Handling High-Dimensional Data in Air Pollution Forecasting Tasks

Diana Domańska[a], Szymon Łukasik[b]

[a]*Institute of Computer Science, University of Silesia, ul. Będzińska 39, 41–200 Sosnowiec, Poland*

[b]*Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01–447 Warsaw, Poland*

Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland

Abstract

In the paper methods aimed at handling high-dimensional weather forecasts data used to predict the concentrations of PM10, PM2.5, SO_2, NO, CO and O_3 are being proposed. The procedure employed to predict pollution normally requires historical data samples for a large number of points in time – particularly weather forecast data, actual weather data and pollution data. Likewise, it typically involves using numerous features related to atmospheric conditions. Consequently the analysis of such datasets to generate accurate forecasts becomes very cumbersome task. The paper examines a variety of unsupervised dimensionality reduction methods aimed at obtaining compact yet informative set of features. As an alternative, approach using fractional distances for data analysis tasks is being considered as well. Both strategies were evaluated on real-world data obtained from the Institute of Meteorology and Water Management in Katowice (Poland), with extended Air Pollution Forecast Model (e-APFM) being used as underlying prediction tool. It was found that employing fractional distance as a dissimilarity measure ensures the best accuracy of forecasting. Satisfactory results can be also obtained with Isomap, Landmark Isomap and Factor Analysis as dimensionality reduction techniques. These methods can be also used to formulate universal mapping, ready-to-use for data gathered at different geographical areas.

Email addresses: `ddomanska@poczta.onet.pl` (Diana Domańska), `slukasik@ibspan.waw.pl` (Szymon Łukasik)

Keywords: big data, multidimensional data, dimensionality reduction, fractional distances, forecasting, pollution

1. Introduction

The time series analysis and their forecasting constitute one of the most important tasks in many fields – starting with finance and economics (Godarzi et al., 2014), (J. et al., 2015) through social (Ozer et al., 2012), political analysis (Royes and Bastos, 2006) and finishing with forecasts relating to phenomena identified in nature (Karatzas et al., 2008). Individuals try to predict success in life or in investments and – on the other hand – organisations make plans, based on forecasts, to build new infrastructure. In economy, planners predict demands on stock and services, auction prices on the stock exchange or currency exchange rates. Thus the possibility of shrewd forecasting became significantly relevant and practices how to create a good forecast have been developed for years. Variety of methods described in the literature are predominantly used to forecasting in finance and economics. Predicting natural phenomena is however also a matter of intensive research. Weather/air condition described by meteorological parameters (Sudheer and Suseelatha, 2015) or pollution concentrations (Sarıand Öztayşi, 2012; Xihao Sun, 2014) represent physical processes frequently studied – as associated with daily life of every individual. Last years brought a variety of new methods and their modifications aimed at improving prediction accuracy. The commonly used recent approaches for air quality forecasting include multiple linear regression (Genc et al., 2010), stepwise regression and wavelet analysis (Chen et al., 2013), artificial neural networks (Feng et al., 2015), Wang et al. (2015), fuzzy logic (Guo et al., 2007), neuro-fuzzy logic (Mishra and Goyal, 2016) and hidden Markov models (Sun et al., 2013). As for other relatively new proposals, in 2010 a geographic based model to forecast the daily average concentrations of SO_2, CO and PM10 three days in advance using MLP was introduced (Kurt and Oktay, 2010). Authors employed three geographic models: the single-site neighborhood model, the two-site neighborhood model and distance-based model. Experimental results showed that geographic based models perform better than the plain ones, especially when distance-based approach is being used. In (Osowski and Garanty, 2007) an alternative approach employing wavelet transformation with ANN ensemble was used to predict the daily average concentrations of PM10. Such newlycreated created hybrid system based on several types of ANN and wavelet transformation proved to be effective tool for air quality forecasting.

In general weather forecasts belong to a class of data sets referred to as multidimensional (or high-dimensional) data. For example, one daily forecast from Consortium for Small Scale Modelling Local Model (COSMO LM) (Doms and Schattler, 2002) being used in this paper is represented by a matrix of 27 parameters captured every hour – for 72 hours. Challenges of analysing such data structures become clear – especially when we want to treat weather forecasting problem as multivariate one, not only as a time series analysis for a set of individual separate parameters.

The most important challenge of multidimensional data analysis is scalability. It is connected with a set of phenomena commonly referred to as a curse of dimensionality. The complexity of many existing data mining algorithms grows exponentially with respect to the number of dimensions. In addition to that the specificity of similarities between points in a high dimensional space diminishes (Maimon and Rokach, 2010). The problems with multidimensional data are usually dealt in two ways. The first is to use fractional distance as a dissimilarity measure among data sample elements. The second one is to apply specialized methods to reduce data dimensionality (Aggarwal et al., 2001; Beyer et al., 1999; Nowakowska et al., 2016).

In the paper we present and analyse the performance of those techniques – when dealing with high dimensionality weather forecast data. The general task of forecasting here is to predict the concentrations of PM10, PM2.5, SO_2, NO, CO and O_3. We build up in this aspect on our previous studies establishing a core of forecasting method. It is based on two interchangeably used techniques employing explorative forecast procedure (Domańska and Wojtylak, 2014). In this scheme forecast factors are represented by data samples concerning the future (forecasts for the meteorological conditions which are already available) which coexist with historical data, i.e. data from the present and the past, and are correlated with them. The idea of prediction using explorative forecast is based on the similarity between the historical data and the data concerning the future. The result of forecasting is time series data for the chosen phenomenon. The diagram of explorative forecast methodology, as defined in (Domańska and Wojtylak, 2014), is presented in Fig. 1.

The goal of the contribution is to study methods capable of alleviating a problem of highly dimensional forecasting data. To our best knowledge it is the first effort to compare efficiency of a vast selection of dimensionality reduction methods, along with fractional distances, for this class of data mining tasks. The paper is organised as follows. In Section 2 we discuss issues related with multidimensional data mining and describe a variety of methods typically used to deal with high dimensionality of datasets. In Section 3 we present selected problem of air pollution forecasting and demonstrate complexity of available input data. Section 4 contains results of experiments evaluating possible solutions of studied data mining problem along with their discussion. Finally, in Section 5 some concluding remarks are given.

2. Dealing with Multidimensional Data

Let us assume that the dataset is of numerical nature, whose n features' values can be represented by real numbers. It consists of m elements $x_1, x_2, ..., x_m$, with $x_i \in \mathbb{R}^n$, $i = 1, ..., m$.

Traditional data mining algorithms employ the distance measure $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$ defined by the Euclidean metric:

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

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.
In the case of multidimensional datasets however, the ideas of proximity and distance, based on such metric could not be qualitatively significant (Aggarwal et al., 2001). It is due to the so called curse of dimensionality, that in general refers to phenomena which occur when analysing datasets in high-dimensional spaces. It includes exponential grow of sample size needed to achieve proper efficiency of data analysis with increasing dimensionality, so called "empty space phenomenon" and vanishing of distances between close and distant points when using typical norms – like Euclidean one (Łukasik and Kulczycki, 2011).

The problem of data dimensionality and its consequences can be tackled by using more sophisticated distance measures: shared-neighbor distance, hubness-aware measures and – most frequently – by employing fractional p-norms. The first relies on rankings of data objects constructed on the basis of some primary distance measure (Houle et al., 2010). The second identifies influential nodes (hubs) in k-nearest neighbour graphs and tries to decrease the impact of those which occur due high dimensionality of the problem at hand (Tomašev et al., 2014). Finally, fractional p-norm constitutes a generalization of (1) formulated as follows:

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

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with $p \in (0,1)$. In practical applications experimental determination of the exact value for p can be recommended (François et al., 2011).

As an alternative to sophisticated dissimilarity measures algorithms reducing the number of features are often being employed. They are commonly referred to as dimensionality reduction algorithms (or in short DR algorithms). This reduction can be achieved in two ways, either by choosing N most significant coordinates/features (feature selection) or through the construction of a reduced data set, based on initial features (feature extraction) (Inza et al., 2000; Xu and Wunsch, 2009). The latter can be treated as more general – the selection is a particularly simple case of extraction. Noteworthy among extraction procedures are linear methods, where the resulting data set Y is obtained through the linear transformation of initial set X, therefore using the formula:

$$Y = X \cdot A, \tag{3}$$

where A is a matrix of dimension $n \times N$, as well as nonlinear methods for which the transformation can be described by the nonlinear function $g : \mathbb{R}^n \to \mathbb{R}^N$. This group also contains the methods for which such a functional dependence, expressed explicitly, does not exist (Kulczycki and Łukasik, 2014). Characteristics of data dimensionality reduction techniques can also include other important features, e.g. authors in the paper (Lee and M. Verleysen, 2007) point out twelve qualification criteria which can be used to describe and categorize them. The ability to create general model which can be used to obtain reduced representation of new data points is one of great importance. This feature is commonly referred to as out-of-sample extension. It not only allows to perform dimensionality reduction once and then use its results for incoming data, but also permits using smaller subset of initial data entries to formulate the mapping. Consequently such a technique is particularly useful in case of large datasets and sophisticated dimensionality reduction procedures. Other important characteristic of algorithms covered here is the ability to work autonomously, without using results of exploration procedures. Techniques possessing this property, through analogy to machine learning methods, are termed as unsupervised (Bartenhagen et al., 2010; Kumar, 2009).

There exists a vast literature providing a survey of the state-of-art in the field of data dimensionality algorithms (e.g. (Burges, 2010) or (Sarveniazi, 2014)). Among them (Maaten van der et al., 2009) deliver the most comprehensive systematic overview of methodological approaches being employed for feature extraction tasks. They are classified into two major groups: convex and non-convex techniques (Maaten van der et al., 2009). It underlines a character of optimized objective function – non-convex techniques optimize functions which contain local optima, whereas convex ones deal with unimodal cost functions. Table 1 provides examples of dimensionality reduction techniques, along with their properties, required parameters (including their values used in the paper established through bibliographical studies) and related references. As all of them will be employed in the subsequent part of the paper they will be covered briefly in the following part of this Section.

Principal Component Analysis (PCA) constitutes standard and most commonly used linear dimensionality reduction. Its roots can be traced back to works of Pearson (Pearson, 1901) and Hotelling (Hotelling, 1933). PCA is based on orthogonal transformation, in a general form given by (3), which transforms the data to a new reduced, feature space characterized by the greatest variance of projected data along coordinate system axes. Technically the transformation is given by principal eigenvectors (i.e. principal components) of the standardized data sample covariance matrix. PCA does not require significant computational effort and needs only one input parameter, shared by almost all DR techniques – dimensionality of reduced feature space N. Original PCA algorithm found numerous modifications and improvements. Notably, Probabilistic Principal Component Analysis (Roweis, 1997) (in short: ProbPCA) generates a linear-Gaussian latent-variable model which allows to tackle problems for which computation of the eigenvectors might be infeasible. It is due to a possibility of using ExpectationMaximization (EM) algorithm for ProbPCA and enables handling missing data appropriately. ProbPCA requires number of EM algorithm iterations i_{max} as an additional input parameter. Factor Analysis, introduced first by Spearman (Spearman, 1904) share the idea of using linear-Gaussian model, it assumes however that model and sample covariances are equal (Burges, 2010). Again providing number of EM algorithm iterations i_{max} is necessary.

Kernel PCA represents another important modification of PCA by using so called "kernel trick" (Schölkopf et al., 1998). Instead of principal eigenvectors of the covariance matrix Kernel PCA uses the eigenvectors of kernel matrix, created by transforming dataset by means of selected positive-semi definite kernel function K, choice of which can be considered as an input parameter (typically normal kernel K_N can be used). Consequently Kernel PCA benefits from a property of constructing nonlinear mappings.

Preserving pairwise distances throughout the dimensionality reduction process represents an intuitive goal of feature set reduction. Multidimensional Scaling (MDS) – procedure which generally aims at embedding sample with given dissimilarity data into a low-dimensional Euclidean space – can be used for that purpose (Cox and Cox, 2000). Classical MDS solves the problem of minimizing sum of squared Euclidean distance between the high and low-dimensional data representation by the eigendecomposition of the Gram matrix of the high-dimensional data. When the number of points is very large it may become problematic. Landmark MDS (LMDS) aims at overcoming those difficulties. It first selects a set of r_L representative sample elements, referred to as landmark points, constructs their embedding in reduced feature space and finally applies distance-based triangulation to obtain new representation of the whole dataset (de Silva and Tenenbaum, 2004). The ratio of landmark points is the only input parameter required for LMDS. Stochastic Proximity Embedding (SPE) constitutes another implementation of MDS concept. Like Sammon Mapping (Sammon, 1969) SPE formulates MDS as an optimization problem with the sum of differences among distances in the initial and reduced feature spaces – so called stress function – being treated as an actual indicator of mapping's quality. SPE can be also aimed at optimizing this function only within neighbourhood of each sample element of specified radius. Furthermore, SPE uses iterative update of the low-dimensional dataset's representation, with two additional parameters – number of points updated in each iteration m_{up} and λ – learning parameter which is decreased in every step of SPE procedure.

Another class of DR algorithms is based on preserving pairwise geodesic (measured over the manifold) distances between data points. Isomap (IM) – established representative of this group of methods – estimates those distances using the shortest path between two points in the neighbourhood graph (every data point is connected with its k neighbours, with k being IM parameter). Formed pairwise geodesic distance matrix is then transformed using classical MDS (Tenenbaum et al., 2000). Landmark Isomap (LIM) constitutes a modification of this approach by creating distance matrix for each data element but only to the selected landmark points (de Silva and Tenenbaum, 2003). Again, a ratio of landmark points needs to be provided.

Local Linear Embedding (LLE) represents a typical method of local non-linear dimensionality reduction. Similarly to Isomap it constructs a neighbourhood graph first. However LLE preserves only a local geometry of the manifold around each data sample element by representing it through a linear combination – the so-called reconstruction weights – of its k nearest neighbours, number of which has to be supplied. Technically low dimensional embedding is constructed by using eigenvectors – corresponding to the smallest non-zero eigenvalues – of the inner product of reconstruction weight matrix W subtracted from the identity matrix I (Roweis and Saul, 2000). Laplacian Eigenmaps (Laplacian in short) is another technique of this kind, i.e. aimed at preserving local properties of the manifold. It includes additional weights corresponding to the proximity index in the k-nearest neighbourhood set. It means that the nearest neighbor contribution to the minimized cost function is the highest. Identifying low dimensional embedding is again formulated as eigenvalue problem by means of spectral graph theory (Belkin and Niyogi, 2003). As the weights of the edges in the neighbourhood graph are computed using the Gaussian kernel function an additional parameter – deviation of this function σ – needs to be provided. Linear variants of this technique – Locality Preserving Projections (LPP) and Neighbourhood Preserving Embedding (NPE) – can also be named (He and Niyogi, 2003; He et al., 2005).

The following two techniques: Local Tangent Space Alignment (LTSA) and Linear Local Tangent Space Alignment (LLTSA) employ a concept of a tangent space. LTSA first identifies local properties of a dataset by constructing tangent spaces for each data element. It is done by means of PCA executed on a set of k-nearest neighbours. Then a mapping from a local tangent space coordinates to the low-dimensional data representation is being performed (Zhang and Hongyuan, 2004). LLTSA constitutes a modification of this approach – it uses linear approximation of LTSA mapping (Zhang et al., 2007). Both algorithms require providing a value of k as an input parameter.

The most recent technique being analysed here is t-Distributed Stochastic Neighbourhood Embedding (t-SNE). It represents an improved variant of Stochastic Neighbourhood Embedding (SNE) introduced by Hinton and Roweis (Hinton and Roweis, 2002). SNE techniques start with calculating similarity matrices in both the original data space and in the low-dimensional embedding space in a way, that the similarities form a probability distribution over pairs of objects (van der Maaten and Hinton, 2008). The probabilities in t-SNE are given by Student-t kernel computed from the input data and from the embedding. The mapping itself is synthesized by minimizing the Kullback-Leibler divergence between the two probability distributions. t-SNE requires providing desired perplexity p of kernels used for forming similarity matrices.

To conclude the list of exemplary dimensionality reduction techniques an unconventional algorithm of Random Projections (RP) will be named. Similarly to PCA it construct linear mapping. It is based however on randomly generated transformation matrix of given size. A rationale behind this technique is based on the Johnson-Lindenstrauss lemma (Johnson and Lindenstrauss, 1984). It states that if initial dataset elements are projected into a randomly selected subspace of appropriately high dimension, then the distances between those points are approximately preserved.

Finally, it is important to observe that all methods of feature extraction covered above require providing dimensionality of reduced feature space N. The problem of estimating intrinsic dimensionality of a dataset to select proper value of N is very important from the data mining perspective. Overview of possible solutions can be found in (Camastra, 2003). Practically a criterion based on explained variance is frequently employed. Alternatively for example the Maximum Likelihood Estimator (MLE) can be used (Levina and Bickel, 2005). This approach was selected to be applied here.

3. Problem of Data Dimensionality and Air Pollution Forecasting

Climate data analysis constitutes one of the areas where effective techniques of multidimensional data mining are of great importance. Climatic models involve using numerous parameters which need to be included into the scheme of forecasting. As such weather and pollution forecasting represents a big data challenge (Rose, 2012; Zheng et al., 2013, 2015) characterized inter alia by plethora of available features. Existing studies dealing with this issue involve employing selected method of dimensionality reduction. In particular the applications of Isomap (Ross et al., 2008), PCA (Lopez, 2006) and Diffusion Maps (Gonzalez et al., 2012) were under investigation. No possibility of using variety of methods – and fractional distances at the same time – was explored up to date.

Weather and pollution level forecasts are computed using mathematical equations for the physics and dynamics of the atmosphere. A variety of numerical atmospheric prediction models (Baer, 2000) and air quality forecasting models (Zhang et al., 2012a,b) are being used. However neither of the used weather forecasts models have accurate verifiability, so various centers are using different models. For example in Poland Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw is using unified model – Coupled Ocean/Atmosphere Mesoscale Prediction System and wind wave model (Jakubiak et al., 2007). Governmental Institute of Meteorology and Water Management on the other hand employs COSMO LM – which will be used in this study as well.

The model considered in the paper forecasts pollution concentrations, e.g. particulate matter PM10 and PM2.5 as well as SO_2, NO, CO, O_3 for any chosen day (usually next day) or hours. Explorative forecast procedures will be realized using e-APFM technique (Domańska and Wojtylak, 2014) – an extension of the older APFM algorithm (Domańska and Wojtylak, 2012a). The input data for APFM model were weather forecasts (past-future), meteorological parameters and pollution concentrations (past-present). In the e-APFM additional information in the form of wind direction in sectors and similarities among meteorological stations is available. The output are predicted values of pollution concentrations. The process of e-APFM model formulation is divided into following steps:

Step 1. Data preparation.
Step 2. Defining the set of similar weather forecasts.
Step 3. Defining the subset of similar meteorological parameters.
Step 4. Defining the set of fuzzy numbers (FN) for the subset of meteorological parameters.
Step 5. Determining the degrees of membership of a subset of meteorological parameters to the fuzzy numbers, and defining the set of similar meteorological parameters.
Step 6. Defining the set of similar pollution parameters.
Step 7. Updating of the similar pollution parameters set with the similarity of a chosen station to the other stations.
Step 8. Calculating the forecast outputs.

The APFM model does not include step number 7. The core of both algorithms is calculation of similarities, in particular between weather forecasts

– data structures used for that purpose will be described below. One hourly forecast is represented as a matrix of 27 parameters. Daily forecasts are formed from hourly forecasts collected for 72 hours, with time step equal to 1 hour. The numeric weather forecasts used in the paper are similar to the ones employed already in (Domańska and Wojtylak, 2012a). They come from the COSMO LM model obtained for Silesian Area in Poland and were collected between January 1st 2005 and December 31st 2012. A summary of variables, including their ranges are given in Table 2.

4. Experimental Setup and Selected Results

4.1. Error measure

The output from the model represents time series data of pollution concentrations for 72 hours with one hour time step. To estimate quality of the forecast RMSE measure has been used (Domańska and Wojtylak, 2012b). Let q be a time series of the forecasted phenomenon in period $\{1, \ldots, N\}$ and h be a measured (actual) course of the forecast phenomenon in the same period.

Then the $RMSE$ (Root Mean Square Error) error is defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(q_i - h_i)^2}. \tag{4}$$

4.2. Meteorological database

The meteorological database used in the experiments consists of numeric weather forecasts. The database composed of meteorological parameters and pollution concentrations was collected from 1 January 2005 to 31 December 2012. All the data was collected with a time step $\Delta t = 1$ hour. It means that for each area it contains 231 786 entries (in other words dataset numerosity $m = 210384$).

In the paper values of pollution concentrations come from three stations: Wodzislaw Slaski, Cieszyn and Katowice. The time horizon for weather forecasts was 72 hours and for all other data 24 hours.

4.3. Experiments

As the core of e-APFM consists of calculating distances between forecasts various strategies of dealing with multidimensional data were implemented in this step. Traditional approach based on the Euclidean distance was compared with fractional distance (with $p = 0.5$ established experimentally) and dimensionality reduction algorithms covered in Section 2. For the latter as intrinsic dimensionality $N = 7$ was used. This specific value was obtained by employing MLE estimator introduced briefly in Section 2.

Experiments were conducted using two schemes of air pollution forecast calculation. In the first (referred to as hourly forecast) the forecasts were obtained with step equal 1 hour, time horizon equal 72 hours and the concentration of pollutants was averaged for each consecutive day (+1 day, +2 day, +3 day). In the second scheme (called daily average) the forecasts were calculated as a single mean value for each day. Additionally three variants of forecasting methods used to create one time series from a set of similar time series using meteorological data were employed: average, α-standardization, $\alpha\beta$-standardization. More details on those procedures can be found in (Domańska and Wojtylak, 2009).

Full set of results for the following air pollution forecasting instances:

- hourly forecast for Katowice station, e-APFM average method
- hourly forecast for Katowice station, e-APFM α-standardization method
- hourly forecast for Katowice station, e-APFM $\alpha\beta$-standardization method
- hourly forecast for Wodzislaw Slaski station, e-APFM average method
- hourly forecast for Wodzislaw Slaski station, e-APFM α-standardization method
- hourly forecast for Wodzislaw Slaski station, e-APFM $\alpha\beta$-standardization method
- daily average forecast for Cieszyn station, e-APFM average method
- daily average forecast for Cieszyn station, e-APFM α-standardization method
- daily average forecast for Cieszyn station, e-APFM $\alpha\beta$-standardization method
- daily average forecast for Wodzislaw Slaski station, e-APFM average method

- daily average forecast for Wodzislaw Slaski station, e-APFM α-standardization method

- daily average forecast for Wodzislaw Slaski station, e-APFM $\alpha\beta$-standardization method

was enclosed in Appendix A. It contains RMSE of forecasts for CO, NO_2, O_3, PM_{10} and SO_2 obtained by using various strategies of dealing with data multidimensionality. Along with error measures their relative values, when compared to the forecasting performance using Euclidean distance, were provided.

Table 3 summarizes the results of experiments. Strategies of alleviating data dimensionality were ranked for each air pollution forecasting instance by their average RMSE for all pollutants (displayed relative to Euclidean distance-based forecasting). Mean and standard deviations were shown in the last columns of the table.

Best performing dimensionality reduction algorithms: Principal Component Analysis, Factor Analysis and Landmark Isomap (as less computationally intensive than similar Isomap) were also evaluated in terms of error expressed in Common Air Quality Index Bands (CAQI). It represents a common method of classifying pollution levels. Pollutants concentration was classified into bands according to CAQI (shown in Tab. 7). Reported error represents the difference between the observed and forecast bands in which the observed and forecast values fall. For example if the actual and forecast pair for PM10 is (14, 31) it is reported as 1 band in the band error method, since 14 falls into the first interval and 31 falls into the second interval. The average results for all methods were shown in Tables 8-10. We selected NO_2, O_3 and PM10 concentrations as the main pollutants for the analysis. For each forecast time frame percentages of forecasts with different error values were provided.

Aforementioned techniques were also studied for the possibility of using out-of-sample extension. For that purpose dimensionality reduction was performed for different area then the one for which the forecast was to be constructed. Obtained mapping was then used for forecasting – to transform the data to new, reduced feature space. Tables 4-6 contains the results of forecasting using this method, with a relative change in the RMSE value when compared with procedure based on dimensionality reduction performed on data used for prediction. Results for which forecasting accuracy was better then the one using Euclidean distance were written with bold font. As generic out-of-sample extension is not a feature of Landmark Isomap the Nyström approximation (Platt, 2005) is used for this algorithm.

4.4. Discussion

Results of experiments demonstrate interesting tendencies and offer a room for notable observations. They will be covered here in detail. First of all it was found that employing fractional distance as a dissimilarity metric causes a significant positive shift in mean RMSE over tested forecasting instances, as well as its standard deviation. Using dimensionality reduction on the other hand does not always ensure an improvement in prediction accuracy. The between methods' span of RMSE varies from 10% to 40% of Euclidean distance algorithm's based RMSE. Only five methods offered mean RMSE lower that standard forecasting algorithm based on Euclidean distance. Here, Isomap, Landmark Isomap and Factor Analysis were shown to be superior for all prediction instances. When taking into account forecasts' time horizon it should be noted that, intuitively, in the reduced feature space short-term prognosis are synthesized more accurately. What is more however, if the mapping does not capture internal data structure properly, accuracy of prediction deteriorates faster. Finally, dimensionality reduction can be more beneficial for e-APFM based forecasting using α standardization and $\alpha\beta$ standardization method. For average method the application of some algorithms (e.g. Kernel PCA, LLE, SPE) brought unsatisfactory results. The analysis using CAQI allows to evaluate relative accuracy of methods with respect to commonly accepted classification of air quality indicators. It was generally observed that most of forecasts were highly accurate – with forecasted CAQI value being the same as the actual one. The most difficult to predict was the concentration of PM10. When observing the results of prediction using out-of-sample extension it could be noted that for most cases using the transformation obtained for different geographical area is sufficient to ensure accuracy which is better than the one attained for Euclidean distance. Best results were identified for CO concentration. Conversely, the level of PM10 was the hardest to predict using this strategy.

5. Conclusion

The paper studied 17 methods which can be used for alleviating the problem of data dimensionality in air pollution forecasting. Past meteorological data from Silesia region in Poland was employed in the experiments. They were aimed at establishing the solution offering best prediction accuracy in terms of RMSE. It was identified that employing fractional distance offers highest accuracy and is the most stable one. Three dimensionality reduction algorithms were also found to be adequately effective, also for establishing mappings which can be universally used for data obtained in different geographical areas. Evaluation of forecasting accuracy using Common Air Quality Index allows to conclude that these methods ensure relatively high conformance of forecasted pollution levels with actual CAQI values. We anticipate that results of our work could be beneficial not only for research dealing with meteorological information systems but also for other studies related to multidimensional data analysis tackling data of the same numerical nature. All of the calculations were implemented in the .NET environment, and the experiments were conducted using a computer with Intel Core I5-2520M 2.50 GHz processor, 4 GB DDR2 RAM memory running Windows 7 (64-bit).

Acknowledgment

Second author's contribution was co-funded by the European Union from resources of the European Social Fund. Project PO KL "Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00.

References

A. A. Godarzi, R. M. Amiri, A. Talaei, T. Jamasb, Predicting oil price movements: A dynamic Artificial Neural Network approach, Energy Policy 68 (0) (2014) 371–382.

P. J., S. Shah, P. Thakkar, K. Kotecha, Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques, Expert Systems with Applications 42 (1) (2015) 259–268.

S. U. Ozer, G. K. Ersoy, D. Tuzunkan, {DARK} {TOURISM} {IN} GALLIPOLI: Forecast Analysis to DeterminePotential of AustralianVisitors, Procedia - Social and Behavioral Sciences 41 (0) (2012) 386–393, the First International Conference on Leadership, Technology and Innovation Management.

G. F. Royes, R. C. Bastos, Uncertainty analysis in political forecasting, Decision Support Systems 42 (1) (2006) 25–35, ISSN 0167-9236.

K. Karatzas, G. Papadourakis, I. Kyriakidis, Understanding and forecasting atmospheric quality parameters with the aid of ANNs, in: Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on, 2580–2587, 2008.

G. Sudheer, A. Suseelatha, Short term load forecasting using wavelet transform combined with HoltWinters and weighted nearest neighbor models, International Journal of Electrical Power & Energy Systems 64 (0) (2015) 340–346.

I. U. Sarı, B. Öztayşi, Forecasting Energy Demand Using Fuzzy Seasonal Time Series, in: C. Kahraman (Ed.), Computational Intelligence Systems in Industrial Engineering, vol. 6 of *Atlantis Computational Intelligence Systems*, Atlantis Press, 251–269, 2012.

Y. L. Xihao Sun, Predicting Dissolved Oxygen Fluctuations in Golden Horn by Fuzzy Time Series, International Journal of Nonlinear Science 17 (3) (2014) 234–240.

D. Genc, C. Yesilyurt, G. Tuncel, Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere, Environmental Monitoring and Assessment 166 (1-4) (2010) 11–27, ISSN 0167-6369.

Y. Chen, R. Shi, S. Shu, W. Gao, Ensemble and enhanced {PM10} concentration forecast model based on stepwise regression and wavelet analysis, Atmospheric Environment 74 (2013) 346 – 359, ISSN 1352-2310, doi:http://dx.doi.org/10.1016/j.atmosenv.2013.04.002, URL `http://www.sciencedirect.com/science/article/pii/S1352231013002471`.

X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, J. Wang, Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, Atmospheric Environment 107 (2015) 118 – 128.

P. Wang, Y. Liu, Z. Qin, G. Zhang, A novel hybrid forecasting model for {PM10} and {SO2} daily concentrations, Science of The Total Environment 505 (2015) 1202 – 1212, ISSN 0048-9697, doi:http://dx.doi.org/10.1016/j.scitotenv.2014.10.078, URL `http://www.sciencedirect.com/science/article/pii/S004896971401523X`.

D. Guo, R. Guo, C. Thiart, Predicting air pollution using fuzzy membership grade Kriging, Computers, Environment and Urban Systems 31 (1) (2007) 33 – 51.

D. Mishra, P. Goyal, Neuro-Fuzzy approach to forecasting Ozone Episodes over the urban area of Delhi, India, Environmental Technology & Innovation 5 (2016) 83 – 94, ISSN 2352-1864, doi:http://dx.doi.org/10.1016/j.eti.2016.01.001, URL `http://www.sciencedirect.com/science/article/pii/S2352186416300013`.

W. Sun, H. Zhang, A. Palazoglu, Prediction of 8A h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models, Atmospheric Environment 81 (2013) 199 – 208.

A. Kurt, A. B. Oktay, Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks, Expert Systems with Applications 37 (12) (2010) 7986 – 7992.

S. Osowski, K. Garanty, Forecasting of the daily meteorological pollution using wavelets and support vector machine, Engineering Applications of Artificial Intelligence 20 (6) (2007) 745 – 755.

G. Doms, U. Schattler, The non-hydrostatic limited-area model LM (LokalModell) of DWD Part I: Scientific documentation, Deutscher Wetterdienst, 2002.

O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, 2nd ed, Springer, 2010.

C. Aggarwal, A. Hinneburg, D. Keim, On the Surprising Behavior of Distance Metrics in High Dimensional Space, Lecture Notes in Computer Science 1973 (2001) 420–434.

K. S. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When Is "Nearest Neighbor" Meaningful?, in: Proceedings of the 7th International Conference on Database Theory, ICDT '99, Springer-Verlag, London, UK, UK, ISBN 3-540-65452-6, 217–235, 1999.

E. Nowakowska, J. Koronacki, S. Lipovetsky, Dimensionality reduction for data of unknown cluster structure, Information Sciences 330 (2016) 74 – 87, ISSN 0020-0255, doi:http://dx.doi.org/10.1016/j.ins.2015.10.009, URL `http://www.sciencedirect.com/science/article/pii/S0020025515007264`, sI Visual Info Communication.

D. Domańska, M. Wojtylak, Explorative forecasting of air pollution, Atmospheric Environment 92 (2014) 19–30, ISSN 1352-2310.

S. Łukasik, P. Kulczycki, An Algorithm for Sample and Data Dimensionality Reduction Using Fast Simulated Annealing, Lecture Notes in Computer Science 7120 (2011) 152–161.

M. Houle, H.-P. Kriegel, P. Kröger, E. Schubert, A. Zimek, Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?, in: M. Gertz, B. Ludscher (Eds.), Scientific and Statistical Database Management, vol. 6187 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 482–500, 2010.

N. Tomašev, M. Radovanović, D. Mladenić, M. Ivanović, Hubness-based fuzzy measures for high-dimensional k-nearest neighbor classification, International Journal of Machine Learning and Cybernetics 5 (3) (2014) 445–458.

D. François, V. Wertz, M. Verleysen, Choosing the Metric: A Simple Model Approach, in: N. Jankowski, W. Duch, K. Grabczewski (Eds.), Meta-Learning in Computational Intelligence, vol. 358 of *Studies in Computational Intelligence*, Springer Berlin Heidelberg, 97–115, 2011.

I. Inza, P. Larranaga, R. Etxeberria, S. B., Feature Subset Selection by Bayesian network-based optimization, Artificial Intelligence 123 (12) (2000) 157–184.

R. Xu, D. Wunsch, Clustering, Wiley, New Jersey, 2009.

P. Kulczycki, S. Łukasik, An Algorithm for Reducing Dimension and Size of Sample for Data Exploration Procedures, International Journal of Applied Mathematics and Computer Science 24 (2014) 133–149.

J. Lee, M. M. Verleysen, Nonlinear dimensionality reduction, Springer, 2007.

C. Bartenhagen, H.-U. Klein, C. Ruckert, X. Jiang, M. Dugas, Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data, BMC Bioinformatics 11, paper no 567.

C. Kumar, Analysis of Unsupervised Dimensionality Techniques, Computer Science and Information Systems 6 (2) (2009) 217–227.

C. Burges, Dimension Reduction: A Guided Tour, Now Publishers Inc., Delft, 2010.

A. Sarveniazi, An Actual Survey of Dimensionality Reduction, American Journal of Computational Mathematics 4 (2014) 55–72.

L. Maaten van der, E. Postma, H. Herik van den, Dimensionality Reduction: A Comparative Review, Technical Report TiCC-TR 2009-005, Tilburg University, 2009.

K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, Philosophical Magazine Series 6 2 (11) (1901) 559–572.

H. Hotelling, Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology 24 (1933) 417–441.

S. Roweis, EM algorithms for PCA and sensible PCA, Technical Report, California Institute of Technology, Computation and Neural Systems, 1997.

C. Spearman, General intelligence objectively determined and measured, American Journal of Psychology 5 (1904) 201–293.

B. Schölkopf, A. Smola, K.-R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (1998) 1299–1319.

T. Cox, M. Cox, Multidimensional Scaling, Chapman and Hall, Boca Raton, 2000.

V. de Silva, J. Tenenbaum, Sparse multidimensional scaling using landmark points, Technical Report, Stanford University, 2004.

J. Sammon, A nonlinear mapping for data structure analysis, IEEE Transactions on Computers 18 (1969) 401–409.

J. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.

V. de Silva, J. Tenenbaum, Global versus local methods in nonlinear dimensionality reduction, in: Advances in Neural Information Processing Systems, vol. 15, The MIT Press, 721–728, 2003.

S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

M. Belkin, P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation 15 (2003) 1373–1396.

X. He, P. Niyogi, Locality Preserving Projections, in: Advances in Neural Information Processing Systems, MIT Press, Cambridge, 153–160, 2003.

X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: Proceedings of the 10th IEEE International Conference on Computer Vision, 1208–1213, 2005.

Z. Zhang, Z. Hongyuan, Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment, SIAM Journal on Scientific Computing 26 (2004) 313–338.

T. Zhang, J. Yang, D. Zhao, X. Ge, Linear local tangent space alignment and application to face recognition, Neurocomputing 70 (79) (2007) 1547–1553.

G. Hinton, S. Roweis, Stochastic Neighbor Embedding, in: Advances in Neural Information Processing Systems, vol. 15, The MIT Press, Cambridge, 833–840, 2002.

L. van der Maaten, G. Hinton, Visualizing High-Dimensional Data Using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605.

W. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemporary Mathematics 26 (1984) 189–206.

F. Camastra, Data dimensionality estimation methods: a survey, Pattern Recognition 36 (2003) 2945–2954.

E. Levina, P. Bickel, Maximum Likelihood Estimation of Intrinsic Dimension, in: Advances in Neural Information Processing Systems, The MIT Press, Cambridge, 777–784, 2005.

S. Rose, Big data and the future, Significance 9 (4) (2012) 47–48.

Y. Zheng, F. Liu, H.-P. Hsieh, U-Air: When Urban Air Quality Inference Meets Big Data, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, ACM, New York, NY, USA, ISBN 978-1-4503-2174-7, 1436–1444, 2013.

Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, T. Li, Forecasting Fine-Grained Air Quality Based on Big Data, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, ACM, New York, NY, USA, 2267–2276, 2015.

I. Ross, P. Valdes, S. Wiggins, ENSO dynamics in current climate models: an investigation using nonlinear dimensionality reduction, Nonlinear Processes in Geophysics 15 (2008) 339–363.

A. J. G. Lopez, Application of nonlinear dimensionality reduction to climate data for prediction, Ph.D. thesis, Universitat Potsdam, 2006.

A. F. A. M. Gonzalez, J. Diaz, J. R. Dorronsoro, Diffusion Maps for the Description of Meteorological Data, in: C. E., V. Snasel, A. Abraham, M. Wozniak, M. Grana, S.-B. Cho (Eds.), Hybrid Artificial Intelligent Systems, vol. 7208 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 276–287, 2012.

F. Baer, Numerical weather prediction, in: M. V. Zelkowitz (Ed.), Fortieth Anniversary Volume: Advancing into the 21st Century, vol. 52 of *Advances in Computers*, Elsevier, 91–157, 2000.

Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, A. Baklanov, Real-time air quality forecasting, part I: History, techniques, and current status, Atmospheric Environment 60 (2012a) 632–655.

Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, A. Baklanov, Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects, Atmospheric Environment 60 (2012b) 656–676.

B. Jakubiak, R. Hodur, P. Flatau, M. Witek, O. Kapala, L. Herman-Izycki, Implementation and Research on the Operational Use of the Mesoscale Prediction Model COAMPS in Poland, Annual Report ADA541252, Naval Research Laboratory, Monterey, 2007.

D. Domańska, M. Wojtylak, Application of fuzzy time series models for forecasting pollution concentrations, Expert Systems with Applications 39 (9) (2012a) 7673–7679.

D. Domańska, M. Wojtylak, Evaluation of the Explorative Forecast, in: A. Wakulicz-Deja (Ed.), Decision Support Systems, 237–249, 2012b.

D. Domańska, M. Wojtylak, Selection criteria of forecast pollution concentrations using collateral informations, in: R. Tadeusiewicz, A. Ligeza, W. Mitkowski, M. Szymkat (Eds.), CMS09 Computer Methods and Systems, 213–218, 2009.

J. Platt, FastMap, MetricMap, and Landmark MDS are all Nyström algorithms, in: Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics, 261–268, 2005.

I. Jolliffe, Principal Component Analysis, Springer, New York, 2002.

M. Tipping, C. M. Bishop, Probabilistic Principal Component Analysis, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 61 (1999) 611–622.

E. Bingham, H. Mannila, Random projection in dimensionality reduction: applications to image and text data, in: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 01), ACM, New York, 245–250, 2001.

D. Agrafiotis, Stochastic proximity embedding, Journal of computational chemistry (2003) 1215–1221.

Appendix A. Detailed results of dimensionality reduction experiments.

Table A.11: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Katowice station (RMSE) for e-APFM average method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	$RMSE$	0.29	0.29	0.31	14.83	14.73	15.51	18.92	18.88	20.25	30.98	30.24	30.88	9.13	8.83	9.24	**1**
	$RMSE_{vsEucl.}$	-17.1%	-14.7%	-8.8%	-15.6%	-12.4%	-7.8%	-10.0%	-11.4%	-11.1%	-7.6%	-7.7%	-6.1%	-6.3%	-4.8%	0.2%	-9.4%
IM	$RMSE$	0.29	0.3	0.31	14.79	15.38	15.97	18.46	19.91	20.91	30.11	30.18	31.44	9.1	9.08	9.4	**2**
	$RMSE_{vsEucl.}$	-17.1%	-11.8%	-8.8%	-15.8%	-8.5%	-5.1%	-12.2%	-6.6%	-8.2%	-10.2%	-7.9%	-4.4%	-6.6%	-2.2%	2.0%	-8.2%
LIM	$RMSE$	0.3	0.3	0.31	14.83	15.38	15.94	18.47	19.91	20.83	30.12	30.1	31.4	9.1	9.11	9.4	**3**
	$RMSE_{vsEucl.}$	-14.3%	-11.8%	-8.8%	-15.6%	-8.5%	-5.2%	-12.1%	-6.6%	-8.5%	-10.2%	-8.1%	-4.6%	-6.6%	-1.8%	2.0%	-8.1%
PCA	$RMSE$	0.29	0.3	0.31	14.87	15.42	16.01	18.5	19.86	20.98	30.23	30.74	32.17	9.05	9.04	9.31	**4**
	$RMSE_{vsEucl.}$	-17.1%	-11.8%	-8.8%	-15.4%	-8.3%	-4.8%	-12.0%	-6.8%	-7.9%	-9.9%	-6.2%	-2.2%	-7.1%	-2.6%	1.0%	-8.0%
FA	$RMSE$	0.3	0.3	0.31	14.98	15.53	16.02	18.42	19.89	20.71	30.57	30.61	31.77	9.22	9.25	9.59	**5**
	$RMSE_{vsEucl.}$	-14.3%	-11.8%	-8.8%	-14.7%	-7.6%	-4.8%	-12.4%	-6.7%	-9.0%	-8.9%	-6.6%	-3.4%	-5.3%	-0.3%	4.0%	-7.4%
ProbPCA	$RMSE$	0.3	0.31	0.32	14.92	15.53	16.12	18.98	20.57	21.62	30.43	30.81	32.12	9.19	9.22	9.5	**6**
	$RMSE_{vsEucl.}$	-14.3%	-8.8%	-5.9%	-15.1%	-7.6%	-4.2%	-9.7%	-3.5%	-5.1%	-9.3%	-6.0%	-2.4%	-5.6%	-0.6%	3.0%	-6.3%
Euclidean	$RMSE$	0.35	0.34	0.34	17.57	16.81	16.82	21.02	21.32	22.77	33.54	32.76	32.9	9.74	9.28	9.22	**7**
	$RMSE_{vsEucl.}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LPP	$RMSE$	0.37	0.36	0.37	17.24	17.32	17.23	25.78	25.77	25.66	35.62	35.73	35.85	12.1	12.27	12.15	**8**
	$RMSE_{vsEucl.}$	5.7%	5.9%	8.8%	-1.9%	3.0%	2.4%	22.6%	20.9%	12.7%	6.2%	9.1%	9.0%	24.2%	32.2%	31.8%	12.8%
tSNE	$RMSE$	0.37	0.37	0.37	17.38	17.45	17.45	26.01	26.34	26.41	35.85	36.01	35.93	12.25	12.45	12.47	**9**
	$RMSE_{vsEucl.}$	5.7%	8.8%	8.8%	-1.1%	3.8%	3.7%	23.7%	23.5%	16.0%	6.9%	9.9%	9.2%	25.8%	34.2%	35.2%	14.3%
Laplacian	$RMSE$	0.36	0.36	0.36	17.36	17.46	17.52	27.31	27.63	27.63	34.45	34.27	34.68	12.47	12.72	12.91	**10**
	$RMSE_{vsEucl.}$	2.9%	5.9%	5.9%	-1.2%	3.9%	4.2%	29.9%	29.6%	21.3%	2.7%	4.6%	5.4%	28.0%	37.1%	40.0%	14.7%
LTSA	$RMSE$	0.38	0.38	0.38	17.25	17.4	17.39	26.8	27.14	26.99	35.98	36.1	35.83	12.58	12.8	12.81	**11**
	$RMSE_{vsEucl.}$	8.6%	11.8%	11.8%	-1.8%	3.5%	3.4%	27.5%	27.3%	18.5%	7.3%	10.2%	8.9%	29.2%	37.9%	38.9%	16.2%
LMDS	$RMSE$	0.37	0.38	0.39	17.28	17.4	17.49	26.64	26.95	27.03	35.78	36.14	36.37	12.67	12.89	12.88	**12**
	$RMSE_{vsEucl.}$	5.7%	11.8%	14.7%	-1.7%	3.5%	4.0%	26.7%	26.4%	18.7%	6.7%	10.3%	10.5%	30.1%	38.9%	39.7%	16.4%
LLTSA	$RMSE$	0.38	0.39	0.38	17.31	17.55	17.42	26.89	27.28	27.15	36.2	36.4	36.06	12.66	12.86	12.64	**13**
	$RMSE_{vsEucl.}$	8.6%	14.7%	11.8%	-1.5%	4.4%	3.6%	27.9%	28.0%	19.2%	7.9%	11.1%	9.6%	30.0%	38.6%	37.1%	16.7%
RP	$RMSE$	0.38	0.38	0.38	17.63	17.65	17.58	25.69	26.6	26.75	36.3	36.2	36.43	13.07	13.07	12.95	**14**
	$RMSE_{vsEucl.}$	8.6%	11.8%	11.8%	0.3%	5.0%	4.5%	22.2%	24.8%	17.5%	8.2%	10.5%	10.7%	34.2%	40.8%	40.5%	16.8%
Kernel PCA	$RMSE$	0.38	0.38	0.39	17.61	17.59	17.7	27.35	27.47	27.61	36	36.35	36.54	12.93	13.18	13.12	**15**
	$RMSE_{vsEucl.}$	8.6%	11.8%	14.7%	0.2%	4.6%	5.2%	30.1%	28.8%	21.3%	7.3%	11.0%	11.1%	32.8%	42.0%	42.3%	18.1%
NPE	$RMSE$	0.38	0.39	0.39	17.47	17.63	17.72	27.41	27.6	27.64	36.18	36.5	36.49	13.05	13.29	13.29	**16**
	$RMSE_{vsEucl.}$	8.6%	14.7%	14.7%	-0.6%	4.9%	5.4%	30.4%	29.5%	21.4%	7.9%	11.4%	10.9%	34.0%	43.2%	44.1%	18.7%
SPE	$RMSE$	0.38	0.38	0.38	17.86	18.03	18.16	30.22	30.5	30.64	35.76	35.69	36.21	14.36	14.4	14.2	**17**
	$RMSE_{vsEucl.}$	8.6%	11.8%	11.8%	1.7%	7.3%	8.0%	43.8%	43.1%	34.6%	6.6%	8.9%	10.1%	47.4%	55.2%	54.0%	23.5%
LLE	$RMSE$	0.39	0.39	0.4	17.65	17.7	17.96	28.74	28.86	29.07	36.28	36.54	37.51	14.74	14.84	15.26	**18**
	$RMSE_{vsEucl.}$	11.4%	14.7%	17.6%	0.5%	5.3%	6.8%	36.7%	35.4%	27.7%	8.2%	11.5%	14.0%	51.3%	59.9%	65.5%	24.4%

Table A.12: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Katowice station (RMSE) for e-APFM α-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.26	0.35	0.39	12.55	16.92	19.17	15.5	19.36	21.43	23.35	33.17	37.56	6.96	8.83	9.72	1
	Relative RMSE [%]	-72.9%	-60.2%	-66.7%	-3.4%	-2.1%	-1.0%	-3.1%	-3.4%	-3.4%	-5.8%	-2.4%	-2.7%	-5.0%	-2.1%	-5.9%	-16.0%
PCA	RMSE	0.26	0.35	0.38	12.69	17.24	19.21	15.35	20.68	22.66	22.78	33.16	37.43	7.03	8.99	9.85	2
	Relative RMSE [%]	-72.9%	-60.2%	-67.5%	-2.3%	-0.2%	-0.8%	-4.0%	3.2%	2.1%	-8.1%	-2.4%	-3.1%	-4.1%	-0.3%	-4.6%	-15.0%
FA	RMSE	0.26	0.35	0.38	12.68	17.18	19.19	15.27	20.56	22.28	23.27	33.5	37.47	7.09	9.04	9.92	3
	Relative RMSE [%]	-72.9%	-60.2%	-67.5%	-2.4%	-0.6%	-0.9%	-4.5%	2.6%	0.4%	-6.1%	-1.4%	-3.0%	-3.3%	0.2%	-4.0%	-14.9%
IM	RMSE	0.27	0.35	0.39	12.57	17.4	19.36	15.21	20.65	22.69	23.3	33.5	37.88	7.01	9	9.91	4
	Relative RMSE [%]	-71.9%	-60.2%	-66.7%	-3.2%	0.7%	0.0%	-4.9%	3.0%	2.3%	-6.0%	-1.4%	-1.9%	-4.4%	-0.2%	-4.1%	-14.6%
LIM	RMSE	0.27	0.35	0.39	12.6	17.39	19.39	15.28	20.72	22.69	23.27	33.76	38.05	7.06	9.02	9.93	5
	Relative RMSE [%]	-71.9%	-60.2%	-66.7%	-3.0%	0.6%	0.2%	-4.4%	3.4%	2.3%	-6.1%	-0.6%	-1.5%	-3.7%	0.0%	-3.9%	-14.4%
ProbPCA	RMSE	0.26	0.35	0.39	12.68	17.43	19.37	15.33	20.87	22.62	23.39	33.42	38.09	7.11	9.13	9.97	6
	Relative RMSE [%]	-72.9%	-60.2%	-66.7%	-2.4%	0.9%	0.1%	-4.1%	4.1%	1.9%	-5.6%	-1.6%	-1.3%	-3.0%	1.2%	-3.5%	-14.2%
Laplacian	RMSE	0.26	0.33	0.35	14.23	17.91	19.1	21.15	24.83	25.87	24.22	33.35	36.96	7.44	9.14	10.05	7
	Relative RMSE [%]	-72.9%	-62.5%	-70.1%	9.5%	3.6%	-1.3%	32.3%	23.9%	16.6%	-2.3%	-1.9%	-4.3%	1.5%	1.3%	-2.7%	-8.6%
LTSA	RMSE	0.27	0.35	0.38	13.93	17.79	19.1	20.29	23.88	24.71	25.52	36.04	39.52	7.91	10.06	11.02	8
	Relative RMSE [%]	-71.9%	-60.2%	-67.5%	7.2%	3.0%	-1.3%	26.9%	19.2%	11.4%	3.0%	6.1%	2.4%	7.9%	11.5%	6.7%	-6.4%
LPP	RMSE	0.27	0.36	0.39	14.06	17.67	19.24	19.81	23.35	24.31	25.73	37.64	42.64	7.77	9.84	10.86	9
	Relative RMSE [%]	-71.9%	-59.1%	-66.7%	8.2%	2.3%	-0.6%	23.9%	16.5%	9.6%	3.8%	10.8%	10.4%	6.0%	9.1%	5.1%	-6.2%
tSNE	RMSE	0.28	0.38	0.41	14.23	18.03	19.36	19.91	23.7	24.72	25.73	36.69	41.63	7.76	9.95	10.78	10
	Relative RMSE [%]	-70.8%	-56.8%	-65.0%	9.5%	4.3%	0.0%	24.5%	18.3%	11.4%	3.8%	8.0%	7.8%	5.9%	10.3%	4.4%	-5.6%
LLTSA	RMSE	0.28	0.38	0.4	14.17	17.96	19.25	19.88	23.67	24.71	26.5	37.49	41.91	7.86	10.11	10.66	11
	Relative RMSE [%]	-70.8%	-56.8%	-65.8%	9.1%	3.9%	-0.6%	24.3%	18.1%	11.4%	6.9%	10.3%	8.5%	7.2%	12.1%	3.2%	-5.3%
NPE	RMSE	0.28	0.37	0.4	14.51	18.24	19.71	20.34	24.26	25.28	25.59	36.83	41.43	7.84	10	11.03	12
	Relative RMSE [%]	-70.8%	-58.0%	-65.8%	11.7%	5.6%	1.8%	27.2%	21.1%	13.9%	3.3%	8.4%	7.3%	7.0%	10.9%	6.8%	-4.7%
LMDS	RMSE	0.28	0.39	0.42	14.1	17.94	19.56	20.25	23.84	24.74	26.05	38.32	43.13	7.7	10.03	10.95	13
	Relative RMSE [%]	-70.8%	-55.7%	-64.1%	8.5%	3.8%	1.0%	26.6%	19.0%	11.5%	5.1%	12.8%	11.7%	5.0%	11.2%	6.0%	-4.6%
Kernel PCA	RMSE	0.29	0.38	0.41	14.71	18.55	19.78	20.4	24.06	25.16	26.57	38.07	42.65	7.78	10	10.87	14
	Relative RMSE [%]	-69.8%	-56.8%	-65.0%	13.2%	7.3%	2.2%	27.6%	20.1%	13.4%	7.2%	12.0%	10.5%	6.1%	10.9%	5.2%	-3.7%
RP	RMSE	0.28	0.36	0.39	14.74	18.4	19.31	20.33	25.18	25.97	26.19	36.4	40.83	8.04	10.44	11.29	15
	Relative RMSE [%]	-70.8%	-59.1%	-66.7%	13.5%	6.5%	-0.3%	27.1%	25.6%	17.0%	5.7%	7.1%	5.7%	9.7%	15.7%	9.3%	-3.6%
SPE	RMSE	0.28	0.36	0.38	15.08	18.58	19.6	22.17	26.38	27.55	27.03	36.57	40.5	8.12	9.99	10.48	16
	Relative RMSE [%]	-70.8%	-59.1%	-67.5%	16.1%	7.5%	1.2%	38.6%	31.6%	24.2%	9.1%	7.6%	4.9%	10.8%	10.8%	1.5%	-2.2%
LLE	RMSE	0.29	0.38	0.41	15.06	18.54	20.14	21.46	25.01	26	27.73	37.83	42.67	8.26	10.44	11.51	17
	Relative RMSE [%]	-69.8%	-56.8%	-65.0%	15.9%	7.3%	4.0%	34.2%	24.8%	17.2%	11.9%	11.3%	10.5%	12.7%	15.7%	11.4%	-1.0%
Euclidean	RMSE	0.96	0.88	1.17	12.99	17.28	19.36	15.99	20.04	22.19	24.78	33.98	38.61	7.33	9.02	10.33	18
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A.13: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Katowice station (RMSE) for e-APFM $\alpha\beta$-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
ProbPCA	RMSE	0.22	0.34	0.37	12.58	18.42	20.76	15.79	22.57	24.21	25.46	38.96	43.5	7.17	9.93	10.79	6
	Relative RMSE [%]	-26.7%	-22.7%	-21.3%	-3.2%	-2.1%	-6.3%	-2.8%	1.9%	0.1%	-0.7%	3.5%	3.1%	-1.6%	0.7%	-3.9%	-5.5%
Fractional	RMSE	0.22	0.34	0.37	12.65	18.19	20.61	16.09	22.21	24.25	26.2	39.25	44.38	7.13	9.79	10.6	1
	Relative RMSE [%]	-26.7%	-22.7%	-21.3%	-2.6%	-3.3%	-7.0%	-0.9%	0.2%	0.2%	2.2%	4.2%	5.2%	-2.2%	-0.7%	-5.6%	-5.4%
IM	RMSE	0.22	0.35	0.38	12.6	18.42	20.86	16.11	22.63	24.33	25.35	38.67	43.68	7.12	9.97	10.88	4
	Relative RMSE [%]	-26.7%	-20.5%	-19.1%	-3.0%	-2.1%	-5.8%	-0.8%	2.1%	0.6%	-1.1%	2.7%	3.6%	-2.3%	1.1%	-3.1%	-5.0%
LIM	RMSE	0.22	0.34	0.37	12.66	18.66	21.01	16.15	22.79	24.42	25.38	38.69	43.45	7.16	10.08	10.94	5
	Relative RMSE [%]	-26.7%	-22.7%	-21.3%	-2.5%	-0.8%	-5.1%	-0.6%	2.8%	1.0%	-1.0%	2.8%	3.0%	-1.8%	2.2%	-2.6%	-4.9%
PCA	RMSE	0.22	0.34	0.38	12.74	18.62	20.69	15.96	22.61	24.22	25.72	38.96	43.74	7.18	9.99	10.99	2
	Relative RMSE [%]	-26.7%	-22.7%	-19.1%	-1.9%	-1.0%	-6.6%	-1.7%	2.0%	0.1%	0.4%	3.5%	3.7%	-1.5%	1.3%	-2.1%	-4.8%
FA	RMSE	0.22	0.34	0.38	12.59	18.61	20.77	15.85	22.44	24.26	25.42	38.78	42.93	7.33	10.32	11.25	3
	Relative RMSE [%]	-26.7%	-22.7%	-19.1%	-3.1%	-1.1%	-6.2%	-2.4%	1.3%	0.3%	-0.8%	3.0%	1.8%	0.5%	4.7%	0.2%	-4.7%
Laplacian	RMSE	0.23	0.33	0.36	14.72	19.69	21.56	19.72	25.89	26.72	25.49	37.22	41.36	6.92	8.95	9.92	7
	Relative RMSE [%]	-23.3%	-25.0%	-23.4%	13.3%	4.7%	-2.7%	21.4%	16.8%	10.5%	-0.5%	-1.1%	-1.9%	-5.1%	-9.2%	-11.7%	-2.5%
Euclidean	RMSE	0.3	0.44	0.47	12.99	18.81	22.15	16.24	22.16	24.19	25.63	37.65	42.18	7.29	9.86	11.23	18
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LPP	RMSE	0.24	0.36	0.39	14.25	19.53	21.5	19.13	24.64	26.31	26.79	39.2	43.7	7.29	9.93	10.98	9
	Relative RMSE [%]	-20.0%	-18.2%	-17.0%	9.7%	3.8%	-2.9%	17.8%	11.2%	8.8%	4.5%	4.1%	3.6%	0.0%	0.7%	-2.2%	0.3%
tSNE	RMSE	0.25	0.36	0.4	14.51	19.71	21.63	19.4	24.55	25.97	26.46	39.29	44.04	7.27	9.92	10.86	10
	Relative RMSE [%]	-16.7%	-18.2%	-14.9%	11.7%	4.8%	-2.3%	19.5%	10.8%	7.4%	3.2%	4.4%	4.4%	-0.3%	0.6%	-3.3%	0.7%
LTSA	RMSE	0.24	0.36	0.39	14.25	19.49	21.3	19.47	24.56	25.79	27.1	39.82	43.47	7.54	10.25	11.23	8
	Relative RMSE [%]	-20.0%	-18.2%	-17.0%	9.7%	3.6%	-3.8%	19.9%	10.8%	6.6%	5.7%	5.8%	3.1%	3.4%	4.0%	0.0%	0.9%
LMDS	RMSE	0.24	0.37	0.4	14.32	19.44	21.63	20.25	25.38	26.49	26.5	39	44.36	7.25	10.13	10.96	13
	Relative RMSE [%]	-20.0%	-15.9%	-14.9%	10.2%	3.3%	-2.3%	24.7%	14.5%	9.5%	3.4%	3.6%	5.2%	-0.5%	2.7%	-2.4%	1.4%
LLTSA	RMSE	0.25	0.36	0.39	14.36	19.53	21.33	20.29	25.12	26.47	27.4	40.17	44.25	7.44	10.16	11.01	11
	Relative RMSE [%]	-16.7%	-18.2%	-17.0%	10.5%	3.8%	-3.7%	24.9%	13.4%	9.4%	6.9%	6.7%	4.9%	2.1%	3.0%	-2.0%	1.9%
RP	RMSE	0.24	0.35	0.39	14.61	19.79	21.76	19.6	26.26	27.46	26.53	39.35	43.86	7.44	10.46	11.26	15
	Relative RMSE [%]	-20.0%	-20.5%	-17.0%	12.5%	5.2%	-1.8%	20.7%	18.5%	13.5%	3.5%	4.5%	4.0%	2.1%	6.1%	0.3%	2.1%
Kernel PCA	RMSE	0.25	0.37	0.41	14.89	20.12	22.38	20.8	25.53	26.91	27.63	40.68	45.81	7.44	10.22	11.15	14
	Relative RMSE [%]	-16.7%	-15.9%	-12.8%	14.6%	7.0%	1.0%	28.1%	15.2%	11.2%	7.8%	8.0%	8.6%	2.1%	3.7%	-0.7%	4.1%
NPE	RMSE	0.25	0.37	0.41	14.56	20.15	22.58	20.93	26.33	27.62	27.71	40.76	45.49	7.45	10.32	11.37	12
	Relative RMSE [%]	-16.7%	-15.9%	-12.8%	12.1%	7.1%	1.9%	28.9%	18.8%	14.2%	8.1%	8.3%	7.8%	2.2%	4.7%	1.2%	4.7%
SPE	RMSE	0.25	0.35	0.38	15.04	20.57	23.14	22.81	28.16	29.23	27.06	38.24	42.2	7.52	10.09	10.87	16
	Relative RMSE [%]	-16.7%	-20.5%	-19.1%	15.8%	9.4%	4.5%	40.5%	27.1%	20.8%	5.6%	1.6%	0.0%	3.2%	2.3%	-3.2%	4.7%
LLE	RMSE	0.26	0.37	0.41	15.26	20.92	23.71	22.16	26.59	27.54	28.84	41.5	46.33	7.92	10.61	11.92	17
	Relative RMSE [%]	-13.3%	-15.9%	-12.8%	17.5%	11.2%	7.0%	36.5%	20.0%	13.8%	12.5%	10.2%	9.8%	8.6%	7.6%	6.1%	7.9%

Table A.14: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM average method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.36	0.35	0.36	11.81	11.74	11.98	18.62	18.47	19.9	43.01	45.53	47.38	12.19	12.05	12.22	1
	Relative RMSE [%]	-7.7%	-10.3%	-7.7%	-6.0%	-6.3%	-7.5%	-9.9%	-9.4%	-8.8%	-14.5%	-11.5%	-8.9%	-10.0%	-8.7%	-6.9%	-8.9%
LIM	RMSE	0.36	0.36	0.36	11.94	12.03	12.03	17.78	19.24	20.43	44.7	47.26	48.27	12.13	12.07	12.25	2
	Relative RMSE [%]	-7.7%	-7.7%	-7.7%	-5.0%	-4.0%	-7.1%	-14.0%	-5.6%	-6.4%	-11.2%	-8.2%	-7.2%	-10.4%	-8.6%	-6.6%	-7.8%
IM	RMSE	0.36	0.36	0.37	11.93	12	12.04	17.77	19.13	20.37	44.8	47.24	48.16	12.16	12.07	12.24	3
	Relative RMSE [%]	-7.7%	-7.7%	-5.1%	-5.1%	-4.2%	-7.0%	-14.0%	-6.1%	-6.7%	-11.0%	-8.2%	-7.4%	-10.2%	-8.6%	-6.7%	-7.7%
FA	RMSE	0.37	0.37	0.37	12.22	12.27	12.26	18	19.41	20.68	45.54	47.93	49.16	12.43	12.28	12.4	4
	Relative RMSE [%]	-5.1%	-5.1%	-5.1%	-2.8%	-2.1%	-5.3%	-12.9%	-4.8%	-5.3%	-9.5%	-6.9%	-5.5%	-8.2%	-7.0%	-5.5%	-6.1%
PCA	RMSE	0.36	0.37	0.37	12.11	12.3	12.28	18.02	19.35	20.48	48.67	49.32	49.69	12.38	12.38	12.52	5
	Relative RMSE [%]	-7.7%	-5.1%	-5.1%	-3.7%	-1.8%	-5.2%	-12.8%	-5.1%	-6.2%	-3.3%	-4.2%	-4.5%	-8.6%	-6.2%	-4.6%	-5.6%
ProbPCA	RMSE	0.38	0.38	0.38	12.47	12.56	12.53	18.76	20.07	21.3	50.11	50.36	50.46	12.91	12.79	12.83	6
	Relative RMSE [%]	-2.6%	-2.6%	-2.6%	-0.8%	0.2%	-3.2%	-9.2%	-1.5%	-2.4%	-0.4%	-2.1%	-3.0%	-4.7%	-3.1%	-2.2%	-2.7%
Euclidean	RMSE	0.39	0.39	0.39	12.57	12.53	12.95	20.67	20.38	21.83	50.32	51.46	52.02	13.54	13.2	13.12	7
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Laplacian	RMSE	0.48	0.48	0.48	15.94	16.04	16.28	29	28.47	28.28	56.71	58.5	58.13	18.02	18.3	18.34	8
	Relative RMSE [%]	23.1%	23.1%	23.1%	26.8%	28.0%	25.7%	40.3%	39.7%	29.5%	12.7%	13.7%	11.7%	33.1%	38.6%	39.8%	27.3%
RP	RMSE	0.5	0.49	0.49	15.99	16.45	16.74	26.71	27.24	26.84	56.62	57.74	58.36	18.59	18.54	18.6	9
	Relative RMSE [%]	28.2%	25.6%	25.6%	27.2%	31.3%	29.3%	29.2%	33.7%	23.0%	12.5%	12.2%	12.2%	37.3%	40.5%	41.8%	27.3%
tSNE	RMSE	0.49	0.49	0.48	15.72	15.99	16.15	27.05	27.04	26.75	61.47	60.73	59.92	18.54	18.71	18.62	10
	Relative RMSE [%]	25.6%	25.6%	23.1%	25.1%	27.6%	24.7%	30.9%	32.7%	22.5%	22.2%	18.0%	15.2%	36.9%	41.7%	41.9%	27.6%
LPP	RMSE	0.5	0.49	0.49	15.97	16.06	16.13	27.81	27.63	27.47	57.84	59.8	59.82	18.82	18.9	18.81	11
	Relative RMSE [%]	28.2%	25.6%	25.6%	27.0%	28.2%	24.6%	34.5%	35.6%	25.8%	14.9%	16.2%	15.0%	39.0%	43.2%	43.4%	28.5%
NPE	RMSE	0.5	0.49	0.49	16.03	16.36	16.65	27.71	27.73	27.45	57.8	60.2	60.49	19.19	19.4	19.48	12
	Relative RMSE [%]	28.2%	25.6%	25.6%	27.5%	30.6%	28.6%	34.1%	36.1%	25.7%	14.9%	17.0%	16.3%	41.7%	47.0%	48.5%	29.8%
Kernel PCA	RMSE	0.5	0.49	0.49	16.11	16.37	16.62	27.93	27.9	27.61	62.45	61.88	61.17	19.08	19.33	19.35	13
	Relative RMSE [%]	28.2%	25.6%	25.6%	28.2%	30.6%	28.3%	35.1%	36.9%	26.5%	24.1%	20.2%	17.6%	40.9%	46.4%	47.5%	30.8%
LLTSA	RMSE	0.5	0.5	0.49	16.09	16.51	16.81	27.96	28.04	27.7	58.61	61.15	61.06	19.25	19.65	19.64	14
	Relative RMSE [%]	28.2%	28.2%	25.6%	28.0%	31.8%	29.8%	35.3%	37.6%	26.9%	16.5%	18.8%	17.4%	42.2%	48.9%	49.7%	31.0%
LTSA	RMSE	0.51	0.51	0.49	16.22	16.48	16.52	28.16	27.93	27.6	59.52	61.89	61.34	19.26	19.57	19.42	15
	Relative RMSE [%]	30.8%	30.8%	25.6%	29.0%	31.5%	27.6%	36.2%	37.0%	26.4%	18.3%	20.3%	17.9%	42.2%	48.3%	48.0%	31.3%
LLE	RMSE	0.51	0.51	0.51	15.91	16.21	16.69	27.74	27.68	27.4	59.17	61.93	62.39	19.55	20.14	20.62	16
	Relative RMSE [%]	30.8%	30.8%	30.8%	26.6%	29.4%	28.9%	34.2%	35.8%	25.5%	17.6%	20.3%	19.9%	44.4%	52.6%	57.2%	32.3%
LMDS	RMSE	0.51	0.51	0.5	16.33	16.7	16.98	28.31	28.31	28	59.6	62.06	61.61	19.75	20.09	20.02	17
	Relative RMSE [%]	30.8%	30.8%	28.2%	29.9%	33.3%	31.1%	37.0%	38.9%	28.3%	18.4%	20.6%	18.4%	45.9%	52.2%	52.6%	33.1%
SPE	RMSE	0.51	0.5	0.5	16.88	17.08	17.64	29.46	29.39	29.15	60.52	62.56	62.91	19.52	19.82	20.24	18
	Relative RMSE [%]	30.8%	28.2%	28.2%	34.3%	36.3%	36.2%	42.5%	44.2%	33.5%	20.3%	21.6%	20.9%	44.2%	50.2%	54.3%	35.0%

Table A.15: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM α-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.31	0.42	0.44	9.73	12.62	13.65	14.94	19.3	21.29	34.53	50.64	54.69	10.14	13.38	14.3	1
	Relative RMSE [%]	-46.6%	-46.2%	-47.0%	-0.2%	-0.9%	-4.3%	-4.1%	-4.5%	-6.1%	-11.7%	-7.5%	-8.1%	-1.7%	-3.7%	-6.0%	-13.2%
IM	RMSE	0.31	0.43	0.46	9.73	12.69	13.56	14.8	20.51	22.61	36.49	54.52	59.17	10.22	13.4	14.2	2
	Relative RMSE [%]	-46.6%	-44.9%	-44.6%	-0.2%	-0.3%	-4.9%	-5.0%	1.4%	-0.3%	-6.7%	-0.5%	-0.6%	-1.0%	-3.6%	-6.7%	-11.0%
LIM	RMSE	0.31	0.43	0.45	9.72	12.79	13.54	14.87	20.67	22.66	36.7	54.75	59.4	10.2	13.38	14.23	3
	Relative RMSE [%]	-46.6%	-44.9%	-45.8%	-0.3%	0.5%	-5.0%	-4.6%	2.2%	0.0%	-6.2%	0.0%	-0.2%	-1.2%	-3.7%	-6.5%	-10.8%
FA	RMSE	0.31	0.43	0.45	9.87	12.91	13.8	15	20.76	22.51	36.5	55.33	60.73	10.33	13.68	14.61	4
	Relative RMSE [%]	-46.6%	-44.9%	-45.8%	1.2%	1.4%	-3.2%	-3.7%	2.7%	-0.7%	-6.7%	1.0%	2.0%	0.1%	-1.6%	-4.0%	-9.9%
PCA	RMSE	0.31	0.43	0.45	9.77	12.82	13.61	14.88	20.73	22.62	38.97	58.26	63.37	10.36	13.56	14.57	5
	Relative RMSE [%]	-46.6%	-44.9%	-45.8%	0.2%	0.7%	-4.6%	-4.5%	2.5%	-0.2%	-0.4%	6.4%	6.4%	0.4%	-2.4%	-4.3%	-9.1%
ProbPCA	RMSE	0.31	0.43	0.45	9.77	12.77	13.78	14.95	20.7	22.61	39.39	57.92	62.03	10.47	13.66	14.63	6
	Relative RMSE [%]	-46.6%	-44.9%	-45.8%	0.2%	0.3%	-3.4%	-4.0%	2.4%	-0.3%	0.7%	5.8%	4.2%	1.5%	-1.7%	-3.9%	-9.0%
Laplacian	RMSE	0.32	0.46	0.47	10.92	13.45	14.44	20.11	24.3	25.75	38.95	57.67	59.73	10.55	14.12	14.86	7
	Relative RMSE [%]	-44.8%	-41.0%	-43.4%	12.0%	5.7%	1.3%	29.1%	20.2%	13.6%	-0.4%	5.3%	0.3%	2.2%	1.6%	-2.4%	-2.7%
LPP	RMSE	0.33	0.45	0.48	10.93	13.43	14.1	19.03	23.22	24.12	39.92	56.28	58.99	11.35	15.22	15.92	8
	Relative RMSE [%]	-43.1%	-42.3%	-42.2%	12.1%	5.5%	-1.1%	22.1%	14.8%	6.4%	2.0%	2.8%	-0.9%	10.0%	9.5%	4.6%	-2.7%
LTSA	RMSE	0.32	0.45	0.46	11.23	13.84	14.48	19.34	23.35	24.24	40.36	56.94	59.08	11.57	15.34	16.17	9
	Relative RMSE [%]	-44.8%	-42.3%	-44.6%	15.2%	8.7%	1.5%	24.1%	15.5%	6.9%	3.2%	4.0%	-0.8%	12.1%	10.4%	6.2%	-1.6%
tSNE	RMSE	0.33	0.46	0.47	10.84	13.45	14.35	18.66	23.1	24.03	43.12	59.98	62.95	11.46	15.28	16.18	10
	Relative RMSE [%]	-43.1%	-41.0%	-43.4%	11.2%	5.7%	0.6%	19.8%	14.2%	6.0%	10.2%	9.5%	5.7%	11.0%	9.9%	6.3%	-1.2%
NPE	RMSE	0.34	0.46	0.48	10.96	13.62	14.63	19.28	23.45	24.28	40.19	56.77	59.82	11.62	15.44	16.53	11
	Relative RMSE [%]	-41.4%	-41.0%	-42.2%	12.4%	7.0%	2.6%	23.7%	16.0%	7.1%	2.7%	3.7%	0.5%	12.6%	11.1%	8.6%	-1.1%
LLTSA	RMSE	0.34	0.47	0.48	11.03	13.78	14.51	19.48	23.73	24.43	40.25	57.15	59.77	11.56	15.73	16.38	12
	Relative RMSE [%]	-41.4%	-39.7%	-42.2%	13.1%	8.2%	1.8%	25.0%	17.4%	7.8%	2.9%	4.3%	0.4%	12.0%	13.2%	7.6%	-0.6%
RP	RMSE	0.34	0.47	0.48	11.18	14.06	15	19.42	24.11	24.68	40.37	56.83	58.61	11.65	15.57	16.5	13
	Relative RMSE [%]	-41.4%	-39.7%	-42.2%	14.7%	10.4%	5.2%	24.6%	19.2%	8.9%	3.2%	3.8%	-1.6%	12.9%	12.0%	8.4%	-0.1%
LMDS	RMSE	0.33	0.46	0.48	11.04	13.84	14.73	19.71	23.87	24.59	40.9	58.14	60.33	11.64	15.66	16.71	14
	Relative RMSE [%]	-43.1%	-41.0%	-42.2%	13.2%	8.7%	3.3%	26.5%	18.1%	8.5%	4.6%	6.2%	1.3%	12.8%	12.7%	9.8%	-0.1%
Euclidean	RMSE	0.58	0.78	0.83	9.75	12.73	14.26	15.58	20.22	22.67	39.12	54.77	59.54	10.32	13.9	15.22	15
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Kernel PCA	RMSE	0.34	0.46	0.48	11	13.73	14.65	19.45	23.43	24.24	42.46	60.88	64.21	11.49	15.53	16.51	16
	Relative RMSE [%]	-41.4%	-41.0%	-42.2%	12.8%	7.9%	2.7%	24.8%	15.9%	6.9%	8.5%	11.2%	7.8%	11.3%	11.7%	8.5%	0.4%
SPE	RMSE	0.34	0.46	0.48	11.35	14.03	15.22	20.47	24.64	25.82	41.32	58.67	61.02	11.49	15.64	16.37	17
	Relative RMSE [%]	-41.4%	-41.0%	-42.2%	16.4%	10.2%	6.7%	31.4%	21.9%	13.9%	5.6%	7.1%	2.5%	11.3%	12.5%	7.6%	1.5%
LLE	RMSE	0.36	0.49	0.5	11.03	13.75	15.04	20.05	24.08	25.08	40.59	58.52	60.84	11.87	16.04	17.35	18
	Relative RMSE [%]	-37.9%	-37.2%	-39.8%	13.1%	8.0%	5.5%	28.7%	19.1%	10.6%	3.8%	6.8%	2.2%	15.0%	15.4%	14.0%	1.8%

Table A.16: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM $\alpha\beta$-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.28	0.42	0.46	9.62	13.15	14.31	15.48	21.73	23.61	35.85	54.33	58.85	10.15	14.2	15.28	1
	Relative RMSE [%]	-42.9%	-38.2%	-33.3%	-4.6%	-8.9%	-9.5%	-3.4%	-2.5%	-3.2%	-8.7%	-6.2%	-7.1%	-2.9%	-8.3%	-8.3%	-12.5%
IM	RMSE	0.29	0.44	0.47	9.63	13.54	14.56	15.61	22.22	24.12	37.79	57.54	63.38	10.37	14.88	16.04	2
	Relative RMSE [%]	-40.8%	-35.3%	-31.9%	-4.5%	-6.2%	-7.9%	-2.6%	-0.3%	-1.1%	-3.8%	-0.7%	0.1%	-0.8%	-3.9%	-3.7%	-9.6%
LIM	RMSE	0.29	0.44	0.46	9.67	13.51	14.56	15.53	22.55	24.33	38.09	57.8	62.07	10.55	15.2	16.27	3
	Relative RMSE [%]	-40.8%	-35.3%	-33.3%	-4.1%	-6.4%	-7.9%	-3.1%	1.2%	-0.2%	-3.0%	-0.2%	-2.0%	1.0%	-1.8%	-2.3%	-9.2%
FA	RMSE	0.29	0.44	0.47	9.78	13.72	14.81	15.98	22.61	24.4	38	58.12	63.37	10.45	15	16.09	4
	Relative RMSE [%]	-40.8%	-35.3%	-31.9%	-3.0%	-4.9%	-6.3%	-0.3%	1.4%	0.0%	-3.2%	0.3%	0.1%	0.0%	-3.1%	-3.4%	-8.7%
ProbPCA	RMSE	0.29	0.44	0.47	9.68	13.4	14.56	15.49	22.19	24.47	39.98	60.93	66.48	10.53	14.89	16.17	5
	Relative RMSE [%]	-40.8%	-35.3%	-31.9%	-4.0%	-7.1%	-7.9%	-3.4%	-0.4%	0.3%	1.8%	5.2%	5.0%	0.8%	-3.8%	-2.9%	-8.3%
PCA	RMSE	0.29	0.43	0.46	9.76	13.66	14.65	15.86	22.59	24.68	41.02	62.26	67.73	10.3	14.65	16.04	6
	Relative RMSE [%]	-40.8%	-36.8%	-33.3%	-3.2%	-5.3%	-7.3%	-1.1%	1.3%	1.2%	4.5%	7.5%	7.0%	-1.4%	-5.4%	-3.7%	-7.8%
LPP	RMSE	0.31	0.45	0.48	10.11	13.63	14.71	19.96	25.33	27.14	38.22	57.52	61.08	10.69	14.98	16.1	7
	Relative RMSE [%]	-36.7%	-33.8%	-30.4%	0.3%	-5.5%	-7.0%	24.5%	13.6%	11.3%	-2.7%	-0.7%	-3.5%	2.3%	-3.2%	-3.4%	-5.0%
Laplacian	RMSE	0.3	0.45	0.47	10.64	13.91	14.96	21.07	26.57	27.82	37.82	59.27	61.03	10.13	13.97	14.89	8
	Relative RMSE [%]	-38.8%	-33.8%	-31.9%	5.6%	-3.6%	-5.4%	31.4%	19.2%	14.1%	-3.7%	2.3%	-3.6%	-3.1%	-9.8%	-10.6%	-4.8%
RP	RMSE	0.32	0.46	0.49	10.16	13.84	15.17	19.47	25.35	27.04	37.43	58.07	60.06	10.63	15.16	16.39	9
	Relative RMSE [%]	-34.7%	-32.4%	-29.0%	0.8%	-4.1%	-4.0%	21.5%	13.7%	10.9%	-4.7%	0.2%	-5.1%	1.7%	-2.1%	-1.6%	-4.6%
NPE	RMSE	0.31	0.46	0.48	10.32	13.79	15.15	19.12	25.49	27.63	38.3	58.88	61.9	10.66	15.04	16.23	10
	Relative RMSE [%]	-36.7%	-32.4%	-30.4%	2.4%	-4.4%	-4.2%	19.3%	14.4%	13.3%	-2.5%	1.6%	-2.2%	2.0%	-2.8%	-2.6%	-4.4%
LTSA	RMSE	0.31	0.45	0.47	10.23	13.75	14.84	19.88	25.68	27.7	38.41	57.97	61.06	10.91	15.11	16.06	11
	Relative RMSE [%]	-36.7%	-33.8%	-31.9%	1.5%	-4.7%	-6.1%	24.0%	15.2%	13.6%	-2.2%	0.1%	-3.6%	4.4%	-2.4%	-3.6%	-4.4%
LLTSA	RMSE	0.33	0.47	0.49	10.18	13.79	14.95	18.78	25.21	27.07	38.37	59.21	62.49	10.88	15.52	16.56	12
	Relative RMSE [%]	-32.7%	-30.9%	-29.0%	1.0%	-4.4%	-5.4%	17.2%	13.1%	11.0%	-2.3%	2.2%	-1.3%	4.1%	0.3%	-0.6%	-3.9%
LMDS	RMSE	0.32	0.45	0.48	10.39	14.07	15.33	19.39	25.66	27.63	38.46	60.11	61.62	10.83	15.19	16.31	13
	Relative RMSE [%]	-34.7%	-33.8%	-30.4%	3.1%	-2.5%	-3.0%	21.0%	15.1%	13.3%	-2.1%	3.8%	-2.7%	3.6%	-1.9%	-2.1%	-3.6%
Kernel PCA	RMSE	0.31	0.45	0.48	10.11	13.65	14.88	19.23	25.33	27.15	41.46	62.34	66.65	10.74	15.17	16.3	14
	Relative RMSE [%]	-36.7%	-33.8%	-30.4%	0.3%	-5.4%	-5.9%	20.0%	13.6%	11.3%	5.6%	7.6%	5.3%	2.8%	-2.0%	-2.2%	-3.3%
LLE	RMSE	0.32	0.46	0.49	10.32	13.89	15.34	19.78	25.44	28.37	38.37	59.11	62.19	10.86	15.63	17.18	15
	Relative RMSE [%]	-34.7%	-32.4%	-29.0%	2.4%	-3.7%	-3.0%	23.4%	14.1%	16.3%	-2.3%	2.0%	-1.8%	3.9%	1.0%	3.1%	-2.7%
tSNE	RMSE	0.32	0.46	0.49	10.03	13.64	14.92	19.63	25.41	27.25	42.07	63.5	67.24	10.77	15.12	16.38	16
	Relative RMSE [%]	-34.7%	-32.4%	-29.0%	-0.5%	-5.5%	-5.6%	22.5%	14.0%	11.7%	7.1%	9.6%	6.2%	3.1%	-2.3%	-1.7%	-2.5%
SPE	RMSE	0.32	0.45	0.47	10.56	14.09	15.49	20.83	27.03	29.8	39.1	60.68	62.71	10.55	14.98	15.95	17
	Relative RMSE [%]	-34.7%	-33.8%	-31.9%	4.8%	-2.4%	-2.0%	29.9%	21.3%	22.2%	-0.4%	4.7%	-1.0%	1.0%	-3.2%	-4.3%	-2.0%
Euclidean	RMSE	0.49	0.68	0.69	10.08	14.43	15.81	16.03	22.29	24.39	39.27	57.93	63.32	10.45	15.48	16.66	18
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A.17: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Cieszyn station (RMSE) for e-APFM average method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.28	0.27	0.27	7.72	7.33	7.73	22.36	22.16	22.96	21.25	20.75	20.34	7.5	7.04	7.4	1
	Relative RMSE [%]	-20.0%	-22.9%	-20.6%	-21.1%	-29.1%	-26.3%	5.4%	8.2%	8.8%	-10.5%	-13.0%	-11.4%	-23.3%	-22.7%	-16.6%	-14.3%
LIM	RMSE	0.3	0.29	0.29	8.2	8.21	8.52	19.87	20.84	22.26	22.3	22.58	22.86	7.83	7.84	8.28	2
	Relative RMSE [%]	-14.3%	-17.1%	-14.7%	-16.2%	-20.6%	-18.8%	-6.3%	1.7%	5.4%	-6.1%	-5.4%	-0.5%	-19.9%	-13.9%	-6.7%	-10.2%
IM	RMSE	0.3	0.29	0.29	8.15	8.17	8.55	20.08	21	22.54	22.41	22.53	22.88	7.8	7.83	8.26	3
	Relative RMSE [%]	-14.3%	-17.1%	-14.7%	-16.7%	-21.0%	-18.5%	-5.3%	2.5%	6.8%	-5.6%	-5.6%	-0.4%	-20.2%	-14.1%	-6.9%	-10.1%
PCA	RMSE	0.32	0.3	0.3	8.41	8.45	8.6	20.69	21.37	22.81	24.53	23.99	25.56	7.82	7.8	8.27	4
	Relative RMSE [%]	-8.6%	-14.3%	-11.8%	-14.0%	-18.3%	-18.0%	-2.5%	4.3%	8.1%	3.3%	0.5%	11.3%	-20.0%	-14.4%	-6.8%	-6.7%
FA	RMSE	0.32	0.32	0.31	8.42	8.58	8.82	19.79	20.8	22.23	24.05	24.5	24.18	8.46	8.27	8.57	5
	Relative RMSE [%]	-8.6%	-8.6%	-8.8%	-13.9%	-17.0%	-15.9%	-6.7%	1.5%	5.3%	1.3%	2.7%	5.3%	-13.5%	-9.2%	-3.4%	-6.0%
ProbPCA	RMSE	0.32	0.31	0.31	8.5	8.57	8.82	20.55	21.07	22.89	24.7	24.56	26.09	8.11	8.14	8.65	6
	Relative RMSE [%]	-8.6%	-11.4%	-8.8%	-13.1%	-17.1%	-15.9%	-3.1%	2.8%	8.4%	4.0%	2.9%	13.6%	-17.1%	-10.6%	-2.5%	-5.1%
Euclidean	RMSE	0.35	0.35	0.34	9.78	10.34	10.49	21.21	20.49	21.11	23.74	23.86	22.97	9.78	9.11	8.87	7
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Laplacian	RMSE	0.35	0.37	0.37	10.32	10.51	10.61	26.92	28.59	28.04	27.86	28.69	27.96	12.41	12.41	12.47	8
	Relative RMSE [%]	0.0%	5.7%	8.8%	5.5%	1.6%	1.1%	26.9%	39.5%	32.8%	17.4%	20.2%	21.7%	26.9%	36.2%	40.6%	19.0%
LLTSA	RMSE	0.33	0.34	0.34	11.13	11.03	11.18	30.63	31.94	31.15	27.17	27.99	27.35	11.93	11.98	12.05	9
	Relative RMSE [%]	-5.7%	-2.9%	0.0%	13.8%	6.7%	6.6%	44.4%	55.9%	47.6%	14.4%	17.3%	19.1%	22.0%	31.5%	35.9%	20.4%
LPP	RMSE	0.35	0.35	0.35	10.78	10.7	10.84	30.19	31.52	30.96	27.42	28.17	27.38	12.12	12.09	12.14	10
	Relative RMSE [%]	0.0%	0.0%	2.9%	10.2%	3.5%	3.3%	42.3%	53.8%	46.7%	15.5%	18.1%	19.2%	23.9%	32.7%	36.9%	20.6%
LTSA	RMSE	0.36	0.36	0.36	10.76	10.65	10.72	29.84	30.93	30.21	27.63	28.52	27.65	12.43	12.36	12.38	11
	Relative RMSE [%]	2.9%	2.9%	5.9%	10.0%	3.0%	2.2%	40.7%	51.0%	43.1%	16.4%	19.5%	20.4%	27.1%	35.7%	39.6%	21.3%
LMDS	RMSE	0.34	0.35	0.35	11.05	10.99	11.08	31.16	32.56	31.54	27.49	28.25	27.55	12.25	12.22	12.3	12
	Relative RMSE [%]	-2.9%	0.0%	2.9%	13.0%	6.3%	5.6%	46.9%	58.9%	49.4%	15.8%	18.4%	19.9%	25.3%	34.1%	38.7%	22.2%
NPE	RMSE	0.34	0.34	0.34	11.31	11.11	11.31	32.29	33.26	32.1	27.37	28.11	27.52	12.23	12.15	12.23	13
	Relative RMSE [%]	-2.9%	-2.9%	0.0%	15.6%	7.4%	7.8%	52.2%	62.3%	52.1%	15.3%	17.8%	19.8%	25.1%	33.4%	37.9%	22.7%
tSNE	RMSE	0.36	0.36	0.36	10.94	10.92	10.87	29.25	30.99	30.26	31.26	31.44	31.48	12.14	12.16	12.15	14
	Relative RMSE [%]	2.9%	2.9%	5.9%	11.9%	5.6%	3.6%	37.9%	51.2%	43.3%	31.7%	31.8%	37.0%	24.1%	33.5%	37.0%	24.0%
LLE	RMSE	0.34	0.34	0.34	11.31	11.17	11.34	33.88	34.78	33.29	27.27	28.18	27.43	12.21	12.19	12.2	15
	Relative RMSE [%]	-2.9%	-2.9%	0.0%	15.6%	8.0%	8.1%	59.7%	69.7%	57.7%	14.9%	18.1%	19.4%	24.8%	33.8%	37.5%	24.1%
Kernel PCA	RMSE	0.36	0.36	0.36	11.23	11.22	11.31	31.15	32.77	31.63	31.53	31.64	31.72	12.26	12.27	12.28	16
	Relative RMSE [%]	2.9%	2.9%	5.9%	14.8%	8.5%	7.8%	46.9%	59.9%	49.8%	32.8%	32.6%	38.1%	25.4%	34.7%	38.4%	26.8%
RP	RMSE	0.35	0.36	0.35	10.96	11.45	11.61	33.36	34.4	33.27	28.06	30.07	29.02	12.95	12.59	12.46	17
	Relative RMSE [%]	0.0%	2.9%	2.9%	12.1%	10.7%	10.7%	57.3%	67.9%	57.6%	18.2%	26.0%	26.3%	32.4%	38.2%	40.5%	26.9%
SPE	RMSE	0.34	0.35	0.34	11.7	11.7	11.79	33.3	34.56	33.25	29.19	29.91	28.66	12.87	12.89	12.9	18
	Relative RMSE [%]	-2.9%	0.0%	0.0%	19.6%	13.2%	12.4%	57.0%	68.7%	57.5%	23.0%	25.4%	24.8%	31.6%	41.5%	45.4%	27.8%

Table A.18: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Cieszyn station (RMSE) for e-APFM α-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
LLTSA	RMSE	0.15	0.26	0.28	4.87	8.36	9.39	14.09	23.78	26.21	10.09	21.91	23.5	4.55	8.72	9.73	1
	Relative RMSE [%]	-54.5%	-21.2%	-56.3%	8.7%	-11.7%	-16.0%	17.8%	3.6%	-7.6%	-12.5%	-2.6%	-17.5%	-1.7%	-14.5%	-17.3%	-13.6%
NPE	RMSE	0.15	0.25	0.28	5.08	8.39	9.34	14.37	23.85	26.37	10.59	22.35	24.77	4.54	9.31	10.36	2
	Relative RMSE [%]	-54.5%	-24.2%	-56.3%	13.4%	-11.4%	-16.5%	20.2%	3.9%	-7.0%	-8.2%	-0.7%	-13.0%	-1.9%	-8.7%	-11.9%	-11.8%
LMDS	RMSE	0.15	0.26	0.28	4.95	8.34	9.32	15.2	23.97	27.09	10.67	21.91	23.59	4.55	8.87	10.43	3
	Relative RMSE [%]	-54.5%	-21.2%	-56.3%	10.5%	-11.9%	-16.6%	27.1%	4.4%	-4.5%	-7.5%	-2.6%	-17.1%	-1.7%	-13.0%	-11.3%	-11.8%
LPP	RMSE	0.16	0.26	0.29	5	8.42	9.4	14.35	24.27	27.78	10.71	21.92	23.72	4.53	8.7	10.17	4
	Relative RMSE [%]	-51.5%	-21.2%	-54.7%	11.6%	-11.1%	-15.9%	20.0%	5.8%	-2.0%	-7.1%	-2.6%	-16.7%	-2.2%	-14.7%	-13.5%	-11.7%
SPE	RMSE	0.14	0.25	0.27	5.14	8.5	9.53	15.01	22.75	25.64	10.62	22.27	24.16	4.91	9.48	10.72	5
	Relative RMSE [%]	-57.6%	-24.2%	-57.8%	14.7%	-10.2%	-14.8%	25.5%	-0.9%	-9.6%	-7.9%	-1.0%	-15.1%	6.0%	-7.1%	-8.8%	-11.3%
LLE	RMSE	0.15	0.26	0.28	5.27	8.59	9.52	14.52	24.02	26.78	10.66	21.99	23.82	4.69	9.35	10.82	6
	Relative RMSE [%]	-54.5%	-21.2%	-56.3%	17.6%	-9.3%	-14.8%	21.4%	4.7%	-5.6%	-7.5%	-2.3%	-16.3%	1.3%	-8.3%	-8.0%	-10.6%
LTSA	RMSE	0.16	0.27	0.29	5	8.51	9.62	14.69	25.39	28.83	10.43	22.26	24.45	4.54	9.47	10.78	7
	Relative RMSE [%]	-51.5%	-18.2%	-54.7%	11.6%	-10.1%	-14.0%	22.8%	10.6%	1.7%	-9.5%	-1.1%	-14.1%	-1.9%	-7.2%	-8.3%	-9.6%
Laplacian	RMSE	0.18	0.29	0.31	4.75	8.61	9.84	13.52	24.63	28.6	10.68	22.83	23.94	4.8	9.55	10.8	8
	Relative RMSE [%]	-45.5%	-12.1%	-51.6%	6.0%	-9.1%	-12.0%	13.0%	7.3%	0.8%	-7.4%	1.5%	-15.9%	3.7%	-6.4%	-8.2%	-9.0%
Fractional	RMSE	0.22	0.34	0.38	4.43	8.95	10.76	12.08	23.62	29.69	8.62	22.19	25.67	4.86	9.39	10.15	9
	Relative RMSE [%]	-33.3%	3.0%	-40.6%	-1.1%	-5.5%	-3.8%	1.0%	2.9%	4.7%	-25.2%	-1.4%	-9.8%	5.0%	-7.9%	-13.7%	-8.4%
tSNE	RMSE	0.16	0.28	0.29	4.94	8.42	9.26	14.56	24.6	27.7	12.78	23.66	25.37	4.88	9.5	11.16	10
	Relative RMSE [%]	-51.5%	-15.2%	-54.7%	10.3%	-11.1%	-17.2%	21.7%	7.2%	-2.3%	10.8%	5.2%	-10.9%	5.4%	-6.9%	-5.1%	-7.6%
Kernel PCA	RMSE	0.15	0.28	0.28	5.06	8.58	9.47	15.19	24.42	27.17	13.07	23.88	25.83	4.95	9.11	10.6	11
	Relative RMSE [%]	-54.5%	-15.2%	-56.3%	12.9%	-9.4%	-15.3%	27.0%	6.4%	-4.2%	13.4%	6.1%	-9.3%	6.9%	-10.7%	-9.9%	-7.5%
RP	RMSE	0.15	0.26	0.28	5.1	9.58	10.35	14.35	23.98	26.75	12.41	24.82	25.84	5	10.91	12.11	12
	Relative RMSE [%]	-54.5%	-21.2%	-56.3%	13.8%	1.2%	-7.4%	20.0%	4.5%	-5.7%	7.6%	10.3%	-9.2%	8.0%	7.0%	3.0%	-5.3%
FA	RMSE	0.22	0.45	0.5	4.28	9.26	10.82	11.09	24.08	29.49	10.31	24.65	27.93	4.55	9.23	9.99	13
	Relative RMSE [%]	-33.3%	36.4%	-21.9%	-4.5%	-2.2%	-3.2%	-7.3%	4.9%	4.0%	-10.6%	9.6%	-1.9%	-1.7%	-9.5%	-15.1%	-3.8%
LIM	RMSE	0.21	0.35	0.37	4.5	10.06	12.27	11.36	24.12	29.96	10.56	25.17	27.89	4.81	9.81	10.93	14
	Relative RMSE [%]	-36.4%	6.1%	-42.2%	0.4%	6.2%	9.7%	-5.0%	5.1%	5.6%	-8.4%	11.9%	-2.0%	3.9%	-3.8%	-7.1%	-3.7%
IM	RMSE	0.23	0.38	0.37	4.55	10.1	12.34	11.41	24.47	29.61	10.71	25.37	27.98	4.75	9.61	10.91	15
	Relative RMSE [%]	-30.3%	15.2%	-42.2%	1.6%	6.7%	10.4%	-4.6%	6.6%	4.4%	-7.1%	12.8%	-1.7%	2.6%	-5.8%	-7.2%	-2.6%
Euclidean	RMSE	0.33	0.33	0.64	4.48	9.47	11.18	11.96	22.95	28.36	11.53	22.5	28.47	4.63	10.2	11.76	16
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
PCA	RMSE	0.24	0.4	0.48	4.8	10.56	13.3	11.51	24.17	29.57	16.58	31.69	32.55	4.93	9.96	11.61	17
	Relative RMSE [%]	-27.3%	21.2%	-25.0%	7.1%	11.5%	19.0%	-3.8%	5.3%	4.3%	43.8%	40.8%	14.3%	6.5%	-2.4%	-1.3%	7.6%
ProbPCA	RMSE	0.33	0.55	0.55	4.84	10.65	13.41	11.72	24.2	29.39	16.85	33.11	32.81	4.85	9.69	10.85	18
	Relative RMSE [%]	0.0%	66.7%	-14.1%	8.0%	12.5%	19.9%	-2.0%	5.4%	3.6%	46.1%	47.2%	15.2%	4.8%	-5.0%	-7.7%	13.4%

Table A.19: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Cieszyn station (RMSE) for e-APFM $\alpha\beta$-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Laplacian	RMSE	0.15	0.28	0.32	4.8	9.18	10.83	13.58	24.57	27.93	12.26	25.95	27.9	4.78	9.82	11.19	1
	Relative RMSE [%]	-55.9%	-48.1%	-33.3%	0.6%	-15.3%	-15.7%	7.1%	0.9%	-2.8%	14.4%	2.9%	-3.0%	-5.2%	-14.2%	-14.6%	-12.1%
Fractional	RMSE	0.15	0.29	0.35	4.43	9.35	11.33	13.52	26.88	34.43	10.02	23.63	26.97	5.45	10.04	10.92	2
	Relative RMSE [%]	-55.9%	-46.3%	-27.1%	-7.1%	-13.7%	-11.8%	6.6%	10.4%	19.8%	-6.5%	-6.3%	-6.3%	8.1%	-12.3%	-16.6%	-11.0%
IM	RMSE	0.15	0.31	0.35	4.8	10.52	13.09	11.82	24.1	29.38	12.69	26.54	28.71	5.47	10.5	11.61	3
	Relative RMSE [%]	-55.9%	-42.6%	-27.1%	0.6%	-3.0%	1.9%	-6.8%	-1.0%	2.2%	18.4%	5.2%	-0.2%	8.5%	-8.3%	-11.4%	-7.9%
LIM	RMSE	0.15	0.31	0.35	4.79	10.56	12.85	11.92	24.28	29.57	12.62	27.13	29.84	5.47	10.72	11.8	4
	Relative RMSE [%]	-55.9%	-42.6%	-27.1%	0.4%	-2.6%	0.1%	-6.0%	-0.2%	2.9%	17.7%	7.6%	3.7%	8.5%	-6.4%	-9.9%	-7.3%
FA	RMSE	0.19	0.36	0.4	4.63	10.1	11.69	11.68	24.36	30.4	11.86	25.48	27.79	5.54	11.13	11.86	5
	Relative RMSE [%]	-44.1%	-33.3%	-16.7%	-2.9%	-6.8%	-9.0%	-7.9%	0.1%	5.8%	10.6%	1.0%	-3.4%	9.9%	-2.8%	-9.5%	-7.3%
NPE	RMSE	0.16	0.29	0.34	4.46	9.46	11.21	13.51	23.88	28.24	16.11	27.96	29.81	5.34	10.52	11.36	6
	Relative RMSE [%]	-52.9%	-46.3%	-29.2%	-6.5%	-12.7%	-12.7%	6.5%	-1.9%	-1.7%	50.3%	10.9%	3.6%	6.0%	-8.1%	-13.3%	-7.2%
LLTSA	RMSE	0.16	0.3	0.35	4.41	9.69	11.36	13.63	24.14	27.35	15.27	29.34	32.05	5.17	10.14	11.18	7
	Relative RMSE [%]	-52.9%	-44.4%	-27.1%	-7.5%	-10.6%	-11.5%	7.5%	-0.8%	-4.8%	42.4%	16.3%	11.4%	2.6%	-11.4%	-14.7%	-7.0%
LPP	RMSE	0.16	0.29	0.34	4.67	9.74	11.37	14.13	23.96	27.77	15.71	27.83	30.92	5.23	10.27	11.46	8
	Relative RMSE [%]	-52.9%	-46.3%	-29.2%	-2.1%	-10.1%	-11.4%	11.4%	-1.6%	-3.4%	46.5%	10.3%	7.5%	3.8%	-10.3%	-12.5%	-6.7%
LTSA	RMSE	0.15	0.3	0.35	4.54	9.5	10.87	14.37	25.23	28.93	14.55	30.14	33.48	4.94	10.35	11.39	9
	Relative RMSE [%]	-55.9%	-44.4%	-27.1%	-4.8%	-12.4%	-15.3%	13.3%	3.7%	0.7%	35.7%	19.5%	16.4%	-2.0%	-9.6%	-13.1%	-6.4%
SPE	RMSE	0.18	0.31	0.38	4.58	9.5	11.43	14.19	22.94	27.54	15.1	29.33	31.69	5.55	10.82	12.18	10
	Relative RMSE [%]	-47.1%	-42.6%	-20.8%	-4.0%	-12.4%	-11.0%	11.9%	-5.8%	-4.2%	40.9%	16.3%	10.1%	10.1%	-5.5%	-7.0%	-4.7%
LMDS	RMSE	0.22	0.34	0.41	4.54	9.61	11.39	14.89	23.88	27.7	15	29.02	32.47	5.03	10.15	11.16	11
	Relative RMSE [%]	-35.3%	-37.0%	-14.6%	-4.8%	-11.3%	-11.3%	17.4%	-1.9%	-3.6%	39.9%	15.1%	12.9%	-0.2%	-11.4%	-14.8%	-4.1%
LLE	RMSE	0.2	0.33	0.39	4.5	9.84	11.47	13.28	23.56	27.89	15.85	29.76	32.67	5.27	10.59	11.77	12
	Relative RMSE [%]	-41.2%	-38.9%	-18.8%	-5.7%	-9.2%	-10.7%	4.7%	-3.2%	-3.0%	47.9%	18.0%	13.6%	4.6%	-7.5%	-10.2%	-4.0%
tSNE	RMSE	0.17	0.3	0.34	4.7	9.66	11.17	14.46	24.88	28.44	17.57	30.99	32.81	5.14	10.16	11.17	13
	Relative RMSE [%]	-50.0%	-44.4%	-29.2%	-1.5%	-10.9%	-13.0%	14.0%	2.2%	-1.0%	63.9%	22.9%	14.0%	2.0%	-11.3%	-14.7%	-3.8%
Kernel PCA	RMSE	0.17	0.31	0.36	4.58	9.58	11.46	14.74	24.1	27.73	18.86	31.85	34.24	5.14	10.39	11.37	14
	Relative RMSE [%]	-50.0%	-42.6%	-25.0%	-4.0%	-11.6%	-10.7%	16.2%	-1.0%	-3.5%	75.9%	26.3%	19.0%	2.0%	-9.3%	-13.2%	-2.1%
Euclidean	RMSE	0.34	0.54	0.48	4.77	10.84	12.84	12.68	24.34	28.74	10.72	25.22	28.77	5.04	11.45	13.1	15
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
PCA	RMSE	0.17	0.34	0.39	4.96	11.14	13.95	11.65	24.17	29.56	18.5	30.79	33.47	5.69	10.79	11.65	16
	Relative RMSE [%]	-50.0%	-37.0%	-18.8%	4.0%	2.8%	8.6%	-8.1%	-0.7%	2.9%	72.6%	22.1%	16.3%	12.9%	-5.8%	-11.1%	0.7%
ProbPCA	RMSE	0.18	0.34	0.38	5	11.04	13.56	12.12	27.97	32.04	15	31.26	33.12	5.63	10.95	11.98	17
	Relative RMSE [%]	-47.1%	-37.0%	-20.8%	4.8%	1.8%	5.6%	-4.4%	14.9%	11.5%	39.9%	23.9%	15.1%	11.7%	-4.4%	-8.5%	0.5%
RP	RMSE	0.16	0.3	0.35	4.64	12.71	15.25	13.68	23.71	27.65	14.67	30.16	31.48	5.2	14.13	15.21	18
	Relative RMSE [%]	-52.9%	-44.4%	-27.1%	-2.7%	17.3%	18.8%	7.9%	-2.6%	-3.8%	36.8%	19.6%	9.4%	3.2%	23.4%	16.1%	1.3%

Table A.20: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM average method for average method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.41	0.41	0.42	10.97	10.69	10.73	15.92	15.69	17.05	49.18	59.33	62.04	14.67	14.07	14.24	1
	Relative RMSE [%]	-21.2%	-18.0%	-16.0%	-10.9%	-9.4%	-9.6%	-13.3%	-12.2%	-10.3%	-25.6%	-18.2%	-15.4%	-22.4%	-22.2%	-20.0%	-16.3%
IM	RMSE	0.41	0.4	0.41	11.05	10.73	10.62	14.64	15.91	17.43	55.13	61.98	63.39	14.34	13.66	13.99	2
	Relative RMSE [%]	-21.2%	-20.0%	-18.0%	-10.2%	-9.1%	-10.5%	-20.3%	-11.0%	-8.3%	-16.5%	-14.6%	-13.6%	-24.1%	-24.5%	-21.4%	-16.2%
LIM	RMSE	0.41	0.41	0.42	11.05	10.78	10.65	14.66	16	17.55	55.21	62.41	64.03	14.46	13.93	14.19	3
	Relative RMSE [%]	-21.2%	-18.0%	-16.0%	-10.2%	-8.6%	-10.3%	-20.2%	-10.5%	-7.7%	-16.4%	-14.0%	-12.7%	-23.5%	-23.0%	-20.3%	-15.5%
FA	RMSE	0.42	0.41	0.42	11.46	11.14	10.94	15.03	16.35	17.98	56.51	63.11	64.68	14.75	13.97	14.14	4
	Relative RMSE [%]	-19.2%	-18.0%	-16.0%	-6.9%	-5.6%	-7.8%	-18.2%	-8.6%	-5.4%	-14.5%	-13.0%	-11.8%	-22.0%	-22.8%	-20.6%	-14.0%
PCA	RMSE	0.41	0.41	0.42	11.17	11.05	10.9	15.15	15.97	17.48	65.64	65.35	66.22	15.04	14.79	14.95	5
	Relative RMSE [%]	-21.2%	-18.0%	-16.0%	-9.3%	-6.4%	-8.2%	-17.5%	-10.7%	-8.0%	-0.6%	-9.9%	-9.7%	-20.4%	-18.2%	-16.1%	-12.7%
ProbPCA	RMSE	0.42	0.42	0.43	11.64	11.48	11.29	16.29	17.38	18.84	66.83	66.69	67.07	15.73	15.25	15.4	6
	Relative RMSE [%]	-19.2%	-16.0%	-14.0%	-5.4%	-2.7%	-4.9%	-11.3%	-2.8%	-0.9%	1.2%	-8.1%	-8.5%	-16.8%	-15.7%	-13.5%	-9.3%
Euclidean	RMSE	0.52	0.5	0.5	12.31	11.8	11.87	18.37	17.88	19.01	66.06	72.57	73.33	18.9	18.09	17.81	7
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Laplacian	RMSE	0.48	0.48	0.48	14.88	14.98	15.25	28.98	28.72	28.16	60.57	69.15	69.74	20.23	20.42	20.5	8
	Relative RMSE [%]	-7.7%	-4.0%	-4.0%	20.9%	26.9%	28.5%	57.8%	60.6%	48.1%	-8.3%	-4.7%	-4.9%	7.0%	12.9%	15.1%	16.3%
LPP	RMSE	0.52	0.51	0.51	14.92	14.87	14.89	27.95	28.03	27.54	59.84	69.72	72.17	21.03	20.98	20.88	9
	Relative RMSE [%]	0.0%	2.0%	2.0%	21.2%	26.0%	25.4%	52.2%	56.8%	44.9%	-9.4%	-3.9%	-1.6%	11.3%	16.0%	17.2%	17.3%
tSNE	RMSE	0.51	0.51	0.51	14.67	14.88	14.98	26.88	27.14	26.47	73.7	73.74	73.5	20.83	20.89	20.78	10
	Relative RMSE [%]	-1.9%	2.0%	2.0%	19.2%	26.1%	26.2%	46.3%	51.8%	39.2%	11.6%	1.6%	0.2%	10.2%	15.5%	16.7%	17.8%
NPE	RMSE	0.52	0.51	0.51	14.95	15.22	15.41	27.63	27.9	27.27	59.35	69.89	72.5	21.18	21.25	21.3	11
	Relative RMSE [%]	0.0%	2.0%	2.0%	21.4%	29.0%	29.8%	50.4%	56.0%	43.5%	-10.2%	-3.7%	-1.1%	12.1%	17.5%	19.6%	17.9%
LLE	RMSE	0.52	0.52	0.52	14.76	14.99	15.33	27.24	27.57	27.01	59.79	70.05	72.84	21.39	21.67	21.99	12
	Relative RMSE [%]	0.0%	4.0%	4.0%	19.9%	27.0%	29.1%	48.3%	54.2%	42.1%	-9.5%	-3.5%	-0.7%	13.2%	19.8%	23.5%	18.1%
LLTSA	RMSE	0.52	0.52	0.51	14.98	15.38	15.58	28.04	28.37	27.66	60.05	70.51	72.96	21.34	21.51	21.48	13
	Relative RMSE [%]	0.0%	4.0%	2.0%	21.7%	30.3%	31.3%	52.6%	58.7%	45.5%	-9.1%	-2.8%	-0.5%	12.9%	18.9%	20.6%	19.1%
LTSA	RMSE	0.53	0.53	0.52	15.15	15.37	15.44	28.38	28.37	27.57	60.5	70.38	72.89	21.32	21.41	21.48	14
	Relative RMSE [%]	1.9%	6.0%	4.0%	23.1%	30.3%	30.1%	54.5%	58.7%	45.0%	-8.4%	-3.0%	-0.6%	12.8%	18.4%	20.6%	19.5%
RP	RMSE	0.52	0.51	0.51	14.81	15.83	16.08	26.25	29.42	28.76	60.35	72.3	73.93	20.91	21.41	21.55	15
	Relative RMSE [%]	0.0%	2.0%	2.0%	20.3%	34.2%	35.5%	42.9%	64.5%	51.3%	-8.6%	-0.4%	0.8%	10.6%	18.4%	21.0%	19.6%
LMDS	RMSE	0.52	0.52	0.51	15.17	15.5	15.69	28.31	28.61	28.07	60.03	70.15	72.33	21.59	21.72	21.61	16
	Relative RMSE [%]	0.0%	4.0%	2.0%	23.2%	31.4%	32.2%	54.1%	60.0%	47.7%	-9.1%	-3.3%	-1.4%	14.2%	20.1%	21.3%	19.8%
Kernel PCA	RMSE	0.52	0.52	0.51	15	15.23	15.36	27.94	28.08	27.51	73.87	73.99	73.63	21.13	21.24	21.2	17
	Relative RMSE [%]	0.0%	4.0%	2.0%	21.9%	29.1%	29.4%	52.1%	57.0%	44.7%	11.8%	2.0%	0.4%	11.8%	17.4%	19.0%	20.2%
SPE	RMSE	0.5	0.5	0.5	15.78	15.9	16.31	29.38	29.5	28.97	62.66	71.59	72.83	21.26	21.26	21.64	18
	Relative RMSE [%]	-3.8%	0.0%	0.0%	28.2%	34.7%	37.4%	59.9%	65.0%	52.4%	-5.1%	-1.4%	-0.7%	12.5%	17.5%	21.5%	21.2%

Table A.21: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM α-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.24	0.54	0.59	6.22	11.58	13.18	10.69	17.75	20.11	24.83	63.07	70.38	7.67	15.79	17.5	1
	Relative RMSE [%]	-78.2%	-56.5%	-60.4%	4.9%	-1.5%	-6.6%	0.5%	-3.1%	-5.1%	-29.8%	-19.4%	-16.0%	-8.0%	-10.7%	-17.3%	-20.5%
LIM	RMSE	0.24	0.54	0.58	6.04	11.69	13.12	10.6	19.01	21.64	29.3	75.31	83.95	7.65	15.56	17.3	2
	Relative RMSE [%]	-78.2%	-56.5%	-61.1%	1.9%	-0.6%	-7.0%	-0.4%	3.8%	2.1%	-17.1%	-3.8%	0.2%	-8.3%	-12.0%	-18.3%	-17.0%
IM	RMSE	0.24	0.55	0.6	6.1	11.51	13.19	10.52	18.83	21.58	28.55	74.8	83.82	7.68	15.76	17.69	3
	Relative RMSE [%]	-78.2%	-55.6%	-59.7%	2.9%	-2.1%	-6.5%	-1.1%	2.8%	1.8%	-19.2%	-4.4%	0.0%	-7.9%	-10.9%	-16.4%	-17.0%
FA	RMSE	0.25	0.55	0.59	6.33	11.99	13.54	10.78	19.09	21.46	27.91	77.48	89.85	7.97	16.65	18.65	4
	Relative RMSE [%]	-77.3%	-55.6%	-60.4%	6.7%	2.0%	-4.0%	1.3%	4.2%	1.2%	-21.0%	-1.0%	7.2%	-4.4%	-5.9%	-11.9%	-14.6%
PCA	RMSE	0.24	0.54	0.58	6.1	11.6	13.25	10.52	18.77	21.5	33	83.82	96.28	8.34	16.33	18.6	5
	Relative RMSE [%]	-78.2%	-56.5%	-61.1%	2.9%	-1.4%	-6.1%	-1.1%	2.5%	1.4%	-6.6%	7.1%	14.9%	0.0%	-7.7%	-12.1%	-13.5%
ProbPCA	RMSE	0.24	0.57	0.6	6.08	11.55	13.41	10.69	18.89	21.31	35.14	84.94	92.76	8.28	16.73	18.82	6
	Relative RMSE [%]	-78.2%	-54.0%	-59.7%	2.5%	-1.8%	-5.0%	0.5%	3.1%	0.5%	-0.6%	8.5%	10.7%	-0.7%	-5.4%	-11.1%	-12.7%
Laplacian	RMSE	0.24	0.58	0.61	6.81	11.42	13.85	15.59	21.92	23.57	28.19	73.36	76.69	7.52	16.74	18.76	7
	Relative RMSE [%]	-78.2%	-53.2%	-59.1%	14.8%	-2.9%	-1.8%	46.5%	19.7%	11.2%	-20.3%	-6.3%	-8.5%	-9.8%	-5.4%	-11.4%	-11.0%
LPP	RMSE	0.24	0.57	0.62	6.8	11.48	12.65	14.86	21.36	22.49	28.58	69.58	74.39	8.82	18.92	20.29	8
	Relative RMSE [%]	-78.2%	-54.0%	-58.4%	14.7%	-2.4%	-10.3%	39.7%	16.6%	6.1%	-19.2%	-11.1%	-11.2%	5.8%	7.0%	-4.2%	-10.6%
LTSA	RMSE	0.23	0.55	0.58	7.09	12.16	13.43	15.14	21.55	22.68	27.91	69.55	73.58	9.2	19.7	20.96	9
	Relative RMSE [%]	-79.1%	-55.6%	-61.1%	19.6%	3.4%	-4.8%	42.3%	17.6%	7.0%	-21.0%	-11.1%	-12.2%	10.3%	11.4%	-1.0%	-9.0%
NPE	RMSE	0.25	0.58	0.61	6.85	11.77	13.39	15.33	21.57	22.65	29.05	70.27	75.32	9.14	19.42	21.63	10
	Relative RMSE [%]	-77.3%	-53.2%	-59.1%	15.5%	0.1%	-5.1%	44.1%	17.7%	6.8%	-17.8%	-10.2%	-10.1%	9.6%	9.8%	2.2%	-8.5%
LLTSA	RMSE	0.27	0.61	0.64	7.03	12.17	13.16	15.65	22.02	22.97	28.41	70.3	74.75	9.2	20.3	21.16	11
	Relative RMSE [%]	-75.5%	-50.8%	-57.0%	18.5%	3.5%	-6.7%	47.1%	20.2%	8.3%	-19.6%	-10.2%	-10.8%	10.3%	14.8%	0.0%	-7.2%
LMDS	RMSE	0.23	0.56	0.6	6.94	12.23	13.5	15.85	22.15	23.05	29.34	72.3	75.73	9.19	20.13	22.33	12
	Relative RMSE [%]	-79.1%	-54.8%	-59.7%	17.0%	4.0%	-4.3%	49.0%	20.9%	8.7%	-17.0%	-7.6%	-9.6%	10.2%	13.8%	5.5%	-6.9%
tSNE	RMSE	0.26	0.6	0.62	6.75	11.51	13.26	14.68	21.37	22.43	38.26	81.53	87.04	9.09	19.06	20.87	13
	Relative RMSE [%]	-76.4%	-51.6%	-58.4%	13.8%	-2.1%	-6.0%	38.0%	16.6%	5.8%	8.2%	4.2%	3.9%	9.0%	7.7%	-1.4%	-5.9%
SPE	RMSE	0.25	0.58	0.62	7.23	12.12	14.11	16.31	22.51	23.81	30.16	74.74	75.17	9.11	19.89	21.12	14
	Relative RMSE [%]	-77.3%	-53.2%	-58.4%	21.9%	3.1%	0.0%	53.3%	22.9%	12.3%	-14.7%	-4.5%	-10.3%	9.2%	12.4%	-0.2%	-5.6%
Kernel PCA	RMSE	0.24	0.58	0.64	6.8	11.98	13.5	15.47	21.54	22.65	32.81	81.02	89.65	9.23	20.09	22.32	15
	Relative RMSE [%]	-78.2%	-53.2%	-57.0%	14.7%	1.9%	-4.3%	45.4%	17.6%	6.8%	-7.2%	3.5%	7.0%	10.7%	13.6%	5.4%	-4.9%
RP	RMSE	0.26	0.61	0.65	6.93	12.94	14.51	15.27	24.81	25.54	31.68	77.23	78.99	9.04	20.51	22.35	16
	Relative RMSE [%]	-76.4%	-50.8%	-56.4%	16.9%	10.0%	2.8%	43.5%	35.4%	20.5%	-10.4%	-1.3%	-5.7%	8.4%	15.9%	5.6%	-2.8%
LLE	RMSE	0.31	0.68	0.69	6.94	11.99	14.07	16.3	22.32	23.22	29.25	74.86	77.8	10.09	21.28	23.93	17
	Relative RMSE [%]	-71.8%	-45.2%	-53.7%	17.0%	2.0%	-0.3%	53.2%	21.8%	9.5%	-17.3%	-4.4%	-7.1%	21.0%	20.3%	13.0%	-2.8%
Euclidean	RMSE	1.1	1.24	1.49	5.93	11.76	14.11	10.64	18.32	21.2	35.35	78.27	83.79	8.34	17.69	21.17	18
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A.22: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) for e-APFM $\alpha\beta$-stand. method.

Method	Measure	CO			NO_2			O_3			PM10			SO_2			**Rank**
		+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	+1 day	+2 day	+3 day	$\overline{RMSE_{vsEucl.}}$
Fractional	RMSE	0.26	0.59	0.71	6.34	12.01	13.86	10.6	18.18	20.51	27.76	67.39	75.3	8.3	17.72	19.68	1
	Relative RMSE [%]	-71.4%	-49.6%	-36.6%	-4.4%	-13.5%	-13.5%	-4.5%	-4.4%	-4.8%	-20.9%	-14.6%	-14.2%	-12.2%	-21.0%	-18.5%	-20.3%
IM	RMSE	0.27	0.61	0.68	6.16	12.5	14.66	10.56	18.79	21.17	31.92	77.5	88.83	8.97	20.31	22.71	2
	Relative RMSE [%]	-70.3%	-47.9%	-39.3%	-7.1%	-9.9%	-8.5%	-4.9%	-1.2%	-1.7%	-9.0%	-1.8%	1.2%	-5.1%	-9.5%	-6.0%	-14.7%
LIM	RMSE	0.28	0.62	0.66	6.23	12.7	14.57	10.7	19.01	21.45	32.82	80.3	88.15	9.4	20.89	22.77	3
	Relative RMSE [%]	-69.2%	-47.0%	-41.1%	-6.0%	-8.5%	-9.1%	-3.6%	0.0%	-0.4%	-6.5%	1.7%	0.4%	-0.5%	-6.9%	-5.7%	-13.5%
LTSA	RMSE	0.27	0.58	0.62	6.23	12.54	14.4	14.71	22.4	25.11	28.06	70.81	77.84	8.91	18.99	21.18	4
	Relative RMSE [%]	-70.3%	-50.4%	-44.6%	-6.0%	-9.7%	-10.2%	32.5%	17.8%	16.6%	-20.0%	-10.3%	-11.3%	-5.7%	-15.4%	-12.3%	-13.3%
NPE	RMSE	0.28	0.61	0.66	6.72	12.86	15.01	13.91	21.45	23.67	28.03	74.67	78.47	8.44	18.86	21.58	5
	Relative RMSE [%]	-69.2%	-47.9%	-41.1%	1.4%	-7.3%	-6.4%	25.3%	12.8%	9.9%	-20.1%	-5.4%	-10.6%	-10.7%	-16.0%	-10.6%	-13.1%
LPP	RMSE	0.28	0.61	0.65	6.2	12.55	14.58	15.29	22.23	24.12	29.06	70.87	77.46	8.26	18.77	21.84	6
	Relative RMSE [%]	-69.2%	-47.9%	-42.0%	-6.5%	-9.6%	-9.0%	37.7%	16.9%	12.0%	-17.2%	-10.2%	-11.8%	-12.6%	-16.4%	-9.6%	-13.0%
FA	RMSE	0.28	0.63	0.7	6.36	12.89	14.92	11.03	19.15	21.5	32.82	80.53	90.68	9.09	20.17	22.67	7
	Relative RMSE [%]	-69.2%	-46.2%	-37.5%	-4.1%	-7.1%	-6.9%	-0.6%	0.7%	-0.2%	-6.5%	2.0%	3.3%	-3.8%	-10.1%	-6.1%	-12.8%
Laplacian	RMSE	0.28	0.58	0.64	6.74	12.31	14.52	16.6	23.47	24.95	28.74	73.73	77	8.34	16.88	19.19	8
	Relative RMSE [%]	-69.2%	-50.4%	-42.9%	1.7%	-11.3%	-9.4%	49.5%	23.5%	15.8%	-18.1%	-6.6%	-12.3%	-11.7%	-24.8%	-20.5%	-12.5%
LMDS	RMSE	0.28	0.61	0.66	6.54	13.23	15.15	14.52	21.88	23.74	29.02	73.86	77.95	8.58	18.89	22.11	9
	Relative RMSE [%]	-69.2%	-47.9%	-41.1%	-1.4%	-4.7%	-5.5%	30.8%	15.1%	10.2%	-17.3%	-6.4%	-11.2%	-9.2%	-15.8%	-8.4%	-12.1%
LLTSA	RMSE	0.32	0.65	0.69	7.02	12.89	14.78	13.75	21.47	23.7	28.64	72.12	77.34	8.63	19.69	22.48	10
	Relative RMSE [%]	-64.8%	-44.4%	-38.4%	5.9%	-7.1%	-7.8%	23.9%	12.9%	10.0%	-18.4%	-8.6%	-11.9%	-8.7%	-12.3%	-6.9%	-11.8%
SPE	RMSE	0.28	0.58	0.62	6.94	13.01	15.27	15.8	23.59	25.62	29.64	77.25	78.34	8.3	17.94	20.65	11
	Relative RMSE [%]	-69.2%	-50.4%	-44.6%	4.7%	-6.3%	-4.7%	42.3%	24.1%	18.9%	-15.5%	-2.1%	-10.8%	-12.2%	-20.1%	-14.5%	-10.7%
LLE	RMSE	0.31	0.67	0.72	6.45	12.89	15.38	14.8	21.63	23.72	29.5	73.32	80.14	8.75	19.85	23.08	12
	Relative RMSE [%]	-65.9%	-42.7%	-35.7%	-2.7%	-7.1%	-4.1%	33.3%	13.8%	10.1%	-15.9%	-7.1%	-8.7%	-7.4%	-11.5%	-4.4%	-10.4%
RP	RMSE	0.3	0.63	0.68	6.25	13.5	15.5	14.6	23.6	25.34	27.92	73.28	77.91	8.39	19.8	23.34	13
	Relative RMSE [%]	-67.0%	-46.2%	-39.3%	-5.7%	-2.7%	-3.3%	31.5%	24.1%	17.6%	-20.4%	-7.2%	-11.2%	-11.2%	-11.8%	-3.4%	-10.4%
ProbPCA	RMSE	0.28	0.63	0.69	6.22	12.3	14.49	10.6	18.77	21.38	39.37	93.01	105.43	9.53	20.83	23.1	14
	Relative RMSE [%]	-69.2%	-46.2%	-38.4%	-6.2%	-11.4%	-9.6%	-4.5%	-1.3%	-0.7%	12.2%	17.8%	20.1%	0.8%	-7.2%	-4.3%	-9.9%
PCA	RMSE	0.26	0.58	0.65	6.34	12.56	14.54	11.13	19.04	21.76	43.19	97.12	108.2	8.71	19.2	22.14	15
	Relative RMSE [%]	-71.4%	-50.4%	-42.0%	-4.4%	-9.5%	-9.3%	0.3%	0.2%	1.0%	23.1%	23.0%	23.3%	-7.8%	-14.4%	-8.3%	-9.8%
Kernel PCA	RMSE	0.29	0.6	0.66	6.21	12.71	14.67	14.25	21.61	23.64	36.66	84.73	97.29	8.27	18.75	22.18	16
	Relative RMSE [%]	-68.1%	-48.7%	-41.1%	-6.3%	-8.4%	-8.5%	28.4%	13.7%	9.7%	4.5%	7.3%	10.8%	-12.5%	-16.4%	-8.2%	-9.6%
tSNE	RMSE	0.31	0.66	0.71	6.15	12.64	14.79	14.98	22	24.06	37.3	88.74	98.1	8.44	19.05	22.54	17
	Relative RMSE [%]	-65.9%	-43.6%	-36.6%	-7.2%	-8.9%	-7.7%	35.0%	15.7%	11.7%	6.3%	12.4%	11.8%	-10.7%	-15.1%	-6.7%	-7.3%
Euclidean	RMSE	0.91	1.17	1.12	6.63	13.88	16.03	11.1	19.01	21.54	35.09	78.93	87.78	9.45	22.44	24.15	18
	Relative RMSE [%]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Tables

Table 1: Selected methods of dimensionality reduction

Method	**Convex**	**Linear**	**Out-of-sample extension**	**Parameters[a] with used values**	**References**
Factor Analysis	Yes	Yes	Yes	$i_{max} = 200$	(Spearman, 1904)
Isomap	Yes	No	No	$k = 12$	(Tenenbaum et al., 2000)
Kernel Principal Component Analysis	Yes	No	No	$K(x) = K_N(x)$	(Schölkopf et al., 1998)
Landmark Isomap	Yes	No	No	$k = 12$	(de Silva and Tenenbaum, 2003)
				$r_L = 0.2$	
Landmark Multidimensional Scaling	Yes	No	No	$r_L = 0.2$	(de Silva and Tenenbaum, 2004)
Laplacian Eigenmaps	Yes	No	No	$k = 12$	(Belkin and Niyogi, 2003)
				$\sigma = 1$	
Linear Local Tangent Space Alignment	Yes	Yes	Yes	$k = 12$	(Zhang et al., 2007)
Locally Linear Embedding	Yes	No	No	$k = 12$ (Roweis and Saul, 2000)	
Local Tangent Space Alignment	Yes	No	No	$k = 12$	(Zhang and Hongyuan, 2004)
Locality Preserving Projection	Yes	Yes	Yes	$k = 12$	(He and Niyogi, 2003)
				$\sigma = 1$	
Neighborhood Preserving Embedding	Yes	Yes	Yes	$k = 12$	(He et al., 2005)
Principal Component Analysis	Yes	Yes	Yes	–	(Jolliffe, 2002)
Probabilistic Principal Component Analysis	Yes	Yes	Yes	$i_{max} = 200$	(Tipping and Bishop, 1999)
Random Projections	Yes	Yes	Yes	–	(Bingham and Mannila, 2001)
Stochastic Proximity Embedding	Yes	No	No	$m_{up} = 20$	(Agrafiotis, 2003)
				$\lambda = 1$	
t-Distributed Stochastic Neighbor Embedding	Yes	No	No	$p = 30$	(van der Maaten and Hinton, 2008)

[a] Does not include dimensionality of the reduced feature space N

Table 2: Parameters of weather forecast collected between January 1st 2005 and December 31st 2012

Variable	Unit	Range
Pressure	hPa	[977.5, 1052.8]
Wind direction at 10m	$^\circ$	[0.0, 359.0]
Wind speed at 10m	m/s	[0.0, 14.0]
Wind direction at 500 hpa	$^\circ$	[0.0, 359.0]
Wind speed at 500 hpa	m/s	[0.0, 38.0]
Wind direction on 700 hpa	$^\circ$	[0.0, 359.0]
Wind speed on 700 hpa	m/s	[0.0, 44.0]
Wind direction on 850 hpa	$^\circ$	[0.0, 359.0]
Wind speed on 850 hpa	m/s	[0.0, 47.0]
Wind direction on 950 hpa	$^\circ$	[0.0, 359.0]
Wind speed on 950 hpa	m/s	[0.0, 68.0]
Ground temperature	$^\circ$C	[-30.7, 35.8]
Temperature at 2m height	$^\circ$C	[-27.2, 35.1]
Dew point temperature at 2m	$^\circ$C	[-29.4, 25.4]
Temperature at 30m	$^\circ$C	[-24.9, 32.1]
Temperature on 850 hpa	$^\circ$C	[-20.1, 25.1]
Temperature on 700 hpa	$^\circ$C	[-29.2, 12.0]
Temperature on 500 hpa	$^\circ$C	[-42.9, 0.0]
Cloud cover from the lower floor	Octant	[0.0, 8.0]
Cloud cover from the medium floor	Octant	[0.0, 8.0]
Cloud cover from the upper floor	Octant	[0.0, 8.0]
Ground fog	Octant	[0.0, 8.0]
Base height of convection cloud above msl	hPa	[0.0, 471.0]
Top height of convection cloud above msl	hPa	[0.0, 1430.0]
Rain amount (grid scale and convective)	mm	[0.0, 103.5]
Snow amount (grid scale and convective)	mm	[0.0, 32.5]
Water content of snow	m	[-0.5, 0.3]

Table 3: Summary of experimental results

		KTW station hourly avg.	KTW station hourly α	KTW station hourly $\alpha\beta$	WOD station hourly avg.	WOD station hourly α	WOD station hourly $\alpha\beta$	CIE station daily avg. avg.	CIE station daily avg. α	CIE station daily avg. $\alpha\beta$	WOD station daily avg. avg.	WOD station daily avg. α	WOD station daily avg. $\alpha\beta$	mean	σ
Fractional	Rank	1	1	2	1	1	1	1	9	2	1	1	1	1.8	2.3
	$\overline{RMSE_{vsEucl.}}$	-9.4%	-16.0%	-5.4%	-8.9%	-13.2%	-12.5%	-14.3%	-8.4%	-11.0%	-16.3%	-20.5%	-20.3%	-13.0%	4.7%
IM	Rank	2	4	3	3	2	2	3	15	3	2	3	2	3.7	3.6
	$\overline{RMSE_{vsEucl.}}$	-8.2%	-14.6%	-5.0%	-7.7%	-11.0%	-9.6%	-10.1%	-2.6%	-7.9%	-16.2%	-17.0%	-14.7%	-10.4%	4.5%
LIM	Rank	3	5	4	2	3	3	2	14	4	3	2	3	4.0	3.3
	$\overline{RMSE_{vsEucl.}}$	-8.1%	-14.4%	-4.9%	-7.8%	-10.8%	-9.2%	-10.2%	-3.7%	-7.3%	-15.5%	-17.0%	-13.5%	-10.2%	4.2%
FA	Rank	5	3	6	4	4	4	5	13	5	4	4	7	5.3	2.6
	$\overline{RMSE_{vsEucl.}}$	-7.4%	-14.9%	-4.7%	-6.1%	-9.9%	-8.7%	-6.0%	-3.8%	-7.3%	-14.0%	-14.6%	-12.8%	-9.2%	4.0%
PCA	Rank	4	2	5	5	5	6	4	17	16	5	5	15	7.4	5.3
	$\overline{RMSE_{vsEucl.}}$	-8.0%	-15.0%	-4.8%	-5.6%	-9.1%	-7.8%	-6.7%	7.6%	0.7%	-12.7%	-13.5%	-9.8%	-7.1%	6.3%
ProbPCA	Rank	6	6	1	6	6	5	6	18	17	6	6	14	8.1	5.2
	$\overline{RMSE_{vsEucl.}}$	-6.3%	-14.2%	-5.5%	-2.7%	-9.0%	-8.3%	-5.1%	13.4%	0.5%	-9.3%	-12.7%	-9.9%	-5.8%	7.3%
Euclidean	Rank	7	18	8	7	15	18	7	16	15	7	18	18	12.8	5.1
	$\overline{RMSE_{vsEucl.}}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Laplacian	Rank	10	7	7	8	7	8	8	8	1	8	7	8	7.3	2.1
	$\overline{RMSE_{vsEucl.}}$	14.7%	-8.6%	-2.5%	27.3%	-2.7%	-4.8%	19.0%	-9.0%	-12.1%	16.3%	-11.0%	-12.5%	1.2%	14.1%
LPP	Rank	8	9	9	11	8	7	10	4	8	9	8	6	8.1	1.8
	$\overline{RMSE_{vsEucl.}}$	12.8%	-6.2%	0.3%	28.5%	-2.7%	-5.0%	20.6%	-11.7%	-6.7%	17.3%	-10.6%	-13.0%	2.0%	14.1%
LTSA	Rank	11	8	11	15	9	11	11	7	9	14	9	4	9.9	3.0
	$\overline{RMSE_{vsEucl.}}$	16.2%	-6.4%	0.9%	31.3%	-1.6%	-4.4%	21.3%	-9.6%	-6.4%	19.5%	-9.0%	-13.3%	3.2%	14.8%
LLTSA	Rank	13	11	13	14	12	12	9	1	7	13	11	10	10.5	3.6
	$\overline{RMSE_{vsEucl.}}$	16.7%	-5.3%	1.9%	31.0%	-0.6%	-3.9%	20.4%	-13.6%	-7.0%	19.1%	-7.2%	-11.8%	3.3%	14.6%
NPE	Rank	16	12	16	12	11	10	13	2	6	11	10	5	10.3	4.2
	$\overline{RMSE_{vsEucl.}}$	18.7%	-4.7%	4.7%	29.8%	-1.1%	-4.4%	22.7%	-11.8%	-7.2%	17.9%	-8.5%	-13.1%	3.6%	14.8%
tSNE	Rank	9	10	10	10	10	16	14	10	13	10	13	17	11.8	2.7
	$\overline{RMSE_{vsEucl.}}$	14.3%	-5.6%	0.7%	27.6%	-1.2%	-2.5%	24.0%	-7.6%	-3.8%	17.8%	-5.9%	-7.3%	4.2%	13.0%
LMDS	Rank	12	13	12	17	14	13	12	3	11	16	12	9	12.0	3.5
	$\overline{RMSE_{vsEucl.}}$	16.4%	-4.6%	1.4%	33.1%	-0.1%	-3.6%	22.2%	-11.8%	-4.1%	19.8%	-6.9%	-12.1%	4.2%	14.8%
RP	Rank	14	15	14	9	13	9	17	12	18	15	16	13	13.8	2.8
	$\overline{RMSE_{vsEucl.}}$	16.8%	-3.6%	2.1%	27.3%	-0.1%	-4.6%	26.9%	-5.3%	1.3%	19.6%	-2.8%	-10.4%	5.6%	13.3%
Kernel PCA	Rank	15	14	15	13	16	14	16	11	14	17	15	16	14.7	1.6
	$\overline{RMSE_{vsEucl.}}$	18.1%	-3.7%	4.1%	30.8%	0.4%	-3.3%	26.8%	-7.5%	-2.1%	20.2%	-4.9%	-9.6%	5.8%	14.2%
LLE	Rank	18	17	18	16	18	15	15	6	12	12	17	12	14.7	3.6
	$\overline{RMSE_{vsEucl.}}$	24.4%	-1.0%	7.9%	32.3%	1.8%	-2.7%	24.1%	-10.6%	-4.0%	18.1%	-2.8%	-10.4%	6.4%	14.7%
SPE	Rank	17	16	17	18	17	17	18	5	10	18	14	11	14.8	4.1
	$\overline{RMSE_{vsEucl.}}$	23.5%	-2.2%	4.7%	35.0%	1.5%	-2.0%	27.8%	-11.3%	-4.7%	21.2%	-5.6%	-10.7%	6.4%	16.0%

Table 4: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Katowice station (RMSE) using out-of-sample extension.

Method	Method	CO			NO_2			O_3			PM_{10}			SO_2		
		+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day
average	LIM	**0.30**	**0.30**	**0.31**	**14.99**	**15.50**	**16.02**	**18.49**	**19.94**	**20.90**	**30.14**	**30.19**	**31.47**	**9.16**	**9.15**	9.45
		+0.0%	+0.0%	+0.0%	+1.1%	+0.8%	+0.5%	+0.1%	+0.2%	+0.3%	+0.1%	+0.3%	+0.2%	+0.7%	+0.4%	+0.5%
	FA	**0.30**	**0.30**	**0.31**	**15.14**	**15.64**	**16.10**	**18.42**	**19.94**	**20.76**	**30.61**	**30.66**	**31.93**	**9.25**	**9.28**	9.61
		+0.0%	+0.0%	+0.0%	+1.1%	+0.7%	+0.5%	+0.0%	+0.3%	+0.2%	+0.1%	+0.2%	+0.5%	+0.3%	+0.3%	+0.2%
	PCA	**0.29**	**0.30**	**0.31**	**14.89**	**15.47**	**16.07**	**18.57**	**19.95**	**21.11**	**30.19**	**30.80**	**32.28**	**9.11**	**9.07**	9.40
		+0.0%	+0.0%	+0.0%	+0.1%	+0.3%	+0.4%	+0.4%	+0.5%	+0.6%	-0.1%	+0.2%	+0.3%	+0.7%	+0.3%	+1.0%
α-stand.	LIM	**0.28**	**0.37**	**0.40**	**12.88**	17.56	19.45	**15.28**	20.80	22.95	**24.22**	34.49	38.95	**7.16**	9.04	**9.97**
		+3.7%	+5.7%	+2.6%	+2.2%	+1.0%	+0.3%	+0.0%	+0.4%	+1.1%	+4.1%	+2.2%	+2.4%	+1.4%	+0.2%	+0.4%
	FA	**0.28**	**0.36**	**0.39**	**12.90**	17.50	19.46	**15.24**	20.51	22.36	23.45	34.06	**38.17**	**7.10**	9.10	**9.99**
		+7.7%	+2.9%	+2.6%	+1.7%	+1.9%	+1.4%	-0.2%	-0.2%	+0.4%	+0.8%	+1.7%	+1.9%	+0.1%	+0.7%	+0.7%
	PCA	**0.27**	**0.36**	**0.39**	**12.74**	**17.28**	**19.23**	**15.31**	20.66	22.79	23.1	**33.19**	**37.23**	**7.12**	9.15	**10.05**
		+3.8%	+2.9%	+2.6%	+0.4%	+0.2%	+0.1%	-0.3%	-0.1%	+0.6%	+1.4%	+0.1%	-0.5%	+1.3%	+1.8%	+2.0%
$\alpha\beta$-stand.	LIM	**0.22**	**0.35**	**0.38**	**12.84**	**18.74**	**21.04**	**15.96**	22.71	24.6	**25.32**	39.44	44.30	**7.15**	9.96	**10.90**
		+0.0%	+2.9%	+2.7%	+1.4%	+0.4%	+0.1%	-1.2%	-0.4%	+0.7%	-0.2%	+1.9%	+2.0%	-0.1%	-1.2%	-0.4%
	FA	**0.23**	**0.35**	**0.39**	**12.85**	18.89	**21.03**	**15.88**	22.41	24.23	**25.31**	39.52	43.78	7.33	10.27	11.25
		+4.5%	+2.9%	+2.6%	+2.1%	+1.5%	+1.3%	+0.2%	-0.1%	-0.1%	-0.4%	+1.9%	+2.0%	+0.0%	-0.5%	+0.0%
	PCA	**0.22**	**0.34**	**0.38**	**12.83**	**18.77**	**21.09**	**16.04**	22.65	24.37	**25.56**	38.8	43.79	7.33	10.18	**11.13**
		+0.0%	+0.0%	+0.0%	+0.7%	+0.8%	+1.9%	+0.5%	+0.2%	+0.6%	-0.6%	-0.4%	+0.1%	+2.1%	+1.9%	+1.3%

Table 5: Results of methods for the daily average forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Cieszyn station (RMSE) using out-of-sample extension.

Method	Method	CO			NO_2			O_3			PM_{10}			SO_2		
		+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day
average	LIM	**0.31**	**0.29**	**0.29**	**8.34**	**8.30**	**8.56**	**20.51**	21.39	22.54	24.72	24.31	25.61	**7.67**	**7.69**	**8.09**
		+3.3%	+0.0%	+0.0%	+1.7%	+1.1%	+0.5%	+3.2%	+2.6%	+1.3%	+10.9%	+7.7%	+12.0%	-2.0%	-1.9%	-2.3%
	FA	**0.32**	**0.30**	**0.29**	**8.54**	**8.61**	**8.72**	**20.19**	21.20	22.53	26.06	25.30	26.36	**7.75**	**7.67**	**8.15**
		+0.0%	-6.3%	-6.5%	+1.4%	+0.3%	-1.1%	+2.0%	+1.9%	+1.3%	+8.4%	+3.3%	+9.0%	-8.4%	-7.3%	-4.9%
	PCA	**0.32**	**0.31**	**0.31**	**8.68**	**8.72**	**8.90**	**21.04**	21.69	22.94	24.75	24.55	26.15	**8.23**	**8.19**	**8.74**
		+0.0%	+3.3%	+3.3%	+3.2%	+3.2%	+3.5%	+1.7%	+1.5%	+0.6%	+0.9%	+2.3%	+2.3%	+5.2%	+5.0%	+5.7%
α-stand.	LIM	**0.24**	0.43	**0.46**	4.69	10.28	12.61	**11.56**	24.96	30.47	17.43	34.32	33.88	4.80	**9.64**	**10.69**
		+14.3%	+22.9%	+24.3%	+4.2%	+2.2%	+2.8%	+1.8%	+3.5%	+1.7%	+65.1%	+36.4%	+21.5%	-0.2%	-1.7%	-2.2%
	FA	**0.25**	0.53	**0.59**	4.74	10.28	12.22	**11.48**	24.49	28.91	16.46	35.16	43.03	4.83	**9.77**	**10.72**
		+13.6%	+17.8%	+18.0%	+10.7%	+11.0%	+12.9%	+3.5%	+1.7%	-2.0%	+59.7%	+42.6%	+54.1%	+6.2%	+5.9%	+7.3%
	PCA	**0.31**	0.52	**0.56**	4.69	10.29	13.22	**11.72**	24.25	29.62	17.19	33.78	35.67	5.01	**9.84**	**10.95**
		+29.2%	+30.0%	+16.7%	-2.3%	-2.6%	-0.6%	+1.8%	+0.3%	+0.2%	+3.7%	+6.6%	+9.6%	+1.6%	-1.2%	-5.7%
$\alpha\beta$-stand.	LIM	**0.18**	**0.35**	**0.39**	5.00	**10.66**	12.90	**12.11**	24.95	30.35	13.93	31.27	32.72	5.50	**10.77**	**11.72**
		+20.0%	+12.9%	+11.4%	+4.4%	+0.9%	+0.4%	+1.6%	+2.8%	+2.6%	+10.4%	+15.3%	+9.7%	+0.5%	+0.5%	-0.7%
	FA	**0.22**	**0.44**	**0.48**	4.96	10.89	**12.79**	**12.18**	25.43	31.09	15.95	34.68	33.95	5.86	**11.00**	**11.65**
		+15.8%	+22.2%	+20.0%	+7.1%	+7.8%	+9.4%	+4.3%	+4.4%	+2.3%	+34.5%	+36.1%	+22.2%	+5.8%	-1.2%	-1.8%
	PCA	**0.19**	**0.37**	**0.39**	4.86	**10.29**	**12.18**	**12.50**	24.48	29.65	17.97	34.16	36.19	5.46	**10.71**	**11.64**
		+11.8%	+8.8%	+0.0%	-2.0%	-7.6%	-12.7%	+7.3%	+1.3%	+0.3%	-2.9%	+10.9%	+8.1%	-4.0%	-0.7%	-0.1%

Table 6: Results of methods for the hourly forecasting values for CO, NO_2, O_3, PM_{10} and SO_2 for Wodzislaw Slaski station (RMSE) using out-of-sample extension.

Method	Method	CO			NO_2			O_3			PM_{10}			SO_2		
		+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day	+ 1 day	+ 2 day	+ 3 day
average	LIM	**0.36**	**0.37**	**0.37**	**11.97**	**12.1**	**12.14**	**17.86**	**19.26**	**20.45**	**48.21**	**48.4**	**48.98**	**12.14**	**12.1**	**12.25**
		+0.0%	+2.8%	+2.8%	+0.3%	+0.6%	+0.9%	+0.4%	+0.1%	+0.1%	+7.9%	+2.4%	+1.5%	+0.1%	+0.2%	+0.0%
	FA	**0.37**	**0.38**	**0.38**	**12.23**	**12.26**	**12.24**	**17.96**	**19.34**	**20.34**	**48.81**	**49.09**	**49.79**	**12.31**	**12.22**	**12.43**
		+0.0%	+2.7%	+2.7%	+0.1%	-0.1%	-0.2%	-0.2%	-0.4%	-1.6%	+7.2%	+2.4%	+1.3%	-1.0%	-0.5%	+0.2%
	PCA	**0.36**	**0.37**	**0.37**	**12.08**	**12.30**	**12.34**	**17.96**	**19.4**	**20.79**	**48.53**	**49.44**	**49.82**	**12.41**	**12.46**	**12.63**
		+0.0%	+0.0%	+0.0%	-0.2%	+0.0%	+0.5%	-0.3%	+0.3%	+1.5%	-0.3%	+0.2%	+0.3%	+0.2%	+0.6%	+0.9%
α-stand.	LIM	**0.32**	**0.44**	**0.46**	9.86	12.85	**13.73**	**14.94**	20.97	22.85	39.33	58.32	63.65	**10.32**	**13.59**	**14.38**
		+3.2%	+2.3%	+2.2%	+1.4%	+0.5%	+1.4%	+0.5%	+1.5%	+0.8%	+7.2%	+6.5%	+7.2%	+1.2%	+1.6%	+1.1%
	FA	**0.32**	**0.44**	**0.46**	9.87	12.8	**13.72**	**15.19**	20.88	**22.4**	39.21	58.35	63.66	**10.31**	**13.67**	**14.55**
		+3.2%	+2.3%	+2.2%	+0.0%	-0.9%	-0.6%	+1.3%	+0.6%	-0.5%	+7.4%	+5.5%	+4.8%	-0.2%	-0.1%	-0.4%
	PCA	**0.32**	**0.44**	**0.46**	9.76	12.84	**13.73**	**14.79**	20.76	22.88	**38.97**	58.47	62.76	10.35	**13.62**	**14.65**
		+3.2%	+2.3%	+2.2%	-0.1%	+0.2%	+0.9%	-0.6%	+0.1%	+1.1%	+0.0%	+0.4%	-1.0%	-0.1%	+0.4%	+0.5%
$\alpha\beta$-stand.	LIM	**0.29**	**0.44**	**0.47**	**9.72**	**13.74**	**14.72**	**15.80**	22.67	24.53	41.76	63.44	68.96	10.57	**15.31**	**16.35**
		+0.0%	+0.0%	+2.2%	+0.5%	+1.7%	+1.1%	+1.7%	+0.5%	+0.8%	+9.6%	+9.8%	+11.1%	+0.2%	+0.7%	+0.5%
	FA	**0.29**	**0.44**	**0.47**	**9.83**	**13.74**	**14.79**	**16.03**	23.32	24.98	40.71	62.38	67.77	10.54	**15.04**	**16.24**
		+0.0%	+0.0%	+0.0%	+0.5%	+0.1%	-0.1%	+0.3%	+3.1%	+2.4%	+7.1%	+7.3%	+6.9%	+0.9%	+0.3%	+0.9%
	PCA	**0.29**	**0.44**	**0.47**	**9.78**	**13.63**	**14.68**	**15.84**	22.62	24.69	40.85	62.18	66.65	**10.38**	**14.75**	**16.19**
		+0.0%	+2.3%	+2.2%	+0.2%	-0.2%	+0.2%	-0.1%	+0.1%	+0.0%	-0.4%	-0.1%	-1.6%	+0.8%	+0.7%	+0.9%

Table 7: Common Air Quality Index bands definition

BAND	CO [8h moving avg.]	SO_2 [1h]	O_3 [1h]	PM10 [1h]	NO_2 [1h]
0	[0,5000]	[0,50]	[0,60]	[0,25]	[0,50]
1	[5001,7500]	[51,100]	[61,120]	[26,50]	[51,100]
2	[7501,10000]	[101,300]	[121,180]	[51,90]	[101,200]
3	[10001,20000]	[301,500]	[181,240]	[91,270]	[201,400]
4	≥20001	≥501	≥241	≥271	≥401

Table 8: Ratio of NO_2 forecasts with errors expressed in CAQI bands.

		+1 day		+2 day			+3 day		
Station	Method	0 band	1 band	0 band	1 band	2 band	0 band	1 band	2 band
Katowice	Fractional	98.26	1.74	92.68	7.32	0.00	89.43	10.48	0.09
	LIM	98.35	1.65	91.45	8.40	0.14	88.30	11.36	0.33
	FA	98.40	1.60	92.21	7.69	0.09	88.08	11.30	0.62
	PCA	98.04	1.96	92.20	7.70	0.10	88.85	10.78	0.37
	Euclidean	95.31	4.69	89.36	10.36	0.28	84.66	14.58	0.76
Cieszyn	Fractional	98.60	1.40	96.27	3.03	0.70	94.02	5.23	0.75
	LIM	98.46	1.54	94.98	3.78	1.24	92.74	6.69	0.57
	FA	98.32	1.68	95.39	4.47	0.14	93.00	6.81	0.19
	PCA	98.10	1.90	94.81	3.88	1.31	93.09	5.83	1.08
	Euclidean	98.60	1.40	94.39	5.52	0.09	92.07	7.74	0.19
Wodzislaw	Fractional	98.72	1.28	96.49	3.46	0.05	94.94	5.01	0.05
	LIM	98.91	1.09	95.95	3.91	0.14	94.78	5.02	0.20
	FA	98.67	1.33	95.69	4.11	0.19	94.55	5.31	0.15
	PCA	98.81	1.19	96.00	3.85	0.15	94.78	5.02	0.19
	Euclidean	98.37	1.63	95.51	4.29	0.20	93.63	6.18	0.19

Table 9: Ratio of O_3 forecasts with errors expressed in CAQI bands.

		+1 day		+2 day			+3 day		
Station	Method	0 band	1 band	0 band	1 band	2 band	0 band	1 band	2 band
Katowice	Fractional	96.94	3.06	91.01	8.99	0.00	88.27	11.73	0.00
	LIM	96.34	3.66	87.68	12.32	0.00	84.74	14.96	0.45
	FA	96.29	3.71	87.94	12.01	0.15	85.56	14.28	0.23
	PCA	96.24	3.76	88.64	11.36	0.00	85.33	14.43	0.37
	Euclidean	93.98	6.02	85.59	14.41	0.00	82.67	17.07	0.39
Cieszyn	Fractional	96.04	3.96	88.75	11.00	0.38	82.19	17.11	1.05
	LIM	96.74	3.26	87.53	12.28	0.29	82.24	17.18	0.87
	FA	96.98	3.02	88.33	11.48	0.29	81.85	17.39	1.15
	PCA	96.27	3.73	88.60	11.16	0.36	80.98	18.74	0.42
	Euclidean	96.41	3.59	88.76	11.20	0.14	83.22	16.40	0.57
Wodzislaw	Fractional	98.05	1.95	93.39	6.24	0.56	90.40	8.86	1.12
	LIM	97.81	2.19	91.60	8.40	0.00	87.20	12.76	0.05
	FA	97.95	2.05	91.15	8.85	0.00	87.54	12.41	0.14
	PCA	97.25	2.75	91.12	8.88	0.00	87.42	12.48	0.14
	Euclidean	96.75	3.25	90.86	8.76	0.57	87.62	11.25	1.13

Table 10: Ratio of PM10 forecasts with errors expressed in CAQI bands.

		+1 day				+2 day				+3 day			
Station	Method	0 band	1 band	2 band	3 band	0 band	1 band	2 band	3 band	0 band	1 band	2 band	3 band
Katowice	Fractional	91.70	7.37	0.89	0.04	66.74	28.46	4.71	0.09	64.26	29.72	5.83	0.19
	LIM	91.71	7.45	0.80	0.05	86.57	11.29	2.00	0.14	86.25	11.28	2.29	0.19
	FA	92.15	7.10	0.70	0.05	87.52	10.16	2.18	0.14	86.53	10.95	2.33	0.19
	PCA	91.86	7.34	0.75	0.05	86.91	10.72	2.28	0.09	86.32	10.89	2.61	0.19
	Euclidean	91.96	7.29	0.75	0.00	87.44	10.42	2.05	0.09	86.23	11.26	2.47	0.04
Cieszyn	Fractional	93.68	5.86	0.45	0.00	89.50	9.16	1.34	0.00	87.82	10.31	1.87	0.00
	LIM	94.00	5.44	0.51	0.05	89.49	8.57	1.94	0.00	88.93	8.65	2.42	0.00
	FA	93.41	5.84	0.71	0.05	90.26	7.92	1.72	0.10	89.61	8.41	1.98	0.00
	PCA	91.90	7.10	0.90	0.10	88.68	8.59	2.52	0.21	87.76	9.08	3.05	0.10
	Euclidean	93.78	5.56	0.61	0.05	90.10	8.46	1.45	0.00	89.32	8.75	1.93	0.00
Wodzislaw	Fractional	94.06	5.70	0.23	0.00	90.84	7.85	1.26	0.05	89.86	8.51	1.63	0.00
	LIM	94.02	5.60	0.37	0.00	91.10	7.16	1.56	0.19	90.06	7.78	1.93	0.23
	FA	94.09	5.53	0.37	0.00	91.18	7.09	1.59	0.14	89.83	8.23	1.84	0.10
	PCA	93.22	5.86	0.83	0.09	90.40	7.13	2.28	0.19	89.28	8.06	2.42	0.23
	Euclidean	94.75	4.83	0.42	0.00	91.38	7.11	1.37	0.14	90.43	7.82	1.56	0.19

Figure Captions

Figure 1. The explorative forecast procedure diagram.

Figures

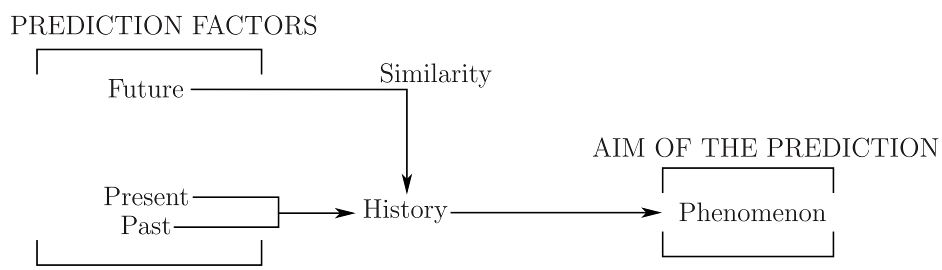

Figure 1: The explorative forecast procedure diagram.