REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
09/27/2017	Final Technical Report	5/1/2015 - 4/30/2017

4. TITLE AND SUBTITLE

Data Analytics for Large Sensor Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER

N00014-15-1-2326

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Eric P. Smith
William N. Alexander
William J. Davenport
Leanna L. House
Scott C. Leman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Virginia Polytechnic Institute and State University
Office of Sponsored Programs
North End Center, 300 Turner Street NW, Suite 4200
Blacksburg, VA 24061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research

10. SPONSOR/MONITOR'S ACRONYM(S)

ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Eric Smith
					19b. TELEPHONE NUMBER *(Include area code)* 540-231-4801

Data Analytics for Large Sensor Systems

Eric Smith, Leanna House and Scotland Leman, Department of Statistics
Nathan Alexander and William Devenport, Department of Aerospace and Ocean Engineering,
Virginia Tech, Blacksburg VA 24061.

Final Report for grant N00014-15-1-2326

Abstract and summary

Modern wind tunnel testing involves multiple diverse sensor systems with anywhere from hundreds to thousands of individual sensors. These large-scale tests can be quite expensive and are usually time-sensitive. As such, any delays due to faulty instrumentation can have serious consequences. Equally serious is the possibility of discovering a sensor failure after the test has been completed, since time and effort will have been spent collecting what amounts to noise. Although sometimes correctable, the time involved in doing so distracts the experimenter from achieving the experimental goals.

As such, any large scale sensor system needs methods to ensure that all the individual sensors are working as intended. Most commercial sensor systems contain rudimentary error detection for sensors within a given system, but these methods typically have no way of incorporating information about the ambient conditions under which they were run, or, more importantly, information from the output of other systems which are used in conjunction. This is a significant problem, as most large sensor systems in wind tunnel tests are made of smaller, unique sensor subsets. By combining the information from diverse sensor systems into a global error detection process, we can measure the extent to which sensors across systems are correlated and use that correlation information to produce more powerful predictions and error detection capabilities.

This report summarizes approaches that have been developed to evaluate sensor systems both for a single type of sensor as well as for systems that include multiple sensor types. A MATLAB macro is described to analyze arrays of data arising from a study utilizing a single sensor type. An approach to analysis of sensor systems using multiple types is described based on a Gaussian process model. Examples are provided to illustrate applications of the methodologies.

Objectives

- Develop statistical measures to rapidly identify suspect sensors or measurement system problems, and identify datasets containing physically important information.
- Evaluate the approach using large microphone array data sets of both aerodynamic and acoustic sources, with varying levels of imperfection, from diverse experimental test programs in the Virginia Tech Stability Wind Tunnel.
- Automate approach to integrate into "live" experiments.
- Generate controlled "faulty" data for inspection of qualities which can be used as identifiers of signal faults and estimate correct detection rates.
- To demonstrate the need for a global error detection scheme that uses a holistic view of the experiment.
- To develop a statistically rigorous foundation for this detection scheme that requires no knowledge of the facility, test articles, or problem physics.
- To demonstrate this method on a large, diverse set of wind tunnel data collected during a Virginia Tech Stability Wind Tunnel experimental campaign.

Approach

Modern wind tunnel testing involves multiple diverse sensor systems with anywhere from hundreds to thousands of individual sensors. These large-scale tests can be quite expensive and are usually time-sensitive. As such, any delays due to faulty instrumentation can have serious consequences. Equally serious is the possibility of discovering a sensor failure after the test has been completed, since time and effort will have been spent collecting what amounts to noise. Although sometimes correctable, the time involved in doing so distracts the experimenter from achieving the experimental goals.

As such, any large-scale sensor system needs methods to ensure that all the individual sensors are working as intended. Most commercial sensor systems contain rudimentary error detection for sensors within a given system, but these methods typically have no way of incorporating information about the ambient conditions under which they were run, or, more importantly, information from the output of other systems which are used in conjunction. This is a significant problem, as most large sensor systems in wind tunnel tests are made of smaller, unique sensor subsets. By combining the information from diverse sensor systems into a global error detection process, we can measure the extent to which sensors across systems are correlated and use that correlation information to produce more powerful predictions and error detection capabilities.

The analytical tools described in this report aim to evaluate the issue of sensor malfunction by automating much of the error identification process. The structure of the tool is described and evaluated using

data from microphone arrays. Although the methods were developed specifically for microphone arrays, the program and tools can be useful in other situations where rapid assessment of sensor quality is required. This report first briefly discusses microphone arrays, as these are the motivation for the work. Then the statistical and graphical methods and their implementation are discussed. Examples using data from a wind tunnel experiment is given to illustrate the package and more general methodology. Finally, some extensions are described.

The Virginia Tech Wind Tunnel and Experiments

The Stability Wind Tunnel, shown in Figure 1, is a continuous, single return, subsonic wind tunnel with removable 7.3-m long square test sections with 1.83-m long edges. The facility has a conventional hard-walled test section and a Kevlar-walled test section ideal for aeroacoustic measurements. The data used in this report were recorded for experiments using the Kevlar-walled test section, which has a well-documented performance (Devenport et al., 2013). The tunnel can reach flow speeds up to 80 m/s for an empty test section. Each side of the test section has an anechoic chamber covered in foam wedges designed to eliminate reflections above 190 Hz. The anechoic chambers are used for measuring far-field sound, as pictured in Figure 2.

For the first application, data are from part of a study that focuses on the acoustic effects of a discontinuity or step in an otherwise smooth surface. For example, the surface of a ship may be mostly smooth except for discontinuities in the hull where plates are joined. Such steps may be acoustically loud and interest lies in the study of the effect of step size on acoustic characteristics of flow (Awasthi et al., 2014) and the design of efficient low-noise vehicles.

In the overall experiment step sizes of 3.6, 14.6 and 58.3 mm are used with flow speeds of 30 and 60 m/s. A variety of measurements are made including the mean wall Cp, mean velocity profiles, 3-D turbulence profiles and wall pressure spectra. Measurements are made using 31 wall-mounted microphones placed forward and aft of a sharp step at a particular location (Figure 3).

For the second application, a 251-microphone array (Alexander et al., 2016) was split into two sub-arrays, and each was placed in an anechoic chamber at 45 degrees relative to the free-stream direction, as seen in Figure 4. The experimental photos describing this are shown in Figure 5. The array was sampled at 51200 Hz for 32 seconds with a custom DAQ system. The system output is a time series signal for every microphone. The acquisition data can create acoustic beamform maps based on these signals, which allow the experimenter to visualize sound sources on the test article (see Figure 6 for an example). For the Gaussian Process (GP) analysis, however, only the root-mean-square of each microphone signal is used as an input, in order to give it the same dimensionality as the metadata. This is equivalent to stating that the Overall Sound Pressure level (OASPL) is used as an input.

The experiment made use of all the available instruments in the facility in order to build a comprehensive data set with which to test the GP methods. It should be noted that, based on the experimental objectives, not all sensor systems were used simultaneously for each measurement run. Following is a description of each sensing method.

In addition to the phased array, acoustic measurements were taken with four standalone Bruel and Kjaer microphones in the port chamber. Data from these microphones was recorded with a Bruel & Kjaer LAN-XI 3050-A DAQ sampling at 65536 Hz for 32 seconds (separate from the phased array acquisition system). The microphones were placed in the port chamber in front of the sub-array, also shown in Figures 4 and 5. Each standalone microphone has a corresponding time series that would typically be associated with it. The data from each microphone would usually be presented as a Sound Pressure Level spectrum, which will be shown in the final paper. As with the phased array, only the root-mean-square (equivalently, OASPL), of each microphone is used as an input.

Hot-Wire Anemometry (HWA) techniques were used to measure local flow velocities. Briefly, the concept of HWA is that convection effects of the flow over a wire change the wire resistance, and these changes can be converted to local velocity fluctuations. Further details can be found in the literature describing the development of this measurement technique (Kegerise and Spina, 2000). For our purposes, it is sufficient to say that the HWA measurements give a voltage time series at each point that can be converted to flow velocity, although the analysis will differ based on the specific measurements.

Measurements of the flow around a 10-bladed rotor ingesting the wake of an upstream cylinder were made with pairs of on-blade probes on adjacent blades (Alexander et al., 2014) sampled at 51200 Hz for 30 seconds, and measurements downstream of a two-bladed marine rotor were made with a conventional single-wire probe sampled at 50000Hz for 30 seconds. Figure 7 shows the experimental set-up for the on-blade measurements, and a typical processed result is included in Figure 8. Single-wire measurements were taken at single points downstream of a two-bladed rotor, as well as in a radial profile. The probe is shown in Figure 9a, and a typical processed result is shown in Figure 10.

For some measurements, a cylinder was placed upstream of the rotor plane to create turbulence for rotor ingestion. The cylinder had pressure ports in a spiral pattern along the circumference, which allowed experimenters to measure a (scalar) mean pressure at each surface point. Typically, the data would be shown as a plot of the pressure coefficient as a function of angular position, which will be included in Crandell et al, (2017, to appear).

Measurements with a drag rake were also performed, consisting of a span-wise array of Pitot and Pitot-static probes that can be traversed in the vertical direction. This device is pictured in Figure 11. Each probe records the mean pressure at a point.

Additional pressure measurements were made with a Pitot probe traversed downstream of a two-bladed marine rotor (Figure 9b). The data has not been processed, but a typical output will be included in the Crandell et al, (2017, to appear), as with the rest of the sensors.

Every quality wind tunnel test will require measurement of the reference conditions, such as the flow speed, temperature, and pressure, in addition to the measurements unique to achieving the experimental objectives (i.e. microphones to measure the acoustics of a rotor ingesting a turbulent wake). In addition to recording the reference conditions, the authors were also interested in recording the metadata that is usually left behind. The term "metadata" is used broadly here to describe data that gives information (indirectly or directly) about the test, such as the reference conditions. Examples of this metadata include the tunnel motor armature voltage, armature current, bearing temperatures, and test section humidity. The idea is that the breadth of data of acquired, along with the Gaussian Process framework, will provide experimenters with a holistic view of the measurement. This idea is explored further in the following section.

Common failures with microphones include complete failure in which the microphone turns off resulting in a flat signal with low variability, drift of the signal and groups of signals that are different from regular signals. In addition, some microphones exhibit over-range and clipping. Over-range is defined here as having values that exceed normal ranges for the microphone. Clipping occurs when a value, usually at the upper-most or lower-most value occurs multiple times.

Statistical Methods

Single type of sensor

The data that are used for a system of single sensor types is viewed as consisting of measurements on K microphones or sensors. Each microphone is assumed to have the same starting and ending value and generate the same number of measurements. The measurements are indexed by a time sequence. Thus, we take xij to be a measurement from microphone i, $i = 1, 2, ..., K$ at time j, $j = 1, 2, ..., T$.

A standard approach for evaluating unusual observations in statistical analysis makes use of means and standard deviations of data. However, this approach is limited because odd values can increase standard deviations and biased means that reduces sensitivity. In addition, the approach requires a large number of sensors relative to the number of observations and the use may be limited due to correlations between measurements on the same sensor. For microphone data it is not uncommon to have as many as 512,000 data values for a single microphone over a 30 second interval. Thus, we are dealing with a relatively small number of long series. In this work, the preference is to make use of a nonparametric approach that is less sensitive to extreme values and reduces correlation over time.

The first step is to divide the series for a sensor into M segments of equal or roughly equal length. If L is the length of the segment, then $M=T/L$. We then index the data values for microphone i x_{ij} as x_{ilm} with $i= 1, 2, ...L$ and $m = 1, 2..., M$. The second step is to define a statistic that summarizes the segment. For this

purpose we use a range that contains a specified percentage of observations, i.e. a range percentile. The range percentile is based on the quantiles from the data in each segment. Specifically, define the range percentile level α, that determines the upper and lower quantiles and hence the quantile range within each segment. For example, for an $\alpha = 90\%$ range percentile, use a quantile interval based the 5th and 95th quantiles of the x_{ilm} for each microphone i and segment m. In applications, the $\alpha = 90\%$ quantile interval is used however in implementation other values may be used. A larger value of α would result in identifying more individual observations while a smaller value would focus more on groups of observations. This value of the percentile range is referred to as $IQR\alpha$. The M segments then produce M values for each microphone, i.e. $IQR\alpha_{im}$. We then calculate the median and pseudo standard deviation (PSD) of these values for each microphone. Extreme values are defined as values that are more extreme than the median plus or minus z times the PSD. The values of z is user defined and is associated with a desired level of certainty in correctly identifying sensor oddities. In addition to these calculations, checks are also made to identify over-range, clipping and microphones that are off. Over-range is evaluated by identifying if observations exceed specified limits, while clipping is determined by the number times the minimum or maximum limits are duplicated. Microphones are determined to be off if a large number of values are below some threshold that is user determined.

Multiple types of sensors

The main predictive engine for this application is the Gaussian process (GP). This is a flexible, non-parametric data fitting and smoothing technique. GPs have seen wide use, originally in mining and geo-spatial analysis (Matheron, 1962), and later in modeling output from computationally demanding computer experiments (Kennedy and O'Hagan, 2001, Grammacy et al., 2015, Higdon et al, 2004). Unlike traditional statistical methods such as least-squares based regression, GPs make very loose assumptions about the data under study. This allows them to infer a great deal of the structure in the data from the data itself, making them well suited to studying complex flow phenomena typically seen in wind tunnel tests.

Central to the idea of GPs is a notion of distance between sensors. When considering sensors within a given system, it is natural to consider the physical separation of the sensors, and to make the assumption that the extent to which the output of two sensors will correlate is a decreasing function of their distance (put another way, that the correlation decreases with increasing distance). While physical separation is a reasonable proxy for correlation within a sensor system, this may not always be the case, particularly for correlation across independent sensor systems. For example, suppose one wishes to measure the far-field sound generated by an ideal dipole source. It is conceivable that a microphone on the null axis of the dipole would record no sound, while a nearby, adjacent microphone may be directly in the sound field. For this case, we need a different notion of correlation. We introduce such a metric by considering the output from

any two sensors across all runs during which they were both active and computing the average $L1$ norm between those observations

$$d(y_1, y_2) = \frac{1}{p}\sum_{i=1}^{p} | y_{1,i} - y_{2,i} |$$

This equation is called the *response distance*. This metric has many advantages as a measure of similarity between sensors. First, by appropriately scaling the output of the sensors, this metric becomes agnostic to the particular system from which the outputs come. This allows comparison between arbitrary systems. Second, the metric makes no assumptions about the dynamics of the phenomenon being measured. Due to the complex and highly non-linear nature of the flow in a wind tunnel test, most simplifying assumptions of the kind usually seen in statistical models can easily break down. This metric learns about correlation entirely from the data. Third, this metric learns correlation not only between sensors in different systems, but also between sensors in the same system. Our method combines this information with information derived from spatial correlation, weighs the two distance metrics appropriately, and combines them to produce a unified correlation measure. This synthesis allows us to estimate correlation and produce predictions after a single run in the wind tunnel experiment.

We will first describe the general framework for Gaussian process regression in the context of a single sensor system. We will then generalize to simultaneous prediction for several systems, and finally extend this framework to prediction across several runs of a unified experimental apparatus. The basic idea of GP regression is to infer the value of a response at a given point through a *weighted average* of the observed responses at all points in a training data set. The extent to which a point in the data set contributes to this average is a function of its distance, physical or otherwise, from the prediction location. The first step in constructing a GP model is to define the precise relationship between distance and correlation. This relationship is encoded in what is known as a *kernel function.*

The most popular GP kernel is the anisotropic Gaussian kernel. This kernel is appropriate for measuring correlation based on physical distance. If we let xi and xj be two length p vectors encoding the positions of two measurements, then this kernel has the form:

$$C(x_i, x_j) = \exp\left(-\sum_{k=1}^{p} \frac{(x_{i,k} - x_{j,k})^2}{\theta_k^2} \right)$$

The θ parameters are known as length scales, and they control the relative importance of distance along the different axes. This allows for the possibility that correlation does not with scale with distance in an isotropic way. Note that without the θ's the argument to the exponential function is the negative of the squared Euclidean distance.

Given n observations, then we can define an nxn matrix, whose $\{i,j\}^{th}$ entry is $C(x_i,x_j)$. By combining this with a few more parameters, we then construct the covariance function. As before, there are many possible choices, but we adopt the following flexible and lightweight parametrization:

$$\mathbf{K}_n = \sigma^2(\mathbf{C_n} + \varphi^2\mathbf{I}_n),$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The parameter φ is known as the *nugget effect*, which augments the diagonal of the C matrix by φ^2, increasing the covariance between an observation and itself. The nugget effect can be thought of as measurement error, and it allows for the possibility that the responses were measured imprecisely. The final parameter σ^2 is known as the *partial sill*, and it controls the overall scale of the correlation. Note that all the above parameters are required to be positive.

Given this covariance matrix, we have the following model for the data generating process. Let $\mathbf{y}$ be the vector of the measured response values. Then:

$$f(y) = 2\pi^{-\frac{n}{2}} |\mathbf{K}_n|^{-1} \exp\left(-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{K}_n^{-1}\mathbf{y}\right)$$

In other words, the vector $\mathbf{y}$ follows a multivariate Gaussian distribution with mean 0 and covariance $\mathbf{K}_n$. In this formulation, the response vector $\mathbf{y}$ is assumed to be a scalar at every spatial location x. It is a substantial approximation to reduce a measured time-series to a scalar, but we will show that even with this highly limiting assumption we are able to find anomalous signals. We will also argue that our method can handle multivariate output, and plan to expand this capability in future work.

There are many methods for choosing appropriate numerical values for the parameters listed above. In this report, we use maximum likelihood estimation. This particular parametrization allows us to estimate σ^2 in closed form, but the nugget and length scale parameters require numerical methods for their estimation. Finally, we can incorporate metadata (for example, the freestream reference conditions) into the model via a linear form for the mean of the Gaussian model. If we let $\mathbf{D}$ be a matrix of metadata measured along with the

response values at every location, and let β be a vector of regression parameters, then we have the following final model for $\mathbf{y}$:

$$f(y) = 2\pi^{-\frac{n}{2}} |\mathbf{K}_n|^{-1} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{D}\beta)^T \mathbf{K}_n^{-1}(\mathbf{y}-\mathbf{D}\beta)\right)$$

As with σ^2, estimates for β are available in closed form.

Once we estimate the model parameters, predictive means and variances are given in closed form. In particular, the predictive distribution for some new response y^* at location x^* with metadata $\mathbf{d}^*$ is a Gaussian distribution with mean $E(y^*|\mathbf{y})$ and variance $\mathrm{Var}(y^*|\mathbf{y})$ given by:

$$E(y^*|\mathbf{y}) = (\mathbf{d}^*)^T\hat{\beta} + \mathbf{C}(y^*,\mathbf{y})K_n^{-1}(\mathbf{y}-\mathbf{D}\hat{\beta}),$$
$$\text{and } Var(y^*|\mathbf{y}) = \sigma^2(1+\hat{\phi}^2 - \mathbf{C}(y^*,\mathbf{y})K_n^{-1}\mathbf{C}(y^*,\mathbf{y}))$$

where $\mathbf{C}(y^*, \mathbf{y})$ is the row vector of covariances between the prediction location x^* and all the training locations, with $\mathbf{C}(\mathbf{y}, y^*)$ being its transpose. While these predictive equations are valid for any location x^*, we will mainly be interested in these predictions at the same locations as the sensors themselves. In that case, the joint distribution of the predictive vector is a multivariate Gaussian with

$$E(\hat{y}|\mathbf{y}) = \mathbf{D}\hat{\beta} + \mathbf{C}_n\mathbf{K}_n^{-1}(y-\mathbf{D}\hat{\beta})$$
$$Var(\hat{y}|\mathbf{y}) = \hat{\sigma}^2\left[(1+\hat{\phi}^2)\mathbf{I}_n - \mathbf{C}_n^T\mathbf{K}_n^{-1}\mathbf{C}_n\right]$$
$$+\hat{\sigma}^2\left[\mathbf{X}_n^T(\mathbf{X}_n^T\mathbf{K}_n^{-1}\mathbf{X}_n)^{-1}\mathbf{X}_n - \mathbf{C}_n^T\mathbf{K}_n^{-1}\mathbf{X}_n(\mathbf{X}_n^T\mathbf{K}_n^{-1}\mathbf{X}_n)^{-1}\mathbf{X}_n^T\mathbf{K}_n^{-1}\mathbf{C}_n\right]$$

The methods described above work well when considering a single sensor system, but require modification for monitoring multiple systems run concurrently. In the case of a typical wind tunnel test, for instance, sensors within an array are on the order of centimeters apart, but separate systems can be separated by several meters. A GP model trained to infer correlation on a scale of centimeters will zero out on the scale of meters (imagine e^{-100}). Thus, the model will infer zero spatial correlation between systems and will do no better than separate GPs trained on separate systems. While that would be a valid approach for detecting local errors within a single system, it is not aligned with the stated objective of developing a framework for a global, holistic detection scheme. In order to unify disparate systems, we need a different notion of distance.

The main idea here is to rethink the space in which the sensor exists. The previous formulation considers a sensor and its corresponding measurement as a point in three-dimensional space. Instead, we will think of a sensor's position as a vector in a space whose dimensionality is equal to the number of

runs during which the sensor produced a measurement. The coordinates of that vector are the response values measured for each run. Then, the distance between two sensors can be computed by considering the values of this response vector.

More formally, consider two sensors i and j, coming from two arbitrary systems. Let $y_{i,g}$ be the *appropriately scaled* univariate response from sensor i during run g and $\mathbf{y}_i$ be the *scaled* vector of such responses. Finally, let $m_{i,j}$ be the number of runs during which both sensors i and j produced a measurement. Then, the response distance between these two sensors is given by:

$$d^{Y}(y_i, y_j) = \frac{1}{m_{ij}} \sum_{g} | y_{i,g} - y_{j,g} |$$

where the sum is taken over all runs during which both sensors were measured. Plugging this into the same kernel function as above gives the response covariance:

$$\mathbf{C}^{Y}(\mathbf{y}_i, \mathbf{y}_j) = e^{-d^{Y}(y_i, y_j)}$$

The response distance, as formulated, can be thought of as the average $L1$ distance between the two response vectors. We chose this norm because it is robust to the presence of outliers. While other choices are possible. The $L1$ norm has proven sufficient for this study.

As well as giving us a way forward with cross system GP modeling, the numerical values in the response covariance matrix are of use in their own right. The matrix can be produced very early in an experimental process, as soon as after a single trial, and gives the experimentalist a quick way to assess the overall performance of their apparatus. In particular, sensors that are offline have a very clear visual signature, which we will illustrate in the example below.

Scaling

Since different sensor systems can measure different physical quantities on different numerical scales, it is essential to scale the output in some way in order to make a sensible comparison. In addition to ensuring that all quantities are unitless and in the same order of magnitude, scaling accentuates some features of the data while hiding others. We consider two scaling regimes, called system scaling and sensor scaling.

Sensor scaling is applied to the output from a single sensor considered across all relevant runs. That is, the output of any given scaled sensor from all its runs has a mean of zero and a standard deviation of one. System scaling is applied to the output of an entire sensor system considered for a single run. Under this

regime, for instance, the output from a microphone array during a given run has mean zero and standard deviation one, as does the output of a pressure rake during a given run.

As well as using information from one system to learn about another, we also want to use information between separate runs of the experimental apparatus. This is also essential for learning the effects that metadata have on sensor outputs. Since the metadata we consider are typically fixed for an entire run, we must consider separate runs in order to gain information about how sensor outputs and metadata co-vary.

Our algorithm handles simultaneous prediction of responses from multiple runs in a natural way. If we let $\mathbf{y}_g$ and $\mathbf{y}_h$ be the response vectors from two runs g and h, then we can let $\mathbf{y}^T = [\mathbf{y}_g^R, \mathbf{y}_h^T]$ and predict the combined vector with a single GP. This combined vector requires a similarly augmented covariance matrix.

One natural choice is to take the covariance matrix from a single run, $\mathbf{K}_n$, and repeat it in block structure:

$$\mathbf{K}_{2n} = \begin{bmatrix} \mathbf{C}_n & \mathbf{C}_n \\ \mathbf{C}_n & \mathbf{C}_n \end{bmatrix} + \phi^2 \mathbf{I}_{2n}$$

We call this a *full block covariance structure*. This is possible because neither the relative physical locations of the sensors with a system change between runs, nor does the response covariance matrix, since it is run-averaged. Outputs from an arbitrary number of trial runs can be stacked in this way.

At first glance, this may appear to be computationally infeasible for even a moderate number of runs. In our experiments we used on the order of 500 different sensors, which is a reasonable amount of data on which to fit a GP. However, GPs scale poorly with larger data sets due the necessity of inverting the $n \times n$ covariance matrix, an $O(n^3)$ operation. The whole experiment consisted of about 400 runs, and to consider all of the output simultaneously would require a response vector of length 200,000. This is far too large for GP methods.

However, due to the highly replicated nature of the cross-run covariance matrix, we can use ideas developed in (Binois, Grammacy and Ludkovski, 2016) to quickly invert this matrix. Rather than inverting the entire matrix, we need only invert the submatrix $\mathbf{K}n$ and use that to reconstruct the entire inverse using relatively quick operations. If we wish to consider m runs in a single GP, then we can let $\mathbf{U} = \mathbf{1}m \otimes \mathbf{I}n$, where $\otimes$ denotes the Kronecker product and $\mathbf{1}m$ a column vector of 1's. Then,

$$\mathbf{K}_{mn} = \phi^2 \mathbf{I}_{mn} + \mathbf{U}\mathbf{C}_n\mathbf{U}^T$$

$$\mathbf{K}_{mn}^{-1} = \phi^{-2}\mathbf{I}_{mn} + \phi^{-4}\mathbf{U}(\mathbf{C}_n^{-1} + \frac{m}{\phi^{-2}}\mathbf{I}_n)\mathbf{U}^T$$

where the inverse is computed using the well-known Woodbury matrix identity (Woodbury, 1950). Note that the inversion steps only involve matrices of size n rather than mn.

Another covariance parameterization we consider has the following form:

$$\mathbf{K}_{2n} = \begin{bmatrix} \mathbf{C}_n & 0 \\ 0 & \mathbf{C}_n \end{bmatrix} + \phi^2 \mathbf{I}_{2n}$$

We call this model a *block diagonal covariance structure.* Similarly easy to compute formulas for the inverse of the covariance matrix will be included in the appendix of Crandell et al, (2017, to appear). In this model, data from one run is not used to estimate correlations from another run. This model is appropriate when, for instance, an experimentalist expects that changes in the experimental conditions make cross-run correlation non-meaningful. We will demonstrate the different kinds of predictions and different kinds of anomalies these covariance structures are able to detect.

Work Completed

Two applications are described below. The first is based on a GUI developed using MATLAB for rapid identification of data collected from arrays of a single types of sensor, in this case microphones. The second application uses a GP approach based on a multiple-sensor system. Data for assessment of the methodology was obtained from the Virginia Tech Wind Tunnel Facility.

Single set of sensors

The statistical methods incorporated into a MATLAB graphical user interface (GUI) are now illustrated. Additional examples are given in Vasta et al. (2017). The interface has two sections or tabs. Analytical methods used with the first tab are implemented to identify channels and segments that are suspect. The second tab is designed to display information about selected channels. This includes a display of the time series and its power spectra as well as information about extreme and unusual segments. Input into the package is a .csv file and is part of the first tab. The .csv file needs to be organized beforehand so that each column is a time series of a specific microphone, e.g. channel. Additional data input includes the record length, number of channels and measure of extreme sensitivity. The record length is the number of time points that is used to determine a natural window for evaluating the interquantile range. For example, if a measurement series consists of 100,000 points, one might divide the series into M=100 segments with record length of size L=1000. The number of channels corresponds to the number of columns in the input file. The input associated with extreme sensitivity is the cutoff for evaluating points that are extreme; it is set at $z = 3$ as a default. From here we have two tabs of the application. When the application is initiated the initial page is blank ad awaiting file input.

The data evaluated below is part of a study on pressure fluctuations produced by forward steps immersed in a turbulent boundary layer (Awasthi et al., 2014). Figure 2 show a schematic of the test section, test wall and anechoic chambers. In a typical study sidewalls are constructed of Kevlar cloth however for this study one wall was removed and replaced with an adjustable wall that is suitable for growing high-Reynolds number two-dimensional turbulent boundary layers.

Diagnostic plots of the 31 channels are presented in Figure 12 and suggest problems with several microphones. Three microphones are identified as being off and five channels suggest clipping. The three channels have power spectra that are considerably different from the other microphones. This is also illustrated in the segment bar chart and the heatmap. Varieties of other channels have minor violations. A plot of channel 4 illustrates clipping (Figure 13). Note that the histogram plotted in the lower left plot shows peaks at the upper and lower values. Channel 6 is one channel that is identified as being off and this is indicated by the flat time series, the narrow histogram and simple power spectra (Figure 14).

The MATLAB GUI allows rapid evaluation of microphone data quality and identifies extreme segments, microphones that are not functioning as well as clipping and overrange. In an example with 31 microphones, we identified three microphones that were not functioning and several microphones with clipping. Identification of these issues may lead to improved analysis of the information in the microphone system. The GUI provides some flexibility for a user. This includes selection of the critical threshold for identifying unusual groups of observations. It is recommended that the user use the GUI with some data of known quality to identify the threshold prior to use on experimental data. We have found the threshold of three to be somewhat liberal with real data, i.e. it tends to identify unusual observations that do not affect resulting analyses. Evaluation of individual series using a Bonferroni correction can also be helpful in selecting the critical threshold.

Having developed the mathematical foundation for the error detection framework, and described the components of the facility sensor system, we apply the GP methodology to our data set to gain some insights about the system performance and anomalies.

Multiple types of sensors

In this section we illustrate the kinds of output our methods produces when considered across multiple runs of a given system. While the kinds of anomalies the algorithm finds are often clear to the human eye, the algorithm can consider the output from an entire experimental trial at once. This saves the experimenter the need to manually analyze the output of every sensor system, a potentially monumental task.

Figure 15 shows the output from the phased array system across eight experimental runs. Specifically, the circles indicate the root-mean-square of the measured time-series for each microphone signal. The dashed lines separate the output from one run to the other (there are 251 microphones in the

phased array; hence, 251 observations per run). There is no natural continuity between runs however, they are presented sequentially for convenience. GP predictions for the same set of data (circles) are shown with lines. Here, we compare the predictions made with two different covariance structures: a block diagonal structure, which does not correlate sensor data across runs, and a full block structure, which does. Essentially, this means that for the eight runs shown in Figure 15, using a block diagonal would only use information from run 1 to predict run 1, whereas the full diagonal would also use information from runs all runs to predict run 1. We chose to include this set of runs to demonstrate two classes of anomalies found thus far and illustrate the performance of the two covariance structures considered.

The first type of anomaly is a recurring sensor failure, shown by the repeatedly low point with an RMS of about -10. That specific microphone was faulty at the start of the experiment and remained faulty throughout. The block diagonal structure predicts a drastically different value for the output of the sensor than the actual measured output. In essence, the prediction is smoothed over by the nearby microphones in the array. Because of the great discrepancy between the prediction and the actual value, our algorithm will flag this output as anomalous. Contrast that to the output of the full block structure. Because the full block uses data from every run, and because the anomaly is consistent across runs, this implementation interprets the disparate output as signal and produces a prediction in line with the observed value.

The second type of anomaly is an idiosyncratic sensor malfunction, one that manifests itself for a short period of time and then vanishes. The exact cause for this is still unknown, but a thorough discussion of both anomalies in the context of this global detection scheme will be included in the final paper and compared to the results of a more typical engineering analyses. The 12 example runs were performed sequentially, so the anomalies in panel 9 were repeated neither before nor after. Because the full block structure learns the relationship between the sensors by considering every run, it does an excellent job of modeling what the proper signal for panel 9 ought to have been. The block diagonal structure has no such recourse, however, as it does not attempt to correlate sensor output across runs.

Figures 16 and 17 show the global correlation maps using the sensor scaling and system scaling approaches, respectively. White indicates a high correlation, and red indicates a low correlation. The dashed lines are used to show the sub-matrices corresponding to the four available sensor systems. The solid white rectangles are runs where data was not taken with that particular sensor group. The difference between the two is in the way the raw sensor outputs are scaled. Figure 16 uses a sensor scaling approach and Figure 17 uses a system scaling approach.

First, note that these plots were produced without using the GP methodology described above. In the final paper, the covariance matrices corresponding to these heat maps will be used for modeling and anomaly detection, but they are revealing in their own right. While the results we will discuss are

preliminary, we wish to highlight that even with this relatively simple construction, we are able to see global patterns in the experimental data and raise interesting questions about the relationships therein.

Next, note the solid red lines crossing all the way across Figure 17 in the rows corresponding to the rake. These lines indicate that the sensors to which they correspond do not correlate with any other sensor in the entire apparatus. While this is not necessarily an aberrant outcome, as it is possible to have sensors which measure a mostly independent quantity, it is unusual that the pattern is so drastically different from the other rake sensors. The reason for this pattern is that these sensors were not used during the experimental trials, a fact known to the engineers but not revealed to the statisticians. This demonstrates that, by examining these figures, one is able to clearly see non-functional sensors without having any knowledge of the physics, or even the instrumentation, of the experiment.

We can readily see additional features in the data by examining Figure 16. Note the distinct looking squares in the corners of the rake by rake section of Figure 16. These blocks correspond to the peripheral rake sensors, the ones in the boundary layer of the wind tunnel. We hypothesize that these sensors are correlated with themselves, but not to with the sensors in the center of test section, because of the different flow regimes in which the sensors are placed (the freestream and the boundary layer).

Finally, note the bands in the rake by phased array submatrix in Figure 16. These bands are brighter at the edges and darker in the interior, indicating that the phased array is more correlated with the sensors in the boundary layer than they are with those in the smooth flow. The reasons for this are not clear and will be explored more fully in future work. This demonstrates that our methodology raises novel questions about the relationships in the experiment, questions that may lead to novel insight about the experimental apparatus, or potentially even the underlying physics.

Opportunities and additional research

The Gaussian process approach provides a valuable tool for the identification of anomalies and is a powerful approach for combining information from different sensor types. There is considerable opportunity for additional research to extend and further investigate the above ideas. These include:

1. Development of the theoretical methods needed to assess sensor fidelity, design sensor systems for error tolerance, and design run schedules for large-scale, fault-tolerant test campaigns that:
    a. Exploit available metadata
    b. Account for data of different dimension
    c. Account for diverse data types
    d. Can be integrated over existing operations
    e. Learn with experience
2. Development of practical software tools that incorporate the above methods into functioning measurement and test planning systems in a realistic diverse and large scale test operation (the VT Stability Tunnel)

3. Demonstration of the effectiveness of the above methods through application to and integration into routine operations of a large-scale facility such as the VT Stability Tunnel.

Figures

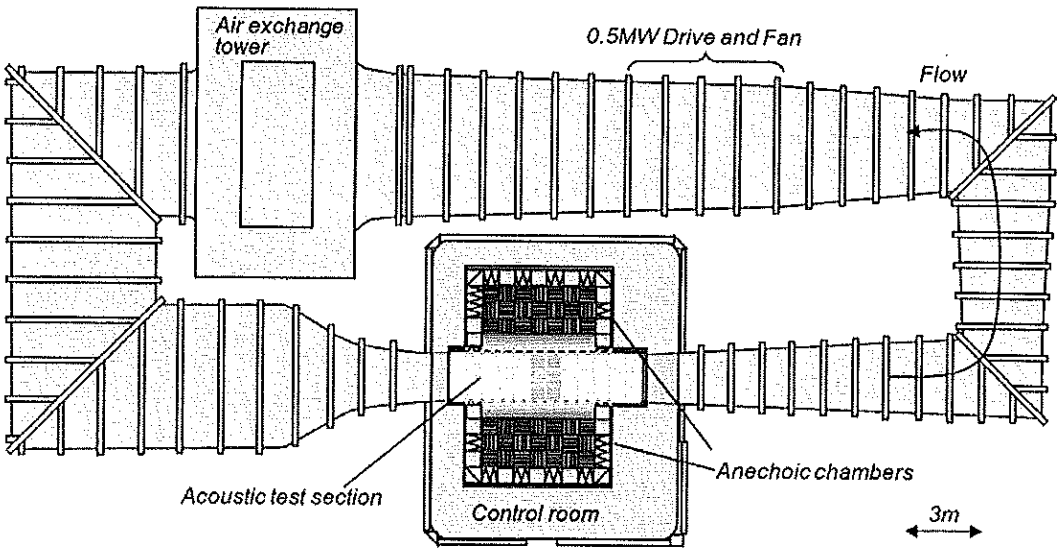


Figure 1: Overhead view of the VT Stability Wind Tunnel.

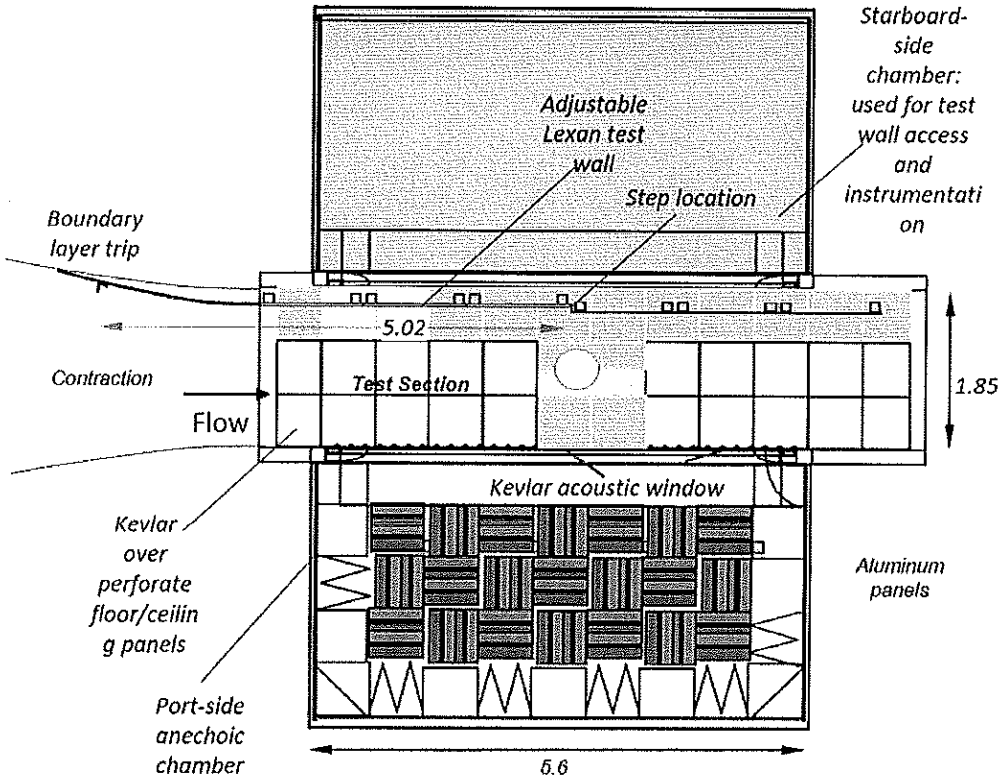


Figure 2: Top view of the test section used in the experiment to collect microphone array data, showing the anechoic chamber. All measurements are in meters.

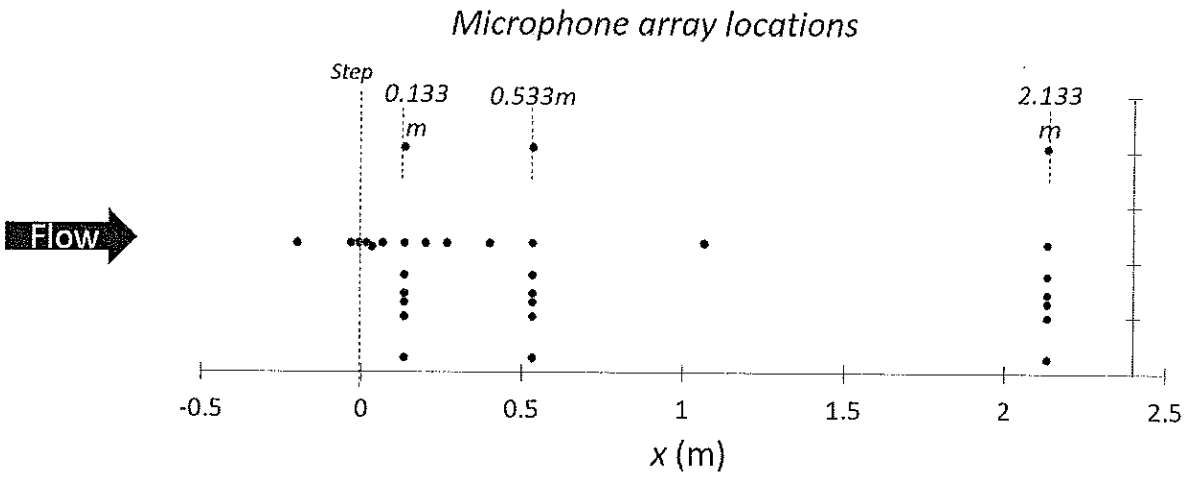


Figure 3: Locations of microphones to collect array data.

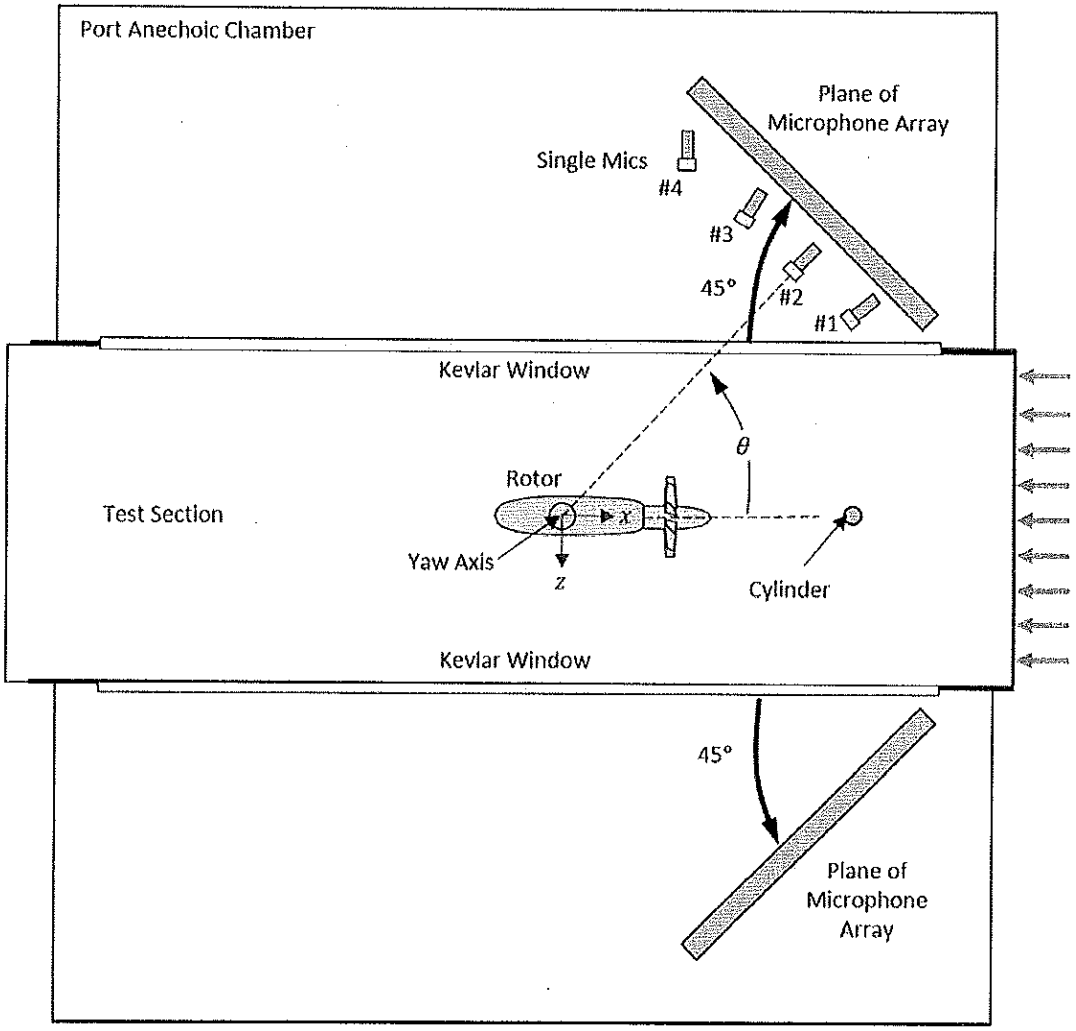


Figure 4: Phased array and standalone microphone placement relative to the free-stream direction.

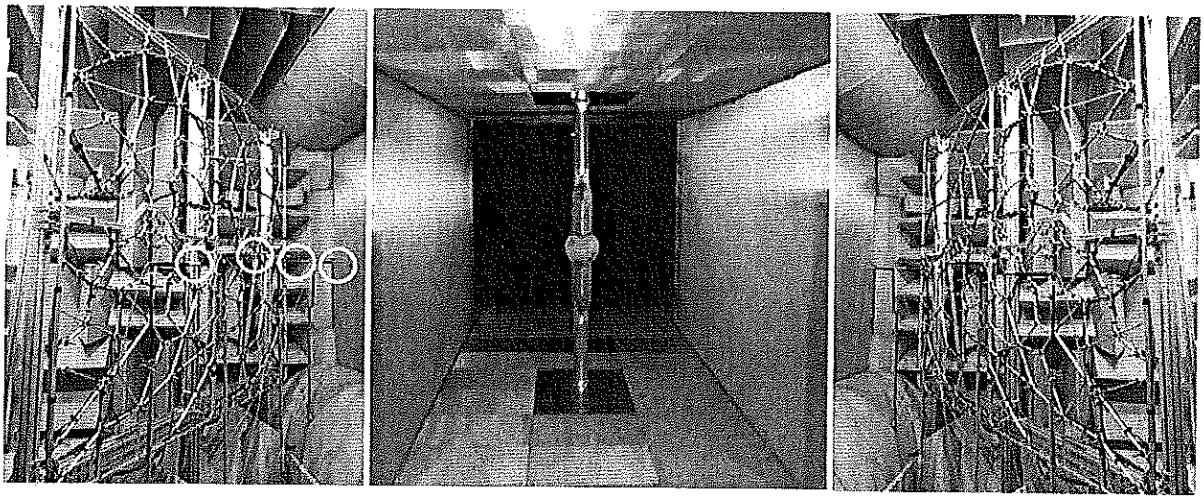


Figure 5: Tunnel view of the test section, phased microphone array, and standalone microphones (white circles added for emphasis). From left to right: port anechoic chamber, test section, and starboard chamber (flow is out of page).

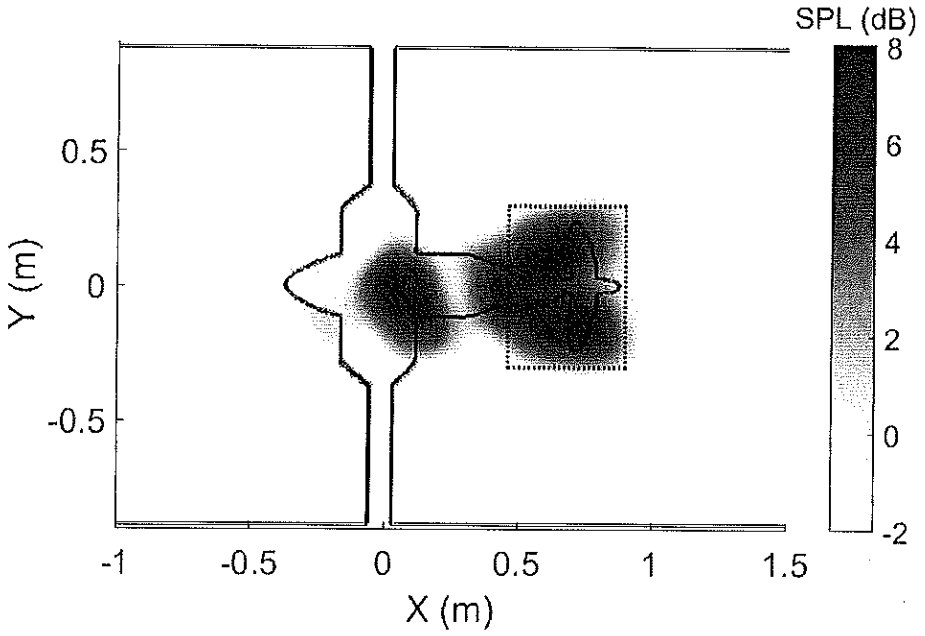
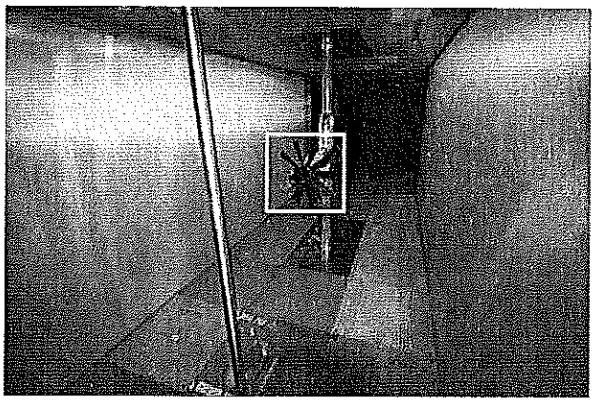
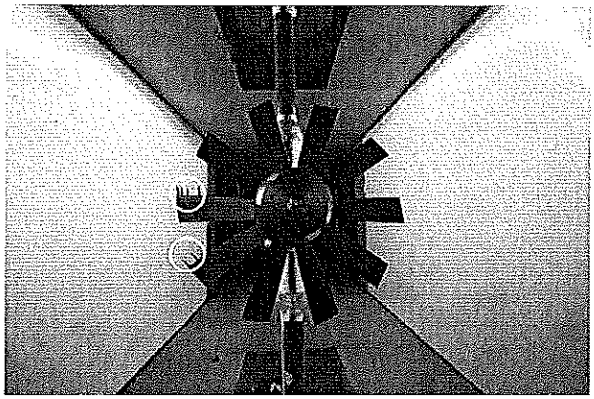


Figure 6: Beamform map at 2500 Hz. The map shows localized noise sources on the blades as well as on the motor housing (likely due to mechanical noise).



(a) Ten-bladed rotor ingesting the wake of an upstream cylinder.



(b) The on-blade hot-wire probes (flow is directed into the page).

Figure 7: Experimental photos of the on-blade HWA probes and upstream cylinder.

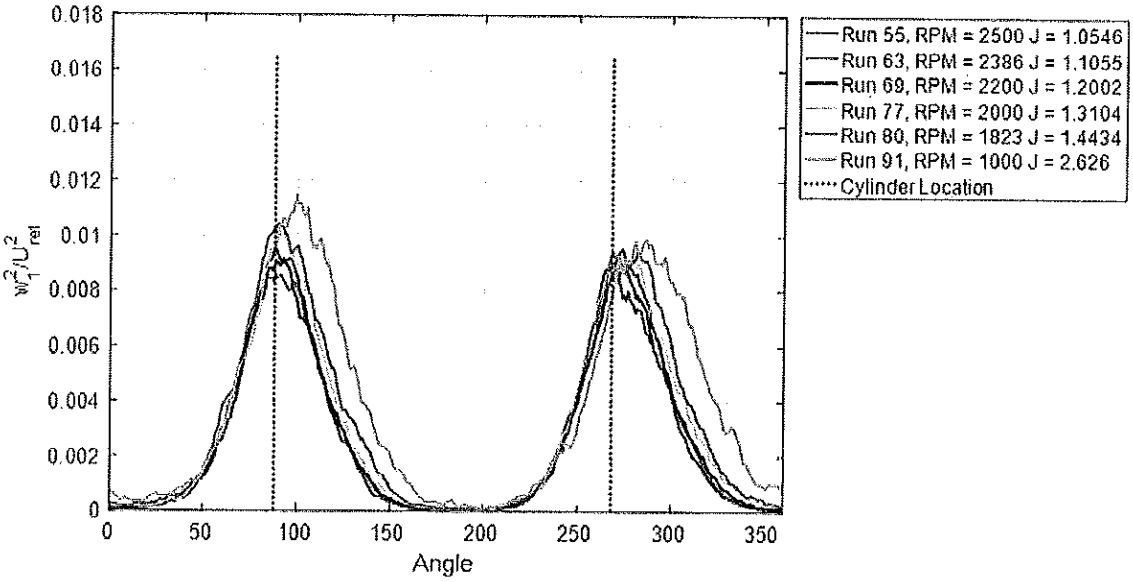
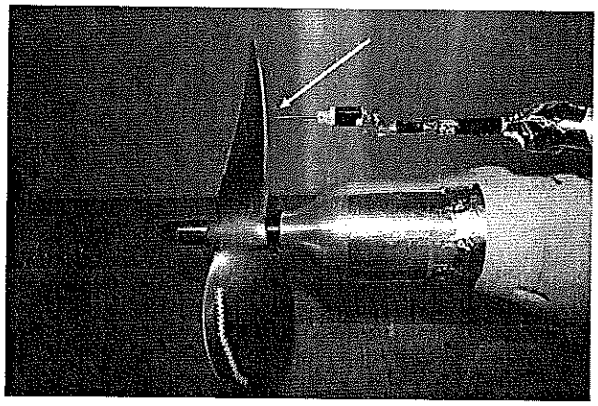
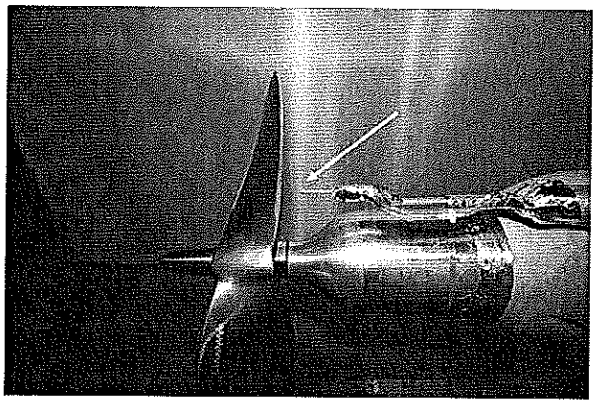


Figure 8: Upwash velocity for various advance ratios. The dashed line shows the cylinder location (upstream of the rotor). The probes cross through the wake center twice per rotation.



(a) Hot-wire probe (indicated by the white arrow) down- stream of a tripped two-bladed rotor, measuring flow ve- locity.



(b) Pitot probe (indicated by white arrow) downstream of tripped two-bladed rotor, measuring pressure.

Figure 9: Experimental photos of the probe measurements made downstream of a two-bladed rotor. The probes could be traversed to obtain a radial profile. Flow direction is from left to right.

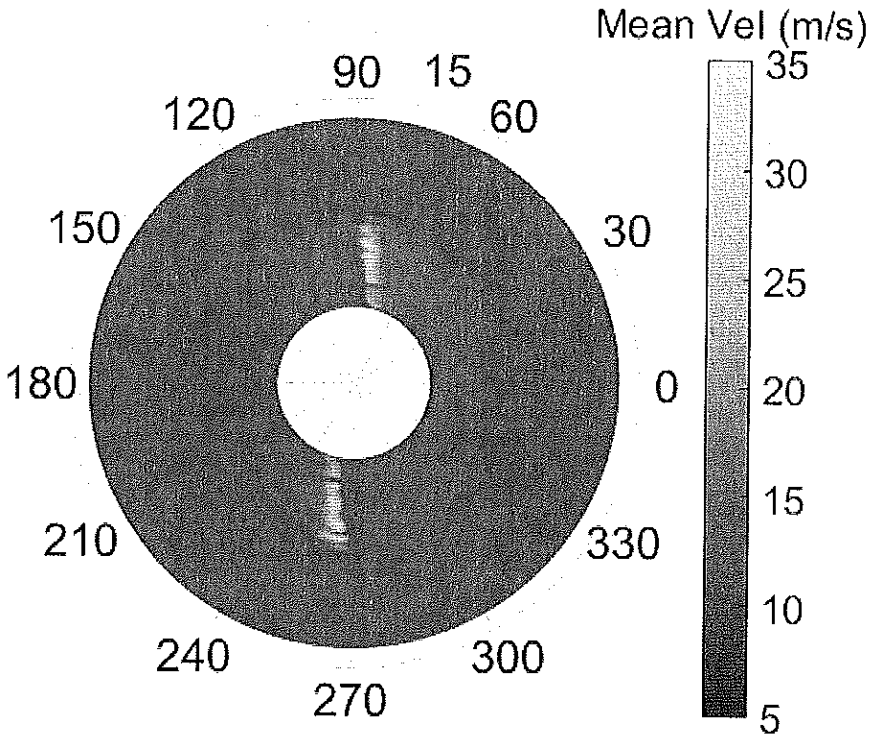


Figure 10: Phase-averaged mean velocity in the rotor plane, with the blades spinning clockwise.

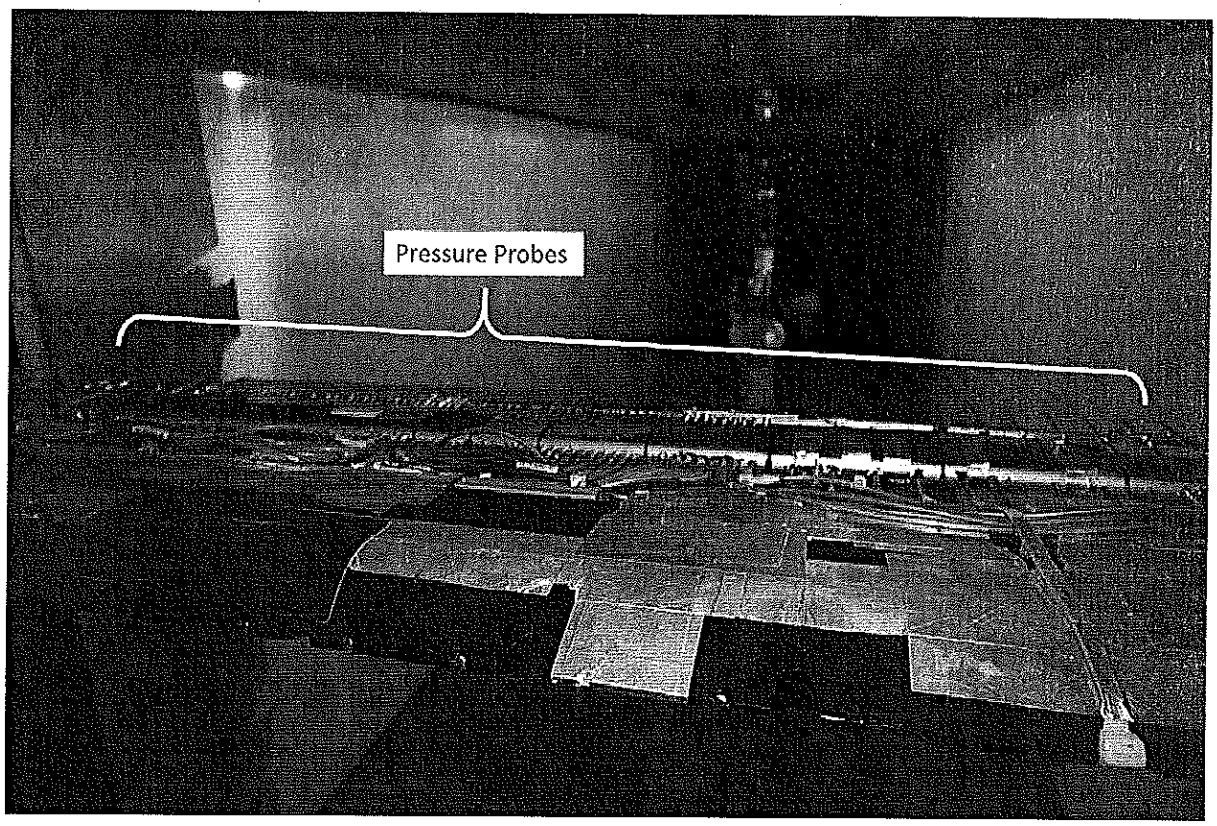


Figure 11: Wake rake of Pitot and Pitot-static probes that can be traversed vertically to give a cross-sectional profile. Flow is out of page.

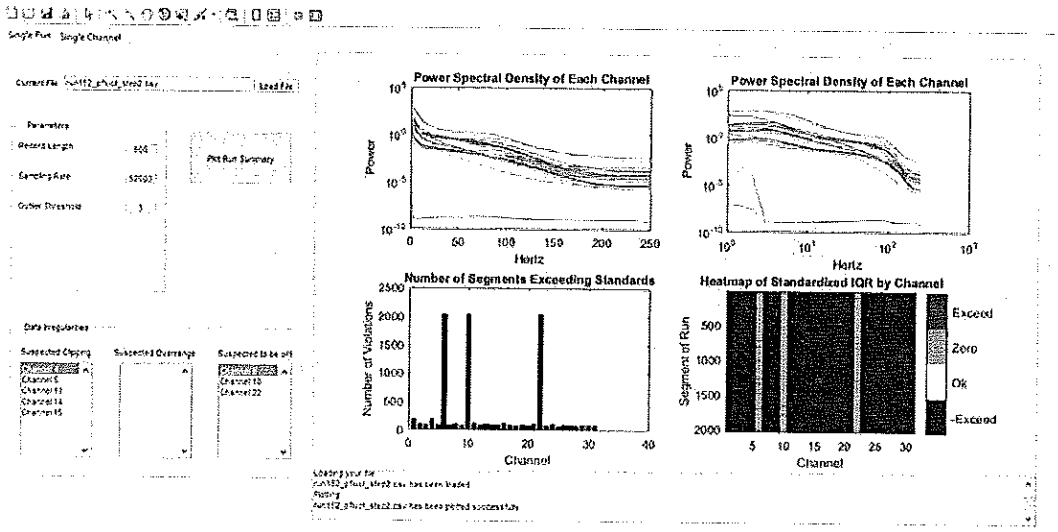


Figure 12: Spectral densities and plots identifying unusual segments in microphone array data.

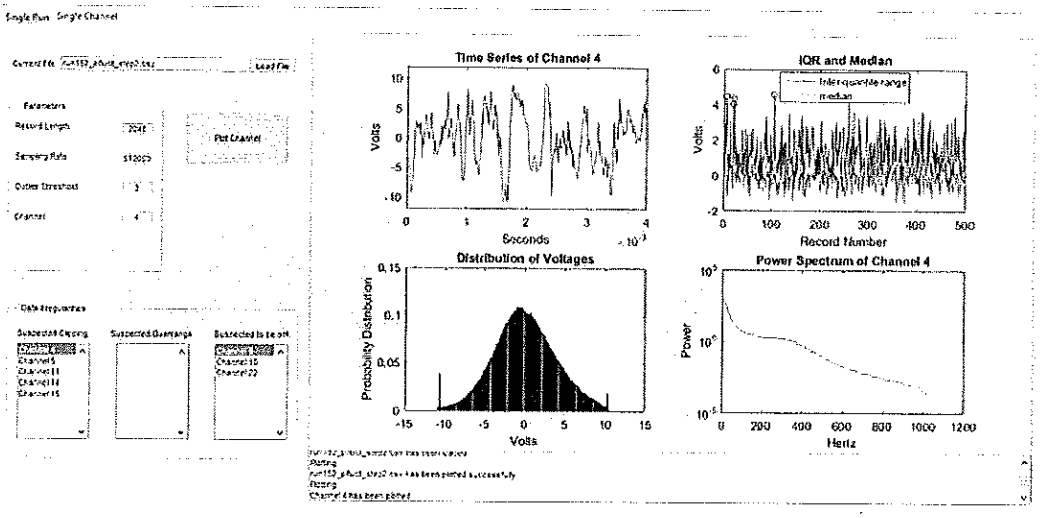


Figure 13: Graphical displays for channel 4 in microphone array data.

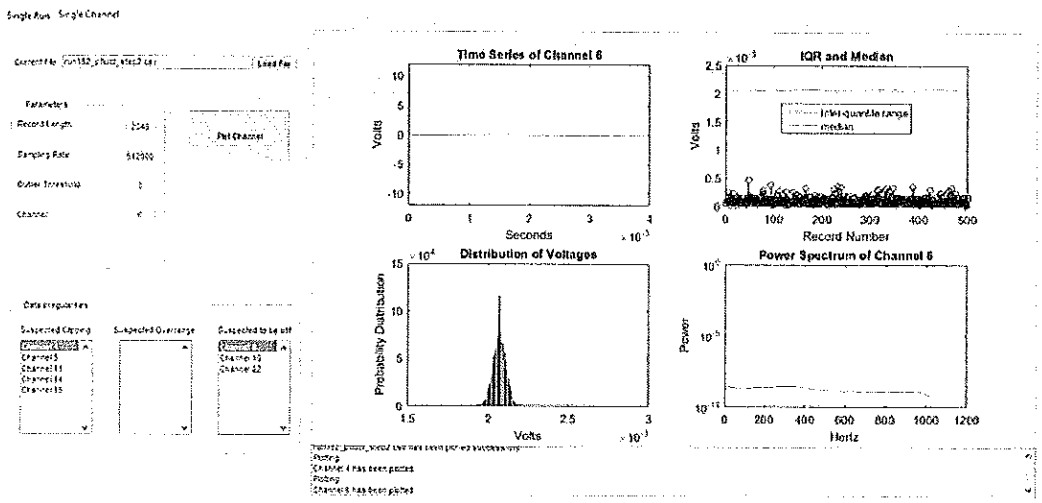


Figure 14: Graphical displays for channel 6 in microphone array data.

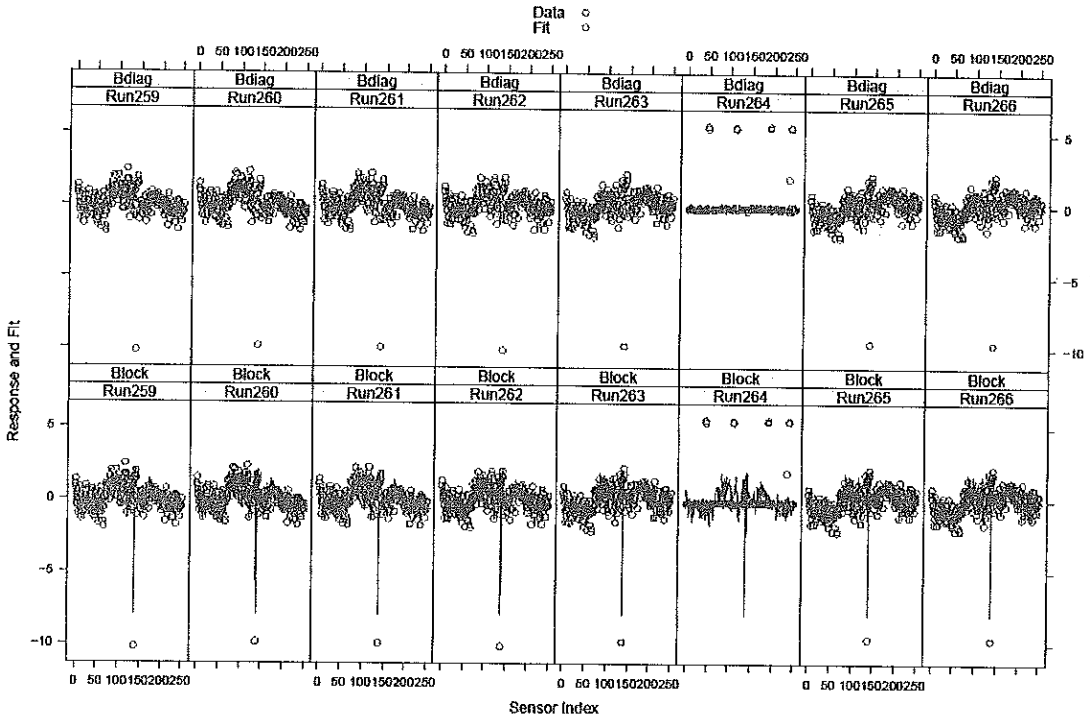


Figure 15: Root-Mean-Square of each of phased array microphone signal (circles), plotted versus microphone number within each run. The dashed vertical lines separate each run (recall that the phased array has 251 microphones. GP predictions are shown in red, using a) a block diagonal covariance matrix b) a full block covariance matrix.

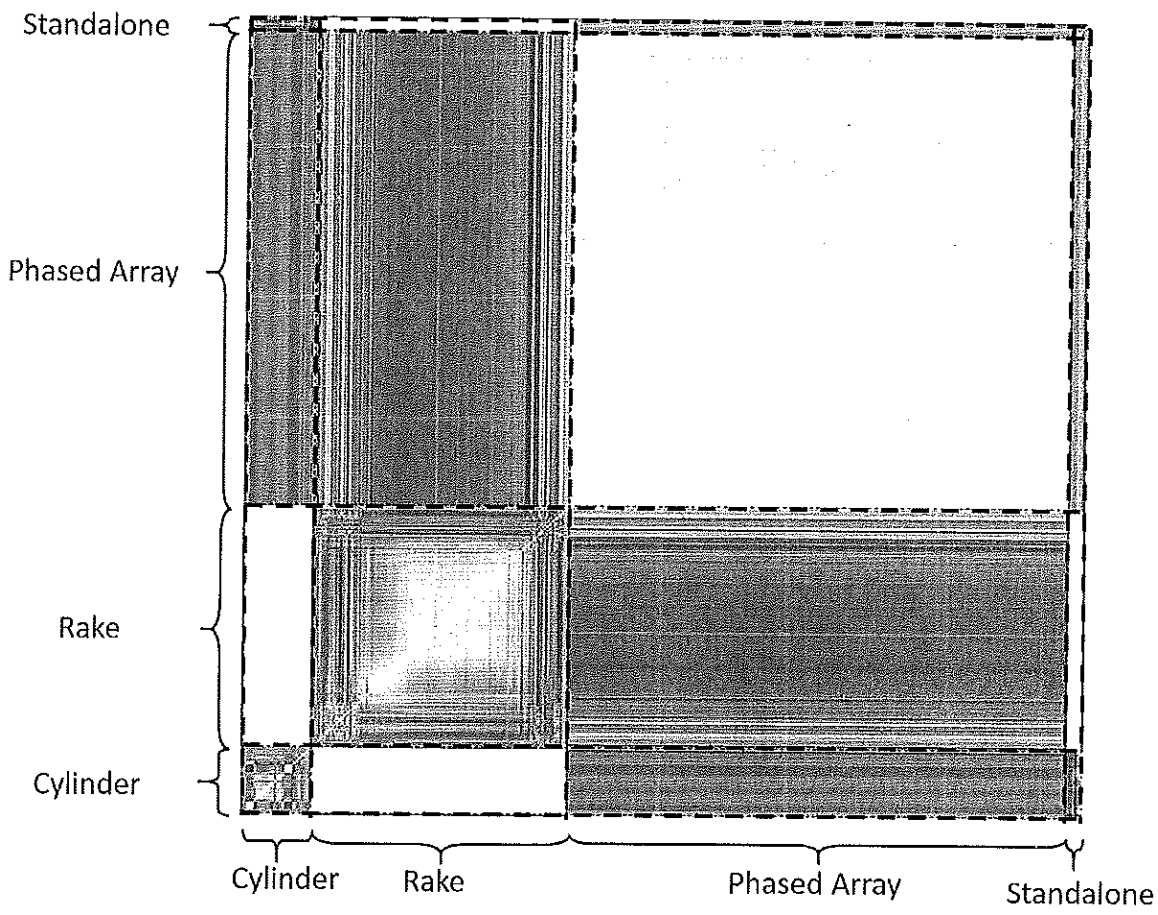


Figure 16: Heat map of the response covariance using sensor scaling

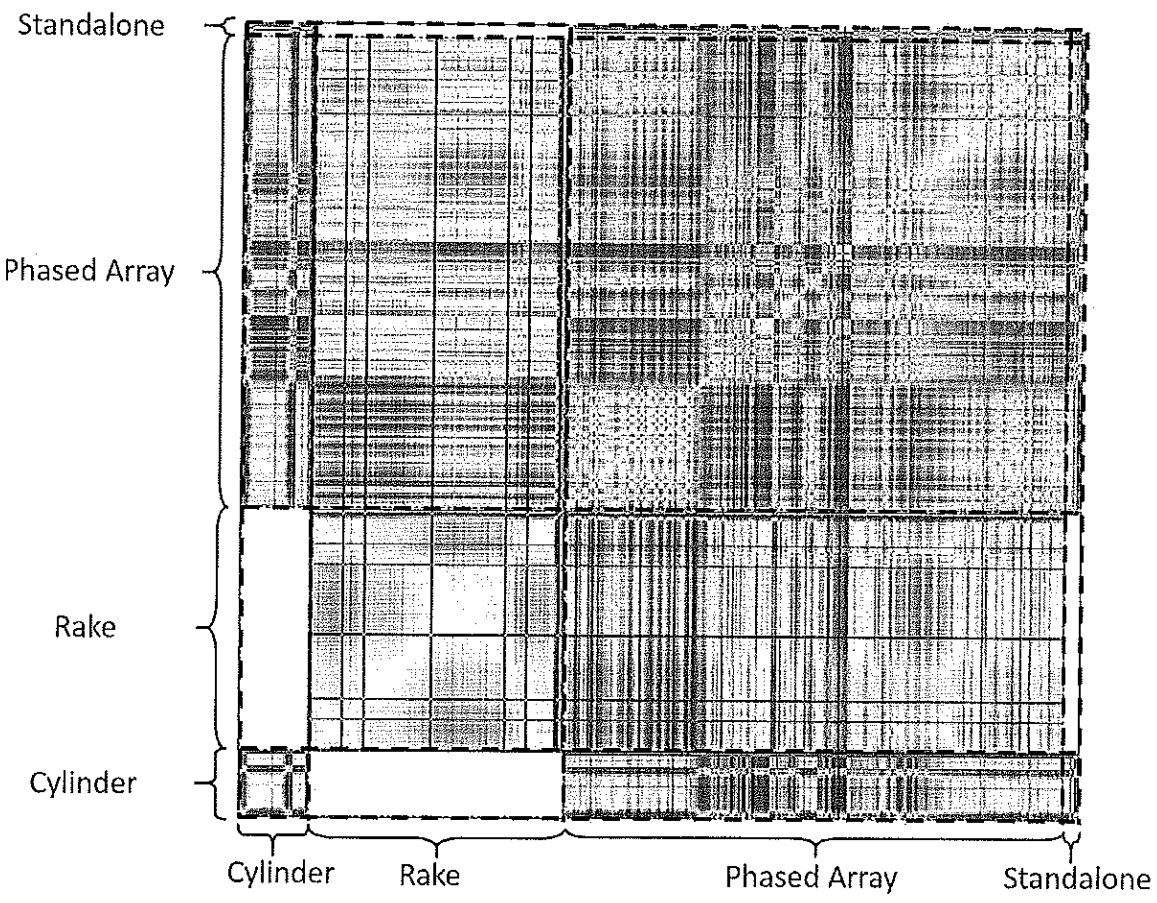


Figure 17: Heat map of the response covariance using system scaling

References

Alexander, W. N., Devenport, W., Wisda, D., Morton, M. A., and Glegg, S. A. L., "Sound Radiated From a Rotor and Its Relation to Rotating Frame Measurements of Ingested Turbulence," *20th AIAA-CEAS Aeroacoustics Conference 2014, 16-20 June 2014* , 20th AIAA/CEAS Aeroacoustics Conference 2014, American Institute of Aeronautics and Astronautics, 2014, p. 18 pp.

Alexander, W. N., Molinaro, N., Hickling, C., Murray, H., Devenport, W., and Glegg, S. A. L., "Phased array measurements of a rotor ingesting a turbulent shear flow," *22nd AIAA/CEAS Aeroacoustics Conference, 30 May-1 June 2016* , 22nd AIAA/CEAS Aeroacoustics Conference, ASCE - American Society of Civil Engineers, 2016, p. 27 pp.

Awasthi, M., Devenport, W. J. Glegg, S. A. L. and Forest, J. B. "Pressure fluctuations produced by forward steps immersed in a turbulent boundary layer". *Journal of Fluid Mechanics*, 756:384–421, 2014.

Binois, M., Gramacy, R. B., and Ludkovski, M., "Practical heteroskedastic Gaussian process modeling for large simulation experiments," *in progress*, 2016.

Clark, I. A., Baker, D., Alexander, W. N., Devenport, W., Peake, N., Glegg, S., and Jaworski, J. W., "Experimental and Theoretical Analysis of Bio-Inspired Trailing Edge Noise Control Devices," 22nd AIAA/CEAS Aeroacoustics Conference, 30 May-1 June 2016 , 22nd AIAA/CEAS Aeroacoustics Conference, ASCE - American Society of Civil Engineers, 2016, p. 15 pp.

Crandell, I., Millican, A.J., Leman, S., Alexander, N., Devenport, W., Vasta, R., Gramacy, R.B., and Binois, M., 2017, "Anomaly detection in large-scale wind tunnel tests using Gaussian processes". Abstract submitted to American Institute of Aeronautics and Astronautics Conference. Accepted for publication.

Cressie, N. A. C., *Statistics for spatial data*, Wiley series in probability and mathematical statistics, J. Wiley & Sons, New York, Chichester, Toronto, 1993.

Devenport, W. J., Burdisso, R. A., Borgoltz, A., Ravetta, P. A., Barone, M. F., Brown, K. A., and Morton, M. A., "The Kevlar-walled anechoic wind tunnel," *Journal of Sound and Vibration*, Vol. 332, No. 17, 2013, pp. 3971–3991.

Gramacy, R. B., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., Rutter, E., Trantham, M., and Drake,P., "Calibrating a large computer experiment simulating radiative shock hydrodynamics," *Ann. Appl. Stat.*, Vol. 9, No. 3, 09 2015, pp. 1141–1168.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D., "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal on Scientific Computing*, Vol. 26, No. 2, Jan. 2004, pp. 448–466.

Kennedy, M. C. and O'Hagan, A., "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464.

Kegerise, M. A. and Spina, E. F., "A comparative study of constant-voltage and constant-temperature hot-wire anemometers. Pt. I. The static response," *Experiments in Fluids*, Vol. 29, No. 2, 2000, pp. 154–64.

Matheron, G., *Trait´e de g´eostatistique appliqu´ee*, Vol. 14, Editions Technip, Paris, 1962.

Woodbury, M. A., *Inverting Modified Matrices*, No. 42 in Statistical Research Group Memorandum Reports, Princeton University, Princeton, NJ, 1950.

Publications submitted

Crandell, I, Millican, AJ, Leman, S, Alexander, N, Devenport, W, Vasta, R, Gramacy, RB, and Binois, M, 2017, *Anomaly detection in large-scale wind tunnel tests using Gaussian processes*. Abstract submitted to American Institute of Aeronautics and Astronautics Conference. Accepted for publication.

Vasta, R. I. Crandell, A. Millican, L. House, E. Smith, 2017. Outlier Detection Sensor System (ODSS): A MATLAB macro for evaluating sensor data quality. American Institute of Aeronautics and Astronautics Student Conference (submitted).

Education

Ian Crandell, Heterogeneous Covariance Estimation for Gaussian Processes, completion of PhD expected August 2017

Anthony Millican MS

Robert Vasta: the grant provided undergraduate research support.