DTU

DTU Library

A Statistical Model for Hourly Large-Scale Wind and Photovoltaic Generation in New Locations

Ekstrom, Jussi; Koivisto, Matti Juhani; Mellin, Ilkka; Millar, John; Lehtonen, Matti

A Statistical Model for Hourly Large-Scale Wind and Photovoltaic Generation in New Locations

Jussi Ekström, Matti Koivisto, Ilkka Mellin, John Millar, *Member, IEEE* and Matti Lehtonen, *Member, IEEE*

Abstract—The analysis of large-scale wind and photovoltaic (PV) energy generation is of vital importance in power systems where their penetration is high. This paper presents a modular methodology to assess the power generation and volatility of a system consisting of both PV plants (PVPs) and wind power plants (WPPs) in new locations. The methodology is based on statistical modelling of PV and WPP locations with a vector autoregressive model, which takes into account both the temporal correlations in individual plants and the spatial correlations between the plants. The spatial correlations are linked through distances between the locations, which allows the methodology to be used to assess scenarios with PVPs and WPPs in multiple locations without actual measurement data. The methodology can be applied by the transmission and distribution system operators when analysing the effects and feasibility of new PVPs and WPPs in system planning. The model is verified against hourly measured wind speed and solar irradiance data from Finland. A case study assessing the impact of the geographical distribution of the PVPs and WPPs on aggregate power generation and its variability is presented.

Index Terms—Monte Carlo simulation, photovoltaic (PV) power generation, renewable energy, vector autoregressive model, wind power generation.

I. INTRODUCTION

THE installed capacities and overall penetration of stochastic renewable energy sources (RES), such as PVPs and WPPs, have been increasing rapidly in past years, and this progress is expected to continue in the future. As the penetration of RES increases, there arise several difficulties related to the operation and planning of power systems due to the variability in the generation caused by the stochastic nature of these energy sources, such as severe power ramps, both up and down, and increased volatility in the power generation in general. Therefore, the modelling of the PV and wind power generation, and especially its variability, together to understand the combined effects of large-scale PV and wind power generation on the power system becomes crucial. This applies for both transmission and distribution system operators, and also for power producers with RES in their portfolio.

Wind power variability modelling and long-term simulations have been a topic of interest for several years, and have been published extensively, [1]–[4]. Large-scale PV generation has also been modelled and studied widely with different

J. Ekström, John Millar and Matti Lehtonen are with the Department of Electrical Engineering and Automation, Aalto University, Espoo FI-00076 AALTO, Finland (e-mail: jussi.ekstrom@aalto.fi; john.millar@aalto.fi; matti.lehtonen@aalto.fi).

M. Koivisto is with the Wind Energy Department, Technical University of Denmark (DTU), Roskilde, Denmark (e-mail: mkoi@dtu.dk).

I. Mellin is with the Department of Mathematics and Systems Analysis, Aalto University, Espoo FI-00076 AALTO, Finland (e-mail: ilkka.mellin@aalto.fi).

methodologies and objectives during past years [5]–[9]. The methodologies to model PV and wind power generation and variability separately in new generation locations without actual measurement data, i.e. in non-measured locations, have been presented in [10], [11]. The modelling of power systems with both PV and wind power is a current topic, and is being actively researched and published, [12]–[15]. The impacts of the geographical distributions of the PV or wind generation have been analysed separately in [10], [11], [16]. However, statistical modelling of systems with both PV and wind generation in non-measured locations and the analyses of different geographical distributions of the generation locations in such systems are scarce.

The copula method has been commonly used with wind and solar modelling to separate the dependency structures in several locations from the marginal probability distributions (margins) of the individual locations [2], [6], [17]. This approach is used as the base of the modelling in this paper.

The dependency structures can be divided into spatial and temporal dependencies, which can be modelled in various ways. The temporal dependencies have been modelled for each individual location with, e.g., autoregressive (AR) models and the spatial correlations between the individual locations with spatial correlation matrices in [10], [18]. Alternative approaches have been to model the temporal and spatial dependencies together with, e.g., artificial neural networks (ANNs) [3], [4], or with vector autoregressive (VAR) models [4], [17], [19].

The paper contributes to the literature by presenting a modular statistical modelling methodology for both WPPs and PVPs. The methodology combines the core functionality of the wind power simulation model introduced in [10] and the PV simulation model presented in [11]. The methodology is based on a VAR model, which is able to capture both the temporal correlations in both types of generation and the spatial correlations between the installation locations. It is used with Monte Carlo (MC) simulations to model the PV and wind energy generation simultaneously, and to evaluate the volatility and energy output of a generation mix consisting of both PVPs and WPPs.

The methodology is able to model non-measured locations, which is crucial for versatile applicability, as new generation is constantly being planned and measurements are not available from all locations of interest. Although applied in to Finnish data in this paper, the methodology is applicable in any geographical area where measurements for parameter estimation are available. One of the key benefits of the methodology is also its modular structure, which allows the usage of different WPP and PVP power generation models within the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2017.2682338, IEEE Transactions on Sustainable Energy

2

methodology.

This paper proceeds as follows. Section II presents the data and the marginal distributions. Section III presents the modelling of the dependency structures by specifying the VAR model and its parameters. Section IV describes the modelling of non-measured wind speed and solar irradiance locations. Section V assesses the simulation results of the out-of-sample test locations. Section VI conducts a case study of the impact of the geographical distribution of the PV and wind power generation locations on the aggregate power generation. Section VII presents the conclusions drawn from the presented models and results.

II. The Data and The Marginal Distributions

This section presents the used datasets and the estimated marginal distributions for wind speed and solar irradiance. Fig. 1 shows the complete simulation procedure. The steps are presented in detail in Sections II, III and IV.
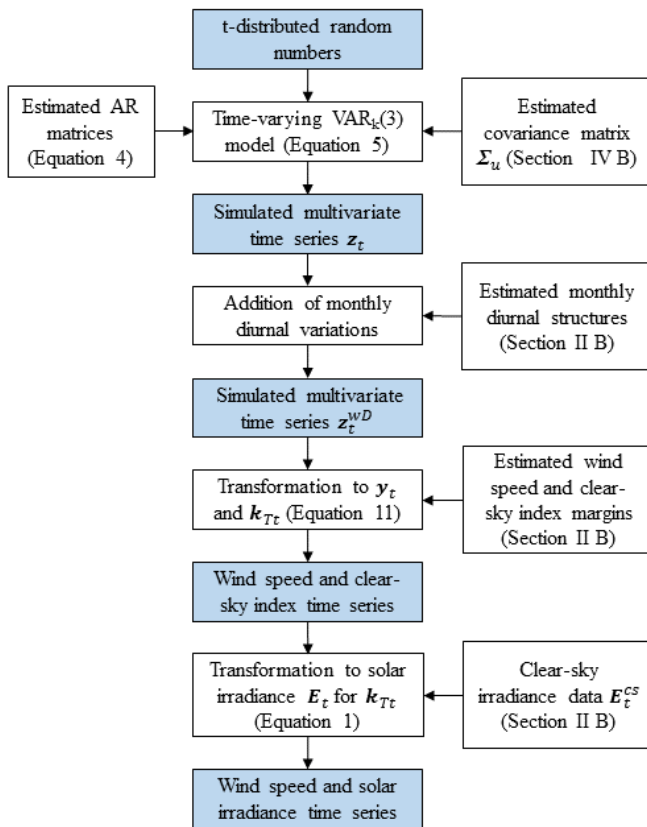

Fig. 1. The flowchart of the MC simulation procedure. The tiles with the blue background indicate simulated data and white background operations.

A. The Data

The wind speed simulation related dataset used in this paper is the same utilized in [10]. Most importantly, it consists of hourly high altitude wind speed time series measured from 12 locations and low altitude measurements from 19 locations in Finland. The high altitude measurement height varies from 74 meters up to 150 meters above the surrounding ground level

and the measurement lengths vary from location to location between one to three years. A more detailed specification of the dataset can be found in [10].

The solar dataset is obtained from the Finnish Meteorological Institute and consists of measured hourly global horizontal irradiance (including both direct normal and diffuse horizontal irradiance) and temperature time series from eight locations in Finland. The measurement lengths vary from location to location between one and five years. A more detailed specification of the dataset can be found in [11].

As the wind speed and irradiance data both have the resolution of one hour, it is the time resolution adopted in this paper. The high altitude wind speed and global irradiance measurement locations used for the estimation of the spatial correlations and the two out-of-sample test locations are shown in Fig. 2.
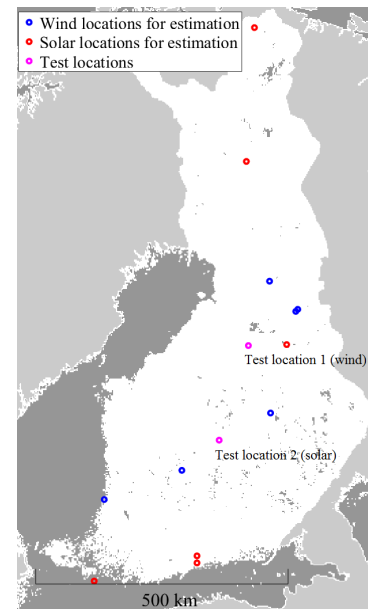

Fig. 2. The wind speed (blue circles) and global irradiance (red circles) measurement locations and the two out-of-sample test locations (magenta circles) in Finland.

B. The Marginal Distributions

Weibull distributions are widely utilized to represent the distribution of wind speeds in a given location [10]. They are also used as margins to describe the local wind speed conditions in this paper.

For solar locations, clear-sky index distributions are used as margins in the simulations, as in [11]. The clear-sky index data are obtained by dividing the measured global irradiance with the estimated clear-sky irradiance. The clear-sky irradiance contains the theoretical maximums, i.e. the deterministic component, caused by the movements of the earth and the sun, of the global irradiance. The clear-sky index data contain the stochastic component (the deviation from that theoretical maximum) of the global irradiance, caused by e.g. the clouds. For location i at time t, the clear-sky index is obtained as

$$k_{Ti,t} = \frac{E_{i,t}}{E_{i,t}^{CS}}, \qquad (1)$$

where $E_{i,t}$ is the measured (or simulated) global irradiance and $E_{i,t}^{CS}$ is the clear-sky irradiance. When the clear-sky irradiance is zero, $k_{Ti,t}$ is not defined, i.e. the clear-sky index data are defined only during the hours when the sun is above the horizon. The clear-sky index data $\boldsymbol{k}_{T,t}$ are modelled as a stochastic variable using the VAR model, and model the deviation from the theoretical maximum irradiance (i.e. the cloudiness).

The VAR model generally assumes normally distributed margins [20]. Therefore, the wind speed and clear-sky index data are transformed to data with normally distributed margins, $\boldsymbol{z}_t^{wD}$ with the probability integral transformation, where wD means *with day structures* (i.e. diurnal structures are not yet removed)

$$z_{i,t}^{wD} = F_N^{-1}[\hat{F}_i(y_{i,t})], \qquad (2)$$

where F_N^{-1} is the inverse cumulative distribution function (CDF) of the standard normal distribution, $\hat{F}_i$ is the estimated clear-sky index or Weibull margin for location i, and $y_{i,t} = k_{Ti,t}$ for solar measurement locations and the wind speed data for wind measurement locations. The transformation to normally distributed margins, despite the t-distributed errors, is an appropriate approach when the error terms are modelled separately from the conditional mean model (the VAR model). Separate modelling of the errors is commonly used, e.g., in ARMA-GARCH modelling, where the GARCH part can include errors that are not normally distributed [21]. A similar approach for modelling the error terms separately from the VAR model is utilized in this paper.

The monthly diurnal day structures are then estimated and removed from $\boldsymbol{z}_t^{wD}$ by calculating hourly mean values for each month in each location and then subtracting these from the respective data, which yields $\boldsymbol{z}_t = [z_{1,t}, z_{2,t}, ..., z_{k,t}]'$. A more detailed explanation of the calculation of the monthly diurnal structures can be found in [11].

The VAR model also assumes stationarity [20]. There were no trends, but seasonality was found from both the wind speed and global irradiance data. However, the transformation to clear-sky index data (1) and the removal of the monthly diurnal structures eliminated the seasonality. To ensure the stationarity of $\boldsymbol{z}_t$, the augmented Dickey-Fuller test [22] was utilized and $\boldsymbol{z}_t$ was found to be stationary, and therefore, suitable to be modelled with the VAR model. $\boldsymbol{z}_t$ is used to estimate the VAR model parameters in the next section.

III. THE DEPENDENCY STRUCTURES

This section presents the time-varying $\text{VAR}_k(3)$ model used in the paper to model the dependency structures of solar irradiance and wind speeds in multiple locations. First, the VAR model is specified and then the calculation of the VAR model parameters is presented.

A. The Time-Varying VAR Model

A VAR model is a multivariate generalization of a univariate AR model. A k-dimensional p-order $\text{VAR}_k(p)$ model for the transformed $\boldsymbol{z}_t = [z_{1,t}, z_{2,t}, ..., z_{k,t}]'$ is defined as

$$\boldsymbol{z}_t = \boldsymbol{c} + \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{z}_{t-i} + \boldsymbol{u}_t, \qquad (3)$$

where $\boldsymbol{c}$ is a k-vector of intercept terms, $\boldsymbol{A}_i, ..., \boldsymbol{A}_p$ are the $k \times k$ AR coefficient matrices and $\boldsymbol{u}_t$ is the error term of the model [20].

In the paper, five appropriate stipulations are made in the VAR model to enable the modelling of both solar irradiance and wind speeds in non-measured locations. First, the intercept terms are assumed to be zero, i.e., $\boldsymbol{c} = \boldsymbol{0}$ (this can be done as in (2) the transformation is to standard normal distribution). Second, the model identification was done by analysing the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of $\boldsymbol{z}_t$. It was concluded that the order of the VAR model should be three, i.e. $p = 3$. The model is thus a $\text{VAR}_k(3)$ model, where k is the number of locations. Third, the AR coefficient matrices are assumed to be time-dependent, i.e., they are $\boldsymbol{A}_{1,t}$, $\boldsymbol{A}_{2,t}$ and $\boldsymbol{A}_{3,t}$, which gives a specific set of coefficients for different t, and thus specifies a time-varying VAR model. Time-dependency is necessary to capture differences in temporal dependency structures in solar irradiance depending on the time of the year [11]. Fourth, $\boldsymbol{u}_t$ is assumed to follow a multivariate t-distribution to handle the peakedness of the residuals, as in [10], [11]. This is justified, as the VAR model parameters are estimated with ordinary least squares (OLS), which do not require normally distributed error terms and allow separate modelling of the errors [21], [23]. Fifth, the off-diagonal components of $\boldsymbol{A}_{1,t}$, $\boldsymbol{A}_{2,t}$ and $\boldsymbol{A}_{3,t}$ are assumed to be zero, yielding

$$\boldsymbol{A}_{i,t} = \begin{bmatrix} a_{i,t,1} & 0 & \cdots & 0 \\ 0 & a_{i,t,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{i,t,k} \end{bmatrix}. \qquad (4)$$

The off-diagonals are assumed to be zero to allow a straightforward modelling of non-measured locations. This matter is discussed in Section IV. Because of specification (4), all spatial correlations between the components of $\boldsymbol{z}_t$ must come from the error terms $\boldsymbol{u}_t = [u_{1,t}, u_{2,t}, ..., u_{k,t}]'$. With the above mentioned specifications, the utilized VAR model can be written as

$$\boldsymbol{z}_t = \sum_{i=1}^{3} \boldsymbol{A}_{i,t} \boldsymbol{z}_{t-i} + \boldsymbol{u}_t, \qquad (5)$$

where $\boldsymbol{A}_{1,t}$, $\boldsymbol{A}_{2,t}$ and $\boldsymbol{A}_{3,t}$ are time-dependant. The adequacy of the model presented in (5) was ensured by the assessment of the ACFs, PACFs and cumulative periodogram of the model residuals, and with the Ljung-Box Q-test [24]. The residuals had no visible autocorrelation in neither of the location types and the residuals passed the Ljung-Box Q-test. Therefore, the model was considered adequate.

The next section introduces an approach to add the required spatial correlations to $\boldsymbol{z}_t$ through $\boldsymbol{u}_t$ using the covariance matrix $\boldsymbol{\Sigma}_u = cov(\boldsymbol{u}_t)$.

B. Specifying the VAR Model Parameters

For solar locations, the time-varying coefficient matrices $\boldsymbol{A}_{1,t}, \boldsymbol{A}_{2,t}, \boldsymbol{A}_{3,t}$ are specified for six different groups, each consisting of two month time periods, which was found to be the most suitable approach to capture the seasons for the AR models in [11]. For wind locations, the coefficient matrices are the same for all t. Estimation of the AR parameters for new locations is presented in Section IV A.

The covariance matrix $\boldsymbol{\Sigma}_u$ is specified as follows. First, the spatial correlations (off-diagonal elements) in the autocorrelation matrices $\boldsymbol{R}_z(h)$ of the process $\boldsymbol{z}_t$ are calculated for all required lags h. The correlations are estimated from the distances between the locations, as shown in Section IV B.

Second, the temporal correlations (the diagonal elements) of $\boldsymbol{R}_z(h)$ of $\boldsymbol{z}_t$ are specified. The temporal correlations for lags $h > 0$ (as autocorrelation is 1 for $h = 0$) can be determined with the AR coefficient matrices $\boldsymbol{A}_i, \ldots, \boldsymbol{A}_p$ using the Yule-Walker equations [20]. To simplify the calculations, and as the AR coefficients do not vary greatly with different t, the averages of $\boldsymbol{A}_{1,t}, \boldsymbol{A}_{2,t}, \boldsymbol{A}_{3,t}$ are used.

Now the autocorrelation matrices $\boldsymbol{R}_z(h)$ are fully determined for all required lags. The autocorrelation matrices are then transformed to corresponding autocovariance matrices $\boldsymbol{\Gamma}_z(h)$, using the transformation

$$\boldsymbol{\Gamma}_z(h) = \boldsymbol{D}\boldsymbol{R}_z(h)\boldsymbol{D}, \tag{6}$$

where $\boldsymbol{D}$ is a diagonal matrix with the standard deviations calculated for the components of $\boldsymbol{z}_t$ (the k solar and wind locations) on the diagonal (these standard deviations are smaller than 1, as the diurnal structures are subtracted from the $N(0,1)$ distributed $\boldsymbol{z}_t^{wD}$).

For further analyses, the VAR model has to be presented in the $\text{VAR}_{kp}(1)$ form (the state-space presentation) [20]. The autocovariance matrix $\boldsymbol{\Gamma}_Z(0)$ of the process $\boldsymbol{z}_t$ in this form is formed from the autocovariance matrices $\boldsymbol{\Gamma}_z(h)$ obtained with (6), in the following manner

$$\boldsymbol{\Gamma}_Z(0) = \begin{bmatrix} \boldsymbol{\Gamma}_z(0) & \boldsymbol{\Gamma}_z(1) & \boldsymbol{\Gamma}_z(2) \\ \boldsymbol{\Gamma}_z(-1) & \boldsymbol{\Gamma}_z(0) & \boldsymbol{\Gamma}_z(1) \\ \boldsymbol{\Gamma}_z(-2) & \boldsymbol{\Gamma}_z(-1) & \boldsymbol{\Gamma}_z(0) \end{bmatrix}. \tag{7}$$

The $kp \times kp$ -dimensional $\boldsymbol{A}$-matrix is then formed from the AR coefficient matrices $\boldsymbol{A}_i$ (where the coefficients are the averages of $\boldsymbol{A}_{i,t}$) as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 \\ \boldsymbol{I}_k & 0 & 0 \\ 0 & \boldsymbol{I}_k & 0 \end{bmatrix}, \tag{8}$$

where $\boldsymbol{I}_k$ is a $k \times k$ identity matrix. Next, the $kp \times kp$ -dimensional covariance matrix $\boldsymbol{\Sigma}_U$ can be calculated with the equation

$$\boldsymbol{\Sigma}_U = \boldsymbol{\Gamma}_Z(0) - \boldsymbol{A}\boldsymbol{\Gamma}_Z(0)\boldsymbol{A}', \tag{9}$$

as shown in [20]. Finally, the $k \times k$ -dimensional covariance matrix $\boldsymbol{\Sigma}_u$ is obtained by taking the first k rows and first k columns from $\boldsymbol{\Sigma}_U$. All parameters for the used VAR model are now specified.

IV. THE SIMULATION OF NON-MEASURED WIND SPEED AND SOLAR IRRADIATION LOCATIONS

The main objective for the methodology is to be able to simulate future scenarios with non-measured wind speed and solar irradiance locations. This section presents the specification of the simulation parameters, the estimation of the covariance matrix for non-measured locations and the simulation procedure.

A. The Simulation Parameters for Non-Measured Locations

As noted in Section III A, the residuals $\boldsymbol{u}_t$ of the $\text{VAR}_k(3)$ model follow a t-distribution. The mean values of the degrees of freedom, estimated from the residuals of the measured locations, as done in [10], are considered for non-measured locations (estimated separately for wind speed and solar irradiance locations).

The mean values of the estimated parameters (separate for wind and solar locations) are used as the time-varying AR parameters for non-measured locations. The monthly diurnal variations are estimated from the geographically closest measured location of the same type.

The wind speed margins in the non-measured locations can be obtained from Wind Atlas or a similar database [25]. The Wind Atlas database, which is used in this paper, provides Weibull distribution parameters, describing the local wind speed conditions according to the provided coordinates.

Empirical cumulative distribution functions (ECDFs) of the clear-sky index data obtained from the geographically closest measured location are considered for the non-measured locations for solar locations, as in [11].

The clear-sky irradiance distributions in non-measured locations are determined using a clear-sky irradiance model [26]. The clear-sky irradiance model estimates the clear-sky global irradiance for the given coordinates, when the Linke turbidity factors for the coordinates are provided [11], [26]. The Linke turbidity factors are obtained from [27].

B. The Estimation of the Covariance Matrix

The underlying correlations, calculated from $\boldsymbol{z}_t$, between wind and solar locations can be linked to the distances between the locations. Fig. 3 shows the correlations, calculated from $\boldsymbol{z}_t$, plotted against the distance between two wind locations. Fig. 4 presents the correlation, calculated from $\boldsymbol{z}_t$, between two solar locations. Fig. 5 shows the correlations, calculated also from $\boldsymbol{z}_t$, between solar and wind locations.

Wind speeds and solar irradiances are negatively correlated, as shown in [13], [14]. According to Fig. 5, wind speed and global irradiance locations are also negatively correlated when calculated from $\boldsymbol{z}_t$, where all deterministic components, caused by the movements of the earth and the sun, in solar irradiance and monthly day structures in solar irradiance and wind speed have been removed from the data. This is a crucial observation, and also indicates why PV and wind power should be modelled together. As the model utilized in the paper is $\text{VAR}_k(3)$, it requires spatial correlations also for lags $h = 0, 1, 2$ (the fitted curves to do this are shown in

Figs. 3−5. Nine curves (lags $h = 0, 1, 2$ for each of the three cases) are fitted to allow the estimation of the non-diagonal components of $\boldsymbol{R}_z(h)$ for non-measured locations.

The diagonal temporal correlations for each non-measured location for lags $h = 0, 1, 2$ are obtained using the average values of the estimated time-dependent AR coefficient matrices $\boldsymbol{A}_{1,t}, \boldsymbol{A}_{2,t}, \boldsymbol{A}_{3,t}$, as explained in Section III B.

The diagonal standard deviation matrices $\boldsymbol{D}$ required to transform $\boldsymbol{R}_z(h)$ to $\boldsymbol{\Gamma}_z(h)$ with (6) are estimated from the measurement data and average values of the standard deviations of the wind speed or global irradiance measurements are used for non-measured locations, respectively. With this approach, the covariance matrix $\boldsymbol{\Sigma}_u$ for the simulation of new locations can be determined.
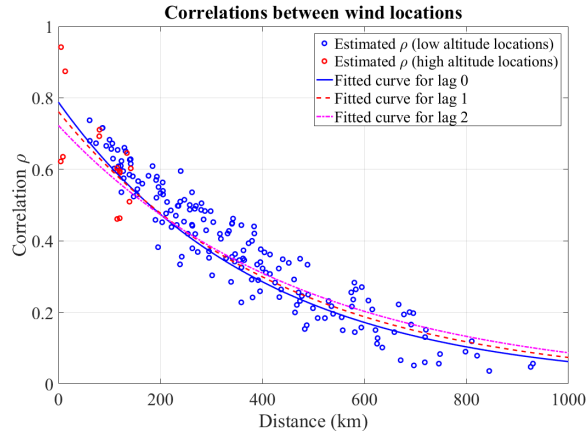

Fig. 3. Spatial correlations ρ estimated from $\boldsymbol{z}_t$ between wind speed locations at high and low altitudes (at lag 0) plotted against the distances between the locations, with fitted curves for spatial correlations at lags 0, 1 and 2.
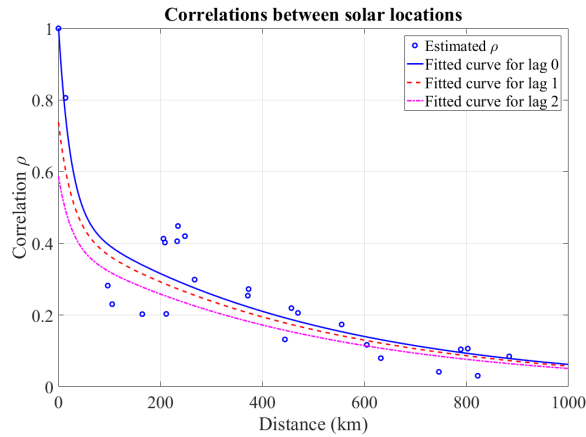

Fig. 4. Spatial correlations ρ estimated from $\boldsymbol{z}_t$ between solar irradiance locations (at lag 0) plotted against the distances between the locations, with fitted curves for spatial correlations at lags 0, 1 and 2.

C. The Simulation Procedure

This section describes the MC simulation procedure for the simulation of non-measured locations illustrated in Fig. 1. First, multivariate normal random numbers, whose margins are then transformed to the appropriate t-distributions, as
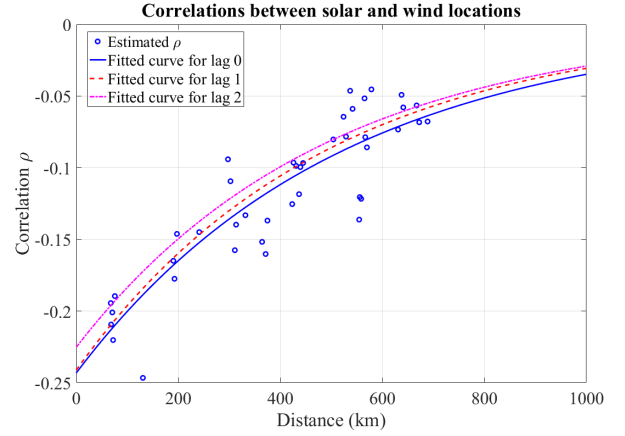

Fig. 5. Spatial correlations ρ estimated from $\boldsymbol{z}_t$ between wind speed and solar irradiance locations (at lag 0) plotted against the distances between the locations, with fitted curves for spatial correlations at lags 0, 1 and 2.

in [16], are generated for all k locations for the desired simulation period $t = 1, ..., T$. These data are then used as the input (innovations) for the time-varying VAR$_k(3)$ model. This results in a multivariate time series $\boldsymbol{z}_t$ containing the desired spatial and temporal correlations.

The estimated monthly diurnal variations (day structures) are then added to $\boldsymbol{z}_t$, and thus, $\boldsymbol{z}_t^{wD}$ is obtained. The next step is to transform the time series $\boldsymbol{z}_t^{wD}$ to the wind speed and clear-sky index domains through the transformation

$$y_{i,t} = \hat{F}_i^{-1}[F_N(z_{i,t}^{wD})], \qquad (10)$$

where $\hat{F}_i^{-1}$ is the inverse of the estimated CDF of the wind speed margin (in wind locations) or the clear-sky index margin (in solar locations) for location i, F_N is the CDF of the standard normal distribution and $y_{i,t}$ are the simulated wind speed or clear-sky index ($y_{i,t} = k_{Ti,t}$) time series. $\boldsymbol{k}_{Tt}$ are further transformed to the global irradiance domain using the clear-sky irradiances obtained with the clear-sky irradiation model [26]. Multivariate time series (with the spatial and temporal correlations) consisting of both wind speed and solar irradiance data from the desired k locations are now obtained for power generation analyses.

V. THE SIMULATION RESULTS FOR THE OUT-OF-SAMPLE TEST LOCATIONS

This section presents the long term MC simulation results for the two out-of-sample test locations shown in Fig. 2. The data from these two locations were excluded from the estimation process. 1000 MC simulation runs, with a length of one year and time resolution of one hour, were carried out, yielding 8.76×10^6 simulated samples for both locations. In addition, 1000 MC simulation runs were also carried out for both locations with the separate wind and solar simulation models (referred to as *separate simulations*) presented for wind in [10] and solar in [11]. The following compares the simulated data to the measured and separately simulated wind and solar data.

A. The Spatial Correlations

The cross-correlation functions (XCFs) between the two test locations, for simulated and measured data, are illustrated in Fig. 6. The correlations between the wind and solar test locations are negative and relatively small, but still notable, especially near lag 0. The simulation model is able to capture the shape of the XCF calculated from the data and the XCFs are relatively similar (especially near lag 0). This applies both when all hours are included in the calculation of the XCFs and when only the hours when the sun is above the horizon are included. It should be noted, though, that the simulated XCF values are slightly lower than the XCF values calculated from measurements. However, some differences are expected in an out-of-sample test, as e.g., the actual monthly diurnal structures in the test locations can vary slightly from the ones used for out-of-sample locations.

It should be noted, that the separate simulations also show some correlation structure as the daily variations in wind speeds and solar irradiance are negatively correlated. However, the impact of the distance between the out-of-sample locations is not modelled, and thus, its effect is left out from the XCF. Consequently, the separate modelling of wind and solar has a major deficiency when modelling both wind and solar locations.
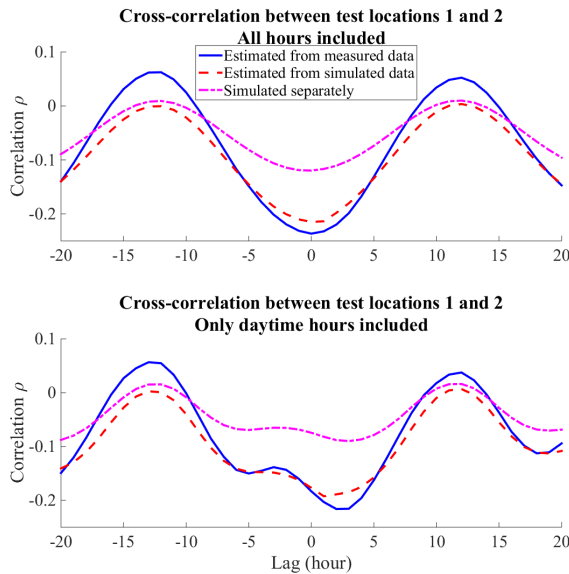

Fig. 6. The XCFs between the simulated and measured wind speed data and the separate simulations from test location 1 and solar irradiance data from test location 2. In the upper figure, all hours are considered and in the lower figure only hours when the sun is above the horizon are considered. The simulation results are averages of the 1000 MC runs.

B. The Temporal Correlations

The ACFs of the simulated and measured data are presented for test locations 1 and 2 in Fig. 7. In can be seen that both of the ACFs calculated from the simulated data are similar to the ACFs calculated from the measurements, although small differences can be seen between the simulated and measured

data near the 24-hour lag. In addition, the simulated data captures the behaviour of the measurement data as closely as the data simulated with the separate simulations.
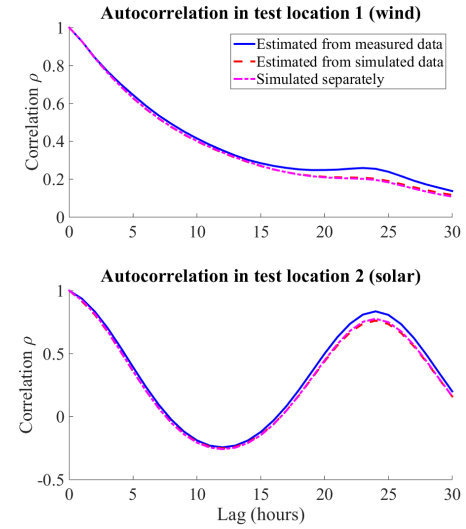

Fig. 7. The ACFs of the simulated and measured wind speed data from test location 1, the simulated and measured global irradiance data from test location 2 and the separate simulations. The simulation results are averages of the 1000 MC runs.

C. The Probability Density Functions

The probability density functions (PDFs) calculated from the simulated and measured data are illustrated in Fig. 8. It can be observed that the shapes of the simulated PDFs are very similar to those estimated from the measurement data for both wind and solar test locations. Additionally, the simulated data depicts the measurement data as accurately as the data simulated with the separate simulations.

Fig. 8. The PDFs of the simulated and measured data and the separate simulations for test locations 1 and 2. The simulation results are averages of the 1000 MC runs.

D. The Ramp Rates

The PDFs for the ramp rates for one hour ramps calculated from the simulated and measured data are illustrated in Fig. 9.

It can be seen that the shapes of the simulated ramp rate PDFs are close to those estimated from the measurement data for both wind and solar test locations. Furthermore, the simulated data represents the ramps of the measurement data similarly as the data simulated with the separate simulations.

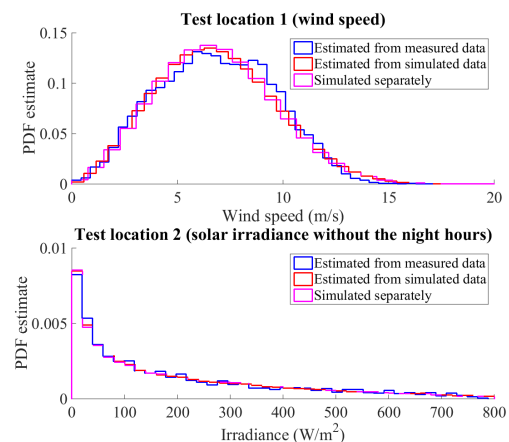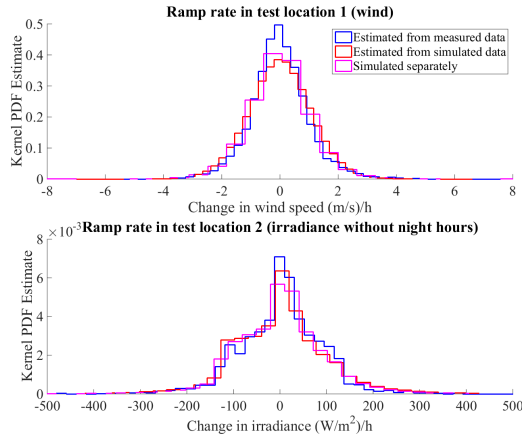Fig. 9. The ramp rate PDFs of the simulated and measured data and the separate simulations for test locations 1 and 2. The simulation results are averages of the 1000 MC runs.

E. The Numerical Results

This section presents the most relevant numerical statistics, shown in Table I, estimated from the simulated and measured data. It can be seen that all of the statistics estimated from the simulated data are similar to those estimated from the measured data. It should be noted that the simulated data performs as well as the separate simulations in each statistic, except the spatial correlations, where the separate simulations perform poorly as the distance related correlation between wind and solar locations is not included in the models.

To summarize, the results for the out-of-sample locations are satisfactory with all measures considered and the model is able to assess the spatial correlations between wind and solar locations and the temporal correlations, margins and ramp rates in individual locations.

VI. THE COMPARISON OF THE GEOGRAPHICAL DISTRIBUTIONS AND PROPORTIONAL YEARLY ENERGIES OF THE WPP AND PV GENERATION

This section presents a case study consisting of 12 scenarios, assessing the effects of different geographical distribution and proportional yearly generated energies of WPPs and PV generation on the volatility of the aggregate power generation. The geographical distributions considered in the scenarios can be seen in Fig. 10. In addition, a short analysis that shows the effect of the proportional yearly installed capacities of WPPs and PVPs on the volatility is also presented.

A. The Simulation Setup for the Scenarios

The scenarios in the case study assess 12 different scenarios, which are presented in Table II. Each scenario consists of 12

WPPs (all with equal capacities) and 12 PVPs (also with equal capacities) located in Finland, so that the expected aggregated annual energy is fixed at 1 TWh, in each case (to provide well comparable scenarios). Due to the fixed generated energy, the installed capacities vary between the cases.
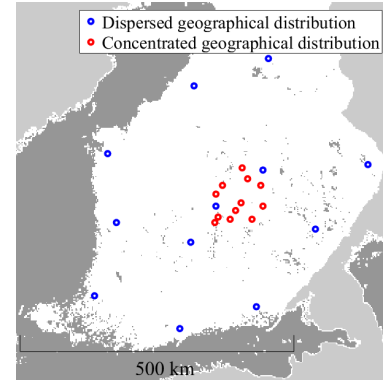

Fig. 10. The dispersed (blue circles) and concentrated (red circles) geographical distributions of wind and PV generation used in the scenarios.

The $VAR_{24}(3)$ model produces wind speed and global irradiance time series, and therefore, these have to be converted to power through wind turbine and PV panel power generation models. For WPPs, a wind turbine model, including the modelling of the wake effect in a wind farm, introduced in [10] is utilized to transform the wind speeds to wind power. Gamesa G128-5.0 turbines [28] with 140 meter towers are used for all turbines. Depending on the scenario, the number of turbines in one WPP varies from 6 to 14. The wake effect inside WPPs is considered as in [10]. To focus on the effects of the geographical distribution of the generation, the locations were chosen so that the wind generation potential (i.e. the Weibull parameters) are similar in all locations, yielding similar annual wind energy generation. The Weibull parameters for the locations are obtained from [25].

For PV power generation, polycrystalline silicone panels facing south and with a tilt angle of 45 degrees are con-

TABLE I
STATISTICS ESTIMATED FROM THE MEASURED AND SIMULATED DATA AND THE SEPARATE SIMULATIONS FOR THE OUT-OF-SAMPLE TEST LOCATIONS. σ DEPICTS STANDARD DEVIATION AND ρ CORRELATION. ALL ESTIMATES FROM THE SIMULATED DATA ARE AVERAGES OF THE 1000 SIMULATION RUNS.

Statistic	Measured Data	Simulated Data	Separate simulations
Mean for test loc 1 (m/s)	6.85	6.84	6.84
Mean for test loc 2 (kJ/m^2)	96.91	97.67	98.12
σ for test loc 1 (m/s)	2.75	2.82	2.82
σ for test loc 2 (kJ/m^2)	167.75	169.06	168.23
Temporal ρ at lag 1 for test loc 1	0.93	0.93	0.93
Temporal ρ at lag 1 for test loc 2	0.94	0.92	0.92
Spatial ρ at lag 0 between locs 1 and 2	-0.23	-0.22	-0.12

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2017.2682338, IEEE Transactions on Sustainable Energy

8

TABLE II
THE SPECIFICATIONS OF THE 12 SCENARIOS.

Scenario	WPP Geograph. Distr.	PV Geograph. Distr.	WPP Yearly Energy (%)	PV Yearly Energy (%)
Scenario 1	Dispersed	Concentrated	70	30
Scenario 2	Dispersed	Concentrated	80	20
Scenario 3	Dispersed	Concentrated	90	10
Scenario 4	Dispersed	Dispersed	70	30
Scenario 5	Dispersed	Dispersed	80	20
Scenario 6	Dispersed	Dispersed	90	10
Scenario 7	Concentrated	Concentrated	70	30
Scenario 8	Concentrated	Concentrated	80	20
Scenario 9	Concentrated	Concentrated	90	10
Scenario 10	Concentrated	Dispersed	70	30
Scenario 11	Concentrated	Dispersed	80	20
Scenario 12	Concentrated	Dispersed	90	10

TABLE III
THE NUMERICAL RESULTS FOR THE 12 SCENARIOS. σ DEPICTS STANDARD DEVIATION. THE $< 10\%$ AND $> 90\%$ CASES PRESENT THE PERCENTAGE OF HOURS IN A SIMULATION RUN WHEN THE HOURLY AGGREGATED GENERATION IS WITHIN THE RESPECTIVE LIMITS. ALL ESTIMATES ARE AVERAGES OF THE 1000 MC RUNS.

Scenario	Hourly σ (MWh/h)	<10%	>90%
Scenario 1	65.0	11.3	0.1
Scenario 2	57.8	7.2	0.1
Scenario 3	58.7	5.0	0.2
Scenario 4	62.4	9.9	0.1
Scenario 5	56.7	6.4	0.1
Scenario 6	58.5	4.6	0.4
Scenario 7	72.2	18.5	0.1
Scenario 8	68.5	14.1	0.1
Scenario 9	72.5	11.5	0.2
Scenario 10	70.5	17.1	0.1
Scenario 11	68.5	14.1	0.1
Scenario 12	72.6	11.1	0.2

sidered and the power transformation is conducted with a widely utilized power generation model for polycrystalline silicon PV panels, presented in [29] and utilized in similar applications e.g., in [11]. The ambient temperatures used for the PV generation models are taken from the closest measured location (measurements from 2013).

B. The Simulation Results

The simulation results using 1000 hourly MC simulation runs with a length of one year for the 12 scenarios are presented in Fig. 11 and in Table III. Fig. 11 illustrates the uncertainty in hourly aggregate generation with estimated PDFs of the generation in different scenarios and Table III presents the key numerical results.

As Table III and Fig. 11 show, there are notable differences between the cases in terms of PDFs, standard deviations and percentage of hours when the generation is less than 10% or more than 90% of the aggregate installed capacity. It should
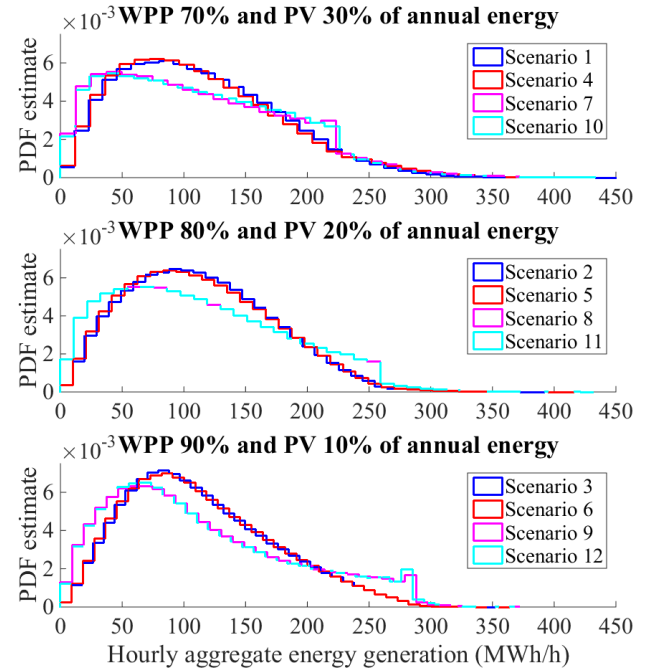

Fig. 11. The PDFs of the hourly aggregate generation in the 12 scenarios. The lines with the same colors have the same geographical distribution. The PDFs are averages of the 1000 MC runs.

be noted, that the values in $> 90\%$ column are small as they require excellent generation conditions in all plants and as the wake effect decreases the efficiency of the WPPs and e.g. the inverters the efficiency of the PVPs. When both WPPs and PVPs have a dispersed geographical distribution (Scenarios 4, 5 and 6), the standard deviation of the aggregate generation is low. However, Scenarios 1, 2 and 3 where WPPs are dispersed and PVPs concentrated, are also good in terms of low volatility. The highest volatilities can be found in Scenarios $7-12$, where wind generation is geographically concentrated. In addition, Fig. 11 also illustrates that in scenarios where the geographical spread of the WPPs is concentrated, a notable drop can be seen in the PDFs when the WPP generation reaches its maximum aggregate capacity.

The most optimal proportion of annual generated energies, considering low volatility of the aggregate generation, can be found in Scenarios 2, 5, 8 and 11, where the WPPs generate 80% and the PVPs 20% of the annual energy. Despite the geographical distribution, volatility increases if the proportion of energy generated with WPPs decreases to 70% or increases to 90%. This is caused by the negative correlation between WPPs and PVPs and the greater hourly variability of the PVPs compared to the WPPs. The relative standard deviation (RSD), which is the sample standard deviation divided by the sample mean, is 0.592 for WPPs and 1.615 for PVPs in Scenario 5.

If the PVP share increases to 30%, the greater variability of the PVPs outweighs the benefits of the negative correlation between wind and solar generation on the volatility of the aggregate generation. If the PVP share decreases to 10%, the benefit of the increased share of the WPPs (which have lower hourly variability than PVPs) is not sufficient to compensate

TABLE IV
THE PROBABILITIES THAT THE AGGREGATED POWER GENERATION
EXCEEDS OR IS LESS THAN A SPECIFIED LIMIT IN THE 12 SCENARIOS.
ALL ESTIMATES ARE AVERAGES OF THE 1000 MC RUNS.

Scenario	Less than 20% of the installed capacity for 3 consecutive hours (%)	Exceeding 80% of the installed capacity for 3 consecutive hours (%)
Scenario 1	27.83	0.0006
Scenario 2	18.83	0.0022
Scenario 3	14.75	0.0081
Scenario 4	24.50	0.0003
Scenario 5	16.18	0.0038
Scenario 6	13.64	0.0245
Scenario 7	30.78	0.0021
Scenario 8	24.49	0.0045
Scenario 9	21.53	0.0452
Scenario 10	28.87	0.0016
Scenario 11	24.48	0.0045
Scenario 12	21.39	0.0726

TABLE V
THE SPECIFICATIONS OF THE FIXED AGGREGATE INSTALLED CAPACITY
SCENARIOS AND RSDS FOR EACH SCENARIO.

WPP Geograph. Distr.	PV Geograph. Distr.	WPP installed cap. (%)	PV installed cap. (%)	RSD
Dispersed	Concentrated	50	50	0.5456
Dispersed	Concentrated	60	40	0.5067
Dispersed	Concentrated	70	30	0.5032
Dispersed	Dispersed	50	50	0.5312
Dispersed	Dispersed	60	40	0.4977
Dispersed	Dispersed	70	30	0.4982
Concentrated	Concentrated	50	50	0.6174
Concentrated	Concentrated	60	40	0.6008
Concentrated	Concentrated	70	30	0.6141
Concentrated	Dispersed	50	50	0.6081
Concentrated	Dispersed	60	40	0.5961
Concentrated	Dispersed	70	30	0.6119

for the lost benefit from the negative correlations between WPPs and PVPs. This is a similar balancing between the volatility in individual locations and the correlations between the locations as was studied with wind generation in [16].

Table IV shows the probabilities for events where the aggregated power generation is less than 20% or exceeds 80% of the installed capacity for three consecutive hours. It can be observed that the probabilities for the generation remaining less than 20% for at least three consecutive hours are large. However, the probabilities for events where the generation remains more than 80% of the capacity for at least three consecutive hours are very low. It can be seen that the dispersed geographical distribution for both generation types and the 90% share of WPPs are both beneficial for reducing the likelihood of the low generation events. On the other hand, the concentrated geographical distribution of WPPs increases the likelihood of the high generation events.

To summarize the results, the smallest volatility is achieved in Scenario 5, where both the WPPs and PVPs are dispersed and the WPPs produce 80% and PVPs 20% of the annual energy. The inclusion of both generation types is beneficial due to the negative spatial correlations between wind speeds and global irradiance, as shown in Fig. 5. In addition, the geographical distribution of the WPPs seems to have a larger impact on the volatility of the generation compared to the distribution of the PVPs.

C. The Fixed Aggregate Installed Capacities Instead of Fixed Yearly Energies

The fixed energy enables a clear comparison between the cases, but the power system operator might be also interested in a similar comparison, but with a fixed installed capacity (power) instead of the yearly energy. Scenarios with the same simulation setup as with the fixed yearly energy were analysed also with an aggregated fixed capacity of 1200 MW in each scenario, so that the proportional installed capacity of WPPs and PVPs varied between the scenarios, as shown in Table V.

The different capacity factors (the ratios between installed capacity and the actual output) made the comparison more difficult, as the PVPs had an average capacity factor of 0.1195 and the WPPs of 0.3132. Therefore, the RSD was used when comparing the volatility between the scenarios. As shown in Table V, the scenarios where both WPPs and PVPs have a dispersed geographical distribution have also the lowest volatility compared to other geographical distributions. This is the same observation as with the fixed yearly energy. In addition, for fixed power the most optimal proportional capacities were 60% of WPPs and 40% PVPs for three out of the four different combinations of geographical distributions. The lowest hourly volatility, in terms of RSD, was achieved in the scenario where both locations had dispersed geographical distribution and the proportional capacities were 60% of the WPPs and of 40% the PVPs.

VII. CONCLUSION

This paper has introduced an MC simulation based methodology to model systems with WPPs and PVPs in multiple existing or new locations. A VAR model was presented to capture the temporal correlations in individual locations and the spatial correlations between the locations. The spatial correlations were linked to the geographical distances between the locations to enable the addition of non-measured generation locations to the model.

The model was verified against two out-of-sample test locations, one wind and one solar location, which were both excluded from the estimation. It was shown that the model was able to produce simulated data with the correct temporal and spatial correlation structures and probability distributions for the test locations.

The methodology can be utilized in long-term simulations and scenario analyses focusing on the variability of the aggregate RES generation in power systems with multiple WPPs and PVPs. The methodology can be applied in any geographical area where measurement data for the estimation of the model parameters are available. The modular structure of the

model allows it to be combined with several different wind turbine or PV panel models, as the model provides simulated wind speed and global irradiance data instead of power data. The methodology can be beneficial for both transmission and distribution system operators alike, and for power producers that have both wind and solar energy generation in their portfolio.

In addition, a case study was conducted with 12 scenarios assessing different geographical distributions of wind and PV generation with various proportional annual generated energies. It was found that the scenario where WPPs and PVPs were both geographically dispersed and the WPPs generated 80% and the PVPs 20% of the annual energy was the most favourable in terms of small variability. The dispersed geographical spread of WPPs was more important than the spread of PV generation for decreasing the variability of the aggregate RES generation.

ACKNOWLEDGMENT

REFERENCES

[1] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, February 2009.

[2] K. Xie, Y. Li, and W. Li, "Modelling wind speed dependence in system reliability assessment using copulas," *IET Renewable Power Generation*, vol. 6, no. 6, pp. 392–399, November 2012.

[3] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, April 2014.

[4] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium (NAPS), 2010*, Sept 2010, pp. 1–8.

[5] D. Lingfors and J. Widén, "Development and validation of a wide-area model of hourly aggregate solar power generation," *Energy*, vol. 102, pp. 559 – 566, 2016.

[6] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1337–1345, Oct 2015.

[7] M. Larrañeta, S. Moreno-Tejera, M. Silva-Pérez, and I. Lillo-Bravo, "An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series," *Solar Energy*, vol. 122, pp. 517 – 528, 2015.

[8] C. M. Fernández-Peruchena, M. Gastón, M. Sánchez, J. García-Barberena, M. Blanco, and A. Bernardos, "Mus: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations," *Solar Energy*, vol. 120, pp. 244 – 256, 2015.

[9] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data," *Solar Energy*, vol. 115, pp. 229 – 242, 2015.

[10] J. Ekström, M. Koivisto, I. Mellin, J. Millar, E. Saarijärvi, and L. Haarla, "Assessment of large scale wind power generation with new generation locations without measurement data," *Renewable Energy*, vol. 83, pp. 362 – 374, 2015.

[11] J. Ekström, M. Koivisto, J. Millar, I. Mellin, and M. Lehtonen, "A statistical approach for hourly photovoltaic power generation modeling with generation locations without measured data," *Solar Energy*, vol. 132, pp. 173 – 187, 2016.

[12] M. Marinelli, P. Maule, A. N. Hahmann, O. Gehrke, P. B. Nørgård, and N. A. Cutululis, "Wind and photovoltaic large-scale regional models for hourly production evaluation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 916–923, July 2015.

[13] P. E. Bett and H. E. Thornton, "The climatological relationships between wind and solar energy supply in Britain," *Renewable Energy*, vol. 87, Part 1, pp. 96 – 110, 2016.

[14] J. Widen, "Correlations between large-scale solar and wind power in a future scenario for Sweden," *IEEE Trans. Sustain. Energy*, vol. 2, no. 2, pp. 177–184, April 2011.

[15] J. Vasilj, P. Sarajcev, and D. Jakus, "Estimating future balancing power requirements in wind–pv power system," *Renewable Energy*, vol. 99, pp. 369 – 378, 2016.

[16] M. Koivisto, J. Ekström, J. Seppänen, I. Mellin, J. Millar, and L. Haarla, "A statistical model for comparing future wind power scenarios with varying geographical distribution of installed generation capacity," *Wind Energy*, vol. 19, pp. 665–679, May 2016.

[17] M. Koivisto, J. Seppänen, I. Mellin, J. Ekström, J. Millar, I. Mammarella, M. Komppula, and M. Lehtonen, "Wind speed modeling using a vector autoregressive process with a time-dependent intercept term," *International Journal of Electrical Power Energy Systems*, vol. 77, pp. 91 – 99, 2016.

[18] D. Villanueva, A. Feijóo, and J. L. Pazos, "Simulation of correlated wind speed data for economic dispatch evaluation," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, January 2012.

[19] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, "Application of auto-regressive models to U.K. wind speed data for power system impact studies," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 134–141, January 2012.

[20] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2007.

[21] J. Hamilton, *Time Series Analysis*. Princeton University Press, 1994.

[22] D. A. Dickey and W. A. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, vol. 49, no. 4, pp. 1057–1072, 1981.

[23] A. Harvey, *Time series models*. Philip Allan, 1981.

[24] G. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Econometrica*, vol. 66, pp. 67–72, 1978.

[25] B. Tammelin, T. Vihma, E. Atlaskin, J. Badger, C. Fortelius, H. Gregow, M. Horttanainen, R. Hyvönen, J. Kilpinen, J. Latikka, K. Ljungberg, N. G. Mortensen, S. Niemelä, K. Ruosteenoja, K. Salonen, I. Suomi, and A. Venäläinen, "Production of the finnish wind atlas," *Wind Energy*, vol. 16, no. 1, pp. 19–35, 2013.

[26] C. Rigollier, O. Bauer, and L. Wald, "On the clear sky model of the {ESRA} — european solar radiation atlas — with respect to the heliosat method," *Solar Energy*, vol. 68, no. 1, pp. 33 – 48, 2000.

[27] Solar radiation data (soda) service. Cited Aug. 2016. [Online]. Available: http://www.soda-is.com/eng/services/

[28] Gamesa G128-5.0 mw wind turbine data sheet. Cited Aug. 2016. [Online]. Available: http://www.gamesacorp.com/recursos/doc/productos-servicios/aerogeneradores/catalogo-g10x-45mw-eng.pdf

[29] M. Mattei, G. Notton, C. Cristofari, M. Muselli, and P. Poggi, "Calculation of the polycrystalline {PV} module temperature using a simple method of energy balance," *Renewable Energy*, vol. 31, no. 4, pp. 553 – 567, 2006.