Wind Turbine Dynamics Identification Using Gaussian Process Machine Learning

Edward Hart, Bill Leithead, Julian Feuchtwang

Wind and Marine Energy Systems CDT
University of Strathclyde,
204 George Street,
Glasgow, G1 1XW

edward.hart@strath.ac.uk , w.leithead@strath.ac.uk

1. Introduction

Huge amounts of data are available to a wind turbine control system, unfortunately this data is usually composed of several additive components, plus noise.

This project seeks to apply machine learning techniques to extract these component parts, this will allow for improved control and a deeper understanding of wind turbine dynamics.

The first focus of this project is determining wind turbine C_Q tables and drivetrain losses from measuring generator speed and reaction torque. In **above rated** operation the wind turbine dynamics are approximately,

$$Q_{aer} - L(\omega_r) = NQgen + J\dot{\omega}_r$$

J is rotor inertia, N the gearbox ratio, L is the drivetrain losses, Q is torque and ω_r is rotor rotational speed. The losses function L is assumed to be linear.

The RHS of the above equation can be determined using measurements available to the control system. Linking the above dynamic equation to the aerodynamics of a wind turbine requires consideration of the effective wind speed, v. However, the only wind speed reading available to the controller will be the nacelle anemometer measurement, $\hat{V}$, which can be interpreted as a very noisy measurement of v,

$$\hat{V} = v + \zeta.$$

The noise here, ζ, is assumed to be iid Gaussian. Using separability of wind turbine dynamics, the LHS can be re-expressed as;

$$Q_{aer} - L(\omega_r) = \tfrac{1}{2}\rho AR\hat{V}^2 C_Q(\hat{\lambda},\hat{\beta}) - L(\hat{\omega}_r) + [g(\hat{V}) - g(v)].$$

Hat symbols, e.g. $\hat{\lambda}$, denote measured quantities and g is the component of the separated wind turbine dynamics which depends on the wind speed (see [1]); the function g is approximately linear and this along with ζ being treated as iid Gaussian implies that the term $[g(\hat{V}) - g(v)]$ will also be approximately iid Gaussian.

Scaling the measured values by $\frac{1}{2}\rho AR^3\hat{\omega}_r^2$ we obtain measured quantities, $\hat{H}$, of the form,

$$\hat{H} = \hat{\lambda}^{-2} C_Q(\hat{\lambda},\hat{\beta}) + L^*(\hat{\omega}_r^{-1}) + \eta(\hat{\omega}_r^{-2}).$$

This is the **above rated regression equation**. Note that since L is linear, L^* is a second order polynomial in $\hat{\omega}_r^{-1}$. The noise, η, is iid Gaussian but scaled this time by $\hat{\omega}_r^{-2}$.

Through a similar but simpler process (Taylor expansion about $\lambda = \lambda_{max}$) the **below rated regression equation** can be derived as,

$$\hat{G} = \hat{\lambda}_{max}^{-3} C_{P_{max}} + L^*(\hat{\omega}_r^{-1}) + \eta .$$

Note that in below rated operation the regression problem has become a polynomial regression.

2. Gaussian Process Machine Learning (GPML)

Gaussian process (GP) machine learning theory allows for regression of noisy function values inside a probabilistic framework by modelling a continuous function, f say, as a collection of multivariate Gaussian distributed random variables corresponding to the function's output at a collection of input variable values ($f(x_i)$ for i = 1,…,N).

A prior distribution over a set of noisy measurement values is formed via a maximum likelihood procedure. The most likely value of noise variance is determined as part of this procedure.

The function output values for a chosen set of input values can then be predicted using the Gaussian conditional distribution, with the mean function being interpreted as giving the most likely function values for the true function and two standard deviations at each point taken as 95% confidence intervals for the prediction. Figure 1 shows an example of a GP prediction and confidence intervals from noisy data.

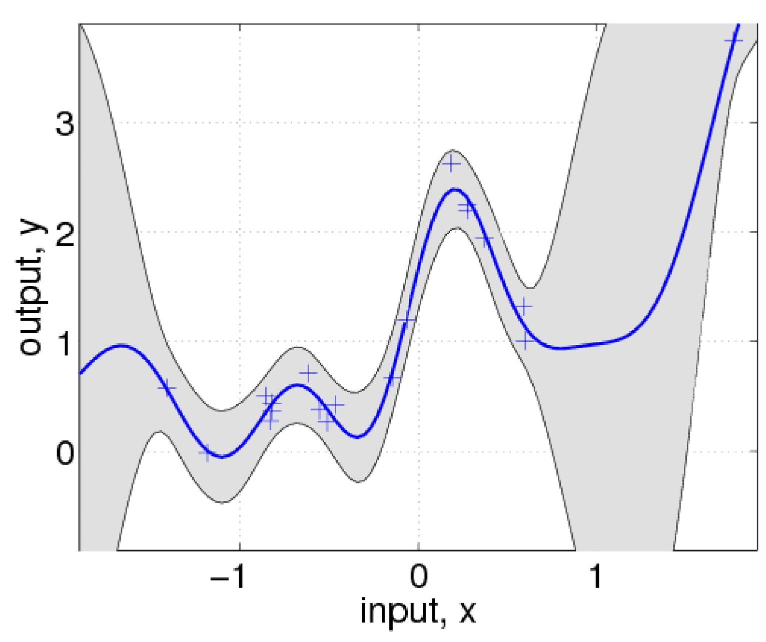

Figure 1. Mean function (blue) and 95% confidence intervals (grey) from the GP predictive distribution obtained by forming a prior and conditioning on the measured values shown by blue crosses.

3. Below Rated Dynamics

As outlined in Section 1, the below rated regression equation is in fact a polynomial regression problem. Attempts to perform regression in above rated conditions showed that the rotor speed varies too little for any changes in the losses values to be detected. Therefore, below rated operation and specifically the $C_{P_{max}}$ tracking region (where the rotor speeds vary across the whole range of values) was identified as being a good candidate for extracting losses information. From the constant polynomial term here can also be extracted the $C_{P_{max}}$ value of the turbine, since λ_{max} will be known for the turbine in question.

GP learning algorithms normally scale cubically in the number of measurements. Fast implementations were developed for the case of GP polynomial regression which instead scale linearly in the number of data points. These new algorithms allow for fast GP polynomial regression on very large data sets. For example a linear regression case on one million data points now takes only a few seconds to complete.

Figures 2 and 3 show GP and Least Squares (LS) polynomial regression predictions for $C_{P_{max}}$ and the linear losses function using simulated data from the Supergen Exemplar 5MW wind turbine model. The true $C_{P_{max}}$ value for the model is 0.4885.

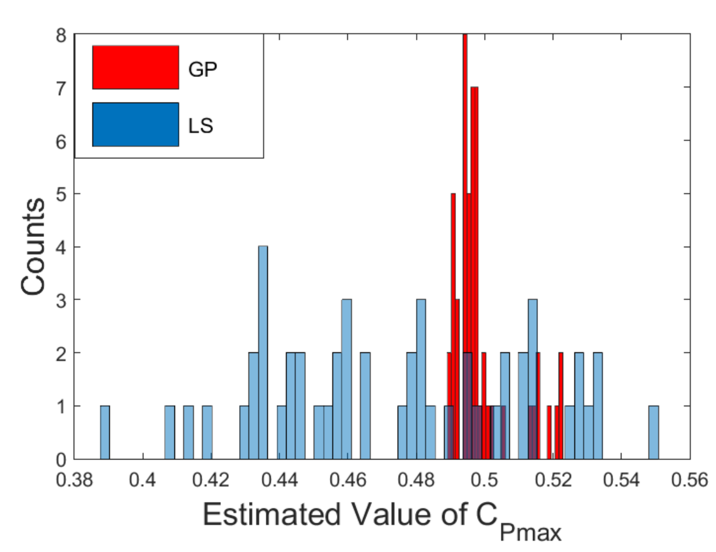

Figure 2. $C_{P_{max}}$ estimates from both GP and LS polynomial regression.

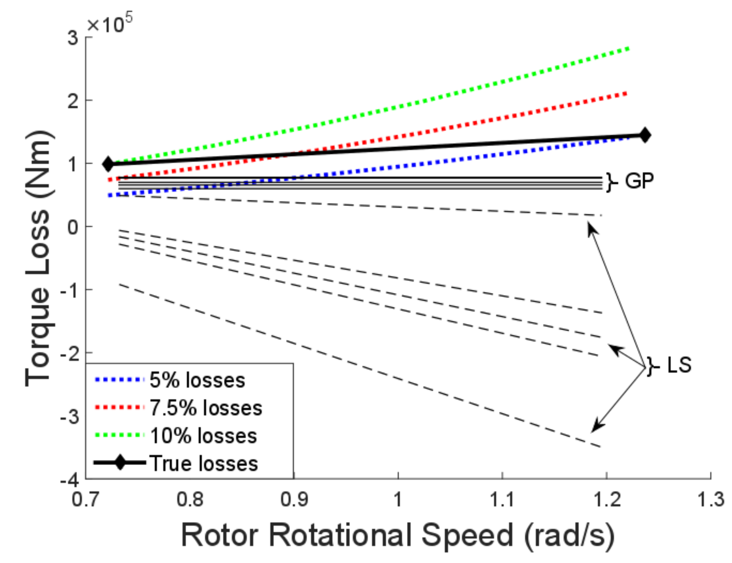

Figure 3. Losses function estimates from both GP and LS polynomial regression. The true linear losses function is also shown.

The GP predictions can be seen to have smaller error than the LS predictions. Furthermore, the GP predictions are tightly clustered, whereas the LS results show very large spreads. Offsets are present in the predictions which are due to the asymmetry of the function $\hat{\lambda}^{-3}$, the offset is generally very small (a few percent) for the $C_{P_{max}}$ prediction but noticeable for the losses prediction.

From an O&M perspective it is changes in these quantities which are important and hence, even with an offset, the tight clustering of the GP predictions means that changes in the losses function or $C_{P_{max}}$ value can be detected. This information could then potentially feed into maintenance scheduling or fault prediction efforts.

4. Batch Processing

The wind field which interacts with a wind turbine is turbulent and non-stationary, i.e. the characteristics of the wind field will change over fairly short timescales (10-30 mins). Changes in the wind field structure will result in different noise characteristics for our regression equations, hence, it is desirable to try and limit the time over which data is collected for a given regression implementation in order to avoid mixing noise types.

However, for the sake of accuracy and to have greater certainty in our results it is also desirable for our predictions to be based on as much data as possible. Within a LS framework these two requirements are irreconcilable. GP regression on the other hand can allow for both constraints as follows: the GP predictions from a given dataset give a probability distribution over the possible functions from which the data could have arisen. Therefore, given a dataset, D_1, the GP predictive distribution GP_1 can be formed from regression on D_1; when a different dataset D_2 becomes available we can similarly form GP_2 from regression on D_2. A refined predictive distribution can then be obtained by probabilistically combining GP_1 and GP_2, in the Bayesian sense, to give $GP_{1\&2}$. As new data becomes available $GP_{1\&2}$ can be further refined and so on. A schematic representation of this process is shown in Figure 4. Note also that this process of refining predictions removes the need for large amounts of data storage since after a prediction has been made on a given dataset, the data itself can be discarded.

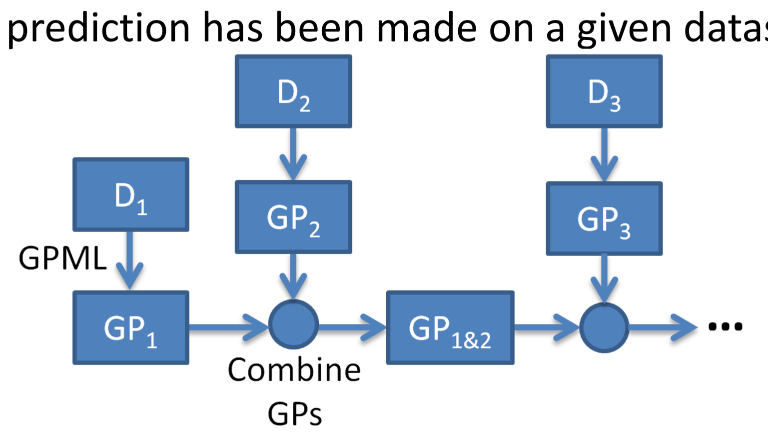

Figure 4. Schematic diagram of GP batch processing.

5. Current and Future Work

Current work is looking to optimise and test batch processing of below rated dynamics predictions. Once this stage has been completed, above rated dynamics identification will be investigated in a similar fashion and batch processing procedures developed for this case. Having developed these techniques to identify wind turbine dynamics from measured data, the final stage will be to understand how best to apply this new information to improve wind turbine control.

References

[1] W. Leithead et al, 'Global Gain-scheduling Control for Variable Speed Wind Turbines' Proceedings of the European Wind Energy Conference, 1999.

This work has been funded by the EPSRC, project reference number EP/G037728/1.