Deep Unsupervised Learning for Condition Monitoring and Prediction of High Dimensional Data with Application on Windfarm SCADA Data

Conference Paper

Author(s):
Mylonas, Charilaos; Abdallah, Imad; Chatzi, Eleni

Publication date:
2020

Permanent link:
https://doi.org/10.3929/ethz-b-000315814

Originally published in:
Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.1007/978-3-030-12075-7_21

Funding acknowledgement:
679843 - Smart Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines (EC)

Deep Unsupervised Learning For Condition Monitoring and Prediction of High Dimensional Data with Application on Windfarm SCADA Data

C. Mylonas[1], I. Abdallah[1], and E. N. Chatzi[1]
[1]Institute of Structural Engineering, ETH Zürich, Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland

ABSTRACT

In this work we are addressing the problem of statistical modeling of the joint distribution of data collected from wind turbines interacting due to collective effect of their placement in a wind-farm, the wind characteristics (speed/orientation) and the turbine control. Operating wind turbines extract energy from the wind and at the same time produce wakes on the down-wind turbines in a park, causing reduced power production and increased vibrations, potentially contributing in a detrimental manner to fatigue life. This work presents a Variational Auto-Encoder (VAE) Neural Network architecture capable of mapping the high dimensional correlated stochastic variables over the wind-farm, such as power production and wind speed, to a parametric probability distribution of much lower dimensionality. We demonstrate how a trained VAE can be used in order to quantify levels of statistical deviation on condition monitoring data. Moreover, we demonstrate how the VAE can be used for pre-training an inference model, capable of predicting the power production of the farm together with bounds on the uncertainty of the predictions.

Examples employing simulated wind-farm Supervisory Control And Data Acquisition (SCADA) data are presented. The simulated farm data are acquired from a Dynamic Wake Meandering (DWM) simulation of a small wind farm comprised of nine 5MW turbines in close spacing using OpenFAST.

The contribution of this work lies in the introduction of state-of-the-art machine learning techniques in the general context of condition monitoring and uncertainty quantification. We show how the high dimensional joint probability distribution of condition monitoring parameters can be analyzed by exploiting the underlying lower dimensional structure of the data imposed by the physics of the problem. The process of making use of the trained joint distribution for the purposes of inference under uncertainty and condition monitoring is clearly exposed.

Keywords: uncertainty quantification, deep learning, variational autoencoder, windfarm SCADA, DWM

INTRODUCTION

Wind turbines are subjected to stochastic loadings throughout their lifetime. Many wind turbines are reaching their end of design life which is 20-25 years. It is of interest to estimate the level of structural damage they have been subjected throughout their lifetime. The turbines can either be refurbished or decomissioned depending on the estimated level of structural deterioration. Due to the requirements for control in the level of electrical grid integration, but also for optimal power production, utility-scale wind turbines contain a supervisory control and data acquisition platform (SCADA) which typically registers 10 minute mean and standard deviation of several quantities of interest. The SCADA stream may contain useful information not only for estimating the current state of the structure based on past measurements, but also for detecting malfunctioning components based on instantaneous SCADA measurements. Both tasks rely on the statistical modelling of the SCADA stream, while at the same time modelling the environmental conditions. That is due to the fact that wind, together with pitch and yaw control, are the primary causes of mechanical straining for wind turbines.

Moreover, for the case of turbines positioned in a park, it is of importance to consider the potential interactions of them through wakes. Wakes in the context of this work, are the result of vortices produced on the tip of horizontal axis wind turbines (HAWT). In this work we have performed a medium-fidelity wake simulation for a windfarm containing 9 turbines with rated capacity of 5MW in a three-by-three arrangement. We used the so-called Dynamic Wake Meandering (**DWM**) simulation model [1] for the effect of wakes as it is implemented in the NTWC simulation suite [2]. The two main parameters characterizing wind statistics in this work are the 10 minute mean windspeed, and turbullence intensity. The mean windspeed distribution typically changes

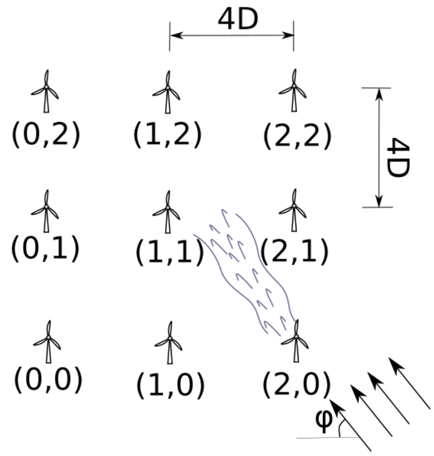

Fig. 1 The layout of the studied simulated farm

with height, however for the simulations that this work considers we have not statistically modelled this effect (typically refered to as *wind shear*).

DESCRIPTION OF THE METHOD

Simulated SCADA dataset

Figure 1 displays a sketch of the turbine layout on the considered windfarm. In total, 600 aero-servo-elastic simulations were run for randomly sampled windpseed and turbullence intensity, according to their joint distribution, as defined in IEC 64100 design standard [3] for a class C turbine. The raw simulated dataset consists of dynamic and operational measurements, such as time-domain tower and blade root moments and power production for the 9 wind turbines for 2000 seconds. The first 400 seconds in the time domain results are ignored in order for the farm to reach approximately stationary operational conditions. Consequently, two 600 second intervals are considered from each simulation, ammounting to 1200 stochastic inputs.

The aero-servo-elastic simulations were performed with NREL-OpenFAST using the simplified ElastoDyn module and the AeroDyn14 module for aerodynamics and the simulations of the downstream wakes, using the DWM simulator. The windfields used for the simulations are created with TurbSim, with the Kaimal turbullence model and using a different random seed for every simulation. Therefore the response of each turbine and the farm as a whole is fully stochastic. The DWM model can capture the meandering and expanding of wakes.

Variational Autoencoder

Relatively recently, in the concurrent works of [4] and [5], an efficient method for building probabilistic latent variable models was proposed. The model form of the so-called *Variational Autoencoder* (**VAE**) is shown in Figure 2.

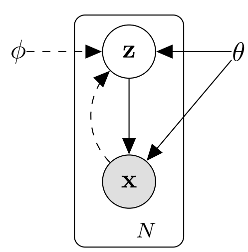

Fig. 2 The variational autoencoder as a probabilistic graphical model. Solid lines denote the generative model (decoder) $p_{\boldsymbol{\theta}}(\mathbf{z})p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$, dashed lines denote the variational approximation $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ to the posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$. The variational parameters $\boldsymbol{\phi}$ to be learned jointly with the generative model parameters $\boldsymbol{\theta}$.

The Variational Autoencoder **(VAE)** is a re-formulation of the autoencoder [6] where the encoder, refered to also as the *recognition model*, parametrizes a known probability distribution $q_{\boldsymbol{\phi}}(\mathbf{x}|\mathbf{z})$ over latent variables $\mathbf{z}$ which is an approximation to the true posterior distribution $p_{\boldsymbol{\phi}}(\mathbf{x}|\mathbf{z})$. The encoder and the decoder, are both implemented as deep neural networks.

The likelihood of the datapoints reads

$$\log_{p_{\boldsymbol{\theta}}}(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = \sum_{i=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) \tag{1}$$

For training the VAE, we optimize a lower bound on the likelihood of each datapoint that is given in Equation 2.

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}^{(i)}) = -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}^{(i)})}\Big[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\Big] \tag{2}$$

The first term of the right-hand-side is the *Kulback-Leibler* divergence between a prior over the latent variable $\mathbf{z}$ and the variational posterior $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$. The second term in the RHS is the expected likelihood of the data given a set of samples from the latent space. The quantity in Equation 2 is referred to as the *Evidence Lower Bound* (**ELBO**). We can chose $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ to be from a parametric family. In this work, we chose a diagonal Gaussian $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mu(\mathbf{x}), \sigma(\mathbf{x}))$. We can re-parametrize samples from a Gaussian distribution, as scaled and shifted samples from a standard Gaussian auxiliary variable $\epsilon \sim \mathcal{N}(0., \mathbf{I})$, and consequently we can train on the scaling and shifting of the distribution, in practice, deterministically. Note, however, that during the evaluation of the network sampling is performed. The computational graph with this re-parametrization is shown in 3. This re-parametrization trick was proposed independently by [4] and [5]. Without this re-parametrization, alternative sampling based estimators for the gradient could have been used for optimization of the hyper parameters. These sampling based estimates are far less efficient and are expected to have much higher variance, especially for cases of large and diverse datasets. This is the main trick that makes training of VAEs efficient and scalable to large datasets.

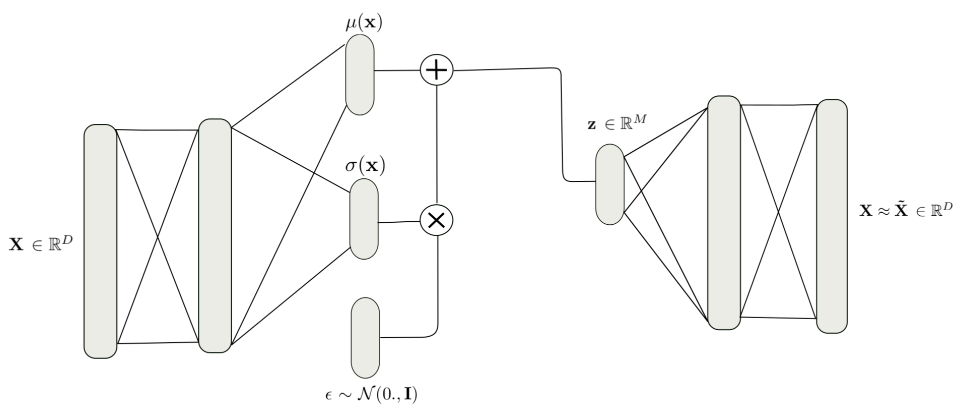

Fig. 3 The computational graph of a VAE with a Gaussian stochastic layer. In essense, we *compress* the raw high dimensional SCADA data $\mathbf{X} \in \mathbb{R}^D$ dimensional SCADA stochastic vectors, to a $\mathbf{z} \in \mathbb{R}^M$ dimensional stochastic vector. ($M << D$)

Sampling from the trained model

In the context of a VAE, the problem of approximating the high dimensional joint distribution of the 45 dimensional raw data vector, is cast as a problem of discovering a distribution over a lower dimensional random vector, that we have assumed is distributed according to a known prior distribution $q(\mathbf{z})$. For the simulated farm studied in this work, a stochastic latent vector of size 3 was found to be adequate to capture most of the variations in the data. The encoder is expected to exploit the correlations between the raw input variables, allowing for a lower dimensional representation. Samples from the training dataset, together with samples from a variational autoencoder trained on the simulated farm data are shown in 4 and 5. It is observed that the approximation is not as good in regions of lower probability mass. Nevertheless, for our purposes the VAE gives a good enough approximation, For clarity, the windspeed and turbullence intensity are shown separately for each turbine, whereas the autoencoder learns them jointly. Moreover, the angle-dependent effect of wakes makes the turbullence intensity higher on waked turbines and this cannot be seen in this figure.

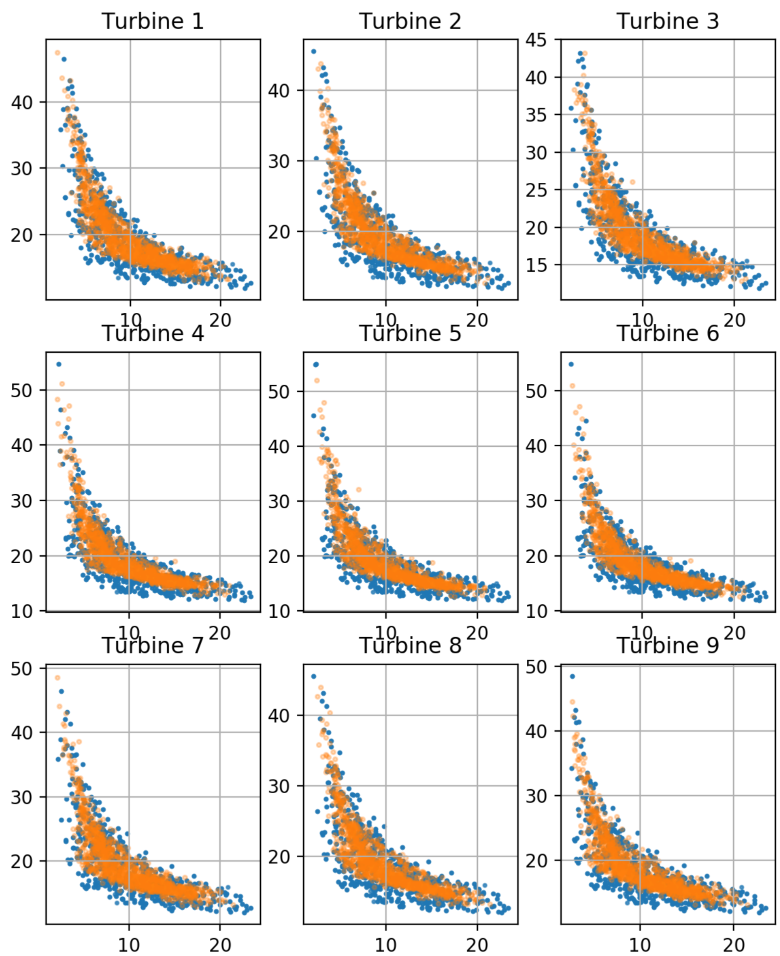

Fig. 4 The samples from the simulation are denoted with blue dots. With orange dots are the samples from the VAE. The autoencoder seems to perform well in capturing the windspeed and turbulence distribution of all the turbines.

For more intuition on the representation that the autoencoder learns, the latent space is presented in Figure 6. The first two columns are the sinus and cosinus of the mean angle of the wind in the farm. It is interesting to note, that the latent factor in column 4 seems to be directly correlated with mean windspeed. The mean inflow angle was kept as a deterministic variable in the latent space. We can easilly control the mean angle and make predictions from the VAE with bounds on the uncertainty, while sampling from the prior distribution $p(\mathbf{z})$.

Using the joint distribution for predictions

In Figure 7, samples from the autoencoder are drawn, conditioned on a windspeed range from the middle of the power curve up to the rated power ($[10m/s, 12m/s]$). This region of the power curve is expected to have the most pronounced wake effects, in terms of power production. That is mostly due to the fact that we are in the below rated regime, and the windspeed is high

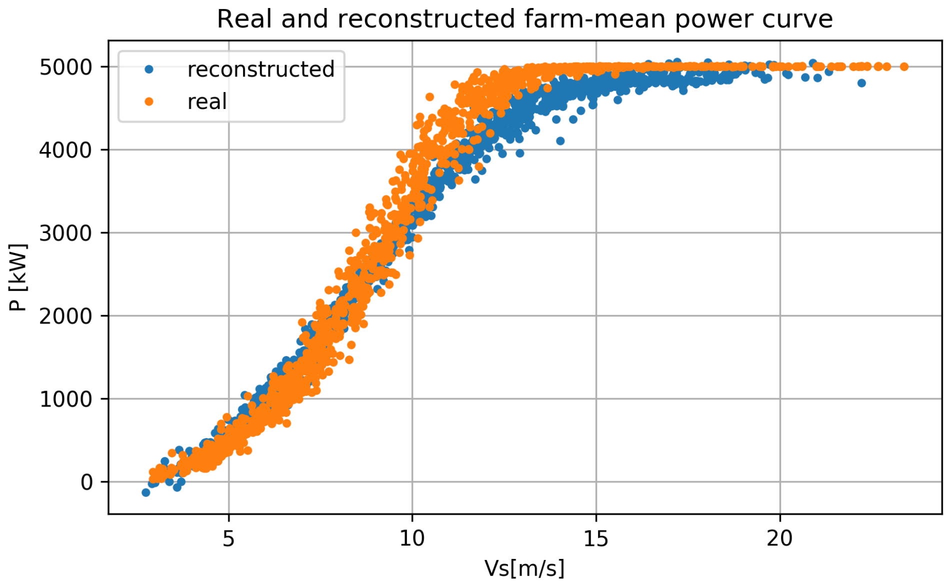

Fig. 5 Recostruction of mean power curve for 3x3 farm.

enough for the effect to propagate in a large distance. Note, that the dots are 5000 samples from the autoencoder for varying wind orientation, and unseen examples. It is observed that the deficits on windspeed present extrema at multiples of $45°$ as it was expected from the configuration of the windfarm. Moreover, at multiples of $90°$ from zero, the farm presents the highest wake deficits, whereas on mupltiples of $90°$ from $45°$ the peaks of wake deficit are lower. This is expected due to the geometry of the farm. However, the autoencoder has not learned a representation that corresponds to the symmetries we would expect, since there is no effect captured in angles $\pm 26.56°$ around every $45°$ spaced point. This is a subtle effect due to the alignment of the center turbine of one side of the farm, with the turbines on the edges of the opposing side. It is believed that this is mostly due to the relatively small number of simulations available for the problem at hand. Finally in Figure 8 the ratio of power production of turbine $T^{(2,2)}$ to the maximum power produced in the farm is given. The correct angles are identified as the peaks of wake deficit, and bounds on the uncertainty of the estimation can be obtained by the samples of the autoencoder. There are some outliers which can be treated either with more training or more input data.

Using the VAE for probabilistic condition monitoring

For a well trained variational autoencoder, where the latent variables have indeed converged to the assumed prior $p(\mathbf{z})$, given a raw SCADA measurement the encoder will produce a point in the latent space $\mathbf{z}$. Since we have endowed the latent space with a known probability distribution, we can compute the likelihood of the raw datapoint as the likelihood of the latent space. It has to be stressed, that this will be a good estimate only if the D_{KL} term in Equation 2 is very low. In our examples we didn't have any simulated *faults* and therefore no results are presented for that application. This may be a complementary approach to fault detection as proposed in [7].

CONCLUSIONS

In this work we have demonstrated how a deep variational autoencoder neural network can be used in order to yield interpretable insights on high dimensional monitoring data. In works to follow, investigations on the architecture of the network are going to be presented, as well as results from the application of the technique on real farm SCADA data.

Acknowledgement The authors would like to gratefully acknowledge the support of the European Research Council via the ERC Starting Grant WINDMIL (ERC-2015-StG #679843) on the topic of Smart Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines.

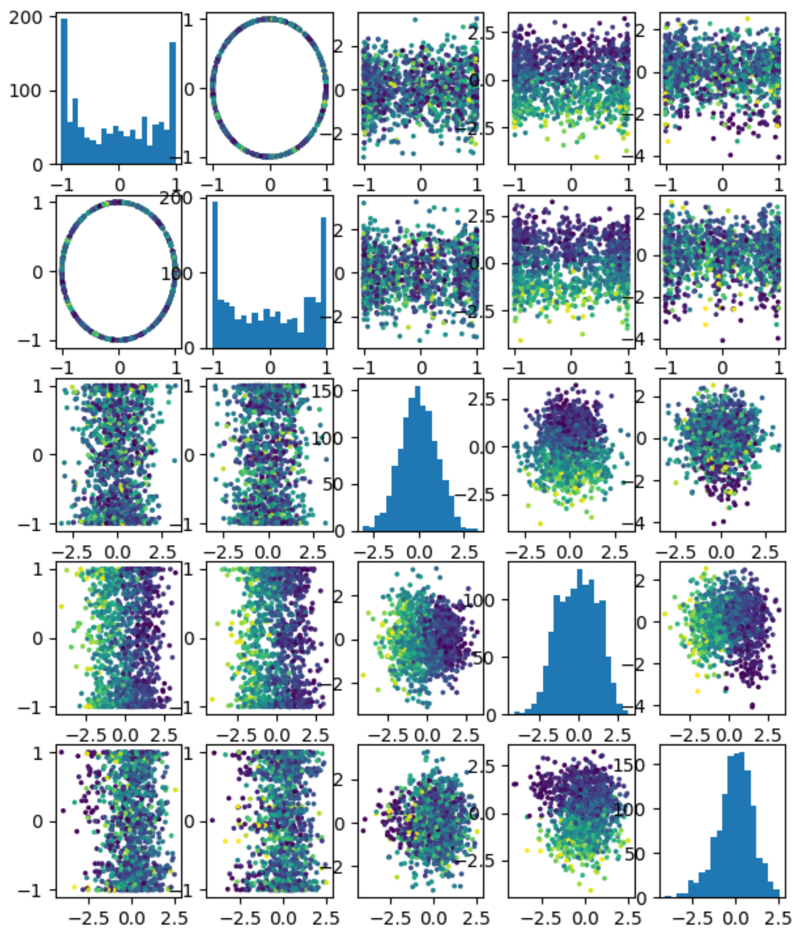

Fig. 6 The latent space variables, colored according to mean farm windspeed. Rows and columns 3 to 5 are the probabilistic latent factors, whereas the mean angle is kept deterministic for easier estimation of predictions w.r.t. angle.

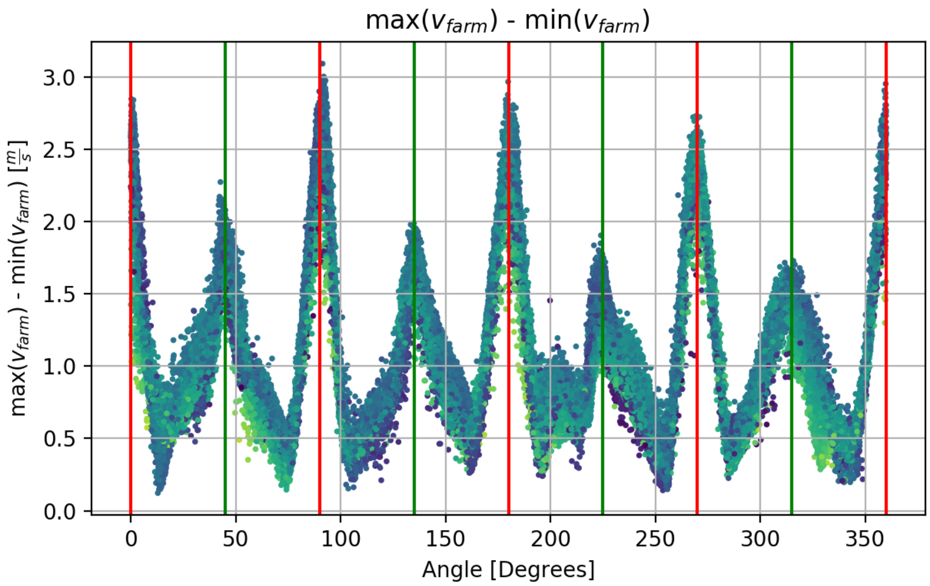

Fig. 7 Estimates of the maximum wake deficit estimated by the autoencoder. Red lines correspond to multiples of 90° for the angle of windspeed and green lines correspond to multiples of 90°but with a 45° shift.

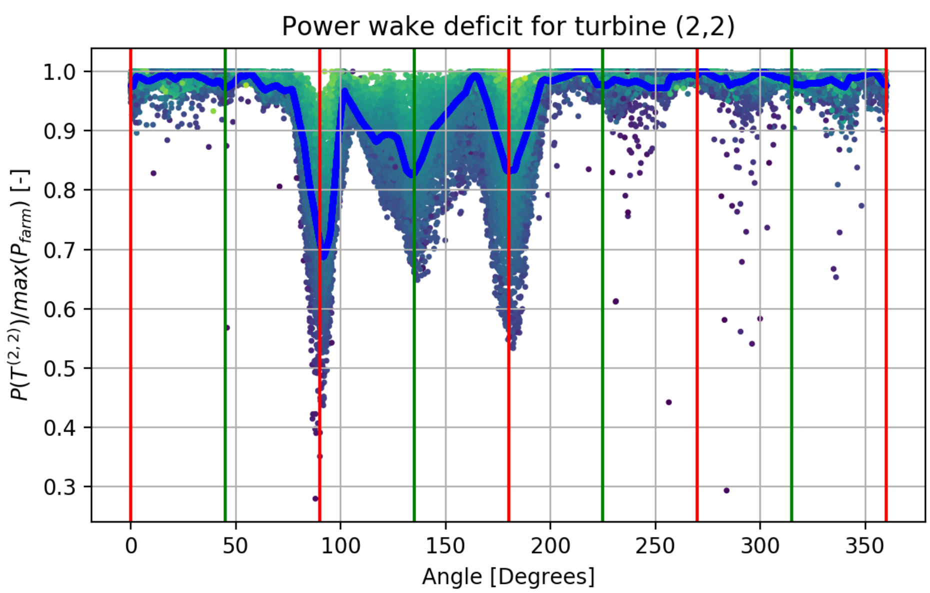

Fig. 8 VAE samples for the wake deficits for turbine (2,2) (see Figure 1 for numbering). The power deficit is defined as the ratio of the power of the turbine, to the maximum power produced from the farm for a given sample. The estimated mean power deficit is shown with the blue curve.

REFERENCES

[1] Larsen, G. C. , Madsen, H. A. , Bingöl, F. , Mann, J. , Ott, S. , Sørensen, J. N. , Okulov, V. , Troldborg, N. , Nielsen, M. , Thomsen, K. , et al. Dynamic wake meandering modeling. *Risø National Laboratory, Technical University of Denmark, Roskilde, Denmark, Risø*, 2007.

[2] Hao, Y. , Lackner, M. A. , Keck, R.-E. , Lee, S. , Churchfield, M. J. , and Moriarty, P. J. . Implementing the dynamic wake meandering model in the nwtc design codes. In *32nd ASME Wind Energy Symposium*, page 1089, 2014.

[3] Turbines, W. . Part 1: Design requirements. *International Standard IEC*, pages 61400–1, 2005.

[4] Kingma, D. P. and Welling, M. . Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[5] Rezende, D. J. , Mohamed, S. , and Wierstra, D. . Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[6] Hinton, G. E. and Salakhutdinov, R. R. . Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[7] Abdallah, I. , Dertimanis, V. , Mylonas, H. , Tatsis, K. , Chatzi, E. , Dervilis, N. , Worden, K. , and Maguire, E. . Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. In *Safety and Reliability–Safe Societies in a Changing World*, pages 3053–3061. CRC Press, 2018.