$. In other words, it is a function of the number of the dimensions δ.

2.2 Experiments

2.2.1 Synthetic Data Experiment

The synthetic data experiment was designed following the methodology detailed in Ahijevych et al. (2009) to illustrate the characteristics of various similarity metrics for assessing spatial distributions. A prescribed geometric spatial distribution mimicking the accumulated surface precipitation distribution was utilized. This distribution is described as follows:

$$\phi(x,y) = \begin{cases} 0, & \left(\frac{x-x_1}{a}\right)^2 + \left(\frac{y-y_1}{b}\right)^2 \geq 1 \\ \Phi_1, & \left(\frac{x-x_1}{a}\right)^2 + \left(\frac{y-y_1}{b}\right)^2 < 1, \quad \left(\frac{x-x_2}{0.4a}\right)^2 + \left(\frac{y-y_1}{0.4b}\right)^2 \geq 1 \\ \Phi_2, & \left(\frac{x-x_2}{0.4a}\right)^2 + \left(\frac{y-y_1}{0.4b}\right)^2 < 1 \end{cases}, \tag{6}$$

where $x_2 = x_1 + 0.4a, \Phi_2 = 2\Phi_1$, and $x = i\Delta x$ and $y = j\Delta x$, with $i = 0, 1, \cdots, 601, j = 0, 1, \cdots, 501$ and $\Delta x = 4$ km.

Six spatial distributions (Fig. 1) were created, including one reference (observation) and five target patterns (forecasts). The parameters (x_1, a, b) for the reference, pattern 1, pattern 2, pattern 3, pattern 4, and pattern 5 are $(200\Delta x, 25\Delta x, 100\Delta x)$, $(250\Delta x, 25\Delta x, 100\Delta x)$, $(400\Delta x, 25\Delta x, 100\Delta x)$, $(325\Delta x, 100\Delta x, 100\Delta x)$, $(325\Delta x, 100\Delta x, 25\Delta x)$, and $(325\Delta x, 200\Delta x, 100\Delta x)$, respectively. In all distributions, $y_1 = 250\Delta x$ and $\Phi_1 = 12.7$ mm.

2.2.2 Synthetic Data Ensemble Experiment

Furthermore, in this study, the geometric distribution (Section 2.2.1) was extended to ensemble forecasts and multiple cases. The observations and ensemble members were generated using specific parameters to simulate diverse scenarios, resulting in 100 cases with 50 ensemble members each.

The parameters for the observations are $(x_1^{\text{obs}}, y_1^{\text{obs}}, a^{\text{obs}}, b^{\text{obs}}, \Phi_1^{\text{obs}}) = (300\Delta x, 250\Delta x, \sqrt{\frac{A}{\pi\alpha}}, \sqrt{\frac{A\alpha}{\pi}}, 2^{\epsilon_3/2})$, where A and α are the area and aspect ratios, respectively, and $(A, \alpha) = (2^{\epsilon_1/2}\pi a_0 b_0, 4^{\epsilon_2/2}\frac{b_0}{a_0})$. The constants were set as $a_0 = 25\Delta x$ and $b_0 = 100\Delta x$ based on the reference in the synthetic data experiment. ϵs are random numbers with a standard normal distribution.

The parameters for ensemble members are $(x_1^{\text{fcs}}, y_1^{\text{fcs}}, a^{\text{fcs}}, b^{\text{fcs}}, \Phi_1^{\text{fsc}}) = (x_1^{\text{obs}} + 50\Delta x\epsilon_4, y_1^{\text{obs}} + 50\Delta x\epsilon_5, \sqrt{\frac{A^{\text{fcs}}}{\pi\alpha^{\text{fcs}}}}, \sqrt{\frac{A^{\text{fcs}}\alpha^{\text{fcs}}}{\pi}}, 2^{\epsilon_8/2}\Phi_1^{\text{obs}})$, where, $(A^{\text{fcs}}, \alpha^{\text{fcs}}) = (2^{\epsilon_6/2}A^{\text{obs}}, 4^{\epsilon_7/2}\alpha^{\text{obs}})$. Note that, for the location difference, i.e., $x_1^{\text{fcs}} - x_1^{\text{obs}}$ and $y_1^{\text{fcs}} - y_1^{\text{obs}}$, 95% of samples (corresponding to 2 standard deviations) statistically range from $-100\Delta x$ to $100\Delta x$. For the area, aspect ratio, and amplitude, the factors of $2^{\epsilon/2}$ and $4^{\epsilon/2}$ were set so that 95% of samples statistically ranged from 0.5 to 2, and 0.25 to 4, respectively.

2.2.3 Real Data Experiment

In order to examine applicability of the method to practical applications, real precipitation data was used. The data had a spatial distribution of radar reflectivity at 2-

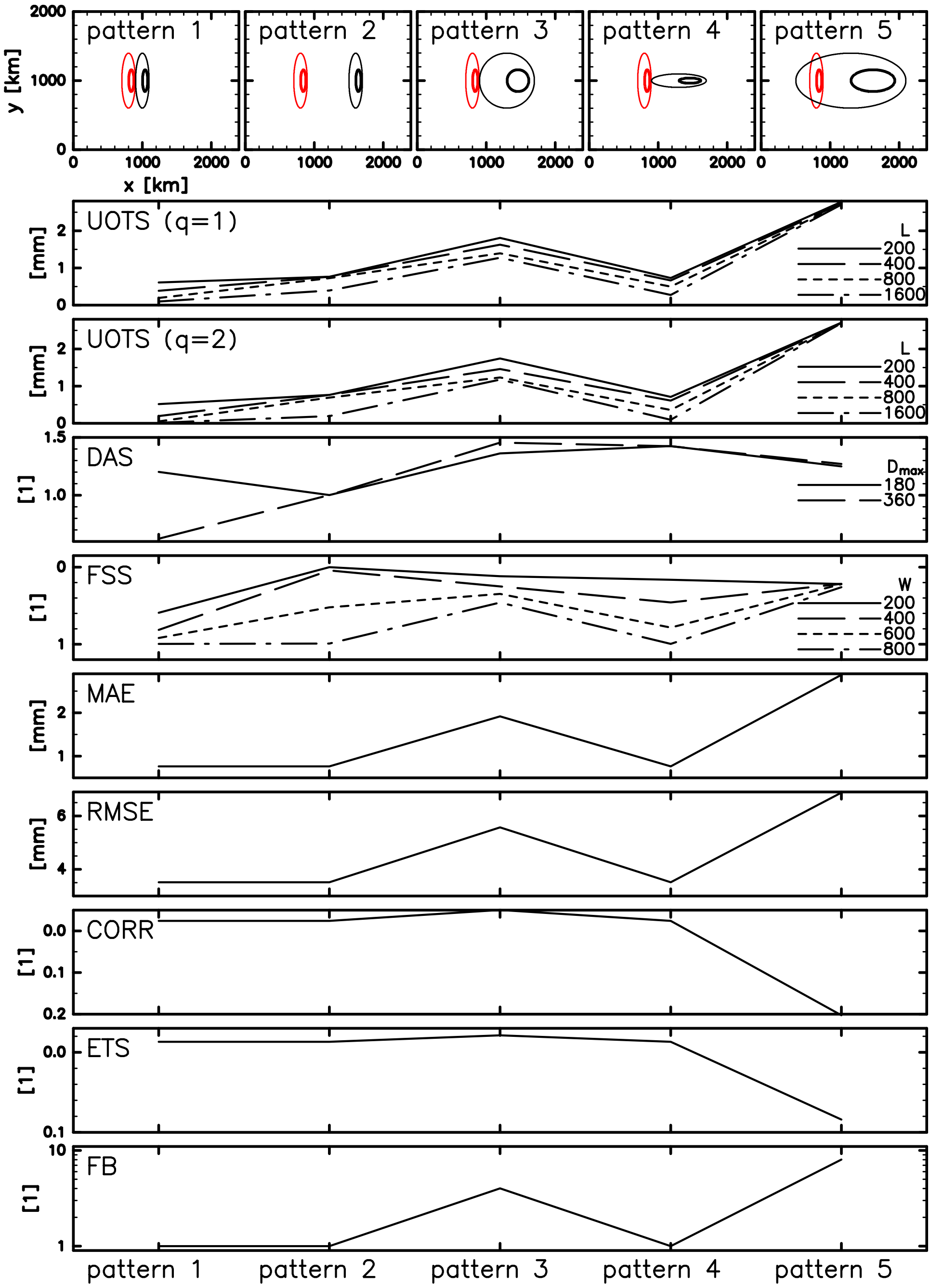


Figure 1. Spatial pattern of the geometric distributions in the synthetic data experiment. The top panels display the reference (observation) and five distributions (forecasts). The red and black color indicates observation and forecasts, respectively. The thin and thick contours represent the area at which $\phi = 12.7$ and 25.4 mm, respectively. The lower panels show the magnitude of similarity metrics for the five distributions with respect to the reference. The vertical coordinates are oriented such that the bottom (top) is more similar (more different). Solid, dashed, dotted, and dash-dotted lines indicate $L = 200, 400, 800,$ and 1,600 km for UOTS, $W = 200, 400, 600,$ and 800 km for FSS, respectively. Solid and dashed lines indicate $D_{\max} = 180$, and 360 km for DAS, respectively.

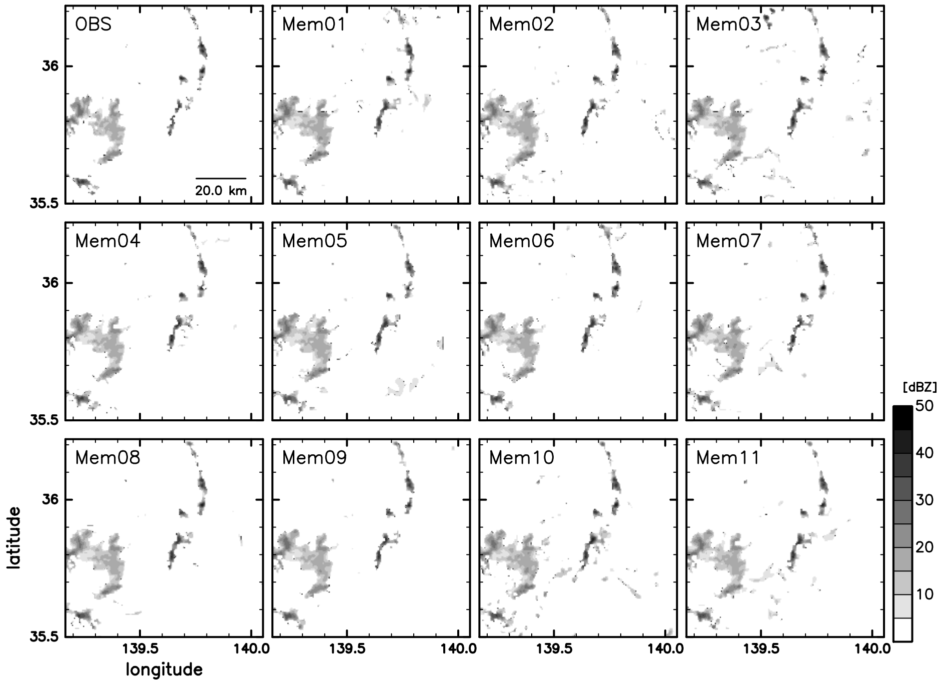


Figure 2. Longitude-latitude cross-section of the radar reflectivity in dBZ in the real data experiment. The top-left panel display the observation. The other panels are of 11 ensemble members out of 50 members.

km height at 18:00 UTC, 29 July 2021 obtained in Miyoshi et al. (2023). A real-time 30-second-refresh numerical weather prediction was conducted at Kanto region in Japan during Tokyo Olympics and Paralympics in 2021 using supercomputer Fugaku. In the prediction system, a regional atmospheric model SCALE-RM (Scalable Computing for Advanced Library and Environment-Regional Model, Nishizawa et al., 2015; Y. Sato et al., 2015) and a data assimilation framework SCALE-LETKF (SCALE-local ensemble transform Kalman filter, Lien et al., 2017) was utilized and generated analysis data every 30 seconds with 1,000 ensemble members of 500-m-mesh simulations assimilating 3-D volume radar observations obtained by the phased array weather radar installed at Saitama University. In this study, data of 50 members out of the 1,000 members was used. The data covers about 80×80 km^2 domain and its spatial resolution is 500 m (161×161 grids). The smaller values less than 5 dBZ were rounded to zero.

Figure 2 shows the horizontal distribution of the radar reflectivity of the observation and some of the ensemble members. The ensemble members have similar patterns to the observation with some differences.

3 Results

3.1 Synthetic Data Experiment

The characteristics of the various similarity metrics were examined using geometric spatial distributions (Section 2.2.1). The experiment involved multiple metrics and the sweeping of their hyperparameters. L and q for UOTS were swept: $L = 200$, 400, 800, and 1,600 km, and $q = 1$ and 2. The FSS also had a hyperparameter W which represents the width of neighborhoods, and it was swept for 200, 400, 600, and 800 km. The parameters for DAS were set to $D_{\max} = 180$ and 360 km, and I_0 were set 15.4 mm according to previous research (Keil & Craig, 2009).

Figure 1 visually demonstrates the magnitude of various similarity metrics applied to the five target patterns (forecasts) with respect to the reference (observation). The difference of the forecast of the pattern 1 from the observation is obviously smaller than that of other patterns. However, the score for the pattern 1 is not the best with the DAS with $D_{\max} = 180$, CORR, and ETS. UOTS displayed consistent rankings across patterns, indicating lower sensitivity against parameter changes. Conversely, DAS and FSS exhibited higher sensitivity to their parameters, signifying the necessity for careful parameter selection. Although the parameters L, W, and $D_{\max}$ for the UOTS, FSS, and DAS, respectively, all indicate the limit distance for location difference, dependency of the scores on the parameter exhibits such significant difference. The lower sensitivity, or higher robustness, is a favorable trait to determine single value representing similarity. Traditional scores, such as RMSE, MAE, CORR, ETS, and FB, as previously reported by Ahijevych et al. (2009), showed limitations in distinguishing between patterns 1, 2, and 4, i.e., different location and aspect-ratio errors. These outcomes emphasize the advantages of UOTS as a more robust similarity metric.

3.2 Synthetic Data Ensemble Experiment

To demonstrate the extraction of latent variables and advantages of the UOTS, a synthetic data ensemble experiment was conducted (Section 2.2.2). Figure 3 presents the distributions of the estimated latent variables in two-dimensional space with the two leading coordinates. With the independent parameters given in the distribution generation, the two coordinates are anticipated to be independent if the latent variables are successfully extracted. When utilizing UOTS, DAS, and FSS with a moderate W, the first and second coordinates appear to be independent. Conversely, in cases employing RMSE, MAE, ETS, CORR, FB, and FSS with small and large W, these two coordinates exhibit a relationship. The FB and FSS with a large W are nearly one-dimensional, relying solely on the first coordinate. FB and FSS with large W depend solely on the area difference and disregard other errors, leading to a one-dimensional latent variable distribution. The dependency on the parameter with UOTS is much lower than with FSS, which is consistent with the result in the previous experiment (Section 3.1). Despite the lower dependency, it is relatively larger with $q = 2$ than with $q = 1$. UOTS with $L = 1,600$ and $q = 2$ shows a relatively one-dimensional structure, since UOTS depend solely on the area difference in the case of $L = \infty$. ETS, CORR, and FSS with small W were distributed in a two-dimensional space however exhibited a rather one-dimensional structure. Reasons for this may be considered as follows: The metrics reach the upper bound value even with a low location error, and many pairs of the ensemble members tend to have the same value, i.e., the upper bound value. In fact, in the synthetic data experiment, these metrics have the value near the upper bound for most of the patterns (Fig. 1). In the two-dimensional latent variable space, they try to locate to have an equal distance, resulting the circular structure. The distributions using MAE and RMSE display intermediate characteristics between the two-dimensional independent structure (e.g., with UOTS) and one-dimensional structure (e.g., ETS). Furthermore, as the ensemble members were generated by adding or multiplying a normal random number to the observation parameters, the observation state was expected to be located near the origin in the latent variable space. With UOTS, FSS with a medium W, and FB, the observation was located near the origin, as expected. However, the observations are not positioned near the origin for the other cases. From this perspective, UOTS and FSS with medium W emerged as favorable similarity metrics among those investigated. Although Fig. 3 represents distributions in a single case, their qualitative characteristics described above are consistent across all the cases.

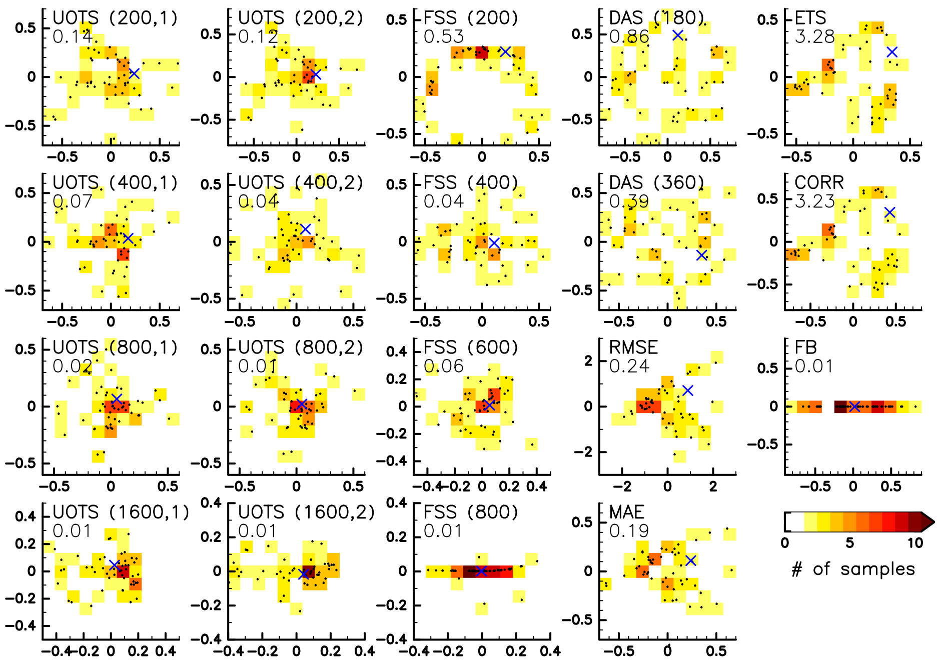


Figure 3. (Scatter plot) Locations of the individual ensemble member and observation in the leading two-dimensional latent variable space and (tone) their two-dimensional histograms. The black dots and blue x symbol indicate the ensemble member and observation, respectively. The numbers in parentheses represent L and q for UOTS, W for FSS, and $D_{\max}$ for DAS. The number under the metric name is the mean distance of the observation from the origin in the two-dimensional space normalized by the standard deviation of the distance of ensemble members from the origin averaged over all the 100 cases.

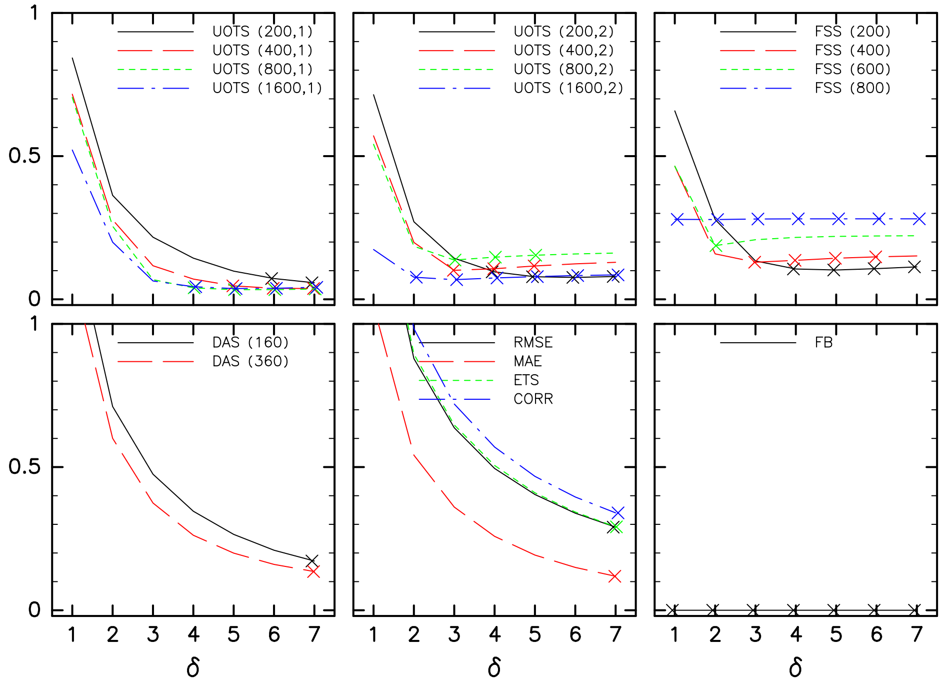


Figure 4. Dependency of the stress on the number of dimension δ. Different color and line types indicate different metrics and parameters. The number in parentheses represent L and q for UOTS, W for FSS, and $D_{\max}$ for DAS. The x symbols indicate the dimension at which the stress is less than the minimum value plus 0.02.

In this experiment, the latent variable was expected to be five dimensions, since the distributions were generated with five independent parameters: the amplitude, two-dimensional location, area, and aspect ratio. The effective dimensionality of the estimated latent variables is explored using the stress function (Fig. 4). The effective dimensionality is estimated as the dimension δ at which the stress becomes constant, i.e., minimum value: For example, when the effective dimension is five, the stress will decrease for $\delta \leq 5$, and remain constant for $\delta \geq 5$. UOTS with $L = 400$ and $q = 1$ exhibited an effective dimensionality of five, aligned with the expectations. However, the effective dimensionality depends on L: it is tend to be smaller as L becomes large. This tendency can be seen for both $q = 1$ and 2. This indicates that some information was being discarded for larger L, since UOTS solely depends on the area in the limit of $L \to \infty$. This information loss is remarkable for the FSS with $W = 800$ and FB, and the stress was almost constant for all δ, corresponding to a one-dimensional structure. On the other hand, some metrics displayed a continuous stress reduction even beyond five dimensions, suggesting an overestimation of dimensionality: UOTS with $L = 200$ and $q = 1$, DAS, RMSE, MAE, ETS, and CORR. With FSS of $W = 200$ and 400, it becomes constant at $\delta = 4$ and 3, respectively. The fact that FSS does not consider the amplitude error is related to this underestimation of the dimensionality. Conversely, the stress increases as δ increases with UOTS with $L = 400, 600, 800$ and $q = 2$, and FSS with $W = 600$. This implies that these metrics are not appropriate for representing the Euclidean distance. These results suggests that a moderate L and $q = 1$ are suitable to obtain dimensionality reasonably.

To investigate the relationship between the five given parameters (x, y, A, α and Φ) and the extracted coordinates, a correlation analysis was performed. In this experiment, if the latent variable is correctly extracted, information of all five parameters should be contained in the five leading coordinates. In each case, the correlation coefficients between the parameters and the elements of the extracted vector corresponding to the five leading coordinates were computed. As the order of the leading coordinates can vary depending on the case, the highest correlation coefficient among the five leading coordinates was selected for each parameter in each case. Figure 5 displays the averaged correlation coefficient in all cases with error bars indicating the 99% confidence level for each parameter. The correlation coefficients for all five parameters were mostly higher than 0.36 for UOTS with larger Ls. The value of 0.36 is the threshold of the correlation coefficient calculated with 50 samples to be statistically significant at the 99% significance level. For other metrics, some of the correlation coefficients are lower than the threshold, indicating that the extracted latent variables do not have information of some given parameters, that is, the metrics lost some information. This confirms that UOTS successfully extracted information of the five parameters.

Overall, UOTS with L =400–800 and $q = 1$ emerged as the most preferred similarity metrics among those investigated, providing insights into the latent variable distribution.

Furthermore, the linearity constraints inherited in MDS were considered. The distributions in two-dimensional space displays a one-dimensional structure with ETS, CORR, and FSS with $W = 200$. This can be attributed to the limitations of linearity inherent to MDS. To address this limitation, the nonlinear method Isomap was employed. However, the distributions in two-dimensional space obtained with Isomap are similar to those obtained using conventional MDS with ETS and CORR. For FSS with $W = 200$, although the shape changed significantly, it still exhibited a one-dimensional structure. This implies that the dimensionality constraint is inherent to the characteristics of the similarity metric. On the other hand, the obtained distributions with UOTS using Isomap are almost identical to that using MDS. Therefore, when UOTS is used as similarity metric, MDS can be used to extract latent variables.

3.3 Real Data Experiment

The synthetic data ensemble experiment in Section 3.2 considers distributions with a single nonzero area, i.e., single phenomena, and practical scenarios involving multiple nonzero areas may require further consideration of the effectiveness of UOTS and appropriate L values. Therefore, in addition to the synthetic data, real application data (Section 2.2.3) was used to examine the applicability of this method. In the real data experiment, we do not know information about the true latent variables. Therefore, in this experiment, consistency of characteristics and dependency on the metric and parameters with those in the synthetic data ensemble experiment was considered.

As in the synthetic data ensemble experiment, the latent variables were extracted by the MDS with variety of similarity metrics. Figure 6 shows the spatial distributions of the extracted latent variables in the leading two-dimensional space. The characteristics are similar to those in the synthetic data ensemble experiment. As shown in the synthetic data ensemble experiment, it is almost one-dimensional with FB and FSS with large W. Sensitivity of the distribution on the parameter is much smaller with UOTS with $q = 1$ than with UOTS with $q = 2$ and FSS. As L becomes larger, the distribution becomes nearly one-dimensional and this is more remarkable for $q = 2$. On the other hand, the one-dimensional structure seen with ETS, CORR, and FSS with small W in the synthetic data ensemble experiment is not clearly seen in the real data experiment. This may be because small-scale noises existing in the original data act as an spatial scatter or smoothing filter in the latent variable space.

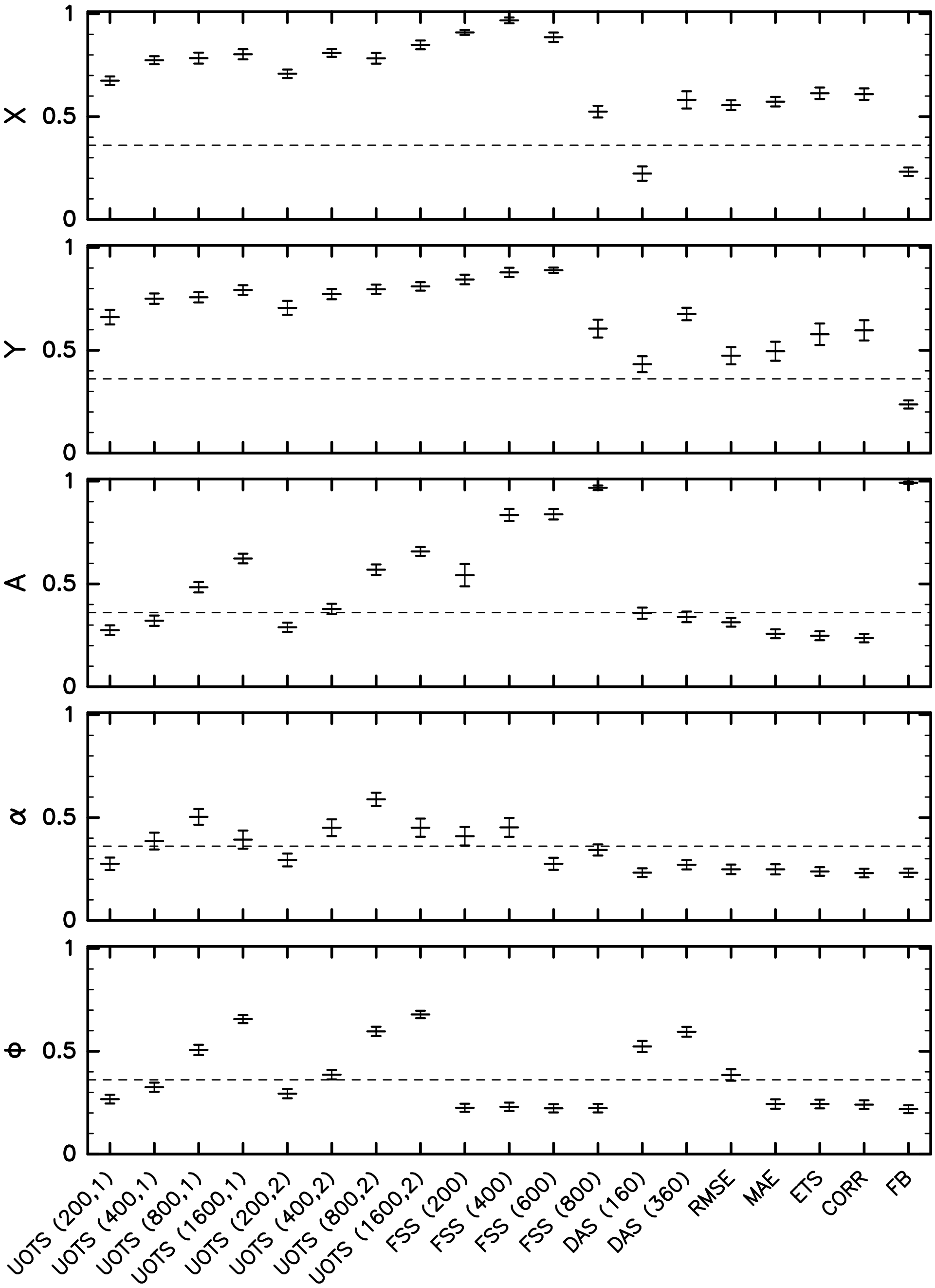


Figure 5. Correlation coefficient between the extracted latent variables and the prescribed five parameters, x, y, A, α and Φ, for each metrics averaged over the 100 cases in the synthetic data ensemble experiment. The error bar represents the 99% confidence interval. The dotted line shows the level above which the correlation calculated from 50 samples is statistically significant at 99% significance level.

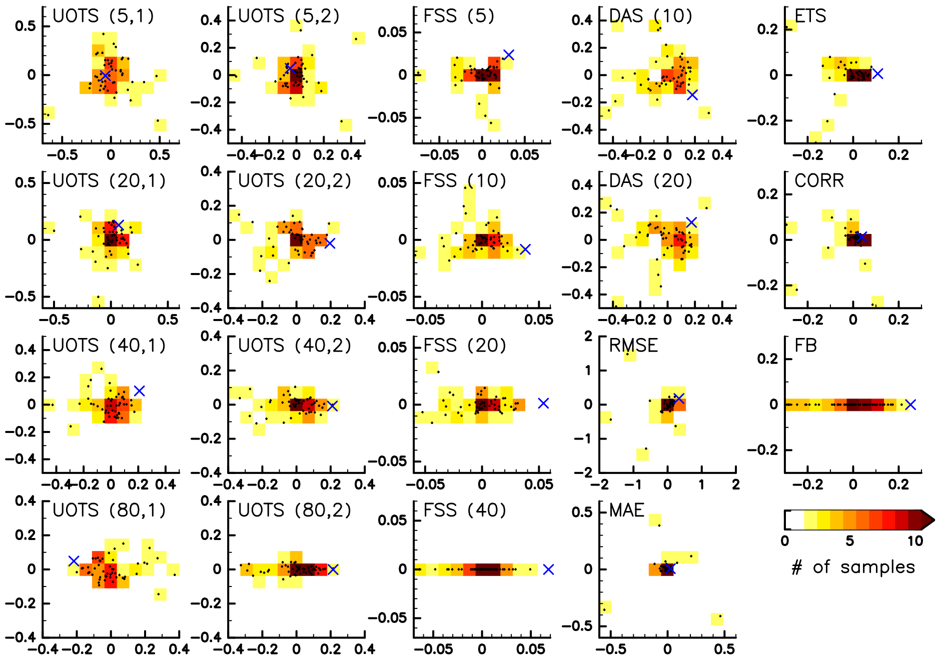


Figure 6. The same figure as Fig. 3 but in the real data experiment.

The dependency of the stress on the dimensionality also shows similar characteristics to the synthetic data ensemble experiment (Fig. 7). The sensitivity of the stress on the parameter is much smaller with UOTS with $q = 1$ than with UOTS with $q = 2$ and FSS. Although the true value of the dimensionality is unknown in this experiment, UOTS with $q = 1$ shows that the effective dimensionality is about 10. The stress show continuous reduction with UOTS with $L = 5$ and $q = 2$, DAS, RMSE, MAE, ETS, and CORR. It increases for large δ with FSS with $W = 10, 20$ and 40. It is almost constant with FB and FSS with $W = 5$.

These consistency of the results support the advantages of UOTS and also suggest that UOTS can be applied to practical data.

4 Conclusions

This study proposes a novel methodology for extracting meaningful latent variables in low-dimensional space from high-dimensional, sparse data, primarily focusing on spatial distributions. The application of multidimensional scaling with a new similarity metric, namely, the UOTS, proves highly effective in achieving this goal. UOTS offers several advantages over traditional metrics, including incorporating amplitude and location errors and preserving physical meaning within its latent variables.

The estimation of probability distributions from these latent variables using density estimation methods, such as histogram or kernel density estimation, offers substantial analytical advantages over the original high-dimensional space. This approach offers several potential advantages for various applications. For example, it enables the determination of the ensemble mean and spread while considering crucial factors such as location differences, which are vital in numerous meteorological applications. The ensemble mean can be determined as the barycenter using the unbalanced optimal trans-

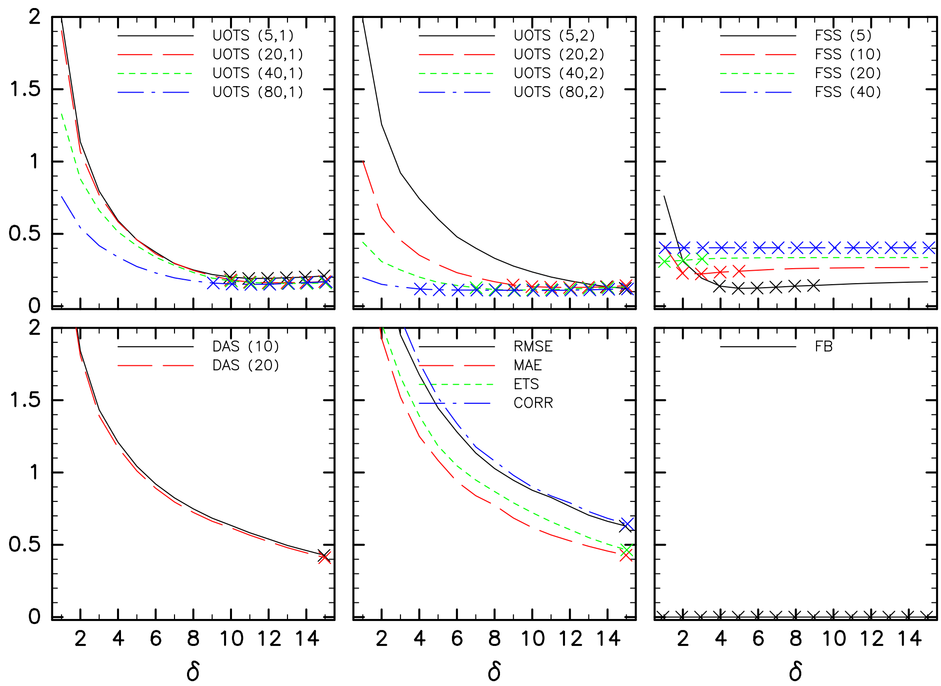


Figure 7. The same figure as Fig. 4 in for the real data experiment.

port theory, whereas the ensemble spread can be derived from the square root of the sum of the eigenvalues obtained by multidimensional scaling. To evaluate the probability distribution and compare distributions in different cases, it is crucial that the Euclidean distance in the latent variable space closely matches the distance in the original high-dimensional space. The UOTS has the same units as the original physical quantity and MDS preserves the units. Therefore, the method using the UOTS and MDS is preferable to consider the probability distribution in low-dimensional space.

The efficacy of this methodology is underscored by its ability to handle discrepancies in spatial distributions by considering the amplitude, location, area, and shape errors. However, the UOTS has two hyperparameters L and q and the efficacy of UOTS depends on these parameters. Therefore, determination of these parameters is one of the challenges of UOTS. Too small L waken the ability to collect the location error, since transport is allowed only within distance of L. On the other hand, too large L makes the score less sensitive to location errors and also creates the danger of equating different phenomena that are far apart. In the limit where L goes to zero and infinity, UOTS is equal to the MAE and the mean mass difference, respectively, and loses its advantages. The synthetic and real-data experiments suggest that a moderate L (around 400 km and 20–40 km in the synthetic data ensemble and real data experiments, respectively) and $q = 1$ lead to the most informative latent variable distribution. Magnitude of the parameter could be guessed based on physical properties of the phenomena of interest such as the spatial extent and typical distance of different phenomena. In the synthetic data ensemble experiment, the standard deviation of center position difference of two ensemble members is 400 km, since the variance of difference in x_1^{fcs} and y_1^{fcs} is $2(50\Delta x)^2$ and $\Delta x = 4$ km. This is almost same scale with the estimated appropriate value of L. In the real data experiment, typical spatial scale of the distribution of the reflectivity (Fig. 2) is estimated to roughly be 10–20 km. From this physical scale, an appropriate L is estimated around 20 km. This is consistent the result of the sweep experiment. These support the validity of choosing L based on the characteristics of physical phenomena of interest.

With q=1, the UOTS is linearly related to the distance between two patterns to be compared, i.e., the location error. This implies that the UOTS with q=1 has similar characteristics to that of the Euclidean distance and would be better matched for the MDS.

Despite of these intuitive consideration, sweep experiment of these parameters may be required to determine the appropriate value in practical cases as well as other similarity metrics having hyperparameters, such as FSS. However, the number of trials for the sweep can be much smaller for UOTS than that for FSS and DAS because of less sensitivity of UOTS to the parameter. Even though the parameters of L, W, and $D_{\max}$ all indicate distance limit for location error or displacement, this smaller sensitivity, in other word stronger robustness, is a preferable feature of UOTS in terms of parameter determination. Furthermore, in the parameter sweep of L, the result of MAE and the mean mass difference may give a hint because they are the limit of the UOTS as L goes to zero and infinity.

One limitation of UOTS is its computational cost compared to conventional metrics. UOTS is obtained by iterative solver, thus the computation time highly depends on input data and parameters. On average, the computation time for the UOTS in the synthetic data experiment was about 5.3 seconds, while it is about 0.06 seconds for FSS on my standard Intel CPU workstation. However, active research in optimal transport is developing faster algorithms (for example, R. Sato et al., 2020). In addition, these algorithms are know to be suitable for graphics processing unit computers. Therefore, these promise future improvement in computational cost of UOTS.

Although the primary focus of this study was on the spatial distributions, this method readily adapts to spatiotemporal distributions with minimal modifications. Incorporating factors, such as advection speed in the temporal direction, into the transport cost of UOTS allows for a seamless extension while maintaining the core methodology.

The versatility of this approach extends to various meteorological applications, for example, comparison of spatial distribution of aerosol and chemical species emitted from specific locations such as Y. Sato et al. (2018). Moreover, this approach is not limited to meteorology and it is also applicable to various fields dealing with sparse spatiotemporal distributions beyond meteorology. Its adaptability to diverse domains and robustness in handling errors makes it a promising tool across scientific disciplines.

Acknowledgments

The author appreciates the associated editor and three reviewers for their efforts and valuable comments. This work was supported by JST [Moonshot R&D Program] Grant Number [JPMJMS2286]. The diagrams in this study were drawn using tools developed by the GFD-Dennou Club (http://www.gfd-dennou.org/).

Open Research Section

The programs for analysis visualization, and input and output data used in this study are available at Nishizawa (2024).

References

Ahijevych, D., Gilleland, E., Brown, B. G., & Ebert, E. E. (2009). Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Weather and Forecasting*, *24*(6), 1485–1497. doi:

10.1175/2009WAF2222298.1

Caffarelli, L. A., & McCann, R. J. (2010). Free boundaries in optimal transport and Monge–Ampere obstacle problems. *Annals of mathematics*, 673–730.

Chizat, L., Peyré, G., Schmitzer, B., & Vialard, F.-X. (2018). Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, *87*(314), 2563–2609. doi: 10.1090/mcom/3303

Constantin, P. (1989). *Integral manifolds and inertial manifolds for dissipative partial differential equations* (Vol. 70). Springer Science & Business Media.

Cox, T. F., & Cox, M. A. (2000). *Multidimensional scaling.* CRC press.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26). Curran Associates, Inc.

De Plaen, H., De Plaen, P.-F., Suykens, J. A. K., Proesmans, M., Tuytelaars, T., & Van Gool, L. (2023). Unbalanced optimal transport: A unified framework for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 3198–3207).

Figalli, A. (2010). The optimal partial transport problem. *Archive for rational mechanics and analysis*, *195*(2), 533–560. doi: 10.1007/s00205-008-0212-7

Foias, C., Sell, G. R., & Temam, R. (1988). Inertial manifolds for nonlinear evolutionary equations. *Journal of differential equations*, *73*(2), 309–353. doi: 10.1016/0022-0396(88)90110-6

Frogner, C., Zhang, C., Mobahi, H., Araya, M., & Poggio, T. A. (2015). Learning with a Wasserstein loss. *Advances in neural information processing systems*, *28*.

Gilbert, G. K. (1884). Finley's tornado predictions. *American Meteorological Journal.*, *1*(5), 166.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., & Ebert, E. E. (2009). Intercomparison of spatial forecast verification methods. *Weather and forecasting*, *24*(5), 1416–1430. doi: 10.1175/2009WAF2222269.1

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151. doi: 10.1146/annurev-statistics-062713-085831

Hanin, L. G. (1992). Kantorovich–Rubinstein norm and its application in the theory of Lipschitz spaces. *Proceedings of the American Mathematical Society*, *115*(2), 345–352. doi: 10.1090/S0002-9939-1992-1097344-5

Keil, C., & Craig, G. C. (2007). A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, *135*(9), 3248–3259. doi: 10.1175/MWR3457.1

Keil, C., & Craig, G. C. (2009). A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, *24*(5), 1297–1308. doi: 10.1175/2009WAF2222247.1

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*. doi: 10.48550/arXiv.1312.6114

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Lee, S.-Y. (2007). *Handbook of latent variable and related models.* Elsevier. doi: 10.1016/B978-0-444-52044-9.X5000-9

Lellmann, J., Lorenz, D. A., Schonlieb, C., & Valkonen, T. (2014). Imaging with Kantorovich–Rubinstein discrepancy. *SIAM Journal on Imaging Sciences*, *7*(4), 2833–2859. doi: 10.1137/140975528

Lien, G.-Y., Miyoshi, T., Nishizawa, S., Yoshida, R., Yashiro, H., Adachi, S. A., ... Tomita, H. (2017). The near-real-time SCALE-LETKF system: A case of the September 2015 Kanto–Tohoku heavy rainfall. *Sola*, *13*, 1–6. doi: 10.2151/sola.2017-001

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis.* Psychology Press. doi: 10.4324/9781410609823

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. doi: 10.48550/arXiv.1802.03426

Miyoshi, T., Amemiya, A., Otsuka, S., Maejima, Y., Taylor, J., Honda, T., ... Uno, A. (2023). Big data assimilation: Real-time 30-second-refresh heavy rain forecast using Fugaku during Tokyo Olympics and Paralympics. In *Proceedings of the international conference for high performance computing, networking, storage and analysis* (pp. 1–10). Association for Computing Machinery. doi: 10.1145/3581784.3627047

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, 666–704.

Murphy, A. H. (1991). Forecast verification: Its complexity and dimensionality. *Monthly Weather Review*, *119*(7), 1590–1601. doi: 10.1175/1520-0493(1991)119⟨1590:FVICAD⟩2.0.CO;2

Narayan, A., Berger, B., & Cho, H. (2021). Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nature biotechnology*, *39*(6), 765–774. doi: 10.1038/s41587-020-00801-7

Nishizawa, S. (2024). *Programs and data used for extracting latent variables from forecast ensembles and advancements in similarity metric utilizing optimal transport [DataSet].* Zenodo. `https://doi.org/10.5281/zenodo.10628522`.

Nishizawa, S., Yashiro, H., Sato, Y., Miyamoto, Y., & Tomita, H. (2015). Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations. *Geoscientific Model Development*, *8*(10), 3393–3419. doi: 10.5194/gmd-8-3393-2015

Nishizawa, S., & Yoden, S. (2004). A parameter sweep experiment on topographic effects on the annular variability. *Journal of the Meteorological Society of Japan. Ser. II*, *82*(3), 879–893. doi: 10.2151/jmsj.2004.879

Peyré, G., & Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, *11*(5-6), 355–607. doi: 10.1561/2200000073

Radanovics, S., Vidal, J.-P., & Sauquet, E. (2018). Spatial verification of ensemble precipitation: an ensemble version of SAL. *Weather and Forecasting*, *33*(4), 1001–1020. doi: 10.1175/WAF-D-17-0162.1

Roberts, N. (2008). Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, *15*(1), 163–169. doi: 10.1002/met.57

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1), 78–97. doi: 10.1175/2007MWR2123.1

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, *290*(5500), 2323–2326. doi: 10.1126/science.290.5500.2323

Sato, R., Yamada, M., & Kashima, H. (2020). Fast unbalanced optimal transport on a tree. *Advances in neural information processing systems*, *33*, 19039–19051.

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., Kajikawa, Y., & Tomita, H. (2015). Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation? *Progress in Earth and Planetary Science*, *2*, 1–16. doi: 10.1186/s40645-015-0053-6

Sato, Y., Takigawa, M., Sekiyama, T. T., Kajino, M., Terada, H., Nagai, H., ... Nakajima, T. (2018). Model intercomparison of atmospheric 137Cs from the Fukushima Daiichi nuclear power plant accident: Simulations based on identical input data. *Journal of Geophysical Research: Atmospheres*, *123*(20), 11–748. doi: 10.1029/2018JD029144

Scott, D. W. (1992). The curse of dimensionality and dimension reduction. In *Multivariate density estimation* (p. 195-217). John Wiley & Sons, Ltd. doi: 10.1002/9780470316849.ch7

Séjourné, T., Peyré, G., & Vialard, F.-X. (2023). Unbalanced optimal transport, from theory to numerics. In E. Trélat & E. Zuazua (Eds.), *Numerical control: Part b* (Vol. 24, pp. 407–471). Elsevier. doi: 10.1016/bs.hna.2022.11.003

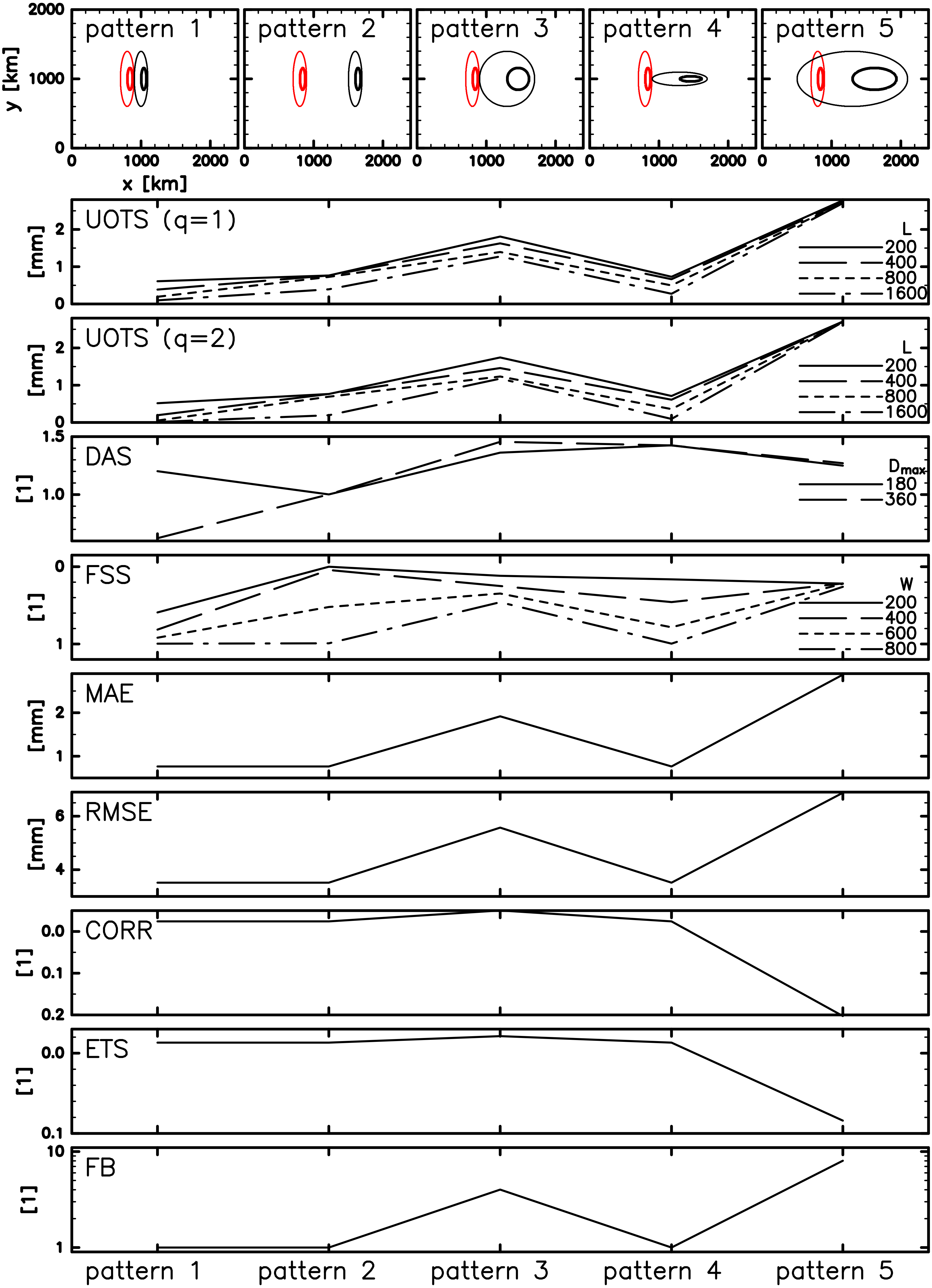
Tenenbaum, J. B., Silva, V. d., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319–2323. doi: 10.1126/science.290.5500.2319

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, *3*(1), 71–86. doi: 10.1162/jocn.1991.3.1.71

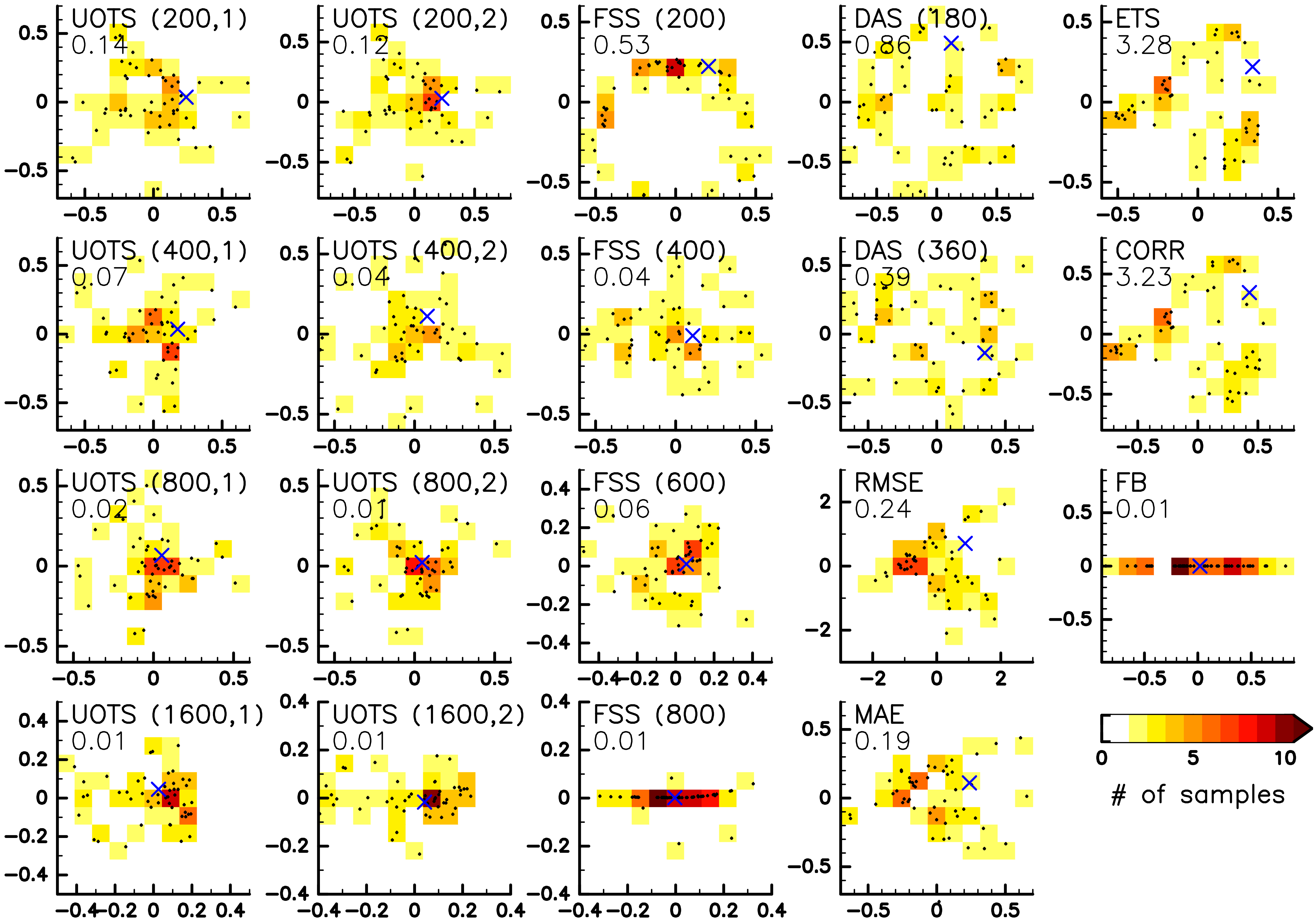
Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*, 2579–2605.

Wernli, H., Paulat, M., Hagen, M., & Frei, C. (2008). SAL—a novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, *136*(11), 4470–4487. doi: 10.1175/2008MWR2415.1

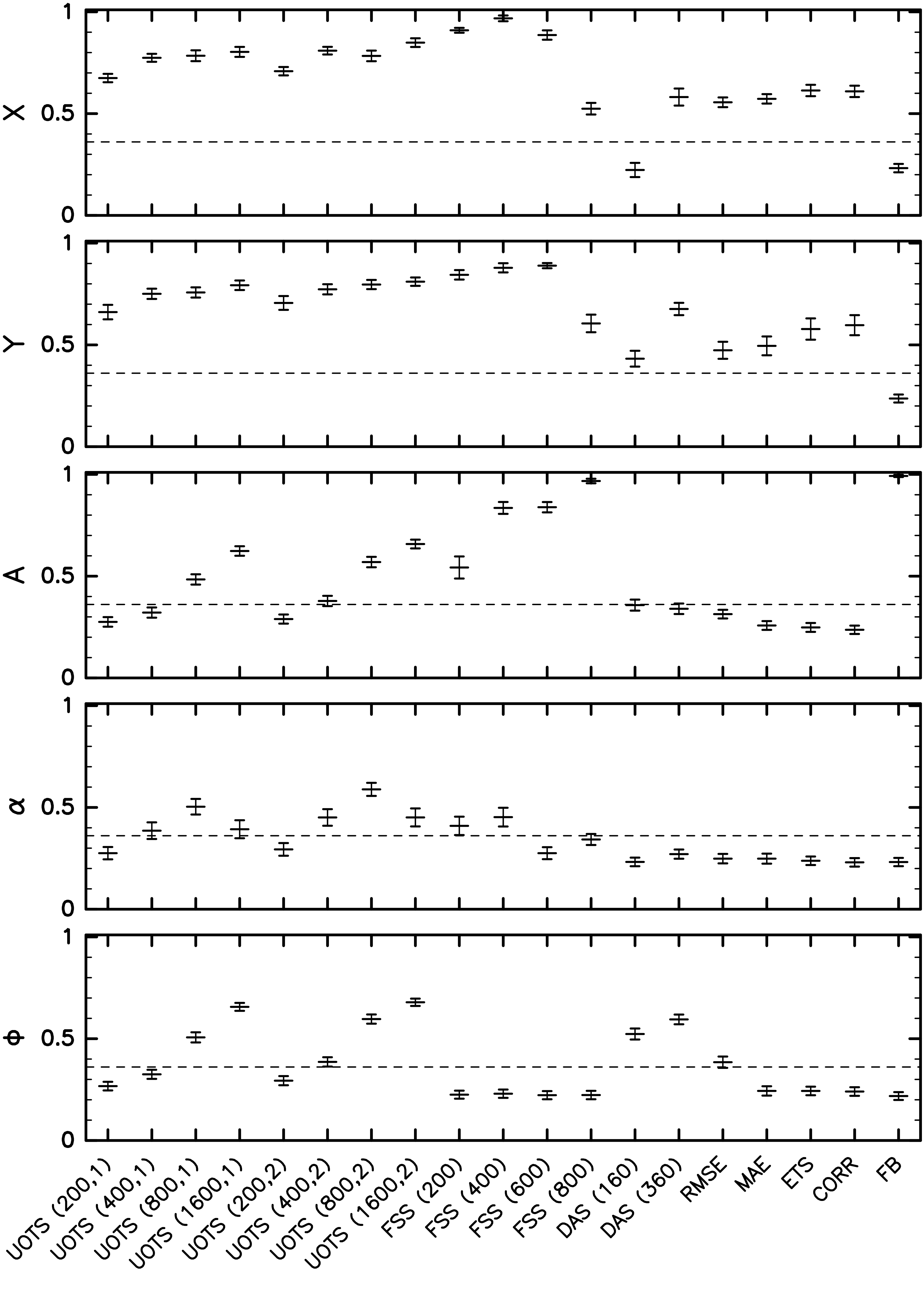
Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (Vol. 91). Academic press.



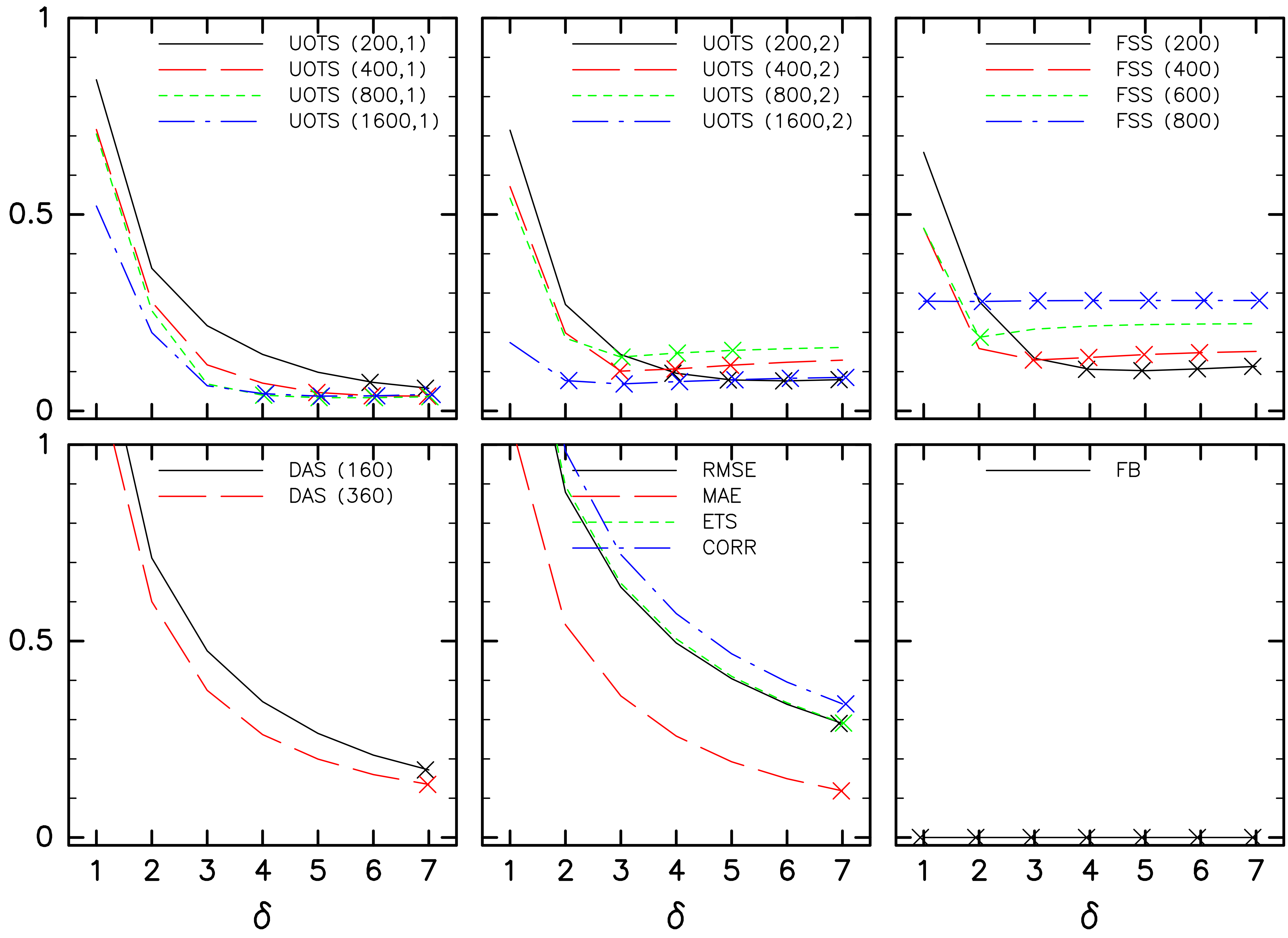
pattern 1
pattern 2
pattern 3
pattern 4
pattern 5
y [km]
x [km]
UOTS (q=1)
UOTS (q=2)
L
200
400
800
1600
[mm]
DAS
D_{max}
180
360
FSS
W
200
400
600
800
MAE
RMSE
CORR
ETS
FB
[1]



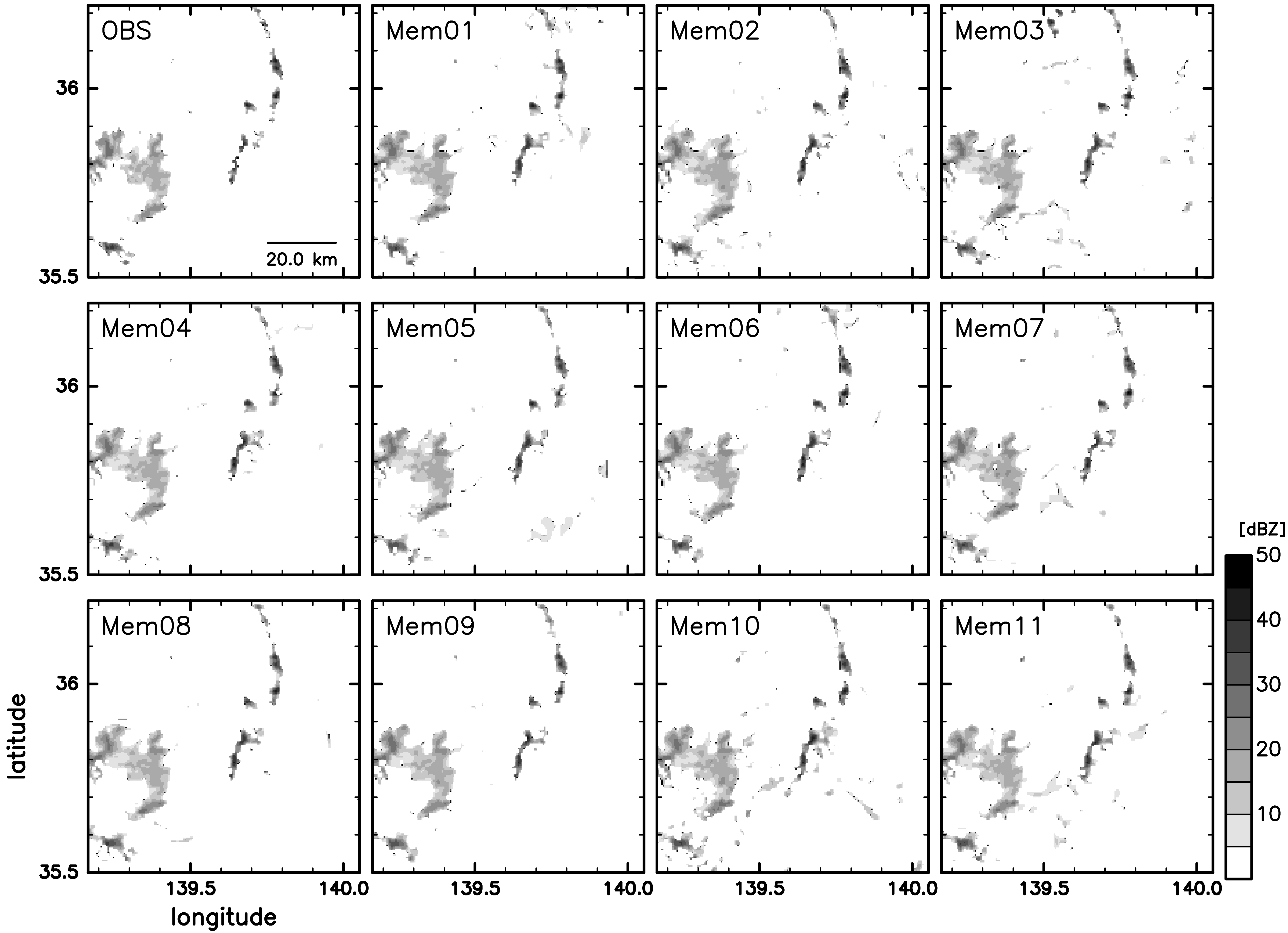
UOTS (200,1)
0.14
UOTS (200,2)
0.12
FSS (200)
0.53
DAS (180)
0.86
ETS
3.28
UOTS (400,1)
0.07
UOTS (400,2)
0.04
FSS (400)
0.04
DAS (360)
0.39
CORR
3.23
UOTS (800,1)
0.02
UOTS (800,2)
0.01
FSS (600)
0.06
RMSE
0.24
FB
0.01
UOTS (1600,1)
0.01
UOTS (1600,2)
0.01
FSS (800)
0.01
MAE
0.19
0
5
10
# of samples



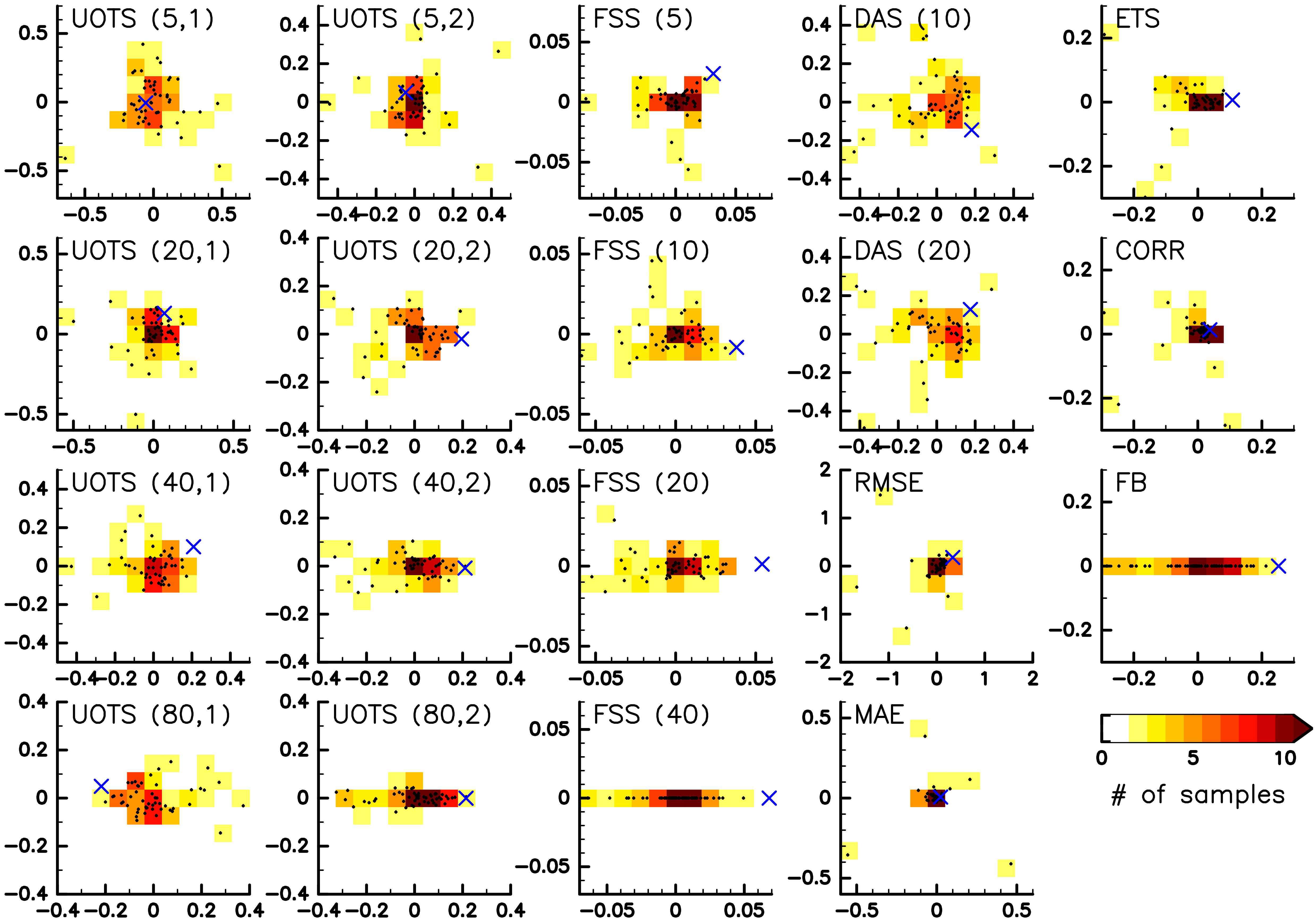
X
Y
A
α
Φ
0
0.5
1
UOTS (200,1)
UOTS (400,1)
UOTS (800,1)
UOTS (1600,1)
UOTS (200,2)
UOTS (400,2)
UOTS (800,2)
UOTS (1600,2)
FSS (200)
FSS (400)
FSS (600)
FSS (800)
DAS (160)
DAS (360)
RMSE
MAE
ETS
CORR
FB

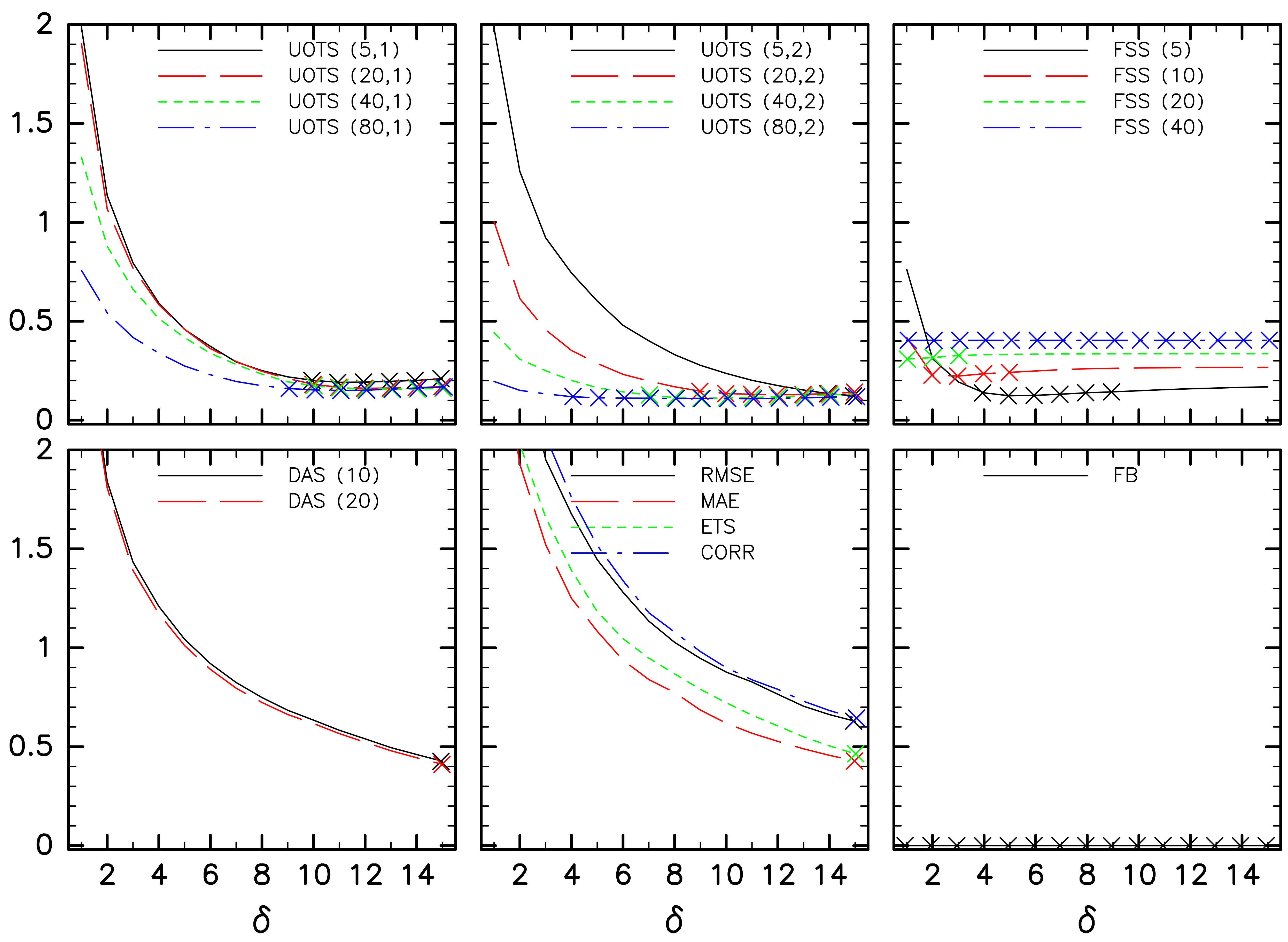


UOTS (200,1)
UOTS (400,1)
UOTS (800,1)
UOTS (1600,1)
UOTS (200,2)
UOTS (400,2)
UOTS (800,2)
UOTS (1600,2)
FSS (200)
FSS (400)
FSS (600)
FSS (800)
DAS (160)
DAS (360)
RMSE
MAE
ETS
CORR
FB
δ



OBS
Mem01
Mem02
Mem03
Mem04
Mem05
Mem06
Mem07
Mem08
Mem09
Mem10
Mem11
20.0 km
36
35.5
139.5
140.0
latitude
longitude
[dBZ]
50
40
30
20
10


UOTS (5,1)
UOTS (5,2)
FSS (5)
DAS (10)
ETS
UOTS (20,1)
UOTS (20,2)
FSS (10)
DAS (20)
CORR
UOTS (40,1)
UOTS (40,2)
FSS (20)
RMSE
FB
UOTS (80,1)
UOTS (80,2)
FSS (40)
MAE
# of samples


UOTS (5,1)
UOTS (20,1)
UOTS (40,1)
UOTS (80,1)
UOTS (5,2)
UOTS (20,2)
UOTS (40,2)
UOTS (80,2)
FSS (5)
FSS (10)
FSS (20)
FSS (40)
DAS (10)
DAS (20)
RMSE
MAE
ETS
CORR
FB
δ
δ
δ