

# Transmission Network Expansion Planning With Stochastic Multivariate Load and Wind Modeling

Mingyang Sun, Ioannis Konstantelos, Goran Strbac

Electrical and Electronic Engineering

Imperial College London

London, UK

{mingyang.sun11, i.konstantelos, g.strbac}@imperial.ac.uk

**Abstract**—The increasing penetration of intermittent energy sources along with the introduction of shiftable load elements renders transmission network expansion planning (TNEP) a challenging task. In particular, the ever-expanding spectrum of possible operating points necessitates the consideration of a very large number of scenarios within a cost-benefit framework, leading to computational issues. On the other hand, failure to adequately capture the behavior of stochastic parameters can lead to inefficient expansion plans. This paper proposes a novel TNEP framework that accommodates multiple sources of operational stochasticity. Inter-spatial dependencies between loads in various locations and intermittent generation units' output are captured by using a multivariate Gaussian copula. This statistical model forms the basis of a Monte Carlo analysis framework for exploring the uncertainty state-space. Benders decomposition is applied to efficiently split the investment and operation problems. The advantages of the proposed model are demonstrated through a case study on the IEEE 118-bus system. By evaluating the confidence interval of the optimality gap, the advantages of the proposed approach over conventional techniques are clearly demonstrated.

**Index Terms**--Multivariate copulas, stochastic optimization, transmission network expansion planning, uncertainty analysis, wind power.

## I. NOMENCLATURE

### Sets and indices

$t \in T$	Scenarios.
$i, j \in I$	Indices of the network buses.
$g \in G$	Indices of all generators.
$\gamma \in \Gamma$	Indices of thermal generators, $\Gamma \subset G$ .
$w \in W$	Indices of wind generators, $W \subset G$ .
$d \in Nd$	Indices of load.
$i, j \in \Lambda$	Existing transmission lines.
$i, j \in \Lambda_{new}$	Candidate transmission lines.

### Input parameters

$n_{ij}$	Candidate lines built between bus $i$ and $j$ , $i, j \in \Lambda_{new}$ .
$C_{ij}$	Cost of building a line from bus $i$ to $j$ (\$/year).

$\pi_g$	Generation cost of generator $g$ (\$/MWh).
$\Upsilon$	Value of lost load (\$/MWh).
$\tau$	Duration of each scenario (hour).
$p^t$	Probability of scenario $t$ .
$B_{ij}$	Susceptance of existing line from bus $i$ to $j$ (p.u.), $i, j \in \Lambda$ .
$B_{ij}^{new}$	Susceptance of candidate line from bus $i$ to $j$ (p.u.), $i, j \in \Lambda_{new}$ .
$f_{ij}^{max}$	Power flow limit of the transmission line from bus $i$ to $j$ (MW).

### Decision variables

$f_{ij}$	Power flow of the transmission line from bus $i$ to $j$ (MW).
$\theta_i$	Phase angle at bus $i$ (rad).
$\mathbf{S}$	Branch-node incidence matrix with elements $s_{ij}$ .
$\mathbf{D}$	Matrix of sampled load (MW).
$\mathbf{P}_W^{max}$	Matrix of sampled wind power (MW).
$\mathbf{P}_\Gamma^{max}$	Matrix of maximum conventional generators capacities (MW).
$\mathbf{U}$	Matrix of load curtailment with elements $U_d$ (MW).
$\mathbf{F}$	Matrix of power flow with elements $f_{ij}$ (MW).
$\mathbf{P}_\Gamma$	Matrix of conventional generator output (MW).
$\mathbf{P}_W$	Matrix of wind generator output (MW).

## II. INTRODUCTION

The increasing integration of large-scale intermittent power plants such as wind farms, the introduction of shiftable load elements, and the growing interconnection that characterizes electricity systems have led to significant uncertainty in planning, operation, and evaluation of transmission networks [1]. Under this new reality, transmission network expansion planning (TNEP) needs to be carried out with the consideration of uncertainties of loads and wind generators. TNEP is a nonlinear, nonconvex, and mixed integer optimization problem (MIP) problem. For practical purposes, it can be approximated via the linearized direct current (DC) power flow [1]. The main purpose of

TNEP is to determine the best time, the most appropriate locations, and the optimal number of new lines that are required to be installed in an electrical network while minimizing the sum of investment and operation costs [3].

In order to solve the TNEP problem, deterministic approaches have been applied in the past [4], [5]. However, the stochastic variables beyond the operator's control (e.g. wind power) introduce substantial power flow fluctuations to the network. Consequently, it is not sufficient to perform the planning with stochastic variables only considering one power system operating profile. In recent years, a series of probabilistic TNEP models (e.g. fuzzy method [7], Monte Carlo simulation [8], and two-point estimation method [9], etc.) have been proposed to take into account the uncertainties stemming from various sources such as forced outage rates of lines [10], liberalized electricity markets [11], and wind farms [12], [13]. Furthermore, the uncertainty of both load and wind have been considered in [14], [15], [16], [8]. Nevertheless, wind and load are assumed to be uncorrelated in [14]. In [15], uncertainties of load and wind are represented by scenarios generated using clustering method. As stated in [8], entirely depending on real observations cannot provide a sufficient number of samples for Monte Carlo simulation.

A parametric model can generate samples that have similar dependence structure but not identical to what has already been encountered, thus interpolating and extrapolating the historical dataset. Another practical advantage is not having to deal with missing data [17]. To this end, total load and available wind power are modeled via a bivariate Gaussian copula based on historical data in [8]. The samples generated from the constructed model are then used to carry out a two-stage stochastic program for transmission planning. However, although this method captures the dependence between total load and total wind output, the individual variables (electricity consumption at each bus and output of an each wind farm) are assumed to be perfectly correlated. This is a substantial simplification undertaken in the interest of relieving the increased computational load that comes with modelling the stochastic variables at the level of individual variables instead of their aggregation. However, capturing the interspatial correlations of load and renewables output at various places can be an important factor for investment planning.

In the past multivariate statistical models have been widely used to model stochastic signals. For example, in [18] power injections from renewable sources of energy are modeled using Vector Autoregressive models to generate multivariate synthetic wind speed time series based on empirical observations. In [19] distribution system loads are represented through a Gaussian mixture model (GMM) whose parameters are estimated by the expectation-maximization (EM) algorithm.

In general, one of the most highly challenging tasks of a multivariate statistical framework, is its ability to model large-scale measurement data by accurately capturing the non-linear dependence structure of the data as well as their non-Gaussian marginal distributions. To this end, a vast number of applications have demonstrated that copulas are capable of accomplishing this task. In [20], a multivariate Gaussian copula is used to fit a model based on historical data

and to generate synthetic wind power output from 15 sites in the Netherlands. Beyond that, a more advanced multivariate modeling framework based on vine copulas has been proposed in [21] to capture spatiotemporal interdependencies of wind power and to determine energy storage requirements. Although the increased computational burden has been resolved through a systematic truncation method, it is still impractical for the modelling of high-dimensional data.

In the present paper, we propose a stochastic TNEP model considering multivariate dependent load and wind modeling based on a multivariate Gaussian copula. Load and wind generation measurements from French electricity system are mapped to the IEEE 118-bus test system, and modeled by the proposed multivariate Gaussian copula. Using the load and wind power samples generated from the constructed model, an appropriate transmission plan for a future year can be obtained by carrying out a Monte Carlo simulation. The proposed TNEP model is solved by using Benders decomposition. The solution quality of the proposed model is evaluated and compared to some conventional methods through constructing a confidence interval on the optimality gap. Results indicate that capturing spatial dependencies at the level of individual variables can have a substantial impact on transmission planning decisions. The proposed framework is shown to outperform existing approaches in this regard.

The outline of the remainder of this paper is as follows: Section II introduces the procedure of stochastic dependence modeling using multivariate copulas. In Section III, we present the mathematical formulation of the proposed TNEP model. Section IV presents the simulation result of the proposed method. A comparison between the results of different methods is also performed here. Section VI summarizes and concludes the paper.

### III. STOCHASTIC MULTIVARIATE DEPENDENCE MODELING USING COPULAS

Dependence modeling of stochastic variables in power systems faces two significant challenges; capturing non-linear dependence structures and non-Gaussian marginal distributions. In Fig. 1. We show histograms and bivariate scatter plots of five stochastic variables taken from the French transmission system; variables L1, L2 and L3 are loads at different locations while W1 and W2 are wind injections from two wind farms.

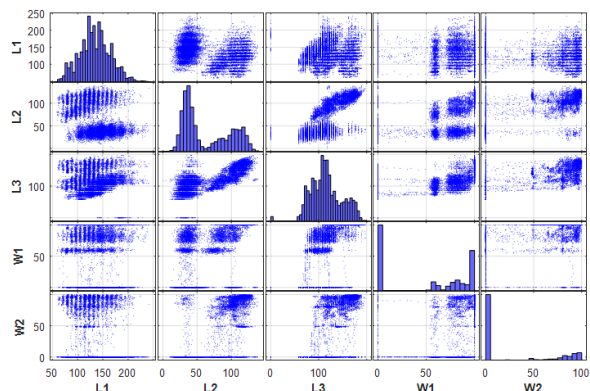


Fig. 1. Histograms of marginals distributions and bivariate scatter plots of active loads (MW) and wind power output (MW) for randomly selected buses in France, measured between Jan 2012 and Feb 2012.

It can be observed that stochastic variables have highly non-normal marginal distributions (diagonal histograms) and non-linear dependencies (scatter diagrams) [17]. The dependence between the different loads, between the outputs of the two wind farms as well as the dependence between individual loads and wind power injections are highly complex and non-standard. As such, it is of great potential benefit to move beyond the assumption of perfect correlation between loads at different locations etc. and investigate the dependence structure at the level of disaggregated variables in more detail. The first step towards this direction is being able to fit a model that can accurately capture all these characteristics, which can then be subsequently used as an engine for the generation of a large number of stochastic scenarios.

Generally speaking, there are three approaches to carrying out statistical modeling: 1) Nonparametric modeling; 2) Parametric modeling; 3) Semiparametric modeling. In general, nonparametric modeling techniques assume that the data do not have a characteristic underlying structure. As a result, such methods do not rely on any specific probability distributions, but the model is constructed directly from the data instead. On the other hand, a parametric model is a distribution that can be described using a finite number of parameters and relies on an analytic description of the joint probability distribution. Finally, semiparametric modeling involves constructing a model that has both parametric and non-parametric components. This is a suitable model for dealing with multi-dimensional data where a combination of different techniques may be essential to capture the important data characteristics. In the present research, the model of choice is semi-parametric due to its primary means of capturing the dependence structure of stochastic variables, as shown in [17]. The procedure of modeling load and wind based on copulas is presented in the following parts.

#### A. Probability Integral Transformation

In this paper, probability density functions and cumulative distribution functions are denoted by  $f$  and  $F$ , respectively, and their copula versions by  $c$  and  $C$ . Consider  $m$  random variables  $X = (X_1, X_2, \dots, X_m) \in \mathbb{R}^m$  with marginal density functions  $f_i(x_i)$  and distribution functions  $F_i(x_i)$ , for  $i = 1, \dots, m$ . For each variable  $X_i$ , its corresponding variable  $U_i$  on  $[0,1]$  can be obtained via the empirical cumulative distribution function (ECDF) transformation, mathematically formulated as:

$$U_i = F_i(X_i) \quad (1)$$

where  $U_i$  is uniformly distributed on  $[0,1]$ . Conversely, given  $U_i$  and the inverse ECDF,  $F_i^{-1}(\cdot)$ , the random variable  $X_i$  can be obtained by:

$$X_i = F_i^{-1}(U_i) \quad (2)$$

Using the equations (1) and (2), samples of stochastic variables with empirical marginal distributions of historical measurements can be generated for Monte Carlo simulation. Examples of this transformation are presented in Fig. 2 based on the historical data shown in Fig. 1. Also, the non-Gaussian margins of the examples show the benefits of using ECDFs instead of fitting parametric distributions.

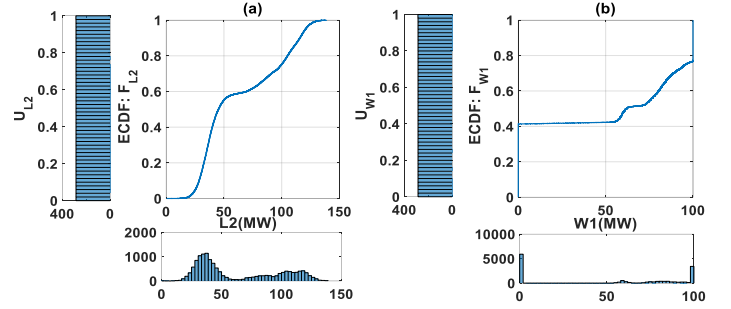


Fig. 2. ECDF transformation of L2 (a) and W1 (b).

#### B. Multivariate Gaussian Copula

According to Sklar's theorem [22], the joint density function of  $m$  random variables  $X = (X_1, X_2, \dots, X_m)$  is given by:

$$f(x_1, \dots, x_m) = \left( \prod_{i=1}^m f_i(x_i) \right) c_{1\dots m}(F_1(x_1), \dots, F_m(x_m)) \quad (3)$$

where the function  $c_{1\dots m}: [0,1]^m \rightarrow \mathbb{R}$  is an  $m$ -dimensional copula with uniform marginals  $\{U_1, U_2, \dots, U_m\} = \{F_1(X_1), F_2(X_2), \dots, F_m(X_m)\}$ . This way, even though a joint density function contains information on both the individual marginal distributions and on the relation between the variables, copulas provide a way for isolating the dependency structure. In general, a copula is an  $m$ -dimensional parametric function defined on the  $[0,1]^m$  space and describing the dependency between  $m$  variables. Copulas have the convenient property that each of their marginal distributions are the uniform distribution on  $[0,1]$ . Moving beyond the above definition, there are many different types of copulas; empirical, parametric, time copulas etc. For the purposes of this research, we will focus on parametric copulas, which are copula functions that can be fully described by a family and corresponding parameters. The idea is to identify the family and parameters that give rise to a copula that best matches the dependency of historical variables.

In terms of the families of copulas, we use a multivariate Gaussian copula to model the stochastic variables. As demonstrated in [8], a Gaussian copula can efficiently capture the first-order notion of dependence. The Gaussian copula is also called normal copula and is usually signified by its defining distribution  $N_m(0, R)$ , where  $m$  is the number of dimensions ( $m = 2$  for bivariate) and  $R$  is the correlation matrix. Gaussian copula is one kind of elliptical copula that only has mean dependence with no upper or lower tail dependencies. The range of pairwise correlations is  $\rho \in [-1,1]$ . Mathematically, when using a multivariate Gaussian copula, equation (3) can be written as:

$$f(x_1, \dots, x_m) = \left( \prod_{i=1}^m f_i(x_i) \right) |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} z^T (R^{-1} - I) z \right\} \quad (4)$$

where  $z_i = \Phi^{-1}(F_i(X_i)) \sim \mathcal{N}_1(0,1)$ ,  $z = (z_1, \dots, z_m)$ , and  $\Phi$  is the standard Gaussian cumulative distribution function.

A visual illustration of the bivariate Gaussian copula is given in Fig. 3.

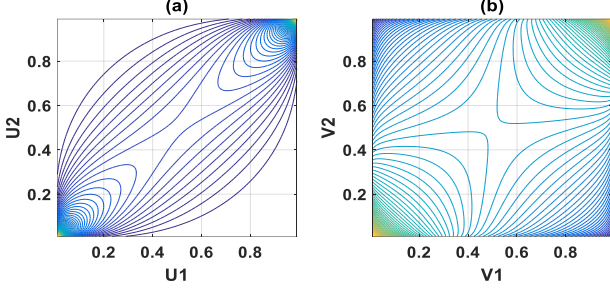


Fig. 3. Examples of bivariate Gaussian copulas with parameter  $\rho = 0.8$  (a) and with parameter  $\rho = 0.2$  (b).

To generate dependent load and wind observations, the procedure for parameterising and sampling from a multivariate Gaussian copula are as follows:

- 1) Using equation (1) to transform marginal distributions of historical variables from actual domain to the uniform domain in  $[0,1]$  through the *ECDFs*.
- 2) Build the multivariate Gaussian copula model based on uniformly distributed variables. Parameters of the constructed model can be estimated by calculating the correlation matrix, as introduced in [20].
- 3) Generate sampled variables in  $[0,1]$  and transform back to actual domain by using equation (2), the inverse *ECDFs* of historical data.

Note that the parameterized model is capable of providing as many samples as required in the Monte Carlo simulation for solving the TNEP problem.

#### IV. PROPOSED TNEP MODEL

##### A. Mathematical Formulation

The TNEP problem aims to minimize the total system cost by determining the optimal transmission expansion decisions along with the optimal operational decisions. Taking into account inter-spatial correlations between load/wind farms across various locations, the proposed TNEP model can be formulated as follows:

$$\min \sum_{i,j \in \Lambda_{new}} C_{ij} n_{ij} + \sum_{t=1}^T \tau \cdot p^t \cdot \Xi(n_{ij}, \mathbf{D}^t, \mathbf{P}_W^{t,max}) \quad (5)$$

s.t.

$$n_{ij} \in \{0,1\}, \quad \forall i,j \in \Lambda_{new} \quad (6)$$

where

$$\Xi(n_{ij}, \mathbf{D}^t, \mathbf{P}_W^{t,max}) = \min \sum_{g \in G} \pi_g \cdot P_g^t + \sum_{d \in Nd} Y \cdot U_d^t \quad (7)$$

s.t.

$$\mathbf{S}(\mathbf{P}_\Gamma^t + \mathbf{P}_W^t + \mathbf{U}^t - \mathbf{D}^t) = \mathbf{F} \quad (8)$$

$$\mathbf{0} \leq \mathbf{P}_W^t \leq \mathbf{P}_W^{t,max} \quad (9)$$

$$\mathbf{0} \leq \mathbf{P}_\Gamma^t \leq \mathbf{P}_\Gamma^{t,max} \quad (10)$$

$$\mathbf{0} \leq \mathbf{U}^t \leq \mathbf{D}^t \quad (11)$$

$$-f_{ij}^{max} \leq f_{ij}^t \leq f_{ij}^{max}, \quad \forall i,j \in \Lambda \quad (12)$$

$$f_{ij}^t = B_{ij}(\theta_i^t - \theta_j^t), \quad \forall i,j \in \Lambda \quad (13)$$

$$\begin{aligned} -n_{ij} f_{ij}^{max} &\leq f_{ij}^t \leq n_{ij} f_{ij}^{max} \\ &: \bar{\mu}_{ij}^t, \underline{\mu}_{ij}^t, \forall i,j \in \Lambda_{new} \end{aligned} \quad (14)$$

$$\begin{aligned} -M(1 - n_{ij}) &\leq f_{ij}^t - B_{ij}^{new}(\theta_i^t - \theta_j^t) \leq M(1 - n_{ij}) \\ &: \bar{v}_{ij}^t, \underline{v}_{ij}^t, \forall i,j \in \Lambda_{new} \end{aligned} \quad (15)$$

In the above model, transmission expansion decisions are determined in the first stage, represented by binary variables,  $n_{ij}$ . In the second stage, the linear objective function includes generation costs as well as the cost of load curtailment, as presented in (7). In (5), the expected operating cost,  $\Xi(n_{ij}, \mathbf{D}^t, \mathbf{P}_W^{t,max})$ , in the proposed model is approximated by the products of the operating hours ( $\tau$ ), the probability ( $p^t$ ) of scenario  $t$ , and the optimal solution of (7), summed across all operating scenarios  $t$ . Note that in this case,  $p^t = 1/T$ , where  $T$  is the number of samples being considered. The branch-node incidence matrix  $\mathbf{S}$  is built based on the configuration of the network. In addition, constraint (8) represents power balance including load curtailment. Maximum capacity limits of thermal and wind generators are given in (9) and (10). Load curtailment  $\mathbf{U}$  is limited by the sampled load  $\mathbf{D}$ , as shown in (11). Finally, constraints (12)-(15) denote thermal capacity limits as well as DC power flow constraints for existing and candidate transmission lines. In particular, dual variables  $\bar{\mu}_{ij}^t, \underline{\mu}_{ij}^t, \bar{v}_{ij}^t, \underline{v}_{ij}^t$  are extracted from constraints (14) and (15). Also,  $M$  is a large value for the disjunctive constraint (15).

##### B. Benders' Decomposition

Benders' decomposition [23] is one of the most widely-used techniques to solve the mix integer problem. In this case, our proposed TNEP model can be decomposed into a master problem and a subproblem via Benders' decomposition. Note that a multi-cut formulation has been adopted, as in [24]. Let  $\kappa \in K$  denote the index iterations, the master problem at the  $\kappa$ th iteration can be presented as:

$$\min \sum_{i,j \in \Lambda_{new}} C_{ij} n_{ij} + z \quad (16)$$

s.t.

$$n_{ij} \in \{0,1\}, \quad \forall i,j \in \Lambda_{new} \quad (17)$$

$$z \geq \sum_{t=1}^T \tau \cdot p^t \cdot \{\hat{y}_\kappa^t + \sum_{i,j \in \Lambda_{new}} [(\bar{\mu}_{ij}^{t,\kappa} + \underline{\mu}_{ij}^{t,\kappa}) f_{ij}^{max} - M(\bar{v}_{ij}^{t,\kappa} + \underline{v}_{ij}^{t,\kappa})] \cdot (n_{ij} - \hat{n}_{ij}^{\kappa})\}, \quad \forall \kappa \in K \quad (18)$$

$$z \geq 0 \quad (19)$$

where  $\hat{n}_{ij}^{\kappa}$  is the solution of the master problem at iteration  $\kappa$ ,  $\hat{y}_\kappa^t = \Xi(\hat{n}_{ij}^{\kappa}, \mathbf{D}^t, \mathbf{P}_W^{t,max})$  and  $\bar{\mu}_{ij}^{t,\kappa}, \underline{\mu}_{ij}^{t,\kappa}, \bar{v}_{ij}^{t,\kappa}, \underline{v}_{ij}^{t,\kappa}$  is the optimal operation cost and dual variables of the subproblem at iteration  $\kappa$  under scenario  $t$ .  $z$  is an approximation of the second stage costs. For each scenario, the objective function of the subproblem is (7) with its constraints from (8) to (15). Note that the binary variables in (14) and (15) are also replaced by  $\hat{n}_{ij}^{\kappa}$ . The procedure of solving the proposed

TNEP problem by using Benders' decomposition can be concluded as follows. Firstly, we initialize  $\varepsilon = 0.02$ , feasible integer solution  $n_{ij} = 1, \forall i, j \in \Lambda_{new}$ , lower bound  $LB = -\infty$ , upper bound  $UB = \infty$ , and iteration  $\kappa = 0$ . After that the solutions of the subproblem with the above initial parameters are used as inputs to solve the master problem to get  $\hat{n}_{ij}^{\kappa}$  and  $z^{\kappa}$ . Accordingly, the lower bound for the  $\kappa$ th iteration can be updated as  $LB = \sum_{i,j \in \Lambda_{new}} C_{ij} \hat{n}_{ij}^{\kappa} + z^{\kappa}$ . The next step is to solve the subproblem using  $\hat{n}_{ij}^{\kappa}$  to get the optimal operation cost  $\hat{y}_{\kappa}^t$  and dual variables  $\bar{\mu}_{ij}^{t,\kappa}, \underline{\mu}_{ij}^{t,\kappa}, \bar{v}_{ij}^{t,\kappa}, \underline{v}_{ij}^{t,\kappa}$ . Also, the upper bound is updated by  $UB = \min\{UB, \sum_{i,j \in \Lambda_{new}} C_{ij} n_{ij} + \sum_{t=1}^T \tau \cdot p^t \cdot \hat{y}_{\kappa}^t\}$ . If the difference between LB and UB is lower than a determined value (e.g.  $\varepsilon \cdot \min\{UB, LB\}$ ), the iteration should be terminated. Otherwise, another benders' cut (18) is built and appended to the master problem for the  $(\kappa + 1)^{th}$  iteration.

## V. SIMULATION STUDY AND RESULTS ANALYSIS

In this section, the performance of the proposed TNEP model along with the multivariate Gaussian copula modeling framework is demonstrated based on a modified version of the IEEE 118-bus system. As detailed in [8], a confidence interval (CI) on the optimality gap of the problem's solution is constructed to determine the solution quality of different planning models.

### A. Test System and Multivariate Load and Wind Modeling

The original IEEE 118-bus system consisting of 54 generators and 186 lines has been modified to also include 10 wind farms of size 100MW as well as 10 more candidate transmission lines with capacity 500 MW. For the historical database of stochastic variables, we map 118 demand buses and 10 wind generators onto the test system from historical load and renewable injection measurements that were provided by RTE, the French Transmission System Operator. The original library includes 14,251 measurements at 15-minute intervals over the course of two months, from January to February 2012, and spans over 7,000 load buses and 200 wind plants [17]. In this paper we focus on a subset; a 128-dimensional dataset consisting of 118 demand buses and 10 wind generators chosen at random. The process of mapping the selected variables is to scale them by a constant that obtained by calculating the ratio between the maximum coincident peak demand of the original data and the sum of active power demand across all buses defined in the system. In this way, the complex dependence structure between the stochastic variables can be maintained. Also note that in this research, value of lost load  $\Upsilon$  is set to \$4,000/MWh. The annualized investment cost for building one candidate transmission line is  $C_{ij} = \$5,000,000/\text{year}$ . Also, we use  $\tau = 8,760$  and  $M = 20,000,000$ .

In order to illustrate the comparative advantage of sampling stochastic scenarios using a multivariate copula model that captures inter-spatial correlations, we generate the load and wind samples by the following methods for comparison;

1) Consider total load and total wind power as two

independent variables (BiInd);

2) Consider total load and total wind power as two dependent variables and model them via a bivariate Gaussian copula (BiGc);

3) Consider load and wind power as multivariate dependent variables and model them via a multivariate Gaussian copula (MGC).

Note that in methods 1 and 2, load and wind at different locations are assumed to be perfectly correlated, respectively, which means static sharing factors between each bus/wind farm are assumed in these cases. An example of comparing historical ( $N_{hist} = 1,4251$ ) and sampled ( $N_s = 2,0000$ ) data generated by the three different methods is shown in Fig. 4 via the scatter plot of a single load variable and a single wind infeed variable chosen at random. In Fig.4 (a), the point-cloud of historical data is shown to be composed of several sub-groups of highly non-linear dependence and complex data features. Moreover, the sampled data points in Fig.4 (b) and Fig. 4(c) are biased towards the bottom with almost zero and non-zero linear correlations, respectively, resulting from the assumption of the static sharing factors. The superiority of the proposed MGC method (Fig. 4(d)) compared to the BiInd method (Fig.5(b)) and the BiGc method (Fig.5(c)), is clearly reflected in the reconstruction quality of the dependence structure. To this end, it is demonstrated that the disaggregated correlations between load/wind can be successfully retained in the sampled data by using the MGC method.

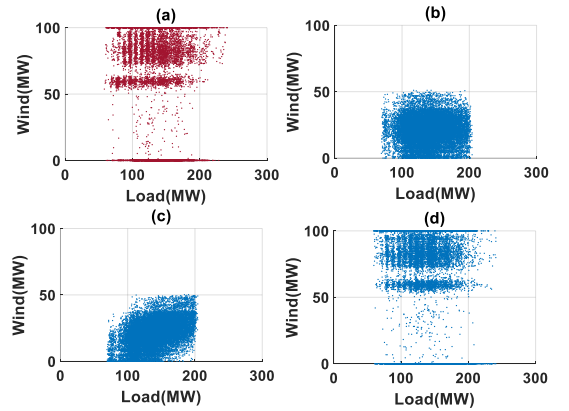


Fig. 4: Scatter plots of historical data (a), scatter plots of generated data using BiInd method (b), BiGc method (c) and the proposed MGC method (d).

### B. Solution Quality Evaluation

We proceed by assessing the impact that the proposed stochastic modelling methodology has on the TNEP problem. For this purpose we make use of a Monte Carlo bounding technique originally proposed in [25] and implemented in [8] to calculate a confidence interval (CI) on the optimality gap of a candidate expansion plan. In this case, the optimality gap is denoted by the difference between a candidate solution  $\hat{\psi} = \sum_{i,j \in \Lambda_{new}} C_{ij} \hat{n}_{ij} + \sum_{t=1}^T \tau \cdot p^t \cdot \Xi(\hat{n}_{ij}, \mathbf{D}^t, \mathbf{P}_W^{t,max})$  and an optimal solution  $\psi^*$ . In this way a 95% CI can be constructed by  $P\{\hat{\psi} - \psi^* \leq \varpi\} \approx 0.95$ , where  $\varpi$  is the CI width; a lower  $\varpi$  value indicates a better solution quality. For the proposed TNEP model with MGC modeling, the

procedure of calculating the CI width  $\varpi$  is briefly summarized as follows.

**Step 1.** Lower bound  $\underline{L}$  construction: Determine the number of samples  $N_s = 500$  as well as the number of replications  $N_L = 50$ . For each replication  $n_L$ , we generate  $N_s$  samples for load  $\mathbf{D}$  and wind  $\mathbf{P}_W^{max}$  of size  $500 \times 118$  and  $500 \times 10$ , respectively, and then solve the TNEP problem to get an optimal solution  $\hat{n}_{ij}^{n_L}$  and the corresponding optimal value  $\hat{\psi}^{n_L}$ . Then the lower bound can be calculated by  $\underline{L} = \frac{1}{N_L} \sum_{n_L=1}^{N_L} \hat{\psi}^{n_L}$ , with its sampling error  $e_L = \frac{std(\hat{\psi}^{n_L})}{\sqrt{N_L}}$ , where  $std(\cdot)$  is the standard deviation, and  $t_{N_L-1, \alpha/2}$  is defined as the  $(1 - \alpha/2)$  level quantile of a Student's t-distribution with  $N_L - 1$  degrees of freedom.

**Step 2.** Candidate solution  $\hat{n}_{ij}$  selection: Use the built MGC model to generate  $N_c = 1000$  sampled load  $\tilde{\mathbf{D}}$  and wind  $\tilde{\mathbf{P}}_W^{max}$ , and solve the TNEP problem with the solution  $\hat{n}_{ij}^{n_L}$  of each replication in Step 1. The candidate solution can be obtained by  $\hat{n}_{ij} \in \operatorname{argmin}_{\hat{n}_{ij}^1, \dots, \hat{n}_{ij}^{N_L}} \{ \sum_{i,j \in \Lambda_{new}} C_{ij} \hat{n}_{ij}^{n_L} + \sum_{t=1}^T \tau \cdot p^t \cdot \Xi(\hat{n}_{ij}^{n_L}, \tilde{\mathbf{D}}^t, \tilde{\mathbf{P}}_W^{max,t}) \}$ .

**Step 3.** Upper bound  $\bar{U}$  construction: Perform the simulation with  $N_U = 20,000$  load and wind samples,  $\bar{\mathbf{D}}$  and  $\bar{\mathbf{P}}_W^{max}$ . The upper bound can be constructed by  $\bar{U} = \frac{1}{N_U} \sum_{n_U=1}^{N_U} \hat{\psi}^{n_U}$  with the sampling error  $e_U = \frac{std(\hat{\psi}^{n_U})}{\sqrt{N_U}}$ . Note that we have  $\hat{\psi}^{n_U} = \sum_{i,j \in \Lambda_{new}} C_{ij} \hat{n}_{ij} + \tau \cdot \Xi(\hat{n}_{ij}, \bar{\mathbf{D}}^t, \bar{\mathbf{P}}_W^{max,t})$ .

**Step 4.** CI width  $\varpi$  calculation. The final CI width can be calculated by  $\varpi = e_U + e_L + [\bar{U} - \underline{L}]^+$ .

Detailed information of the above steps are illustrated in [8]. In addition, the CI widths of the TNEP solutions using BiInd and BiGc modeling algorithms are also calculated following the above procedures with the results shown in Table I. Note that the modeling methods (i.e. BiInd, BiGc, and MGC) and the TNEP optimization problem were implemented in MATLAB and FICO Xpress, respectively, and run on an Intel Xeon E5-2690 PC with 8 cores. For comparison, the time consumption of these three methods for generating 500 samples are given in Table II.

TABLE I. SOLUTION QUALITY EVALUATION FOR DIFFERENT METHODS (500 SAMPLES FOR EACH METHOD)

	Investment cost (\$ million/year)	$\bar{U}$ (\$M/year)	$\underline{L}$ (\$M/year)	$\varpi$
BiInd	30	1,727.41	1,684.72	231.78
BiGc	30	1,625.93	1,649.83	57.83
MGC	15	1,201.87	1,195.76	21.60

TABLE II. COMPUTATION TIMES FOR DIFFERENT METHODS (500 SAMPLES FOR EACH METHOD)

	CPU Times (s)
BiInd	2.24
BiGc	3.12
MGC	5.21

When using the proposed MGC method, only three lines are required to be built, whereas the other two methods both cost double (i.e. \$30 million/year) in the investment stage. Furthermore, regarding the upper bound  $\bar{U}$  and the CI width  $\varpi$ , the superiority of the proposed MGC method can be well demonstrated by exhibiting much lower cost values than the other two methods. Compared to the BiInd method, there are approximately 6% and 75% expected cost reductions in  $\bar{U}$  and  $\varpi$ , respectively, when taking into account the bivariate dependence between total load and total wind.

Most importantly, a substantial improvement on the solution variability (i.e. 30% lower  $\bar{U}$  and 91% lower  $\varpi$ ) is obtained by capturing the inter-spatial dependence between stochastic variables through the MGC. The other methods examined give rise to unrealistic and highly divergent scenarios which lead to inefficient investment decisions. The gradual improvement, from independent to bivariate to multivariate and from aggregated to disaggregated variables that pertain to the specific network locations highlights the importance of capturing stochastic dependence at the level of individual variables when performing transmission expansion planning. The amount of operational variability and importance of capturing spatial correlation is bound to increase in the future due to higher penetration of renewables and the looming electrification of the transport and heating sectors [26].

## VI. CONCLUSIONS

The present paper proposes a two-stage TNEP model integrated with a multivariate Gaussian copula modeling framework. A modified version of IEEE 118-bus system is used as the test network in this research. Based on real historical stochastic variables from the French transmission system, multivariate dependent load and wind samples of 128 dimensions are generated by the proposed MGC algorithm, and compared with the results of BiInd and BiGc methods. A visual comparison between these three methods shows that the inter-spatial dependence can be re-constructed much more accurately when employing a multivariate Gaussian copula model. Furthermore, to evaluate the solution quality, a Monte Carlo bounding technique is implemented to build a 95% CI on the optimality gap of the solution of each method. The results of the comparison suggest that the CI width and the expected total cost of the TNEP problem can be substantially reduced when considering the multivariate dependence of load and wind.

Further research will focus on applying some more advanced high-dimensional modeling techniques in the TNEP problem to deal with more complex dependence structure between stochastic variables. In the future, contingency constraints could also be appended to the TNEP model. Another application of interest is to solve a multistage TNEP problem considering load and wind farm uncertainties.

## REFERENCES

- [1] P. Panciatici, G. Bareux, and L. Wehenkel, "Operating in the fog: Security management under uncertainty," *IEEE Power Energy Mag.*, vol. 10, no. 5, pp. 40–49, Sep. 2012.

- [2] H. Park and R. Baldick, "Transmission Planning Under Uncertainties of Wind and Load: Sequential Approximation Approach," in *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2395-2402, Aug. 2013.
- [3] R. Hemmati, R. A. Hooshmand and A. Khodabakhshian, "Comprehensive review of generation and transmission expansion planning," in *IET Generation, Transmission & Distribution*, vol. 7, no. 9, pp. 955-964, Sept. 2013.
- [4] L. L. Garver, "Transmission network estimation using linear programming," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 7, pp. 1688-1697, Sep./Oct. 1970.
- [5] R. Romero, A. Monticelli, A. Garcia, and S. Haffner, "Test systems and mathematical models for transmission network expansion planning," *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, vol. 149, no. 1, pp. 27-36, Jan. 2002.
- [6] I. De, J. Silva, M. J. Rider, R. Romero, and C. A. F. Murari, "Transmission network expansion planning considering uncertainty in demand," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1565-1573, Nov. 2006.
- [7] A.S. Sousa and E.N. Asada, "Combined heuristic with fuzzy systems to transmission system expansion planning" *Elect. Power Syst. Res.*, vol. 81, no. 1, pp 123-128, Jan. 2011.
- [8] H. Park, R. Baldick and D. P. Morton, "A Stochastic Transmission Planning Model With Dependent Load and Wind Forecasts," in *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3003-3011, Nov. 2015.
- [9] A. Hajebrahimi; A. Abdollahi; M. Rashidinejad, "Probabilistic Multiobjective Transmission Expansion Planning Incorporating Demand Response Resources and Large-Scale Distant Wind Farms," in *IEEE Systems Journal*, vol. PP, no.99, pp.1-1,2015.
- [10] J. Choi, T. Tran, A. A. El-Keib, R. Thomas, H. Oh, and R. Billinton, "A method for transmission system expansion planning considering probabilistic reliability criteria," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1606-1615, Aug. 2005.
- [11] P. Sanchez-Martin, A. Ramos, and J. F. Alonso, "Probabilistic midterm transmission planning in a liberalized market," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2135-2142, Nov. 2005.
- [12] M. Moeini-Aghtaie, A. Abbaspour and M. Fotuhi-Firuzabad, "Incorporating Large-Scale Distant Wind Farms in Probabilistic Transmission Expansion Planning—Part II: Case Studies," in *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1594-1601, Aug. 2012.
- [13] G. A. Orfanos, P. S. Georgilakis and N. D. Hatzigiorgiou, "Transmission Expansion Planning of Systems With Increasing Wind Power Integration," in *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1355-1362, May 2013.
- [14] H. Yu, C. Y. Chung, K. P. Wong and J. H. Zhang, "A Chance Constrained Transmission Network Expansion Planning Method With Consideration of Load and Wind Farm Uncertainties," in *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1568-1576, Aug. 2009.
- [15] H. Park and R. Baldick, "Transmission Planning Under Uncertainties of Wind and Load: Sequential Approximation Approach," in *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2395-2402, Aug. 2013.
- [16] C. Li, Z. Dong, G. Chen, F. Luo and J. Liu, "Flexible transmission expansion planning associated with large-scale wind farms integration considering demand response," in *IET Generation, Transmission & Distribution*, vol. 9, no. 15, pp. 2276-2283, 11 19 2015.
- [17] M. Sun, I. Konstantelos, S. Tindemans, and G. Strbac, "Evaluating Composite Approaches to Modelling High-Dimensional Stochastic Variables in Power Systems," 19<sup>th</sup> PSCC conference, Genoa, Jun. 2016.
- [18] B. Klöckl and G. Papaefthymiou, "Multivariate time series models for studies on stochastic generators in power systems", *Electric Power Systems Research*, vol. 80, no. 3, pp. 265-276, 2010.
- [19] R. Singh, B. C. Pal and R. A. Jabr, "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model," in *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29-37, Feb. 2010.
- [20] G. Papaefthymiou and D. Kurowicka, "Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis", *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 40-49, 2009.
- [21] H. Valizadeh Haghi and S. Lotfifard, "Spatiotemporal Modeling of Wind Generation for Optimal Energy Storage Sizing," in *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 113-121, Jan. 2015.
- [22] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges" *Publ. Inst. Statist. Univ. Paris* 8: 229-231, 1959.
- [23] J. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numer. Math.* vol. 4, no. 1, pp. 238-252, 1962 [Online]. Available: <http://dx.doi.org/10.1007/BF01386316>.
- [24] I. Konstantelos and G. Strbac, "Valuation of Flexible Transmission Investment Options Under Uncertainty," in *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 1047-1055, March 2015.
- [25] W. Mak, D. P. Morton, and K. Wood, "Monte Carlo bounding techniques for determining solution quality in stochastic programs," *Oper. Res. Lett.*, vol. 24, pp. 47-56, 1999.
- [26] G. Strbac, I. Konstantelos, M. Pollitt and R. Green, Delivering future-proof energy infrastructure, report for the National Infrastructure Commission, Feb. 2016.