COMMISSION OF THE
EUROPEAN COMMUNITIES

Equitable Testing and Evaluation of Marine Energy Extraction Devices in terms of Performance, Cost and Environmental Impact

Grant agreement number: 213380

Deliverable D3.4

Best practice for tank testing of small marine energy devices

Grant Agreement number: 213380

Project acronym: EQUIMAR

Project title: Equitable Testing and Evaluation of Marine Energy Extraction Devices in terms of Performance, Cost and Environmental Impact

Deliverable D3.3

Best practice for tank testing of small marine energy devices

T. McCombes, C. Johnstone
University of Strathclyde

B. Holmes
University College Cork

L. E. Myers, A.S. Bahaj
University of Southampton

J.P. Kofoed
Aalborg University

December 2010

Summary

At present no common practices are adopted to assess the performance and operational characteristics of conceptual and small prototype wave and tidal energy devices when tested within controlled laboratory environments. Information acquired from this early stage assessment may be used to secure development funding or promote a specific wave or tidal energy device. Since no standards exist, the data produced may be misinterpreted or inaccurately presented, which in turn may lead to failure to live up to performance expectations, as devices scale up in size.

This report builds on Deliverable 3.3 which identified limitations of current practices adopted for tank testing of small prototype devices. The recommendations contained herein constitute minimum set of best practices for device testing and benchmarking. The protocol contains explicit Design of Experiment and Uncertainty Analysis techniques. Particular emphasis has been placed on repeatability, quantification of uncertainty, estimation of accuracy and elimination of laboratory specific effects.

Outline

Definitions:

Model: A mathematical approximation to the behaviour of the prototype. In general, a useful model may be a very simple series of algebraic expressions or an appropriate computational code (such as Blade Element Method) such that performance data may be quickly generated

Prototype: The device or component to be tested. This will be proportionately scaled down from full size, to an extent commensurate with the limitations of the test facility.

Deterministic: Of a process or system, one where future states are wholly determined by those preceding, i.e. in which there is no random variability. Mathematical models, computer simulations and theoretical calculations are generally completely deterministic (within the bounds of rounding errors and where random processes have not been artificially added).

Statistical: Of a process or system, one which includes random variation. Statistical methods are then used to extract the response of a system hidden in the data to 'noise' in the data; the major tenet is to locate and describe repeatable underlying phenomena. Repeated measurements yielding different results under fixed conditions are an indication that the process is statistical, although the reverse is *not necessarily true.*

Test: A process of measurements leading to some result.

Measurement: An indication of the state or a property of some object**.**

Confidence level: The level of certainty that the true value of a measurement lies within a particular margin.

Interval: The margin within which the true value of a measurement is said to exist.

Error: The difference between measured and true value of a thing being measured.

Uncertainty: The level of doubt in the measurement result.

RSS: Root-sum-of squares. The means by which uncertainty is propagated.

DRE: Data reduction equation.

1. Introduction - Experimental Good Practice

This document sets out a Protocol for tank testing wave and tidal marine energy converters. It contains explicit Design of Experiment and uncertainty analysis methodologies which should be considered the minimum requirement for tank testing work. This places particular emphasis on repeatability, quantification of uncertainty, estimation of accuracy and elimination of laboratory specific effects.

Purpose of this Document

Protocol Requires

Developer must deliver bounded uncertainty:

Final result [e.g. C_P, C_T, η, ... inflow] ± 5% full scale with a 95% confidence level

Protocol Provides

Guidelines for and Means of identifying, reducing and reporting uncertainty in experimental results:

- Minimum Uncertainty Analysis Requirements
- Potential Design of Experiment Methodologies
- Specific techniques for Wave and Tidal technologies

The purpose of an experiment is to generate physical data to test a hypothesis. The purpose of experimental good practise, manifest in Design of Experiment (DoE), is to optimise in advance an experimental process in order to generate the maximum quantity of high quality data – in other words maximising *value for money* for a particular experiment.

In the context of EquiMar, the experimental procedures are those which will provide performance data on the performance of conceptual marine energy devices, however the DoE process as well as that of the Uncertainty Analysis are common to a very wide range of engineering fields and thus the domain is well documented and processes and procedures are widely accepted.

The procedures outlined in this document are primarily a synthesis of those promulgated in

- the proceedings and procedures of the International Towing Tank Conference (henceforth referred to as ITTC) [1]
- the American Institute of Aeronautics and Astronautics (henceforth AIAA) [2]
- the ISO Guide to the Expression of Uncertainty in Measurement [3] and the NIST Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results (henceforth GUM) [4]
- the ISO International Vocabulary of Basic and General Terms in Metrology (henceforth VIM) [5]
- the NIST Reference on Constants, Units and Uncertainty [6]
- the NIST/SEMATECH e-Handbook of Statistical Methods [7].

The core purpose of these procedures is thus to allow an experimental test result to be stated in the standard form (as defined in the VIM) of either a **standard uncertainty**, i.e.:

$$\text{(Result): } x \text{ (units) [with a] standard uncertainty of} \, u_c \text{(units)}$$

or an **expanded uncertainty**, i.e.:

$$\text{(Result): } x \ \pm U \text{(units)}$$

such that the uncertainty is a combination of all identified, reduced where possible and accounted for uncertainties associated with the experiment. The expanded uncertainty is related to the standard uncertainty via the coverage factor *k*, itself calculated (under the assumption of normally distributed data) from the Student *t*-statistic where degrees of freedom *υ* is the number of samples or tests minus 1:

$$U = ku_c \text{ where } k = t_{\alpha/2}\left(\nu\right) \text{ for } \nu \text{ degrees of freedom}$$

Qualitative and quantitative guidance is provided in Deliverable 3.3 [8] on potential sources of experimental error, and minimum requirements are contained herein as to how these are included and presented in a result statement. Further generic guidance is given to identifying common error types, and on to how to design and undertake an experiment to reduce or eliminate these errors and on how to present the result. Later sections of this document detail specific requirements, tests and methods associated with experiments on wave and tidal devices to further support these generic procedures.

The recommendation of this Protocol is that all experiments are conducted in such a manner that the reported performance of a prototype device is stated with a precision of 5% at a confidence level of 95%. **In plain English this requires that 95 times out of 100 the error of a reported value is no greater than 5% of the true value.** This requires that the standard uncertainty is calculated for large degrees of freedom such that the coverage factor, *k*, is 0.96 (approximately 2), corresponding to approximately 95% coverage.

Therefore this document should be used alongside Deliverable 3.3 [8] to identify error sources, calculate their contribution to overall uncertainty and focus efforts in the most efficient manner into reducing it in line with this recommendation. Furthermore, it is essential that experiments are conducted in such a way that any reported causal relationship is actually present in the physics. To this end, some basic design of experiment methodologies are described to aid construction of the test schedule.

1.1. Purpose of the Test

It is likely that the majority of tank test programmes will be undertaken in order to achieve one of the following objectives.

Proof of Concept	Unstructured experiment to answer the question **"does it work?"** at some fundamental level. Very short tests likely to proceed in a trial and error manner, often unaccompanied by a mathematical model. *Example: determine whether a wave device moves in a wave field.*
Comparison	Determination of the **significance of levels of a single variable**, identified in advance, on the response. *Example: determine the* $C_P - \lambda$ *characteristic of a simple rotor*[1].
Screening	Identification of **a subset of the most important variables**, from a larger set of candidate variables that have been identified in advance, on the response. *Example: identify the key geometric performance variables for a novel wave energy device.*
Response Surface Modelling	**Optimisation** via identifying relationships and estimate interactions between multiple variables and responses, and specifically identify the levels of the important variables which would produce an optimum response. Quadratic surfaces can be fitted to data providing local maximum/minimum. *Example: reduce the two responses* pitch magnitude *and* roll magnitude *as a function of* H_s *and* T_z *for a wave energy device.*
Model Fitting	Identification of a high quality **mathematical model** in terms of goodness of model parameter estimates. *Example: estimate the numerical models of the two responses* C_P *and* C_T *as a function of* blade pitch *and* TSR *for a novel tidal turbine rotor.*

Once the objectives of the test have been identified, it is possible to select appropriate uncertainty analysis and design of experiment methodologies.

1.2. Outline Test Procedure

Although every test procedure is individual, this protocol recommends the adoption of the general outline test process formulated by the AIAA [2] and adopted by the ITTC [10]. This provides a means of introduction and integration of uncertainty assessment into each phase of the experimental process, with appropriate decision points and reporting.

The ITTC states that "*this philosophy of testing, rigorous application/integration of uncertainty assessment methodology into the test process and documentation of results should be the foundation of all [towing] tank experiments.*"

[1] λ is actually made up of 2 factors: inflow and angular velocities.

1.2.1. Description

The stages of the flow chart in Figure 1 correspond to the various subsections in this document, which are identified in the following extended summary. This breakdown of the pre-test procedures identifies a list of what should be considered mandatory stages, but which will generally be undertaken automatically as part of a well thought out experimental process, therefore do not constitute a significant or onerous burden in time or resource. In common with the technical documents of the EquiMar project, this document considers parts of the 5 stage development schedule, specifically Stage 1: Concept Appraisal and Stage 2: Large Scale Tank Testing.

The following list sets out the stages whose documentation is required under this Protocol.

1. **Requirement:** Identify test objectives
   - Stage 1: Functionality/Proof of Concept; Comparison; Factor Screening…
   - Stage 2: Optimisation; Variation Reduction; Adding Robustness; Model Identification…
2. **Requirement:** Identify facility and process (e.g. towing tank –> thrust and power measurements whilst towing)
   - Ascertain capabilities and proficiencies of facility
     - Range and quality of measurements & instrumentation
     - Range and types of tests
     - Calibration capabilities
3. **Requirement:** Identify primary and secondary model(s)
   - Write data reduction equation(s)
     - Perform sensitivity analysis using instrument tolerances & estimated experimental biases -> estimate, tolerate & correct
   - Focus resources on reducing estimated result bias below 5%
4. **Requirement:** Design of Experiment
   - Statistical Design of Experiment for maximum quality (minimised uncertainty) of data and maximum robustness of interpretation of results
   - Different DoE approaches to be taken depending on objectives, number of factors etc.

1.2.2. Flow Chart

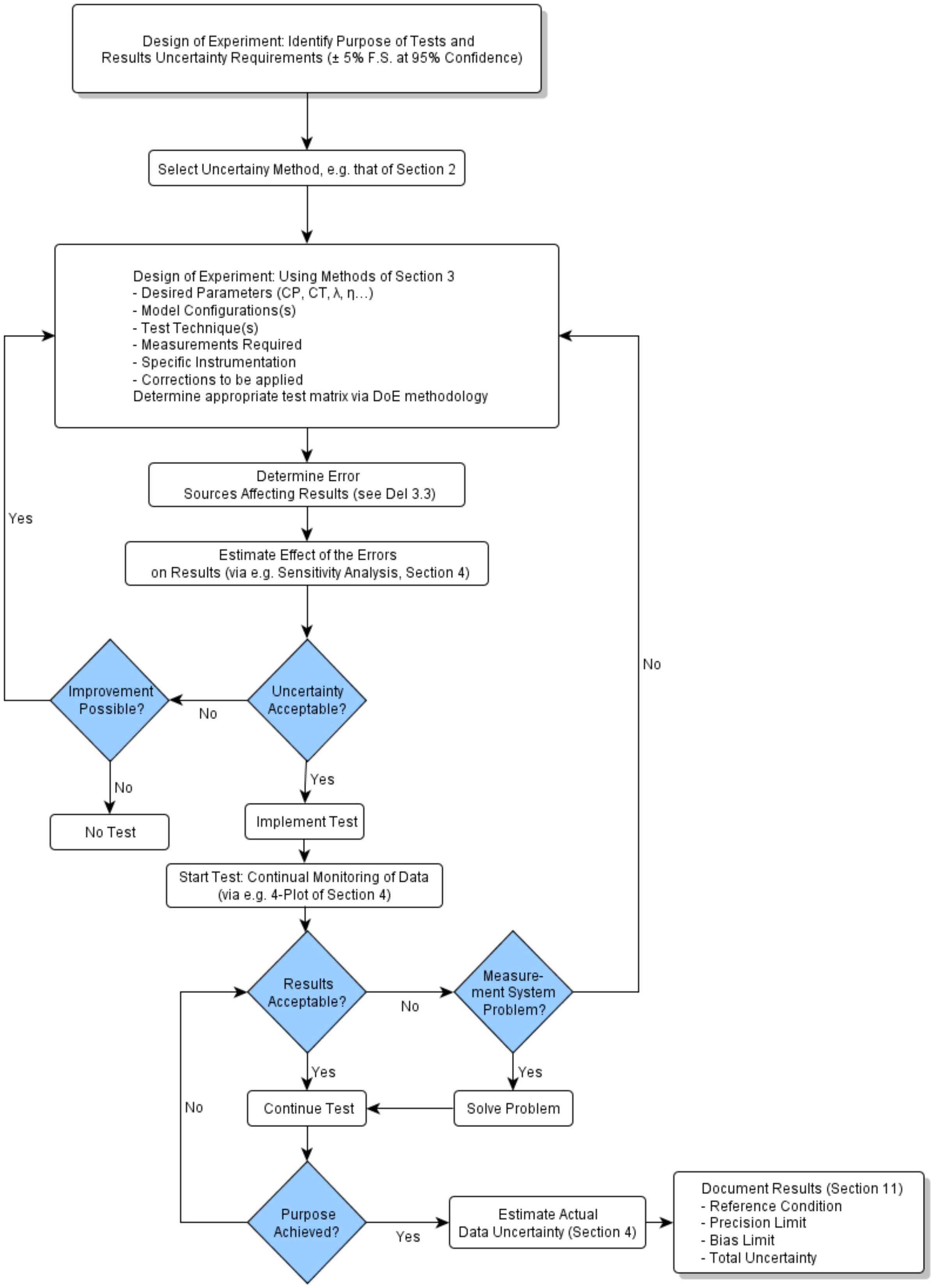


Figure 1: Flow chart of experimental process, indicating decision points and information sources. Adapted from ITTC Procedure 7.5-02-01-02 [9].

2. Quality and Accuracy of Results

The quality of an experimental measurement can be quantified in terms of how much an estimate of a parameter value is in error from the true value. Unfortunately the true value of the parameter is either not known (in the case of an experimental test) or is only known for a limited subset of a much larger population of influencing factor values (in the case of a calibration test). Therefore quality of a measurement must be estimated using an uncertainty analysis to compute the error in the result.

Traditionally, the error is decomposed into a precision component, accounting for random scatter about some mean value, and an accuracy component, which when quantified is known as bias, that accounts for a shift in the mean value from the true value. These concepts are illustrated in the following figures:

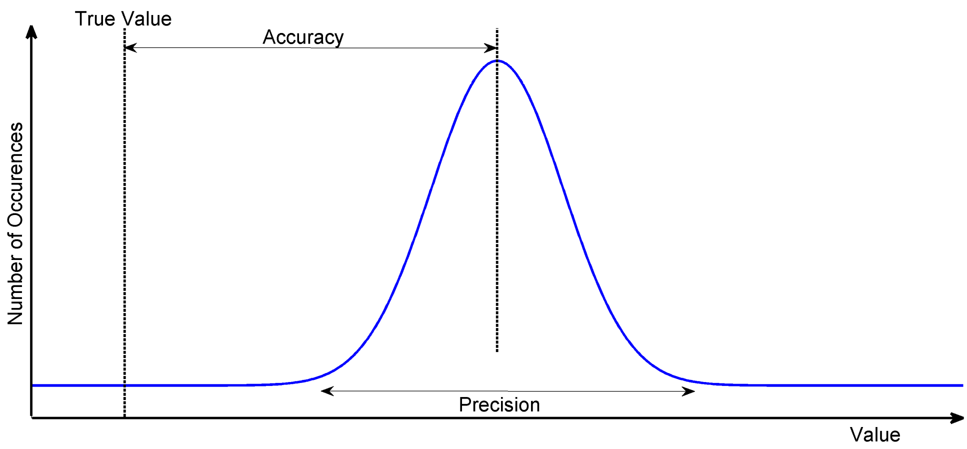


Figure 2: Illustration of Precision and Accuracy components of exerimental uncertainty, and also the notion of a probability density function.

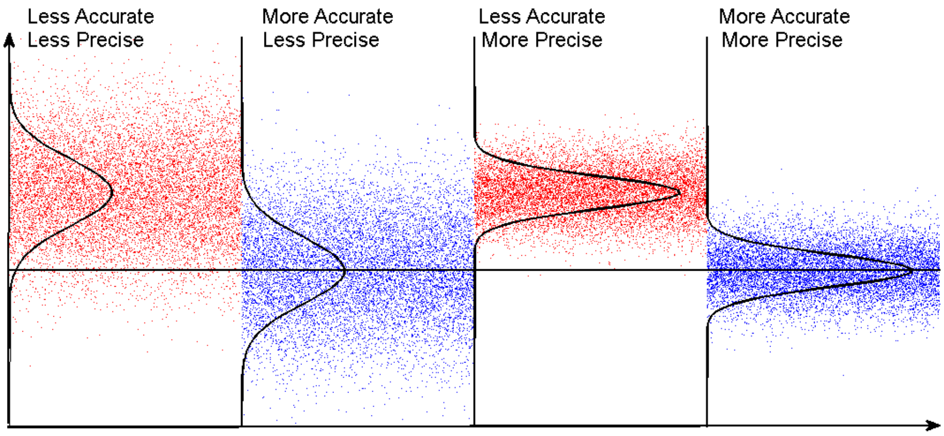


Figure 3: Illustration of qualitative Precision and Accuracy components of exerimental uncertainty, and also the effects on a probability density function.

When considering time series data, the distribution around the mean values may appear as follows:

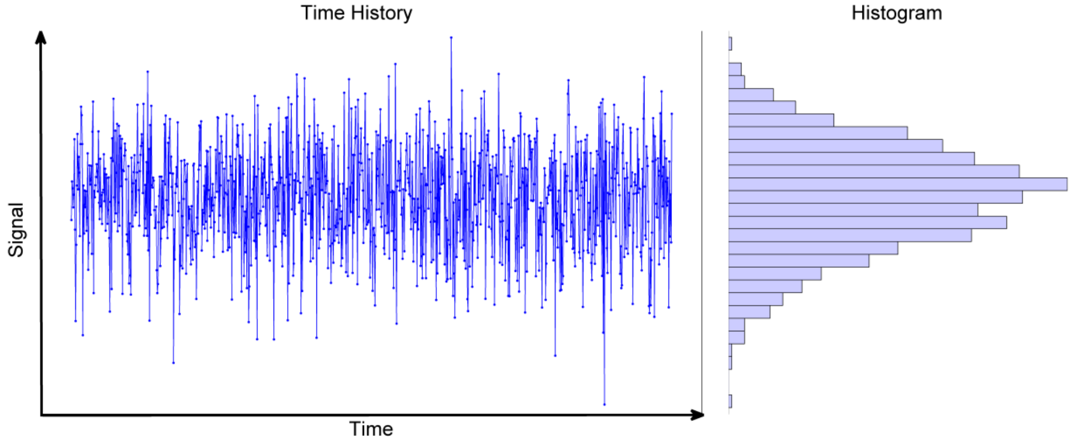


Figure 4: Illustration of a noisy time series of data, and the associated histogram of values.

The uncertainty analysis (UA) methodology of Section 4 provides a basic framework for estimation of the uncertainty terms (e.g. the combined precision and bias errors) in an experimental measurement.

2.1. Statistical Definitions

A number of statistical properties can be used to describe the data:

Given a series of *n* measurements of the same measurand, in a single test, the **sample mean**. is the mathematical average of the data values, and is the point at which the expected value would occur:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{... 1}$$

The experimental **sample standard deviation** is the mean distance between the mean of the data values and the values themselves and is a measure of the spread of the values in the sample:

$$s = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{... 2}$$

If a number, *M,* of tests have been performed in order to reduce the variance in *x* and thus the standard deviation, then a number of quantities describing the sample distribution, that is the distribution of a particular test statistic calculated for a test sample of set size, can be determined. The Central Limit Theorem posits that as the number, *M*, of tests are increased, the distribution of test statistics will tend to normal, in other words that the distribution of a test statistic, e.g. the mean, is normally distributed, regardless of the distribution of the population from which it was drawn. This allows the use of an assumption of normality on test statistics, greatly easing statistical uncertainty analysis.

The experimental **standard deviation of the mean** is the standard deviation of the different experimental test sample means. This is also sometimes (mistakenly) known as the standard error of the mean, and is equivalent to the standard uncertainty:

$$u(x_i) = s(\bar{x}) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{... 3}$$

with x_i being the result of the i^{th} measurement. If multiple test results are not available, then an estimate of the standard deviation of the mean can be computed from a single sample of n observations by the following:

$$u(x) = \frac{s(x)}{\sqrt{n}} \qquad \text{... 4}$$

Clearly, as the number of samples in the test increase, the estimated standard deviation of the mean decreases. This **equation** is strictly only valid for normally distributed data or data made up of statistical quantities adhering to the central limit theorem.

If the standard deviation is sought over M tests, which in practice requires that the sample standard deviations are of the same magnitude, then the pooling formula may be used to calculate the weighted average of the standard deviations:

$$s_p = \sqrt{\frac{\sum_{j=1}^{M}\left((n_j - 1)\, s_j^2\right)}{\sum_{j=1}^{M}(n_j - 1)}} \qquad \text{... 5}$$

2.1.1. Coverage & Expanded Uncertainty

A coverage factor, k, is used to re-scale the combined uncertainty – which is essentially one standard deviation of the result – such that the uncertainty may be stated at other confidence levels than approximately 68% (the confidence level for one standard deviation). This results in the expanded uncertainty, as defined in section 1.

Coverage factors typically applied to a normal distribution are:

k = 1	For a confidence level of 68%
k = 1.64	For a C.L. of 90%
k = 2	For a C.L. of 95% [the requirement of this protocol]
k = 2.58	For a C.L. of 99%
k = 3	For a C.L. of 99.7%

These coverage factors correspond to the Students t statistic, for an infinite number of degrees of freedom. In cases when the number of samples is significantly fewer than that which may be considered "infinite" (fewer, say, than 30 observations) then the t-statistic with the appropriate degrees of freedom should be used to scale the combined uncertainty.

2.2. Frequency Domain Analysis

In performing the discrete signal analysis using FFT some assumptions are implicitly made:

- The N digital data is decomposed into N linear components, each at a frequencies going from 0 to N/T_0 (T_0 being the length of the time series).
- The Nyquist frequency (corresponding to the $(N/2\text{-}1)^{th}$ component, which roughly half the sampling frequency) sets the limit of what frequency components can be resolved. In others words, the first half of the linear components represents the 'true' part of the spectrum, while the second half contains the folding components (aliasing). Thus only phenomena with a frequency lower than the Nyquist frequency can be detected by the FFT analysis. Any phenomena at higher frequencies will be 'folded' back into the true components, which will then be contaminated. Another way of expressing the Nyquist frequency criteria, is to say that at least two data points are needed to describe an oscillation. This is enough because it is already assumed that the oscillation is harmonic.
- The signal is assumed to be periodic with the period T_0.
- The data is describing a stationary process.
- The signal is assumed to be a sum of harmonics.

From the above is follows that the obtained 'true' (single sided) spectrum will have $N/2$ components, which corresponds to a frequency width of $1/T_0$. However, it can be shown that the spectral components in this case will have as large a variance as the value of the component itself

$$\mathrm{Var}\left(S'\left(f\right)\right) = \frac{S\left(f\right)}{\sqrt{P}} \qquad \text{... 6}$$

P being number of estimates of $S'(f)$ (subtimeseries)). To lower the variance on the individual spectral components more estimates are needed per frequency. This is normally obtained by division of the original timeseries into a number of subtimeseries, performing the FFT on the individual subtimeseries, and then averaging the spectral estimates for the individual frequencies. (An alternative can be to simply averaging neighbouring frequency components, but this is very inefficient in terms computational effort.) Hereby, the variance on the individual spectral estimate is reduced, but at the 'expense' of frequency resolution. Using the above given expression for the variance dependency on number of subtimeseries, it can be seen that using e.g. 30 subtimeseries reduces the variance on the individual spectral estimate to 18.3 %, which for most purposes can be considered acceptable.

Due to the need for division of the timeseries into several subtimeseries (windowing), together with the assumption of periodicity of the signal, there will be a need for tapering of the signal within each end of the subtimeseries (which will force the subtimeseries for start and end at the same value). Failing to apply tapering will generally give raise to high frequency noise in the spectra. When apply tapering the tapered parts of the signal will be given less 'importance' compared to the non-tapered parts. Therefore, overlapping of the subtimeseries should be applied, so that the tapering up of the next subtimeseries is started at the same time step as the tapering down is initiated for the previous on.

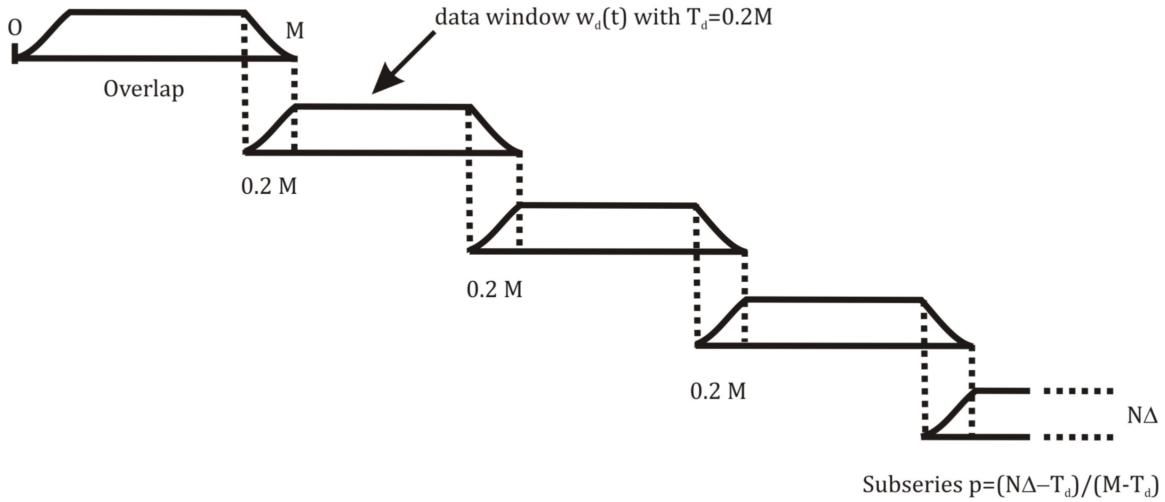


Figure 5: Illustration of windowing, tapering and overlapping.

The needed accuracy of the spectral estimates depends on for what the spectra is to be used. Often the characteristic parameters derived from the spectra are based on spectral moments of n^{th} order. The higher order moments the higher sensitivity to the accuracy of the individual spectral estimates. Thus, calculation of the spectral estimate of the significant wave height (H_{m0}), which is based on the 0^{th} (zeroth) moment (which is actually just the variance of the signal), is insensitive to the uncertainty of the individual spectral estimates. On the other hand, if looking at parameters based on higher order moments, such as spectral width parameters, or identification of peak frequencies (and thereby periods), definition of transfer functions etc. a larger accuracy is needed. This is ultimately going to define the needed duration of each test.

2.2.1. Time domain analysis

When performing a time domain analysis of a discrete signal, no assumptions have been made on the shape or characteristics of the signal. The needed sample frequency will therefore in this case be significantly higher than what is derived from the Nyquist frequency based on the frequency domain analysis. In order to obtain a reasonable description of a waveform signal (including water surface waves), a resolution of at least 20 data points per wave is needed. For other signals, the demand can be higher, e.g. monitoring of impulse pressure loads etc. Failing to use sufficiently high sample frequency will mean failing to identifying the ‘real’ crests and troughs of the acquired signal.

Needed duration of test linked to how well the ‘tail’ of the distribution function should be described. For tests in irregular waves, a min. of 500-1000 waves should be used in order to have a fair representation of the extremes (typically interesting for design). If interest is only in average power production capabilities shorter durations (200-300 waves) can be sufficient.

2.2.2. Assessing Minimum Quality and Quantity of Data

Using the definitions of Expanded and Standard uncertainty, we can write the confidence limits of the sample mean as:

$$\begin{array}{ll} 90\% & mean \pm 1.64\, u_c \\ \mathbf{95\%} & \mathbf{mean \pm 1.96\, u_c} \\ 99\% & mean \pm 2.58\, u_c \end{array} \qquad \text{... 7}$$

Because the standard uncertainty can be related to the number of samples, it is possible to define the number of samples which are required to achieve a particular confidence interval. Based on results generated during the test programme, or from prior experience, it is possible to determine the total number of samples which are required to achieve a particular confidence interval. Defining the interval over which the confidence limits hold:

$$\begin{array}{ll} 90\% & u_c = Interval / 1.64 \\ \mathbf{95\%} & \mathbf{u_c = Interval / 1.96} \\ 99\% & u_c = Interval / 2.58 \end{array} \qquad \text{... 8}$$

the value of u_c can be calculated and used along with the sample standard deviation to estimate the required sample size:

$$n = \left(\frac{s}{u_c}\right)^2 \qquad \text{... 9}$$

In the situation where the data can be analysed as it is collected, or where the deviation in conditions between runs is known to be very small (for example in flume facilities), then the required quantity of data is that which would yield a statistical stationary dataset. Suggested methods of analysing the stationarity of a time series are:

- Comparison of the mean of the first half of the data with the second half;
- Autoregression on the running mean (to observe the running mean at instant k and compare it with that at k-1. If the difference is less than some arbitrary ε then say the mean has converged.)

An example of this would be data collected from a flume via an ADV (acoustic Doppler velocimeter) probe over a period of a few minutes, sampled at 50Hz.

Figure 6 illustrates a slow forward run in a towing tank. In this case the tank was heavily seeded with backscattering material in order to improve measurement quality from the acoustic probe. Only the section of the trace where the towing carriage is up to speed is shown. Sample mean is 0.4606 ms^{-1}. Mean values for the first and second half of the trace are 0.4595 ms^{-1} and 0.4617 ms^{-1} respectively. Thus the corresponding errors from the sample mean are approximately 2.3% in both cases.

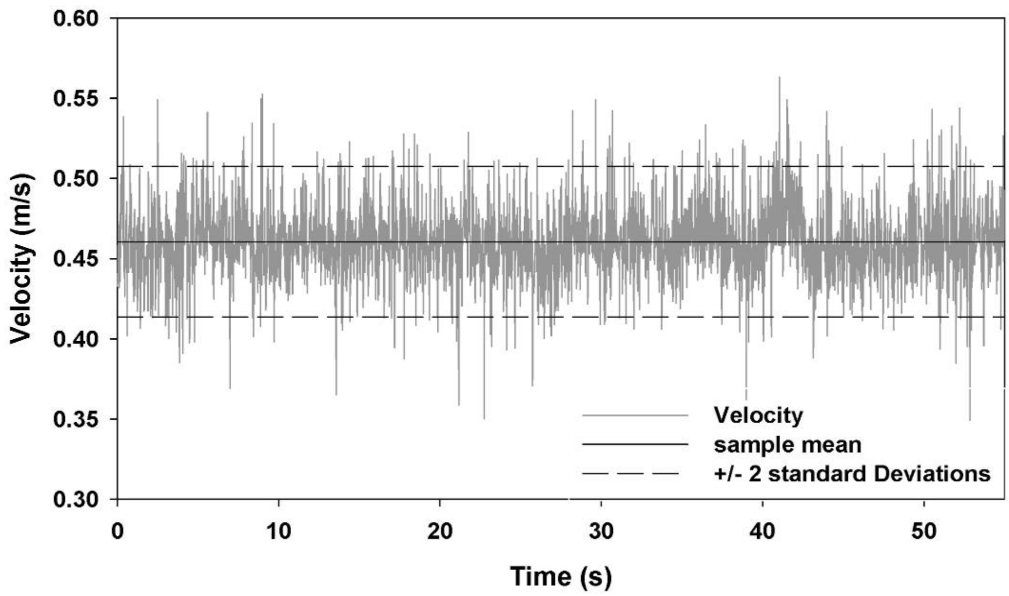


Figure 6: Velocity measurement in a towing tank facility

It is demonstrated that the steadiness of the carriage forward towing velocity was acceptable. In this case a further 9 runs were acquired and compared. Only the first run was deemed unacceptable, the most likely explanation was that it was the first towed run of that day. Therefore it is essential that all data sets are continuously assessed during testing.

In a similar manner rotor torque and thrust was measured at the hub of a horizontal axis tidal turbine. The turbine was advanced at a constant speed with varying blade tip speeds between sets of towed runs.

Table 1: Summary of example results.

Run #	***Measured mean Thrust (N)***	***Sample standard deviation (N)***	***Standard deviation between 3 data sets***
RPM 1	176.71	0.012	**0.85**
	176.75	0.011	
	175.26	0.017	
RPM 2	148.23	0.010	**1.04**
	149.85	0.010	
	147.91	0.010	

Table 1 illustrates the low variability of each run. For each separate run the sample variation was very low. Inter-sample standard deviation was slightly greater indicating that greater variation arising from this aspect of the testing. However, it was still well within limits which in this case were that the thrust measurement should be described within the interval ±5% with 95% confidence. Thus the stability of the performance facilitated a low number of towed runs (in this case 3) in order to accurately quantify performance at each set operational point.

It should be noted that in some cases variability was close to or exceeded 5%. In such cases data was collected until the variability fell within the set limits. Reasons for this could include increased unsteadiness due to high motion velocities on the hydrodynamic subsystem or resonance effects.

In cases where there is periodicity or other time dependency in the observed quantities – examples could be wave tank measurements or measurements on a tidal turbine rotor operating in yaw – then one of the following methods may be assumed in order to qualitatively determine whether sufficient data have been gathered:

- If the observed data are a function of an underlying signal which is known to repeat, then phase locking can be used and the averages of numerous cycles can be taken. It is up to the experimenter to decide if it is appropriate to describe each cycle as an individual test.
    - E.g. oscillating hydrofoil lift and drag histories phase locked with hydrofoil pitch and plunge oscillations.
- If the observed data are a function of a non-repeating signal synthesized, drawn or expected to be drawn from a specified or expected distribution where all bins of the distribution are expected to be represented an integer number of times, then frequency domain analysis, spectral moments should be calculated from zero to order 2, where convergence to 5% should be expected as per statistically stationary signal requirements for time series data.
- If the observed data are a function of a non-repeating signal synthesized, drawn or expected to be drawn from a specified or expected distribution as a truly random drawing, where the number incidences of a representation of a particular bin is either zero or a not necessarily integer multiple, then windowing should be applied to the signal and spectral moments should be calculated from zero to order 2, where convergence to 5% should be expected as per statistically stationary signal requirements for time series data.

3. Design of Experiment

The primary purpose of Design of Experiment in the context of EquiMar is the assurance that results generated from experimental tests have been derived in such a manner that the described response is satisfactorily linked to the explanatory variable(s). In other words, what is stated as having occurred due to an input *did* actually occur due to that input in a cause-and-effect relationship. This applies even if the precise physical mechanism is not fully understood.

Invoking the classic metaphor, this purpose is achieved in DoE by reducing the "signal to noise" ratio of the experiments, where the signal is the idealised measurement output and the noise is the corruption of this by nuisance factors. This is done by enhancing the signal and reducing the noise. Signal enhancing designs work by ensuring that the causality relationships are made as prominent as possible, and are built upon drawing these relationships out through the programme schedule (e.g. **factorial designs** and **Taguchi** designs). Noise reducing designs diminish the impact of extraneous nuisance factors by using insight into the sources of variability (e.g. **randomised** and **blocking designs**). In practise both signal enhancing and noise reducing designs are combined.

This document does not seek to outline all, nor necessarily the define optimal designs of experiment for a given situation. In general, the designs below assume a high level of factor orthogonally in the response and also (factor) linearity in the model. In situations where this is not the case or where certain combinations of factors are prohibitively expensive or hard to measure then more advanced methods such as Optimal Designs should be used. Descriptions of these can be found in NIST/SEMATECH e-Handbook of Statistical Methods and also in standard textbooks on statistics.

3.1. Choice of an Experimental Design

Based on the objectives of the test as described in Section 1.1. the following proposed experimental designs are suggested for all except *Proof of Concept* type objectives, where the test will proceed in an ad-hoc manner:

Table 2: Appropriate Design of Experiments methods by purpose and number of factors.

	Number of Factors			
	1	*2 - 4*	*5 - 50*	*>50*
Comparison	Randomised	Randomised & Blocking	Randomised & Blocking	Randomised & Blocking
Screening	Not Applicable	Full Factorial or Taguchi	Taguchi	Random Design
Response Surface Modelling	N/A	Full Factorial or Taguchi	Screen to reduce factors	Screen to reduce factors
Model Fitting	Randomised	N/A	N/A	N/A

The choice of the DoE methodology is therefore related to the objectives of the test and the number of factors under consideration. As it is uncommon that *purely physical* experimental tests for marine renewable energy device performance evaluation have very large numbers of factors, the Random design is most useful in combined computational/physical studies or when there are large numbers of control system parameters which can be changed as required.

3.2. Description of Experimental Designs

The following are some elementary DoE methods. The list is by no means exhaustive but these designs have been outlined since they have proven effective in R&D. Even though these designs are proven to be robust, they are not supposed to replace common sense or experience.

3.2.1. Randomised Design

Rationale: By randomising the order in which a series of tests are performed rather than performing them sequentially, such that as parameter values are altered the value of the factor is assigned at random from the pre-determined test conditions, errors introduced due to drift in the measurement apparatus will not be masked by trends due to altering the input parameter. If drift errors are introduced into the results and masked by a trend, they are impossible to remove. If the test sequence is randomised, analysis such as a **run order plot of residuals** will provide indication of the presence of drift. In practice, all experimental designs should include an element of randomisation for this purpose.

Process: Each case in the test series should be assigned a run-order number drawn at random from e.g. a table or random numbers, a random number generator or a hat.

Example*: choosing the order of angles of attack when performing a lift/drag test on a hydrofoil section at random will allow identification of possible drift in the tunnel speed due to the facility's motor temperature increasing over the course of the test. The test matrix is then of the form:*

Table 3: Randomised Design of Experiment Test Matrix

α [°]	*11*	*4*	*3*	*8*	*7*	*6*	*9*	*3*	*2*	*1*	*10*
Run order	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*

3.2.2. Blocking Design

Rationale: In cases where there are factors which are not of primary interest under the objectives of the test, but must be included as they have been identified as having significant, albeit secondary, effects on the experimental outcome, these nuisance factors can be accounted for by using them as blocking factors.

The nuisance factors can take on values which are:

- **continuous**, e.g. daily trends in outdoor air temperature or barometric pressure. For example the outdoor air temperature and pressure might be important in determining cavitation and free surface effects during a marine turbine test.

- **piecewise continuous**, e.g. day of the week or time of day. For example the first run on Monday morning may not be performed optimally, and similarly runs before lunch or the final runs on Friday afternoon may be performed with undue haste.
- **discontinuous**, e.g. due to facility/equipment or operator. For example facilities and equipment differ in capability, even when calibrated, in terms of the quantity, type and quality of measurements available. Operators might read scales slightly differently, or have a predilection or bias towards certain parameter values (e.g. determining how long it takes for a towing tank to settle).

The basic rule of research methods applies here: reduce or randomise variation in nuisance factors as far as is possible, manipulate through blocking what is difficult or expensive to control.

Accounting for these requires that homogeneous blocks of test runs are made in which nuisance factors are held constant, but in which the primary variable(s) can be allowed to change according to a randomised design internal to the block. Blocking allows analysis of the results to account for the effects of the most important nuisance factors while the internal randomised design accounts for the less important variables.

Process: Blocks within the test programme are established (in an externally randomised manner) in which the blocking factor value is held constant. The blocking process requires every level of a primary, non-nuisance factor to occur the same number of times for each level of the blocking factors. Inside each block the variables of interest are allowed to vary according to an internal randomisation.

Example: *Consider the hydrofoil in the example in 0. Due to the level of uncertainty, 3 repeat test runs are required, and the test programme is scheduled over an extended period of time such that three different operators are employed in the tunnel tests. In this example the nuisance factors are the operators. The test runs are then performed allocated to operators in "chunks" as follows (colour coded per operator):*

Table 4:Blocking Design of Experiment Test Matrix

	Operator 1	*Operator 2*	*Operator 3*
α [°]	*Randomised 0 ≤ α≤10*	*Randomised 0 ≤ α≤10*	*Randomised 0 ≤ α≤10*
Run order	*1, 2, 3, 4, 5, 6, 7, 8, 9, 10*	*1, 2, 3, 4, 5, 6, 7, 8, 9, 10*	*1, 2, 3, 4, 5, 6, 7, 8, 9, 10*

3.2.3. Full Factorial

Full factorial designs test all possible combinations of variables and as such are the most robust. It is assumed that every variable contributes significantly and all pairwise interactions are strong and important. The problem is that as the number of factors or the number of levels of these factors increases, the number of permutations increases very quickly: 4 factors at 3 levels equates to $3^4 = 81$ tests; 2 factors at 12 levels is $12^2 = 144$ tests etc. As such, in most situations a full factorial design may be prohibitively time consuming to perform, especially if multiple runs at each test condition are desired. If this is the case, Fractional Factorial methods such as the Taguchi method for orthogonal arrays should be considered.

Process: All combinations of factors and levels are listed and then performed in a random manner.

Example*: A turbine test designed to test 3 blade types each at 3 pitch settings and 6 tip speed ratios would require 54 tests, which may be at the boundary of what is feasible. The following test matrix illustrates this process:*

Table 5: Full Factorial Design of Experiment Test Matrix

Pitch / ID		*Blade Type A*			*Blade Type B*			*Blade Type C*		
	$\lambda 1$	$\theta 1$ / 1	$\theta 2$ / 2	$\theta 3$ / 3	$\theta 1$ / 19	$\theta 2$ / 20	$\theta 3$ / 21	$\theta 1$ / 37	$\theta 2$ / 38	$\theta 3$ / 39
	$\lambda 2$	$\theta 1$ / 4	$\theta 2$ / 5	$\theta 3$ / 6	$\theta 1$ / 22	$\theta 2$ / 23	$\theta 3$ / 24	$\theta 1$ / 40	$\theta 2$ / 41	$\theta 3$ / 42
	$\lambda 3$	$\theta 1$ / 7	$\theta 2$ / 8	$\theta 3$ / 9	$\theta 1$ / 25	$\theta 2$ / 26	$\theta 3$ / 27	$\theta 1$ / 43	$\theta 2$ / 44	$\theta 3$ / 45
	$\lambda 4$	$\theta 1$ / 10	$\theta 2$ / 11	$\theta 3$ / 12	$\theta 1$ / 28	$\theta 2$ / 29	$\theta 3$ / 30	$\theta 1$ / 46	$\theta 2$ / 47	$\theta 3$ / 48
	$\lambda 5$	$\theta 1$ / 13	$\theta 2$ / 14	$\theta 3$ / 15	$\theta 1$ / 31	$\theta 2$ / 32	$\theta 3$ / 33	$\theta 1$ / 49	$\theta 2$ / 50	$\theta 3$ / 51
	$\lambda 6$	$\theta 1$ / 16	$\theta 2$ / 17	$\theta 3$ / 18	$\theta 1$ / 34	$\theta 2$ / 35	$\theta 3$ / 36	$\theta 1$ / 52	$\theta 2$ / 53	$\theta 3$ / 54

The run order is made up of a random drawing of test ID numbers. Obviously there is a pragmatic balance between optimal randomness and time taken to alter each factor value, and in this case the blade type would appear a good blocking variable.

3.2.4. Fractional Factorial

In the example of Section 3.2.3, the total number of tests might become prohibitively large, especially when each test must be performed a number of times. The **Taguchi method** for orthogonal arrays allows a subset of the tests to be performed, under a number of assumptions, and still yields the important results. Taguchi methods assume few interactions between variables and only draws out pairwise interactions. It also assumes only a few variables contribute significantly, thus helps identify large effects more significantly. Also, the number of factor values is assumed to be low: for example two level designs have only "high" and "low" factor values, whereas three level designs also have a "centre" or "medium." Arrays *can* be found for higher numbers of parameter values.

Process: In the circumstance where the experimenter does not need to know how everything affects everything else it is expedient to not test all combinations but instead test 'edges' by finding pairwise combinations. In creating the test matrix every (most) 2 way combination of variables should be represented across all experiments. A table of experimental conditions, known as an orthogonal array, is determined based on the known number of factors and also the number of values each factor assumes. Tables are named "L#" where the number replacing the hash is the number of experiments which must be performed. For example, if there are 4 parameters each with 2 levels, a L9 table is appropriate (indicating that 9 tests will be required). If each parameter can assume 5 values, a L25 tables is required.

There exist numerous tables of Taguchi designs – small ones can be made by hand, larger ones can be constructed using statistical analysis software or found online. L9 = 4 factors each of 3 states.

3.2.5. Random Design

Random designs can be near optimal, however they do not tend to work well on small experiments (<50 factors) but work well for large systems. Random design assumes few interactions between variables and pulls out only a random sample of combinations, thus assumes very few of the many variables contribute significantly to the output.

Process: The process is very simple: choose number of experiments to run and assign to each variable a state based on a uniform sample of the variable values.

3.3. Replication and Choice of Factor Values

Given the implications of the Uncertainty Analysis, that is a diminishing uncertainty with increasing number of measurements, it behoves the test programme to allow some amount of repetition in measurements. Even if large quantities of time series data are collected at an experimentally stationary point, it is possible that unnoticed errors (especially operator/human error) render the reported value of that series incorrect, even though the reported uncertainty in the measurement is within the required bounds. To this end, it is advantageous to allow additional time in the test programme, not only for repetition of tests, but to determine the source of and correct for any disparity in results.

A suggested minimum number of repetitions is three. If the first and second tests do not match, then a third will be required by default. Where sample tests have been performed, it will be possible to calculate the number of data required by the method in Section 4.6.

The ITTC Procedure 7.5-02-01-01 [10] makes the following definitions about repeatability and reproducibility of experimental results. For results to be considered repeatable the following conditions must be met:

1. The same measurement procedure is used using
2. the same instrument under the same environmental conditions in
3. the same facility over
4. a short time period, i.e. the same day.

If one or more are violated, then the results may be said to be reproducible at best.

4. Uncertainty Analysis (UA) and Data Quality Assurance Procedures

The uncertainty associated with a measurement can be described under one of three categories:

1. **Standard uncertainty *u*** is made up of Type A (statistical) and Type B (non-statistical) estimates and are expressed as:

   Type A (see 4.3.): the standard deviation of the mean, equal to the standard deviation (positive square root of the variance) for a single test;

   Type B (see 4.4.): as the approximation of the standard deviation, calculated from the approximate variance determined from an assumed probability distribution.

   i.e. it is an expression of an estimation of the standard deviation of the measurement.

2. **Combined uncertainty u_c** brings the standard uncertainty of a number of measured variables together as, typically, an uncertainty in the expression of a derived quantity via the uncertainty of numerous factors propagated through data reduction equations.

3. **Expanded uncertainty *U*** qualifies the combined uncertainty by including a coverage factor k such that an interval defined about a reported measured value has a specific probability of containing the true value of the measurement.

The uncertainty is calculated as the root square sum of the contributing bias and precision uncertainty:

$$U^2 = B^2 + P^2 \quad \text{... 10}$$

In estimating bias, B, it is generally simpler to estimate elemental contributions than it is to attempt to identify bias through single sets of experimental results (if it is possible at all). Calibration simplifies this and allows instruments which would have a large number of contributing elemental biases to be treated as a single bias.

In estimating precision, P, it is often more cost and time effective to analysis the precision errors *en masse*, that is as if propagated through the DREs to some extent, rather than on an individual basis, and as such this protocol follows ITTC 7.5-02-01-01 [10] and recommends direct computation of the precision of the final result.

The approaches used to estimate standard uncertainty differ depending on whether the observations are consistent with a single test with multiple samples at a fixed test condition, or whether they are from multiple independent tests.

Single tests are those where the test duration is relatively instantaneous by comparison with the timescales associated with any of the experimental variables. In this situation the test can be assumed statistically stationary and the measurement of any variables over the relatively short duration of the test is considered a single test, even if the number of readings used to generate the mean is (significantly) greater than one.

Examples of measurements from single tests are those associated with single tow tank runs where a large number of samples are taken once the carriage is at the desired velocity.

Multiple tests are those which are performed in order to attempt to extend the test duration so as to capture sufficient variation of experimental factors. This is done by repeating the test over sufficient time periods to capture any long-period variation, or where multiple subsamples are taken from a single test run in order to capture any variation which is on a timescale of less than the test duration.

Examples of measurements from multiple tests are those associated with the means of the values measured during individual tow tank runs.

4.1. Methodology

The overall Uncertainty Analysis methodology can be split into pre-test and post-test phases as follows:

Pre-test

1. Write down Data Reduction Equations (DREs) and draw a data flow block diagram.

   Outputs:

   - a list of measurement systems and variables.

2. Perform a sensitivity analysis to determine where best to focus efforts.

   Outputs:

   - a table listing the sensitivity coefficients for terms in the DREs;
   - Partial derivatives of terms appearing in the DREs.

3. Estimate the bias and precision limits using information from the sensitivity study and any available Type A or B sources.

4. Modify and improve the test setup and programme so as to reduce estimated uncertainty below the required level.

At this point the test should be carried out. During the test it is up to the developer and facility operator to determine at which point sufficient data has been collected. Ideally, if the test progresses well according to the plan, then sufficient data will be collected if all the tests required by the Design of Experiment are performed, for example if during runs the standard deviations are sufficiently small and the tests are highly repeatable. However, it may be necessary to perform additional tests in order to reduce the uncertainty, and it is expedient to budget time for these from the outset.

Once the experimental data are collected, the following Post-test UA may proceed:

5. Perform a visual inspection of the data, identifying outliers, unexpected correlation, drift, autoregression etc., using the Data Quality Assurance methods of Section 4.7. **Error! Reference source not found.**;

6. Calculate the results of the experiment via the DREs, and apply any correction factors;

7. Determine the combined uncertainty for all factors in the experiment propagating uncertainties using the root sum square (RSS) methods of Section 4.2.

8. Express the uncertainty as an expanded uncertainty, with a coverage factor, the size of the uncertainty and state the level of confidence.

4.2. Propagation of Uncertainty

Once precision *or* bias[2] uncertainty components have been evaluated, by Type A and/or B means, the combined standard uncertainty is evaluated by using the **law of propagation of uncertainty**. Given a system modelled using a DRE of the form

$$y = f\left(x_1, x_2, x_3, \ldots x_N\right) \qquad \ldots 11$$

where *f* is a functional relationship describing some performance parameter based on measured quantities x_i in such a manner that any known covariance is captured in the equation. The combined standard uncertainty of y is linearised by considering the first 2 terms (but the constant is discarded) of a Taylor expansion of *f*:

$$\begin{aligned} u_c^2\left(y\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u\left(x_i, x_j\right) \\ &\equiv \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2\left(x_i\right) + 2\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u\left(x_i, x_j\right) \end{aligned} \qquad \ldots 12$$

The partial derivatives are the sensitivity coefficients, used when performing a sensitivity analysis (see 4.5. below) and can be evaluated analytically or numerically. The terms $u(x_i,x_j)$ are the estimated covariance and $u(x_i)$ is the estimated uncertainty. These values are determined using the methods below.

4.2.1. Monte-Carlo Simulation

Some situations might arise whereby a numerical rather than analytical (as above) approach may be undertaken. This might be because it is impractical to evaluate the propagation of uncertainties analytically or where it is desired to validate analytical results. In these situations, the Monte-Carlo Simulation (MCS) sampling methodology should be applied to the DRE. A basic MCS methodology is as follows:

1. Given a variable represented by a DRE of the form $y = f(x_1, x_2 \ldots x_i \ldots x_n)$, generate N samples for each of the measured variables x_i, randomly drawn from assumed or known (possibly joint) distributions, e.g. normal, triangular or rectangular, around a nominal operating point.
2. For k = 1 to N, evaluate the DRE to yield y_k
3. Calculate sample standard deviation for the N results y_k and hence $u(y)$

[2] Bias and precision uncertainty contributions should be calculated by independent root-square-sums

4. Confidence intervals may be found by sorting the list of y_k into ascending order and determining the required upper and lower percentiles.

The quality of MCS results is dependent on the number N of samples: a number between 10,000 and 100,000 is suggested, and given the low computing overhead, entirely feasible. The MCS method has the advantage that it can produce an output distribution which can be compared to experimental results. In addition, due to the nature of the method, limits will be constrained to physical (according to the DRE) values.

4.3. Standard Uncertainty - Type A UA

A Type A uncertainty analysis is one based on valid statistical method(s) applied to the data. In particular, precison components of uncertainty, P, are evaluated based on the number and scatter of measuremnts. The uncertainty associated with a Type A UA is simply represented by the estimated standard deviation, as given by equation 2 in Section 2

Type A analysis can also calculate bias, litterally the difference between the mean of a series of measurements and the expected mean.

4.4. Standard Uncertainty - Type B UA

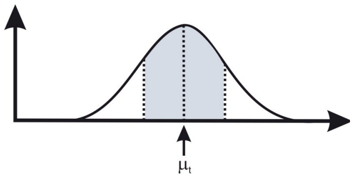
A Type B Uncertainty Analysis is one that is NOT based on statistical methods applied to the data. Type B Uncertainty Analysis relies on experience and judgement, and is thus more subjective than a Type A UA, and is reliant on assimilation and consideration of all relevant information, such as:

- Previous experience, either mathematical or physical, with the specific test subject;
- Previous experience or familiarity with the test procedure, facility, instrumentation etc.;
- Knowledge of existing test data including use of data from different facilities;
- Specifications and data provided by facility, instrument or test piece manufacturer;
- Calibration and specifications/certification requirements data;
- Any other data source, e.g. textbooks, handbooks, rules-of-thumb.

There are a number of means of performing a Type B evaluation, the most common being where uncertainty is evaluated based on reported values from an outside source, e.g. quoted by an instrument manufacturer, or determined by calibration of the facility. These are generally supplied either as a multiplier of standard deviation or a confidence interval.

Other means of obtaining a Type B uncertainty are by considering the data as being from an assumed distribution. Examples of this are shown in Table 6.

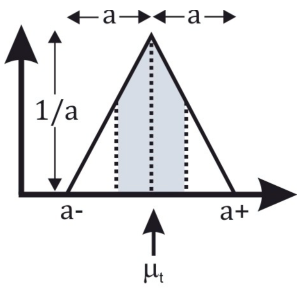
Table 6: Uncertainty Estimates from Assumed Distributions



Normal Distribution

Define a fractional region of the PDF such that the odds of the true value lying in the interval $\mu \pm a$ are

- **“50-50”** then $u \cong 1.48a$ (i.e. $a \cong u/1.48 = s/1.48$)
- **“2 times in 3”** then $u \cong a$
- **“99.73%”** then $u \cong a/3$

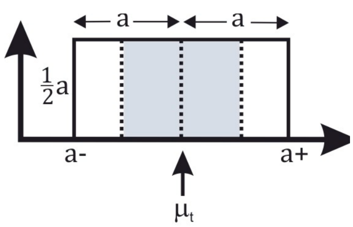


Triangular Distribution

If the true value is 100% contained within interval $\mu \pm a$ but there is a central tendency, then

- $u = a/\sqrt{6}$ where a is the **half-width** of the interval

Triangular is more conservative than normal.



Rectangular Distribution

If the true value is 100% contained and equally likely to fall (uniformly distributed or distribution not known at all) anywhere in the interval $\mu \pm a$ then

- $u = a/\sqrt{3}$ where a is the **half-width** of the interval

Applications are resolution uncertainty of digital displays and generating random numbers. Often used with information from calibration certificates/specidications. The GUM [4] suggests that this is used as the “worst case” distribution if the actual distribution of the data is not known, as the rectangular distribution is more conservative than both triangular and normal.

4.5. Sensitivity Analysis and Estimation of Bias Limits

Bias limits are the bounds of the bias component, B, of experimental uncertainty, and the magnitude can be estimated and included in the uncertainty statement, and in the expanded uncertainty via a root square sum. Bias limits are defined by the bound that the magnitude of the true value of experimental bias is expected to be less than, 95% of the time.

The methods and calculations used to perform a Sensitivity Analysis can be recycled to estimate the bias limits.

The process for evaluating a linearised approximation is as follows:

1. Write down the DRE in the form $y = f(x_1,x_2,x_3,\ldots,x_N)$ where f is a functional relationship describing some performance parameter based on measured quantities $x_1,x_2,x_3,\ldots,x_N$ in such a manner that any known correlations are captured in the equation.

2. Write down the total derivative of the function *f* with respect to some variable, then divide through by the differential:

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \frac{\partial f}{\partial x_3}dx_3 + \ldots \frac{\partial f}{\partial x_N}dx_N \qquad \ldots 13$$

At this point it may be convenient to use a direct calculation of the deviation due to small perturbations equal to tolerances as determined by Type B analysis, evaluated about a nominal operating condition, to yield the bias limits. If so, go to step 4, otherwise perform step 3.

3. Calculate an approximation for the by substituting in the finite differences $\Delta x_{1 \ldots n}$ then dividing through by the original expression *f* gives the expression for the fractional changes in *Y* due to small changes in $x_1, x_2, x_3, \ldots, x_N$:

$$\frac{\Delta f}{f} = \frac{1}{f}\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\Delta x_i\right) \qquad \ldots 14$$

4. At this stage nominal values (estimates or predictions of anticipated test conditions) should be tabulated, along with the estimated bounds, attained via tolerances associated with a Type B UA on the nominal values. Checks should be performed to ascertain how symmetric or otherwise the individual bias limits are, and it is suggested that if bias limits are asymmetric, then the limit with the maximum effect on the DRE should be used.

5. Once the coefficients have been calculated, a root-square sum of the form **of the law of propagation of uncertainty** provides the bias limits for *y*. Again, if the bias limits for *y* are found to by asymmetric, then it is the maximum bias which should be quoted and used in equation 14.

As an example, consider a proposed test on a model tidal turbine which will use the current through a resistor to measure turbine power. This gives rise to the following DRE for the power coefficient:

$$C_P = \frac{2I^2R}{\rho U^3 \pi r^2} \qquad \ldots 15$$

Following steps 2 and 3 above, taking the necessary derivatives gives

$$\begin{aligned}\Delta C_P &= \frac{\partial C_P}{\partial I}\Delta I + \frac{\partial C_P}{\partial R}\Delta R + \frac{\partial C_P}{\partial \rho}\Delta\rho + \frac{\partial C_P}{\partial U}\Delta U + \frac{\partial C_P}{\partial r}\Delta r \\ &= \frac{4IR}{\rho U^3 \pi r^2}\Delta I + \frac{2I}{\rho U^3 \pi r^2}\Delta R - \frac{2I^2R}{\rho^2 U^3 \pi r^2}\Delta\rho - \frac{6I^2R}{\rho U^4 \pi r^2}\Delta U - \frac{4I^2R}{\rho U^3 \pi r^3}\Delta r\end{aligned} \qquad \ldots 16$$

Which when divided through by the original expression (equation 15) gives the following simple form of the fractional difference in power coefficient:

$$\frac{\Delta C_P}{C_P} = \frac{2\Delta I}{I} + \frac{\Delta R}{R} - \frac{\Delta \rho}{\rho} - \frac{3\Delta U}{U} - \frac{2\Delta r}{r} \qquad \text{... 17}$$

Based on nominal values, the following table can be created using data from various Type B analyses:

Table 7: Sensitivity Analysis for a Tidal Turbine

		Nominal	Bounds	Source	Coefficient	$\Delta C_P / C_P$	ΔC_P	%
I	A	0.63	± 9.45 x 10^{-4}	Voltmeter documentation (± 0.015%)	1.5 x 10^{-3}	3 x 10^{-4}	1.2 x 10^{-4}	< 1
R	Ω	1,000	± 50	Gold band tolerance (± 5%)	0.05	0.05	0.02	5
ρ	kgm^{-3}	998.2	± 0.04464	ITTC	4.5 x 10^{-5}	4.5 x 10^{-5}	1.8 x 10^{-5}	< 1
U	ms^{-1}	1.0	± 0.01	Calibration documentation (± 1%)	0.01	0.03	0.012	3
r	m	0.8	± 0.0001	CNC machine tolerance (±0.01mm)	1.25 x 10^{-4}	2.5 x 10^{-4}	1 x 10^{-4}	< 1
C_P	**[1]**	**0.4**	**± 0.023**	**Root Square Sum**	-	**0.058**	**0.023**	**5.8**

By examining these results, it is apparent that maximum benefit in this case will be achieved by focussing effort on resistor tolerance and inflow (carriage) velocity measurement. The bias limits for C_P are ± 5.8%, and the assumption here is that they are symmetric.

4.6. Estimating Precision Limits

Precision limits are the bounds of the precision component, P, of experimental uncertainty, and the magnitude can be estimated and included in the uncertainty statement and the expanded uncertainty via a root square sum (RSS). Since precision errors present themselves as the scatter, due to random errors, of measurements around some biased mean, the precision limits give the interval in which the (biased) results are expected to fall 95% of the time under repeatability conditions.

The ITTC Procedure 7.5-02-03-01 suggests using a series of 5 sets of tests each with 3 speed measurements where the model is removed and reinstalled between each set of measurements as a means of determining the precision limits from the standard deviation. It is suggested that this is a suitable means of including random errors such as misalignment etc.

The precision limits for a measured variable obtained from a series of experimental tests with *n* observations are simply the experimental standard deviation of the mean, equation 4, multiplied by the coverage factor, *k*, and can be written as:

$$P(x) = k\frac{s(x)}{\sqrt{n}} \quad \text{... 18}$$

The value of *k* is determined from Student's *t*-distribution, using t-tables and is assumed to be equal to ≅ 2 (as per the GUM [3], NIST guidelines [4] and ITTC 7.5-02-01-01 [10]) for large sample sets where *n* tends to infinity (in practise $n > 30$) for a 2-tailed t-distribution and 95% confidence. This equation assumes a normal distribution and that the central limit theorem applies. For further information, consult the GUM**.**

4.6.1. Single Test with a Single Sample

The worst case scenario is precision limits for single tests with a single sample. These are impossible to estimate using this method since *k* is not defined for zero degrees of freedom. This might arise, for example, in the case where a single measured value for swept area is used to calculate turbine performance using time series data from a towing tank, or when time series data are from such a small interval that factor variation is not adequatly represented at all. In this case a methodology similar to that applied to estimate bias limits (Section 4.5.) is used to estimate the first order Taylor expansion:

$$P(x_i) = \sqrt{\sum_{i=1}^{J}\left(\frac{\partial x_j}{\partial x_i}\cdot P(x_j)\right)^2} \quad \text{... 19}$$

This now becomes a Type B uncertainty analysis, with estimates for the precision limits of measured variables coming from sources on the list in 4.4. , e.g. instrument precision error.

In single tests where some or all measured variables are available as averages over the test period, then precision limits in equation 19 from the subsample approach to the multiple test analysis should be considered, if the test period is sufficiently long.

4.6.2. Multiple Tests

With multiple tests (where *s* is determined using the pooling formula equation 5) or multiple subsamples from a single long (relative to factor variation) test, the number of observations *n* scales as the number of tests. *n* is thus greater than 1 and the experimental standard deviation of the mean and hence the precision limits of the estimated value decrease with the inverse square of the number of measurements. Thus to reduce precision limits by half (i.e. halving the uncertainty in the absence of any uncertainty contributions due to bias), four times the number of tests are required. In order to reduce the precision limits to 1/10th their original value, 100 times the number of tests are required. This clearly has the potential to be extremely expensive, and as such it is up to the developer and facility to determine the cost compromise between additional tests and upgrading the experimental process or instrumentation.

4.6.3. Standard Forms Table

The following table lists the combined uncertainty for some standard DRE forms and lists the RSS in terms of coefficients $\partial x_i/\partial x_j$ where the terms u are synonymous with the uncertainty in the subscripted factor. The approximations are exact if there is no correlation.

Table 8: Combined Uncertainty for some Standard Forms of DRE

Addition/Subtraction $f(x_1, x_2, \ldots, x_n) = \sum_i^n x_i$	$u_c^2(f) = \sum_i^n u_{x_i}^2 \approx u_{x_1}^2 + u_{x_2}^2 + \ldots u_{x_n}^2$
Multiplication/Division $f(x_1, x_2, \ldots, x_n) = \prod_i^n x_i^{p_i}$	$\left[\frac{u_c(f)}{f}\right]^2 = \sum_i^n p_i \frac{u_{x_i}}{x_i} \cdot \sum_j^n p_j \frac{u_{x_j}}{x_j} \approx \sum_i^n \left[p_i \frac{u_{x_i}}{x_i}\right]^2$
Reynolds Number $\mathrm{Re} = f(\rho, V, l, \mu) = \frac{\rho V l}{\mu}$	$\left[\frac{u_c(\mathrm{Re})}{\mathrm{Re}}\right]^2 = \frac{(u_\rho)^2}{\rho^2} + \frac{(u_V)^2}{V^2} + \frac{(u_l)^2}{l^2} + \frac{(u_\mu)^2}{\mu^2} + \left(\frac{2u_V u_\rho}{V\rho} - \frac{2u_V u_\mu}{V\mu} - \frac{2u_\mu u_\rho}{\mu\rho} + \frac{2u_l u_\rho}{l\rho} + \frac{2u_l u_V}{lV} - \frac{2u_l u_\mu}{l\mu}\right) \approx \frac{(u_\rho)^2}{\rho^2} + \frac{(u_V)^2}{V^2} + \frac{(u_l)^2}{l^2} + \frac{(u_\mu)^2}{\mu^2}$
Froude Number $\mathrm{Fr} = f(V, g, l) = \frac{V}{\sqrt{gl}}$	$\left[\frac{u_c \mathrm{Fr}}{\mathrm{Fr}}\right]^2 = \frac{u_g^2}{4g^2} + \frac{u_V^2}{V^2} + \frac{u_l^2}{4l^2} + \left(\frac{u_g u_l}{2gl} - \frac{u_g u_V}{gV} - \frac{u_l u_V}{lV}\right) \approx \frac{u_g^2}{4g^2} + \frac{u_V^2}{V^2} + \frac{u_l^2}{4l^2}$

4.7. Data Quality Assurance Procedures

Once measurement data are obtained as a time series of values from an instrument, before a detailed analysis is undertaken it is good practice to perform a quality check to ensure the experiment is proceeding as anticipated. Graphical methods, for example the 4-Plot espoused by in the NIST eHandbook [7], provide quick and easy checks on data quality. The 4-Plot consists of plotting:

1. Run order plot	Plotting ***i* vs. *y_i*** provides an early visual indication of: • Data randomness or variation – vertical spread; • Underlying trends or drift – "flatness" and apparent gradients; • Possible outliers.
2. Lag plot	Plotting ***y_{i-n}* vs. *y_i*** determines if an observation is related to an observation *n* previous: • Indication as to whether there is an underlying function and; • Indication of underlying function form. • If there is an underlying form, indication of outliers.
3. Distribution histogram	Plotting ***y* vs. counts** will provide: • Indication of data mean (the centre) and spread and skew; • Indication of outliers; • Indication of what sort of (possibly multimodal) distribution the data follow.
4. Normal probability plot	Plotting a **normal distribution of theoretical values for ordered *y* vs. ordered *y* values** aids in: • Determining if the data are normally distributed and thus if the normal distribution is a good model for the data; • If the distribution has fat tails, long tails or skew;

The plots shown in Figure 7 are all basic data analysis methods, and as such guidance on interpretation may be sought from any standard data analysis or statistics textbook, or online at, e.g. the NIST eHandbook. They should be considered the minimum requirement in analysis a time series of data. Further analysis proceeds with analysis of variance, correlations and frequency characteristics.

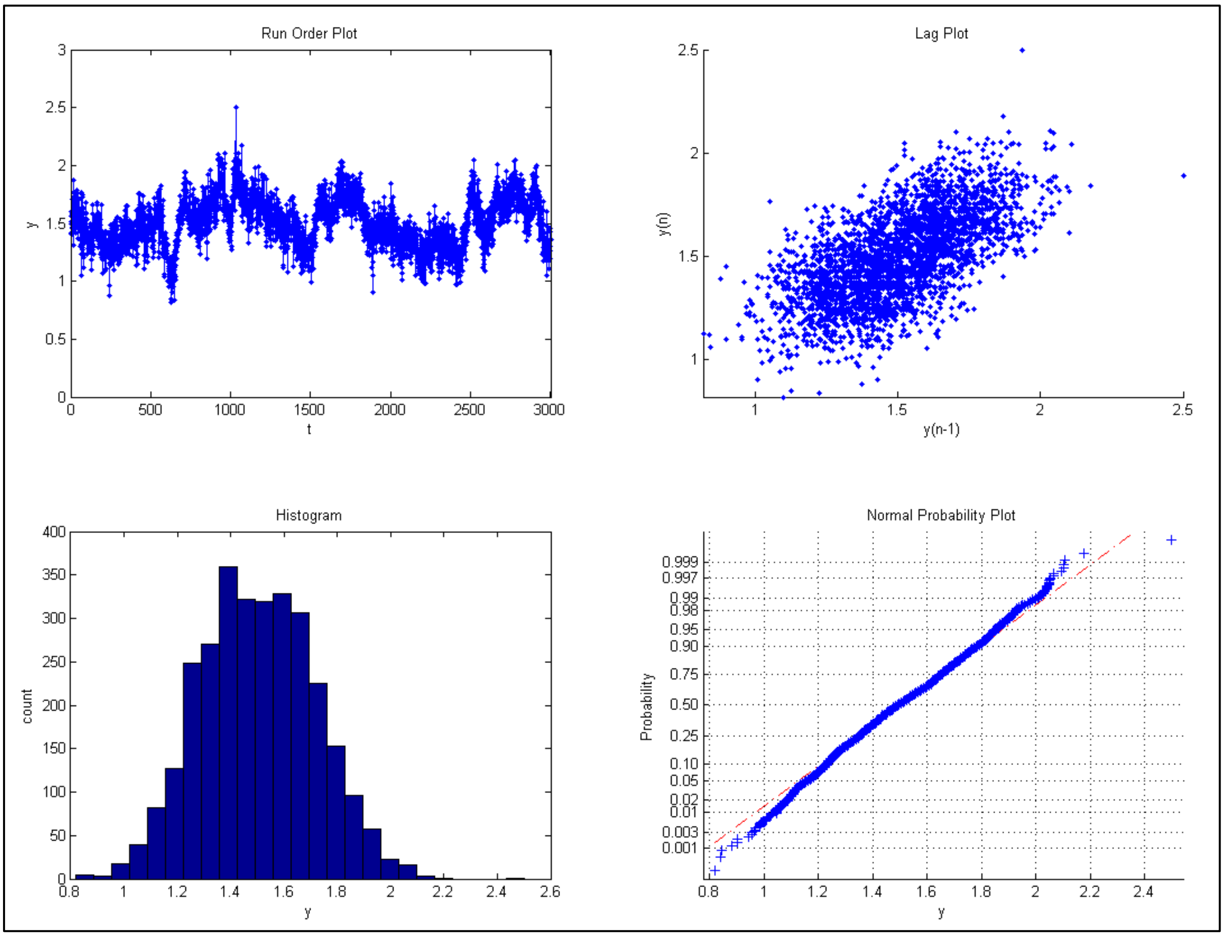


Figure 7: The 4-Plot for a data series.

4.1. Correlation

Scatter diagrams where the value of one factor is plotted against the value of another can be used to qualitatively demonstrate whether there is any correlation between the data. Since the methods Section 4.5. and the DREs in Table 8 simplify somewhat if the correlation between factors can be neglected (in other words there is no or very little joint variation between factors allowing various of the terms $\partial x_i/\partial x_j$ to ignored) it is suggested that as a minimum a matrix of scatter diagrams be used to ascertain whether further statistical evaluation of the correlation is required.

The following image (Figure 8) illustrates the process. The matrix is symmetric (in the diagonal) and the results are plotted in the lower region and the correlation coefficient is shown in the top region. The diagonal shows a correlation coefficient of 1, since the results are plotted against each other. Correlation coefficients magnitude shows the strength of the pairwise correlation, with the plot corresponding to 0.96 demonstrating high correlation thus the $\partial x_i/\partial x_j$ cannot be neglected, whereas the low correlation (0.3) would allow $\partial x_i/\partial x_j$ to be justifiably ignored.

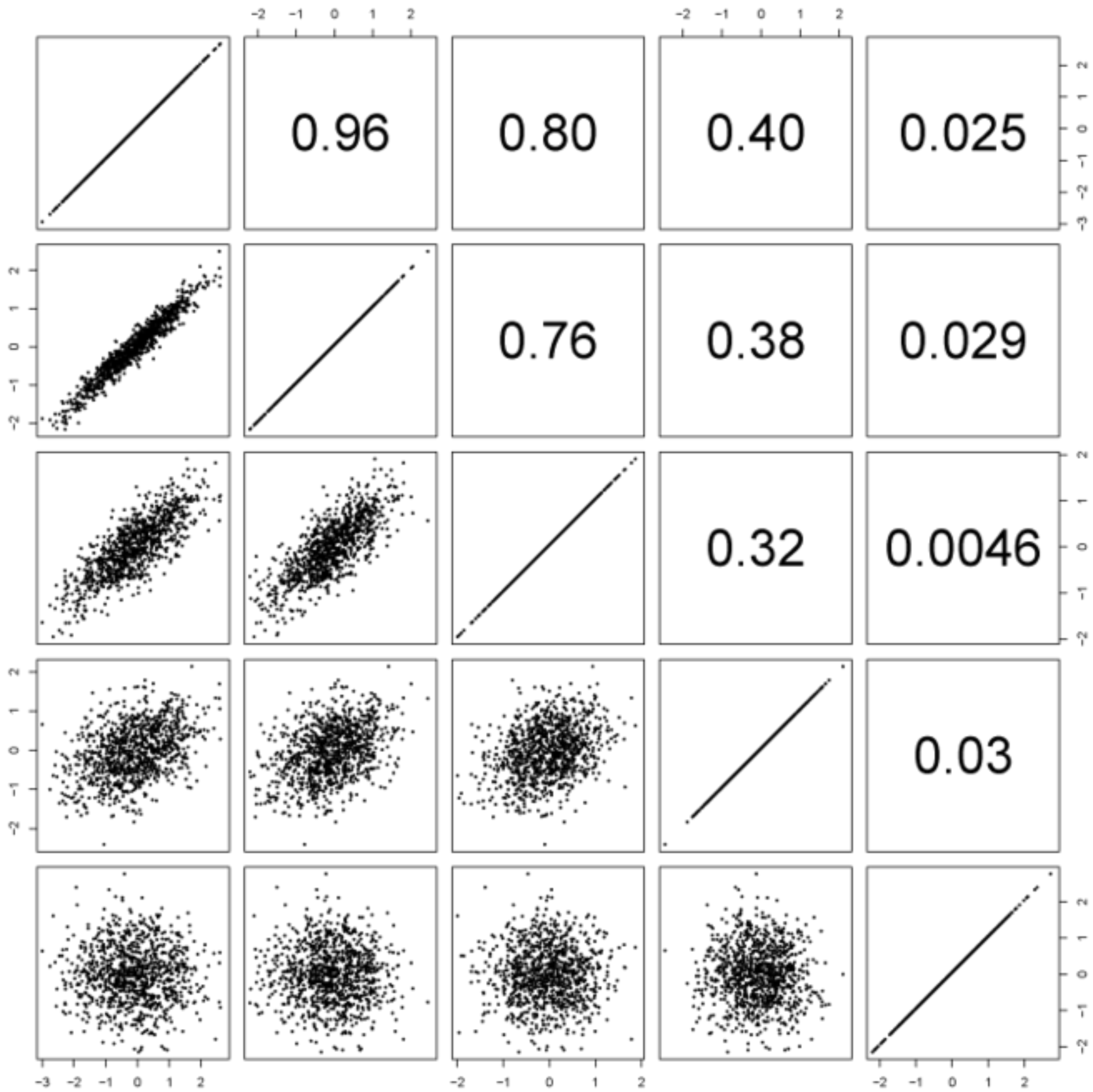


Figure 8: Linear Correlation Example. Taken from Wikimedia.

4.2. Outliers

Outliers are data that have been recorded which fall outwith the *expected* range of values. They can be incorporated within experimental results by means of hardware glitches or errors, which can produce impulses, spikes or short time constant responses, which in general saturate the measurement system. Examples of this are poor electrical connections and electromagnetic interference or instantaneous loads associated with slamming waves or slippage in a mechanical fixing.

Outliers can also exist due to inaccuracies in the assumption of the expected data distribution. For example, a distribution can have "fat"-tails by comparison with the normal distribution, and therefore it is an error in expectation rather than in measurement. It is important to properly inspect outliers to determine if they are a member of the underlying population distribution, which may differ from that expected, or a member of some other population.

Noise can also present itself as outliers. In this situation it is suggested that the source of the noise be identified and remediated rather than excessive smoothing/filtering be applied. This will reduce the possibility of noise masking outliers due to unexpected fat-tailed distributions. Remediation can be accomplished by the systematic removal of experimental apparatus until the source is identified.

4.2.1. Identification

The standard methods for identification of outliers are:

- Visual inspection and rule of thumb;
- Student's t-Test;
- Chauvenet's criteria;
- Grubb's test.

These methods are well established and simple to apply, and are also facilitated in standard data analysis packages. Care should be exercised when removing data: the source of the outliers should be sought and an explanation for their presence and justification of their removal offered when reporting.

5. Calibration [Of Sensors]

Calibration is the process by which systematic errors leading to bias uncertainty are identified and removed from measurement equipment. Once equipment has been calibrated, there is always bound to be a random error associated with using the equipment to make a measurement, and the random error of a calibrated instrument is as likely to be positive as negative.

5.1. Force

Static

Following ITTC Procedure 7.5-01-03-01 [12], section 5, force calibrations on e.g. load cells or strain gauges are generally performed under static conditions, in the dry using traceable masses. The relation between force *F,* gravitational acceleration *g* and mass *m* is then given by

$$F = mg\left(1 - \frac{\rho_a}{\rho_w}\right)$$

Where the term in parenthesis is a buoyancy correction. The reference masses have been calibrated at a traceable laboratory. If force multipliers (levers) are used, then appropriate terms should be incorporated into this relationship.

The reference loads are then compared against the indicated loads and the characteristic equation of the instrument can be determined.

Dynamic

No guidance on calibration for dynamic loads is given in this protocol, excepting the requirement for sample rate.

5.2. Air Temperature, Density, Pressure and Viscosity

Temperature

Thermometers must have a traceable calibration route to a recognised calibration laboratory. If using a thermocouple, the calibration must be carried out against a calibrated glass thermometer.

Density

Air density follows from the perfect gas equation of state if temperature and pressure are known via:

Pressure

Air pressure is obtained via local airfields or weather reports from news stations or their websites. Local laboratory measurement requires a traceable barometer, adjusted correctly for altitude.

5.3. Speed and Velocimetry

Fluid Velocity

Portable electronic velocimetry equipment, specifically ADV (acoustic Doppler velocimeters), suitable for use in water is generally sealed and factory calibrated, therefore no calibration is possible.

Calibration of Pitot probes, as defined in ISO 3966:2008 [14] must allow for wall effects, unsteady flow or velocity gradients and turbulence, the effects of temperature and particle entrainment and the flow direction. The water density must be also known with the appropriate level of accuracy**.**

Tow Carriage Velocity

Tow-tank facility velocities are calibrated via a number of internal cross-checks:

- a trailing wheel with a magnetic or optical sensor;
- optical or magnetic sensors on components of the drive system;
- optical or proximity sensors at known displacements along the rail;
- carriage mounted immersed pitot static probe (gives relative fluid point velocity);
- laser displacement measurements against fixed points.

Accuracy of tow-tank velocities is generally very good, and a survey by the ITTC [13]Error! Reference source not found. found that in for all respondents tow-tank speed measurement accuracy is below 0.1% of the maximum speed. The majority of respondents also report calibrating their facility once or twice per year.

Blockage

For flumes and towing tanks, the inflow or carriage velocity can be used to provide the incident velocity to the data-reduction-equations for thrust and power, however, in doing so a conceptual bias error will be introduced. This is due to the possibility of there being velocity gradients across the power capture area caused by the presence of the device itself, known as blockage.

Correction for this follows from either a non-intrusive (optical) measurement with the device *in situ*, or if this is impractical via blockage correction factors for example that of Barnsley-Wellicombe [15] as modified and applied by Bahaj et al [16].

6. Experiments on Extreme or Rarely Occurring Events

Testing of the device performance in extreme conditions is a desired requirement in ensuring performance as a vessel – i.e. ensuring safe station keeping, seaworthiness, survivability and validating failure modes. Results will be required by investors, insurers and engineers considering underwriting development, deployment and operation and produced through due-diligence investigations, undertaken by commissioned third parties. For wave devices, station keeping data while operating in various sea states must be generated: combinations of wave height, grouping, steepness should be examined based on theoretical considerations (identified sensitivities to slamming loads, overtopping/shipping of greenwater, capsize/stability, etc.). Tidal energy converters require examination in various flow speeds and directions (as well as wake and turbulence levels if possible) beyond the design envelope. Tidal devices may also require an examination of wave/structure interactions if the technology will be significantly affected by waves.

Survival tests are therefore an essential requirement before progression to Stage 3. This document provides qualitative guidance and recommendations only and does not seek to provide guidance on what constitutes an extreme condition as these are highly site and device specific in terms of which particular combination of conditions produce the worst loading conditions.

6.1. Rarely Occurring Wave Conditions

Deterministic analysis of the effects of various extreme wave events requires the synthesis of wavefields with one or more extraordinarily high waves in an otherwise ordinary regular or irregular sea. In general, numerical results and Stage 1 experiments performed in a small tank will produce conditions which must be examined at Stage 2 to test sea keeping in specific, pre-determined wavefields. Metocean conditions corresponding to the proposed deployment site should be used if available with extreme conditions calculated as per Deliverable 2.2.

6.2. Rarely Occurring Tidal Conditions

Tidal devices at Stage 2 should be tested in a facility capable of generating fluid velocities greater than the appropriately scaled velocities of proposed test sites; large scale, high intensity turbulence should also be added and where possible wakes induced into the flow. It is appreciated that the scaling of turbulence is problematic and it is not entirely clear as to how this can be done in a rigorous manner – therefore this is an option that will provide qualitative behaviour only. Turbulence kinetic energy spectra generated should be broadly similar to the Batchelor spectrum (measured via LDV etc.) or a site specific spectrum if available.

6.3. Scale Recommendation

In performing tests of extreme conditions there is generally a major issue that a scale prototype will have been scaled for power performance measurements in seastates which are readily and manageably produced by a particular test facility. It is therefore unlikely that the same facility will be able to produce the large waves or currents required to adequately test the prototype at that scale. This being the case, it is recommended that as a developer moves from Stage 1 to Stage 2 testing, which will involve moving from a facility which can produce appropriate conditions for 1:50 testing to one which can facilitate 1:10 testing, that the small-scale prototype be reused in the large-scale facility.

6.4. Reporting

The reporting of tests in extreme conditions will be as per other tests, however any device failures, or observations of abnormal performance or behaviour will be noted and explained along with a brief description of whether the failure was due to erroneous scaling used for model sizing or the design being tested (i.e. relative to full scale).

6.5. Required Measurements

The following are a minimal list of the types of measurement for a generic device: some will not be applicable in a particular case.

6.5.1. Stability and Trim

For a device which floats e.g. many wave devices, or is acting as a vessel e.g. under tow during deployment, the metacentric height, device trim and water level(s) are of critical importance in determining stability. These data are used in verifying computer simulations of extreme operating modes.

6.5.2. Accelerations

Acceleration measurements are required alongside stability and trim data as inputs for the validation of computational predictions of device behaviour in various failure, survival and operational modes. Accelerometers and inclinometers should be placed carefully such that correct transformation can be made between the local and global coordinate systems via e.g. an Euler transform.

6.5.3. Displacements and Attitudes

Generally, device free-response should be measured using an appropriate system incorporating measurements in as many degrees of freedom as possible. For devices on the surface, optical systems can be incorporated using infra-red reflectors to determine the device motion. If the device is submerged alternative approaches must be sought, e.g. inclinometer or accelerometer traces.

It is also often desirable to test the device while restricted in some axis, for example in testing longitudinal or lateral stability in the absence of heave motion. Such tests require that the device be secured in such a manner that only the motion of interest is permitted, and therefore it is essential that the displacements and attitudes of the device are measured correctly before, during (if possible) and after the test.

6.5.4. Overtopping Volume and Frequency

In devices where overtopping is a possible or actual operating mode, wave probe measurements of overtopping height, or flow-meter measurements of discharge should be sought.

6.5.5. Impact Loads and Vibration

Slamming is accompanied by a very rapid spike in local pressure at the slamming location which peaks and falls very quickly. Therefore, in order to adequately measure this phenomena sample rates must be very high. Further complications are that the sensors themselves have resonant frequencies and as such special care must be paid in selecting appropriate transducers. Further guidance is provided in ITTC 7.5-02-07-02.3 [18].

Device natural vibration frequencies are also likely to be very high for stiff metal structures, and measurements must thus be taken at appropriate sampling rates.

6.5.6. System Dynamics

For devices composed of numerous interconnected reacting subsystems, e.g. pitching blades on a rotating hub, then the device dynamics must be recorded. This includes displacements, if any, from trim or operating points, any actuator loads, control system inputs etc. It is especially important to identify actuators which become rate-limited in non-standard operational modes, since any scaling of the control system implemented under the assumption that these actuators are performing fully will result in problems. Early identification is the key.

Operational modes simulating the effects of failed actuators or PTO should be performed, e.g. tests of infinite or zero damping for wave devices, or with runaway or stopped rotor for tidal turbines.

6.5.7. Waves

The EMEC Wave Tank Testing Standard [19] suggests the following wave conditions:

- For seakeeping tests high energy as well as short period, steep waves (approaching the breaking limit). A suggested minimum scaled time series of a 3 hour storm is advised (bearing in mind the recommendations of Section 2.2. A H_{mo}/H_{max} ratio of 1:1.8 – 1:2 is required.
- For extreme loading the device characteristics need to be determined over a range of seastates and appropriate extreme conditions based on these must be modelled. It is suggested that wave periods close to device resonance periods will produce maximum loads.
- For extreme motions of the device, wave frequencies according to some or many of the device natural frequencies as well as breaking waves should be tested.

7. Qualities of Prototype Models

Ideally, a prototype model would be dynamically scaled against the full scale device. However, limitations in the ability to dynamically scale all components, coupled to the cost of ensuring that all components are built to exacting tolerances lead to the following pragmatic guidance.

It is essential that the hydrodynamic subsystem is scaled appropriately, either by kinematic or dynamic means. Other subsystems can be substituted, simulated or removed, depending on the nature of the test and scale of the model. In general, best practice is to ensure that components other than the hydrodynamic subsystem do not hamper the operation of the hydrodynamic subsystem in unrealistic ways, but other than that no requirements on build quality are made. Therefore, once the components are fabricated, they can be measured, and the protocol requires the following geometric tolerances, after ITTC Procedure 7.5-01-01-01 [20] and ITTC Procedure 7.5-01-02-02 [21]:

- ± 1mm or ± 0.05% for hull, reaction subsystem etc. components
- ± 0.1mm for rotor diameter, thickness and chord
- ± 0.5% for pitch at each rotor radius

7.1. Power Take Off

At Stage 1, in order that the effects of power take-off (PTO) systems on hydrodynamic performance are adequately represented it is likely that the scale PTO be significantly different from both Stage 2 and prototype scale PTO systems. This is due to significant disparity in the scaling laws for power and geometry. The situation is often further complicated by non-transparent boundaries between hydrodynamic and PTO subsystems, and PTO and control subsystems.

At early stages, concept appraisal tests will not necessarily have a complete power conversion chain: i.e. the PTO may be simulated by mechanical or viscous means. If a complete PTO is present, then there will almost certainly be no attempt at connecting this into the electrical distribution network, power may simply be dissipated into a controlled load, or some other mechanism may be employed. There will be losses associated with every stage of the power conversion process. The number of links in the power chain may be large, depending on the complexity of the PTO process being adopted:

Incident hydrodynamic to

- ➔ Aero/hydrodynamic motion (conversion)
- ➔ Mechanical (conversion)
- ➔ Viscous [damper/orifice plate] (power dissipation)
- ➔ Head (energy storage)

Mechanical to

- ➔ Viscous [damper] (power dissipation)
- ➔ Electrical (conversion)
- ➔ Mechanical motion [linear ⇔ rotational, direction/sign change] (power conversion)
- ➔ Mechanical storage [spring, mass elevation] (energy storage)

Electrical to

- ➔ Heat [resistor] (power dissipation)
- ➔ Electrical [DC ⇔ AC] (power conversion)
- ➔ Electrical [DC ⇒ sink device] (power dissipation)
- ➔ Electrical storage [capacitance] (energy storage)

In general, the flow of energy will be uni-directional as the scale of losses decreases the further down the PTO process.

At each conversion, in addition to losses, there will generally be a change in the time response shape of the power signal. In other words, at each conversion there will be a relationship between the input power signal and the modified output power signal carrier. This relationship will, in general, be non-linear and potentially difficult to model mathematically. As such, for the purposes of Stage 1 & 2 tests (proof of concept and controlled preliminary power performance evaluation) it is critical that performance measurements *used to characterise device power performance (rather than whole system performance)* must be taken at the notional hydrodynamic subsystem/PTO interface. This can be defined as the point where the hydrodynamic power conversion takes place. In some cases this might not be obvious or measurements at this point may be impractical. Therefore, the best compromise between a meaningful and sensible power indicator, intrusion into the operation of the device and the requirements of fabricating and instrumentation of the test prototype should be sought.

8. Archival and Storage of Data

Data will be generated during the experimental tests in a number of formats, however it is anticipated that standard analysis packages will be present in most facilities. As such, it is likely that data will be recorded in file-types associated with those packages.

Occasions where this may not be the case are when data is dumped directly from the instrument onto some internal or external media (for example certain acoustic Doppler velocimetry devices, and also custom/bespoke experimental setups).

Data will be recorded in either binary (computer readable) or ASCII/UTF plain text (human readable) and the requirements for good practice in data storage are similar.

ASCII/UTF: The advantages of human readable formats is that the information contained within is available and (given sufficient metadata) usable without the creating software. "Off-the-shelf" human readable files can be created in formats such as comma/tab-separated values (filename.csv) which can be read and produced by most software, including Excel, LabVIEW and MATLAB.

Binary: The advantages of binary formats is a speed advantage in computer input/output and also that they can often contain the same information as a human readable file, but occupy significantly less storage space. Binary formats are essentially closed unless sufficient information is given in the metadata and as such are generally locked to the software that created them. For this reason **human readable data-storage formats are to be preferred**.

Metadata are lines of human readable information containing data such as the date, time, operator, equipment type and direct measurement scale (e.g. mV) and converted scale (e.g. kg). Metadata appear as a header in both binary and human readable files. Optimally, metadata should also be included as a footer, whereby missing or corrupt footer metadata indicate file truncation.

An example of metadata from the header of a file is shown below:

```
LabVIEW Measurement
Writer_Version 0.92
Reader_Version 1
Separator      Tab
Multi_Headings No
X_Columns      No
Time_Pref      Absolute
Operator       luke
Date    10/06/2010
Time    57:26.6
***End_of_Header***

Channels       4
Samples 100    100     100     100
Date    10/06/2010     10/06/2010      10/06/2010      10/06/2010
Time    57:27.5 57:27.5 57:27.5 57:27.5
Y_Unit_Label   Volts   Volts   Volts   Volts
X_Dimension    Time    Time    Time    Time
X0      0.00E+00       0.00E+00        0.00E+00        0.00E+00
Delta_X 0.01   0.01    0.01    0.01
***End_of_Header***
X_Value Thrust         Torque          Thrust (N)      Torque (Nm)     Comment
        -0.003687      -0.003687       -1.691090481    -0.065322579    10/06/2010 09:57:26\09
        -0.024714      -0.0211         -11.33539738    -0.3738287
        -0.003359      -0.004016       -1.540649017    -0.071151472
```

Figure 9: Example metadata from a LabVIEW file.

As a suggested minimum requirement the following must be stored in the metadata:

- (Local) Time and Date stamp, minimally for the initiation of the test, optimally for each record;
- Facility, Test Process and Operator;
- Software used (e.g. LabVIEW, MATLAB) and computer architecture and OS (Intel x86, AMD64, PPC; Unix/Linux, MacOS, Windows…) – especially important in binary files;
- Number of channels (streams), data sampling frequency and number of samples;
- Data type, for example integer (int8, int16…), float, double. This is also especially important in binary data files.
- If non-proprietary or in house software is being used to generate binary data files, metadata should include a field indicating the length of the metadata itself, in bytes, and also the “magic word” which is used to indicate the end of the metadata.

9. Documentation

It is recognised that in general research groups, companies and individuals will each have their own style, template or standard form of a report. The experimental results should be presented in a concise report whose structure can vary, where the following guidelines may be adopted to give a consistent "feel" or function as an aide-mémoire:

Executive Summary	• What the client wanted • What was done to achieve the client objective • What the most important results were.
Introduction	• What the client wanted and some background to the clients requirements, • What information was available to the project team, • What approach was agreed with the client, • What tools were to be used, • What key outcomes were expected from the work.
Analysis Method	• Facility information, including capabilities and limitations; • Description of mathematical models; • Description, diagrams and information on scaling of physical test pieces; • Description of DREs and statistical analysis. • Sources and characteristics of input/inflow conditions; • Description and diagrams of process, measurement systems, data-stream in block-diagram; • Outline of Sensitivity Analysis and Design of Experiment including test matrices; • Detailed description of error sources considered and methods of Uncertainty Analysis; • Description and rational of tools, apparatus, equipment and procedures; • Images that usefully illustrate important aspects of models, or other subjects of investigation. Avoid complex diagrams with illegible text or figures that are difficult to decipher. • Relevant construction information • Time periods simulated, monitored, etc. • Rational and description of any correction factors.
Results	• Results presented in a standard form for the device class (if it exists) as defined in sections ## and ## including the uncertainty expressed in the forms outlined in section **## above**. • Presentation of main outcomes and discussion of the meaning of the results in the context of the client brief.
Conclusions	• To what extent the results of the project have met the client's needs. • Any other significant findings that arose during the course of the project. • Guidance to the client as to how the results should be used or interpreted. • Recommendations for further work if appropriate.

10. References and Bibliography

[1] ITTC Quality Systems Manual Version 2008

[2] AIAA Standard — *Assessment of Experimental Uncertainty with Application to Wind Tunnel Testing* (S-071A-1999)

[3] ISO/IEC Guide 98-3:2008 Uncertainty of measurement -- Part 3: *Guide to the expression of uncertainty in measurement* (GUM:1995)

[4] NIST Technical Note 1297: *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*

[5] ISO/IEC Guide 99:2007 - *International Vocabulary of Metrology* (VIM)

[6] NIST Reference on Constants, Units and Uncertainty, http://physics.nist.gov/cuu/Constants/, December 2010

[7] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, December 2010

[8] EquiMar Deliverable 3.3 *"Assessment of current practice for tank testing of small marine energy devices"*

[9] ITTC Procedure 7.5-02-01-02 *"Testing and Extrapolation Methods, General Uncertainty Analysis in EFD, Guidelines for Resistance Towing Tank Tests"*

[10] ITTC Procedure 7.5-02-01-01 *"Guide to the Expression of Uncertainty in Experimental Hydrodynamics"*

[11] ITTC Procedure 7.5-02-03-01.1 *"Testing and Extrapolation Methods: Propulsion: Performance Propulsion Test"*

[12] ITTC Procedure 7.5-01-03-01 *"Uncertainty Analysis: Instrument Calibration"*

[13] ITTC Proceedings 2008, Volume 1 Section 4: Resistance Committee

[14] ISO 3966:2008 *"Measurement of fluid flow in closed conduits. Velocity area method using Pitot static tubes"*

[15] Barnsley, M.J., Wellicome, J.F., (1990) Final report on the 2nd Phase of development and testing of a horizontal axis wind turbine test rig for the investigation of stall regulation aerodynamics. Carried out under ETSU Agreement E.5A/CON5103/1746.

[16] Batten, W.M.J., Bahaj, A.S., molland, A.F., chaplin, J.R., (2007) Experimentally validated numerical method for the hydrodynamic design of horizontal axis tidal turbines Ocean Engineering, v 34, n 7, pp. 1013-1020.

[17] EquiMar Deliverable 2.2 *"Draft Protocols on Physical Environment Specification"*

[18] ITTC Proceedure 7.5-02-07-02.3 *"Loads and Responses: Seakeeping: Experiments on Rarely Occurring Events"*

[19] EMEC Marine Renewable Energy Guide: *"Tank Testing of Wave Energy Conversion Systems,"* 2009

[20] ITTC Procedure 7.5-01-01-01 *"Model Manufacture: Ship Models"*

[21] ITTC Procedure 7.5-01-02-02 *"Model Manufacture: Propeller Models: Propeller Model Accuracy"*