WIND FARM DIVERSIFICATION AND ITS IMPACT ON POWER SYSTEM RELIABILITY

A Thesis

by

YANNICK DEGEILH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2009

Major Subject: Electrical Engineering

WIND FARM DIVERSIFICATION AND ITS IMPACT ON

POWER SYSTEM RELIABILITY

A Thesis

by

YANNICK DEGEILH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Chanan Singh
Committee Members,	Guy L. Curry
	Garng M. Huang
	Robert D. Nevels
Head of Department,	Costas N. Georghiades

August 2009

Major Subject: Electrical Engineering

ABSTRACT

Wind Farm Diversification and Its Impact on Power System Reliability. (August 2009)

Yannick Degeilh, Graduate in Engineering, Ecole Spéciale des Travaux Publics, Paris

Chair of Advisory Committee: Dr. Chanan Singh

As wind exploitation gains prominence in the power industry, the extensive use of this intermittent source of power may heavily rely on our ability to select the best combination of wind farming sites that yields maximal reliability of power systems at minimal cost.

This research proposes a general method to minimize the wind park global power output variance by optimally distributing a predetermined number of wind turbines over a preselected number of potential wind farming sites for which the wind patterns are statistically known. The objective is to demonstrate the benefits of diversification for the reliability of wind-sustained systems through the search for steadier overall power outputs.

Three years of wind data from the recent NREL/3TIER study in the western US provides the statistics for evaluating each site for their mean power output, variance and correlation with each other so that the best allocations can be determined. Some traditional reliability indices such as the *LOLP* are computed by using sequential Monte Carlo simulations to emulate the behavior of a power system uniquely composed of wind turbines and a load modeled from the 1996 IEEE RTS.

It is shown that configurations featuring minimal global power output variances generally prove the most reliable for moderate load cases, provided the sites are not significantly correlated with the modeled load. Under these conditions, the choice of uncorrelated/negatively correlated sites is favored. The correlations between the optimized global wind power outputs and the modeled load are studied as well.

DEDICATION

To my parents, for they have been contributing to the present work in many subtle ways since I was born…!

ACKNOWLEDGEMENTS

I would like to express profound gratitude and appreciation to my advisor, Dr. Chanan Singh, not only for his invaluable insights, expertise and patience, but also for his availability, authenticity, open-mindedness and willingness to embark on any kind of conversation! I thankfully acknowledge my committee members, Dr. Curry, Dr. Nevels and Dr. Huang for their guidance and support throughout the course of my Masters Degree.

Thanks also to my friends for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the ESTP, my French engineering school, for the opportunity to complete a double degree program here at Texas A&M. Thanks to Aggieland as well, for I will keep precious memories of my time spent in Bryan - College Station.

Finally, thanks to my mother and father for their encouragement and to my girlfriend for her patience and love.

NOMENCLATURE

n	Total number of wind farms considered
m	Total number of wind turbines to be dispatched
n_i	Number of wind turbines belonging to wind farm *i*
P_i	Random variable describing the power output (MW) of farm number *i*
p_i	Random variable describing the power output in MW of a single wind turbine located in farm *i*. Assuming perfect correlation of wind turbines belonging to the same site, the following relation holds: $P_i = n_i \times p_i$
G	Random variable describing the global power output in MW, i.e. the sum of the power outputs supplied by the n wind farms; $G = \sum_{i=1}^{n} P_i$
EXP	Desired global power output mean value
$E(\,)$	Expected value operator
$Var(\,)$	Variance operator
$Std(\,)$	Standard deviation operator
$Cov(\,,\,)$	Covariance operator
$Corr(\,,\,)$	Correlation operator
$\nabla_{n_1 \dots n_n} f = \begin{bmatrix} \frac{\partial f}{\partial n_1} \\ \vdots \\ \frac{\partial f}{\partial n_n} \end{bmatrix}$	The gradient vector of function *f* with respect to each element n_i
λ_f	Failure rate of a wind turbine (per hour)
t_f	Time before failure (hours)
μ_r	Recovery rate of a wind turbine (per hour)
t_r	Time before recovery (hours)
MDT	Average mean down time of a WT (hours)

MUP	Average mean up time of a WT (hours)
f	Frequency of failure/recovery of a WT (per hour)
$\widetilde{LOLP}$	Loss of load probability estimate over an iteration of MCS
td	Total number of system down hours over an iteration of MCS
$\widehat{LOLP}$	Loss of load probability estimate over all the MCS iterations carried out up to that time
N_T	Number of system down times (all iterations taken into account)
td_i	System down times (hours) (all iterations taken into account)
T	Actual simulation time (number of hours simulated up to that time)
$\overline{td_\iota}$	Actual estimate of the system mean down time (hours) (calculated over all iterations up to that time)

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

I. INTRODUCTION

The basic goal of wind power exploitation lies in the economical, sustainable and environmentally-friendly replacement of conventional energy sources such as fossil fuels or nuclear. However, such ambitions assume large scale deployments of wind turbines in order to significantly impact national economies. Many developed countries are already engaged in policies privileging high wind power penetration, setting objectives to be achieved in the decades to come.

Due to the intermittent nature of the wind itself and the lack of efficient way of storing energy, the main challenge lies in ensuring power system reliability standards. This research investigates a methodology to achieve this by exploiting wind park diversification as a means to reduce wind power unpredictability and as shown later, loss of load indices for moderate load cases. Besides, the methods presented in this study should also permit an efficient and reasoned partition of wind turbines across the land, which is of interest for large scale wind power integration studies.

To the author's knowledge, no paper published in the literature has previously investigated how different wind parks could, through the correlations of their power outputs, complement each other so as to ultimately smooth the global power output. Reported studies are generally concerned about the selection of a given potential wind farming site based on its wind patterns [1], but not about the beneficial interactions that various power outputs from various wind parks may yield. Estimating the wind capacity of a large system area, in the following case a country, has already been investigated in [2]: the wind speed characteristic of each Belgian geographical area was taken into account so as to reflect the power production capacity of each region. From this information, a global wind power distribution was then convoluted considering two cases: the first case assumed that the wind parks belonging to the same region were totally correlated, while the second case assumed their independence, which gave

---

This thesis follows the style of *IEEE Transactions on Power Systems*.

slightly (but not significantly) better results in terms of capacity credit according to the reported study.

The mix of various wind speed patterns originating from various regions is the basic idea of the present research: to examine if it is possible to take advantage of the diversity of wind speed profiles among various sites in order to balance the global wind power output. The main purpose of this study is to find the best distribution of a given number of wind turbines over preselected sets of potential wind farming sites so that the global wind power output gets smoother and thus more predictable and reliable in the long run. The sole enhancement of wind power output predictability is in itself desirable, as it would permit the accurate design of thermal conventional units dedicated only to the compensation of wind power erratic behavior.

The turbines used in the studies are 3 MW Vestas V90. For illustrating the methodology, two sets of 7 sites have been investigated. In one set, wind farm wind distributions/power outputs are relatively positively correlated to each other whereas in the other set, they appear almost completely uncorrelated. It is assumed that power outputs supplied by wind turbines belonging to the same site are perfectly correlated. This is not strictly true on a very short term basis according to [3, 4, 5, 6] as land features, hub heights, wind turbulence, wake effects and spacing between wind turbines (among other reasons) affect the particular power output of any individual unit [5, 6]. However, it is a reasonable assumption here since this study addresses planning issues and as such is based on 10 minute step data, as seen later in Part II.4, to consider overall correlation effects on an hourly basis. In addition, data is generally scarce or unknown in the planning stages, which underlines the practical aspect of the hypothesis.

The final goal is to outline the benefits of wind farm diversification on the results provided by both sets of analysis and provide more insights for the selection and combination of suitable wind farming sites. Although generally wind farms would be embedded in a power grid containing conventional units, here we assume wind power to be the only source of power in order to more clearly observe the advantages and disadvantages of diversification. The primary purpose of this study being early stage planning, no power transmission considerations have been made.

The presentation in the thesis proceeds as follows. In the second section, the basic ideas promoting a diversification of the wind farms are discussed and formalized mathematically so as to provide the third section with the basic elements necessary for a clear and rigorous definition of the wind farm optimal distribution problem. This part also discusses the supporting data from the NREL/3TIER project [3] and the main hypothesis consisting in assuming that wind power outputs from turbines of the same farm are perfectly correlated. In the third section, the optimization process is described, analyzed and implemented. The fourth section gathers numerous application studies. The first study introduces figures providing visualization of the various possible configurations including the optimal one(s). In the second application study, a sequential Monte Carlo simulation emulates the failures and recoveries of every single wind turbine to finalize their behavioral model and permits an accurate simulation of 3 years of wind turbine power output history. The reliabilities of the configurations are then analyzed in the third application study and compared using the hourly load model proposed in the 1996 IEEE Reliability Test System [7], which also permits the search for the best configurations in terms of loss of load probability *LOLP*. The last application study compares wind power outputs behaviors with that of the modeled load. Their correlation is notably studied.

II. WIND FARM DIVERSIFICATION THEORY

II. 1. The Concept of Diversification: Dice Game Illustration

Diversifying wind farming shares many similarities with diversifying one's investment portfolio through methods developed by the Mean-Variance Portfolio Theory [8]; the basic idea consists in avoiding a risky dependence on only one source profit/power because of its unpredictability. Drawing upon many independent or negatively correlated sources helps ensure a steady, more predictable outcome. Such a phenomenon directly results from a good diversification and can be illustrated by dice games.

Consider two simple dice games of same expected outcome: Game A consists in rolling an unbiased die of 6 faces, twice the result of which gives the amount of tokens a player earns. Similarly, Game B has the player roll two unbiased dice of 6 faces; the sum of their rolls gives the player's gain. Note that the dice rolls are supposed independent from each other. Both games obviously have the same expected outcomes, i.e. $E(A) = E(B) = 7$ (tokens), where A is the random variable representing Game A outcome, and B the random variable for Game B outcome.

Yet, do these games "feel" the same? Are their results equivalent? The answer is no, as the probability distribution of their outcome present some differences. Let us have a look at the variance (standard deviations) of their outcomes to get a better idea of the gain dispersion around their mean:

- In Game A, the outcome A has the following variance and standard deviation:

$$Var(A) = \frac{1}{6}\sum_{x=1}^{6}\left(2x - E(A)\right)^2 \cong 11.667 \qquad (1)$$

Where x is a possible outcome.

$$Std(A) = \sqrt{Var(A)} \cong 3.416 \tag{2}$$

- In Game B, the outcome *B* has the following variance and standard deviation:

$$Var(B) = \frac{1}{36}\sum_{x=1}^{6}\sum_{y=1}^{6}\left(x + y - E(B)\right)^2 \cong 5.833 \tag{3}$$

Where x is a possible outcome of the first die and y a possible outcome of the second die. Note that the dice outcomes are supposed independent from each other, though their probabilistic distributions are the same.

$$Std(B) = \sqrt{Var(B)} \cong 2.415 \tag{4}$$

The standard deviations of both games differ, although they feature the same expected outcomes. Game B actually proves more predictable, or less “risky”, as shown by its lower standard deviation. One can expect Game B outcomes to be, in a large number of occurrences, closer to their mean than Game A outcomes. Game B incorporates in fact more diversification than Game A as it calls for two *independent* dice rolls (which may yield different results each time) whereas game A can be seen as the throw of two dice that will always yield the *same* result for a given occurrence. As such, these two imaginary dice could be said perfectly positively correlated. If we now introduce random variable *X*, designating the outcome of Game X consisting in rolling a single unbiased die of 6 faces, the result of which gives the number of tokens earned, one can rewrite results for Games A and B the following way:

$$E(A) = E(B) = 2E(X) = 7 \tag{5}$$

$$Var(A) = Var(2X) = 4Var(X) \cong 4 \times 2.917 \cong 11.667 \tag{6}$$

$$Var(B) = Var(X) + Var(X) = 2Var(X) \cong 5.833 \tag{7}$$

$$Var(A) = 2Var(B) \text{ and } Std(A) = \sqrt{2}\, Std(B) \tag{8}$$

If we now generalize the previous equations to order *n*, i.e. a gain multiplier *n* for Game A and *n* dice rolls for Game B, one finally obtains:

$$Var(A) = n\, Var(B) \text{ and } Std(A) = \sqrt{n}\, Std(B) \tag{9}$$

The above relations are true provided each dice roll is independent from the others (Game B). What if they were not? In the case of Game B, which calls for the sum of *n* random variables, the variance would be given as follows:

$$Var(B) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}^{n} Cov(X_i, X_j) \tag{10}$$

Where $Cov(X_i, X_j)$ is the covariance of random variables X_i and X_j.

One can observe that if the sum of the covariance terms is negative, the variance of *B* can decrease even further. This remark prompts the wind farm diversification theory: if one diversifies farms, i.e. distributes the wind turbines in various geographical areas featuring different wind patterns, one can expect a lowering of the variance of the global power output.

II.2. Diversification Applied to Wind Farm Planning

Diversification is therefore of interest in reducing the variance of a sum of random variables, provided they are uncorrelated/negatively correlated. What if we now replace the number of rolls (or dice) by wind turbines? Power outputs of these latter heavily depend on the localization of the wind turbine itself as wind patterns differ from one region to another. One can make the most of this observation by distributing wind turbines in such a way that their power outputs eventually prove uncorrelated/negatively correlated with respect to each other.

Hence for a given hour of the day, wind-deficient farms could be compensated by wind-benefiting farms, thus always securing a minimum global power output at anytime. Intuitively, two positively correlated wind farms - i.e. wind farms showing very similar wind patterns over time- will behave much the same way, meaning that a drop in the wind speed will cause a drop in global power production and vice versa. On the contrary, two negatively correlated wind farms, showing almost complete opposite wind patterns, will compensate each other all the time, meaning that the power output will remain almost constant around its mean value (assuming wind farms of comparable power outputs). Intuitively, the second situation is much more satisfying in terms of reliability as the power output remains steady all the time. The lower the correlation coefficient of two statistical series (here wind speed or power output of wind farms), the better it is for steadiness and reliability.

Correlation is not all, however. The power output of a given wind farm also tends to vary around its mean value according to the wind pattern. The variance of a statistical series is a good indicator of how much the statistical values are distributed around their mean. The greater the variance, the more dispersed the values around the mean. The lower the variance, the more concentrated the values around the mean. Once again, it seems preferable to favor wind farms showing smaller power output variance, as it means a steadier power output.

The concepts of individual wind farm power output variance and wind farm correlation both appear in the quantification of global power output variance, i.e. “the variance of the sum of the various wind farm power outputs”. The formula of the global power output variance $Var(G)$, considering n wind farms, is then given by the following:

$$Var(G) = Var\left(\sum_{i=1}^{n} P_i\right) \tag{11}$$

Now, knowing that:

$$Var\left(\sum_{i=1}^{n} P_i\right) = \sum_{i=1}^{n} Var(P_i) + 2\sum_{i<j} Cov\left(P_i, P_j\right) \tag{12}$$

We eventually have:

$$Var(G) = \sum_{i=1}^{n} Var(P_i) + 2\sum_{i<j} Cov\left(P_i, P_j\right) \tag{13}$$

The covariance between the power outputs of two wind farms is related to their correlation coefficients $Corr\left(P_i, P_j\right)$ as follows:

$$Corr\left(P_i, P_j\right) = \frac{Cov\left(P_i, P_j\right)}{\sqrt{Var(P_i)\, Var\left(P_j\right)}} \tag{14}$$

It is obvious from (13) and (14) that the smaller the correlations between sites (between -1 and 1), the smaller the global variance, and intuitively, the steadier the global power output. Same remarks can be made concerning the individual variances. An appropriate way of distributing wind turbines over a preselected number of sites would then consist in dispatching them so that the global power output variance is eventually as small as possible. As a matter of fact, if in (13) we detail the expressions of the variances $Var(P_i)$ and covariances $Cov(P_i, P_j)$ by introducing the number n_i of wind turbines belonging to farm *i*, we obtain:

$$Var(G) = \sum_{i=1}^{n} n_i^2 Var(p_i) + 2\sum_{i<j} n_i n_j Cov\left(p_i, p_j\right) \tag{15}$$

Equation (15) is only valid assuming a *complete correlation of wind turbines within a wind farm/site*. Otherwise, one cannot write the following equation relating a wind farm power output to the power outputs of its wind turbines: $i \epsilon [1 \ldots n], P_i = n_i \times p_i$ The relevance of such a hypothesis is in fact a sensitive question that is discussed in Part II.4 right after the presentation of the NREL/3TIER supporting data.

It can be seen from (15) that the wind turbine distribution has a considerable influence on the global power output variance, not making its minimization trivial. If we are to map what one can expect from the best allocation of *m* wind turbines (the total number of wind turbines standing for a constraint) over *n* sites, we have to minimize the global variance for as many mean global power output values as desired (introduction of a second constraint). This mapping aims at illustrating the best wind turbine configurations in terms of small variance - and supposedly better reliability and predictability - with respect to their average power output levels. To proceed, we need data from multiple wind farming sites. The knowledge of the statistical power outputs p_i from single wind turbines located in farms $i \in [1 \ldots n]$ is required so as to assess terms such as $Var(P_i)$, $Cov(P_i, P_j)$ and $E(p_i)$ (expected value of p_i being necessary for the formulation of the global mean value constraint).

II.3. The NREL/3TIER Data

The recent Western Wind and Solar Integration Study conducted by the National Renewable Energy Laboratory (NREL) in collaboration with the group 3TIER [3] can provide the aforementioned information for more than 30,000 US western sites over a period of 3 years. Figure 1 depicts the interface through which the data can be accessed.

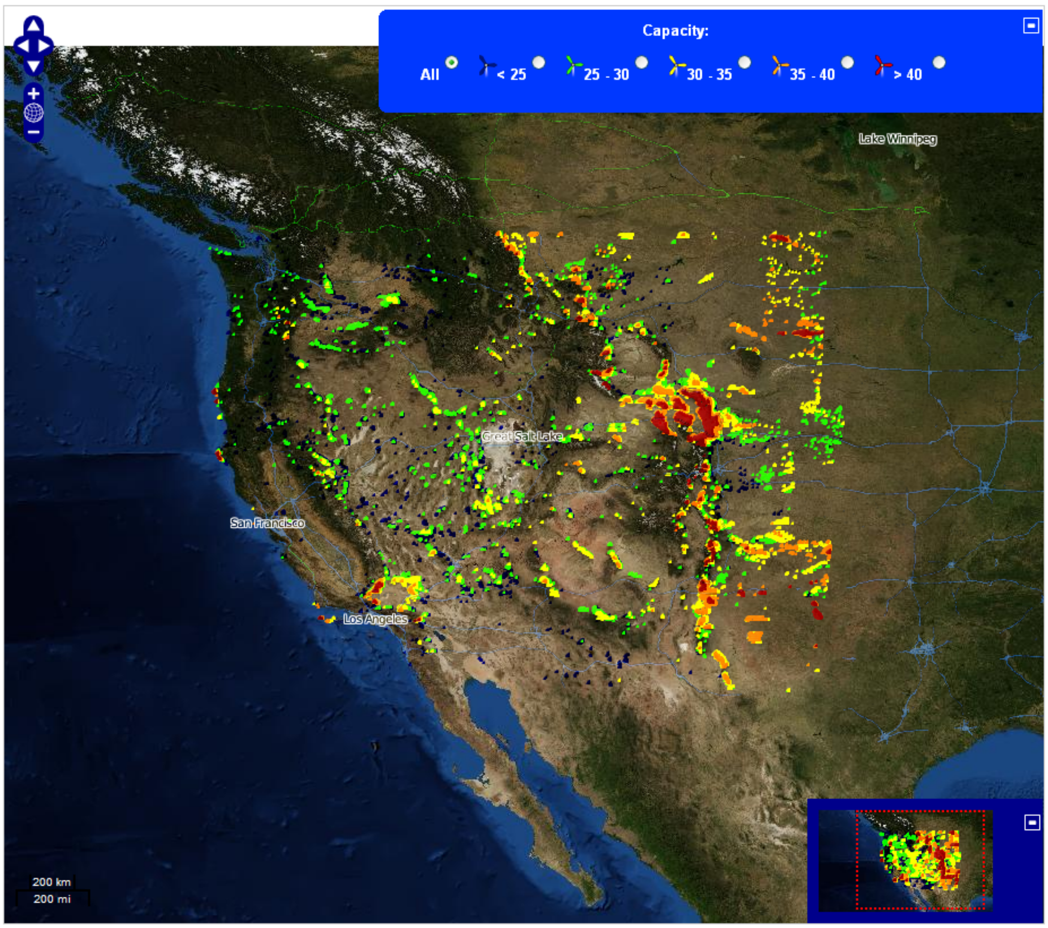

Fig. 1. Interface granting access to the 30,000 sites for which NREL/3TIER offer wind power output data [9]

Wind speed and power output data was actually mesomodeled, meaning it is based on the output of a numerical weather prediction model relying on physical conservation equations [3]. The weather was realistically recreated from year 2004 to 2006 so that wind speeds and 3MW Vestas V90 wind turbine power outputs (each site comprising of 10 turbines) were eventually estimated for every 10 minutes span of the covered period. These assessments were then adjusted by MOS-correction [3] in order to be as close as possible to the actual measurements (few being actually available).

Absolute power output measurements are not critical here, for in this thesis, we primarily strive to show the benefits of wind farming using geographical diversification. Realistic power output time series are largely sufficient for the job as we only need some coherent estimations of the wind behavior over many areas to assess variances and correlations.

With availability of 3 years of power output data from 30,000+ wind farming sites, reasonable *estimates* of the expected value and variance of a site power output can be statistically computed, as well as the correlation coefficient (and covariance) between any pair of wind farms. This can be easily achieved thanks to software such as MATLAB. With such information available, minimizing the global variance by picking the best wind turbine configuration over a preselection of wind farms becomes possible. The thesis will then focus on the general method to optimize the distribution of *m* wind turbines over *n* preselected sites and discuss the optimal distributions of 40 wind turbines (3 MW Vestas V90) over 2 different sets of 7 sites, the sites being strongly positively correlated in the first case, almost uncorrelated in the second.

It should be noted that in order to simplify the notation, further references to the *sample estimates* of mean values, variances and so on will be simply named or noted by their probabilistic counter parts i.e. “expected value”, “variance” etc., of a random variable. This makes sense in regard of the law of large numbers and the very large and representative samples we are dealing with. As a result, operator notation such as $E(\,)$, $Var(\,)$, $Cov(\,,)$ or $Corr(\,,)$ are used throughout the study, although in practice they represent sample estimates.

II.4. Discussion Over "the Perfect Correlation of Wind Turbine Power Outputs within the Same Wind Farm" Hypothesis

The NREL/3TIER wind turbine power output data has been assessed for groups of 10 turbines in order to better account for smoothing effects that affect the global power output of wind turbines pertaining to a same wind farm. Individual wind turbines actually show independent power output at the minute scale because of numerous factors (detailed later in this part) in addition to the fact that the conversion from wind speed to power output is not entirely deterministic, as shown in Figure 2.

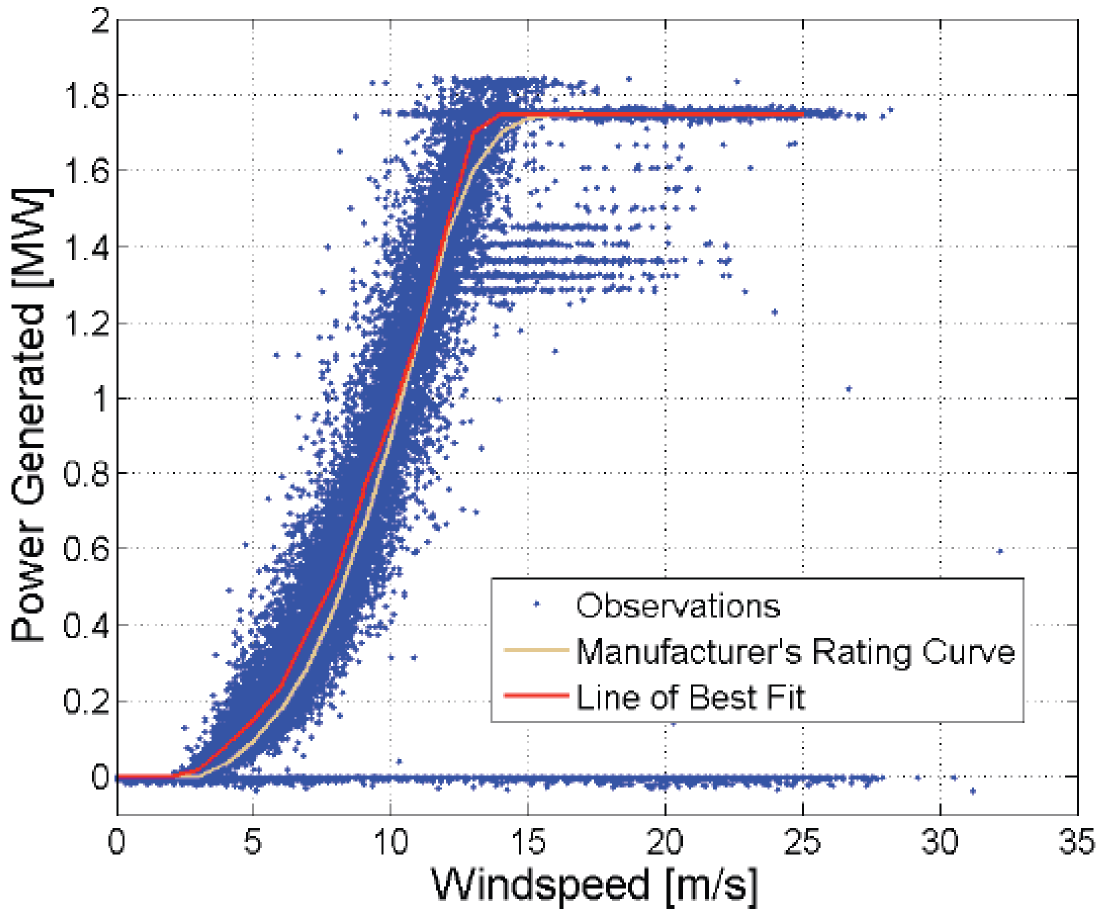

Fig. 2. The conversion of wind to power. (wind turbine of 1.75 MW max capacity) [10]

These facts have many consequences upon the validity of our assumption stating that the wind turbine power outputs within a given wind farm are perfectly positively correlated with each other. Few in the literature have actually numerically assessed the correlations among different wind turbines within a wind farm but all agree that on short times (second-minute scale essentially [4] and [6], though none of the references clearly define shorter and longer scales), the wind turbine power outputs are not correlated, thus leading to the smoothing of the wind park global power output on short time spans [1,3,4,5,6].

The whole problem primarily lies in the fact that it is necessary, for practical and financial reasons, to evaluate a site based on some very scarce data. Land features, hub heights, wind turbulence, wake effects, spacing between wind turbines are some parameters (among others) that need to be considered in the assessment of an aggregated power output (i.e. the power output of a wind park) [1,5,6]. This is why at first sight (statistical sight) wind turbine power outputs are statistically independent for very shorts time spans (minute scale time). This also means that at the minute scale, their aggregated power output will tend to be smoother, thus making the output actually closer to its mean value. So over 10 minutes (which is the step time of our data), these statistically independent variations actually tend to make a 10 minute estimation more representative of the minute scale trend and thus more reliable. All in all, it does not really hurt the practicality of our hypothesis of perfect correlation. This proves probably more problematic when working on very short terms operations. This thesis however solely focuses on wind power planning issues and thus on the long term.

The MATLAB program used to process the data assesses one site's characteristics based on the power output expected value, variance and covariances of a *single* wind turbine of reference. To do so, the program simply takes the 10th of the NREL/3TIER power output series of 10 wind turbines for a given site. We finally get an estimation of what the time series of a single wind turbine would be. Given the non perfect correlation of wind turbines within a site, this evaluation is not completely correct. However, it acts like a mean value of what can be expected from a wind turbine of a given site in terms of wind power. Nevertheless, the power output of a small number of installed wind turbines may appear excessively smooth, therefore reducing the overall variance and making our evaluation optimistic. For 10 installed wind turbines, this would be a perfect match with the NREL/3TIER data. For more, the aggregated power output would in reality keep getting smoother and smoother (as a result of the increase in the number of wind turbines) which could actually benefit the global reliability, thus making our evaluation rather pessimistic. For a number of wind turbine so large that space may be lacking, meaning by that that the wind patterns may significantly change over the same wind park because of its size, then a new wind park with its own data should be defined and the process executed again.

As stated previously, the impact of these variations should be minimal on the long term analysis presented in this thesis. The optimization process as presented later clearly cannot be executed if we do not assume perfect correlation of the wind turbines within a site which leads to the relation: $i \epsilon [1 \dots n], P_i = n_i \times p_i$. The hypothesis of perfect correlation seems however strong for the kind of long term studies investigated in the thesis.

The impact of these minute scale variations could be verified as follows, provided the necessary data was actually available: after optimization, we could enhance the Monte Carlo simulations (introduced in Section IV.2) by not only emulating the mechanical/electrical failures of the wind turbines, but also by simulating their minute scale statistical behavior (meaning that we would not work any more on hourly or 10-minute scale data, but on minute scale data whose mean value would have to be interpolated from the greater time scales). To do so, we could then define the minute scale statistical behavior by Gaussian probabilistic distributions having for mean values the numbers used until now in the more simplistic simulations (10 minute scale data) and for variance a quantity that needs to be defined from what operators actually observe on site (the literature does not tell much about it). With such a Gaussian distribution, we can then pick random values and match simulated power output values via the aforementioned probabilistic distributions. A minute scale Monte Carlo simulation can then be performed, leading to the creation of one minute scale power output series for every single wind turbines that can be thereafter aggregated to obtain the minute scale wind park power output series. The problem is, when it comes to compare such a minute scale power series to an hourly scale load series, it cannot be done seriously without knowing the minute variations of the load itself.

III. FORMALIZATION AND TREATMENT OF THE OPTIMIZATION PROBLEM

III. 1. Problem Formalization

Now that the main ideas have been outlined, let us define the optimization problem mathematically. Essentially the problem is finding the optimal distribution of *m* wind turbines of the same kind over *n* sites. We consequently have *n* unknowns describing the number of wind turbines actually assigned to a given site. The global power output variance $Var(G)$ is the objective function to be minimized with respect to the wind turbine dispatch. The constraints consist of the total number of wind turbines *m* to be installed and the global power output mean value *EXP* (*EXP* standing for Expected Power) the configuration is expected to supply. As a matter of fact, this optimization needs to be carried out many times for different mean values *EXP* so as to make clear what can be expected from the optimal distribution of *m* wind turbines. This will be illustrated later in Section IV in the figures on pp.24 and 25 respectively.

If we now formalize the problem mathematically, we have:

$$Min\ \ Var(G) = \sum_{i=1}^{n} n_i^2\, Var(p_i) + 2\sum_{i<j} n_i n_j Cov\big(p_i, p_j\big) \tag{16}$$

subject to

$$\sum_{i=1}^{n} n_i = m \tag{17}$$

$$\sum_{i=1}^{n} n_i E(p_i) = EXP \tag{18}$$

$$i \epsilon [1 \dots n], \;\; n_i \geq 0 \tag{19}$$

Here $\sum_{i=1}^{n} n_i E(p_i) = E(G)$ by linearity of the "expected value" operator $E(\,)$.

The problem clearly calls for the optimization of a quadratic objective function (16) subject to a set of two linear equality constraints (17) and (18) and *n* linear inequalities (19). In the next part, we shall determine whether we are dealing with a maximization or a minimization.

III. 2. Nature of the Optimization

Our problem qualifies as a convex programming problem. This particular setting comes with two notable properties: first, any stationary point within the feasible region is deemed to be a local minimizer; second, a local minimizer of the problem is also a *global minimum* [11]. In other words, the finding of a stationary point within the feasible region supposes the finding of the problem global minimum. Let us now show that our problem abides by the criteria defining a convex programming problem.

An optimization problem is said convex if the objective function is convex over a convex feasible region. In our problem, the feasible region is defined by a set of *linear* constraints (no matter what their nature, i.e. equality or inequality) that inherently makes it convex. This is straightforward since the intersection of convex subspaces (defined by the linear constraints) results in a convex subspace (the feasible region). Mathematically, the convexity of a subspace can be written as follows:

A set (of constraints in our case) *S* is convex if, for any elements *x* and *y* of *S*:

$$\alpha x + (1 - \alpha) y \in S \; for \; all \; 0 \leq \alpha \leq 1 \tag{20}$$

One can then easily verify that subspaces defined by linear constraints abide by that definition, and so does their intersection, the feasible space.

The convexity of the objective function can be established by looking at its Hessian matrix, which, in the multi-dimensional case, contains the second derivative information needed to conclude. The Hessian of an n-dimensional function f_n is defined as follows:

$$H(f_n) = \begin{bmatrix} \frac{\partial^2 f_{n1}}{\partial x_1^2} & \frac{\partial^2 f_{n1}}{\partial x_2^2} & \cdots & \cdots & \frac{\partial^2 f_{n1}}{\partial x_n^2} \\ \vdots & \frac{\partial^2 f_{n2}}{\partial x_2^2} & . & . & \vdots \\ \vdots & . & \ddots & . & \vdots \\ \vdots & . & . & \frac{\partial^2 f_{nn-1}}{\partial x_{n-1}^2} & \vdots \\ \frac{\partial^2 f_{nn}}{\partial x_1^2} & \cdots & \cdots & \frac{\partial^2 f_{nn}}{\partial x_{n-1}^2} & \frac{\partial^2 f_{nn}}{\partial x_n^2} \end{bmatrix} \quad (21)$$

In the case of a one dimensional function f_1, the Hessian matrix reduces to the second derivative f_1''. Such a function f_1 can easily be determined to be convex if it has two continuous derivatives and verifies:

$$f_1''(x) \geq 0 \; for \; all \; x \in S \quad (22)$$

Similarly, a multi-dimensional function f_n is convex on set S if its Hessian matrix is positive semi-definite for all $x \in S$. In the case of objective function $Var(G)$, the Hessian calculation yields:

$$H\big(Var(G)\big) = 2 \; CovM \quad (23)$$

With *CovM* the covariance matrix of the statistical single wind turbine power outputs p_i (i designating the farm number):

$$CovM = \begin{bmatrix} Var(p_1) & Cov(p_1,p_2) & \cdots & \cdots & Cov(p_1,p_n) \\ Cov(p_2,p_1) & Var(p_2) & \ddots & . & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & . & \ddots & Var(p_{n-1}) & Cov(p_{n-1},p_n) \\ Cov(p_n,p_1) & \cdots & \cdots & Cov(p_n,p_{n-1}) & Var(p_n) \end{bmatrix} \quad (24)$$

Objective function $Var(G)$ is, therefore, convex because of the positive semi-definiteness of any covariance matrix. This concludes the proof that our optimization problem is convex. Let us now focus on a method searching efficiently for the feasible stationary point that will prove to be the global minimizer of our problem.

III. 3. Resolution of the Optimization Problem

The optimization problem can be solved efficiently by one of the many quadratic programming algorithms available today. For instance, MATLAB features a built-in function called *quadprog* that resorts to an active set method to solve such problems. The active set method is a procedure that searches for a stationary point along a set of supposedly active constraints (i.e. equality plus “activated” inequality constraints). The active set constraint is appropriately redefined until obtaining of an optimum. This method works consistently well for both sets of potential wind farming sites tested later in the thesis.

Another resolution method has also been implemented to solve this kind of problem. It experimentally yields the same results as the MATLAB active set method while remaining perhaps easier to understand. Note however, that the program has only been tested on the experimental sets of wind farming sites discussed later in the thesis and that it has not been shown to converge for any general case. The procedure extensively uses the *Lagrange multipliers theory* and its first order necessary condition to find stationary points on equality constrained sets updated after each iteration.

Practically, the inequality constraints are first assumed to be inactive. If after resolution, the optimal solution is not feasible (i.e. some n_i - number of wind turbines at site *i* - are found negative), the appropriate inequality constraints are activated by setting the negative n_i to 0. This comes down to eliminate the poor sites by literally removing them from the process, thus reducing the dimension of the problem. The procedure must then be reiterated until a feasible optimum that verifies the *KKT* first order necessary conditions of the initial (convex) problem is eventually found, i.e. in our case: $-\nabla_{n_1 \dots n_n} f = \lambda \nabla_{n_1 \dots n_n} g \;\; with \;\; \forall i \in [\![1;n]\!],\ \lambda_i \geq 0 \;\; and \;\; g_i \leq 0$ (f is taken as the objective function, g_i as the constraints, λ as the vector of corresponding Lagrangian multipliers). Note that the maximum iteration number of such an algorithm is *n* (*n* being the number of selected sites) if one site is to be removed at each iteration (if not then the algorithm terminates with a solution). Moreover, it may be preferable to solely remove the worst site at each iteration (instead of removing all negative n_i sites) so as to make sure the program strives for a diversification that includes a maximum number of sites. This last approach has finally been retained for the tests.

Next follows a complete description of how the algorithm finds a stationary point at the first iteration. The method is exactly the same for later iterations, except that some variables n_i will have been set to 0, ultimately reducing the problem dimension.

If $L(n_1 \dots n_n, \lambda, \mu)$designates the Lagrangian with λ and μ the multipliers, we then have:

$$L(n_1 \dots n_n, \lambda, \mu) = Var(G) - \lambda \left[m - \sum_{i=1}^{n} n_i \right] - \mu \left[EXP - \sum_{i=1}^{n} n_i \, E(p_i) \right] \qquad (25)$$

With the following notations for the objective function and the constraints: $Var(G) = Obj$, $[m - \sum_{i=1}^{n} n_i] = C_1$ and $[EXP - \sum_{i=1}^{n} n_i \, E(p_i)] = C_2$, Equation (25) can be rewritten:

$$L(n_1 \dots n_n, \lambda, \mu) = Obj(n_1 \dots n_n) - \lambda\, C_1(n_1 \dots n_n) - \mu\, C_2(n_1 \dots n_n) \qquad (26)$$

And the optimization problem comes down to solving the following system (first order necessary condition):

$$\begin{cases} \nabla_{n_1 \dots n_n} L = 0 \\ \nabla_\lambda L = 0 \\ \nabla_\mu \mathrm{L} = 0 \end{cases} \qquad (27)$$

$$\Leftrightarrow \begin{cases} \nabla_{n_1 \dots n_n} Obj = \lambda\, \nabla_{n_1 \dots n_n} C_1 + \mu\, \nabla_{n_1 \dots n_n} C_2 \\ \sum_{i=1}^{n} n_i = m \\ \sum_{i=1}^{n} n_i\, E(p_i) = EXP \end{cases} \qquad (28)$$

$$\Leftrightarrow \begin{cases} 2 n_i Var(p_i) + 2 \sum_{j \neq i} n_j\, Cov(p_i, p_j) + \lambda + \mu\, E(p_i) = 0 \\ \sum_{i=1}^{n} n_i = m \\ \sum_{i=1}^{n} n_i\, E(p_i) = EXP \end{cases} \qquad (29)$$

System (29) actually makes up a system of (n+2) linear equations that can be reformulated into the following matrix equation (30):

$$\begin{bmatrix} . & \cdots & . & 1 & E(p_1) \\ \vdots & 2\, CovM & \vdots & \vdots & \vdots \\ . & \cdots & . & 1 & E(p_n) \\ 1 & \cdots & 1 & 0 & 0 \\ E(p_1) & \cdots & E(p_n) & 0 & 0 \end{bmatrix} \times \begin{bmatrix} n_1 \\ \vdots \\ n_n \\ \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ m \\ EXP \end{bmatrix} \qquad (30)$$

$$\begin{bmatrix} n_1 \\ \vdots \\ n_n \\ \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} . & \cdots & . & 1 & E(p_1) \\ \vdots & 2\,CovM & \vdots & \vdots & \vdots \\ . & \cdots & . & 1 & E(p_n) \\ 1 & \cdots & 1 & 0 & 0 \\ E(p_1) & \cdots & E(p_n) & 0 & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} 0 \\ \vdots \\ 0 \\ m \\ EXP \end{bmatrix} \tag{31}$$

As seen previously, the solution vector $[n_1 \dots n_i \dots n_n, \lambda, \mu]$ may contain some negative n_i values, which suggests that some sites are of no interest compared to the others. This may be due to a high variance and/or a low mean power output along with a high correlation with the other sites. As a matter of fact, some site wind turbine numbers n_i are found negative so as to provide the best sites with even more wind turbines despite the limitation number *m*. In this case, the wind turbine numbers n_i previously found negative must be set to 0 and the optimization process reiterated until a feasible solution is reached (i.e. $i \in [1 \dots n],\ n_i \geq 0$ and first order necessary KKT conditions satisfied). This way, the “poor sites” are basically removed from the preselected set.

The n_i can also be rounded up so as to obtain an exact number of wind turbines to be installed in every site (keeping in mind that *m* must remain the total number of wind turbines). Nevertheless, the decimal part of n_i can also be interpreted as the max capacity (in percent of 3MW) of a smaller wind turbine which features a *proportional* characteristic and that could be installed in addition to the usual 3MW Vestas V90 (and so on with other types of turbines). In the rest of the study, however, the calculations will be made on the basis of the rounded up results.

The resolution method can also be extended to *m* wind turbines of different characteristics, provided their power output statistical data is available. In this case, *n* takes on a more general meaning as it then designates the total number of wind/power output patterns to be considered. As a matter of fact, these patterns differ according to the geographical location (site), but also to the wind turbine characteristics. Each turbine type possesses its own statistical series p_i (also depending on the place) and must be consequently included in the equations formulating the optimization problem.

IV. APPLICATION STUDIES

The main goal of these application studies is to eventually compare the optimal distributions of 40 wind turbines over two different sets of 7 potential sites. The sets differ in the type of correlation the potential sites feature with respect to each other. Set #1 features positively correlated sites whereas Set #2 features uncorrelated sites.

IV. 1. Visualization of Optimal and Random Configurations in Terms of Global Power Output Coefficient of Variation vs. Expected Value

In this study, the optimization process was applied for the distribution of *40* wind turbines over two different sets of *7* preselected sites. The set/site characteristics will be thoroughly discussed and analyzed in the following section. The results are shown in Figures 3 and 4. The “dot cloud” is made of 15000 random distributions of *40* wind turbines, presumably accounting for the space of possible solutions. The plain line edging the bottom of the cloud (red thick line) is made of points obtained by running the optimization process for *EXP* values ranging from 48 to 57 MW (*EXP* lower and upper bounds in this case) with incremental steps of 0.1MW. The resulting line can be seen as the efficient frontier [8] standing for the optimal solutions given by the process described above. This mapping of configurations according to their coefficient of variation (defined by *Standard Deviation/Expected Value*) and their expected value for global wind power output is the key contribution of the present research, as it yields least variance configuration for any expected value of power output. The most reliable configurations, or at least the most predictable, are believed to be and will be searched for, in the later sections, on the efficient frontier.

As shown on the graphs and as expected, the optimal solutions show the lowest possible global variance (or on the graph, coefficient of variation) for a given global mean power output value and number of installed wind turbines, i.e. *40* here. One can observe that the efficient frontier also features a minimum. It may not be always the case, depending on the site selection. It will be also shown later that this is not always the most reliable configuration as a result of a trade-off between global power output maximization and global variance minimization. Computationally, the easiest way to get the most reliable configuration (which is not always the minimum of the efficient frontier as seen later) consists in asking the computer program to look for and store "the best $Std(G)/EXP$ minimum" as it is solving system (30) for different power output mean value EXP. Directly dealing with the optimization of the coefficient of variation would yield a non linear system which would be much less practical to solve.

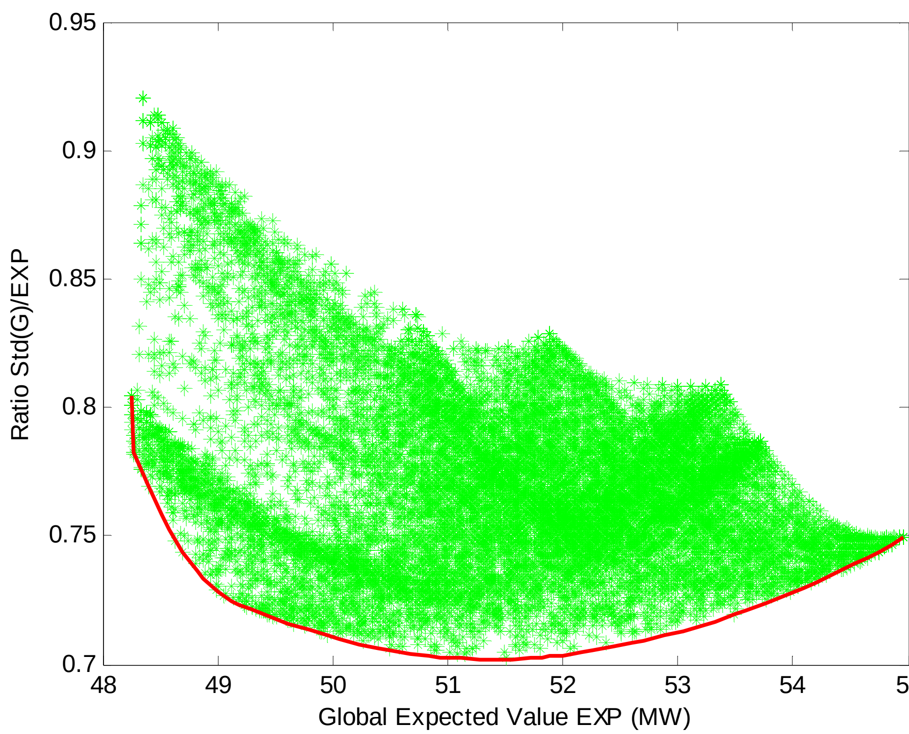

Fig. 3. Optimal and random configuration characteristics for set # 1 (positively correlated sites)

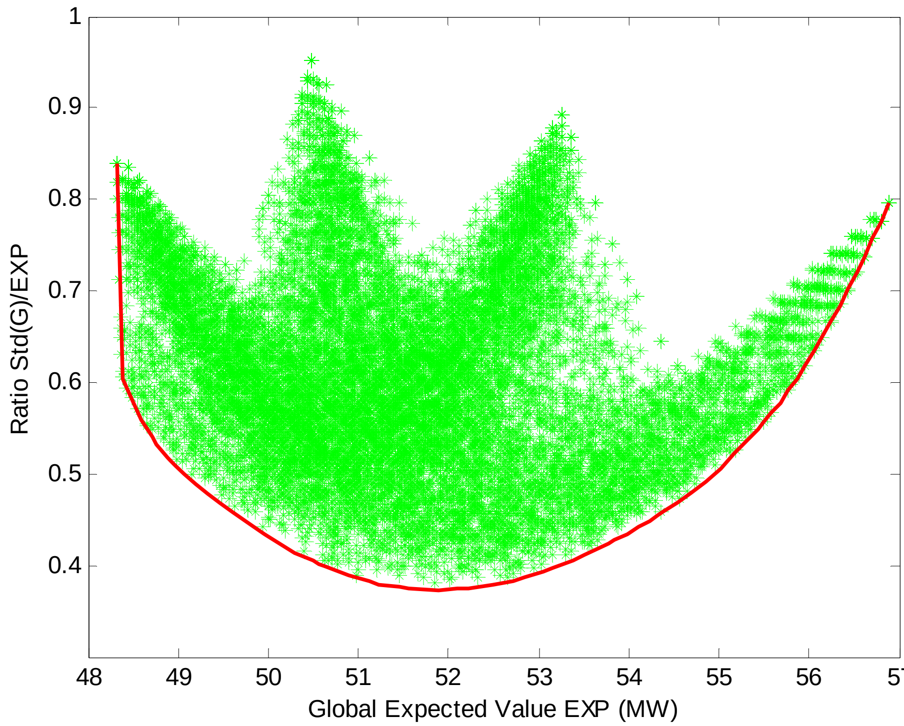

Fig. 4. Optimal and random configuration characteristics for set # 2 (uncorrelated sites)

Note that the first set of sites features mostly positively correlated sites in terms of power output, whereas the second comprises almost totally uncorrelated sites. It is already obvious from Figures 3 and 4 that the optimal configurations of Set #2, i.e. uncorrelated sites, have much smaller variances than their counter parts of Set #1 for the same expected values. One can then expect the uncorrelated sites to offer a much better reliability than more positively correlated configurations.

IV. 2. Sequential Monte Carlo Simulation Aimed at Emulating the Mechanical/Electrical Failures and Recoveries of Wind Turbines

Several methods have been proposed [12 – 15 are some examples] for the reliability analysis of wind power systems. Some methods use analytical approach and others use Monte Carlo Simulation. Here sequential Monte Carlo is used for its ease of implementation and because it also generates plausible histories that could prove useful in further research. It is possible to simulate the global power output of a so-called “optimal configuration” for every single hour of the 2004-2006 period. In the end, a comparison between global power output and load over the 3 years period will determine the level of reliability one can expect from a given configuration. However, wind turbines being vulnerable to mechanical and electrical failures as any other systems, their actual availability still remains to be modeled.

The basic idea here consists in running a sequential Monte Carlo Simulation over the 3 year period to emulate the potential failures and recoveries of every single wind turbine. To this end, Next Event procedure is implemented on MATLAB. The program basically samples (or updates) the system state every time a wind turbine fails or recovers; as such it waits for the next event. The system state is therefore known at all time during the simulation and the actual global wind power output originally calculated from the NREL/3TIER data is adjusted to take into accounts wind turbines that may be down. The time to next event is calculated as follows: let λ_f be the failure rate of a wind turbine and μ_r its recovery rate. The probabilistic distribution characterizing the time before failure/recovery of a wind turbine can be modeled as an exponential law:

$$Prob\left(t \leq t_f\right) = z_f = 1 - e^{-\lambda_f t_f} \tag{32}$$

$$Prob(t \leq t_r) = z_r = 1 - e^{-\mu_r t_r} \tag{33}$$

If we now express t_f and t_r respectively as a function of z_f and z_r, it comes:

$$t_f = -\frac{ln\left(1 - z_f\right)}{\lambda_f} \tag{34}$$

$$t_r = -\frac{ln(1 - z_r)}{\mu_r} \tag{35}$$

Here the notation *ln* designates the natural logarithm.

Using equations (34) and (35), one can simulate wind turbine times before failure (t_f) and recovery (t_r) by randomly drawing z_f and z_r between 0 and 1 when necessary, i.e. after an actual failure/recovery, a new time before recovery/failure is to be calculated. The MATLAB program can therefore be used to generate a complete history of any wind turbine provided λ_f and μ_r are known.

Reference [16] can help establish some estimates of MW class wind turbine transition rates based on German extended experience of wind turbine operation. B. Hahn et al. provides the annual frequency of failure f of various wind turbine power classes for up to 14 years of operability (7 years for the MW class) and the machine downtimes for every possible mechanical or electrical causes of failure, along with their statistical occurrence. From this information, we can compute the average mean down time (*MDT)* of a wind turbine and deduce its average mean up time *(MUP)*:

$$MUP = \frac{1}{f} - MDT \tag{36}$$

Which leads us to λ_f and μ_r, knowing that per definition:

$$\lambda_f = \frac{1}{MUP} \tag{37}$$

$$\mu_r = \frac{1}{MDT} \tag{38}$$

Downtimes due to preventive maintenance were not considered here. In a fully integrated power system, they should not significantly affect the wind park capacity credit as one can proceed to the preventive maintenance of one unit at a time, when the contribution of this latter is least needed, or plan for a conventional unit to compensate temporarily.

The following data (Table 1) describe the behavior of a MW class wind turbine in its first three years of operation (during which the failures are most frequent):

Table 1: Failure/recovery characteristics of a MW class wind turbine

	MUP	*MDT*	f	λ_f	μ_r
1st year	0.1430 years 1254.73 h	0.00695 years 60.89 h	6.663 /year 0.00076 /h	6.98158 /year 0.000797 /h	143.866 /year 0.016423 /h
2nd year	0.176607 years 1548.14 h	Same	5.448 /year 0.000621 /h	5.66228 /year 0.000646 /h	Same
3rd year	0.151157 years 1325.04 h	Same	6.325 /year 0.000722 /h	6.61566 /year 0.000755 /h	Same

With the knowledge of wind turbine statistical power outputs p_i and the failure/recovery rate λ_f and μ_r, we can obtain a complete reconstitution of the behavior of any of the *m* wind turbines distributed over the *n* preselected sites during the 2004-2006 period. Practically, the statistical power outputs p_i are extracted from the NREL/3TIER study. The sequential Monte Carlo simulation is then run so as to determine the hours during which a wind turbine should be declared out of order and its power output consequently set to 0.

With this data, the following step consists in assessing the wind turbine distribution loss of load probability (*LOLP*) by comparing the global power output to a given level of load. To this end, the 1996 IEEE Reliability Test Study [7] can be used to accurately model the hourly variation of the load over the 3 years and MATLAB be programmed to review the 26304 hours (about 3 years) of data and determine the number of hours td the system (solely composed of the *m* wind turbines) did not manage to meet the load. As a result, the *LOLP* estimates of any configurations over *any one iteration* of MCS, are computed as follows:

$$\widehat{LOLP} = \frac{td}{26304} \tag{39}$$

Other estimates such as the expected unserved energy, the system mean up/down time or the failure frequency can similarly and straightforwardly be evaluated.

The Sequential Monte Carlo Simulation is run as many times as required by the convergence criterion formalized in inequality (40), though a minimum of 10 iterations is imposed in order to ensure the plausibility of the results. In this study, the *LOLP* has been chosen as the reliability index to be used for judging the convergence of the simulation. As a matter of fact, the condition under which the simulation is acknowledged to have converged is given by:

$$\frac{\widehat{Std}\left(\widehat{LOLP}\right)}{\widehat{LOLP}} \leq tol \tag{40}$$

Where $\widehat{Std}(\widehat{LOLP})$ stands for the estimate of the standard deviation of the *LOLP* estimate $\widehat{LOLP}$ and *tol* designates the tolerance (fixed at 2.5% in the rest of the study). Also note that in (40), all the estimators are assessed upon the results provided by the current number of MCS iterations carried out up to that time.

This criterion is derived from the following consideration: its goal being to make sure the simulation provide realistic and reliable figures, it strives to get the standard deviation of the *LOLP* estimate (the true value being unavailable) to be *tol* % inferior to the estimate of its expected value, which basically means that once this criterion is satisfied, the value of the reliability index *LOLP* is very unlikely to vary significantly with further rounds of simulation. The reliability index is then seen as representative of the system overall behavior.

Inequality (40) can be rewritten the following way; starting with:

$$\widehat{LOLP} = \frac{1}{T}\sum_{i=1}^{N_T} td_i \tag{41}$$

With N_T as the number of system down times td_i, observed during the actual simulation time T encompassing all the iterations already run at this point. We obtain, after simplification (and denoting $\overline{td_\iota}$ the actual estimate of the system mean down time):

$$\widehat{Std}\left(\widehat{LOLP}\right) = \sqrt{\frac{\sum_{i=1}^{N_T}\left(td_i - \overline{td_\iota}\right)^2}{T^2}} \tag{42}$$

$$\Rightarrow \frac{\sqrt{\sum_{i=1}^{N_T}\left(td_i - \overline{td_i}\right)^2}}{\sum_{i=1}^{N_T} td_i} \leq tol \tag{43}$$

The last lines of the tables on pp.34 and 35 indicate the number of iterations that have been carried out right before the criterion was found satisfied with a tolerance *tol* of 2.5% and convergence has been achieved. Notice that high $\widehat{LOLP}$ actually cause the Monte Carlo simulations to converge fast, so that 10 minimum iterations usually suffice.

IV. 3. Wind Park Global Power Output vs. Load

a. Characteristics of the potential wind farming sites studied in this research

This research primarily examines 2 configurations of *m=40* Vestas V90 wind turbines (3MW capacity) scattered over *n=7* sites. The sites show very similar mean power outputs (about 1.3MW for an installed capacity of 3MW per turbine) and variances (each wind turbines having a standard variation of about 1 MW) for us to better highlight the influence of diversification. The first set of preselected sites shows some strong positive correlation between every local power output (see correlation matrix *CorrM1*, fig 5), all sites being chosen in west Texas. The second configuration consists of sites distributed all over the West US (namely in Oklahoma, California, Oregon, Wyoming, Montana and North Dakota). As a consequence, the power outputs of these sites appear almost completely uncorrelated (see correlation matrix *CorrM*2, fig 6).

$$CorrM1 = \begin{bmatrix} 1.0000 & 0.5544 & 0.6614 & 0.7154 & 0.7625 & 0.7610 & 0.7383 \\ 0.5544 & 1.0000 & 0.7016 & 0.6985 & 0.6932 & 0.6833 & 0.7248 \\ 0.6614 & 0.7016 & 1.0000 & 0.8681 & 0.8022 & 0.8276 & 0.7830 \\ 0.7154 & 0.6985 & 0.8681 & 1.0000 & 0.9179 & 0.8971 & 0.9080 \\ 0.7625 & 0.6932 & 0.8022 & 0.9179 & 1.0000 & 0.9502 & 0.9393 \\ 0.7610 & 0.6833 & 0.8276 & 0.8971 & 0.9502 & 1.0000 & 0.8911 \\ 0.7383 & 0.7248 & 0.7830 & 0.9080 & 0.9393 & 0.8911 & 1.0000 \end{bmatrix}$$

Fig. 5. Correlation matrix of set #1 - positively correlated sites

$$CorrM2 = \begin{bmatrix} 1.0000 & 0.2157 & -0.0635 & 0.0307 & 0.1562 & -0.0981 & -0.0553 \\ 0.2157 & 1.0000 & -0.0854 & 0.1494 & 0.1472 & 0.0496 & 0.0434 \\ -0.0635 & -0.0854 & 1.0000 & -0.0583 & -0.0401 & -0.0717 & -0.0999 \\ 0.0307 & 0.1494 & -0.0583 & 1.0000 & 0.2476 & 0.4411 & 0.2545 \\ 0.1562 & 0.1472 & -0.0401 & 0.2476 & 1.0000 & 0.1862 & 0.0925 \\ -0.0981 & 0.0496 & -0.0717 & 0.4411 & 0.1862 & 1.0000 & 0.3057 \\ -0.0553 & 0.0434 & -0.0999 & 0.2545 & 0.0925 & 0.3057 & 1.0000 \end{bmatrix}$$

Fig. 6. Correlation matrix of set # 2 - uncorrelated sites

Tables 2 and 3 outline the main characteristics of each site (within each set) by giving, over the 2004-2006 period, the mean power output $E(p_i)$ of a single wind turbine, as well as its standard deviation $Std(p_i)$ and its coefficient of variation $Std(p_i)/E(p_i)$ (which permits to assess *relatively* the dispersion of statistical values around their mean).

Table 2: Characteristics of set #1 (positively correlated sites) in terms of wind turbine power output mean value, standard deviation and ratio Std(pi)/E(pi)

Set #1	Site # 763 (TX)	Site # 5073 (TX)	Site # 2296 (TX)	Site # 1884 (TX)	Site # 1329 (TX)	Site # 1368 (TX)	Site # 1506 (TX)	Set Mean
E(pi) (MW)	1.2670	1.2062	1.2080	1.3342	1.3420	1.2970	1.3740	1.2901
Std(pi) (MW)	1.0590	0.9704	1.1120	1.0788	1.0560	1.0747	1.0298	1.0546
Std(pi)/E(pi)	0.8350	0.8045	0.9200	0.8085	0.7870	0.8287	0.7495	0.8191

Table 3: Characteristics of set #2 (uncorrelated sites) in terms of wind turbine power output mean value, standard deviation and ratio Std(pi)/E(pi)

Set #2	Site # 4262 (CA)	Site # 7186 (OK)	Site # 23098 (OR)	Site # 20179 (WY)	Site # 26796 (OR)	Site # 28134 (MT)	Site # 30353 (ND)	Set Mean
E(pi) (MW)	1.2621	1.2081	1.4223	1.3263	1.2666	1.3317	1.2111	1.2897
Std(pi) (MW)	1.2018	1.0133	1.1319	1.1128	1.1261	1.1881	1.0124	1.1123
Std(pi)/E(pi)	0.9522	0.8388	0.7959	0.8391	0.8890	0.8922	0.8359	0.8633

One can notice that both configurations are very comparable in terms of mean power output and standard deviation (see “Set Mean” in Tables 2 and 3). Set #1 even looks slightly better on the whole, especially in terms of ratio $Std(p_i)/E(p_i)$. The main difference actually comes from the site correlation, as shown by the correlation matrices *CorrM1* and *CorrM2*. (Resp. fig 5 & 6)

b. Simulation results

Now if we run the simulation for various peak loads ranging from *5MW* to *120MW* (*120MW* being the installed capacity), compute reliability indexes and search for the optimal configurations in terms of *LOLP* (and therefore, in a way, reliability), we can assess and compare the capacity credits of both sets. The following estimates

(Tables 4 and 5) were derived for the most reliable configuration that could be found on the efficient frontier (the program being asked to compute the *LOLP* for every point of the efficient frontier and then find the best match), as well as for 50 random configurations for which the power output expected values were in the vicinity of the efficient frontier best configuration's. The four *most* reliable random configurations were eventually retained so as to figure in Tables 4 and 5 and illustrate the fact that the efficient frontier best solution generally stands for one of the most reliable possible distribution, if not the best of all.

Practically, the indexes were calculated based on the comparison of the hourly available global power output with the corresponding load. The hourly global power output was taken as the minimum value of the six 10 min-span power output estimations so as to be as close to the reality as possible. As a matter of fact, it would not have been relevant to compare power output and load every 10 minutes because of the given hourly scale of the load data. Furthermore, from an operational point of view, it seems hard to exploit an energy so inherently intermittent that it may be able to supply the entire system for 20 or 30 minutes before fading again… Consequently, if the global power output was found unable to supply the entire load for a given 10 minutes span, it was declared unable to do so for the whole corresponding hour (this is a pessimistic model at worst, a realistic one at best).

Table 4: Simulated indexes for period 2004 to 2006 – set #1 (positively correlated sites)

	Set #1 - Positively Correlated sites - Total Number of Wind Turbines: 40 units							
		Annual Peak Load (MW)						
		5	10	25	50	75	100	120
"Efficient Frontier" Most Reliable Configuration	Wind Turbine # Site 763 (TX)	11	12	9	9	0	0	0
	Wind Turbine # Site 5073 (TX)	16	18	13	11	0	0	0
	Wind Turbine # Site 2296 (TX)	0	0	0	0	0	0	0
	Wind Turbine # Site 1884 (TX)	0	0	0	0	0	0	0
	Wind Turbine # Site 1329 (TX)	0	0	0	0	0	0	0
	Wind Turbine # Site 1368 (TX)	0	0	0	0	0	0	0
	Wind Turbine # Site 1506 (TX)	13	10	18	20	40	40	40
	Mean Power Output without Mechanical/Electrical failures (MW)	51.1086	50.6667	51.8245	52.1601	54.9619	54.9619	54.9619
	Mean Power Output with Mechanical/Electrical failures MPO (MW)	43.458	43.0179	44.0553	44.3935	46.893	46.8368	46.8368
	Ratio: Standard Deviation / MPO	0.7687	0.7701	0.7707	0.7728	0.8262	0.8258	0.8258
	LOLP	0.0852	0.1419	0.2752	0.4317	0.5441	0.6365	0.7057
	Failure Frequency (per year)	163.3667	226.1667	308.1333	351.2333	418.2667	398.8333	378.5667
	Mean Up Time (hours)	147.0061	99.9838	63.0331	42.8289	28.5178	23.8229	20.5513
	Mean Down Time (hours)	13.7223	16.5068	23.4953	32.3344	34.2158	41.9829	49.041
	Expected Unserved Energy (EUE) (*10^4 MWh/year)	0.1517167	0.4821333	2.3400667	7.4203333	14.984	23.227	30.703333
4 "Most Reliable" Configurations picked among 50 random Configurations	Mean Power Output without Mechanical/Electrical failures ExP (MW)	51.1139	51.0836	51.5003	53.5868	54.7739	54.6485	53.4567
		51.5164	51.5293	50.1736	54.0032	54.9619	54.6511	53.3026
		51.6433	52.4014	50.6955	53.9015	54.5179	53.9611	53.9611
		51.4530	52.2612	50.9845	54.1442	54.5255	54.6699	53.4405
	Mean Power Output with Mechanical/Electrical failures MPO (MW)	43.3881	43.4442	43.9305	45.6557	46.8181	46.7598	45.9576
		43.7853	43.8354	42.7013	46.0684	46.9132	46.7148	45.7721
		44.1132	44.6789	43.2166	46.0981	46.5701	46.3707	46.3707
		43.8963	44.5414	43.4528	46.1390	46.4797	46.7079	45.9484
	Ratio: Standard Deviation / MPO	0.7859	0.7724	0.7782	0.7984	0.8250	0.8258	0.8785
		0.7912	0.7846	0.7850	0.8077	0.8259	0.8250	0.8816
		0.7992	0.7866	0.7987	0.8095	0.8258	0.8369	0.8369
		0.8182	0.7942	0.7985	0.8133	0.8215	0.8183	0.8597
	LOLP	0.0959	0.1456	0.2817	0.4350	0.5432	0.6348	0.7033
		0.1038	0.1536	0.2901	0.4355	0.5438	0.6360	0.7043
		0.1041	0.1567	0.2945	0.4379	0.5454	0.6366	0.7053
		0.1065	0.1601	0.2953	0.4386	0.5465	0.6377	0.7063
	Failure Frequency (per year)	185.1667	217.2667	300.0333	386.2000	403.0333	374.9667	345.5000
		197.1333	231.3000	300.0667	394.6667	419.2333	385.7333	341.9000
		182.8333	231.3667	318.3333	374.2333	396.4667	357.8000	334.4333
		184.7667	228.6333	299.7667	393.0000	406.4333	387.8000	343.0000
	Mean Up Time (hours)	128.4336	103.4474	62.9759	38.4856	29.8120	25.6218	22.5937
		119.5915	96.2576	62.2309	37.6226	28.6254	24.8211	22.7532
		128.8989	95.8810	58.2960	39.5060	30.1622	26.7142	23.1781
		127.1998	96.6322	61.8407	37.5738	29.3510	24.5744	22.5237
	Mean Down Time (hours)	13.6276	17.6245	24.6982	29.6260	35.4543	44.5323	53.5445
		13.8454	17.4673	25.4338	29.0279	34.1197	43.3727	54.1853
		14.9772	17.8133	24.3393	30.7831	36.1866	46.8054	55.4781
		15.1676	18.4190	25.9089	29.3604	35.3696	43.2561	54.1674
	Expected Unserved Energy (EUE) (*10^4 MWh/year)	0.1721	0.4943	2.3992	7.7263	14.9660	23.1930	31.6427
		0.1884	0.5369	2.4645	7.8520	14.9767	23.2213	31.7587
		0.1869	0.5497	2.5165	7.8803	15.0013	23.4443	30.8917
		0.1958	0.5619	2.5577	7.9610	14.9770	23.1227	31.3903
Number of Iterations at MCS convergence - Tolerance=2.5% - 10 iterations minimum		10	10	10	10	10	10	10

Table 5: Simulated indexes for period 2004 to 2006 – set #2 (uncorrelated sites)

	Set #2 - Uncorrelated sites - Total Number of Wind Turbines: 40 units							
		Annual Peak Load (MW)						
		5	10	25	50	75	100	120
"Efficient Frontier" Most Reliable Configuration	Wind Turbine # Site 4262 (CA)	6	6	6	6	3	0	0
	Wind Turbine # Site 7186 (OK)	8	7	6	5	0	0	0
	Wind Turbine # Site 23098 (OR)	8	9	10	12	20	40	40
	Wind Turbine # Site 20179 (WY)	2	2	3	4	8	0	0
	Wind Turbine # Site 26796 (OR)	4	4	4	3	1	0	0
	Wind Turbine # Site 28134 (MT)	3	3	4	5	8	0	0
	Wind Turbine # Site 30353 (ND)	9	9	7	5	0	0	0
	Mean Power Output without Mechanical/Electrical failures (MW)	51.3	51.444	51.8938	52.4993	54.7617	56.8901	56.8901
	Mean Power Output with Mechanical/Electrical failures MPO (MW)	44.9044	45.1838	45.6387	46.1553	48.3441	50.6961	50.725
	Ratio: Standard Deviation / MPO	0.4036	0.399	0.3948	0.4024	0.5201	0.85	0.8505
	LOLP	0.0012	0.0061	0.0441	0.2244	0.4832	0.5884	0.6464
	Failure Frequency (per year)	4.1533	17.0175	74.1667	265.1333	334.1667	253.6	262.7333
	Mean Up Time (hours)	6569.3	1508.5	341.1448	78.8698	39.627	42.2308	36.1392
	Mean Down Time (hours)	7.3602	9.4078	15.6364	22.2693	38.0413	61.0356	64.7152
	Expected Unserved Energy (EUE) (*10^4 MWh/year)	0.0009249	0.0111676	0.2205933	2.0967	8.7503333	23.082333	29.943333
4 "Most Reliable" Configurations picked among 50 random Configurations	Mean Power Output without Mechanical/Electrical failures ExP (MW)	53.0139	51.1510	51.0688	53.2869	54.1772	56.3870	56.7300
		50.5193	52.0712	52.9844	53.8303	53.7089	56.0577	56.3282
		52.2024	52.6013	51.0234	53.1953	53.6960	55.8344	55.8475
		52.5537	52.2957	51.3078	53.5125	53.4548	55.8605	52.9873
	Mean Power Output with Mechanical/Electrical failures MPO (MW)	46.6427	44.7205	44.3877	46.8488	47.8433	50.2745	50.6319
		43.8906	45.6516	46.6410	47.5348	47.3906	49.8974	50.2270
		46.2162	46.1857	45.0780	47.0984	47.3909	49.7171	49.7040
		46.6240	46.0175	45.3940	47.2317	47.1080	49.7532	46.4282
	Ratio: Standard Deviation / MPO	0.4302	0.4180	0.4577	0.4611	0.5608	0.7706	0.8300
		0.5044	0.4058	0.4826	0.4635	0.4904	0.7314	0.7705
		0.4491	0.4305	0.4736	0.5049	0.4645	0.7550	0.6776
		0.4696	0.4371	0.5065	0.4832	0.4830	0.7200	0.8912
	LOLP	0.0018	0.0074	0.0730	0.2483	0.4812	0.5959	0.6503
		0.0021	0.0083	0.0757	0.2576	0.4880	0.6024	0.6654
		0.0024	0.0114	0.0774	0.2718	0.4904	0.6024	0.6975
		0.0027	0.0114	0.0912	0.2769	0.4906	0.6031	0.7036
	Failure Frequency (per year)	6.4444	19.5088	127.6667	279.3667	359.0333	259.0333	265.0667
		7.2989	22.6842	121.6000	281.6667	334.0333	268.9000	265.4000
		5.9042	24.8070	127.1333	257.9000	373.7333	264.7333	274.7000
		7.3180	25.2281	140.3000	275.0333	345.0667	268.9333	275.5667
	Mean Up Time (hours)	4097.0	1340.7	191.0610	70.7834	38.0099	41.0380	34.7033
		3625.9	1151.3	199.9900	69.3348	40.3234	38.8994	33.1607
		4464.9	1050.0	191.0003	74.2723	35.8674	39.5074	28.9729
		3601.3	1031.7	170.4112	69.1658	38.8340	38.8192	28.2912
	Mean Down Time (hours)	7.3823	9.9326	15.0449	23.3850	35.2617	60.5127	64.5399
		7.5906	9.6290	16.3723	24.0586	38.4258	58.9294	65.9560
		10.7082	12.0827	16.0152	27.7274	34.5190	59.8604	66.7901
		9.5908	11.9004	17.0970	26.4909	37.4015	58.9969	67.1663
	Expected Unserved Energy (EUE) (*10^4 MWh/year)	0.0016	0.0135	0.3806	2.5890	9.3187	21.5813	29.5967
		0.0016	0.0156	0.4448	2.6763	8.4327	20.9120	28.6327
		0.0032	0.0221	0.3837	3.1726	8.1463	21.4093	27.0960
		0.0028	0.0221	0.4820	3.1567	8.4107	20.6390	31.2787
Number of Iterations at MCS convergence - Tolerance=2.5% - 10 iterations minimum		87	19	10	10	10	10	10

c. Interpretation and comments

Tables 4 and 5 highlight the results of the optimization process and the advantages of wind farm diversification across uncorrelated (or probably even better: negatively correlated) sites. The indexes are significantly improved when considering Set #2 configurations, especially for reasonable peak loads. A small power output variance will certainly improve the reliability of systems whose load is less than the power output expected value; however, for higher loads, it is unlikely to go well beyond its power output mean value, thus meaning it may never have the opportunity to catch up with the load, contrarily to a high variance system which is so unpredictable that it can go anywhere. Still, the small variance system remains preferable in terms of operability as it proves more predictable, though this may be at the price of some power capacity. Figures 7 and 8 respectively depict the probability densities of global power output of Set #1 and Set #2 based on the NREL/3TIER statistical data for years 2004 to 2006. The graphs were plotted for the optima curve most reliable dispatch and random configurations of $m=40$ wind turbines given a peak load of 50MW. It appears that the optimal configuration (red thick curve) features smaller probabilities for supplying very small or very high amount of power than the others. However, the probabilities for delivering medium amount of power are far greater. This illustrates what was previously said: small variance means staying close to the mean value most of the time. Also note that in the case of Set #2, the probabilistic distribution of the optimal configuration looks like a Weibull distribution, or perhaps even a Normal Gaussian distribution, which hints at the Central Limit theorem stating that the sum of a large number of random variable (with finite variances) tends to a Normal distribution. Here, the number n of random variables p_i may not be sufficient to shape a nice Gaussian (or the p_i may not have the same importance enough in the sum, thus jeopardizing the sum of large number of uncorrelated random variables), but the probabilistic distribution of the global power output of a *large* number of *uncorrelated* sites may indeed be modeled as a Normal distribution.

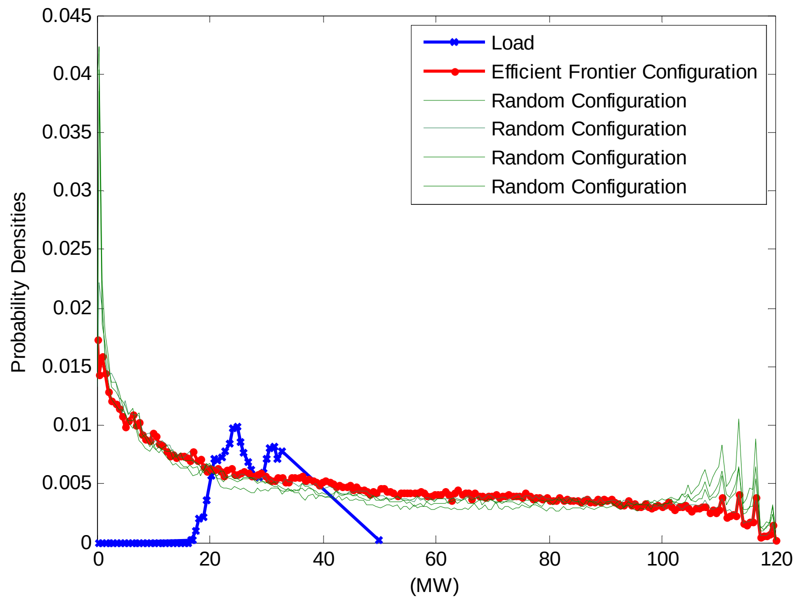

Fig. 7. Global power output and load probability densities for set #1 (positively correlated sites) - Efficient frontier distribution and some random configurations for a peak load of 50 MW

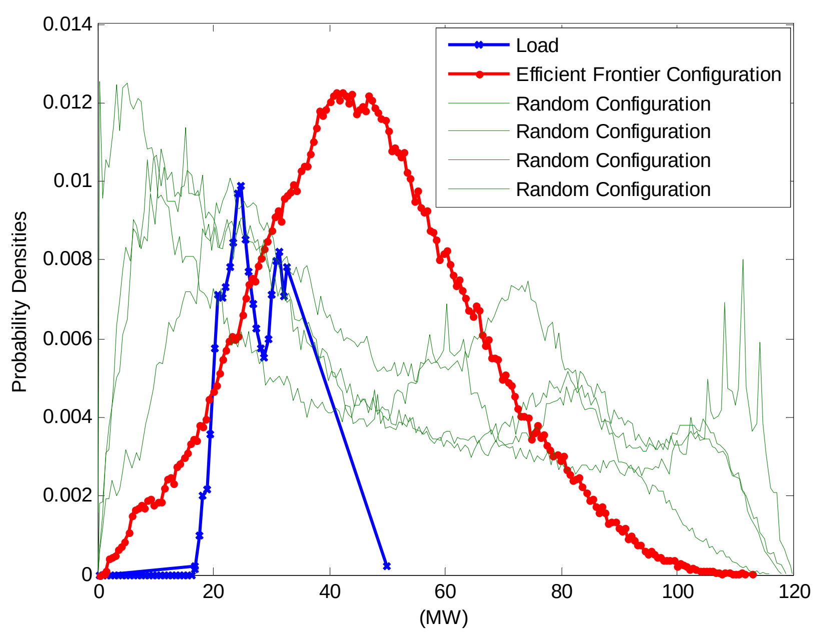

Fig. 8. Global power output and load probability densities for set #2 (uncorrelated sites) - Efficient frontier distribution and some random configurations for a peak load of 50 MW

As for the search for the most reliable configuration, one can observe (Tables 4 and 5) that different peak loads call for different optimal configurations, which actually highlights an optimization trade-off between minimization of the global variance and maximization of the global expected value. As a matter of fact, for small and medium peak loads, the efficient frontier most reliable dispatch can usually be considered as the best configuration in terms of *LOLP*. Reasonable loads can be met by the many configurations featuring a medium power output mean value. Such a large number of possible combinations favors diversification and thus allows significant reductions of variance. Hence, as small variances tend to prevent the power output from falling behind a too small value, the most suitable configuration for low-medium loads turns out to be the one that best minimizes the variance. However, for high peak loads, the best possible reliability can only be achieved via high power output mean values. This considerably limits the number of possible combinations and consequently hinders the diversification process. Actually, the constraint on mean power output is so important that it completely overshadows variance reduction. This is illustrated by cases for which peak loads exceed 100 MW: the load is so high that the 40 wind turbines are attributed to wind farm #3 (Set #2 case), the site showing the best power output mean value per turbine (1.4223MW). Even then, the global power output expected value does not go beyond 57 MW, which is small compared to load values revolving around 100 MW. Here then, a high variance may be desirable so as to reach (at least occasionally) very high values and catch up with the load. Minimizing the variance does improve the reliability when the power output expected value is comfortably greater than load values. This can be seen as well on Figure (8): the method presented in this paper actually strives to reduce the overlap between power output and load probabilistic densities as long as the power output is well ahead in terms of Mega Watts.

The optimization process can also be made slightly more efficient (as of the search for the optima curve) by integrating the load data into the equations depicted in Section III. Instead of solely minimizing the global power output variance, one can strive to minimize the variance of the random variable defined by the difference between the power output and the load (provided this latter is statistically well known) under the same constraints as previously described. This method has the advantage of taking into account the correlation between the load and the wind turbine power output profiles. However, correlation coefficient calculations show that (in this case at least, the load being taken from the 1996 IEEE RTS), load and power outputs are uncorrelated, which eventually comes down to solve the optimization problem described in Section III.

Back to Tables 4 and 5, one can also examine the wind distributions of Sets #1 and #2 and see that some sites remain unused within a set showing too much positive correlation (Set #1). In this case, it seems the best sites are selected because of their small ratio $Std(p_i)/E(p_i)$ and relatively small correlation with each other. On the other hand, Set #2 features a quite homogeneous repartition of its sites for small-medium peak loads, meaning the site non-correlation fact is fully exploited.

d. Conclusion

On the whole however, wind-sustained systems do not score very well in terms of reliability. Even by selecting uncorrelated sites (and better if possible: negatively correlated) and proceeding to the optimization of their wind turbine repartition, the loss of load probability (*LOLP*) remain high despite low peak loads. Referring to Table 5, one can observe that for a peak load of 50 MW, which is less than half of the installed capacity of 120 MW, the *LOLP* is still greater than 20%, while the system failure frequency amounts to about 265 loss of load events a year (not that far from a loss event per day!). A solution to these poor ratings may lie in the selection of a higher number of sites *n*, knowing that the more *negatively* correlated, the better.

However, the main interest in diversification may rather lie in the enhanced predictability and stability of wind power outputs. In a fully integrated system, some thermal conventional units will be committed to the compensation of wind erratic behavior. The main question remains the design of these generators, which heavily depends on potential wind fluctuations. As such, a decrease in wind power variance should alleviate system rates requirements and “bound” them. Some more studies could be envisioned so as to evaluate ramping, i.e. the variation speeds of power swings.

Note that the results of the optimization process are intimately related to the mean values, variances and covariances used into equations (30). Also, if these values are calculated for one year instead of three (as it is done in this paper), the optimal configurations may significantly change. Tables 6 and 7 show the optimal configurations given by the optimization process for both sets with data based solely on year 2004, 2005 or 2006 and a peak load of 50 MW.

Table 6: Optimal distributions for set #1 (total of 40 wind turbines)

Set #1	Site # 763 (TX)	Site # 5073 (TX)	Site # 2296 (TX)	Site # 1884 (TX)	Site # 1329 (TX)	Site # 1368 (TX)	Site # 1506 (TX)
2004	8	7	0	0	0	0	25
2005	5	12	0	0	0	0	23
2006	10	7	0	0	0	0	23
3 years	9	11	0	0	0	0	20

Table 7: Optimal distributions for set #2 (total of 40 wind turbines)

Set #2	Site # 4262 (CA)	Site # 7186 (OK)	Site # 23098 (OR)	Site # 20179 (WY)	Site # 26796 (OR)	Site # 28134 (MT)	Site # 30353 (ND)
2004	7	5	12	4	3	5	4
2005	9	2	12	6	2	7	2
2006	4	5	13	4	3	5	6
3 years	6	5	12	4	3	5	5

The results actually change considerably from year to year, though some tendencies remain. It proves that in the long run, it is important to have statistical data over as many years as possible to obtain a well optimized system. But once again, this is a long run consideration based on the assumption that history will repeat itself.

IV. 4. Correlation between Load and Global Wind Power Output as Given by Optimal Configurations

The NREL/3TIER data coupled with the 1996 IEEE RTS can also be of use in the study of the correlation relating wind power with load. MATLAB can directly calculate the correlation coefficient characterizing the statistical hourly global power output and load data. The results of such a calculation prove unequivocal for both sets: about -0.0594 for Set #1 and -0.0516 for Set #2 on average (the hourly global power output was derived from the optimal configurations seen in Tables 4 and 5). This means that wind power, as given by the aforementioned configuration, and load behave independently over the course of the year. In other words, they are totally uncorrelated. Though this is the case over long time periods such as a year, this may not be true for smaller time spans, as suggested by Figures 9 and 10, representing the mean values of the load (for a peak load of *50MW*) and the global power outputs of Sets #1 and #2 most reliable configurations with respect to hours of the day and weeks of the year.

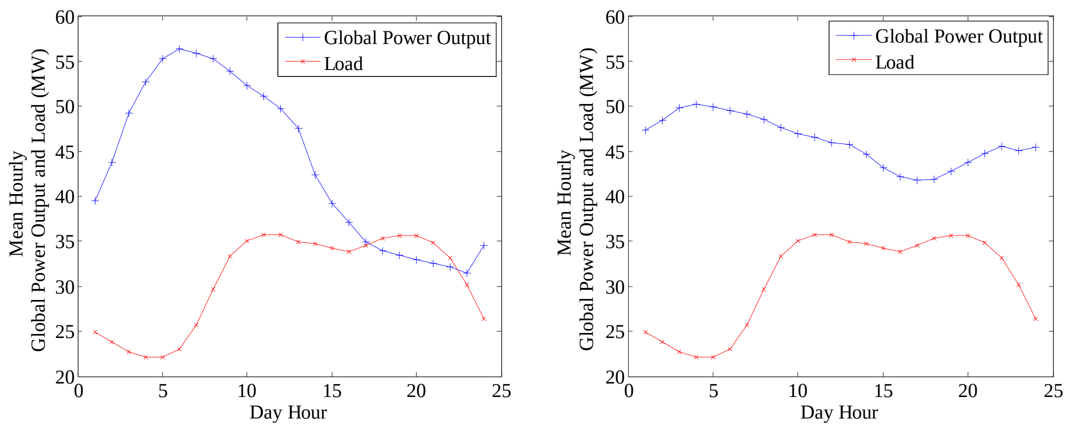

Fig. 9. Wind power output and load with hours of the day (respectively sets #1 and #2 most reliable configurations for a peak load of 50MW)

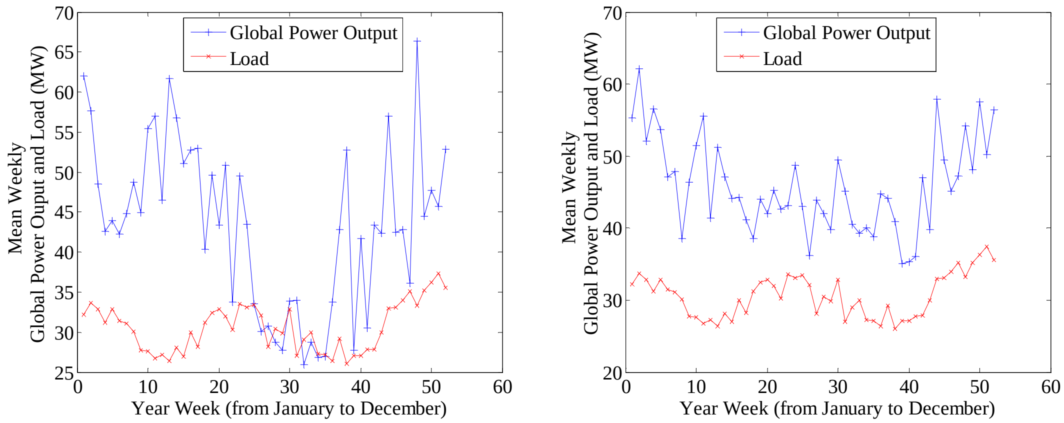

Fig. 10. Wind power output and load with weeks of the year (respectively sets #1 and #2 most reliable configurations for a peak load of 50MW)

One can quickly notice the reduction in variance induced by the good diversification of wind farming sites in the case of Set #2 with respect to that of Set #1, while the expected global power output remains almost the same in both cases. Note that the global power output and load averages calculated for a given time slot are based on the mean of all the corresponding values available over the 3 years of data. All graphs exhibit (about equally distributed) time periods for which wind power and load show positive or negative correlation. However, on the scale of a year, these “alternating” correlations cancel each other out so that the yearly correlation eventually goes to 0. Still, Figures 9 and 10 can help predict some overall loss-of-load patterns for short spans of time. For example, a positive correlation between wind power and load may indicate that the loss of load frequency should be about the same from one hour (resp. week) to another. Conversely, negative correlation may point at a worsening in existing discrepancies. These tendencies may be visualized in Figures 11 and 12 respectively depicting estimations of Sets #1 and #2 annual loss of load frequency for any hour of the day (most reliable configurations) in two cases: one for which the load sticks to its usual pattern (fig 11) (the peak load being fixed at *50*MW) and the other for which the load is kept constant throughout the year (fig 12) (fixed at *30.7MW*, which is the mean value of the load throughout the year given a peak load of *50MW*).

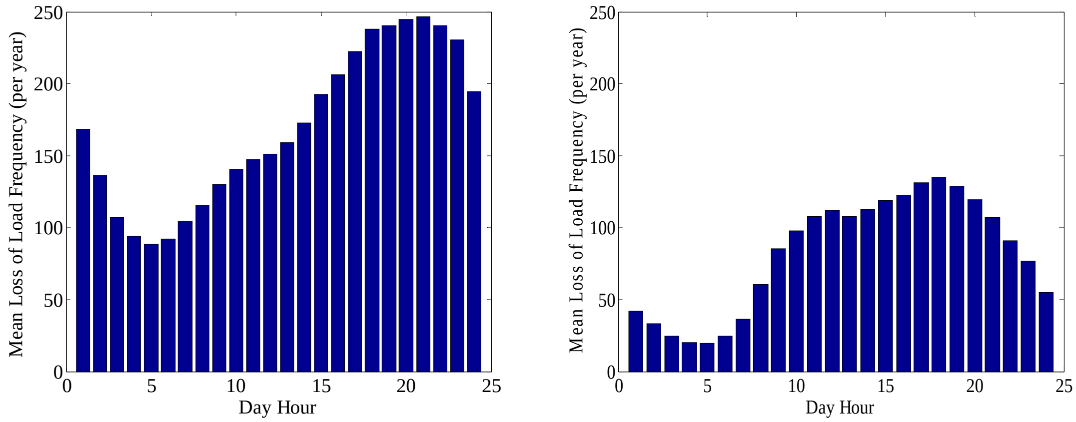

Fig. 11. Annual loss of load frequency for any hour of the day (respectively based on sets #1 and #2 results, most reliable configurations); usual load pattern

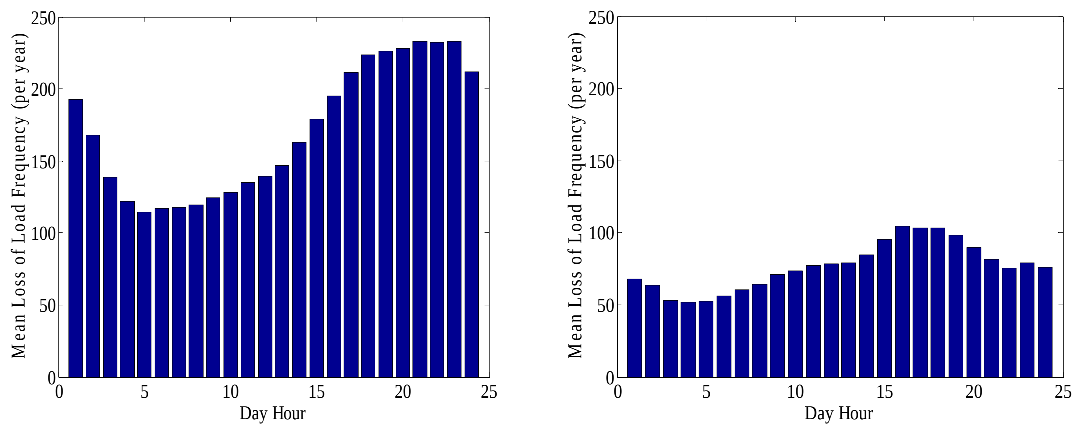

Fig. 12. Annual loss of load frequency for any hour of the day (respectively based on sets #1 and #2 results, most reliable configurations); constant load of 30.7 MW

Overlooking better performance of Set #2, all figures show quite similar patterns. However, when the load is allowed to vary, the pattern generally looks more “extreme”, meaning more discrepancies between hourly mean loss of load frequency. This can be explained by looking at Figure 9: in the case of Set #2 for example, in the beginning/middle of the night (from 21:00 to 3:00) and morning (from 6:00 to 10:00), power output and load are clearly negatively correlated: they vary in opposite directions. As a result, the mean loss of load frequency experience wild variations from an hour to another. Between 17:00 and 20:00 however, power output and load show some positive correlation, which results in much less significant variation of the mean loss of load frequency. As a matter of fact, one can see the negative correlation between wind power and load as a sign of fast change of the current system state, thus designating “hot spots” to the system operators. On the other hand, positive correlation accounts for stability, steadiness of the situation.

Figure 12 roughly illustrates the fact that wind is generally more powerful at night, which could be seen as another incentive for wind energy night storage. In any case, it shows that wind patterns statistically differ over time. Figures 13 and 14 describe the same thing as Figures 11 and 12, but at weekly (season) scale. In this case however, given the few differences between graphs of a same case, it seems like the load has less influence over the loss of load frequency than the wind which varies more intensely and “dictates” the system state.

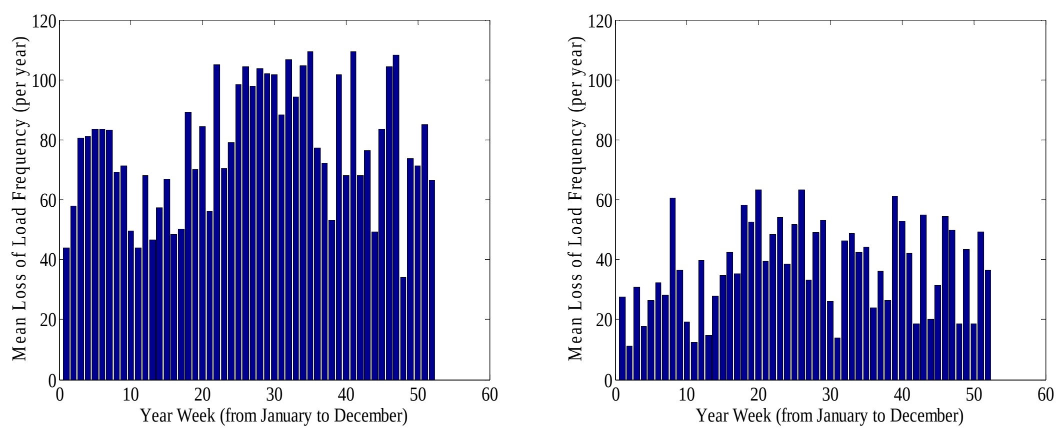

Fig. 13. Annual loss of load frequency for any week of the year (respectively based on sets #1 and #2 results, most reliable configurations); usual load pattern

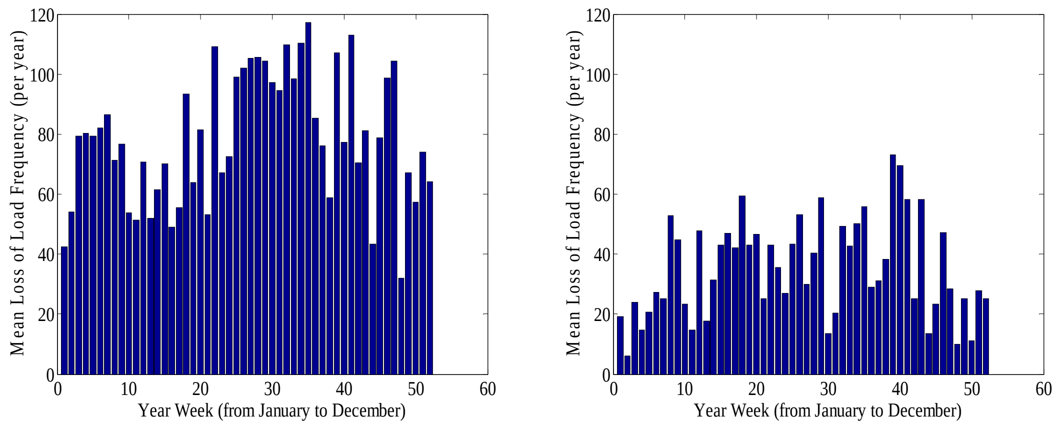

Fig. 14. Annual loss of load frequency for any week of the year (respectively based on sets #1 and #2 results, most reliable configurations); constant load of 30.7 MW

These graphs could prove useful in a thorough planning study that would consider distributing wind turbines in order to alleviate "hot spots" threats in priority. As a matter of fact, it appears, on the overall and given the particular wind farming locations that have been studied in this thesis, that wind mainly blows in the night while slightly abating when most needed before the load peaking of 8:00 pm. Also, on a seasonal scale, wind power tends to decrease in summer. It could then be interesting to run the optimization on fragments of data that mainly includes day "hot spot" hours in summer time and see if the reliability gets improved.

V. CONCLUSION

This research has introduced a general method for improving wind energy based system reliability through the optimal distribution of a given number of wind turbines over a given number of preselected sites. The illustration of the method is based on 3 years of projected wind power output statistics provided by the NREL/3TIER Western Wind and Solar Integration Study. The optimization process strives to minimize the global power output variance for fixed expected power, primarily taking advantage of uncorrelated or negatively correlated sites to "smooth" the global power output and make it more predictable.

It has been shown that the global minimum of the (convex) optimization problem could be found via the use of an active set method or similar algorithm that proceeds by activating progressively the inequality constraints. The latter procedure has then been applied to two sets of potential sites featuring various degrees of correlation. The first set consists of highly positively correlated sites while the second comprises almost independent sites so as to study the potential benefits of diversification relatively to system reliability.

After obtaining the optimal partitions of 40 wind turbines for both sets, a sequential Monte Carlo simulation has been run so as to take into account the mechanical/electrical failures any wind turbine may undergo throughout its operational period. Various reliability indexes such as the loss of load probability *LOLP* have been calculated by recreating 3 years of statistical wind power output and comparing it to the yearly load model proposed the 1996 IEEE RTS. A search for the most reliable distribution has also been carried out and it has been determined that configurations featuring the lowest power output variances also yielded the smallest *LOLP* for peak loads of the same (or lesser) order as their mean power output.

Optimal wind power outputs from both sets and the load model from the 1996 IEEE RTS have been compared over different time scales and their correlation has been found to be close to 0 on the overall. However, some specific time spans have been shown to pose more threats than others to the system reliability, and further research could be made on optimizations primarily prone to deal with those hot spots.

In the end, the results of this study show that the correlations between wind farm global power outputs significantly impact the system reliability. Sets of uncorrelated/negatively correlated sites are to be preferred in the planning of large wind turbines installments. Some more research on the nature of low variance power outputs can be done in order to establish if they can improve the success rates of actual wind predictors and help design conventional units dedicated to wind power output swings compensation.

REFERENCES

[1] R. D. Prasad, R. C. Bansal, M. Sauturaga, “Some of the Design and Methodology Considerations in Wind Resource Assessment”, *IET Renew. Power Gener.*, vol. 3, no. 1, pp. 53-64, 2009.

[2] F. Vallée, J. Lobry, O. Deblecker, “System Reliability Assessment Method for Wind Power Integration”, *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1288-1297, Aug. 2008.

[3] C.W. Potter, D. Lew, J. McCaa, S. Cheng, S. Eichelberger, E. Grimit, “Creating the Dataset for the Western Wind and Solar Integration Study”, Presented at the *7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms*, Madrid, Spain, 26-27 May, 2008, pp 7.

[4] C. W. Potter, H. A. Gil and J. McCaa, “Wind Power Data for Grid Integration Studies”, *Proc. 2007 IEEE Power Engineering Society General Meeting*, Tampa, FL, USA, Jun. 2007, Paper no. 07GM0808.

[5] P. Norgaard, H. Holttinen, “A Multi-turbine Power Curve Approach”. *Nordic Wind Power Conf.*, Chalmers University of Technology, Goteborg, Sweden, 1-2 Mar. 2004.

[6] G. McNerney, R. Richardson, “The Statistical Smoothing of Power Delivered to Utilities by Multiple Wind Turbines”, *IEEE Trans. Energy Convers.*, vol. 7, no. 4, pp. 644–647, 1992.

[7] RTS Task Force of the Application of Probability Methods Subcommittee, “The IEEE Reliability Test System-1996”**,** *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010-1020, Aug. 1999.

[8] S. Awerbuch, M. Berger, “Applying Portfolio Theory to EU Electricity Planning and Policy-Making”, *IEA/EET Working Paper*, Paris Feb. 2003, Report Number EET/2003/03.

[9] (Online resource) Western Wind Resource Dataset, NREL. Available: http://wind.nrel.gov/Web_nrel/

[10] C. Potter, “Wind Power Data for Grid Integration Studies”, *Panel on Wind Integration Studies*, IEEE PES General Meeting, Tampa, Florida, USA, June 2007.

[11] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*. New York McGraw-Hill, 1996, Chap. 2, pp. 20-25.

[12] C. Singh and Y. Kim, “An Efficient Technique for Reliability Analysis of Power Systems Including Time Dependent Sources,” *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 1090–1096, Aug. 1988.

[13] S. Fockens, J. M. Van Wijk, W. C. Turkenburg, and C. Singh, “Reliability Analysis of Generating Systems Including Intermittent Sources,” *Int. J. Elect. Power Energy Syst.*, vol. 14, no. 1, pp. 2–8, 1992.

[14] C.L.T. Borges, J.P. Galvão, “Evaluation of Wind Energy Sources Influence on Composite Generation and Transmission Systems Reliability”, *Proceedings of the Power Systems Computation Conference* , Liège, Belgium, Aug. 2005.

[15] R. Allan, R. Billinton, “Probabilistic Assessment of Power Systems”, *Proceedings of IEEE*, vol. 88, no. 2, Feb. 2000.

[16] B. Hahn, M. Durstewitz, K. Rohrig, “Reliability of Wind Turbines – Experiences of 15 years with 1500 WTs”, *Wind Energy, Proceedings of the Euromech Colloquium*, Springer-Verlag, Berlin Heidelberg, Feb. 2007, pp. 329-332.

VITA

Yannick Degeilh received his French “Diplôme d’ingénieur” degree in mechanical/electrical engineering at the Ecole Spéciale des Travaux Publics (ESTP), Paris, France, and his Master of Science in electrical engineering at Texas A&M University, College Station, Texas, USA, in August 2009. He is currently pursuing a Ph.D. degree in Electrical Engineering at the University of Illinois, Urbana Champaign, Illinois, USA, His research interests include power system reliability, integration of intermittent sources of energy, probability theory, statistics, economics and system optimization and control.

Mr. Degeilh may be reached at the Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Everitt Laboratory, MC-702, 1406 W. Green St. , Urbana, IL 61801-2918. His email is ydegeilh@gmail.com.