Renewable and Sustainable Energy Reviews 186 (2023) 113662

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Near real-time predictions of renewable electricity production at substation level via domain adaptation zero-shot learning in sequence

S.Y. Teng [a,*], C. Cambier van Nooten [b,c], J.M. van Doorn [b], A. Ottenbros [d], M.A.J. Huijbregts [d], J.J. Jansen [a]

[a] Institute for Molecules and Materials, Radboud University Nijmegen, the Netherlands
[b] Alliander N.V., P.O. Box 50, 6920 AB, Duiven, the Netherlands
[c] Institute for Computing and Information Sciences, Radboud University Nijmegen, the Netherlands
[d] Department of Environmental Science, Radboud Institute for Biological and Environmental Sciences, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Renewable energy
Missing data
Zero-shot learning
Transfer learning
Energy network

ABSTRACT

With the urgency of the transition to a resilient low-carbon economy, the monitoring and prediction of regional renewable energy generation over time have become increasingly important. The difficulties of renewable energy data transfer between multiple stakeholders have also caused elevated the need for more trustworthy data analytics within the energy grid. However, only a few voltage transformation facilities in the grid (i.e. substations) contain complete information about the renewable energy generated within the region. A large number of incomplete-information substations with fully-missing renewable energy data limits the analysis and policy-making related to renewable energy. This work studies the potential to transfer information from perfect-information substations to incomplete-information substations with fully-missing renewable energy data. To preserve the practicality of renewable energy prediction, a domain adaptation for zero-shot learning in sequence (DAZLS) strategy is proposed for fully-missing renewable energy prediction in substations. DAZLS is a model agnostic technique which can utilize any base model within its framework for the prediction of renewable energy generation. Using total measured power and weather information (solar irradiation and wind speed) in information-complete substations (8831 timestamps for 10 substations) within the Netherlands, we developed a model to predict solar and wind power from energy producers associated with information incomplete substations via additional real-time weather data, metadata information (e.g. geospatial position, existence and capacity of renewable facilities). Using DAZLS, the average root-mean-squared error for prediction (RMSEP) is 0.07, while that of a default transfer learning model is 0.70. This meant that renewable energy sources in information-incomplete substations could be reliably monitored using weather data, meta-data and physical data, resulting in lesser investment in power meters. This approach was demonstrated for the highest frequency prediction possible in the grid, with a near-real-time frequency of 15 min. Our method can be effectively used for renewable energy grid optimization, planning and analysis.

1. Introduction

The United Nations has provided sustainable development goals (SDGs) to be reached by 2030 [1]. SDG 7 is targeted to provide affordable, reliable, sustainable and modern energy for all. In this context, the Netherlands pledged to combat climate change by reducing at least 40% of greenhouse gas (GHG) emissions and increasing renewable energy shares to at least 32% compared to 1990 levels [2]. Since the year 2000, the Netherlands has promoted energy transition initiatives [3] to achieve sustainable development via micro novel configurations, patchworks of regimes, and evolving socio-technical landscapes. Kemp et al. [4] discussed that the challenges of national governance for energy transition include ambivalence about goals, uncertainty about cause-effect relations, distributed power of control, political myopia, determination of short-term steps for long-term change and the danger of lock-in to new systems. With more investments in onshore wind energy [5] and the diffusion of photovoltaic systems among consumers [6], more renewable energy innovations are emerging [7].

The stability of the energy generation process is affected by the

* Corresponding author.
E-mail address: SinYong.Teng@ru.nl (S.Y. Teng).

https://doi.org/10.1016/j.rser.2023.113662
Received 27 August 2022; Received in revised form 6 August 2023; Accepted 16 August 2023
Available online 25 August 2023

List of abbreviations

Abbreviations	*Definition*
AI	Artificial intelligence
CORAL	Correlation Alignment
DAZLS	Domain Adaptation for Zero-Shot Learning in Sequence. The strategy proposed in this work (pronounced as "dazzles")
GHG	Greenhouse gases
HAL	An anonymized substation in the mid-north of the Netherlands
HFDP	An anonymized substation in the west of the Netherlands
LCA	Lifecycle assessment
LLS	An anonymized substation in the center of the Netherlands
MCA	Misestimates in carbon avoidance
MNZL	An anonymized substation in the mid-north of the Netherlands
NAP	Normal Amsterdam level
NARX	Nonlinear Autoregressive with eXogenous inputs
NRYN	An anonymized substation in the south of the Netherlands
openSTEF	Open Short-Term Energy Forecasting package. Sometimes also referring to the default unsupervised energy splitting method within the package
OWD	An anonymized substation in the east-north of the Netherlands
PCA	Principal component analysis
PLS	Partial least squares
PLS-PM	Partial least squares – Path modelling extension
Process PLS	Process partial least squares
RMSE	Root-mean-square error
RMSEP	Root-mean-square error for prediction
SDG	Sustainable development goals
WEW	An anonymized substation in the west-north of the Netherlands
WHF	An anonymized substation in the west-north of the Netherlands
WLS	An anonymized substation in the east-south of the Netherlands
WWF	An anonymized substation in the west-north of the Netherlands
XGBoost	eXtreme Gradient Boosting. A supervised machine learning model

equilibrium of energy demand and supply [8]. Energy security and stability has been a long-term concern in Europe [9,10] and many parts of the world (e.g. Africa [11,12], Brazil [13], China [14], Malaysia [15], Singapore [13], United States [16], Japan [17], etc.). Recent pandemic [18] and warfare [19] highlighted the importance of digitalization for energy systems. The COVID-19 pandemic has significantly impacted the energy sector by increasing energy consumption for plastic production, energy and waste due to hygiene supplies, and altered energy supply industries [20]. Further considering recent European energy politics during Ukraine warfare, the supply of natural gas has become unreliable and expensive [21]. This relates to consumers changing behaviour towards energy saving and consumption, creating interest in a transition towards clean and decarbonized energy [22]. With growing renewable energy sources as supplies, the requirements for predicting renewable energy generation have been increasing in interest [23].

Benefits of digitalization can improve aspects related to renewable energy integration [24], reduces energy consumption, decreases energy intensity and optimizes energy structure [25]. Baidya et al. [26] discuss that digitalization in energy research has benefits for grid data integrity, protection against cyber-attack, privacy protection, trust management, grid up-scaling, authentication, grid data monitoring, improved energy demand response, prosumer support, enhanced distributed energy system, open energy market design, and enhanced environmental impacts. Artificial intelligence (AI) algorithms [27,28] has also became an important part of renewable energy digitalization to reduce uncertainty for renewable energy quantification, stabilizing energy operation and management. The combination of high fidelity AI models, information infrastructure and hardware is also the key cornerstone for energy digital twin technologies [8], which may cover various scales of process level including nano (molecular level), micro (single operation level), meso (multiple operation level) and macro (regional level).

Near real-time frequency energy prediction is crucial for energy digital twin at macro-level as the intraday electricity trading system in Europe [9,10] (and many other regions [11,12]) is based on 15-min interval. Vaz et al. [13] demonstrated a Nonlinear Autoregressive with eXogenous inputs (NARX) model for photovoltaic system power production forecasting at 15-min interval, demonstrating usefulness of the algorithm over persistence model. Previous work Teng et al. [14] demonstrated the usefulness of hierarchical temporal memory coupled with dual-mode optimization for waste-to-energy forecasting and optimization with 15-min interval. O'Dwyer et al. [15] also demonstrate that optimal multi-vector energy systems can be coordinated via a digital twin with 15-min interval. From a broader view, near real-time frequency applications also aligns with various digital twin solutions towards the '15-min city' [16] with the goal of increasing urban efficiency, resilience and sustainability [17].

2. Literature review

Data analysis and data infrastructure are being developed for energy planning [18] and energy saving [19]. Machine learning algorithms have been used to predict renewable energy generation [20]. For example, multiple machine learning methods were used for wind energy forecasting [21] and solar irradiation [22]. As an example, Li et al. [23] used a multi-verse optimizer to optimize a support vector machine in predicting photovoltaic power generation. Díaz–Vico et al. [24] used a deep neural network for the predictions of solar and wind energy. Lasso, k-Nearest Neighbour, XGBoost, Support Vector Regression and Random Forest were evaluated by Demolli et al. [25] in providing accurate wind power forecasting. Despite the recent usage of machine learning [26], deep learning and statistical approaches [27] for direct renewable energy forecasting, most predictions rely on high-quality temporal information to carry out predictions.

Missing data commonly occurs in renewable energy systems and can be categorized as (i) partially missing values and (ii) fully missing data. The impact of missing data on renewable energy is critical as it deviates from the performance estimation of the system [28,29]. Further effects of missing data also create misestimates in renewable energy resources leading to bias during policy-making [30]. For example, missing data in renewable energy systems can directly affect the calculated impact in life-cycle assessment (LCA), resulting in a difference between LCA integrated certification and LCA of buildings [31]. Such missing data problems also make the establishment of an adequate digital twin for renewable energy systems highly challenging [32]. This limits the potential of using digital twins for applications such as augmentation of predictability, improvement of asset economics, optimization of unit control, and reduction of unplanned outages [33].

The high-frequency quantification of carbon dioxide emission mitigation by relying on renewable energy instead of fossil fuels is crucial for international energy decision-making. However, high-frequency monitoring of solar and wind power generated is commonly missing in regional substations due to (i) costs and management of additional sensors [34], (ii) privacy challenges related to different companies operating renewable energy systems [35], and (iii) data infrastructure difficulties on large-scale [36]. Therefore, there are commonly only a few perfect-information substations and many incomplete-information substations (with fully-missing renewable fraction data) for renewable energy monitoring. In incomplete-information substations, the total load is measured while there is no information about renewable energy generated. For such cases, renewable energy quantification is challenging and rule-of-thumbs estimation is commonly used [37]. Research on addressing partially missing values includes popular approaches of using data interpolation [38] or imputation [39] to manage missing values within a short time span. Apart from partially-missing values, problems related to fully-missing renewable energy data are often overlooked and rarely studied in the literature.

In the real grid management scenario, many substations often have fully missing data (i.e. information-incomplete substations) instead of partially-missing values, as the substation does not contain the sensors necessary for renewable energy data collection. In such cases, missing data imputation cannot be performed due to the fully missing historical renewable energy data. Related metrological analysis from Moustris et al. [40] proposed to use 7 representative locations in Greece to estimate the solar irradiation within the country. Although the work did not relate hourly-frequency solar irradiation to energy systems, it suggests that missing locations could be predicted from measurement point locations. Furthermore, Tasnim et al. [41] also recommended a cluster-based transfer learning method to be used to predict new stations from existing stations. Although the problem statements from these works are different, they suggest the potential of transfer learning (domain adaptation) and fixed representative locations for renewable energy prediction. Nevertheless, the application to provide near real-time predictions for fully-missing renewable energy generation data at the lowest substation level has never been addressed.

Domain adaptation [42] is a subfield within machine learning which deals with (i) different source and target marginal distribution on input data and (ii) same task on source and target domain. Domain adaptation is generally used for image recognition [43], language processing [44], and other traditional machine learning applications [45]. Although many researchers solve domain adaptation problems with deep neural network [46–48], domain adaptation can sometimes be elegantly solved via straightforward algorithms [49–51] depending on the application leading to improved model interpretability. For tasks related to natural language processing, Daume III [49] proposed using kernel to map augmented features to target domain. Blitzer et al. [51] proposed structural correspondence learning for domain adaptation. More recently, Sun et al. [50] proposed an approach based on correlation alignment (CORAL) for object detection and sentiment analysis. Kouw et al. [52] discussed that domain adaptation strategy can be generally categorized as sample-based, feature-based and inference-based methods.

The use of weather information, geospatial metadata, and physical metadata to simultaneously predict the energy generation of multiple target renewable energy generation via domain adaptation approach is still a research gap in the literature. This is an urgent problem because most energy monitoring substations do not record continuous data for the renewable fractions, causing fully-missing renewable energy data. In this work, we propose a framework to utilize metadata and apply transfer learning for fully-missing data by utilizing information from fixed representative substations. The framework relies on a new inference-based domain adaptation strategy via sequential boosting for the specific application of renewable energy prediction. With this, a high-level understanding of the underlying mechanism and performances of the regional renewable systems can be obtained to provide high-frequency reconciliation for environmental assessment or policy-making [53].

The main novelty of this work is that it is a model-agnostic framework to predict renewable energy generation in incomplete-information substations in near real-time frequency. The technique assigns information to 3 different parts (domain, adaptation, and physical information) of the model sequentially to achieve renewable energy prediction on target substations with no additional renewable energy meters required. This domain adaptation technique can be used on any data-driven model and is fully scalable to achieve renewable predictions towards sustainability. To provide a high-frequency prediction of renewable energy generation, perfect-information substations are selected and the expected performance on fully-missing data substations is validated via the leave-one-out method [54]. The combined use of weather data, geospatial and system metadata to perform zero-shot learning [55] is studied for 10 perfect-information substations in the Netherlands (see Fig. 1). Subsequent sections of this paper is arranged as methodology (Section 3), results (Section 4), discussion (Section 5) and conclusion (Section 6).

3. Methodology

3.1. Framework

The overall methodology of this work is illustrated in Fig. 2. The main task is to transfer knowledge from complete information to incomplete-information substations for renewable energy prediction. Firstly, exploration analysis is carried out using an unsupervised learning approach by combining principal component analysis (PCA) and the Ledoit-Wolf covariance shrinkage method to uncover the underlying covariance between all the substations.

Next, a proposed domain adaption for zero-shot learning in sequence (DAZLS) strategy was deployed to exploit the domain information, metadata and physical constraints of the problem. Multiple base models and moving window sizes are tested in DAZLS, while the best model is selected and used to quantify GHG emissions and compared to an unsupervised learning approach in the OpenSTEF package [56] which was used as default in the case study. The improvement of misestimates in carbon avoidance due to using renewables (as opposed to non-renewables) is then compared for the default OpenSTEF method and the best method obtained by DAZLS.

Difficult-to-predict substations were explored using a supervised explained variance approach using Process Partial Least Squares (PLS) as a *post-hoc* analysis. The most important pathway with the most sum of explained variances and a low number of nodes in the path is selected via the Pareto no-preference method. The identified pathway provides a sequential interpretation of the explained variances for the prediction.

3.2. Data

For analysis, the target substation will be rotated amongst the 10 studied substations (see Fig. 1) as a leave-one-out validation approach. A total of 8831 timestamps for 10 substations with 15-min frequency were used. Metadata is categorized into two types of metadata: (i) target metadata and (ii) source metadata (see Table 1). The input data of the model consist of weather data and the mandatorily measured total load at the substation. The purpose of the model is to provide a 15-min interval prediction for solar and wind power generated by using information from perfect information substations to predict fully missing target substations.

3.3. Explorative modelling and analysis

To explore the relationships between multi-block and multivariate input data of all substations (8831 timestamps for 10 substations), a combined principal component analysis [59] and Ledoit-Wolf covariance shrinkage estimator [60]. Firstly, for each block of input data the substation, the multivariate input data $(\boldsymbol{X})$ with number of samples (s) and number of variables (v) is dimensionally reduced using principal component analysis (PCA) to give its scores, $\boldsymbol{T}$. The general PCA

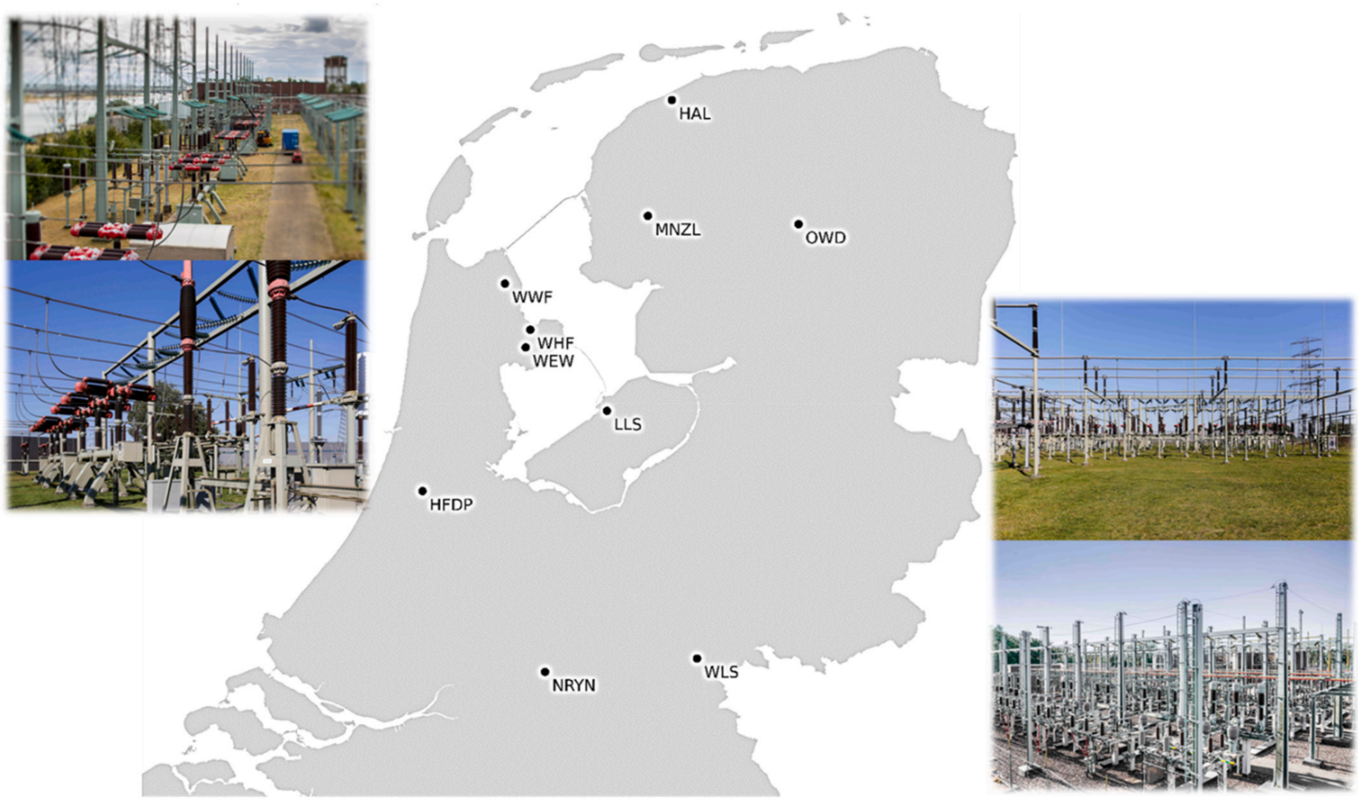

Fig. 1. The location of the studied 10 perfect-information substations in the Netherlands. The substations have their names encoded as HAL, MNZL, OWD, WWF, WHF, WEW, LLS, HFDP, WLS and NRYN. The four subpictures represent visual examples of substations in the Netherlands with their exact location anonymized.

equation can be provided as:

$$X = TP^T + E \tag{1}$$

This shows that the X can be decomposed into T and its transposed vector of the regression coefficient, P^T. Here, we consider only the dimension reduction to 1 principal component per block for simple interpretation and to fit the covariance shrinkage analysis. This means that the score matrix is $T = [t_1]$ with size $[s \times 1]$ and the regression coefficient matrix is $P = [p_1]$ with size $[v \times 1]$.

For blocks of data in each substation in $N = [1, \dots, n]$, the multi-block scores matrix Z can be constructed as the following, with the size $[s \times n]$.

$$Z = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{s,1} & \cdots & t_{s,n} \end{bmatrix} \tag{2}$$

Covariance is an important property of understanding the local control of data between blocks, it also provides insight into the effects of environmental variations [61]. Covariance shrinkage methods are used to provide a covariance estimator that is more well-conditioned and accurate than directly using the sample covariance matrix [60]. The Ledoit-Wolf covariance shrinkage estimator [60] provides a robust estimation of the covariance (Σ^*) by solving the minimization problem below.

$$\min_{\alpha,\beta,\delta,\mu} \frac{\alpha^2\beta^2}{\delta^2} \tag{3}$$

$$s.t. \quad \Sigma^* = \frac{\beta^2}{\delta^2}\mu I + \frac{\alpha^2}{\delta^2} S$$

$$\delta^2 = \alpha^2 + \beta^2$$

Where $\alpha, \beta, \delta, \mu$ are parameters to be optimized, S is the sample covariance matrix and $S = \frac{ZZ^T}{s}$. To complete the equation, I is an identity matrix of the same size as S. For interpretation, only the relative magnitude of the covariance matrix is taken into account. Hence, the absolute value of the covariance matrix, $|\Sigma^*|$ is min-max normalized between 0 and 1 and the lower triangle of the matrix is used as an adjacency matrix for an undirected graph as visualization.

3.4. Domain adaptation for zero-shot learning in sequence (DAZLS)

Zero-shot learning for renewable prediction on unseen incomplete-information substations is challenging and can involve models with highly parameterized and nonlinear methods, which devastates the possibilities for complete interpretation. Furthermore, the variance and bias trade-off problem with zero-shot learning causes more difficulties in prediction [62]. Often in prediction, when the variations in the data are well-predicted, there is an offset in the prediction and vice-versa. This work proposes DAZLS as a generic framework that is effective even for simple machine learning methods.

The task is to predict the output data (y_t) of the target substation using data from multiple sources of other information-complete substations. In such cases, target metadata (N_s), source metadata (M_s), input data (X_s), output data (y_s) will be available for the source substation (s). The corresponding target metadata (N_t), source data (M_t), and input data (X_t) are present for the target substation, however no output data (y_t) is available (see Fig. 3(a)). This problem formulates as a zero-shot learning problem [63] where the model has to carry out a prediction for the unseen target substation, using training data from other substations.

The DAZLS approach proposed here is a 3-step approach which consists of the domain model, adaptation model and physical correction model deployed in sequence (see Fig. 3(b)). The DAZLS strategy is agnostic to the base model of the domain model and the adaptation model, which means any data-driven model can be plugged and played

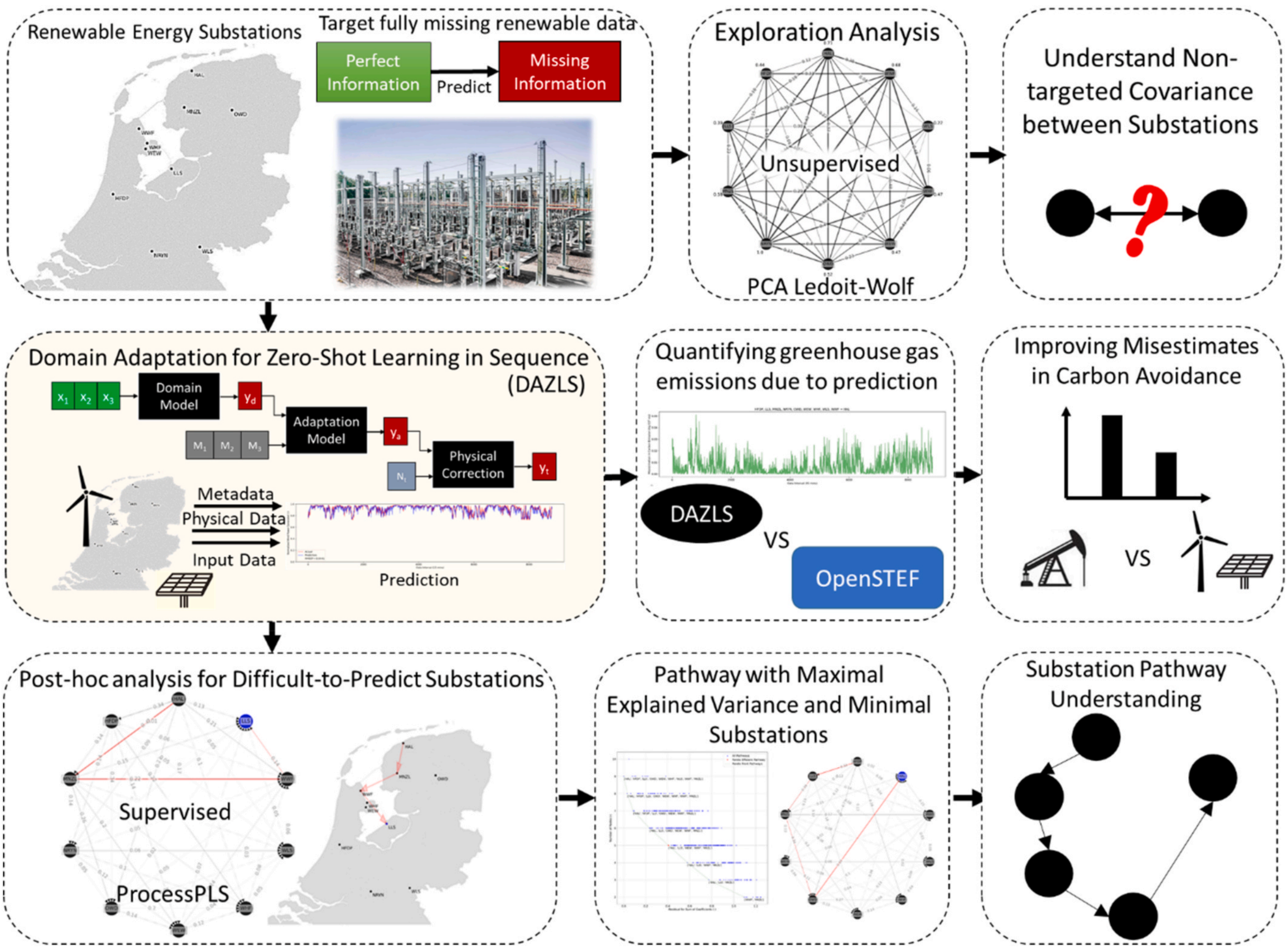

Fig. 2. Overall methodology for data analysis and model construction.

to serve this purpose. The physical correction model is specific to this task and can be easily adapted for similar applications. Firstly, the domain model (f_d) maps the input data (X_s) to the output data (y_s) in the information-complete source substations. The model gives predicted output data (y_d) as the following.

$$y_d = f_d(X_s) \tag{4}$$

Next, the predicted output data (y_d) from the domain model is concatenated with the source metadata (M_s) and acts as input for the adaptation model (f_a) to predict the output data (y_s) again. This step maps the input-output data in the domain model into the metadata domain and provides an output prediction (y_a) as shown below.

$$y_a = f_a([y_d, M_s]) \tag{5}$$

Lastly, physically impossible predictions are removed and renormalized using the physical correction module. In this task, there is binary target metadata for which solar or wind facility is present (B). For this, the element-wise multiplication of the output prediction from the adaptation model (y_a) and the binary vector is taken as the new output (y_t). Also, the predicted power for solar or wind is min-max renormalized to fit between the maximum (C_{max}) and minimum capacity (C_{min}) of the facility, constraining the prediction to fall out of the physical capacity range.

$$y_t = y_a * \mathbf{B} \tag{6}$$

These 3 steps complete the DAZLS strategy in using domain data, metadata and physical constraints sequentially to achieve zero-shot learning in incomplete-information substations. The base model for DAZLS that was considered in this work consists of commonly used supervised learning models including random forest regression, K-neighbors regression, multi-task lasso with internal cross-validation and partial least squares (PLS) with internal cross-validation. These base models were chosen with consideration of the balance between interpretability and nonlinear modelling capabilities. Furthermore, to account for historical dynamics in the data, a moving window is optimized within the pipeline and used to incorporate temporal information into all the base models. The window size of the mowing window is automatically optimized by validation. For validation of the substations, a substation-wise leave-out-out (LOO) approach [54] is used to cycle through each possible substation as the target block for the whole DAZLS pipeline. After considering a variety of base models and moving windows, a single best model is selected based on the prediction error to be used.

3.5. Important pathway exploration via process PLS

The Process Partial Least Squares (PLS) model [64] uses an inner and outer model for the supervised modelling of multi-block data. The Process PLS model is a multi-dimensional latent variable model that models explained variances (P^2) as effects between blocks in a directed acyclic graph structure. Each block of the data is modelled using a PLS model using the SIMPLS algorithm [65] and its network structure is similar to the PLS-PM (Partial Least Squares – Path Modelling Extension) model [66]. The Process PLS model has also been demonstrated to be

Table 1
Structure of metadata, input data and output data.

Variable	Rationale	Data source	Fall-backs
Target Metadata			
Maximum Capacity of Solar Power; Minimum Capacity of Solar Power	The capacity of solar energy generation facilities is a physical aspect that is independent of weather, load and geospatial location.	This metadata can be obtained from reported or engineering design documents of solar facilities and stored as a metadata table.	This metadata can be estimated via an unsupervised learning algorithm. Directly removal still results in a functional algorithm.
Maximum Capacity of Wind Power; Minimum Capacity of Wind Power	The capacity of wind energy generation facilities is a physical aspect that is independent of weather, load and geospatial location.	This metadata can be obtained from reported or engineering design documents of solar facilities and stored as a metadata table.	This metadata can be estimated via an unsupervised learning algorithm. Directly removal still results in a functional algorithm.
Is Solar Power present? (1 or 0)	The target location might have suitable weather for solar generation, however, no physical facility was built there.	The existence of solar generation facilities can be identified from the facilities connected to the grid in the substation region.	If this metadata is not provided, a prediction for solar power will always be made.
Is Wind power present? (1 or 0)	The target location might have suitable weather for wind generation, however, no physical facility was built there.	The existence of wind generation facilities can be identified from the facilities connected to the grid in the substation region.	If this metadata is not provided, a prediction for wind power will always be made.
Source Metadata			
Difference of Latitude in Substations; Difference of Longitude in Substation	The differences in the location latitude and longitude give geospatial correlation for the model.	The latitude and longitude can be pinpointed using global positioning systems [57].	This metadata can also be obtained using other web mapping systems.
The Variance of Historical Input Data	The historical input data (weather and total load), and the historical variance is used as an additional feature to gauge how noisy the substation is.	This is a feature engineering technique that can be performed on the input data.	Directly removal still results in a functional algorithm.
Standard Error of Mean in Historical Input Data	The historical input data (weather and total load), and the historical standard error of mean are used as an additional feature to gauge how easily the substation deviates from the mean.	This is a feature engineering technique that can be performed on the input data.	Directly removal still results in a functional algorithm.
Input Data			
Wind speed at the location	The wind speed provides direct weather correlation to the wind power generated.	This data can be obtained in near-real-time from the Royal Netherlands Meteorological Institute [58].	Other weather forecasting and monitoring services can be used.
Irradiation at location	Solar irradiation provides direct weather correlation to the solar power generated.	This data can be obtained in near-real-time from the Royal Netherlands Meteorological Institute [58].	Other weather forecasting and monitoring services can be used.
Total Load at location	The total load at the substation involves renewables and non-renewables.	This is a mandatory variable recorded in substations.	Data extrapolation is possible, however highly not recommended.
Hour-of-day	This feature is used to quantify the day-night cycle in the region.	Directly be obtained from a clock.	External clocks can be used.
Minute-of-Hour	This feature is used to quantify the short-term dynamics within the hour.	Directly be obtained from a clock.	External clocks can be used.
Output Data			
Solar Power Generated at Substation	The solar power generated is used for target fitting.	This variable is only available for information-complete substations.	Not provided for incomplete-information substations.
Wind Power Generated at Substation	The wind power generated is used for target fitting.	This variable is only available for information-complete substations.	Not provided for incomplete-information substations.

useful for multi-block process modelling in improving conditional interpretability [67]. The details of the full algorithm can be found in Ref. [64].

In this work, Process PLS serves as a supervised learning method to explore the most important pathway that is related to the target substation of interest. We define an important pathway from Process PLS as the path which gives the most sum of explained variances ($\mathscr{S}(p)$), which uses the least number of nodes ($\mathscr{N}(p)$), both depending on the selected pathway. The Pareto efficient point was computed using the no-preference method [68] as follows.

$$\min_{\boldsymbol{p}} \left\| \boldsymbol{f}(\boldsymbol{p}) - \boldsymbol{z}^{ideal} \right\| \tag{7}$$

$$s.t. \quad \boldsymbol{p} \in \mathscr{P}$$

$$\boldsymbol{f}(\boldsymbol{p}) = \begin{bmatrix} \mathscr{S}(\boldsymbol{p}) \\ \mathscr{N}(\boldsymbol{p}) \end{bmatrix}, \boldsymbol{z}^{ideal} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}$$

Where p is the specific pathway, $\mathscr{P}$ is the collection of all the possible pathways, and the sum of explained variances ($\mathscr{S}$) is computed by min-max normalizing $\sum_p P^2$ between 0 and 1. The number of nodes across the selected pathway is $\mathscr{N}(p)$ can be obtained by counting the number of elements in the specific pathway, $count(p)$ and min-max normalizing it similarly. This Pareto efficient pathway is the most important pathway to be studied for a specific substation, as it gives contributes to the most sum of explained variances using a small number of substations (data blocks). The most important pathway provides insight into the sequential combination of substations that gives the most correlation in a parsimonious way.

3.6. Greenhouse gas emissions

The greenhouse gas (GHG) emission includes sources from solar

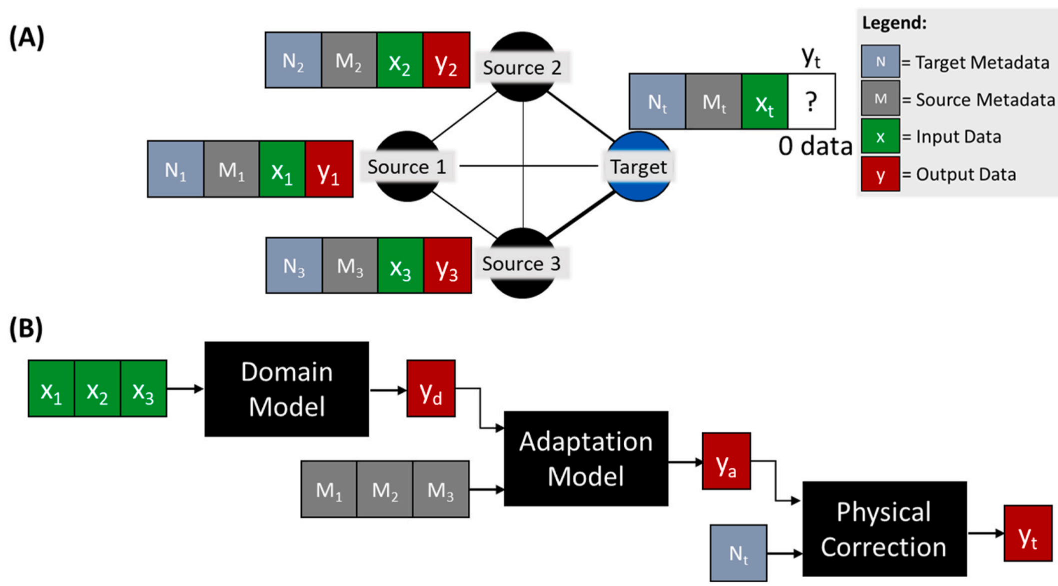

Fig. 3. (a) Illustration of the zero-shot learning problem for renewable energy quantification in an incomplete-information substation. (b) Diagram illustrating domain adaptation for zero-shot learning in sequence (DAZLS) approach.

energy, wind energy, natural gas and coal. The carbon intensity for the non-renewable average is taken from the Ecoinvent 3.5 database [69] and re-averaged using relative energy fraction from CE Delft's report to the ministry of infrastructure and water management [70].

The misestimates in carbon avoidance (*MCA*) due to generating renewables as opposed to non-renewables are calculated as follows:

$$MCA = (y_t - y_t') \bullet (CE_{NA} - CE) \tag{8}$$

In which CE_{NA} is the carbon emission related to the non-renewable average in the Netherlands, y is the ground truth output variable (wind and solar power generation), y' is the predicted output variable, CE is the corresponding carbon emission from Table A1.

4. Results

4.1. Substation explorative covariance analysis

From Fig. 4(a) the network representation of the shrinkage covariance estimator is presented in an undirected graph of the substations.

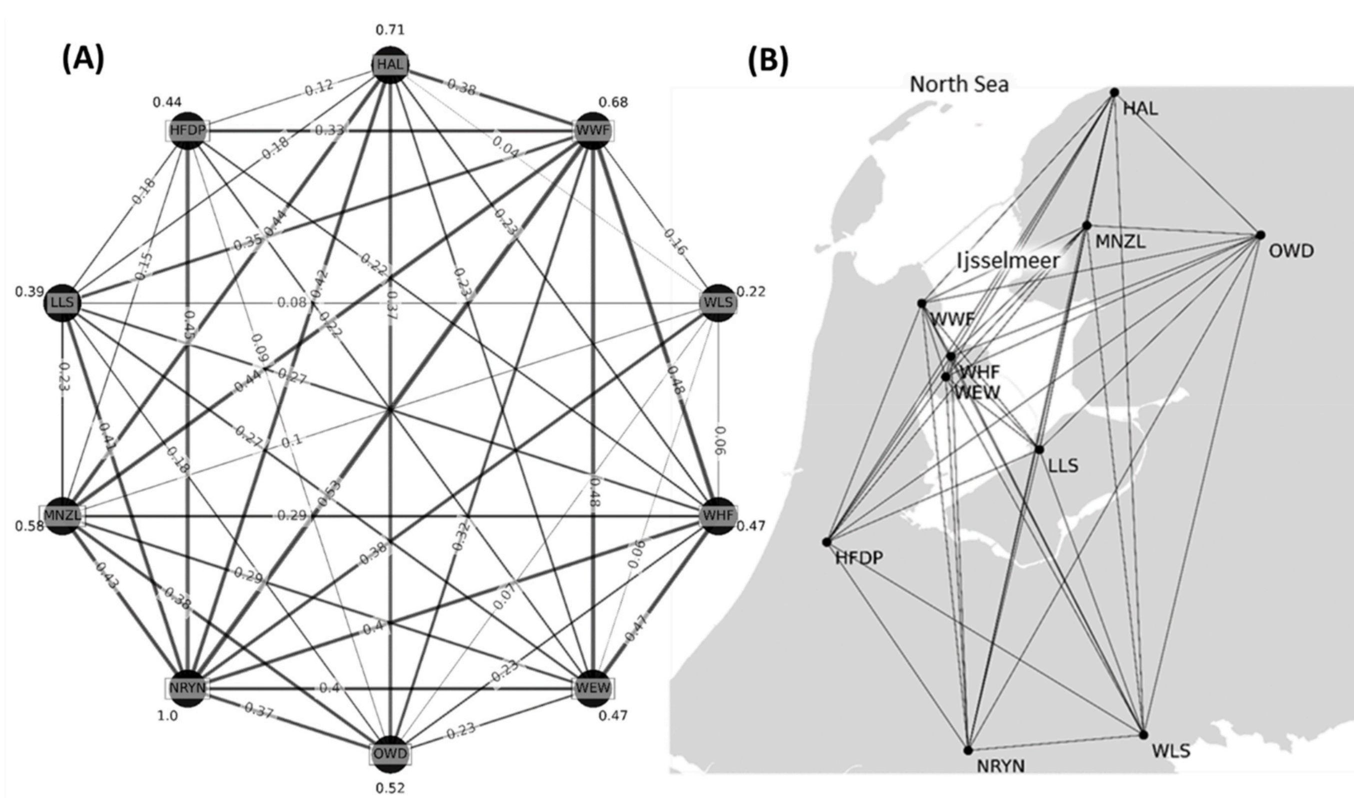

Fig. 4. (a) Shrinkage covariance estimator represented in an undirected graph. (b) The corresponding undirected graph on a geospatial map.

The covariance estimator represents the tendency for the data within the substations to vary together. The number beside the nodes of the network represents the variance of the PC1 in the substation. The corresponding network is embedded on a geospatial map to demonstrate the effects of location in covariance (see Fig. 4(b)). From Fig. 4(a, c), it is observed that NRYN, HAL and WWF have high variance in the substations. However, only NRYN and WWF had larger covariance with other substations. This highlights the importance of NRYN and WWF as a perfect-information substation for renewable energy reconciliation. From the geospatial location, it is expected that NRYN represents the variance due to inland weather dynamics of the country while WWF represents the combined variance at the Ijsselmeer reservoir and also from the North Sea. The substation HAL captures the variance of the Northern part of the country near the North Sea, however, other substations do not share such characteristics leading to lower covariances across other substations. There is also large covariance between WEW and WHF, in which they are located close to each other geospatially. In general, PCA-Ledoit Wolf has shown that variables within substations have direct covariance between each other which is caused by geospatial influence, including weather effects.

4.2. DAZLS and models for zero-shot renewable prediction

In Fig. 5(a), the standard deviation of the error (RMSEP) and mean of error are plotted to represent the variance-bias trade-off. It can be observed that the models that were using DAZLS had much better (lower) standard deviation and mean of error compared to the same models (without applying DAZLS). Models with DAZLS applied had a mean of error with a range between 0.07 and 0.18 (7–18%) while that of the same models via default transfer learning was at around 0.71 to 0.73 (71–73%). The standard deviation of errors is plotted in log-scale (due to models with DAZLS being too much better than the normal transfer learning model), having models using DAZLS with a standard deviation under 0.01, while that of the normal models around 0.009 to 0.025.

The application of DAZLS has also removed unnecessary nonlinearity from the base models by separating the task of domain modelling, adaptation modelling and physical corrections (see Figure (b)). The top 20 methods with DAZLS applied were plotted in Fig. 5 (c), and the less complex model (i.e. K-Neighbors Regression) was performing better considering the sum of RMSEP. Although the nonlinear method started to show up as the 5th best method, the first 4 methods all consist of K-Neighbour Regressions with varying window sizes. The usage of DAZLS simplifies the need for nonlinearity in zero-shot learning for renewable energy prediction in substations.

The best model (1-KNeigborsRegressor-KNeighborsRegressor) from Fig. 5(b) was further compared to an open-source unsupervised learning model, the energy splitting model in the OpenSTEF software package [56] for the same dataset using block-wise leave-one-out validation. From Fig. 6(a) it can be observed that the best method from DAZLS significantly outperforms the openSTEF method where all predictions from DAZLS were below 0.15 RMSEP. OpenSTEF had predictions around 0.2 to 0.3 RMSEP when stable, however, the prediction shoots up to very large values for stations that are not easily predicted. It is also worth mentioning that openSTEF is an unsupervised decomposition method instead of a one-shot learning method.

The model prediction errors for the best method of DAZLS were further studied by embedding the geographical contour map onto the elevation map of the Netherlands (see Fig. 6(b)). It can be seen that the prediction error forms a peak (yellow lines) in the middle, forming a higher error area between 0.12 and 0.135. This region of the Netherlands has a negative Normal Amsterdam Level (NAP) elevation and is also highly affected by the weather from the Ijselmeer reservoir. Predictions were becoming more accurate the further the substation is away from this inaccuracy region (orange-yellow lines).

Nevertheless, we can still observe that some of the substations are still less accurately predicted than others. Looking at Fig. 6(a), although the best method using DAZLS was low on average (blue bars), the substations LLS, HFDP and OWD had relatively higher prediction RMSEP than other stations. Two predictions from solar in WHF (Fig. 7(a)) and wind in WWF substation (Fig. 7(b)) prediction are shown to demonstrate easy-to-predict cases, while solar from OWD and wind from LLS substation were demonstrating difficult-to-predict cases.

For the difficult-to-prediction substations, the main misprediction was in the prediction of the peaks in the time series, this is a generic artefact for models predicting time series in real-time. Furthermore, it can be also identified that the variances in difficult-to-predict

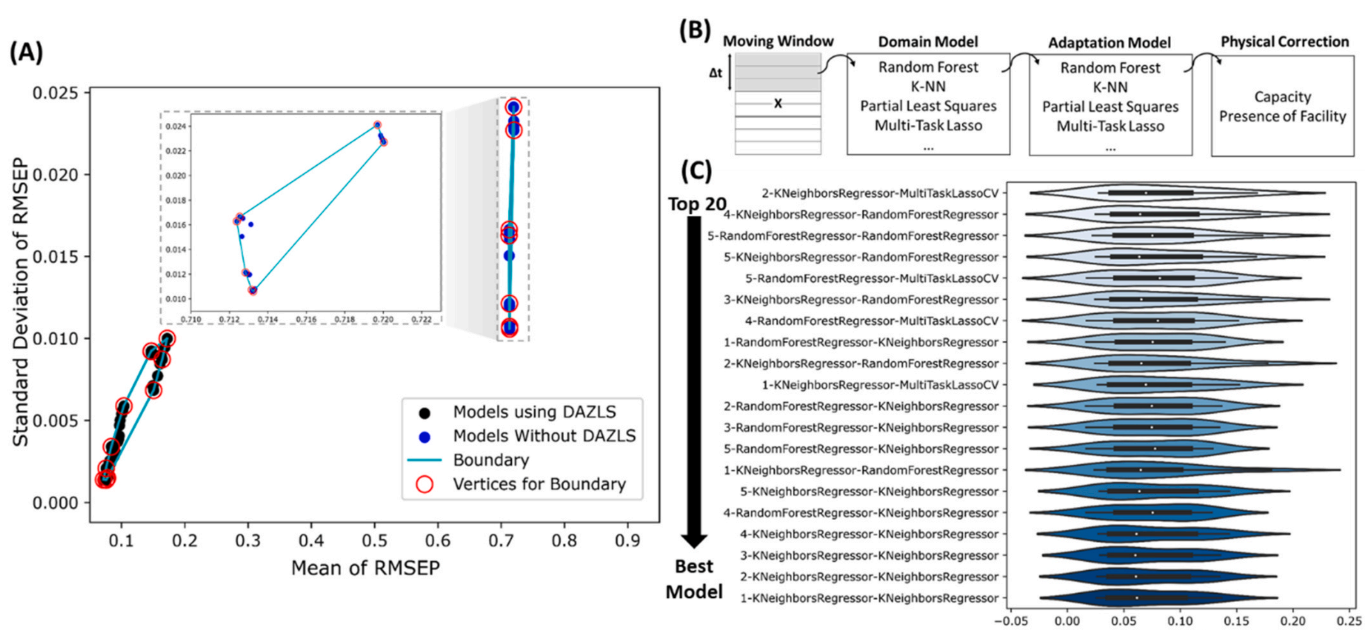

Fig. 5. (a) Standard deviation of root mean squared error of prediction (RMSEP) against the mean of RMSEP for models using DAZLS and without DAZLS (b) Applied scheme of the DAZLS framework with base models within. (c)Top 20 methods out of 80 models using DAZLS sorted by sum of errors from bottom to top. The name of the models was structured as (window size)-(domain model)-(adaptation model).

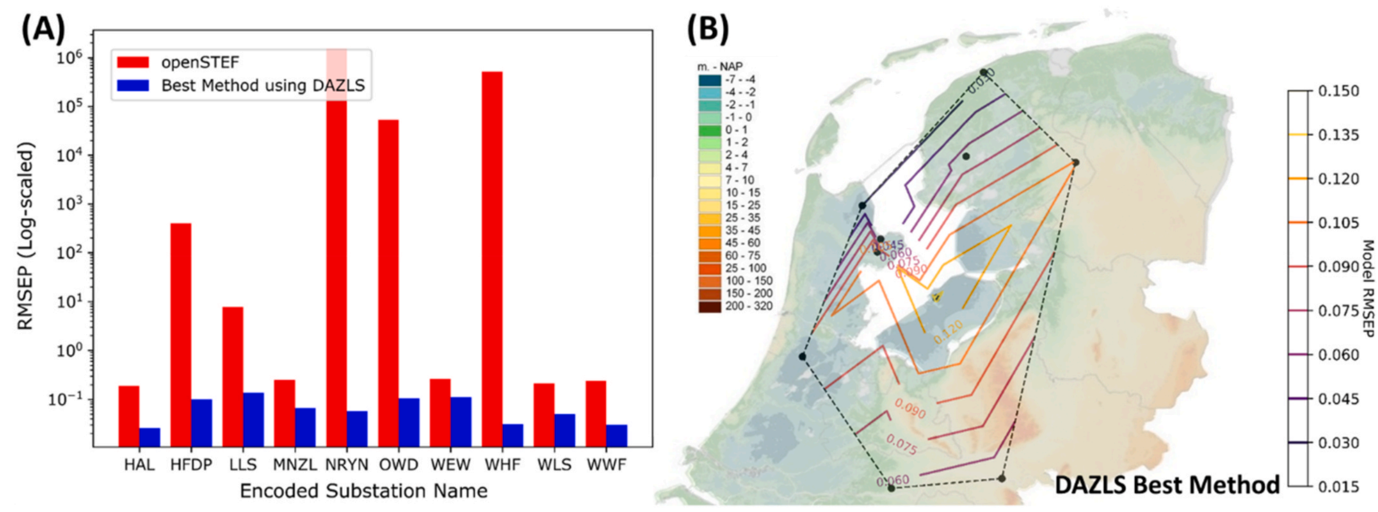

Fig. 6. (a) Comparison of RMSEP for best method from DAZLS and openSTEF energy splitting. (b) Contour map of model error embedded onto an elevation map of the Netherlands. NAP of the Netherlands adapted from Blom-Zandstra et al. [71].

substations were significantly larger than that of the easy-to-predict stations. Although variance correction was already included in the DAZLS framework, some prediction performance in such difficult-to-predict substations was inevitably inflated by the prediction variances.

4.3. Critical pathway exploration for difficult-to-predict substations

For the LLS substation (see Fig. 8(a)), the most important pathway comes from north to south in a sequence of HAL, MNZL, and WWF to LLS. This suggests that LLS is affected by weather dynamics from the North Sea (from HAL), and also affected by substations that are nearby the reservoir region. In Fig. 8(b), it can be observed that the renewable energy generation in the HFDP station has a pathway from HAL, MNZL, LLS, NRYN, and OWD to HFDP. This long sequence suggests that the HFDP substation has a more complex relationship with the other substations. This suggests that the HFDP substation is affected by weather dynamics from the North Sea (from HAL), and also shares similarities with inland substations which are equally distanced from the reservoir region (NRYN, OWD). Furthermore, the LLS substation was within the most important pathway of the HFDP substation, which itself is also a difficult-to-predict substation. This creates more uncertainty and variance in the HFDP substation. Next, the OWD substation (Fig. 8(c)) also has a most important pathway that starts from HAL. The full pathway is HAL, HFDP, WHF, and OWD. The pathway is very similar to that of HFDP, however, WHF was selected instead of substations far down South. WHF and OWD are quite similar in latitude, which provides reasoning on why it is on the path of most explained variance.

In general, the difficult-to-predict stations have the most up north substation (HAL) as the starting node of their most important pathway, suggesting that their energy generation might be related to the weather due to the North Sea. Furthermore, two of the difficult-to-predict substations (HFDP and OWD) have each other on their most important pathway. This might suggest that there is some effect on renewable energy generation that both input data and metadata cannot capture. Nevertheless, the prediction accuracies resulting from the best model using DAZLS can outperform non-one-shot-learning (typical transfer learning) approaches and even commercial models.

4.4. Improvement in environmental impact reconciliation

From utilizing the best method of DAZLS, the difference between cumulative misestimates in GHG emissions with respect to other methods using DAZLS, methods without DAZLS and unsupervised openSTEF method was compared in Fig. 9(a). Overall, the method applying DAZLS significantly provided lower misestimates in GHG emissions compared to models without using DAZLS. Normal machine learning models (without DAZLS) were also performing generic transfer learning by default. Hence, they give lower misestimates in GHG emissions when compared to the openSTEF method, which was based on unsupervised decomposition. The best method applying DAZLS gave a 24 times lower cumulative MCE than the generic transfer learning method (without DAZLS) and 706,436 times lower than when compared to the unsupervised openSTEF energy splitting method.

For the MCA from using renewables (as opposed to non-renewables), all the values were positive, which meant that prediction models were constantly under-estimating MCA values. Underestimation of MCA values meant that GHG values avoidance due to renewables were not properly accounted for, seemingly giving higher GHG emissions. The best method using DAZLS had a monthly average MCA of between 0 and 130 tons CO_2-eq/month (Fig. 9(b)). The openSTEF method was stable in a few substations but had large spikes up to 5×10^6 ton CO_2-eq/month. The percentage improvement in MCA by using DAZLS with the openSTEF as a basis is illustrated in Fig. 9(c). There are very large improvement possibilities (close to 100%) for substations of OWD, WHF, LLS, and HFDP, despite OWD, LLS and HFDP being listed as difficult-to-predict substations. Interestingly, the substation HAL and MNZL which were highly influenced by the weather in the North Sea did not have much potential improvement in MCA. Other substations ranged between 30% and 50% improvement in MCA.

5. Discussion

The proposition of DAZLS allows for reliable renewable predictions in substations with fully missing renewable energy data. This allows for new possibilities for renewable energy-related applications in near real-time frequencies. The direct impact is that near real-time renewable energy quantification in any substations within the boundary is fully possible with appropriate perfect-information substations acting as standard [40], global weather prediction services [72], and geographical information systems [73].

5.1. Reconciliation for digital twins in smart grid

In general, the new approach allows for various applications requiring scalable renewable data from the substation level. From the results of our work, we estimate that previously unpredicted or poorly predicted missing data substations will have at least a 10-fold improvement in terms of performance. This provides accurate

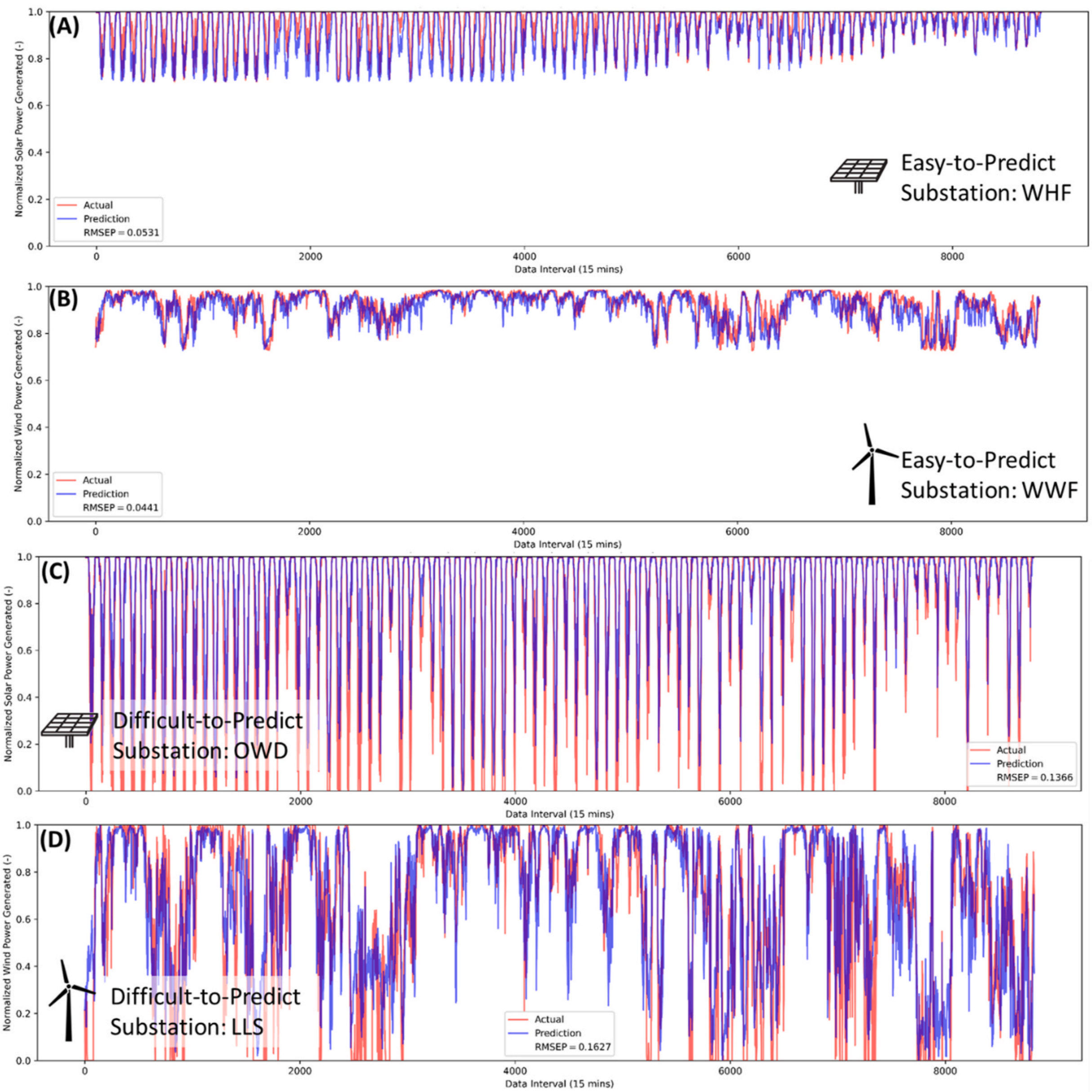

Fig. 7. Actual and DAZLS best model prediction for (a) solar energy generation in WHF substation, (b) wind energy generation in WWF substation, (c) solar energy generation in OWD substation, and (d) wind energy generation in LLS substation.

quantification for previously unknown renewable energy generation in near-real time resolution. Our method directly allows for a near-real time digital twin of region-wide renewable energy production at the substation level, provided that there are a few perfect information substations in the region. Improved renewable data prediction also provides additional value for reliable grid control [74]. This gives improvements in optimal demand side management, renewable energy dispatch, resolving grid conflicts, provide grid autonomy [75] for smart grid and intelligent systems. Nevertheless, the integration of such models remains an unexplored expedition and it allows for various applications related to smart cities [76], applications of digital twins [15], energy management systems [77], etc. Our work provides the basis of a transferable region-wide digital twin for renewable energy prediction, providing accurate control and optimization for the grids of the future and unlocking various peer-to-peer (P2P) prosumer networks [78].

Energy planning is essential towards the long-term development of climate-resilient smart cities [79], industrial symbiosis [80], and circular urban metabolism [81]. Different energy planning models rely on data accuracy to address temporal features, stochastic variability, epistemic uncertainties, multi-objectivity and multi-agent problems [18]. Holistic methods such as carbon emission pinch analysis (CEPA) [82] can benefit from improved grid information for renewable energy prediction from DAZLS, thus giving more understanding of renewable energy planning at the substation level. The improved time resolution of DAZLS can also provide near real-time analysis for time-dependent energy planning tools such as Power Pinch Analysis (PoPA) [83], which previously only relied on hourly time interval data. Multi-criteria strategies for energy planning [84] can also capture more variance of the high-frequency prediction data from DAZLS, relating to more critical multi-objective energy planning.

5.2. Environmental implications

For environmental impact quantification, the near real-time prediction allows precise and dynamic quantification of GHG that is avoided by using renewables as opposed to non-renewables. This opens up a

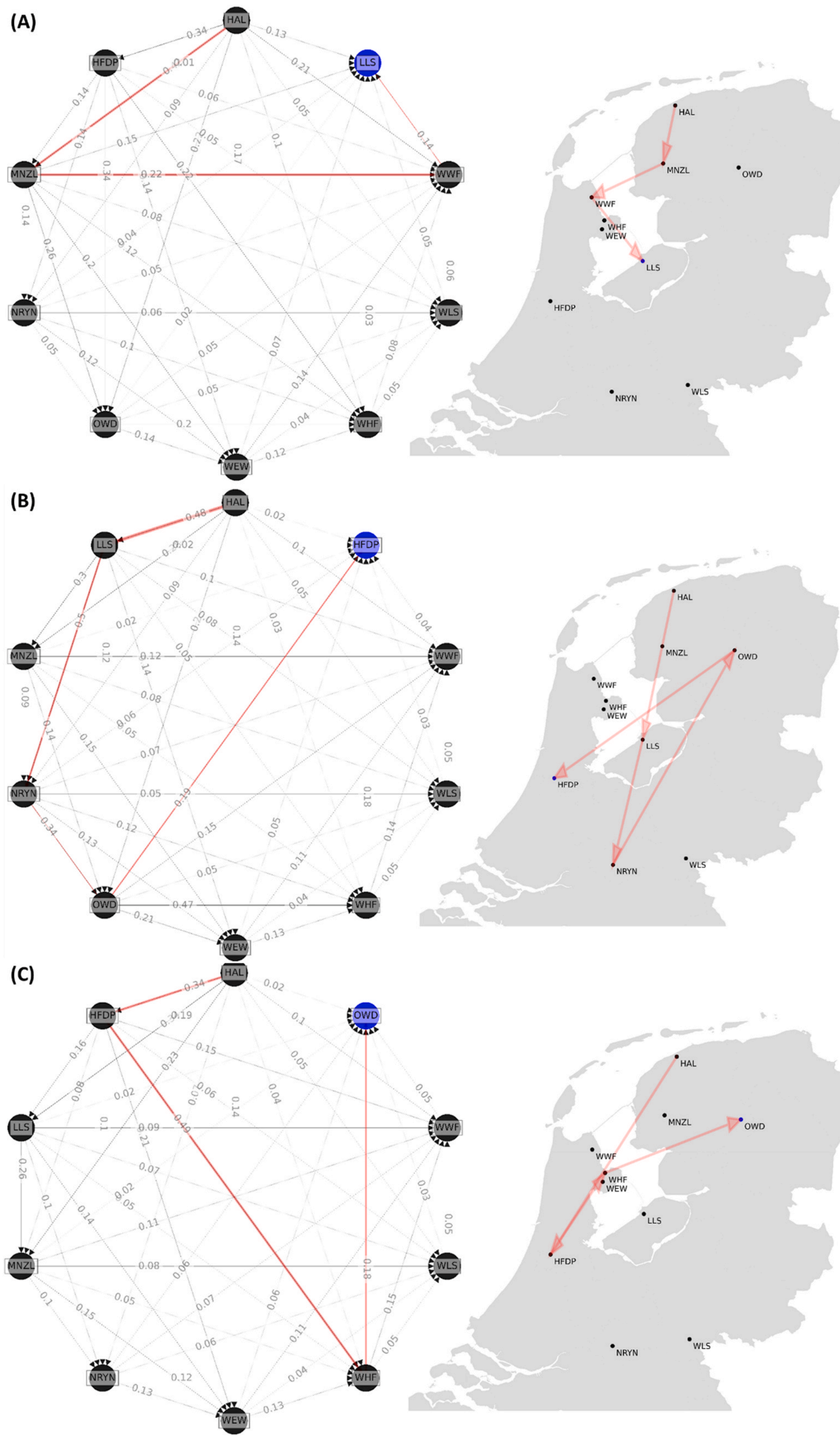

Fig. 8. Critical pathway and its corresponding geospatial map for (a) LLS, (b) HFDP, (c) OWD as the target substation.

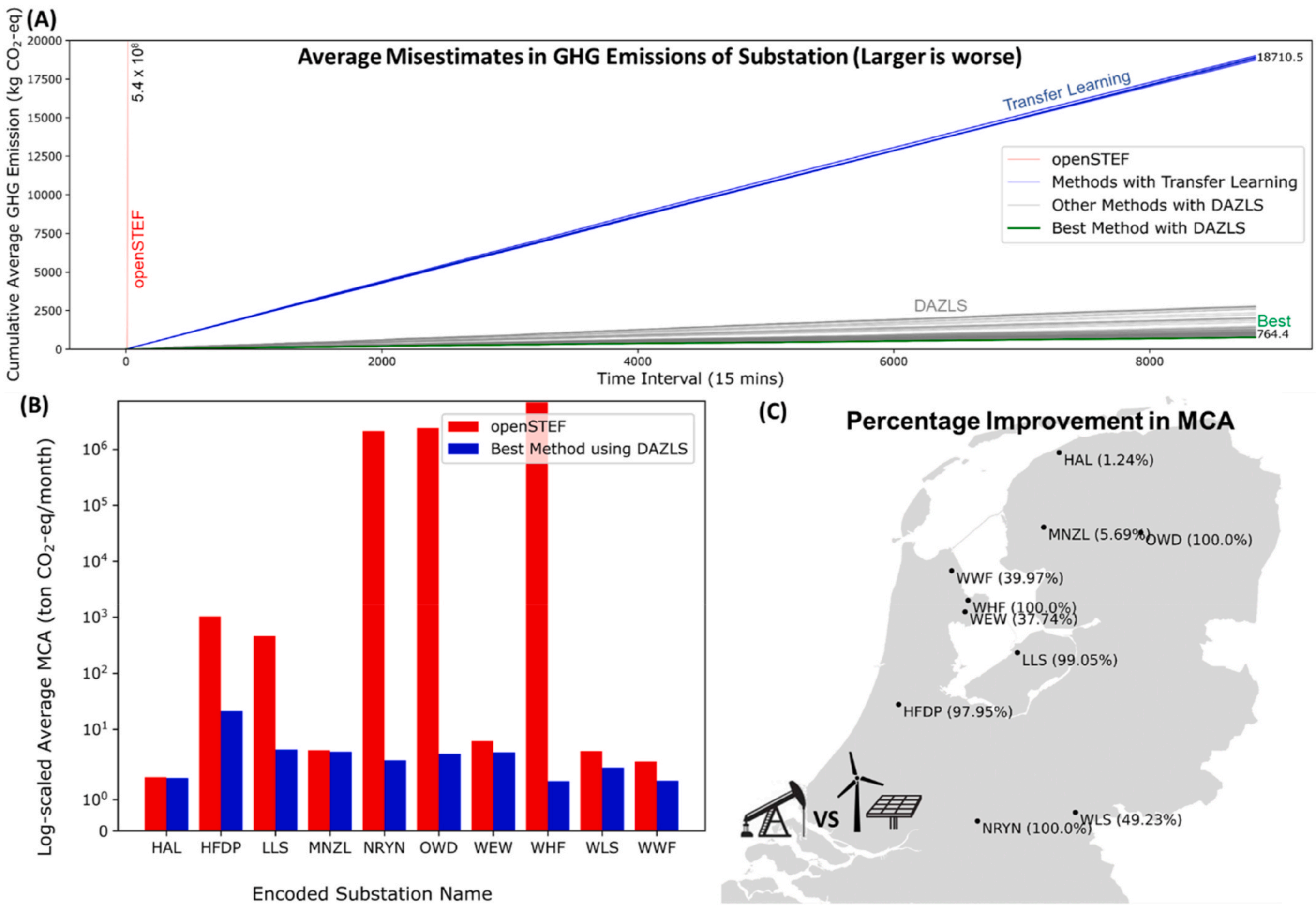

Fig. 9. (a) Average misestimates in GHG emissions of substations over time due to using openSTEF, other methods with DAZLS, best method with DAZLS and methods without DAZLS (b) Average misestimates in carbon avoidance (MCA) (renewables from non-renewables) in different substations of openSTEF and the best method applying DAZLS. (c) Percentage improvement potential (rounded to 2 decimals) in a geospatial map for the best method using DAZLS with openSTEF as a basis.

bottom-up data-driven reconciliation for GHG produced by renewables. As the renewable energy predicted directly affects the energy payback time for LCA [85], substation level prediction information will be an additional consideration for system-level GHG basis in renewables quantification [86,87]. Furthermore, the incorporation of DAZLS can also include substation-level information in dynamic carbon accounting in regional energy markets [88]. In such cases, more accurate quantification of GHG emissions due to renewable energy can directly affect policy-making [89], energy resilience [90], and energy security [91]. This will also provide clearer targets for regional pathways toward United Nation's 2050 GHG emission goals [92].

5.3. Pros, cons and future perspectives

DAZLS provides a low-cost strategy to transfer information from perfect-information substation to imperfect-information substations without additional sensors installed. Additionally, this approach also avoids any data confidentiality issues between stakeholders [93], reducing any security risks. This implies that any substation without renewable energy sensors can still have an accurate prediction in near-real time resolution by utilizing load-weather information, geospatial metadata, weather metadata, and physical data. The region of study is also scalable, and DAZLS can be extended to a global scale to monitor global renewable energy production. DAZLS can be transferable to any other countries or expanded to a wider region (e.g. whole European Union) since the strategy is model-agnostic. A limitation of DAZLS is that it relies on consistent weather information being provided in real-time. Disrupted weather information or low time frequency weather information likely lowers the prediction accuracy of DAZLS. Nevertheless, there are many weather data streaming services that are reliable and readily available [94,95].

For future applications, the incorporation of automatic global substation detection via satellite can be deployed to detect all the substations of interest [96]. This would directly allow the detection and prediction of all substations within a large region. As aforementioned, DAZLS will also serve as a flexible base model for energy digital twins, allowing for more efficient grid control, energy management, and long-term energy planning. The environmental impact due to more precise quantification of renewable energy will also directly influence many LCA-related purposes, policy-making and sustainability measures. As the way forward, more countries should gather near real-time or real-time data from perfect-information renewable energy. This would allow DAZLS to enlarge the effective prediction region, and provide accurate renewable energy quantification globally.

6. Conclusions

In conclusion, we developed an effective procedure to improve any basic machine learning model for zero-shot renewable energy prediction in incomplete information substations. The proposed technique is called domain adaptation for zero-shot learning in sequence (DAZLS) in which input data and metadata are sequentially assigned to a domain model, adaptation model and physical correction model. By testing DAZLS on commonly used machine learning models, it was found that DAZLS reduces the mean of prediction error by 4 times and the variance of prediction error by 3 times. DAZLS favours simple algorithms because the task of mapping the domain data and metadata has been separated into two models of different functionality. This reduces the need for highly nonlinear methods. For unsupervised substation interpretation, this work also proposed using a PCA-combined Ledoit-Wolf shrinkage covariance estimator to explore the underlying nature of the renewable energy generation within various substations. Difficult-to-predict substations can also be explored using a supervised learning approach using process PLS and studying a parsimonious pathway with the highest sum of explained variances. The implementation of the best model using DAZLS has also provided 24 times lower cumulative misestimates in GHG emissions when compared to generic transfer learning models, and 700 thousand times lower than when compared to an open-sourced unsupervised energy splitting method. The possibility of using DAZLS with any data-driven models allows for scalable improvement in carbon avoidance estimation due to using renewables as opposed to non-renewables, ensuring a more informative and resilient energy transition aimed towards sustainability. For broader practical applications, DAZLS provides a cornerstone for region-wide energy digital twins by providing accurate predictions in incomplete-information substations without the need to invest in additional power meters. The benefit of using DAZLS is that the strategy is model-agnostic, so any specific data-driven models of preference can be incorporated into the application. Furthermore, DAZLS can also be extended to any region without major modifications. The main requirement is that information from representative perfect-information renewable energy substations and real-time weather information needs to be provided for the algorithm to perform optimally. These conditions are frequently found in the electricity grids of numerous countries, and if not already present, they can be established at considerably reduced investment expenses. Thus, the usage of DAZLS can act as a cheap and effective building block for accurate renewable energy information prediction.

Credit author statement

Teng, S·Y.: Conceptualization, Methodology, Software, Investigation, Project Administration, Visualization, Writing - Original Draft. Cambier van Nooten, C.: Validation, Software, Writing - Review & Editing. van Doorn, J.M.: Data Curation, Resources, Writing - Review & Editing. Ottenbros, A.: Data Curation, Writing - Review & Editing. Huijbregts, M.A.J.: Resources, Writing - Review & Editing. Jansen, J.J.: Resources, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The research contribution from S.Y. Teng is supported by the European Union's Horizon Europe Research and Innovation Program, under Marie Skłodowska-Curie Actions grant agreement no. 101064585 (MoCEGS). The financial support from the Green Information Technology division of the Interdisciplinary Research Platform, Radboud University Nijmegen under the voucher "RISAI: Renewable Integrated Systems Accelerate by Artificial Intelligence" is appreciated. MA.J. Huijbregts received support from a grant from the Dutch research foundation for the project Global environmental trade-offs of renewable energy technologies (016.Vici.170.190). The photographers Hans Peter van Velthoven, Jack Tillmans and Foto Focus via Alliander for Fig. 1 were also appreciated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2023.113662

The basic DAZLS algorithm can be found online at https://doi.org/10.5281/zenodo.8269094

Sin Yong Teng. (2023). tsyet12/DAZLS: DAZLS release v0.1 (release). Zenodo. https://doi.org/10.5281/zenodo.8269094

References

[1] Sustainable SDG. Development goals: sustainable development knowledge Platform. Sustain Dev Goals; 2019. https://sustainabledevelopment.un.org/?menu=1300. [Accessed 2 May 2020].
[2] Goverment of the Netherlands. Dutch goals within the EU. Clim Chang; 2020. https://www.government.nl/topics/climate-change/eu-policy. [Accessed 12 August 2021].
[3] van der Loo F, Loorbach D. The Dutch energy transition project (2000-2009). Gov. Energy Transit. Reality, Illusion or Necessity?. 2012. https://doi.org/10.4324/9780203126523.
[4] Kemp R, Rotmans J, Loorbach D. Assessing the Dutch energy transition policy: how does it deal with dilemmas of managing transitions? J Environ Pol Plann 2007;9: 315–31. https://doi.org/10.1080/15239080701622816.
[5] Niesten E, Jolink A, Chappin M. Investments in the Dutch onshore wind energy industry: a review of investor profiles and the impact of renewable energy subsidies. Renew Sustain Energy Rev 2018;81:2519–25. https://doi.org/10.1016/j.rser.2017.06.056.
[6] van der Kam MJ, Meelen AAH, van Sark WGJHM, Alkemade F. Diffusion of solar photovoltaic systems and electric vehicles among Dutch consumers: implications for the energy transition. Energy Res Social Sci 2018;46:68–85. https://doi.org/10.1016/j.erss.2018.06.003.
[7] de Boer J, Zuidema C, van Hoorn A, de Roo G. The adaptation of Dutch energy policy to emerging area-based energy practices. Energy Pol 2018;117:142–50. https://doi.org/10.1016/j.enpol.2018.02.008.
[8] Yu W, Patros P, Young B, Klinac E, Walmsley TG. Energy digital twin technology for industrial energy management: classification, challenges and future. Renew Sustain Energy Rev 2022;161:112407. https://doi.org/10.1016/j.rser.2022.112407.
[9] Kiesel R, Paraschiv F. Econometric analysis of 15-minute intraday electricity prices. Energy Econ 2017. https://doi.org/10.1016/j.eneco.2017.03.002.
[10] Märkle-Huß J, Feuerriegel S, Neumann D. Contract durations in the electricity market: causal impact of 15 min trading on the EPEX SPOT market. Energy Econ 2018. https://doi.org/10.1016/j.eneco.2017.11.019.
[11] Liu W, Zhang X, Wu Y, Feng S. Economic analysis of renewable energy in the electricity marketization framework: a case study in guangdong, China. Front Energy Res 2020. https://doi.org/10.3389/fenrg.2020.00098.
[12] Oprea SV, Bâra A. Devising a trading mechanism with a joint price adjustment for local electricity markets using blockchain. Insights for policy makers. Energy Pol 2021. https://doi.org/10.1016/j.enpol.2021.112237.
[13] Vaz AGR, Elsinga B, van Sark WGJHM, Brito MC. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, The Netherlands. Renew Energy 2016. https://doi.org/10.1016/j.renene.2015.06.061.
[14] Teng SY, Máša V, Touš M, Vondra M, Lam HL, Stehlík P. Waste-to-energy forecasting and real-time optimization: an anomaly-aware approach. Renew Energy 2022;181:142–55. https://doi.org/10.1016/j.renene.2021.09.026.
[15] O'Dwyer E, Pan I, Charlesworth R, Butler S, Shah N. Integration of an energy management tool and digital twin for coordination and control of multi-vector smart energy systems. Sustain Cities Soc 2020;62:102412. https://doi.org/10.1016/j.scs.2020.102412.
[16] Allam Z, Bibri SE, Chabaud D, Moreno C. The '15-Minute City' concept can shape a net-zero urban future. Humanit Soc Sci Commun 2022. https://doi.org/10.1057/s41599-022-01145-0.
[17] Allam Z, Bibri SE, Jones DS, Chabaud D, Moreno C. Unpacking the '15-minute city' via 6G, IoT, and digital twins: towards a new narrative for increasing urban efficiency, resilience, and sustainability. Sensors 2022. https://doi.org/10.3390/s22041369.
[18] Kong KGH, How BS, Teng SY, Leong WD, Foo DC, Tan RR, et al. Towards data-driven process integration for renewable energy planning. Curr Opin Chem Eng 2021;31:100665. https://doi.org/10.1016/j.coche.2020.100665.
[19] Teng SY, Touš M, Leong WD, How BS, Lam HL, Máša V. Recent advances on industrial data-driven energy savings: digital twins and infrastructures. Renew Sustain Energy Rev 2021;135:110208. https://doi.org/10.1016/j.rser.2020.110208.
[20] Perera KS, Aung Z, Woon WL. Machine learning techniques for supporting renewable energy generation and integration: a survey. Lect. Notes comput. Sci.

(including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2014:81–96. https://doi.org/10.1007/978-3-319-13290-7_7.
[21] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. Renew Energy 2012;37:1–8. https://doi.org/10.1016/j.renene.2011.05.033.
[22] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, et al. Machine learning methods for solar radiation forecasting: a review. Renew Energy 2017;105:569–82. https://doi.org/10.1016/j.renene.2016.12.095.
[23] Li L-L, Wen S-Y, Tseng M-L, Wang C-S. Renewable energy prediction: a novel short-term prediction model of photovoltaic output power. J Clean Prod 2019;228:359–75. https://doi.org/10.1016/j.jclepro.2019.04.331.
[24] Díaz–Vico D, Torres–Barrán A, Omari A, Dorronsoro JR. Deep neural networks for wind and solar energy prediction. Neural Process Lett 2017. https://doi.org/10.1007/s11063-017-9613-7.
[25] Demolli H, Dokuz AS, Ecemis A, Gokcek M. Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Convers Manag 2019;198:111823. https://doi.org/10.1016/j.enconman.2019.111823.
[26] Wang Y, Shen Y, Mao S, Chen X, Zou H. LASSO and LSTM integrated temporal model for short-term solar intensity forecasting. IEEE Internet Things J 2019;6:2933–44. https://doi.org/10.1109/JIOT.2018.2877510.
[27] AlKandari M, Ahmad I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. Appl Comput Inform 2020. https://doi.org/10.1016/j.aci.2019.11.002.
[28] Romero-Fiances I, Livera A, Theristis M, Makrides G, Stein JS, Nofuentes G, et al. Impact of duration and missing data on the long-term photovoltaic degradation rate estimation. Renew Energy 2022;181:738–48. https://doi.org/10.1016/j.renene.2021.09.078.
[29] Tawn R, Browell J, Dinwoodie I. Missing data in wind farm time series: properties and effect on forecasts. Elec Power Syst Res 2020;189:106640. https://doi.org/10.1016/j.epsr.2020.106640.
[30] Coville A, Siddiqui A, Vogstad K-O. The effect of missing data on wind resource estimation. Energy 2011;36:4505–17. https://doi.org/10.1016/j.energy.2011.03.067.
[31] Anand CK, Amor B. Recent developments, future challenges and new research directions in LCA of buildings: a critical review. Renew Sustain Energy Rev 2017;67:408–16. https://doi.org/10.1016/j.rser.2016.09.058.
[32] Teng SY, Touš M, Leong WD, How BS, Lam HL, Máša V. Recent advances on industrial data-driven energy savings: digital twins and infrastructures. Renew Sustain Energy Rev 2021;135:110208. https://doi.org/10.1016/j.rser.2020.110208.
[33] Brosinsky C, Westermann D, Krebs R. Recent and prospective developments in power system control centers: adapting the digital twin technology for application in power system control centers. In: 2018 IEEE int. Energy conf. ENERGYCON 2018; 2018. https://doi.org/10.1109/ENERGYCON.2018.8398846.
[34] Akhtar F, Rehmani MH. Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: a review. Renew Sustain Energy Rev 2015;45:769–84. https://doi.org/10.1016/j.rser.2015.02.021.
[35] Ul Hassan M, Rehmani MH, Kotagiri R, Zhang J, Chen J. Differential privacy for renewable energy resources based smart metering. J Parallel Distr Comput 2019;131:69–80. https://doi.org/10.1016/j.jpdc.2019.04.012.
[36] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25. https://doi.org/10.1016/j.rser.2015.11.050.
[37] Gagne D. Tribal options analysis rules of thumb : solar. Wind , and Biomass; 2019.
[38] Yang Z, Liu Y, Li C. Interpolation of missing wind data based on ANFIS. Renew Energy 2011. https://doi.org/10.1016/j.renene.2010.08.033.
[39] Weber M, Turowski M, Cakmak HK, Mikut R, Kuhnapfel U, Hagenmeyer V. Data-driven copy-paste imputation for energy time series. IEEE Trans Smart Grid 2021;12:5409–19. https://doi.org/10.1109/TSG.2021.3101831.
[40] Moustris K, Paliatsos AG, Bloutsos A, Nikolaidis K, Koronaki I, Kavadias K. Use of neural networks for the creation of hourly global and diffuse solar irradiance data at representative locations in Greece. Renew Energy 2008;33:928–32. https://doi.org/10.1016/j.renene.2007.09.028.
[41] Tasnim S, Rahman A, Oo AMT, Haque ME. Wind power prediction in new stations based on knowledge of existing Stations: a cluster based multi source domain adaptation approach. Knowl Base Syst 2018;145:15–24. https://doi.org/10.1016/j.knosys.2017.12.036.
[42] Redko I, Habrard A, Morvant E, Sebban M, Bennani Y. Advances in domain adaption theory. Elsevier; 2019. https://doi.org/10.1016/C2016-0-05108-2.
[43] Panareda Busto P, Iqbal A, Gall J. Open set domain adaptation for image and action recognition. IEEE Trans Pattern Anal Mach Intell 2020. https://doi.org/10.1109/TPAMI.2018.2880750.
[44] Tseng BH, Budzianowski P, Wu YC, Gašić M. Tree-structured semantic encoder with knowledge sharing for domain adaptation in natural language generation. In: SIGDIAL 2019 - 20th annu. Meet. Spec. Interes. Gr. Discourse dialogue - proc. Conf.; 2019. https://doi.org/10.18653/v1/w19-5920.
[45] Singhal P, Walambe R, Ramanna S, Kotecha K. Domain adaptation: challenges, methods, datasets, and applications. IEEE Access 2023;11:6973–7020. https://doi.org/10.1109/ACCESS.2023.3237025.
[46] Xie G-S, Zhang X-Y, Yan S, Liu C-L. Hybrid CNN and dictionary-based models for scene recognition and domain adaptation. IEEE Trans Circ Syst Video Technol 2017;27:1263–74. https://doi.org/10.1109/TCSVT.2015.2511543.
[47] Xiong L, Ye M, Zhang D, Gan Y, Liu Y. Source data-free domain adaptation for a faster R-CNN. Pattern Recogn 2022;124:108436. https://doi.org/10.1016/j.patcog.2021.108436.
[48] Tas Y, Koniusz P. CNN-Based action recognition and supervised domain adaptation on 3D body skeletons via kernel feature maps. Br Mach Vis Conf 2018. BMVC 2018 2018.
[49] Daumé H. Frustratingly easy domain adaptation. ACL 2007 - Proc 45th Annu Meet Assoc Comput Linguist 2009.
[50] Sun B, Feng J, Saenko K. Return of frustratingly easy domain adaptation. 30th AAAI Conf Artif Intell AAAI 2016 2015. https://doi.org/10.1609/aaai.v30i1.10306.
[51] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning. In: Proc. 2006 conf. Empir. Methods nat. Lang. Process. - EMNLP '06, morristown, NJ, USA: association for computational linguistics; 2006. p. 120. https://doi.org/10.3115/1610075.1610094.
[52] Kouw WM, Loog M. A review of domain adaptation without target labels. IEEE Trans Pattern Anal Mach Intell 2021. https://doi.org/10.1109/TPAMI.2019.2945942.
[53] Kiss B, Kácsor E, Szalay Z. Environmental assessment of future electricity mix – linking an hourly economic model with LCA. J Clean Prod 2020;264:121536. https://doi.org/10.1016/j.jclepro.2020.121536.
[54] Zhang P. Model selection via multifold cross validation. Ann Stat 1993;21. https://doi.org/10.1214/aos/1176349027.
[55] Sun X, Gu J, Sun H. Research progress of zero-shot learning. Appl Intell 2021;51:3600–14. https://doi.org/10.1007/s10489-020-02075-7.
[56] Alliander NV. OpenSTEF - open short term energy forecaster. 2022.
[57] Google. Google. Maps. Google maps Netherlands. 2022. https://www.google.com/maps.
[58] KNMI. The Royal Netherlands Meteorological Institute. Página web KNMI. 2021.
[59] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. London, Edinburgh Dublin Phil Mag J Sci 1901;2:559–72. https://doi.org/10.1080/14786440109462720.
[60] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. J Multivariate Anal 2004;88:365–411. https://doi.org/10.1016/S0047–259X(03)00096-4.
[61] Cochran WG. Analysis of covariance: its nature and uses. Biometrics 1957;13:261. https://doi.org/10.2307/2527916.
[62] Le Cacheux Y, Le Borgne H, Crucianu M. From classical to generalized zero-shot learning: a simple adaptation process. Lect Notes Comput Sci 2019. https://doi.org/10.1007/978-3-030-05716-9_38.
[63] Wang W, Zheng VW, Yu H, Miao C. A survey of zero-shot learning. ACM Trans Intell Syst Technol 2019;10:1–37. https://doi.org/10.1145/3293318.
[64] van Kollenburg G, Bouman R, Offermans T, Gerretzen J, Buydens L, van Manen H-J, et al. Process PLS: incorporating substantive knowledge into the predictive modelling of multiblock, multistep, multidimensional and multicollinear process data. Comput Chem Eng 2021;154:107466. https://doi.org/10.1016/j.compchemeng.2021.107466.
[65] de Jong S. SIMPLS: an alternative approach to partial least squares regression. Chemometr Intell Lab Syst 1993;18:251–63. https://doi.org/10.1016/0169-7439(93)85002-X.
[66] Kroonenberg PM, Lohmoller J-B. Latent variable path modeling with partial least squares. J Am Stat Assoc 1990;85:909. https://doi.org/10.2307/2290049.
[67] Offermans T, Hendriks L, van Kollenburg GH, Szymańska E, Buydens LMC, Jansen JJ. Improved understanding of industrial process relationships through conditional path modelling with process PLS. Front Anal Sci 2021;1. https://doi.org/10.3389/frans.2021.721657.
[68] Miettinen K. No-preference methods. Nonlinear Multiobjective Optim.; 1998. p. 67–76. https://doi.org/10.1007/978-1-4615-5563-6_3.
[69] Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B. The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 2016. https://doi.org/10.1007/s11367-016-1087-8.
[70] Wielders L, Nusselder S. Emissiekentallen elektriciteit. Delft, The Netherlands. 2020.
[71] Blom-Zandstra M, Paulissen M, Agricola H, Schaap B. How will climate change affect spatial planning in agricultural and natural environments? Examples from three Dutch case study regions. IOP Conf Ser Earth Environ Sci 2009;8:012018. https://doi.org/10.1088/1755-1315/8/1/012018.
[72] Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature 2015;525:47–55. https://doi.org/10.1038/nature14956.
[73] Shafiullah M, Rahman SM, Mortoja MG, Al-Ramadan B. Role of spatial analysis technology in power system industry: an overview. Renew Sustain Energy Rev 2016;66:584–95. https://doi.org/10.1016/j.rser.2016.08.017.
[74] Ourahou M, Ayrir W, El Hassouni B, Haddi A. Review on smart grid control and reliability in presence of renewable energies: challenges and prospects. Math Comput Simulat 2020;167:19–31. https://doi.org/10.1016/j.matcom.2018.11.009.
[75] Panda DK, Das S. Smart grid architecture model for control, optimization and data analytics of future power networks with more renewable energy. J Clean Prod 2021;301:126877. https://doi.org/10.1016/j.jclepro.2021.126877.
[76] Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezgui Y, et al. Next-generation energy systems for sustainable smart cities: roles of transfer learning. Sustain Cities Soc 2022;85:104059. https://doi.org/10.1016/j.scs.2022.104059.
[77] Olatomiwa L, Mekhilef S, Ismail MS, Moghavvemi M. Energy management strategies in hybrid renewable energy systems: a review. Renew Sustain Energy Rev 2016;62:821–35. https://doi.org/10.1016/j.rser.2016.05.040.
[78] Kong KGH, Lim JY, Leong WD, Ng WPQ, Teng SY, Sunarso J, et al. Fuzzy optimization for peer-to-peer (P2P) multi-period renewable energy trading planning. J Clean Prod 2022. https://doi.org/10.1016/j.jclepro.2022.133122.

[79] Kim KG. Planning models for climate resilient and low-carbon smart cities: an urban innovation for sustainability, efficiency, circularity, resiliency, and connectivity planning. Urban B. Ser.; 2018. https://doi.org/10.1007/978-3-319-59618-1_2.

[80] Fraccascia L, Yazdanpanah V, van Capelleveen G, Yazan DM. Energy-based industrial symbiosis: a literature review for circular energy transition. Environ Dev Sustain 2021. https://doi.org/10.1007/s10668-020-00840-9.

[81] Lucertini G, Musco F. Circular urban metabolism framework. One Earth 2020. https://doi.org/10.1016/j.oneear.2020.02.004.

[82] Tan RR, Foo DCY. Pinch analysis approach to carbon-constrained energy sector planning. Energy 2007. https://doi.org/10.1016/j.energy.2006.09.018.

[83] Wan Alwi SR, Mohammad Rozali NE, Abdul-Manan Z, Klemeš JJ. A process integration targeting method for hybrid power systems. Energy 2012. https://doi.org/10.1016/j.energy.2012.01.005.

[84] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning - a review. Renew Sustain Energy Rev 2004;8:365–81. https://doi.org/10.1016/j.rser.2003.12.007.

[85] Varun, Bhat IK, Prakash R. LCA of renewable energy for electricity generation systems—a review. Renew Sustain Energy Rev 2009;13:1067–73. https://doi.org/10.1016/j.rser.2008.08.004.

[86] Dammeier LC, Loriaux JM, Steinmann ZJN, Smits DA, Wijnant IL, van den Hurk B, et al. Space, time, and size dependencies of greenhouse gas payback times of wind turbines in northwestern Europe. Environ Sci Technol 2019;53:9289–97. https://doi.org/10.1021/acs.est.9b01030.

[87] Bosmans JHC, Dammeier LC, Huijbregts MAJ. Greenhouse gas footprints of utility-scale photovoltaic facilities at the global scale. Environ Res Lett 2021;16:094056. https://doi.org/10.1088/1748-9326/ac1df9.

[88] Tranberg B, Corradi O, Lajoie B, Gibon T, Staffell I, Andresen GB. Real-time carbon accounting method for the European electricity markets. Energy Strategy Rev 2019;26:100367. https://doi.org/10.1016/j.esr.2019.100367.

[89] Patlitzianas KD, Doukas H, Kagiannas AG, Psarras J. Sustainable energy policy indicators: review and recommendations. Renew Energy 2008;33:966–73. https://doi.org/10.1016/j.renene.2007.05.003.

[90] Dong K, Dong X, Jiang Q, Zhao J. Assessing energy resilience and its greenhouse effect: a global perspective. Energy Econ 2021;104:105659. https://doi.org/10.1016/j.eneco.2021.105659.

[91] Radovanović M, Filipović S, Pavlović D. Energy security measurement – a sustainable approach. Renew Sustain Energy Rev 2017;68:1020–32. https://doi.org/10.1016/j.rser.2016.02.010.

[92] Fragkos P, Tasios N, Paroussos L, Capros P, Tsani S. Energy system impacts and policy implications of the European Intended Nationally Determined Contribution and low-carbon pathway to 2050. Energy Pol 2017;100:216–26. https://doi.org/10.1016/j.enpol.2016.10.023.

[93] Hu J, Vasilakos AV. Energy big data analytics and security: challenges and opportunities. IEEE Trans Smart Grid 2016. https://doi.org/10.1109/TSG.2016.2563461.

[94] Williams M, Cornford D, Bastin L, Jones R, Parker S. Automatic processing, quality assurance and serving of real-time weather data. Comput Geosci 2011. https://doi.org/10.1016/j.cageo.2010.05.010.

[95] Thornton PE, Shrestha R, Thornton M, Kao SC, Wei Y, Wilson BE. Gridded daily weather data for North America with comprehensive uncertainty quantification. Sci Data 2021. https://doi.org/10.1038/s41597-021-00973-0.

[96] Peng J, Liu Y-B, Yima, He Y, Zheng Z-Z, Zhang Y-M, et al. Deep transfer learning for power substation recognition with google earth. In: 2018 15th int. Comput. Conf. Wavelet act. Media technol. Inf. Process. IEEE; 2018. p. 54–7. https://doi.org/10.1109/ICCWAMTIP.2018.8632585.