



Wind speed and global radiation forecasting based on differential, deep and stochastic machine learning of patterns in 2-level historical meteo-quantity sets

Ladislav Zjavka¹

Received: 28 February 2022 / Accepted: 20 September 2022
© The Author(s) 2022

Abstract

Accurate forecasting of wind speed and solar radiation can help operators of wind farms and Photo-Voltaic (PV) plants prepare efficient and practicable production plans to balance the supply with demand in the generation and consumption of Renewable Energy (RE). Reliable Artificial Intelligence (AI) forecast models can minimize the effect of wind and solar power fluctuations, eliminating their intermittent character in system dispatch planning and utilization. Intelligent wind and solar energy management is essential in load scheduling and decision-making processes to meet user requirements. The proposed 24-h prediction schemes involve the beginning detection and secondary similarity re-evaluation of optimal day-data sequences, which is a notable incremental improvement against state-of-the-art in the consequent application of statistical AI learning. 2-level altitude measurements allow the identification of data relationships between two surface layers (hill and lowland) and adequate interpretation of various meteorological situations, whose differentiated information is used by AI models to recognize upcoming changes in the mid-term day horizon. Observations at two professional meteorological stations comprise specific quantities, of which the most valuable are automatically selected as input for the day model. Differential learning is a novel designed unconventional neurocomputing approach that combines derivative components produced in selected network nodes in the weighted modular output. The complexity of the node-stepwise composed model corresponds to the patterns included in the training data. It allows for representation of high uncertain and nonlinear dynamic systems, dependent on local RE production, not substantially reducing the input vector dimensionality leading to model over simplifications as standard AI does. Available angular and frequency time data (e.g., wind direction, humidity, and irradiation cycles) are combined with the amplitudes to solve reduced Partial Differential Equations (PDEs), defined in network nodes, in the periodical complex form. This is a substantial improvement over the previous publication design. The comparative results show better efficiency and reliability of differential learning in representing the modular uncertainty and PDE dynamics of patterns on a day horizon, taking into account recent deep and stochastic learning. A free available C++ parametric software together with the processed meteo-data sets allow additional comparisons with the presented model results.

Keywords Statistical modeling · Machine learning · Operator calculus · Partial differential equation · Laplace transform · PDE polynomial conversion · Inverse formulation · Angular representation

Arabic keywords الشبكة العصبية . التعلم الالي . نموذج معياري . المعادلة التفاضلية

✉ Ladislav Zjavka
ladislav.zjavka@vsb.cz

¹ Department of Computer Science, Faculty of Electrical Engineering and Computer Science, VŠB-Technical University of Ostrava, 17. listopadu 15/2172, Ostrava, Czech Republic

Introduction

Wind and PV power generation capacity is restricted by local or regional barriers (e.g., building structures or terrain obstacles). RE sources exhibit seasonal and diurnal addition with stochastic and high-frequency variability. Wind speed

and solar radiation are the complex results of global atmospheric convection processes, primarily caused by pressure and temperature differences or anomalies [1]. Wind and solar energy are the most important sources in off-grid systems located in mountains or coastal backcountry domains with inadequate conventional electricity supply, grid structure, or other inconsistent power sources. Wind turbine or PV panel arrangements are affected by a set of environmental factors, e.g. contour line, surface complexity, roughness or temperature stratification around the location. Their configuration can be established by modeling the characteristics of individual components or integrating information into the total energy production of a plant or farm [2]. The prediction of wind and solar parameters can be broadly divided into two main categories using [3]:

- Physical approach simulating wind formation processes through complex mathematical physics, such as numerical weather prediction (NWP).
- Statistical consideration using data-driven models for the related factors supposing the stochastic nature of the processes

Physical models are able to solve the equations of fluid mechanics and thermodynamics for the future atmospheric motion state in a certain time-step and resolution to simulate the variation tendency of meteo-factors. NWP systems attempt to model global or local weather patterns, starting with the observed input data [4]. They solve the hydrodynamic and thermodynamic equations of atmospheric flow in models initialized with specific starting and boundary constraints. This modeling approach is based on discretized conservation equations of mass, momentum, and energy in several atmospheric layers. This approach is usually efficient at long-term time horizons, but its applications are limited by a large number of multistep iterations and inadequacies in the short-term definitions. Difficulties are usually encountered in solving numerical models with higher resolution due to limitations in the complexity and time-consumption. NWP requires a lot of computing time and variables, which restricts its practicability in short-term forecasting [5].

Statistical data-driven techniques refer to the use of mathematical knowledge, such as statistics, chaos, and probability theory. AI forecasting models, formed with iterative learning, generally obtain better results in solving problems that cannot be defined analytically [3]. Their adaptability and robustness have higher potential in dealing with non-stationary disturbances. The statistical approach is usually more efficient than the physical simulation in a specific terrain allocation. Hybrid solutions usually include ensemble learning and metaheuristic optimization. Short-term AI prediction mostly outperforms the original NWP forecast accuracy, capturing temporal dynamics of the wind turbine with relationships

among the local meteo-quantities [6]. However, the loss in causality between wind speed or solar radiation and other meteorological factors results in shortcomings in statistical reliability using only historical data. AI usually considers meteorological factors only from the point of view of correlations, potentially losing a lot of useful physical information [7].

The proposed AI hybrid method, based on differential learning, combines numerical mathematics with neural computing to form progressive PDE-modular models. Its component PDE-formularization using evolving dynamic tree structures is supposed to solve some problems in statistical weather forecasting (e.g. model oversimplification, pattern variability and uncertainty, high-dimensionality reduction, feature extraction, data transformation, model composition, structure self-organization, etc.). The experiments start with two different level ground data sets whose differentiated information allows for an early recognition of changeovers in weather patterns in the 24-h prediction horizon. Specific meteo-quantities, recorded in 2 professional automatic meteo-stations (on lowland and peak bases), are examined to what extent they contribute to the overall forecast reliability. Self-detection of the most valuable data inputs from the days and advanced self-optimization reduce the structural complexity and uncertainties in the model initialization. The initial rough detection of applicable day training sequences is enhanced additionally by a reassessment processing of the interval sample records, one by one, according to pattern similarity in the determinate last observational time. No architecture hand design or training/testing initialization is necessary, as is common in deep learning.

The new designed differential learning is used in RE prediction with significant incremental innovations in its model definition, optimization and initialization (pre-processing):

- Periodical variables (radiation, temperature, etc.) are modeled using sine and cosine PDE-conversion functions, in combination with wind azimuth or time-stamp radius data (Section “[Differential learning—a novel hybrid neuro-math computing approach](#)”).
- A pareto list of the best input combination couples is initially determined in each layer learning cycle (separate model components are tested for the error minima) to be examined by inserting their node PDE modules into the gradually expanded complete model (Section “[Wind & solar day-ahead forecasting—methodology and data acquisition](#)”).
- Binary-tree structures (producing model PDE-components) are dynamically evolved and modified in each training cycle.

- Backpropagation is used in polynomial parameter post-adaptation in binary-tree nets to improve the model development (Section “[Differential learning—a novel hybrid neuro-math computing approach](#)”).
- Supplement power functions are used in model definition to refine the final PDE-formulation.
- Initially estimated record series of applicable daily training intervals are reassessed, one by one, according to pattern similarity in the last observation times (Section “[Wind & solar day-ahead forecasting—methodology and data acquisition](#)”).
- The C++ parametric application software with examined data sets is provided with publication ([C] D-PNN application C++ parametric software with Solar, Wind & Meteo-data sets: https://drive.google.com/drive/folders/1Q9m09bZ6LIQ2Up2_oJ0vDQpceYvXROrN).

The PDE-modular tree representation of weather pattern dynamics allows for free-standing RE statistical prediction in an increased mid-term day horizon, which is a significant advancement in comparison of recent AI (the problem solution is analogous to NWP). High-dimensionality data are sequentially processed (without information loss), in step-by-step model expansion and adaptation to the defined constraints. The Laplace transformation is automatically applied in 2-variable node PDE-conversion, which eliminates unexpected wind and solar parameter data fluctuations. The optimal model definition is automatically performed by selection from several types of base approximation function (rational, periodical, power), which allows a high diversity in model combination forms. Composite PDE modules are back-composed in multi-layer tree structures in the products of determined sub-PDE images in previous layers. Redundant components of PDE-transform products (using the same input variables) are automatically detected and removed from the structural model in learning to eliminate their undesirable interference and increase in complexity. Advanced external complement testing prevents the insertion of unacceptable non-generalizing PDE solutions.

State-of-the-art in wind and solar prediction

Decomposition techniques usually transform the original non-stationary series into a more applicable sub-series. Data decomposition analysis can be used to distribute the original series into specific frequency signal bands [8]. Empirical mode decomposition can recognize specific signal samples in unknown data that are forecasted and restored in an ensemble output of the target wind series using adaptive wavelet models in a particular time horizon [9]. Wavelet analysis decomposes the mother functions into several wave levels, which appear as the most critical parameters. A predictor

is constructed for each forecasting component, so the training time is multiplied. Causalities between wind & solar and other meteorological factors (e.g., pressure, temperature, etc.) can be recognized to divide data into several equivalent classes: central, chained, ring, tree, and network, according to the topological data structure. The Deep Learning (DL) network is dynamically formed with respect to the recognized causality data categories [7]. Meta-heuristics can adaptively estimate the optimal predictor parameters. These techniques can reconstruct missing information for specific data. Heuristics multi-objective optimizer usually begins with one initial estimate solution. The algorithm efficiently searches a defined search space to iteratively update a set of solution states and modify key parameters until the convergence criterion is satisfied. It can be applied to find the optimal weights of the ensemble forecast models or predictor parameters [10]. Fuzzy C-mean clustering can assess the difference in wind turbine output from day-ahead wind prediction. The principle of minimum distance allows for the selection of initial rough cluster centers in the data [11].

Ensemble learning is applied to integrate multiple parts of predictors in hybrid models and guarantee diversity. Ensemble strategies in forecasting can be used to form two types of AI models:

- weighted ensembles,
- learner-based aggregation,

The weighted output is a simple summary of single-model estimates. The learner approach combines multiple sets of forecast series generated by different predictors. The final output learner captures the relationships between the individual forecasts. Diversity-based methods partition data sets into training samples with different statistical distributions to form predictors, using bootstrap aggregation in bagging or boosting models:

- Boosting is applied in training base predictors to combine their best parameters, estimated by heuristics, in an aggregated output. It attempts to improve weak learners by building an integrated stronger predictor. It continuously modifies the distribution of data in the training of individual learners to achieve better performance. Data samples with higher prediction errors are assigned higher weights.
- Bootstrapping searches for a residual distribution function using resampling of the original data to construct Prediction Intervals (PIs). It repetitively draws commensurate samples to replace the original training data and partition them into several groups.

Multi-step prediction procedures based on machine learning are mostly applied in increased time horizons. Larger errors can be induced due to the incorrect initialization time

of wind or solar models in synoptic processes [12]. They are broadly categorized into the following:

- Recursive approach (using the iterative learning).
- Direct strategy (calculating the output series separately).

Chaos theory based on methods can capture linear and nonlinear characteristics in structural data. One-dimensional time series can be extended to a multi-dimensional matrix form using phase-space reconstruction to better represent the characteristics used as model inputs in forecasting [5]. The atmospheric stability factor is determined by its tendency to encourage or deter vertical air motion or flow. NWP data are grouped into several meaningful sets according to the atmosphere stability class, based on the comparison of observations at the prediction times. Gaussian Process Regression (GPR) is a nonparametric Bayesian modeling approach which attempts to detect complicated nonlinear relations between model inputs and output variables (predictors), based on the standard statistical distribution in data observations. Its models can recognize anomalies in ground wind speed [13]. Singular spectrum analysis can detect the periodic, quasi-periodic, and trend components in data series. The base and detail wind components are learned and predicted by separate models to generalize the inherent depth and long-term data relations. The Gray prediction method can solve high-uncertain problems with a significant lack of data or information, based on sample modeling [14]. The validity of the forecasting models for different RE generation scenarios can be determined by their testing approach. Decomposition-based models can analyze robustness and performance under different environmental conditions [15].

Probabilistic models provide information on the uncertainty of the calculated forecasts. They provide PIs where the point estimations are expected, as compared to the point forecasts produced by deterministic models. The distribution of output data can be estimated from the given training samples. Output errors, resulting from incorrect assumptions of distribution shapes, can be eliminated in this way [16]. Lower–upper bound output estimations can be used to construct PIs. Ensemble forecasts integrate several models using the different approaches or initial conditions. They can estimate a probability distribution of data of random weather quantities. Quantile regression forms ‘quantiles’ to estimate the conditional probability distribution of a random variable. Conditional quantiles are functions of independent explanatory variables, used as input to the model. Explanatory variables can be the result of an analysis of the NWP model. The kernel density can be used to estimate a probability distribution for random variables. A kernel is applied to each data point for a given variable to highlight its contribution and relevance in the density probabilistic function. After that, the sum of all kernel functions gives a smooth curve in

the final output to determine the kernel density. Probability solutions are easy to transform into a stochastic Markov chain frame [3]. Clustering and K-fold cross-validation in ensemble learning are used to generate multiple training subsets with the same distribution for the Bayesian base learner combining strategy to increase the diversity of input–output samples [17].

Wind & solar day-ahead forecasting—methodology and data acquisition

The available data record series were first examined using tentative initialization models. Their test error minima, obtained in the latest 8–12 h in the step-by-step increasing learning day interval, give the first rough approximate of the practicable init-time range in formation of prediction models. The applicability of predetermined sequence data was then re-evaluated according to Pearson’s Correlation Distance (CD), based on the Correlation Similarity (CS) measure (1), to particularize the most valuable samples, one by one, computed for the time-counterparts of the last day pattern. Generally, similarity is measured in the range 0 to 1, that is, the value of CS is ‘1’ if the vectors are exactly identical $P = Q$ and converges to ‘0’ if the vector P is totally different from Q.

$$CD' = 1 - CS = 1 - \frac{\text{cov}(P, Q)}{\sqrt{\text{var}(P)} \cdot \sqrt{\text{var}(Q)}} = 1 - \frac{\sum_{k=1}^n \left(p_{ij} - \frac{1}{n} \sum_{j=1}^n p_{ij} \right) \cdot \sum_{k=1}^n \left(q_{ij} - \frac{1}{n} \sum_{j=1}^n q_{ij} \right)}{\sqrt{\sum_{i=1}^n \left(p_{ij} - \frac{1}{n} \sum_{j=1}^n p_{ij} \right)^2} \cdot \sqrt{\sum_{i=1}^n \left(q_{ij} - \frac{1}{n} \sum_{j=1}^n q_{ij} \right)^2}} \quad (1)$$

$P(p_1, p_2, \dots, p_n)$ and $Q(q_1, q_2, \dots, q_n)$ are 2-point data in the space of n-dimensions, $\text{cov}(P, Q)$ and $\text{var}(P/Q)$ are covariance and variance of P, Q data

12 meteorological quantities were selected from the available observational data recorded in 10 min. series in the experimental 2-level automatic stations allocated in the Kopisty plain (240 m above sea level) and Milešovka peak (837 m attitude) points (Meteorological observational stations of Czech Academy of Sciences in Milesovka and Kopisty www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/meteorological-observatory-milesovka/milesovka-current-weather, www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/kopisty-weather-station/actual-weather) from 1 to 31 December, 2017:

- Global Radiation (GR), Height of the Condensation Output Level.

- Ground Temperature, Relative Humidity in 2 m, See Level Pressure.
- Wind Speed (WS) aver., Wind Direction aver., Time of Maximal Wind Speed, Wind Trajectory (integral).
- Visibility aver., Height of the 1st Cloudiness Base, Height of the 2nd Cloudiness Base.

The above variables were selected as input to the model with the highest relevance. Combining data from two-attitude-based stations allows modeling the relations in 2 atmospheric layers, which contributes to the acceptable mid-term forecasting accuracy in 24-h output time horizon using no NWP processing data ('Aladin' regional meso-scale NWP-model produced every 6 h ('Meteograms' are in Czech language) www.chmi.cz/files/portal/docs/meteo/ov/aladin/results/public/meteogramy/mhtml/m.html) (Fig. 3). Pattern variations or instabilities in the 2-level pattern zones indicate eventual frontal disturbances or break changeovers, whose evidence can be registered in the next day hours. The two-layer data relations are identified and incorporated by the multilevel forecasting model to reflect the dynamics of global progress [18]. The Clear Sky Index (CSI) factor, a fundamental ratio parameter formulated in the relative GR (2), was used to norm GR input–output.

$$\text{CSI} = \text{GR}(t)/\text{GR}_{\text{cls}}(t), \quad (2)$$

GR and GR_{cls} is the denotes measured and clear sky (peak) irradiance in time t series.

Figure 1 illustrates the identification search for the first applicable day-data sequences using the single-time modeling initialization in the gradually increasing starting-day interval in each validation experiment. Data records in the determined time range were after reassessed, one by one, in computing the sufficient pattern similarity with equivalent data points in the last-day hours (antecedent the prediction time). If tentative models cannot get a satisfactory approximation of the last-day test data (in the case of a frontal break), the start and end daytimes can be gradually shifted in searching appropriate learning multi-base records in the available set of data.

Figure 2 shows the day-ahead forecasting training scheme applicable in wind and solar parameter AI modeling without using NWP data ('Aladin' regional meso-scale NWP-model produced every 6 h ('Meteograms' are in Czech language) www.chmi.cz/files/portal/docs/meteo/ov/aladin/results/public/meteogramy/mhtml/m.html). The evolved models are secondary tested in the last hours and applied to unseen data to calculate their approximation of the response times of the target output in the response times of the following day [19]. If a prediction model cannot obtain a defined test error threshold, then its statistical prognosis is obviously impracticable and should

not be used as a basis in RE planning. Figure 3 shows a visualization of the situation in the 2-level localization area of the two flat and peak professional observation meteorological facilities (Meteorological observational stations of Czech Academy of Sciences in Milesovka and Kopisty www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/meteorological-observatory-milesovka/milesovka-current-weather, www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/kopisty-weather-station/actual-weather).

AI computing methods used in day-ahead statistics model development

Differential learning—a novel hybrid neuro-math computing approach

Differential Learning (DfL) is a novel hybrid soft-computing procedure, proposed by the author, which integrates ML with mathematical principles of solving Partial Differential Equations. Differential Polynomial Neural Network (D-PNN) is a DfL-based regression method that decomposes and solves the general linear PDE of the k th order into reduced 2-variable PDEs of the determined low order (3) converted in nodes. D-PNN allows modeling of complex high-nonlinear systems, described by a number of variables, which cannot be completely defined by conventional physical equations or represented by AI computing. Its model development is based on the optimal self-selection of applicable 2-inputs, so that it need not use the initial pre-extraction. D-PNN forms step by step multi-layer Polynomial Neural Network (PNN) tree structures, extending it node by node in the layers. Each selected node can produce PDE-model components, which are selected to be involved (or removed) in (or from) the sum output to progressively improve the approximation of target data. The gradual extension modeling procedure usually leads to the best acceptable solution in the use of the theorem defined by Kurt Goedel's incompleteness. PNN nodes, included in the back-computing extended D-PNN architecture, process the most proper 2-input data to pre-define and re-substitute the particular PDEs in the sum combinatorial model development, according to Operation Calculus (OC). The polynomial processing order evaluated for the component model substitutes for the PDE-transformation order [18].

$$A \frac{\partial^2 u}{\partial x_1^2} + B \frac{\partial^2 u}{\partial x_1 x_2} + C \frac{\partial^2 u}{\partial x_2^2} + D \frac{\partial u}{\partial x_1} + E \frac{\partial u}{\partial x_2} + Fu = G \quad (3)$$

, where A, B, \dots, G is the parametric coefficient of x_1, x_2 independent variables of the unknown u function.

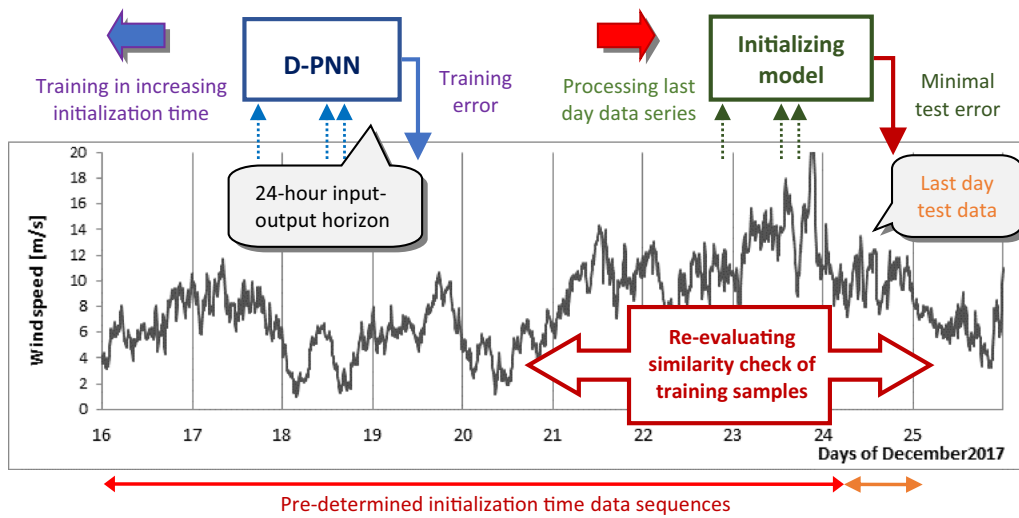


Fig. 1 Predetermination of the modeling initialization time and reappraising data samples, one by one

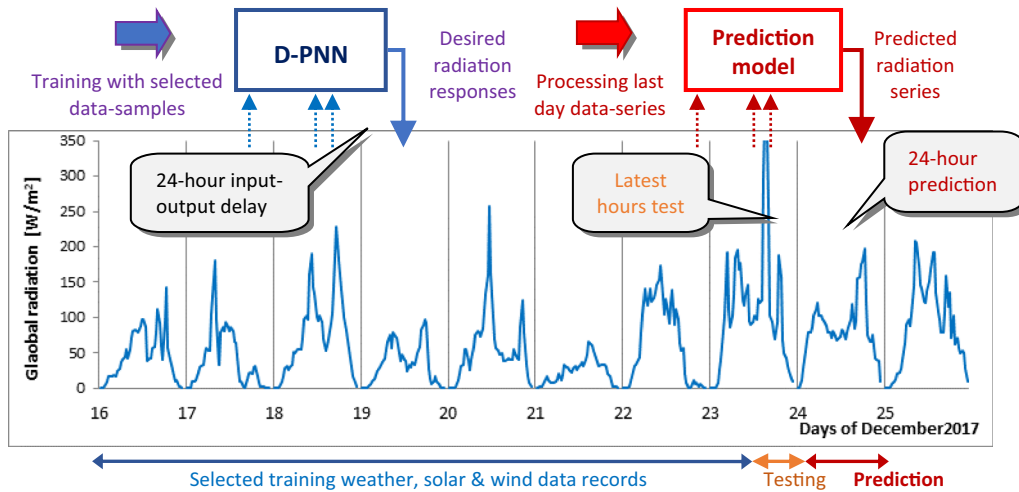


Fig. 2 Training and testing schemes in a day-ahead prediction horizon for wind and solar pre-assessed data

OC conversion of n th order derivatives for the $f(t)$ function is founded on the assumption that it can be replaced by the Laplace transformations (L-transforms) supposing defined initial conditions (4).

$$L\{f^{(n)}(t)\} = p^n F(p) - \sum_{k=1}^n p^{n-k} f_{0+}^{(k-1)} L\{f(t)\} = F(p), \tag{4}$$

$f(t), f'(t), \dots, f^{(n)}(t)$ is the continuous originals in $\langle 0+, \infty \rangle$, t is the complex and real variables, L is the transform

The $f(t)$ derivatives are L-transformed to define a system of linear Eq. (5), where the transform $F(p)$ is expressed with the imaginary conjugate p and separated in pure rational form (3).

$$F(p) = \frac{P(p)}{Q(p)} = \frac{Bp + C}{p^2 + ap + b} = \sum_{k=1}^n \frac{A_k}{p - \alpha_k} \tag{5}$$

B, C, A_k is the coefficients of elementary fractions, a, b is the polynomial parameters $\alpha_1, \alpha_2, \dots, \alpha_k$ is the simple real roots.

The resulting ratios correspond to the L-transforms of the original $f(t)$, to which can be applied the inverse L-operation of OC (6) to calculate the $f(t)$ function defined by the initial PDE (3).

$$F(p) = \frac{P(p)}{Q(p)} = \sum_{k=1}^n \frac{P(\alpha_k)}{Q_k(\alpha_k)} \cdot \frac{1}{p - \alpha_k} \quad f(t) = \sum_{k=1}^n \frac{P(\alpha_k)}{Q_k(\alpha_k)} e^{\alpha_k t} \tag{6}$$

$P(p), Q(p)$ is the multinomials of degree $s-1, s, \alpha_1, \alpha_2, \dots, \alpha_n$ is the simple real roots of $Q(p)$



Fig. 3 Low- and high-land locations of professional meteo-observational base stations in the west-Bohemian region

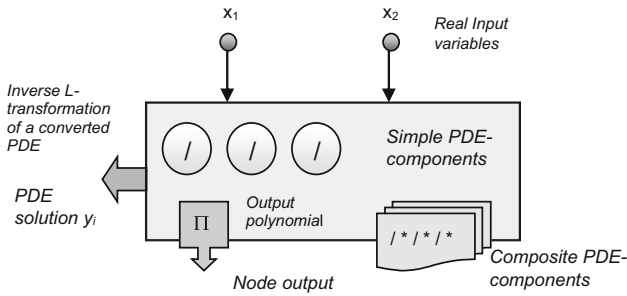


Fig. 4 D-PNN blocks form and solve sub-PDEs in the nodes to compose the component model

If $f(t)$ is supposed to be a circulator function, its derivatives are converted into the sine and cosine L transform and the original calculated by the reverse inverse L-operation (7). This expanded definition can apply the amplitudes and phases of periodical data variables to obtain the node function L-images, i.e. convert sub-PDEs into periodic functions.

$$f(t) = \sum_{k=1}^r \frac{P(\alpha_k)}{Q'(\alpha_k)} e^{\alpha_k t} + 2 \sum_{k=1}^s e^{\beta_k t} (a_k \cdot \cos \gamma_k t - b_k \cdot \sin \gamma_k t) \tag{7}$$

The inverse OC L-transform recovery operation is used for the reduced rational (6) or periodic Eq. (7), obtained by the first PDE-conversion. The sum of 2-variable u_k originals, formed in D-PNN nodes (Fig. 4), represents a final PDE-

model of the n-variable u function (3).

The expression of imaginary Euler conjugate numbers (8) is related to the OC definitions for the original $f(t)$ (6). Radius r (amplitude) can replace the rational component while phase (frequency) $\phi = \arctg(x_2/x_1)$ of variables $x_1, \times 2$ can replace the inverse L transform for $F(p)$.

$$c = \underbrace{x_1}_{\text{Re}} + i \cdot \underbrace{x_2}_{\text{Im}} = \sqrt{x_1^2 + x_2^2} \cdot e^{i \cdot \arctan\left(\frac{x_2}{x_1}\right)} = r \cdot e^{i \cdot \phi} = r \cdot (\cos \phi + i \cdot \sin \phi). \tag{8}$$

D-PNN main characteristics, allowing (D-PNN application C++ parametric software with Solar, Wind & Meteo-data sets: https://drive.google.com/drive/folders/1Q9m09bZ6LIQ2Up2_oJ0vDQpceYvXROrN):

- Splitting the n-variable general-order PDE into a defined set of reduced PDE converts
- Developing PNN structures by inserting node by node into the back-computing structure
- Producing PDE-components in each added PNN node to be involved in the sum model
- Several types of PDE conversions using OC base functions to define its computing frame
- Using L-transforms of PDE-derivatives and the inverse OC recovering of node originals

- (Re)selecting dynamically optimal 2-inputs to expand the parallel PDE-component model
- Non-downsizing significantly data dimensionality leading to an over-reduction in models
- A great variety in selecting the optimal combination of model components

Deep learning—Matlab Toolbox

Deep Learning is a computing method that learns patterns directly from data samples, utilizing its specific architecture designed in several types of structural layers, not relying on a predefined particular modeling approach. The Matlab Deep Learning Toolbox (DLT) provides a framework for the design and implementation of deep neural network algorithms. It uses the Long-Short-Term Memory (LSTM) network in sequence-to-sequence regression (Matlab—Deep Learning Toolbox (DLT) for sequence-to-sequence regression www.mathworks.com/help/deeplearning/ug/sequence-to-sequence-regression-using-deep-learning.html), usually consisting of these multilevel parts:

- Sequence layer of inputs.
- LSTM.
- Fully-connected layer.
- Drop-out layer.
- Fully-connected layer.
- Regression layer—output.

The key part of the LSTM regression is mostly the LSTM layer. The sequence layer feeds a succession of input time series into the LSTM structure. The LSTM layer is used to learn long-term data relations in time steps of sequenced series. The LSTM blocks combine their current state (c_{t-1} , h_{t-1}) with the next time X sequenced data to calculate the h_t output (that is, hidden inner state) and an update of the cell c_t state in a time $t - 1$, before the next update. This means that the information is represented by the hidden h_t state (i.e. output) and the cell c_t state. The dropout layer can set some random inputs to the zero values using a probability function to prevent model overfitting. The gradient of a function loss is calculated in consideration of the pre-assessed minimal batch length of sub-set data in training to optimize the updating of weights.

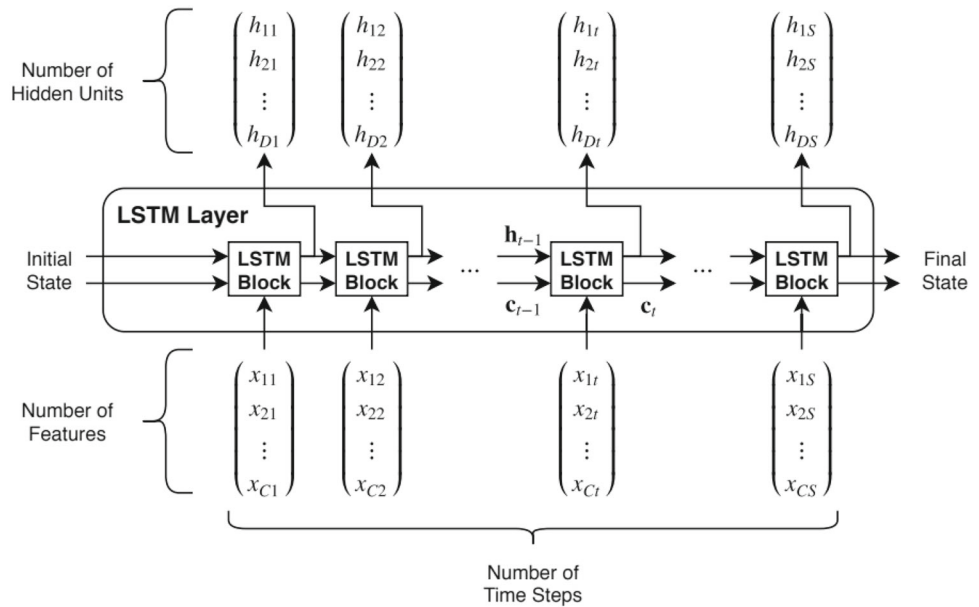
DLT uses LSTM networks to learn useful pattern representations from input–output data. Its networks integrate multiple nonlinear processing layers using simple operating and computing elements with connections inspired by biological nervous systems. DLT architectures consist mainly of a hand-defined many-layer replicate structure, including convolution and other layer types.

Statistics and machine learning—Matlab toolbox

The Matlab Statistics and Machine Learning Regression Tool-Box (SMLT) was used in the comparative evaluation of the statistical WS and GR forecasting results. SMLT includes and aggregates several efficient conventional, soft-computing, and stochastic-tree AI methods (Matlab—Statistics and Machine Learning toolbox (SMLT) for regression www.mathworks.com/help/stats/choose-regression-model-options.html):

- Linear regression (interaction, stepwise, robust)—uses linear equations with the parameters, which are easily adaptable to be interpreted in the most simplified processing form.
- Regression Tree (fine, medium, coarse)—uses the binary 2-branch form, which is easily to interpret, fast in fitting, and adaptation. Input data are processed step by step from the initializing root in a particular way to the terminal leaf considering the recognized states of predictors. Data are checked in all binary nodes to determine which of the two ways is applied as the computed way. The terminal leaf values correspond to the calculated output model response. Fine trees usually include many node leaves. These detail models may be less accurate in testing for an unlearned data validation set and fall into overfitting with essentially higher errors than those obtained in the training. Coarse trees use mostly a small amount of larger leaves, which usually do not yield a higher training accuracy, but they are more robust in applying to unseen data in the testing (forecasting).
- Support Vector Machine (SVM)—uses the linear, cubic, square, Gaussian, or Radial Basis Function (RBF) of a kernel form to define the specific transformation of data, which is initially applied before starting the learning process. The linear ε -vector training can be applied to eliminate or ignore output errors that are outside the interval defined by ε values (which are assumed to be zero). The support vectors are the computing output intervals whose errors are larger than the defined ε range.
- Gaussian Process Regression (GPR) uses a probability data distribution in a space of definition to calculate the output where base functions (e.g. linear, constant, zero, etc.) supply the prior mean model. Kernel functions (rational, exponential, squared exponential, quadratic, or matern) are applied to define response relations in the model output, according to a distance space vector for the predictors.
- Ensemble-Boosted/Bagged Trees (EBT, EBoosT / EBaggT) – aggregate outputs of several weak-learner tree-based models. The least squares principles of bagging, boosting, and bootstrapping can be applied to compose the final output ensemble using a probabilistic approach based on training data statistics (see Sec.2).

Fig. 5 The flow of input time series X with C features (channels) of length S in the LSTM layer



Principal component analysis (PCA), a built-in SMLT tool, did not produce better WS and GR average forecasts by processing the selected data inputs. The final forecast models, applied to the last available data, were detected from the best approximation results obtained in the testing interval hours.

Statistics data experiments in day-ahead wind & solar AI forecasting

The 2-level ground station data sets (described in Section “Wind & solar day-ahead forecasting—methodology and data acquisition”) were used in the development and verification of 2 different AI forecast models, using the proposed time-initialization and training-evaluation schemes (Figs. 1 and 2). If the pattern similarity (1) of the data records, included in the initial detection interval (Fig. 1), to those of the reference last time is lower than a defined correlation limit (0.5), then their samples are excluded from the training process. An extension in the initializing data range may be considered (in searching for other day points) in the event of insufficiency (or impracticability) in the applicable learning samples. D-PNN automatically searches for the best input couples by trying to include initially scored components (predetermined separately in each layer list) in the expansion of PDE-structural models, node by node. No hand-made parameter preset or architecture design (as used by DLT) was needed in the self-organizing and model-composing learning process of D-PNN and SMLT. After model complement testing and verification (in NWP or additional data comparison, if available), it can be applied in a definite 24 h computing of all-day forecast output series. This one-flush processing essentially reduces computational complexity, since the same

prediction model is applied sequentially to the last available day input series at each reference time. The optimal model is finally chosen by considering its test error minima (in all applied AI strategies), resulting from random or user-adaptive start-ups.

The demo graphs in Figs. 6 and 7 reflect the desired and prediction series of 0–24 h wind speed and 8–16 h irradiance in the 1st and 2nd examination days, produced by the day-ahead component D-PNN and comparison DLT and SMLT models in the 10-day monitor autumn–winter season.

All the AI compared models mostly properly approximate ramping series of the target GR or WS quantity in unsettled weather (Figs. 7 and 6), although the PDE modules allow D-PNN to better adapt its combinatorial solutions to sudden variances in pattern dynamic next-time progress. Characteristics of the following-day patterns remain unchanged and similar to those in the day illustrative graphs with a significant exhibit change coming through 21–22 December and the following days (Figs. 1 and 2) in a gusty and unsettled cloudiness period. The solar radiation series were normalized using the CSI ratio factor to eliminate day-period alterations related to the actual solar sky-horizon. CSI nominal series are calculated with respect to the ideal maximal ‘clear sky’ GR assumption values, allowing forecasting regardless of the day season. The target GR output is recovered from the nominal CSI forecast series [20]. Periodical or angular data quantities (such as GR, temperature, humidity, or wind azimuth) are automatically recognized and related to the time or amplitude represented by the L-conversion functions (sine, cosine) of the cycle (7) in the definition of D-PNN models [21]. All SMLT-evaluated methods are self-optimizing (as D-PNN) and do not require any hand design in architecture or training settings (as DLT does).

Fig. 6 17.12.2017, Milešovka, Wind speed day-ahead prediction RMSE: D-PNN = 1.516, DLT = 1.987, SMLT (GPR matern) = 1.490 [m/s]

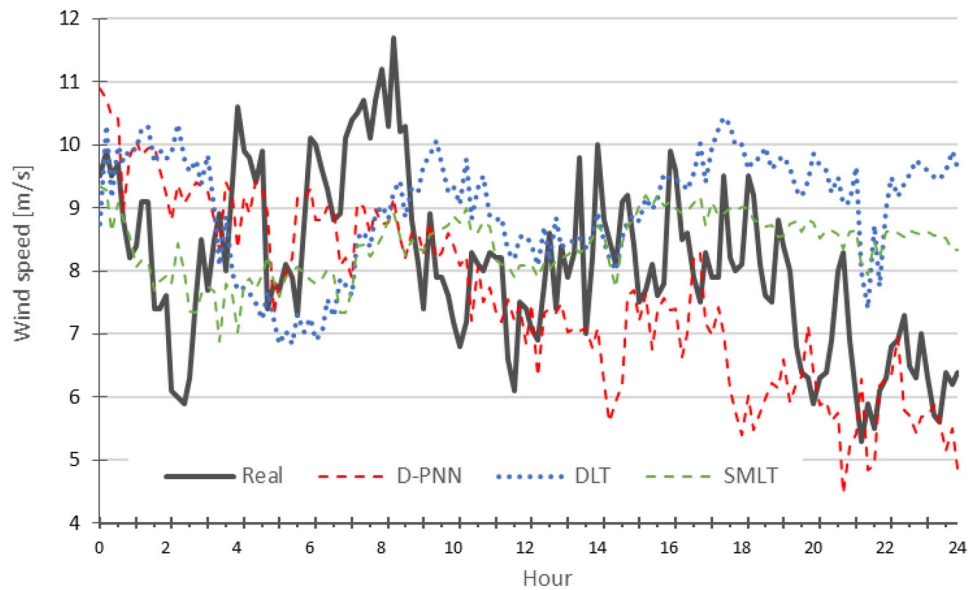
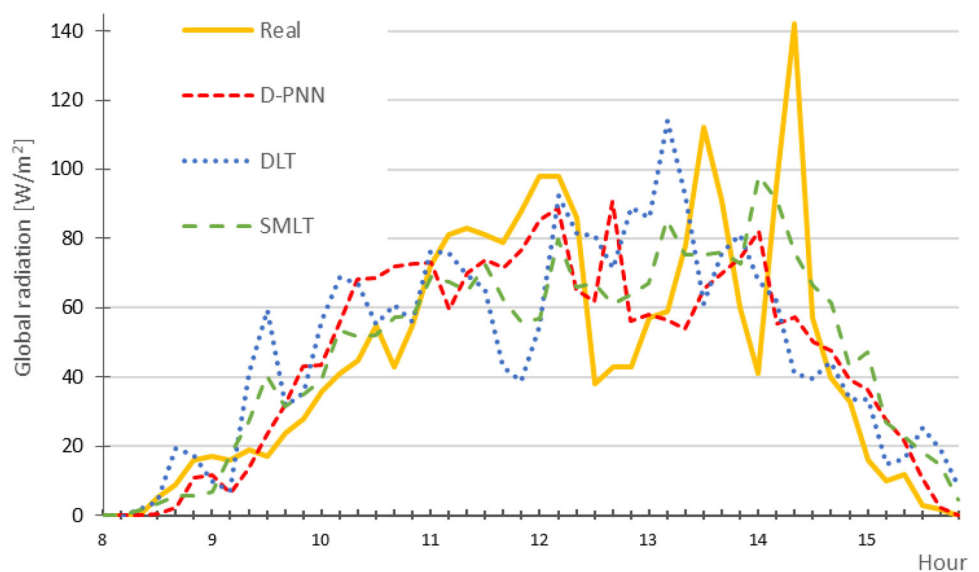


Fig. 7 16.12.2017, Kopisty, Global radiation day-ahead prediction RMSE: D-PNN = 21.46, DLT = 27.51, SMLT (EBoosT) = 20.42 [W/m^2]



Evaluation of day-ahead wind & solar forecasting using AI

Figures 8 and 9 resume average daily errors of the compared neuro- and soft-computing models in wind and solar parameter AI statistical day-ahead forecasting in the 10-day evaluating autumn–winter season.

Notable changes in wind data patterns (a growth in gusts) beginning on 21 December (Fig. 1) effect an accumulative increase in the model errors (Fig. 8). An evident drop in GR on 21 and its analogous characteristic turnover on 22 December (Fig. 2) results in larger prediction errors of the compared AI models in the same and subsequent days (Fig. 9). Some unexpected forecast errors can be eliminated by an extensive sequence selection search for applicable training records (1)

in a larger monthly database (in several years). The results of statistical models are determined primarily by the optimal pre-detection of data samples in training and testing sets [18]. Figures 10 and 11 compare the Pearson correlation determination R^2 coefficient of the single-day D-PNN, DLT and SMLT models applied throughout statistical forecasting in the 10-day experiment period. The significant decrease in the values of the R^2 coefficients of SMLT implies an undesirable WS output averaging of the probabilistic GPR and distribution EBT models on some days in the 24-h prediction horizon (Fig. 6). The results of the compared AI models are more or less the same in all 10-day GR and partly avg. WS 24-h forecast, with only slight variations in daily accuracy. More elaborate robust D-PNN modular solutions partly outperform DLT and SMLT in WS and also slightly in GR.

Fig. 8 Milešovka, 10-day Wind speed day-avg. prediction
RMSE: D-PNN = 2.038, DLT = 3.158, SMLT (GPR vs. EBT 7:3) = 2.571 [m/s]

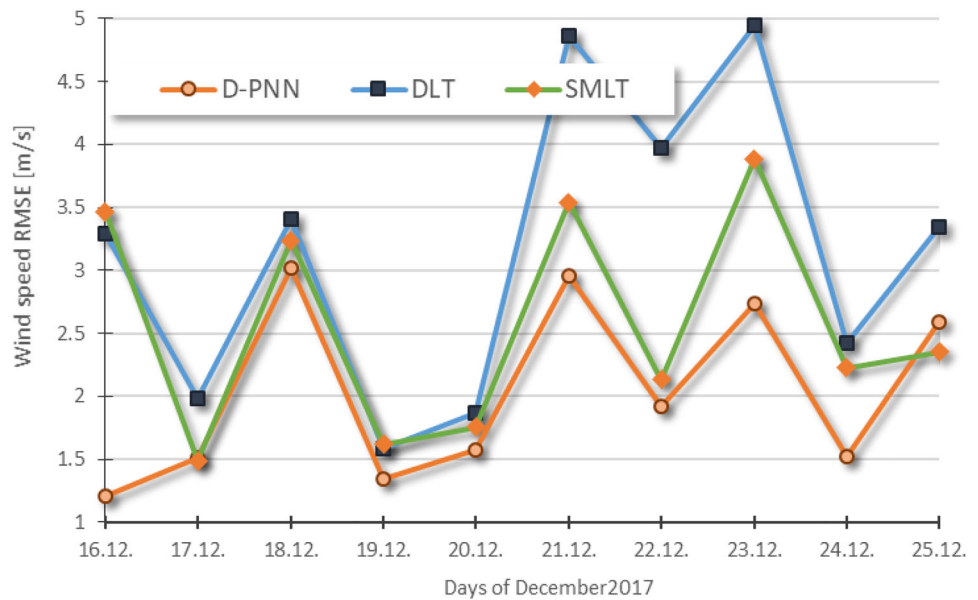
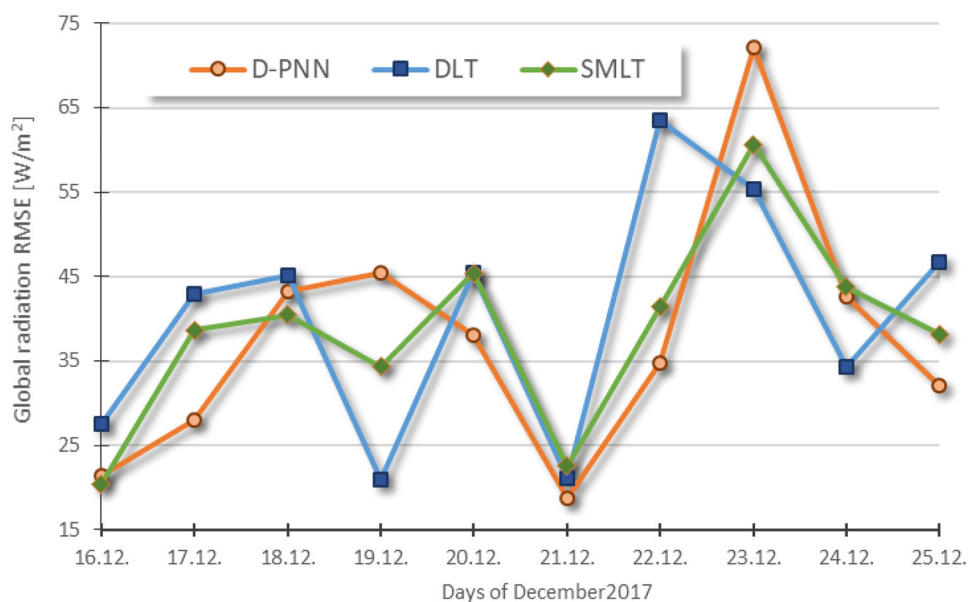


Fig. 9 Kopisty, 10-day Global radiation day-avg. prediction
RMSE: D-PNN = 37.68, DLT = 40.30, SMLT (GPR vs. EBT 5:5) = 38.59 [W/m²]



Each of the compared computing approaches gets a better day-ahead approximation in the cycle GR alterations than chaotic WS fluctuations. More leveled spatial observational points in data acquisition can naturally contribute to more reliable AI day-ahead forecasting in early statistics learning of unexpected change variations in multi-layer data correlations.

Figures 12 and 13 indicate the first identified modeling initial day time used in the starting search for similarity distance reinterpretation of applicable learning record sequences.

The forecast results of all applied AI techniques are mostly correlated, which denotes analogies in the forecasting methodology and model development. The PDE-modular approach of D-PNN obtains better accuracy in each mean

error evaluation (Figs. 10 and 11); however, SMLT and DLT outperform it slightly on some days. All the model-type day-initialization time is more or less analogous (Figs. 12 and 13) except DLT in some unsettled day patterns, including an overall extension/reduction in the same consequent day interval, which implies a second significant time point in weather over-change characteristics. DLT in general requires more precise estimations in training periods (Sec.3), compared to the more robust and resistant SMLT-probabilistic and PDE-modular form, to supply reliable 24-h forecasts. The D-PNN limits rest in a large search space in the combinatorial component alternatives (explosion), which allows, on the other hand, a great variety in modular component production to represent system uncertainty. The D-PNN error oscillations,

Fig. 10 Milešovka, 10-day Wind speed day-avg. prediction
Correlation determin. coeff.:
D-PNN = 0.145, DLT = 0.120,
SMLT = 0.062

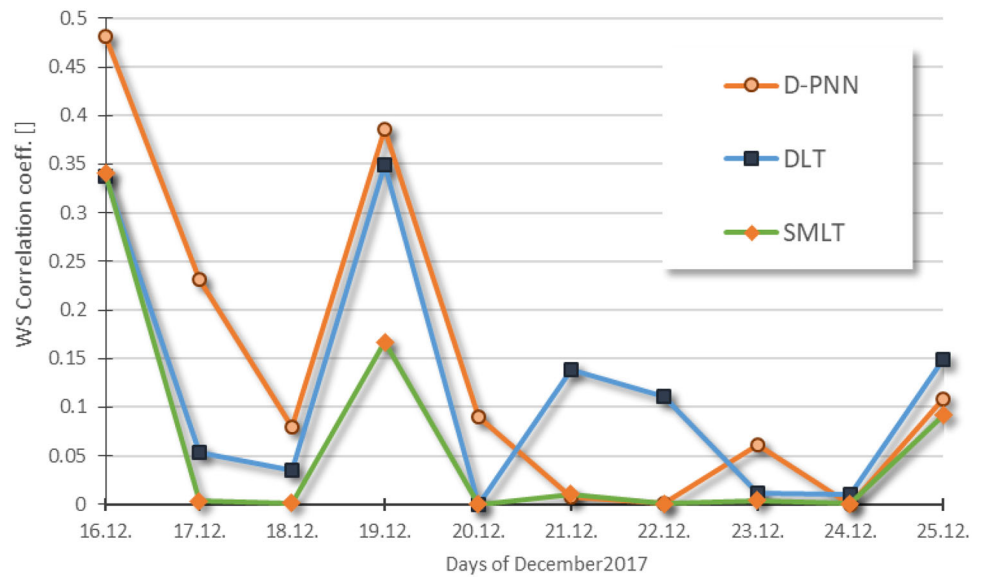
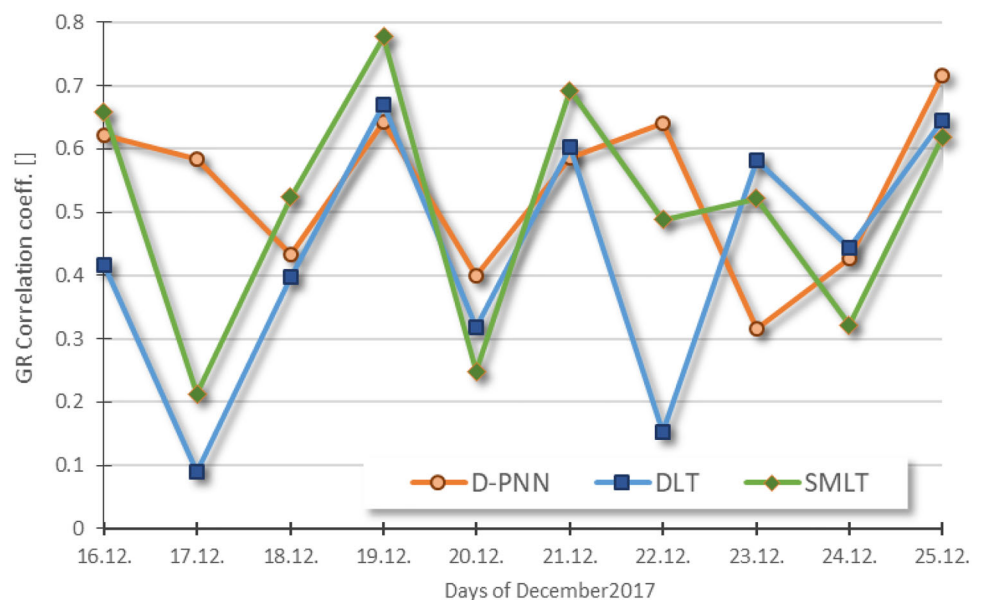


Fig. 11 Kopisty, 10-day Global radiation day-avg. prediction
Correlation determin. coeff.:
D-PNN = 0.537, DLT = 0.432,
SMLT = 0.506



resulting from unexpected pattern variances, are lower than the DLT or SMLT ones (Figs. 8 and 9), as the L-transformed data contribute to its more stable model output. The most successful SMLT methods (reaching test error minima) were found to be GPR and EBT using the probabilistic and distribution statistics approach (Section “[State-of-the-art in wind and solar prediction](#)”). The ratios of the final models were 7: 3 pro GPR chosen in the WS and 5: 5 in the GR forecasting. The stochastic GPR and EBT ensemble computing bases of SMLT ([E] Matlab—Statistics and Machine Learning toolbox (SMLT) for regression www.mathworks.com/help/stats/choose-regression-model-options.html) were approved to be very efficient in a fast and detailed approximation to GR real-day cycles in denormalized CSI forecast series (Figs. 7 and 9). Their WS prediction results are partly debased by

undesirable averaging some interval time series in the computed output, evident in the low R^2 correlation with target data (Fig. 10).

Discussion

Sudden variances in training and prediction data patterns are mainly induced by irregular surface interactions in air flow parameters and chaotic instabilities. These unstable oscillating states can result in uncertain practicability of the training data with respect to the 24 h model input delay (applied in forecasting) [4]. The adequate approximation of ramping series is largely related to predetermined learning samples in

Fig. 12 The predetermined initial numbers of learning day-data sequences used in the pattern sample similarity selection for wind speed model development

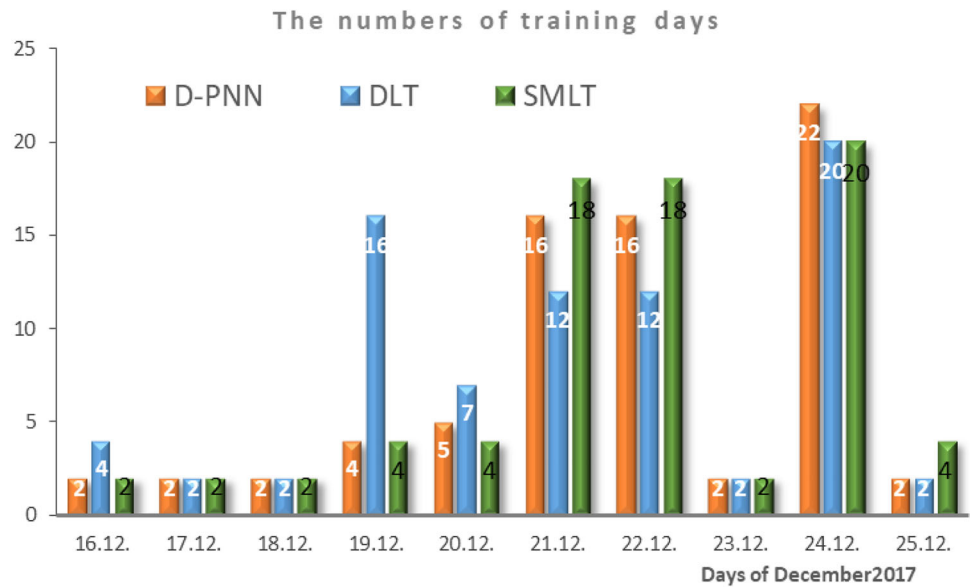
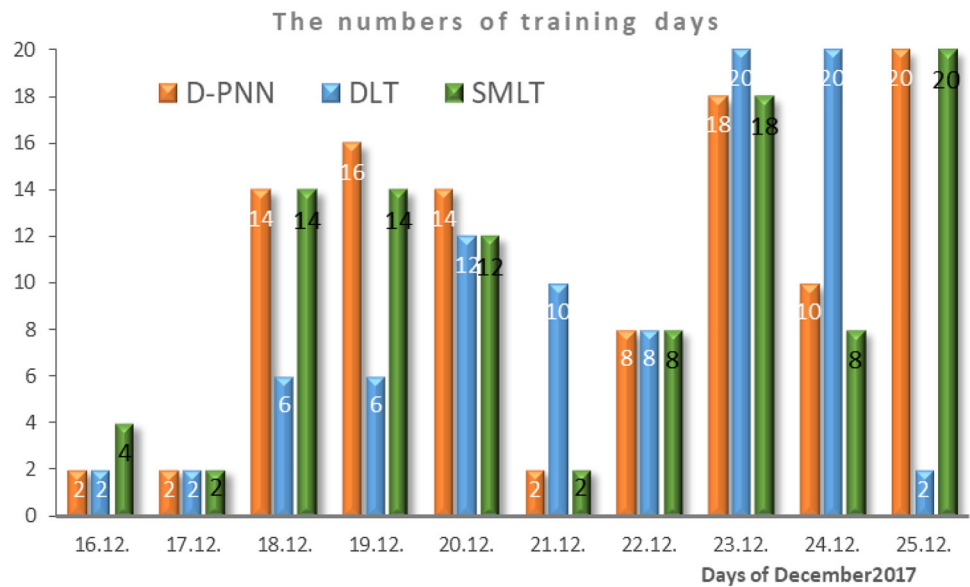


Fig. 13 The predetermined initial numbers of learning day-data sequences used in the pattern sample similarity selection for solar radiation model development



similar weather patterns. Their unexpected short-term fluctuation (Fig. 6) inducing rapid abrupt alterations in WS or GR results mainly in prediction troubles and model failures. The optimization extraction of training records (one by one) is determined by the chosen processing strategy in pattern similarity formulated by a correlation distance in the input space of n-dimensionality (1). A more complex or hybrid measure can improve the selective search for applicable training data. Statistical predictions can be wildly flawed on days of overnight frontal breaks in weather. The WS or GR patterns in the forecast time are totally uncorrelated with those in the test hours. A fixed threshold in test accuracy can be estimated in relation to previous misfires in day-ahead computing model forecasts or comparative NWP data. If AI models cannot be statistically validated in a comparative error limit or in an

NWP test, a series of transformed NWP cloud cover or wind parameters can provide the forecast [22]. Previous break over-changes in patterns can be examined and detected in a large-scale database in a point-by-point assessment in several initialization time ranges, applicable in daily training (Fig. 10). Additional input data that are delayed throughout the day cycle can be used, analogously to humidity [23] or electrical load [24].

Limitations and future scope of research

A more consistent approach in detection pattern similarity can be applied, involving all the determine time series (6–8 h.) in computing a comparative mean measure for

each individual training sample according to CSI and wind data. Additional normalized periodical quantities (temperature, relative humidity, etc.) can be used in this all-time re-evaluation process in future experiments. The initial pre-assessment of applicable data intervals is necessary, as statistical AI modeling is unable to represent (comprise, nowadays) all the weather dynamics in a global (earth) scale. Thus, a precise determination of optimal learning sequences is necessary in larger historical sets, and this procedure would naturally require extra processing time. On the other hand, model development is simplified in a sequential process. Validating pattern similarity between training/testing samples and NWP data can also contribute to optimized sample extraction and model verification, vital in frontal break changeovers. Although NWP tabular data records ([B] ‘Aladin’ regional meso-scale NWP-model produced every 6 h (‘Meteograms’ are in Czech language) www.chmi.cz/files/portal/docs/meteo/ov/aladin/results/public/meteogramy/mhtml/m.html) are not available, contrary to accessible free observational archives (Meteorological observational stations of Czech Academy of Sciences in Miletovka and Kopisty www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/meteorological-observatory-miletovka/miletovka-current-weather, www.ufa.cas.cz/en/institute-structure/department-of-meteorology/observatories/kopisty-weather-station/actual-weather). The D-PNN computing time is naturally higher, comprising a few minutes as compared to the DLT second-order model adaptation (but using its fixed-layer design with the entire input vector). However, D-PNN performs automatic day input/PDE module selection in its stepwise model optimization, which was shown to be efficient in representing weather dynamics and uncertainties. Extension of the D-PNN input vector is limited to dozens of variables (owing to the higher computing costs at this time). Processing and extracting time-lagged series would naturally improve the model performance. Component heuristics and model optimization algorithms can be further improved (in future work) to approach the standard soft-computing time.

Node-by-node development using the D-PNN binary back-selective architecture in the stepwise expanding additive model allows incremental learning. This means inserting new or removing useless PDE-components in the sum model re-adjusting according to an up-dated training set without resetting the present structure and model combination form. New assigned day samples can be additionally learned to readapt the same D-PNN model for each new situation in the next partial training step, to achieve by degrees greater robustness and stability for unknown prediction patterns and parameter uncertain variances. The complexity of PDE models is gradually increased and refined for new knowledge, in addition to retaining previously learned skills [25].

Conclusions

The effective all-day schemes in statistics GR and WS forecasting were implemented with the recent neuro- and soft-computing compared approaches. The advantages of all-day one-sequence procedures are apparent in their processing efficiency and computation time reduction, using single autonomous AI models that provide the complete day output series in the same fixed time horizon. This day-ahead iterative approach allows operational on-time forecasting with an acceptable reliability, the results of which are comparable to the intra-hourly or NWP-model results in most of the day-conducted experiments. The early-produced and transformed prognoses of GR or WS in the evening on a day horizon are helpful in planning and using the RE supply. Advantages of the physical NWP simulations are apparent in break changeovers in weather, where AI models with the 24-h input delay can be out-dated; however, their data are usually charged. Single-time AI models can correct the prognoses of more effective all-day forecasting schemes on the reduced horizon of a few hours in these doubtful situations, using an early warning notification based on available NWP pattern analysis. Inconsistent output estimations of AI models in subsequent hours may denote their incompetence in the applicable statistical prediction and alternative usage of NWP. Parametric modeling C++ software, historical solar, wind and weather sets are available free in data repositories to allow further comparative reinterpretation of the forecast procedure and model performance ([C] D-PNN application C++ parametric software with Solar, Wind & Meteo-data sets: https://drive.google.com/drive/folders/1ZA8KcvDEDm-i7ifVe_hDoS35nI64-Fh?usp=sharing).

Acknowledgements The work was supported by SGS, VSB-Technical University of Ostrava, Czech Republic, under the grant ‘Parallel processing of Big Data IX’ [No. SGS2022/12]. No conflict of interest or ethical issues. Informed consent and ethical approval. Compliance with Ethical Standards. 100% Authorship sole contribution.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yang D, Wang W, Xia X (2022) A concise overview on solar resource assessment and forecasting. *Adv Atmos Sci* 40:1–12
2. Wang Y, Zou R, Liu F, Zhang L, Liu Q (2021) A review of wind speed and wind power forecasting with deep neural networks. *Appl Energy* 304:1–25
3. Bazionis IK, Georgilakis PS (2021) Review of deterministic and probabilistic wind power forecasting: models, methods, and future research. *Electricity* 2:13–47
4. Lorenz E, Kühnert J, Heinemann D, Pagh Nielsen K (2016) Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions. *Progress Photovolt* 24:1–10
5. Jiang P, Wang B, Li H, Haiyan Lu (2019) Modeling for chaotic time series based on linear and nonlinear framework: application to wind speed forecasting. *Energy* 173:468–482
6. Meka R, Alaeddini A, Bhaganagar K (2021) A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables. *Energy* 221:1–13
7. Zhang Z, Qin H, Liua Y, Wang Y, Yao L (2019) Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction. *Energy Convers Manag* 192:37–51
8. Qian Z, Pei Y, Zareipour H, Chene N (2019) A review and discussion of decomposition based hybrid models for wind energy forecasting applications. *Appl Energy* 235:939–953
9. Santhosh M, Venkaiah C, Vinod Kumar DM (2018) Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. *Energy Convers Manag* 168:482–493
10. Liu H, Lv X, Chen C, Wu X, Liu M (2019) Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. *Energy Convers Manag* 195:328–345
11. Yang M, Shi C, Liu H (2021) Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy* 218:1–10
12. Hanifi S, Liu X, Lin Zi, Lotfian S (2020) A critical review of wind power forecasting methods—past, present and future. *Energies* 13:1–24
13. Hoolohan V, Tomlin AS, Cockerill T (2018) Improved near surface wind speed predictions using gaussian process regression combined with numerical weather predictions and observed meteorological data. *Renew Energy* 126:1043–1054
14. Liu H, Mi X, Li Y, Duan Z, Yan Xu (2019) Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression. *Renew Energy* 143:842–854
15. Gonzalez-Sopena JM, Pakrashib V, Ghosha B (2021) An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renew Sustain Energy Rev* 138:1–22
16. Afrasiabi M, Mohammadi M, Rastegar M, Afrasiab S (2021) Advanced deep learning approach for probabilistic wind speed forecasting. *IEEE Trans Industr Inf* 17:720–727
17. Gang Wang R, Jia JL, Zhang H (2020) A hybrid wind power forecasting approach based on bayesian model averaging and ensemble learning. *Renew Energy* 145:2426–2434
18. Zjavka L (2020) Photo-voltaic power daily predictions using expanding pde sum models of polynomial networks based on operational calculus. *Eng Appl Artif Intell* 89:1–10
19. Zjavka L, Mišák S (2018) Direct wind power forecasting using a polynomial decomposition of the general differential equation. *IEEE Trans Sustain Energy* 9:1529–1539
20. Zjavka L (2021) Photo-voltaic power intra-day and daily statistical predictions using sum models composed from l-transformed pde components in nodes of step by step developed polynomial neural networks. *Electr Eng* 103:1183–1197
21. Zjavka L (2021) Photovoltaic energy all-day and intra-day forecasting using node by node developed polynomial networks forming pde models based on the l-transformation. *MDPI Energies* 14:1–14
22. Zjavka L, Krömer P, Mišák S, Snášel V (2017) Modeling the photovoltaic output power using the differential polynomial network and evolutionary fuzzy rules. *Math Model Anal* 22:78–94
23. Zjavka L, Sokol Z (2018) Local improvements in numerical forecasts of relative humidity using polynomial solutions of general differential equations. *Q J R Meteorol Soc* 144:780–791
24. Zjavka L, Snášel V (2016) Short-term power load forecasting with ordinary differential equation substitutions of polynomial networks. *Electr Power Syst Res* 137:113–123
25. Zjavka L (2021) Power quality multi-step predictions with the gradually increasing selected input parameters using machine-learning and regression. *Sustain Energy Grids Netw* 26:2–10

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.