Eseye, Abinet Tesfaye; Lehtonen, Matti; Tukia, Toni; Uimonen, Semen; John Millar, R.

Adaptive Predictor Subset Selection Strategy for Enhanced Forecasting of Distributed PV Power Generation

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2019.2926826

Published: 01/01/2019

Please cite the original version:
Adaptive Predictor Subset Selection Strategy for Enhanced Forecasting of Distributed PV Power Generation

ABINET TESFAYE ESEYE, MATTI LEHTONEN, TONI TUKIA, SEMEN UIMONEN, AND R. JOHN MILLAR
Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland
Corresponding author: Abinet Tesfaye Eseye (abinet.eseye@aalto.fi)

This work was supported by the Department of Electrical Engineering and Automation, Aalto University, Finland.

ABSTRACT
Distributed photovoltaic (PV) solar power plants are playing an increasing role as a power generation resource in the modern electricity grid. However, PVs pose significant challenges to grid planners, operators, owners, investors, aggregators, and other stakeholders. This is due to the high uncertainty of the PV output power, which is caused by its entire dependence on intermittent environmental factors. This has brought a serious problem to the power industry to integrate and manage power grids containing significant penetration of PVs. Thus, an enhanced PV power forecast is very important to operate these power grids efficiently and reliably. Most previous methodologies have focused on predicting the aggregate amount of potential solar power generation at the national or regional scale and ignored the distributed PVs that are installed primarily for local electric supply. Furthermore, a few research groups have carried out predictor selection before training predictive models. This paper proposes an adaptive hybrid predictor subset selection (PSS) strategy to obtain the most relevant and nonredundant predictors for enhanced short-term forecasting of the power output of distributed PVs. In the proposed strategy, the binary genetic algorithm (BGA) is applied for the feature selection process and support vector regression (SVR) is used for measuring the fitness score of the predictors. In order to validate the effectiveness of the proposed strategy, it is applied to actual distributed PVs located in the Otaniemi area of Espoo, Finland. The findings are compared with those achieved by other PSS techniques. The proposed strategy enhances the quality and efficiency of the predictor subset selection, with minimal chosen predictors to achieve enhanced prediction accuracy. It outperforms the other prediction selection methods. Besides, a configuration of an adaptive forecasting model is introduced and the performance tests are presented to further validate the impact of the PSS results for the PV power prediction accuracy enhancement.

INDEX TERMS
Adaptive model, big data, binary genetic algorithm, distributed PV, forecasting, fitness evaluation measure, predictor subset selection, renewable energy, smart grid, solar energy, support vector regression.

NOMENCLATURE

\[ \Delta \text{acc} \quad \text{Prediction accuracy enhancement} \]
\[ \Delta D \quad \text{Dimensionality reduction} \]
\[ \Delta \text{fit} \quad \text{Fitness value improvement} \]
\[ \Delta \text{comp} \quad \text{Computation time reduction} \]
\[ \varepsilon \text{-SVR} \quad \text{Linear-epsilon-insensitive SVR} \]
\[ \xi_1, \xi_2 \quad \text{SVR model auxiliary parameters} \]
\[ \gamma \quad \text{SVR model normalization parameter} \]
\[ \Phi \quad \text{SVR model mapping function} \]
\[ \varepsilon \quad \text{SVR model loss parameter} \]

\[ \alpha, \alpha^* \quad \text{Lagrange multipliers} \]
\[ \beta \quad \text{Percentage forecast error} \]
\[ \sigma \quad \text{Gauss parameter (width of RBF kernel)} \]
\[ \lambda \quad \text{NCA model regularization parameter} \]
\[ \rho \quad \text{NCA model probability parameter} \]
\[ \text{acc}_{\text{with PSS}} \quad \text{Prediction accuracy with PSS} \]
\[ \text{acc}_{\text{without PSS}} \quad \text{Prediction accuracy without PSS} \]
AI: Artificial intelligence
ACO: Ant colony optimization
ANN: Artificial neural network
Installation of renewable energy resources, in particular solar energy, has received much attention globally due to several environmental protocols agreed by almost all countries as primary directives of the United Nation (UN). This is because electricity generation from solar energy is clean, accessible nearly everywhere, has a simple structure, and does not require a prime-mover. Besides, the advent of power electronics and its associated control technology has further accelerated the rapid deployment of solar generation systems globally. Although solar power generation has significant environmental advantages and is a promising source of energy for the future, its uncertainty due to intermittency of weather variables makes it more challenging to utilize than the conventional generation sources. This is due to the uncertainty of the generation causes large problems on the power grid stability and control.

However, this problem is not insurmountable. To harness the benefit and increase the competitiveness of the solar energy, accurate forecast of the solar generation is essential. Accurate solar power forecast enhances the control, stability, reliability and flexibility of power grids with a large penetration of PV power. Accurate forecasts assist the various stakeholders involved in the power industry to make better decisions on power system investment, planning, operation, management, economics, market, strategy, and risk analysis. Thus, accurate prediction of PV power plays a key role in power grids containing a huge penetration of PV solar power.

Selecting suitable input variables or a predictor subset is currently a very important research and development (R&D) topic in the field of PV power forecasting. Choosing the best predictor subset from a large number of predictors to constitute the input dataset for PV power forecasting enhances the prediction performance.

This calls for R&D in effective and applicable predictor subset selection (PSS) strategies and enabling tools for enhancing the existing accuracy levels of PV power forecasting.

Predictor subset selection is a process of picking a subset of most important predictors (features, attributes, or variables) for use in forecasting model development.
Different research groups have performed various PSS methods for various applications and scenarios. However, very few of them have coupled and investigated PSS tools and forecasting models. Moreover, there is no standard and universally agreed PSS method so far. The R&D for finding the most effective PSS tools is still ongoing by various independent research groups and institutions.

Predictor subset selection strategies are important for forecasting problems and big data analysis because they:
- Decrease computation time, storage requirement and overfitting
- Simplify models and evade curse of dimensionality
- Enhance data understandability, interpretability and generalization

The core argument when applying a PSS method is that the original dataset holds some variables that are either duplicated or not important, and can therefore be eliminated without inducing ample damage in the information. It has been proved by several research groups that redundant and irrelevant features reduce the accuracy and generalization capability of forecasting models. That is why, nowadays, PSS studies have become very popular in AI, Machine Learning (ML), Deep Learning (DL), and Statistics.

The techniques being used for forecasting the future power production in distributed PV systems have a great impact on achieving the best economic benefit and energy flexibility of PV systems. However, PSS strategies and enabling tools for the PV power forecasting models have not yet been investigated deeply and the results so far in this regard are not adequate.

Most prior works on PV power forecasting approaches used a predetermined and user-defined set of variables as inputs for the forecast models. They did not utilize PSS techniques to choose the forecast model input variables or predictors, which would have a significant improvement on the obtained forecasting accuracy.

Therefore, the goal of this paper is to propose and implement a predictor subset selection approach for modeling and forecasting the uncertain generation power in distributed PVs in general, and building rooftop PVs in particular. The results will assist the various PV stakeholders in having accurate PV power forecast models that will aid the efficient use of limited energy resources and the regulation of dispatchable generation and flexible demand levels.

Prediction accuracy is the indispensable target in forecasting studies. It is soundly revealed in [1] and [2] that the accuracy of prediction models not only relies on the models’ configurations and associated learning methods but also on the predictor domain, which is established via the initial predictor space and PSS techniques. PSS is mostly applied in ML implementations as one of the preprocessing steps, where a predictor subset (independent attributes) is found by removing predictors with lower or irrelevant information and highly redundant information [3]. However, very few forecasting techniques perform PSS before training the prediction models.

Meta-heuristic optimization algorithms have become very popular and significantly effective for various problems in the power/energy sector, especially in the field of renewable energy generation [4], [5]. They have been effectively implemented as searching techniques for PSS problems. For instance, these methods contain Particle Swarm Optimization (PSO) [6], Ant Colony Optimization (ACO) [7] and the Genetic Algorithm (GA) [8]. GA has gained extensive consideration due to its operability and robust searching ability. GA is one of the artificial intelligence (AI) probabilistic searching algorithms, and has been broadly implemented for several optimization problems [9]. BGA is a special version of GA which operates by first representing the given predictor space (chromosomes or candidate solutions) in binary bitstrings. This makes the BGA better suited for PSS problems than the conventional GA.

PSS methods are classified as filter, wrapper and embedded techniques [1].

Filter techniques do not depend on any prediction model and they sort features depending on statistical characteristics. They utilize a correlation score to grade a feature subset. Filter technique based PSS methods are generally fast. The Filter PSS approach includes correlation-based [10], mutual information-based [11], and principal component analysis-based methods [12]. Filters generally require less computation time than other PSS techniques, but they generate a predictor set which is not fitted to a particular forecast model. Wrapper techniques evaluate predictor subsets based on their worth to a specific forecaster or classifier. Wrapper techniques assume the PSS to be a searching problem that prepares various mixes of predictors, which are assessed and contrasted with other mixes. The common heuristic AI-based optimization methods mentioned above are used to monitor the searching procedure. Compared to filter techniques, wrapper techniques reveal improved performance, since various predictor sets are assessed by a predictive model or fitting method in every iteration [13]. Embedded techniques merge the predictor selection process into the training task of prediction models. For instance, the regularization approaches in [1] are one example of an embedded type PSS method. Table 1 presents recent works on PSS strategies for forecasting problems.

<table>
<thead>
<tr>
<th>Application</th>
<th>PSS Type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power systems</td>
<td>Filter</td>
<td>[10], [14], [15]</td>
</tr>
<tr>
<td></td>
<td>Wrapper</td>
<td>[16], [17]</td>
</tr>
<tr>
<td>Energy systems</td>
<td>Filter</td>
<td>[18], [19]</td>
</tr>
<tr>
<td></td>
<td>Wrapper</td>
<td>[20], [21]</td>
</tr>
</tbody>
</table>

The works proposed in [22]–[30] have implemented the GA-based PSS in different application domains and scenarios.

Following a comprehensive assessment of the above-mentioned genetic algorithm based PSS techniques, we discover that a conventional genetic algorithm with the usual
framework (conventional GA configuration) has been used in most research. For instance, the initial population (initial chromosome set) is arbitrarily created where the population variety cannot be guaranteed and the occurrence of duplicated predictors may influence the quality of the search procedure. Moreover, the conventional GA works with the continuous features themselves to minimize the desired fitness function (PSS evaluation measure). This reduces the efficiency of the algorithm and causes computational complexity and increased total computation time.

Assuming adaptive heuristic algorithms should be among the best options to determine the search target; a research problem exists and can be addressed by replacing the conventional GA with the BGA and hybridizing it with robust fitness evaluation measures. The BGA first represents the predictors as encoded binary strings, and works with the binary strings to minimize the SVR-based evaluation measure to obtain a relevant and nonredundant predictor subset at the end. BGA is more efficient and stable than the conventional GA. It also reduces computational complexity and execution time compared to the conventional GA.

Therefore, this paper proposes an adaptive hybrid predictor subset selection strategy to obtain the most relevant and nonredundant predictors for enhanced short-term forecasting of the power output of distributed PVs. In the proposed strategy, the Binary Genetic Algorithm (BGA) is applied for the predictor selection process and Support Vector Regression (SVR) is used for measuring the fitness score of the features.

To the best of our understanding, there exist very few research works that have performed PSS work before fitting or training forecasting models. Moreover, as far as we have investigated, the BGA-SVR based hybrid machine learning approach has never been applied for PSS problem in the domain of renewable energy generation forecasting. Generally, the paper’s contribution can be considered as (1) modeling, parameterization and implementation of the BGA and SVR algorithms to suit the predictor selection problem in question, and (2) establishment of seamless combination of the two algorithms to work in unison for solving the predictor selection problem. Therefore, from application and hybridization point of view, this is the first work to hybridize the BGA and SVR algorithms for PSS problem in the domain of electric power system research.

Specifically, the paper contributions can be summarized as follows:

- Analyze and recommend the relevance of an effective PSS strategy and enabling tools for enhanced performance or accurate PV power forecasting;
- Present an effective and efficient machine learning-based adaptive PSS strategy for PV power forecasting;
- Enhance PV power forecasting accuracy through the application of PSS before training forecasting models.

The rest of the paper is organized as follows. Section II describes the dataset and states the PSS problem. Section III presents the brief working principle of the BGA. Similarly, the theory and mathematical modeling of the SVR model used for the fitness measure in the BGA is described in Section IV. Section V presents the proposed BGA-SVR based PSS strategy. The achieved experimental results and validations are presented in Section VI. The paper is concluded in Section VII.

II. DATASET AND PREDICTOR SELECTION PROBLEM

The original predictor set is constructed through basic assessment of the characteristics of the power production of distributed PV systems and its association with external agents. The external agents are seasonality (minute/hour, month and season) and weather factors. The availability of the data sources for these external agents affecting the PV power production is also another major factor to construct the original predictor space.

The candidate original predictor set for the PV power forecasting in this PSS work consists of seasonal (or calendar) parameters and weather parameters. The variables \( f_i, i = 1, 2, \ldots, 20 \), in Table 2 represent the original predictor dataset (predictor space) required for the PSS work in this paper.

### TABLE 2. Predictor space of the PSS problem.

<table>
<thead>
<tr>
<th>Predictor Index</th>
<th>Predictor (f)</th>
<th>Unit/Scale</th>
<th>Data Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hour of the day</td>
<td>1-24</td>
<td>Seasonality/calendar</td>
</tr>
<tr>
<td>2</td>
<td>Month of the year</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Season of the year</td>
<td>1-4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Ambient air temp.</td>
<td>°C</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Dew-point temp.</td>
<td>°C</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Relative humidity</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Precipitation</td>
<td>mm/h</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Snow depth</td>
<td>cm</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Ambient air pressure</td>
<td>kPa</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Horizontal visibility</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Wind direction</td>
<td>deg</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Wind speed</td>
<td>m/s</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Gust speed</td>
<td>m/s</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Cloud cover</td>
<td>0-8</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Diffuse solar radiation</td>
<td>Watt/m²</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Direct solar radiation</td>
<td>Watt/m²</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Global solar radiation</td>
<td>Watt/m²</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Long wave solar radiation</td>
<td>Watt/m²</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>UV radiation index</td>
<td>≥ 0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Sunshine duration</td>
<td>s</td>
<td></td>
</tr>
</tbody>
</table>

Therefore, the predictor space of the PSS is an \( R^{m \times n} \) matrix, where \( m = 192 \) is the number of samples of the predictors and \( n = 20 \) is the size of the predictor space (original dataset). The samples are hourly observations of the predictors collected for two randomly chosen days from each season of a year. That means hourly values of the predictors for a total of eight days are used to form the predictor space. These days are:

- Fall: October 4, 2017 and October 12, 2017,
- Winter: January 7, 2018 and January 11, 2018,
- Spring: April 21, 2018 and April 26, 2018, and
- Summer: July 10, 2018 and July 11, 2018.

In this paper, the following optimization problem is solved to find the best (relevant and nonredundant) predictor subset from the original dataset given in Table 2.
PSS Problem:
Given that:

\[ f_r \in \mathbb{Z}^+, \quad 1 \leq f_r \leq 20, \quad \text{and} \quad \beta \in \mathbb{R}^+ \ni 0 \leq \beta \leq 100 \quad (1) \]

where, \( f_r \) is the number of predictors in the lower-dimension (reduced) predictor subset and \( \beta \) is the percentage forecast error. Find a predictor subset of \( f_i \) from Table 2 such that the objective \( \beta \) and \( f_r \) are reduced.

III. BINARY GENETIC ALGORITHM (BGA)
GA is a population-based heuristic optimization method that is inspired by the survival of the fittest principle of the Charles Darwin theory of evolution and genetics [31]. The GA operating mechanism involves iterative steps processing a set of chromosomes (candidate solutions) to generate a new population (offsprings) via genetic operators - selection, crossover and mutation. The fitnesses of the nominee solutions (chromosomes) are calculated employing an objective or fitness function, meaning that the objective function provides scores (numeric values) which are employed for grading the existing solutions in the population. BGA is an extended version of the standard GA. BGA first represents the candidate solutions as encoded binary strings (binary search space) and works with the binary strings to minimize or maximize the fitness function. BGA is more efficient and stable. It also reduces computational complexity and execution time. Figure 1 shows the flowchart of BGA.

IV. SUPPORT VECTOR REGRESSION (SVR)
SVR is a non-parametric method that essentially depends on a kernel function. Vapnik [32] established the essentials of SVRs in 1995. SVRs are gaining significant credit at the time of writing due to a number of noticeable characteristics and promising hands-on performances. SVR has been effective implemented to perform prediction tasks and pattern classifications, mainly the clustering of two unlike pattern categories. Their formulation comprises the structural-risk-minimization (SRM) theory, which has been proved to be superior to the standard empirical-risk-minimization (ERM) theory utilized by conventional ANNs [33].

Linear-regressions in the upper-dimension hyperplanes are associated with non-linear regressions in the lower-dimension plane, and are articulated below [34].

\[ y(x) = w \cdot \Phi(x) + b; \ \Phi : \mathbb{R}^n \rightarrow \mathbb{R}^N \]

where, \( y \in \mathbb{R}^N \) is a training target; \( x \in \mathbb{R}^n \) is a training input (predictor); \( b \) is a bias parameter; \( w \in \mathbb{R}^N \) is weight/coefficient parameter; \( \Phi(x) \) is a non-linear mapping-function; and \( \Phi: \mathbb{R}^n \rightarrow \mathbb{R}^N \) is a non-linear mapping that converts the initial training inputs to the upper-dimension characteristic hyperspace.

Figure 2 illustrates the configuration of a SVR, where input \( x \) is transformed into output \( y \) via the mapping-function \( \Phi(\cdot) \). The yield of the regression \( y \) is the linear integration of scaled \( \Phi(x) \).
solving its equivalent dual-problem defined below.

\[
\min \left\{ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (a_i - a_j^\ast) \cdot \Phi \left( x_i, x_j \right) \cdot (a_j - a_j^\ast) \right\}
\]

subject to : \( \sum_{i=1}^{N} (a_i - a_i^\ast) = 0; a_i, a_i^\ast \geq 0 \) (4)

Solving for the positive Lagrange-multipliers \((a_i- a_i^\ast)\), the final formulation of the SVR output \(\hat{y}\) is described by:

\[
\hat{y} (x) = \sum_{i=1}^{N} (a_i - a_i^\ast) \cdot K \left( x, x_i \right) + b
\] (5)

where, \( K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \) is known as the kernel-function of the SVR model.

From the Karush-Kuhn-Tucker (KKT) optimality condition [34] for quadratic-programming type objective functions, all the terms \((a_i - a_i^\ast)\) cannot have non-zero values. The radial basis function (RBF) kernel expressed below is used in this study.

\[
K \left( x_i, x_j \right) = \exp \left( -\frac{\| x_i - x_j \|^2}{\sigma^2} \right)
\] (6)

where, \(\sigma\) is a Gauss parameter (width of RBF kernel) and defines the impact area of the support-vectors in the training window domain.

The SVR model parameters are obtained by solving the optimization problem formulated in (3).

V. PROPOSED BGA-SVR BASED PREDICTOR SUBSET SELECTION STRATEGY

As shown by the flowchart in Figure 3, there are five key sub-operations in BGA: chromosome encoding, objective value calculation, selection methods, genetic operators and stopping condition. The BGA works in a binary search domain (chromosome bitstrings), and operates the finite binary chromosome set based on the survival of the fittest principle. A starting population is generated and assessed using an objective function (BGA driver) should be specified to calculate the fitness of each chromosome. For the binary chromosome employed in this paper, a gene value of ‘1’ indicates that the specific feature pointed to by the location of the ‘1’ is chosen. Else, (if ‘0’), the feature is not chosen for the fitness evaluation.

Employing the place pointer of the variables pointed by the ‘1s’, the individuals are then ordered and the \(k\) fittest offsprings (Elitism of size \(k\)) are chosen to persist in the succeeding generation. Once the selected offsprings are moved directly to the succeeding generation, the other offsprings in the present solution space are permitted to genetically move via the crossover and mutation operators to create crossover and mutation offsprings respectively [31]. The three offsprings, selection, crossover and mutation, then establish the new solution space (new generation). The crossover operator is a fusion of two chromosomes to create crossover offsprings. The mutation operator is employed for genetic disorder (diversity) of the genes in the chromosomes by tossing the bits based on the mutation likelihood. Following the procedures outlined in Figure 3, the detailed operating mechanisms of the proposed BGA-SVR PSS are described in the following subsections.

A. INITIAL POPULATION

The BGA starting solution space used in this work is a matrix of size \(p \times q\), where \(p\) is the number of chromosomes and \(q\) is the chromosome length (called Genomelength). \(p\) equals the population size and \(q\) equals the number of bits or genes in each individual. It is recommended to let the number of chromosomes be equal to at least the chromosomes’ length, such that the chromosomes in every population encompass the search domain [35].

B. FITNESS EVALUATION

For the BGA to choose the predictor subset, an objective function (BGA driver) should be specified to calculate the discriminative power of each predictor subset.

The fitness of each chromosome in the population is evaluated employing an SVR-based fitness function. In this paper, the fitness of the various subsets of predictors is evaluated...
using the MSE (mean squared error) of the SVR model residuals. The SVR model output \( y(x) \) is fitted for every predictor subset.

Hence, the MSE of the training target and the SVR model estimate evaluated for each predictor subset in the predictor search space defined in Table 2 are used as the fitness evaluation measure, defined as follows.

\[
fit = \frac{1}{N} \sum_{i=1}^{N} (T_i - y_i)^2
\]  

(7)

where, \( T \) is a vector of training target (PV power) and \( N \) is the number of training samples or observations.

The aim of the BGA is to minimize the fitness function (MSE) defined in (7) by choosing a subset of input predictors having the best fitness over subsequent iterations. In each chromosome, a gene value of ‘1’ shows the specific predictor pointed by the place of ‘1’ is chosen. If it is ‘0’, the predictor is not chosen for assessment of the chromosome in question. The chromosomes representing the predictors are encoded as bitstrings.

While the BGA runs, the individual chromosomes (feature subsets) in the present population are assessed, and their fitnesses are graded based on the SVR model residual or error. Chromosomes with smaller fitness (smaller residual or error) have a greater probability of persisting in the next population or mating-pool.

Each iteration of the BGA running guarantees that the BGA decreases the error level and classifies the chromosome with the lowest (best) objective function value as Elite. This is because the error level is stated for each individual engaged and the least error level is obtained by the BGA at the end. The individual chromosome corresponding to the least error level of the fitness evaluation contains the most relevant desired predictors.

C. REPRODUCTION

Table 3 presents the parameters of the BGA used in this paper. From Table 3, the chromosome length equals 20, as there are an overall number of 20 predictors nominated for the PSS work in this paper. Following the fitness evaluation, a new population is produced for the next generation through elitism, crossover and mutation.

In BGA, three kinds of sequential offsprings are formed to create the new population [35]. They are:

1) **Elite offspring**: Tournament Selection Mechanism (with size 2) is used in this study because of its ease-of-use, swiftness and efficiency [29], [36]. Hence, the upper 2 offsprings with the best fitness scores are directly taken into the following generation. Therefore, the quantity of the elite offsprings (elite count) = \( O_1 = 2 \). That is, there are 18 (i.e. 20 - O_1) chromosomes in the population in addition to the elite offsprings. From the other 18 individuals, crossover and mutation offsprings are then generated.

2) **Crossover offspring**: The crossover function used in this paper is of the arithmetic type, which applies a logical XOR operation on the chromosomes of the two parents, as they are represented in binary form. The portion of the following generation (excluding the elite offsprings) created by the crossover operator is known as crossover offspring. The crossover fraction, which refers to the ratio of the number of crossover offsprings, is taken as 0.8. With the crossover ratio of 0.8, the number of the crossover offsprings is \( O_2 = \text{round}(18 \times 0.8) = 14 \).

3) **Mutation offspring**: The BGA implemented in this study uses uniform mutation. Using uniform mutation, the BGA creates a set of uniformly distributed random numbers whose size equals the length of the chromosomes. The quantity of mutation offsprings is \( O_3 = 20 - O_1 - O_2 = 20 - 2 - 14 = 4 \). This is verified by \( O_1 + O_2 + O_3 = 20 \).

D. CONVERGENCE CONDITION

The BGA terminates when it converges to the desired optimal solution. The optimal solution corresponds to the desired predictor subset for the PSS problem in question. The termination condition where the BGA ends running is known as the convergence or stopping condition. The two convergence conditions used in this paper are the following:

1) Maximum number of generations or iterations
2) Stalled generation limit

The values used for these convergence conditions are given in Table 3.

E. FINAL PREDICTOR SUBSET

After the BGA attains convergence, the chromosome that resulted in the best fitness score is chosen and decoded as the final predictor subset, shown in Figure 4.

VI. EXPERIMENTAL RESULTS AND VALIDATION

In this section, the case study for the proposed PSS work and the results obtained are discussed. Comparative validation, configuration of an adaptive PV power forecasting model based on the PSS results and quantitative relevance analysis of the PSS results are also presented in this section.
In this paper, the hybrid BGA-SVR based PSS strategy is developed and implemented based on a pilot distributed PV system installed on a building rooftop located in the Otaniemi area of Espoo, Finland. The PV system has a peak generation capacity of 4.3kW.

The original predictor space for the PSS work is described in Table 2. The amount of the PV power production is the desired target variable in the proposed PSS strategy. Hourly samples from eight days, 192 values, of both the predictor set and target variable are used in the PSS.

The empirical results achieved by the proposed PSS method are presented in Table 4. As is clearly observed from the PSS result in Table 4, the number of predictors chosen by the proposed PSS strategy is considerably lower than the size of the predictor space (the number of predictors in the original dataset is given in Table 2). This can be due to irrelevant and redundant information in most of the variables in the original predictor space. The BGA-SVR finally selects the predictor subset which contains the most relevant and nonredundant variables. A predictor subset consisting of predictors 1, 2, 3, 4, 8, 14, 17, and 20, which represent hour of the day, month of the year, season of the year, ambient air temperature, snow depth, cloud cover, global solar radiation, and sunshine duration, respectively, is selected by the devised BGA-SVR based PSS method. This selected predictor subset can therefore establish an appropriate input dataset for improved PV power forecasting.

Figure 5 shows the BGA objective function value (SVR model based MSE function formulated in (7)) over generations. Besides, the average computation time of the devised integrated BGA-SVR based PSS algorithm with eight-days long hourly sample of 20 initial predictors is about 5 minutes, using MATLAB simulation environment on a research workstation with Intel Core i7-6820HQ Processor, 2.70 GHz CPU, 16 GB RAM.

To validate the BGA-SVR PSS work in this paper, the predictor subset result by the proposed BGA-SVR PSS is compared with predictor subset results using two other common PSS techniques, namely: Correlation-based predictor subset selection (C PSS) and Neighborhood Component Analysis Regression-based predictor subset selection (NCA PSS).

The Correlation-based PSS first calculates the Pearson and Spearman correlations of each predictor with the target, and it then takes the maximum of the two correlation coefficients. The Pearson correlation ($r_P$) evaluates the linear relationship between two variables, while the Spearman correlation ($r_S$) estimates the monotonic relationship between two continuous or ordinal variables.

- The Pearson correlation ($r_P$) is defined as:

$$r_P = \frac{n \left( \sum_{i=1}^{n} x_i y_i \right) - \left( \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} y_i \right)}{\sqrt{\left[n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 \right] \left[n \sum_{i=1}^{n} y_i^2 - \left( \sum_{i=1}^{n} y_i \right)^2 \right]}}$$

where $n$ is sample size, $x$ is the value of the predictor and $y$ is value of the target variable.

- The Spearman correlation ($r_S$) is defined as:

$$r_S = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

where $n$ is the number of samples, $d_i$ is the difference between the predictor value $x$ and the target variable value $y$.

The values of these two correlation coefficients are the same if and only if there exists a linear relationship between the variables. The values of $r_P$ and $r_S$ in the range $[-1 \ 1]$, where a very strong negative relationship exists between the variables, is indicated by ‘−1’ while a very strong positive relationship is indicated by ‘+1’. Table 5 presents the
interpretation of the correlation values, used in this paper, to evaluate the strength of the existing relationship between the predictors and the target variable.

Either of the correlation values can be higher based on the nature of the relationship of the variables. In this paper, the maximum of the two correlation coefficients is used to measure the strength of the relationship between the predictors and target variable as defined below:

\[
Corr = \max\{|r_p|, |r_s|\}
\]  

The correlation coefficient can accurately indicate the strength of the dependency between the predictors and target variable when the existing relationship is linear or monotonic. However, correlation based methods may fail to measure the strength of the dependency between two variables when the relationship is neither linear nor monotonic. They cannot guarantee the existence of redundant information in the predictor set as well. Table 6 presents the correlation coefficient values between the predictors and target variable (PV power produced).

A predictor with correlation value greater than a given threshold value can be selected as a relevant predictor and included in the final predictor subset. The SVR model based fitness evaluation measure (MSE) formulated by equation (7) can be calculated in order to determine the threshold correlation value to select the relevant predictors affecting the PV power production. Table 7 provides the values of the fitness measure for the various predictor subsets for the different correlation coefficients given in Table 6.

As clearly observed in Table 7, the predictor subset consisting of predictors with correlation values greater than or equal to 0.20 achieves the best fitness value, lowest MSE, \(3.8 \times 10^{-3}\). Thus, 0.20 is taken as the threshold correlation value to choose the relevant predictors for the correlation based benchmarking method in this paper. Hence, according to the correlation-based method, predictors 4, 5, 6, 8, 15, 16, 17, 18, 19, and 20 are selected to constitute the input variables for the PV power forecasting. The result of the correlation-based PSS is illustrated in Figure 6.

The NCA PSS is based on the neighborhood component analysis (NCA) regression model fitted over the predictor subsets versus target dataset. The NCA PSS obtains the predictor weights (for reduced predictor subsets) using a diagonal adaptation of the NCA regression model. The NCA model realizes PSS by regularizing the predictor weights. The predictor weight indicates the strength of the relationship of the predictor with the target variable. The predictor weights are obtained by solving the following NCA regression model.
based on the unconstrained stochastic minimization problem:

$$\min f(w) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \rho_{ij} l(y_i, y_j) + \lambda \sum_{r=1}^{p} w_r^2$$  \hspace{1cm} (11)

where, $y_i$ and $y_j$ are the target variable values for predictors $x_i$ and $x_j$ respectively, $w$ is the predictor weight vector, $n$ is the number of samples, $p$ is the number of predictors, $\lambda$ is the regularization parameter, $\rho_{ij}$ is the likelihood that $x_j$ is reference for $x_i$, and $l$ is the loss function. A mean absolute loss function defined below is used:

$$l(y_i, y_j) = |y_i - y_j|$$  \hspace{1cm} (12)

The predictor weights are therefore obtained by solving (11). The limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm is used to solve the stochastic minimization objective function formulated in (11). Figure 7 shows the objective function value as a function of the number of iterations.

The limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm is used to solve the stochastic minimization objective function formulated in (11). Figure 7 shows the objective function value as a function of the number of iterations.

The predictors and their associated weight values are plotted in Figure 8.

As shown in Figure 8, the irrelevant predictors that are not selected by this method are indicated with zero weight values. Predictors whose weight value is not indicated by zero in Figure 8 are chosen. Hence, according to the NCA regression model based PSS, predictors 12, 17 and 19 are selected to constitute the input variables for the PV power forecasting.

Table 8 provides the performance comparison of the PSS result by the proposed method and the other two methods. For the purpose of suitability of comparison, the same fitness function (MSE) modeled as the residual of the SVR model is used. That means that each selected predictor subset by the respective method is evaluated for fitness using the SVR model residual.

As shown in Table 8, the proposed BGA-SVR based PSS achieved the predictor subset with the best fitness value (lowest MSE). Hence, the predictor subset selected by the proposed PSS strategy contains more relevant and non-redundant features than the other PSS methods. That means, a PV power forecasting model whose input dataset constitutes the predictor subset found by the proposed BGA-SVR PSS strategy can achieve accurate prediction results.

### D. PSS RESULTS FOR ENHANCED-ACCURACY AND ADAPTIVE PV POWER FORECASTING

For further validation of the effectiveness of the obtained PSS results, a Feedforward Artificial Neural Network (FFANN) based 24h-ahead PV power forecast model was developed for the case study PV system. The eight predictors selected by the devised BGA-SVR PSS, presented in Section VI B, form the training input dataset for FFANN forecast model. The training target variable is the output power of the PV plant. Eight month’s time series hourly data of the selected predictors and target variable were used to train the FFANN model. The FFANN model parameters was found experimentally. A hidden layer of 10 neurons was used. Moreover, the conventional GA was used to find the optimal weight parameters of the FFANN model.

The proposed model is adaptive, such that it can learn or adapt continuously the changes in the values of the predictor and target variables. It can be retrained periodically when new input datasets are available. This way it can acquire continuous knowledge about the predictor versus PV power production characteristics, and hence improves its prediction performance for future times.

Figure 9 shows the configuration of the PV power forecasting model that uses the selected predictors by the predictor selection strategy proposed and implemented in this paper. The prediction performance of the developed FFANN forecast model was verified with an out-of-sample hourly testing data of four randomly chosen days representing the four seasons of a year. The model testing (forecasting) results are presented with one-hour time resolution, and they are depicted in Figures 10 to 13, for the winter, spring, summer and fall testing days, respectively.
As shown in Figures 10-13, the forecasts follow the actual PV power production trends with smaller gaps (errors) in between. This further verifies the effectiveness of the proposed PSS approach in selecting the best predictor subset that enables the forecast model to achieve improved forecasts that are more accurate.

Furthermore, the following criteria were employed to evaluate the accuracy of the obtained forecasts:

- **Error**

  \[
  \text{Error} = P^a_h - P^f_h \quad (13)
  \]

  where, \( P^a_h \) and \( P^f_h \) are the actual and forecasted values of the PV power production at hour \( h \), respectively.

- **Mean absolute error (MAE)**

  \[
  MAE = \frac{1}{NH} \sum_{h=1}^{NH} | P^a_h - P^f_h | \quad (14)
  \]

  where, \( NH \) is the forecasting horizon and its value is 24 for 24h-ahead forecast.

- **Mean absolute percentage error (MAPE)**

  \[
  MAPE = \frac{100}{NH} \sum_{h=1}^{NH} \left| \frac{P^a_h - P^f_h}{P^a_h} \right| \quad (15)
  \]

  Average MAE of 16.13kWh, MAPE of 4.64%, and daily peak MAPE of 4.72% are obtained for the forecasts of the four testing days using the proposed BGA-SVR PSS results as input dataset for the FFANN based forecast model of the case study local PV system. Hence, the obtained results validate the quality of the predictions and effectiveness of the implemented PSS method, compared to the existing accuracy levels for day-ahead prediction of solar power generation. The numerical analysis of the prediction accuracy improvement is presented next.
E. QUANTITATIVE RELEVANCE ANALYSIS OF PSS RESULTS

In order to quantify the benefits and relevance of the proposed hybrid BGA-SVR based PSS method and the selected predictors, the following metrics are used:

- **Computation time reduction:**
  \[ \Delta t_{\text{comp}} = \frac{t_{\text{without PSS}} - t_{\text{with PSS}}}{t_{\text{without PSS}}} \]  
  \[(16)\]

  where, \( t_{\text{without PSS}} \) is the total computation time which includes data preprocessing, forecasting model training, validation, and prediction using the original predictor space without PSS, \( t_{\text{with PSS}} \) is the total computation time with the use of the obtained PSS results, and \( \Delta t_{\text{comp}} \) is the change in total computation time due to PSS. Positive value of \( \Delta t_{\text{comp}} \) indicates the reduction of computation time requirement of the PV power forecasting model due to making use of PSS results.

- **Dimensionality reduction:**
  \[ D = \frac{R_{m \times n}^{\text{without PSS}} - R_{m \times n_{r}}^{\text{without PSS}}}{R_{m \times n}^{\text{without PSS}}} \]  
  \[(17)\]

  where, \( R_{m \times n}^{\text{without PSS}} \) is a matrix of predictor space without PSS with \( m \) number of samples and \( n \) number of predictors, \( R_{m \times n_{r}}^{\text{without PSS}} \) is a matrix of the reduced predictor space with PSS with \( m \) number of samples and \( n_r \) number of reduced predictors, and \( D \) is the change in data dimension due to PSS. Positive value of \( D \) (\( n - n_r \)) indicates the reduction of input data dimension for the PV power forecasting model.

- **PSS fitness value enhancement:**
  \[ \Delta \text{fit} = \frac{\text{fit}_{\text{without PSS}} - \text{fit}_{\text{with PSS}}}{\text{fit}_{\text{without PSS}}} \]  
  \[(18)\]

  where, \( \text{fit}_{\text{without PSS}} \) is the fitness value of the predictors without PSS with respect to a predefined fitness function (MSE of SVR output and actual target formulated in equation (7)), \( \text{fit}_{\text{with PSS}} \) is the fitness value of the selected predictors with PSS, \( \Delta \text{fit} \) is the change in fitness value due to PSS. Positive value of \( \Delta \text{fit} \) indicates the improvement in fitness value (reduction in MSE value) due to PSS.

- **Prediction accuracy enhancement:**
  \[ \Delta \text{acc} = \frac{\text{acc}_{\text{with PSS}} - \text{acc}_{\text{without PSS}}}{\text{acc}_{\text{without PSS}}} \]  
  \[(19)\]

  where, \( \text{acc}_{\text{without PSS}} \) is the accuracy of the predictions without making use of PSS results (using the original predictor space as training input) and \( \text{acc}_{\text{with PSS}} \) is the accuracy of the predictions with PSS results (using the reduced predictor space as training input). \( \text{acc}_{\text{without PSS}} \) and \( \text{acc}_{\text{with PSS}} \) are defined as follows:

  \[ \text{acc}_{\text{without PSS}} = 100 - \text{MAPE}_{\text{without PSS}} \]  
  \[(20)\]

  \[ \text{acc}_{\text{with PSS}} = 100 - \text{MAPE}_{\text{with PSS}} \]  
  \[(21)\]

  where, \( \text{MAPE}_{\text{without PSS}} \) is the mean absolute percentage error of the predictions without PSS and \( \text{MAPE}_{\text{with PSS}} \) is the mean absolute percentage error of the predictions with PSS. \( \Delta \text{acc} \) is the change in prediction accuracy due to PSS. Positive value of \( \Delta \text{acc} \) indicates the improvement of prediction accuracy due to making use of PSS results in the forecasting process.

Table 9 presents the values of the metrics defined in (16) to (19) to determine the benefits achieved due to the implementation of the devised PSS method for short-term PV power forecasting. It also shows the performance comparison of the PSS results by the proposed method and other conventional counterparts.

As shown in Table 9, the implementation of the PSS and its integration to the forecasting model has resulted in much improvements compared to the forecasting performance using the original dataset without PSS. For example, the enhancement in fitness value (MSE) using the BGA selected predictors to fit the PV power by the SVR model is 64.5% over the original predictor set (without PSS). Similarly, the reductions in computation time and data dimensionality over the original predictor space are 53% and 60%, respectively.

Primarily, the enhancement of the prediction accuracy is the most important and major objective of this paper. The enhancement in prediction accuracy using the BGA-SVR PSS selected predictors to constitute the forecasting model training inputs is 58.4%, compared to prediction accuracy using the original predictor space. Moreover, it is shown that the proposed PSS has given higher performance improvement compared to the other, conventional, counterparts, regarding prediction accuracy and fitness value metrics.

Therefore, the above quantifications and experimental results further validate the relevance and effectiveness of the PSS work for the enhancement of the PV power forecasting.

TABLE 9. Quantitative relevance analysis and comparison of PSS results.

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>BGA-SVR PSS</th>
<th>C PSS</th>
<th>NCA PSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computation time reduction, ( \Delta t_{\text{comp}} )</td>
<td>53</td>
<td>44.2</td>
<td>75</td>
</tr>
<tr>
<td>Dimensionality reduction, ( AD )</td>
<td>60</td>
<td>50</td>
<td>85</td>
</tr>
<tr>
<td>Fitness value enhancement, ( \Delta \text{fit} )</td>
<td>64.5</td>
<td>50</td>
<td>27.6</td>
</tr>
<tr>
<td>Prediction accuracy enhancement, ( \Delta \text{acc} )</td>
<td>58.4</td>
<td>45.3</td>
<td>25</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS

This study devised and developed a BGA based predictor subset selection strategy for enhanced short-term PV power forecasting. The strategy includes the use of an SVR fitness function to choose a combination of predictors from a given original predictor space. A real local PV output power measurement data is used for the PSS work. The devised BGA-SVR PSS has given a predictor subset that resulted in better fitness (lower MSE value) than the original predictor space with all the initial predictors. It achieved the best predictor subset, which can constitute the input variables for accurate forecasting of distributed PV systems. For comparison and validation purposes, predictors selected...
by two other PSS methods were investigated. The BGA-SVR selected predictors outperformed the other predictors with respect to the MSE fitness function defined using the SVR framework. Besides, a FFANN based 24-hour PV power forecast model was developed to evaluate effectiveness of the PSS results. The PV power forecasting model developed using the obtained PSS results has achieved a prediction accuracy improvement of 58.4% compared to forecasting based on the original predictor space without PSS. The devised PSS has also achieved 53%, 60% and 64.5% improvements in computation time, data dimensionality and MSE fitness value, respectively, compared to the original predictor space without PSS. The paper findings confirm that the combination of effective PSS method and forecasting models owns robust forecasting power, compared to forecasting with arbitrary predictors without predictor selection methods. This work is both new and effective from the viewpoints of application in the renewable energy sector and hybridization of algorithms for performance improvement. It contributes a novel and robust predictor selection tool by combining BGA and SVR for enhanced and more accurate forecasting of short-term solar power forecasting.

REFERENCES


ABINET TESFAYE ESEYE received the B.Sc. degree in electrical engineering from Hawassa University, Ethiopia, in 2010, the M.Sc. degree in electrical power systems engineering from Bahir Dar University, Ethiopia, in 2012, and the Ph.D. degree in electric power system and its automation from North China Electric Power University, China, in 2018. He was an Assistant Lecturer with the School of Computing and Electrical Engineering, Bahir Dar University, Ethiopia, from 2011 to 2012. He was a Lecturer and the Head of the Electrical and Computer Engineering Department, Mettu University, Ethiopia, from 2012 to 2014. He was a Renewable Energy and Smart Microgrid R&D Engineer with Goldwind Science and Technology Company Ltd., China, from 2015 to 2018. He is currently a Postdoctoral Researcher with the Department of Electrical Engineering and Automation, Aalto University, Finland. His research experiences and interests include electric power systems, distributed generation, hybrid energy systems, microgrid, renewable energy, the energy Internet, demand response, energy economics and market, techno-economic and feasibility analysis, the Internet of Things (IoT), optimization, computational intelligence, forecasting, state estimation, artificial intelligence, machine learning, and deep learning in energy systems.

MATTI LEHTONEN received the M.S. and Licentiate degrees in electrical engineering from the School of Science and Technology, Aalto University (formerly Helsinki University of Technology), Espoo, Finland, in 1984 and 1989, respectively, and the D.Sc. degree from the Tampere University of Technology, Tampere, Finland, in 1992. Since 1987, he has been with VTT Energy, Espoo. Since 1999, he has been with the School of Electrical Engineering, Aalto University, where he is currently a Professor of IT applications in power systems. His main research interests include earth fault problems, and harmonic-related issues and applications of information technology in distribution automation and distributed energy management.

TONI TUKIA was born in Kirkkonummi, Finland, in 1988. He received the M.Sc. degree in electrical engineering from the School of Electrical Engineering, Aalto University, Espoo, Finland, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University. His research interests include measuring and modeling the energy and power balance of buildings, distributed generation, and the effect of people on the building power consumption.

SEMEN UIMONEN was born in Petrozavodsk, Russia, in 1990. He received the M.Sc. degree in electrical engineering from the School of Electrical Engineering, Aalto University, Espoo, Finland, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University. His research interests include energy and power balance of buildings, the effect of people in the power consumption, and the Internet of Things (IoT) in power systems targeted at energy-efficiency improvements.

R. JOHN MILLAR received the B.Eng. degree (mechanical) from The University of Auckland, New Zealand, in 1984, and the master’s degree in electrical engineering and the D.Sc. (Tech) degree from the Helsinki University of Technology, Finland, in 2002 and 2006, respectively. He is currently a Researcher and a Lecturer with Aalto University. His main research interests include distribution network planning, underground cable rating, and pedagogical development.