Spatial variability of wave fields over the scale of a wave energy test site

Submitted by Ian Gerard Ashton to the University of Exeter
as a thesis for the degree of

Doctor of Philosophy in Earth Resources
In July 2011

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Abstract

Accurate wave measurements are required for wave energy applications, including resource assessments and performance assessments. In response, wave data are measured from deployment sites, commonly using wave buoys or other point wave sensors. Spatial variability in the wave field will introduce inaccuracies to the analysis of data captured from a single point to represent a separate location or area. This thesis describes research undertaken to quantify the effect of spatial variability on the accuracy of direct wave measurements taken at a wave energy site. An array of four time-synchronised wave buoys were deployed, separated by 500m, in a location close to the Wave Hub wave energy test site in Cornwall, UK. These data were subject to close scrutiny in terms of data processing and quality control, which raised specific issues regarding data processing and the validation of wave data for a new measurement facility. Specific recommendations are made for data captured from this facility, and bespoke quality control routines were developed. This process minimises the possible contribution of errors to the processed data, which is observed to be of the highest importance when analysing simultaneous data sets, and provides a data set that is particularly suited to the examination of the spatial characteristics of ocean waves. The differences between simultaneous data demonstrated local physical processes to be causing a deterministic difference between the waves at the measurement sites, which contributed to a significant difference between the power statistics at different locations within the site. Instantaneous differences between measurements were observed to agree well with theoretical estimations of random error, based on sampling theory. The culmination of the research is a unique analysis of the spatial properties of ocean wave fields on the scale of a wave energy test site, of direct relevance to the development and monitoring of wave energy test sites.

Acknowledgments

Undertaking this PhD has been a challenging and difficult process but ultimately, incredibly rewarding. I have progressed a long way and learnt many things since starting out, and I would like to extend my gratitude to all those who have made it possible.

In particular, my primary supervisor, Prof George Smith, who's knowledge and enthusiasm have propelled me through the past few years. Furthermore, the team that Prof Smith has built around him are inspiring to work with, and have supported me throughout. In particular, my supervisors (both official and unofficial), Dr Dean Millar, Dr Lars Johanning and Prof Julian Wolfram, and colleagues, Dr Helen Smith and Dr Jean-Baptiste Saulnier, who have had patience with my frequent questions. In addition, I must say thank-you for the great deal of support which I received from friends and associates, particularly Ted Pitt, who has always had time for a discussion about waves which were invaluable for developing my understanding. At this point, I must mention my sister, Dr Christine Currie, and her colleague, Prof Russell Cheng of Southampton University, for taking the time to help a an oceanographer in his first major foray into the world of statistics.

The research included a significant amount of marine operations, and I have had to learn quickly (in at the deep end?!). This would not have been possible without the support and patience of technical support colleagues, including David Parish, David Raymond, Neill Wood and Andrew Vickers. I am continually impressed by their skill, knowledge and working attitude without which, no wave buoys would leave the docks, and if they did, they are unlikely ever to return. I would like to say they have made a sailor out of me, but they haven't, I have simply learnt to endure. I would also like to mention the support from Hans Ornes of Oceanor who provided advice and instructions in the most difficult of circumstances.

This research would not have been possible without PRIMaRE, which was supported by the European Regional Development fund and the South West RDA.

Throughout the process, I have benefited greatly from the love and support of my family.

It is difficult to put into words how grateful I am to my wife, Sarah, who's unwavering patience and support has helped me at every stage. Thank-you for helping me remain positive through the ups and downs and for picking up the slack when I was immersed in the thesis. Thank-you also to my daughter, Megan, who arrived during my studies and whose enthusiasm for life and learning is inspiring.

Contents

List of Tables

List of Figures

Nomenclature

λ	Wavelength (m)
λ_0	$= \frac{gT^2}{2\pi}$ Deep water wavelength (m)
h	Wave height (trough to crest) of individual wave (m)
h_c, h_t	Height of crest and trough above mean water level (positive or negative) (m)
a	Amplitude of wave above mean water level (m)
η	Instantaneous surface elevation, above mean water level (positive or negative) (m)
$H_s, H_{1/3}$	Significant wave height, mean wave height, h, of the highest 1/3 waves in a record (m)
H_{max}	Maximum height, h, measured in a record (m)
T	Wave period of individual wave T (s)
T_z	Mean zero-crossing wave period (s)
f	Frequency (Hz)
$S(f)$	Spectral variance density (m^2/Hz)
p	The number of averages used to derive a spectrum, either the number of segments, or the number of harmonics averaged
D	The duration of the time-series used in spectral analysis (s)
f_p	Frequency associated with the maximum value of $S(f)$ (Hz)
$a_i(f), b_i(f)$	Fourier harmonics
c_i	$= \sqrt{a_i^2 + b_i^2}$
$\theta(f)$	Direction from which a spectral wave component is propagating (^{o})
m_n	Spectral moment, of the order n
H_{m0}	$= 4\sqrt{(}m_0)$, Spectral approximation of $H_{1/3}$ (m)
H_σ	$= 4(r.m.s(\eta))$, H_s calculated using the rms value of the displacement time series (m)

H_N	Return value of significant wave height associated with return period, N, in years (m)
T_p	$= 1/f_p$ Inverse of peak frequency (s)
T_{m02}	$= \sqrt{m_0/m_2}$, Spectral approximation of T_z (s)
T_{mm10}	$= m_{-1}/m_0$, Energy period (s)
S	$= h/\lambda$ Wave steepness
S_s	$= H_s/1.56T_z^2$, Significant steepness (ms^{-2})
S_{m02}	$= H_{m0}/1.56T_{m02}^2$, Significant steepness (spectral estimate) (ms^{-2})
E	Total energy in a sea state (J)
P_t	$= \rho g \sum_{i=1}^{N} c_g(f_i) S(f_i) \Delta f$, Total power in a sea state (W)
P_{tot}	$= \sum P_t$, Total power in a data set, used for calculating a power matrix (W)
ρ	Density $\mathrm{kg/m}^3$
g	$= 9.81$, Acceleration due to gravity (ms^{-2})
σ	$= \sqrt{\frac{(x-\bar{x})}{n-1}}$, Standard deviation
μ	Mean
$c.o.v.$	$= \sigma/\mu$, Coefficient of variation
υ	Degrees of freedom of a spectral estimate, $S(f)$
$d.o.f.$	Degrees of freedom
$dofW$	$= \frac{\upsilon(\sum_1^N S(f))^2}{\sum_1^N S^2(f)}$, Effective degrees of freedom in a spectral parameter.
	$= \frac{1}{1+\frac{2}{\upsilon}} dofW$, Accounts for the bias in the effective degrees of freedom where discrete frequency bands are used to estimate spectral parameters.
z	A statistic defined such that it follows a Gaussian distribution
Z	$Z(X)_i = \frac{\hat{X}_{Ai} - \hat{X}_{Bi}}{\sqrt{\sigma(X_{Ai})^2 + \sigma(X_{Bi})^2}}$, A specific statistic used to evaluate differences between wave parameters
SST	$= \sum_0^i (y_i - \bar{y})^2$,Total sum of squares ($SST = SSR + SSE$)
SSE	$= \sum_0^i (y_i - \hat{y}_i)^2$, Sum of squares for error
SSR	$= \sum_0^i (\hat{y}_i - \bar{y})^2$, Sum of squares for regression
MS	$= SS/dof$, Mean squares (MS) is a sum of squares divided by the appropriate degrees of freedom.
R^2	$= \frac{SSE}{SST}$, Proportion of variability in a data set that is explained by a linear model

w	Weighting applied during weighted least squares, commonly the inverse of the error variance
ϵ	Error term in a given parameter
$N(\mu, \sigma)$	Indicates a Gaussian distribution with mean, μ and standard deviation, σ.
θ	Wave direction (^{o})
$G(f, \theta)$	Directional spectral variance density
$\Delta\theta$	$\approx spr_{Tp}$, The directional beam width over which $G(f, \theta)$ can be taken as constant (^{o})
spr_{Tp}	Directional spreading at the spectral peak (^{o})
f_0	Resonance frequency of a spring (Hz)
R_{lf}	Calculated as the ratio of S(f) values at low frequencies (< 0.04) to $S(f_p)$
r_{lf}	$= \frac{m_{0lf}}{m_{0filt}}$, Ratio of variance of low frequencies ($f < 0.04Hz$), m_{0lf}, and variance of the rest of the spectrum, m_0
U_{10}	Wind speed at $10m$ above sea level (ms^{-1})
$\tilde{x}$	$= gx/U_{10}^2$, Non dimensional fetch
D_{par}, D_{perp}	Distance separation of two measurement points in a direction parallel and perpendicular to the wave propagation direction respectively (m)
V	Tidal parameter calculated as the velocity component of the tidal flow along the principal direction of flow on a flood tide (ms^{-1})
F	A parameter expected to exhibit a F distribution
$F_{\sigma^2}(X)$	A specific parameter calculated from the measured or expected variance between simultaneous measurements of a given parameter, X

Chapter 1

Introduction

Ocean surface gravity waves comprise the highest energy density of all renewable energy sources (Clement et al., 2002). The UK is particularly well situated, with a national resource estimated as $58TWH/yr$, which has the potential to provide in excess of 10% contribution to the national electricity supply (Mueller, 2008). This is comparable with the highest resources available to countries worldwide, and the mean annual wave power in excess of $60kW/m$ is amongst the largest in the Northern Hemisphere (Cruz, 2008). The scale of the resource, which seems obvious to any casual observer of the sea during bad weather, has often led scientists and engineers to develop methods of converting ocean wave energy into electrical energy. To date, there exists a wide variety of devices designed to do this, with over 1000 patented devices in Japan, America and Northern Europe (Clement et al., 2002) and circa 90 devices currently in development around the world (Mueller, 2008).

Figure 1.1 describes progression of device development from conception, through test phases to demonstration at full-scale in the ocean. This is very much a nascent industry and only a handful of devices have reached the demonstration stage. Among these, Pelamis wave power, Aquamarine, OPT, and others have operated grid-connected devices, although a commercially viable project generating profit through energy production is not yet a reality.

Infrastructure requirements increase through the categories in figure 1.1 and full-scale, grid connected testing requires a significant outlay compared to previous test phases. Undersea cables can be costly, and shoreline grid connections may require re-engineering of the distribution network, especially in remote areas. Thus, connection costs for a new project will be high (Little, 2005). Furthermore, working offshore significantly increases deployment, repair and maintenance costs, and failures at sea are potentially very costly.

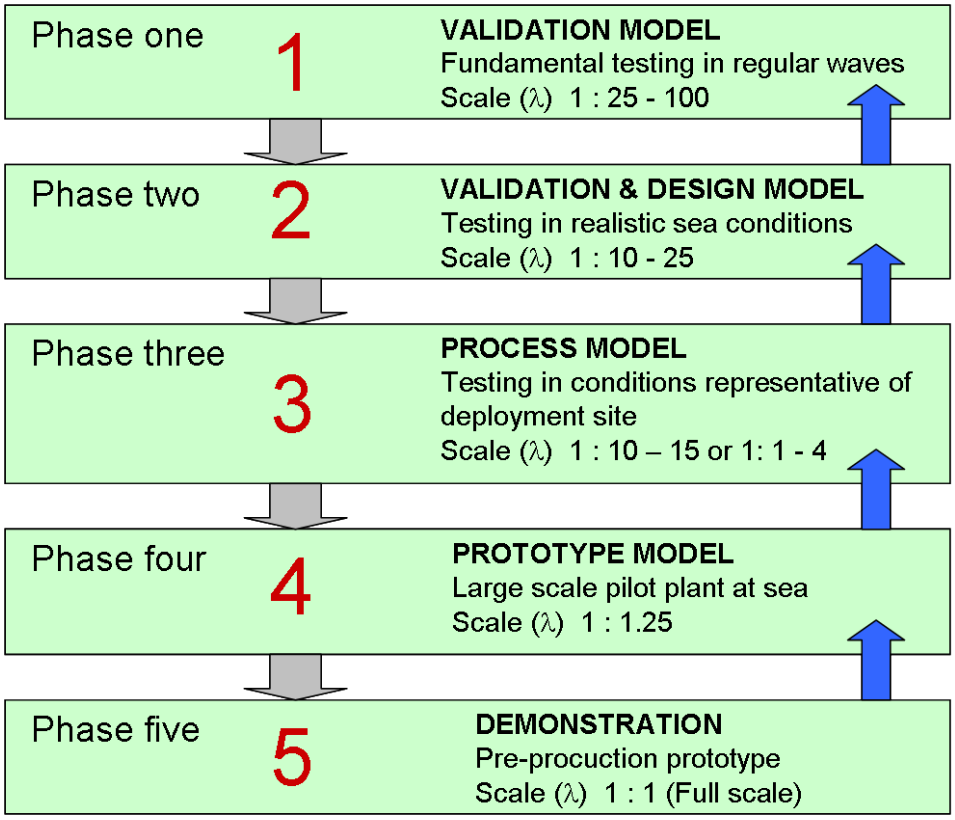

Figure 1.1: The development phases for WEC. Adapted from WavePlam (2009).

At prototype stage, revenue generated remains small and developers continue to rely on investment and governmental support to cover the costs of testing. Wave energy test sites backed with public funding have been developed to alleviate some of the costs of deployment and to support progression of the technology. Figure 1.2 represents the progression of prototype testing by a developer, where the blue ellipses show the facilities available in the UK for each phase of testing. Other countries around the world are also developing a framework of test-sites.

The European Marine Energy Centre (EMEC) has two test berths for full-scale wave energy converters (WEC) of capacity 1MW each, with a grid connection to shore. They also operate tidal energy test berths elsewhere on Orkney. Planning and permissions exist for installations at the site, and developers can access the marine renewables proving fund (MRPF) to gain government support (EMEC, 2010). Pelamis wave power have successfully tested their device there, and at the time of writing are testing their newer, P2 device. Aquamarine Power also installed their Oyster device installed on the sea bed during summer 2009, which was connected in November 2009 and continues to produce power to the grid.

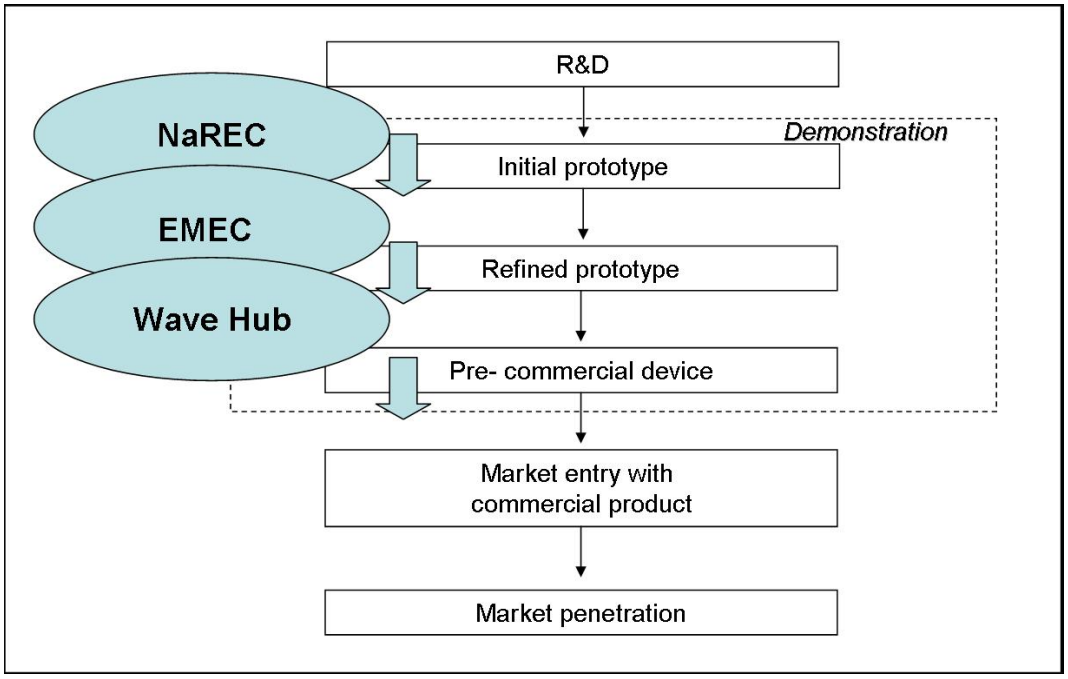

Figure 1.2: The phases of prototype testing and the test site facilities available in the UK to accommodate testing at each phase.

Much of the recent research and development (R&D) effort has been focused on developing installations which can provide large-scale energy production and make a significant contribution to a national energy framework. To achieve higher levels of power consumption, multiple devices will be deployed close enough to each other that their output to the grid can be combined. It is these arrays, or wave farms, that are commonly seen as the commercial future of the wave energy industry. Further description and analysis of the leading devices and their plans for development of wave farms can be found in reports reviewing the state of the industry (Previsic et al., 2004; WavePlam, 2009), and on their web-sites. With this next stage of progression in mind, the Wave Hub development was conceived as a test site to accommodate developers wishing to test arrays of devices. It has four sub-sea electrical grid connections of 1MW[1], each associated with a $1km \times 1km$ area of sea. It is located approximately $20km$ north of the town of Hayle on the North Cornwall coast, situated in a depth of $50m$. The mean annual wave power has been estimated as between $21kW/m$ and $25kW/m$ (Halcrow, 2006b). The cable lay was completed during October 2010, and all permits and consents are in place for deployment of WEC.

The Peninsular Research Institute for Marine Renewable Energy (PRIMaRE) was started at the same time, to provide research support to the Wave Hub. A key research stream

[1]Installed infrastructure has been designed to support an upgrade of capacity to 4-5MW each.

identified was to provide wave measurement and analysis in support of the assessment of a spatial site and testing WEC deployed as arrays. Wave measurements play a key role in the successful operation of a wave energy test site, and the research presented in this thesis addresses the issue of spatial variability in the wave field affecting the accuracy of site measurements. A key stimulus to this research arose from the simultaneous deployment of two buoys at EMEC, separated by $1500m$ and subsequently $600m$ (EMEC, 2010). Analysis of the recovered data highlighted that the observed differences in simultaneous wave measurements between these buoys were statistically significant, indicating spatial variability of wave fields across a wave energy site. It was speculated that this could affect resource and performance assessments conducted using data from point wave sensors applied to WEC positions separated from these measurements on various scales (Smith et al., 2006a).

Later, in the development of the Wave Hub test site, considerable public opinion demonstrated the importance of accurate estimations of the amount of energy removed from the propagating wave field by operating devices. With no operational data, early work lacks evidential support for generalisations made when predicting the magnitude and nature of wave energy absorption by WEC using modelling (Millar et al., 2007; Smith et al., 2007; Venugopal and Smith, 2007; Beels et al., 2010). Predictive models of wave/device interactions would benefit from validation using direct measurements of the changes in the waves across a wave energy site due to operational devices. A first step in this process is the quantification of the baseline wave climate where no WEC are in place, which the EMEC data have suggested may contribute significant variability on the scale of a wave energy site.

In response to these stimuli, four directional wave buoys were deployed at a locations close to the Wave Hub site, in a square formation of sides $500m$, with the aim of quantifying the variability of wave conditions on a scale representative of a multiple deployment of WEC. The exact scale of commercial wave farms is yet to be defined, however this research took the size of the Wave Hub berths as a guide, which measure $1km^2$. The results from this facility will support the design and planning of future sites in order to optimise the measurement and analysis techniques. In particular this will provide a case study of the accuracy of standard procedure for resource and performance assessments when applied to an array of devices, and a useful data source for the validation of wave models at these scales. In addition, it is possible to derive an independent assessment of measurement accuracy using multiple devices. This is a key piece of information to refine the

statistics used in assessment of the wave conditions for wave energy and all other offshore applications.

When developing a research data set, consideration must be given to ensuring a detailed understanding of data processing methods. This is most effectively demonstrated through standardisation of the processing with the wider research community and development of the data archive such that it is accessible and contains all relevant information for the use of these data in the future. Published guidelines for the capture and storage of wave data exist for the oceanographic community (CDIP, 2003) and for offshore engineering, (BSI, 1984; Tucker, 1993; Tucker and Pitt, 2001). More recently, standards have been developed in the UK for the wave energy industry specifically, (Smith et al., 2006a; Pitt, 2009) and the EQUIMAR project has published standards developed from a european basis that will feed into the formation of international standards (Ingram et al., 2011). The data collected from the array of four wave buoys is a new data set for which, quality control, data processing and data storage were developed according to these sources. Development of these processes in order to provide an accessible, clearly documented and high quality data set is essential to support future research. In particular, it is essential to ensure robust conclusions can be drawn from analysis of the measured wave fields by the wave buoy array.

The primary research aims of this thesis, using the data set developed, are to establish the magnitude of the random error term for verification of theoretical variability, identify the underlying processes contributing to deterministic differences and quantify their effect. These can be summarised in three key research questions for which answers are sought in this thesis,

- Is there evidence of deterministic differences between co-temperous data?
- Can a correlation be identified between a non-random signal of deterministic differences and any physical factor that may affect the spatial variability of the wave field at the measurement site?
- Do theoretical estimations of random statistical variability account for the observed differences?

Differences between the data measured at each buoy can be expected to be caused by three principal factors; random variability or error, unequal calibration of the measurements systems and deterministic differences between the waves at each buoy. A difference in the underlying wave field at two separate measurement sites, caused by different physical

conditions can be termed a deterministic difference. These are related to the spatial distribution of wave fields across an area, which will be caused by the interaction of waves and other physical processes such as the local bathymetry, tidal flow and the wind field. Evidence of deterministic processes in the data set will be manifest as differences that cannot be explained by random variability, and are correlated to other processes. Previous studies to quantify spatial differences using simultaneous measurements have been unable to rule out differences between measurement technologies (Sova and Wyatt, 1995). For this research, bias between data sets is minimised through accurate calibration of sensors, and by deploying the same measurement devices throughout the array, which is a clear benefit of the buoy array developed in the course of this research over previous data sets.

Importantly, a mathematical description of the interactive relationship between all contributing factors, along with sufficient data for these factors could predict deterministic differences, which would be of direct benefit to wave energy applications. In fact, various methods have been developed to perform this role, including computational wave models and data interrogation approaches such as multivariate regression, neural networks and genetic programming. However, all of these methods will involve simplification of complex interactions and their application on this scale will require an understanding of what level of simplification can be applied whilst still accounting for all factors having a significant influence on the measured data. Furthermore, validation and calibration of any such procedure requires robust estimations of the deterministic differences between wave measurements derived from measured data. This data set represents a new, unique opportunity for this analysis and research described in this thesis develops and applies statistical methodologies to derive estimations of both deterministic differences and their interaction with physical conditions at the site.

Random variability in wave measurements can be predicted based on sampling theory for a statistical description of ocean waves (see section 2.2), and it is these predictions that are applied to the analysis of wave data in the form of confidence limits or error variance. However, these theoretical predictions of random error do not include instrument errors and critically, do not have a reference data set on which to validate predictions. The deployment of multiple sensors offers an opportunity to quantify variability based on the differences between direct measurements (e.g. (Mackay, 2009)). Through consideration of the predicted sampling theory, the contribution to errors in the data from other sources can be quantified giving an absolute measure of variability. This analysis has the potential to validate theoretical predictions of random error, and provide a direct measurement of

the accuracy of the wave buoy measurement system.

Results will allow conclusions to be drawn regarding the accuracy of wave measurements when separated from the point of application, which is of direct relevance to resource and performance assessments. The differences between measurements at two buoys will also provide a useful indication of the differences that can be expected between two measurement sites with no interference from devices, which is indicative of environmental assessments to be performed. Quantification of the random error and the validation of theoretical errors will have implications for all analysis undertaken on wave buoy data, although it will be of particular relevance to this measurement technology. Critical to these results is ensuring that predictions are robust, and that observed differences can be attributed to differences in the wave field as opposed to errors in the data set. It was with these considerations that the research project was designed and implemented.

Chapter 2

Review

2.1 The description of surface water waves

A simplified surface gravity wave is demonstrated in figure 2.1 using a sinusoidal form. Linear wave theory shows that with certain assumptions, the equations describing a sinusoidal wave form propagating in water can be considered linear (Dean and Dalrymple, 1984). However, sea states encountered in nature do not follow this form, rather they are a complex combination of waves of various shapes and frequencies (fig. 2.2). If each contributing wave is assumed to be sinusoidal and the descriptive equations linear, then the surface of the water under a complex wave field can be calculated by the superposition of sinusoidal waves of varying frequency. This method is known as Linear Wave Theory (LWT), or small amplitude wave theory and a full discussion of the assumptions can be found in the major textbooks (Tucker and Pitt, 2001; Holthuijsen, 2007; Dean and Dalrymple, 1984). The latter name arises from a key assumption that the amplitude of each wave is infinitely small. This simplification, and others involved in linearisation of the wave equations, have been seen to closely represent observed waves under many conditions (Tucker and Pitt, 2001). However, when the waves become very steep, such as in shallow water, and in the presence of strong currents, their amplitudes become large and the simplifications , particularly the small amplitude assumption, become inaccurate. In these conditions a non-linear theory should be applied.

When the wave field is considered as a combination of a large number of contributing wave trains, the properties of individual waves vary randomly from one wave to the next (Ochi, 1998). In turn, the probability density function of η has a Gaussian probability distribution with zero mean and variance dependent on the energy in the waves. By considering the waves as a Gaussian stochastic process, the properties of a wave field can

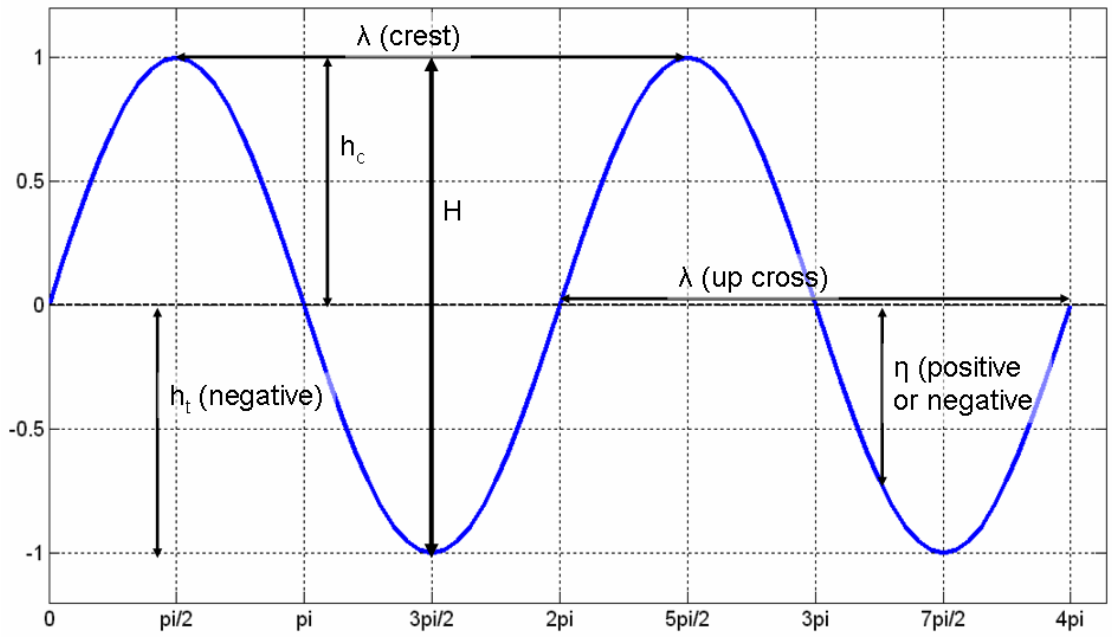

Figure 2.1: A sinusoidal wave, where h_t is trough height, h_c is crest height, H is wave height, λ is wave length, and η is instantaneous wave elevation.

be described by the statistical properties of such a process. For example, Longuet-Higgins (1952) showed that H will follow a Rayleigh distribution. Many aspects of wave statistics are based on this description and great care must be taken when analysing situations where the assumptions become inaccurate.

To estimate the statistics of an incident wave field, a sample of measurements that span either a length of time or an area are required, known as a wave record. In statistical terms, a wave record is a sample of a population, and calculations based on the sample will engender a certain amount of random error. In theory, the longer the wave record, the greater the accuracy to which the statistics can be estimated. However, accurate estimation requires that the statistical properties remain constant over the sample. The length of time (or size) of the record, must be long enough to allow an accurate estimation of wave statistics, whilst not seriously violating the assumption that the statistics remain constant within the record. Sampling variability is discussed in more detail in section 2.2.

2.1.1 Direct analysis

Direct time-series analysis involves the identification of individual waves from a wave record and the measurement of their properties. Today, it is not commonly applied to the routine analysis of principal wave parameters such as the significant wave height, H_s,

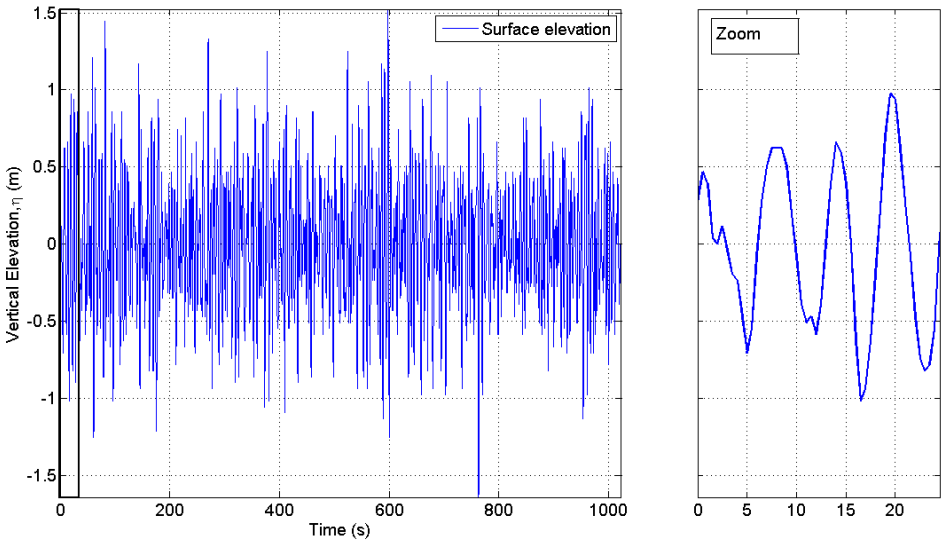

Figure 2.2: A sample time series of vertical displacements captured by a Seawatch Mini II directional wave buoy during December 2009. $H_{m0} = 1.7m$ and $T_Z = 6.5s$.

however it is the only method available for extreme parameters such as the maximum wave height H_{max}. It also finds use in quality control procedures as it is affected by errors in different ways to parameters derived from other methods. The process is described here for analysis of time-series of surface elevations, as derived from wave-buoys, pressure sensors or ADCP's.

Each wave is identified as a complete cycle, which can be defined between wave crests (as a 'crest to crest wave'), or between zero crossings (e.g. down crossing to down crossing). The dimensions and temporal properties of each wave can then be measured directly from the time-series (see fig. 2.1 for definitions). Summary parameters are calculated using the statistics (mean, max etc...) of the relevant property of all of the waves in a record. Definitions of parameters that are commonly calculated using direct time-series analysis can be found in the following chapter (tab. 4.1).

Identification of individual waves relies on the processing of the time-series accurately identifying zero-crossings. Tucker (1993) explains how the sensors digitising rate of the sea surface will affect the ability of the analysis to identify zero crossings, particularly in low sea states. A trend in the data will lead to the zero-level changing through the record. If this is significant, individual waves may not cross the mean level for the record. Again this is more critical for smaller waves, and can be prevented by de-trending the data prior to processing.

2.1.2 Spectral analysis

Spectral analysis is the process of breaking complex signals down into harmonic components of varying frequencies that, when combined constitute the original signal. The application of this method to water waves assumes that the waters surface is a combination of harmonic waves of sinusoidal form with random phases which oscillate about a zero mean. The validity of this assumption is discussed above. The output of spectral analysis is a measure of the variance of the sea surface (proportional to the energy) as a function of frequency. This measure itself is an excellent tool for describing a sea, and can also be used to calculate descriptive parameters. Where the sea surface is Gaussian, as described in section 2.1, the probability density function of wave heights can be predicted and parameters calculated in the frequency domain can be seen to be equivalent to those in the time-domain. It is from this association that the definition of spectrally-derived significant wave height is derived (as well as other spectral parameters). Since the development of computers with sufficient power to perform the processing, the analysis of wave surface elevation time-series measurements into summary parameters commonly uses frequency domain analysis. These data sets are used for calibration of new measurement technologies (Wyatt et al., 1999; Cotton and Carter, 1994) and are recommended for the assessment of a wave climate as a wave energy resource (Smith et al., 2006a). Protocols for the determination of the performance of wave energy devices are also based on frequency domain analysis of measured time-series of surface elevations into wave spectra (Smith and Taylor, 2007). To allow conclusions to be drawn relevant to these procedures, this thesis is primarily focused on data derived from frequency domain analysis. For this reason, the procedure was reviewed and is described in detail below.

Spectral analysis theory

The sea surface encountered by a wave sensor is a finite example of the infinite possible surface profiles that could result from the harmonic components present in the sea. Spectral analysis is the procedure of breaking a surface profile down into the frequency of components contributing to the sea and the energy associated with each. The waves are assumed to oscillate about a zero mean and the summation of all components (the surface profile) is assumed to follow a Gaussian probability distribution about a mean of 0. Details of the theory and procedure relevant to ocean waves is found in the major textbooks (Dean and Dalrymple, 1984; Tucker and Pitt, 2001; Holthuijsen, 2007)). The methodology applied depends on the wave sensor from which the measurements are derived. Perhaps the most

common form is the surface elevation time-series, as derived from floating wave buoys, which is described here. If applied to a spatial record of the sea surface, this methodology would yield the wavenumber spectrum. In an idealised case with no spatial variability, and no residual currents, the wavenumber and frequency spectra would be equivalent. There are many differences in the form and notation used in the literature to describe Fourier transforms. At a time, t, the instantaneous surface elevation, η can be expressed as the sum of N sine waves with different amplitudes, a, frequencies, f and phase, ϕ. For $i = 1, 2...N$,

$$\eta = \sum_{i=1}^{N} a_i sin(2\pi f_i t + \phi_i) \tag{2.1.1}$$

In practice, the variance of the sea surface, $\frac{a^2}{2}$, rather than the amplitude is most commonly used as this is proportional to the energy in the sea state. Taking a long term average, the variance becomes:

$$E = \overline{\eta^2(t)} = \sum_{n} \tfrac{1}{2} a_i^2 \tag{2.1.2}$$

which shows that the averaged variance of the sea surface is the sum of the variance of its components (Tucker and Pitt, 2001). However, surface elevation is a continuous variable, meaning there will be an infinite number of frequency components and summing all of these would lead to an infinitely large variance, suggesting an infinite wave amplitude. Instead, the density is used, which is calculated as the variance for small bands of frequency, Δf and plotting the variance per Δf. The frequency band is then allowed to approach zero to make a continuous variance density spectrum with the definition, (Holthuijsen, 2007)

$$E(f) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \tfrac{1}{2} \overline{a^2} \tag{2.1.3}$$

where $E(f)$ is the variance for frequencies in the range Δf. A key assumption in application to surface gravity waves is that the phase, ϕ can be considered equally likely to take any value (between $0 < \phi < 2\pi$). Thus, the phase spectrum can be ignored and the wave record can be characterised by the amplitude at each frequency.

The above formulation relies upon linearity in the wave field and stationarity of conditions across the wave record (be it in time or space). In practice, these assumptions are violated, but experience has shown that this approach remains valid for weakly non-linear or weakly non-stationary wave records (Ochi, 1998; Tucker and Pitt, 2001)

Application of spectral analysis

Application of spectral analysis to time-series is based on applying the theory described above to finite, discontinuous records of surface elevation. The development of computing has led to spectral analysis usually being performed by computers on digital data, most commonly using the fast Fourier transform (FFT) method.

Equation (2.1.3) defines the spectrum where a is the amplitude of a component wave. This amplitude is estimated from the fourier series of the surface elevation time-series ($\eta(t)$) as follows (Holthuijsen, 2007),

$$\eta(t) = \sum_{i=1}^{N} c_i sin(2\pi f_i t + \phi_i) \qquad \text{with } f_i = \frac{i}{D} \qquad \text{so that } \Delta f = \frac{1}{D} \tag{2.1.4}$$

where D is the duration of the time-series, and c_i represents the amplitude of the i^{th} harmonic. From this equation, it is apparent that the duration of the measurement period governs the frequencies that can be resolved and therefore, the frequency resolution. It is also evident that the components to be resolved have an exact number of cycles in the duration, because $T = 1/f$ and therefore, $D = iT_i$. Using trigonometric identities, this can also be written as:

$$\eta(t) = \sum_{i=1}^{N} [a_i cos(2\pi f_i t) + b_i sin(2\pi f_i t)] \tag{2.1.5}$$

where the amplitude and phase can be calculated as:

$$c_i = \sqrt{a_i^2 + b_i^2} \tag{2.1.6}$$

$$tan\phi_i = -\frac{b_i}{a_i} \tag{2.1.7}$$

Subsequently, the amplitudes a_i and b_i can be calculated from the time-series with fourier integrals

$$a_i = \frac{2}{D} \int_D \eta(t) cos(2\pi f_i t) dt \qquad \text{for } f_i = i/D \tag{2.1.8}$$

$$b_i = \frac{2}{D} \int_D \eta(t) sin(2\pi f_i t) dt \qquad \text{for } f_i = i/D \tag{2.1.9}$$

For a digital time-series, the record is discontinuous and therefore the integrals are converted to sums and a_i and b_i are calculated for each frequency band up to 1/2 the sampling frequency. The estimate of the spectral variance density is then calculated as, (Tucker and Pitt, 2001)

$$\hat{S}(f)\Delta f = \frac{1}{2} \sum (a_i^2 + b_i^2) \tag{2.1.10}$$

$$\hat{S}(f)\Delta f = \frac{1}{2} \sum c_i^2 \tag{2.1.11}$$

2.1.3 Spectral moments and parameters

The variance density spectrum can also be used to derive descriptive parameters, some of which are seen to approximate those defined in section 2.1.1, as well as parameters that describe the spectral shape. These parameters are derived from the spectral moments which can be defined as:

$$m_n = \int_0^\infty f^n S(f) df \tag{2.1.12}$$

Therefore the zeroth spectral moment, $m_0 = \int_0^\infty S(f)df$, or the area under the spectrum, represents the variance of the surface elevation.

Spectral moments can be used to calculate summary parameters including:

- **Significant wave height**, $H_{m0} = 4\sqrt{m0}$
- **Wave periods**,
  - $T_p = 1/f_{max}$
  - $T_{m02} = \sqrt{\frac{m_0}{m_2}}$, approximation of T_z, mean zero-crossing period
  - $T_{m10} = \frac{m_{-1}}{m_0}$, approximation of T_E, energy period
  - $T_{m01} = \frac{m_0}{m_1}$, mean period, 1/average frequency of the spectrum

Longuet-Higgins (1952) first demonstrated that for a linear wave field and a narrow band spectrum, wave heights follow the Rayleigh distribution. Where these assumptions hold, spectral summary parameters represent estimates of the associated time-series parameters and can be shown to be equivalent for a narrow-banded sea with a Rayleigh distribution of wave heights. As the sea becomes more broad-banded, this assumption becomes less accurate and discrepancies between the frequency-domain and time-domain parameters will grow.

Energy and power

Tucker and Pitt (2001) give an equation for the total energy in a wave field as,

$$E = \frac{1}{2}\rho g a^2 \tag{2.1.13}$$

They go on to show that substituting from equation 2.1.2 this becomes,

$$E = \rho g \int S(f) df = \rho g m_0 \tag{2.1.14}$$

Therefore, the energy in a sea is directly proportional to the variance of the vertical displacement of the sea surface, m_0. Wave power is dependent on the speed of propagation

of the energy. Energy will propagate at the speed of the wave groups, rather than phase speed. The measure of power transported per metre of wave crest is,

$$P = \rho g \int c_g(f) S(f) df \tag{2.1.15}$$

Spectral bandwidth

The spectral bandwidth describes the spread of energy within a spectrum. There are a number of parameters that represent the bandwidth (tab. 2.1), which have been derived for various ocean wave investigations, particularly those concerned with wave groupiness (e.g. Goda (1977)); in general, the smaller the bandwidth, the more pronounced the groupiness (Saulnier et al., 2007).

Table 2.1: Spectral bandwidth parameters, their formulation and source.

Parameter	Formula	Author
ε	$\sqrt{1 - \frac{m_2^2}{m_0 m_4}}$	Cartwright & Longuet-Higgins (1956)
ν	$\sqrt{\frac{m_0 m_2}{m_1^2}}$	Longuet-Higgins (1984)
ε_0	$\sqrt{\frac{m_0 m_{-2}}{m_{-1}^2}}$	Mollison (1985)
ε_1	$\sqrt{\frac{m_1 m_{-1}}{m_0^2}}$	Smith *et al.* (2006)
Q_P	$\frac{2}{m_0^2 \int_0^\infty f E^2(f) df}$	Goda (1976)

As can be seen in table (2.1), the derivation of the spectral bandwidth originally used values for the 4th order spectral moments. These moments have been found to have a high dependency on high frequency noise and the high frequency cut-off, making them unstable (Rye and Svee, 1976; Woolf, 2002; Smith et al., 2006b).

2.2 Sampling variability

Scientific descriptions of ocean waves, calculated as some form of statistical average of a finite record of the surface elevation, are estimates of the equivalent measure in the seaway. Their accuracy depends on the length, duration and sampling frequency of the measured record. In theory, uncertainties can be reduced by increasing the duration of the wave record. However, during the measurement period, conditions are assumed to be stationary and this assumption imposes limits on the length of measurement. In reality, the duration of measurement is a balance between being long enough to reduce statistical

uncertainties and being short enough to maximise stationarity. In the case of measuring ocean waves, a wave record of 17mins - 30 mins has become standard, however, there remain significant uncertainties associated with this time period.

As described above, the sea surface encountered by a wave sensor is a finite example of the infinite possible surface profiles that could result from the harmonic components present in the sea. After an FFT procedure, the two parts of the resultant complex number, a_i and b_i can be shown to be uncorrelated random variables that have a mean of zero and equal variances (Bendat and Piersol, 2000). Thus, from equation 2.1.10 $S(f_i)\Delta f$ is calculated as the sum of two squares of normally distributed variables and the probability distribution of their vector sum will follow a chi-squared probability distribution with 2 degrees of freedom (Tucker and Pitt, 2001; Bendat and Piersol, 2000). In the estimate of an individual harmonic, this represents an error term with variance of approximately 100%.

In order to reduce this variability, spectral estimates are commonly averaged over a number of harmonics. It is here that the length of the time-series plays a role in reducing variability. A long time-series, or one with high sampling frequency, can be averaged over a larger number of Fourier estimates than a short time-series, whilst retaining reasonable frequency resolution of the spectral estimates. Harmonics are averaged using one of two methods. The first is to average over a number of frequency bins. The second is by splitting the original time series into a number of segments before Fourier estimates are calculated for each. The spectral density estimate is then given by,

$$\hat{S}(f)\Delta f = \bar{S}_p(f) \tag{2.2.1}$$

where $\bar{S}_p(f) = \frac{1}{2}c_i^2$ for the p^{th} segment.

This involves averaging the sine and cosine components, making twice the number of harmonics (p) of squares of normally distributed variables. Thus, the averaged spectrum, $\hat{S}(f)$ remains chi-squared distributed, with $2p$ degrees of freedom, where p is either the number of segments, or the number of individual harmonics averaged. A fundamental property of the Chi-squared distribution is such that,

$$\frac{\sigma^2}{\mu^2} = \frac{2}{\upsilon} \tag{2.2.2}$$

where μ is the mean, σ^2 the variance about that mean and υ is the degrees of freedom. Substituting $\upsilon = 2p$ before re-arranging this equation demonstrates that the co-efficient

of variation ($c.o.v.$) in a spectral estimate will be,

$$\frac{\sigma}{\mu} = \sqrt{\frac{1}{p}} \tag{2.2.3}$$

$$c.o.v.(\hat{S}(f)) = \frac{1}{\sqrt{p}} \tag{2.2.4}$$

Where a large number of estimates are averaged, large values of p, the chi-squared distribution of each harmonic approaches a Gaussian distribution and confidence limits can be calculated based on this assumption (Bendat and Piersol, 2000). Therefore, the 95% confidence limit for each harmonic is estimated at $+/-2/\sqrt{p}$ from its mean.

$$\left[\left(1 - \frac{2}{\sqrt{p}}\right)\hat{S}(f) \leq S \leq \left(1 + \frac{2}{\sqrt{p}}\right)\hat{S}(f)\right] \tag{2.2.5}$$

Notably, the variability in the spectral estimate for a single harmonic does not depend on the length of the time-series; this is only variable when averaging a number of spectral estimates to estimate a final spectrum. A longer time-series will increase the frequency resolution of the spectral estimates, therefore allowing a greater number of individual estimates to be used for the same resolution in the final spectrum, and hence reducing the variability of these averaged estimates. However, assuming a finite record, increasing the number of averages used for each estimate will decrease the spectral frequency resolution of the resultant spectrum. This demonstrates the trade off between the variability and resolution in the spectrum.

Example of the effect of averaging on the resultant spectrum

Using a sample 30 minute record from a Datawell Waverider buoy operating at its standard sampling frequency of $1.28Hz$, yields 2300 data points. For the FFT, this is reduced to 2048 data points. Using segments of 128 data points (100 seconds) gives $p = 16$ and the $c.o.v. = 0.25$ (25%). Applying varying number of segments during processing to a sample of real data is demonstrated in figure 2.3. With no segmentation, the high variability in the spectrum can be observed. As segmentation is applied, the energy is spread between the harmonics, and the spectrum becomes smoother, with tighter confidence limits. The bi-modality of this spectrum is seen in all estimates. However, some structure is lost as more averages are used, and the frequency of the swell peak can be seen to change when 32 segments are used. Thus, the choice of averaging regime is a balance between the structure of the spectrum, and its variability. The choice will commonly depend on the application for which the spectrum is required.

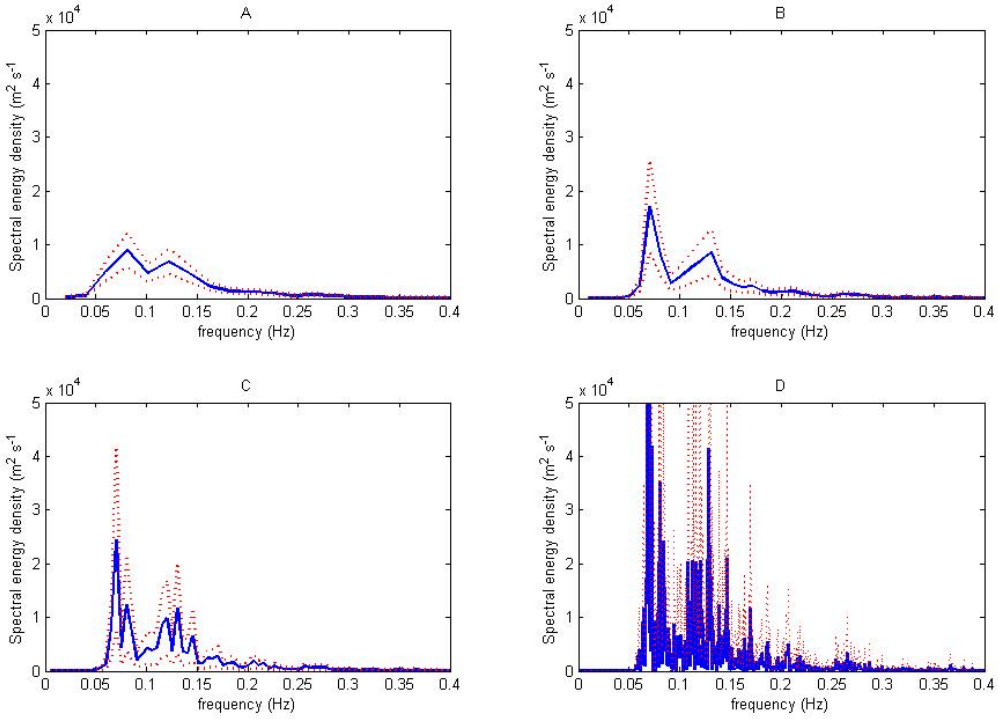

Figure 2.3: Wave spectra plotted for a sample of real data collected at the EMEC test site on the 1st September 2006. A: 32 segments, B: 16 segments, C: 8 segments, and D: 1 segment. The plots demonstrate the effect of changing the number of segments used during processing, and show their error bars.

2.2.1 Variability of spectral moments and parameters

The variability in spectral moments can be calculated from the variability in the spectral estimates as follows (Tucker and Pitt, 2001; Krogstad et al., 1999),

$$\sigma^2(m_n) = \frac{1}{D}\int_0^{f_u} \left[f^{2n}S(f)\right]^2 df \tag{2.2.6}$$

where D is the duration of the record, and f_u is the upper limit of the spectrum [1]. This equation forms the basis for estimating the sampling variability in all parameters that are derived from the spectral moments.

Krogstad et al. (1999) define the covariance of spectral moments as,

$$cov(m_r, m_s) = \frac{1}{D}\int_0^{f_u} f^{r+s}S^2(f)df + O(N^{-2}) \tag{2.2.7}$$

where N is the number of data points in the time-series, and $O(N^{-2})$ is a variable to the order of N^{-2}. It can be seen that if $n = r = s$ then equation (2.2.7) simplifies to equation

[1] Tucker and Pitt (2001) note that when this equation is applied to raw spectral data in its finite form, there is a factor of 1/2 required on the right hand side. This factor is reduced when spectral estimates are averaged, to 1.1 or less when averaging over 10 or more values.

(2.2.6) and gives the variance of m_n. Krogstad et al. (1999) go on to derive sampling variability of summary parameters based on equation (2.2.7). Those for H_{m0} and T_{m02} are as follows,

$$\sigma^2(H_{m0}) = 4cov(m_0, m_0)/m_0 \tag{2.2.8}$$

$$\sigma^2(T_{m02}) = \frac{1}{4}\left(\frac{cov(m_0, m_0)}{m_0 m_2} - 2\frac{cov(m_0, m_2)}{m_2^2} + \frac{m_0 cov(m_2, m_2)}{m_2^3}\right) \tag{2.2.9}$$

Where the spectrum from which H_{m0} (or other spectral parameters) is calculated is averaged over p harmonics (or segments), $c.o.v.(H_{m0})$ will be reduced by a factor of $1/\sqrt{p}$. Tucker and Pitt (2001) do not give a formula for the variance in H_{m0} estimates explicitly. Rather, they take the assumption that the $c.o.v.(H_{m0} \approx \frac{1}{2}c.o.v.(m_0))$, where

$$c.o.v.(m_0) = 100\left(\frac{\sqrt{\sigma^2(m_0)}}{m_0}\right) \tag{2.2.10}$$

and $\sigma^2(m_0)$ is calculated using equation (2.2.6). Both of these methods were then tested for a range of Pierson-Moskowitz spectra, where $H_{m0} = 0.5, 1, 1.5...10m, T_p = 10s, D = 1024s, \Delta f = 2Hz$ and no difference is seen between the results from these two formulations. The variability of H_{max} in a narrow band swell depends upon the number of waves passing the sensor. However, it will also depend upon the correlation of crest heights and the trough depth, as these will combine to form the maximum wave height in a record. In a wide band sea, these will not be 100% correlated, and the variability of this parameter will also depend on the correlation. Tucker and Pitt (2001) offer values for the variability in a narrow band sea, but with the caveat that these will be reduced by a factor of 0.85 for a locally-generated sea.

Pitt (2009) gives a formulation for the variance in an estimate of P_T from the spectrum,

$$\sigma^2(P_T) = \frac{1}{D}\sum_{i=1}^{N} P_i^2 \Delta f_i \tag{2.2.11}$$

where P_i is the power per m for a frequency bin in the variance spectrum, calculated as $P_i = \rho g c_g(f_i) S(f_i)$.

2.2.2 Degrees of freedom of spectral parameters

It has been described above that the degrees of freedom in individual spectral estimates is equivalent to twice the number of averages used in their estimation, $\upsilon = 2p$. Young (1986) shows that the effective degrees of freedom for a spectral parameter such as m_0, is not simply the sum of the number of degrees of freedom for each spectral estimate, rather

it is calculated as a weighted sum that depends on the spectral shape (Elgar et al., 1985).

$$d.o.f. = \frac{\upsilon(\sum_{1}^{N} S(f))^2}{\sum_{1}^{N} S^2(f)} \tag{2.2.12}$$

Elgar et al. (1985) demonstrate that this formulation, using discrete frequency bands, will introduce a bias of 2 for unsmoothed spectra, which reduces with increasing averages. An equation to calculate this bias is given as,

$$d.o.f. = \frac{1}{1 + \frac{2}{\upsilon}} d.o.f. \tag{2.2.13}$$

which shows that the maximum bias is a factor of 2, when $\upsilon = 2$ (no averaging), and decreases as the number of averages is increased. By using both equations (2.2.12) and (2.2.13), the number of degrees of freedom for spectral moments and other parameters estimated from the whole spectrum can be calculated.

2.3 Comparing spatially separated wave measurements

Goda (1977) used simulations to estimate the variability of the parameter $H_{1/3}$ and its effect on differences between simultaneous measurements from spatially separate sensors. The simulations are representative of two independent variables with errors that are equal and Gaussian distributed (see subsequent section, 2.4 for more detail). As such, the variance of the differences between the variables will be the sum of their respective error variances. Therefore, the simulations predict that the standard deviation of differences between two separate measurements from sensors exposed to the same wave field (but not the same waves) will be $\sqrt{2}\sigma$ where σ is the standard deviation in the original measurements.

Goda (1977) found the standard deviation of differences in empirical data sets approximately 50% greater than those in the simulations. The measurements used were taken from shallow water locations (8m - 24m), separated by 100m and situated close to a harbour. Data were subject to unspecified transformations in order to produce the deep water wave height. The increased variability was attributed to instrument noise, and the transformation applied. In a coastal location such as this, interactions with the local coastline and bathymetry would be expected to cause deterministic differences in the wave fields, and this may contribute to the increased variability. Furthermore, coherence between error terms was highlighted as an influencing factor (although neglected in the simulation),

particularly in the coastal zone, where shoaling and diffraction decrease the directional spreading (see section 2.4).

The coefficient of variation (*c.o.v.*) can be defined as follows,

$$c.o.v. = \frac{\sigma}{\mu} \tag{2.3.1}$$

where μ is the mean value and σ is the standard deviation of a given parameter, X. Mackay (2009) examined variability using the proportional differences between pairs of spatially separated simultaneous measurements, dX. For a data set containing n pairs of data, dX for each pair, i, is calculated as,

$$dX_i = \frac{(X_{1i} - X_{2i})}{\frac{1}{2}(X_{1i} + X_{2i})} \tag{2.3.2}$$

He then used numerical simulation to demonstrate that $\sigma_{dX} \approx 1.4(c.o.v.)$, of dX; because $\sqrt{2} \approx 1.4$, this result is equivalent to that found in (Goda, 1977).

Mackay (2009) used this methodology to estimate the *c.o.v.* values for the parameters H_{m0} and T_{m02} using two concurrent data sets captured at the EMEC site. However, estimations of variability calculated in this way represent the average variability for all values in the data set. Section 2.2 has demonstrated that variability in wave measurements will be dependent on the spectral shape and therefore on the sea conditions. It follows that estimates of *c.o.v.* derived in this way can be expected to be dependent on the wave conditions contributing to the data set.

Equation (2.3.2) was applied to empirical results in (Barrett et al., 2009) for three sets of simultaneous data measured at varying distance separations. This research identified a reduction in the variance of dX with distance in the empirical data. These data were sourced from different data sets captured at two different site. Therefore, factors including differences in the wave conditions, site-specific spatial distributions of the wave field or varying sensor performance will potentially influence the measured differences, not solely random variability.

A bootstrap technique applied in (Barrett et al., 2009) was used to classify the significance of observed differences between wave measurements. However, a shortcoming in the technique was that the variable sampling errors in H_{m0} estimates were not accounted for directly. Therefore, the results remain dependent on the wave conditions measured within the data set.

A more robust technique was applied by Sova and Wyatt (1995), who relate measured differences to estimated sampling error in each datum for the purpose of studying the spatial stationarity of wave fields using simultaneous measurements taken during the WADIC

(Wave data intercomparison study) experiment. Here, a Z statistic was defined as the difference between each individual pair of parameters, X_A and X_B, divided by the sum of their theoretical error variances, $\sigma^2(X)$, as predicted by sampling theory (see section 2.2).

$$Z(\Delta X)_i = \frac{\hat{X}_{Ai} - \hat{X}_{Bi}}{\sqrt{\sigma(X_{Ai})^2 + \sigma(X_{Bi})^2}} \tag{2.3.3}$$

Where no deterministic differences exist between the parameters X_A and X_B, and they are independent, the variable Z would be expected to follow a Gaussian distribution with mean 0 and standard deviation of 1, $p(Z) = N(0, 1)$.

Sova and Wyatt (1995) applied this analysis to pairs of data captured during the WADIC experiment (Allender et al., 1989). The results of this analysis found differences between wave measurements separated by distances between $200m$ and $400m$ to exceed the expected sampling variability more frequently when the winds were variable or strong. Measurements for this study were taken from different sensors, and this was found to affect the conclusions that could be drawn. Normalisation by variance (Z analysis) provides a robust analysis of the random errors. Unlike the estimation of *c.o.v.* described above, it is not dependent on the sea conditions within the data set. Although a bias will be indicated in the results, the normalisation applied means that conclusions cannot be drawn regarding its magnitude.

2.3.1 Regression techniques for wave data

When comparing two data sets, it is common practice to plot corresponding values as a scatter diagram on axis with one parameter as the x variable and the other as the y variable. A 'line of best fit' for these data describes the relationship between these parameters and most commonly takes linear form, $y = \beta_1 x + \beta_2$. The standard method for calculating the linear fit is to use an ordinary least squares (OLS) technique. This finds the β values that minimise

$$(\hat{y}_i - y_i)^2 \tag{2.3.4}$$

$$\text{where, } \hat{y}_i = \beta_1 x_i + \beta_2 + \epsilon \tag{2.3.5}$$

and ϵ is the error term associated with the y values, assuming that the error term in the x parameter are negligible. The R^2 term is commonly calculated for the line of best fit and represents the proportion of variability in the data that is explained by the model. In other words, it is the proportion of the sum of squares of the residuals (difference between model and data point on the y axis) to the total variation in the data. R^2 is calculated as

follows,

$$R^2 = \frac{SSE}{SST} \tag{2.3.6}$$

$$= \frac{\sum_0^i (y_i - \hat{y}_i)^2}{\sum_0^i (y_i - \bar{y})^2} \tag{2.3.7}$$

where SSE and SST are the residual sum of squares and the total sum of sqares respectively.

Linear regression using OLS requires the assumption that the error variance of data points is constant. Section 2.2 has demonstrated that this assumption will be violated when applied to wave data. A variable error term can be identified in the residuals as their spread about 0 will not be constant. In this case, weighting methods can be applied, which determines the contribution of each data point in the parameter estimation relative to the other data points. The function to be minimised becomes (Carroll and Ruppert, 1988),

$$w_i(\hat{y}_i - y_i)^2 \tag{2.3.8}$$

where w_i is the weight applied to each data point. These are commonly defined as the inverse error variance, so that the value with the largest errors has the least influence on the fit. Calculation of equation (2.3.8) for each data point gives the weighted residuals. Where weightings have been applied correctly, the spread of the weighted residuals should be approximately equal throughout the data set. Where the weighted regression is calculated using a least squares technique, it is known as Weighted Least Squares, or Generalised Least Squares. Weighting in regression is discussed in major textbooks e.g. (Draper and Smith, 1981), and a thorough review and description can be found in (Carroll and Ruppert, 1988).

Least squares techniques can include co-variates in the model to give a linear equation dependent on multiple parameters,

$$y = \beta_0 + \beta_1 x_i + \beta_2 x_i + \beta_j x_i ... + \beta_k x_i + \epsilon \tag{2.3.9}$$

where there are $j = 0, 1...k$ correlation co-efficients, β_j. Estimation of these parameters follows the same principles as the univariate method, and the error between the multivariate model and the observations is minimised (Draper and Smith, 1981; Montgomery, 1982). With multiple regressors, each β_j represents the expected change in y with changing x_j, when all other x values are kept the same.

When comparing statistical parameters describing wave fields, both parameters are likely to contain a significant error term, in which case, a least squares analysis may be biased

(Draper and Smith, 1981; Krogstad et al., 1999; Smith, 2009). In response to the problem of errors in both variables, Tucker and Pitt (2001) recommend that a reduced major axis (RMA) fit technique be applied. In its formulation, RMA accounts for errors in both the X and Y variables by minimising the horizontal distance, as well as the vertical distance, between itself and the 'line of best fit'; thus, taking into account errors in both variables. Smith (2009) discusses the choice between OLS and RMA techniques, showing that for data where the error is split evenly between the two variables, RMA is the more appropriate choice. However, this technique does not explicitly account for heteroscedastic variables (changing variance). Kalantar (1990) describes an alternative WLS method that applies a weight to both the x and the y variable, for a case where errors are Gaussian distributed and known for each data point. This method does not appear to consistently provide a more sensitive analysis, or reduction in bias than standard WLS.

Krogstad et al. (1999) compare the performance of alternatives to OLS for the purpose of comparing wave statistics between different sensors. Methods compared include a two-sided linear regression, which takes into account errors in the x variable as well as the y variable although again, the error term is assumed to be constant for the whole data set. Perhaps more relevant for application to wave data, described in Krogstad et al. (1999) is a maximum likelihood (ML) method, suggested by Sova (1995). This assumes that the two parameters have a Gaussian error term and a linear relationship, but accounts for variable errors in both data sets. Krogstad et al. (1999) applies a range of regression methods to intercomparisons of data sourced from wave buoys and HF radar. The ML method is seen to provide a more consistent estimation of bias than other regression techniques.

2.4 Coherence between sampling errors in wave measurements

Correlation between time-series will cause coherence between the error term in derived summary wave parameters. When examining the differences between simultaneous records, this coherence will affect the results. Methodologies reviewed in section 2.3 to define variability in parameter measurements rely on an assumption that measured time-series are independent, which is considered in more detail in this section.

When separation of measurement points occurs perpendicular to the wave propagation direction (angle of incidence, $i = 90^o$), correlation between surface profiles will occur when the crest length of individual waves spans the distance between sensors. Where

the separation is parallel to the direction of propagation (angle of incidence, $i = 0^o$), persistence of individual phase waves will cause correlation between wave sensors that is diminished with distance as the phase of individual waves is altered relative to those of other frequencies through the process of dispersion.

This correlation can be defined through expressing an individual measurement of a given wave parameter, X, as a combination of the true value and a random error component,

$$\hat{X} = X + \epsilon \tag{2.4.1}$$

where $\hat{X}$ indicates a measurement of the true parameter, X, and ϵ is a random error term, assumed to be drawn from a Gaussian distribution with variance equal to the error variance of an individual estimate of $\hat{X}$ (defined in section 2.2.1),

$$\epsilon = N(0, \sigma(X_i)) \tag{2.4.2}$$

For two simultaneous measurements, X_A and X_B, captured by two sensors measuring close to each other, the true sea conditions measured by the buoys will be the same (or similar) and therefore, values of $\hat{X_A}$ and $\hat{X_B}$ will be highly correlated. However, independence of the time series would mean that ϵ_A is not correlated with ϵ_B. Therefore, for the methodologies reviewed in section 2.3 to be valid, the measurements would be sited close enough together that the sea conditions are the same, whilst separated by sufficient distance that the time-series are independent.

Tucker (1993) defines the distance, D, over which the errors in two spatially separated wave measurements can be considered independent for individual frequency bands, as that over which the relative phase of component waves at each edge of the frequency band have changed by 2π. For simplicity, the frequency band is assumed small, such that frequency is assumed constant, and $G(f, \theta)$ is assumed constant across a bandwidth of $2\Delta\pi$. When the buoys are separated along the line parallel to the direction of propagation, D_{par} can be calculated using,

$$k_n D_{par}(1 - cos\Delta\theta) = 2\pi \tag{2.4.3}$$

where $k_n = \omega^2/g$, for the range of frequencies in a given band, and $\Delta\theta$ is the directional spreading. When the sensors are positioned perpendicular to the direction of propagation of the waves, Tucker (1993) defines correlation as zero when,

$$k_n D_{perp} sin(\Delta\theta) = \pi \tag{2.4.4}$$

For distances greater than D, coherence is assumed to remain as zero. For the comparison of summary parameters, these formulae should be applied using the range of active

frequencies that contribute to the parameter estimates. However, extending the frequency band reduces the accuracy of the assumption that frequency is constant across the band. Goda (1977) also examined correlation of spatially separated wave profiles by simulating wave fields using directional spectra based on a Bretschneider-Mitsuyasu formulation. Thus, the spectral energy densities represent that of a developed sea,

$$S(f,\theta) = S(f)G(f,\theta) \tag{2.4.5}$$

$$G(f,\theta) = \alpha cos^{2s}\left(\frac{\theta}{2}\right) \tag{2.4.6}$$

where α is a constant and θ is the direction of propagation. Directional spreading is therefore governed by the function cos^{2S} and a parameter s_{max} was defined to describe the level of spreading in the spectrum as follows,

$$s = s_{max}(f/f_p)^5 : [f \leq f_p] \tag{2.4.7}$$

$$s = s_{max}(f/f_p)^{-2.5} : [f \geq f_p] \tag{2.4.8}$$

where f_p represents the peak spectral frequency. Therefore, $s = s_{max}$ at the peak spectral frequency. s and $\Delta\theta$ have a fixed relationship, described in (Tucker and Pitt, 2001). Goda (1977) uses this formulation to estimate the distance above which correlation between surface profiles becomes negligble for two levels of directional spread and for separations both parallel and perpendicular to the direction of wave propagation (table 2.2).

Results from Goda (1977) and values calculated using the theoretical formulation of Tucker (1993) are shown in table 2.2. Note that the values from Tucker (1993) represent approximations for a narrow frequency band in deep water. These results demonstrate that the level of correlation can be expected to diminish rapidly perpendicularly to the direction of propagation, whilst it will persist for significantly greater ditances in the direction of propagation. Furthermore, low directional spreading results in correlation at greater separations for both angles of incidence. Maximum stated levels for correlation of $2km$ are greater than the separation for the buoy array, whilst minimum separations for larger directional spreading values are roughly equivalent to the separation of buoys within the array. Therefore, correlation in the error terms can be expected in the measured data, the level of which will be dependent on the wave conditions prevalent at the time. Tucker (1993) stresses that equations (2.4.3) and (2.4.4) have not been tested empirically.

Results from the two sources are of the same order of magnitude and both reflect differences caused by the alignment of separation (table 2.2). Estimates from Goda (1977) predict independence of wave fields over shorter distances than the formulation of Tucker (1993).

The differences between the two methodologies relates to their definition of independent records. Tucker (1993) uses the relative phase of the waves whilst Goda (1977) defines it empirically from the simulation as the distance, D, at which correlation co-efficients between records were less than 0.3.

Table 2.2: The spatial limit of independence of surface elevation time-series.

Perpendicular separation		
$\Delta\theta(^o)$	Goda (1977)	Tucker (1993)
45	0.3λ	0.7λ
22.5	0.8λ	1.3λ
Parallel separation		
45	2λ	3.4λ
22.5	10λ	13λ

Chapter 3

The PRIMaRE buoy array

As part of its research support of the Wave Hub project, PRIMaRE purchased 8 Fugro Oceanor SeaWatchMini II (SW) directional wave buoys for deployment around the proposed wave energy test site. All eight buoys are the same model and when operational they are time-synchronised so that measurement periods co-incide across all buoys deployed. The author played a key role in the design of a large-scale experiment using these buoys, and subsequently led the field operations and data processing. This chapter describes the site, experimental design and equipment used in this experiment.

3.1 Buoy Site

The choice of deployment site for the buoy array took into consideration the proposed uses for the data captured by the facility. The final deployment must be a compromise that accommodates as many areas of research as possible, whilst offering data of the highest quality for the key research aims. These decisions must be taken within the framework of the resources available and considerations for marine operations and other marine activities in the area. The discussions centred on two key points, the location of the buoys relative to each other, and their geographic location. A key decision taken early on was that the first deployments would use 4 buoys. This limited the operational requirements for service and maintenance and ensured that replacement buoys were available should anything happen to those in operation, which would limit gaps in the data.

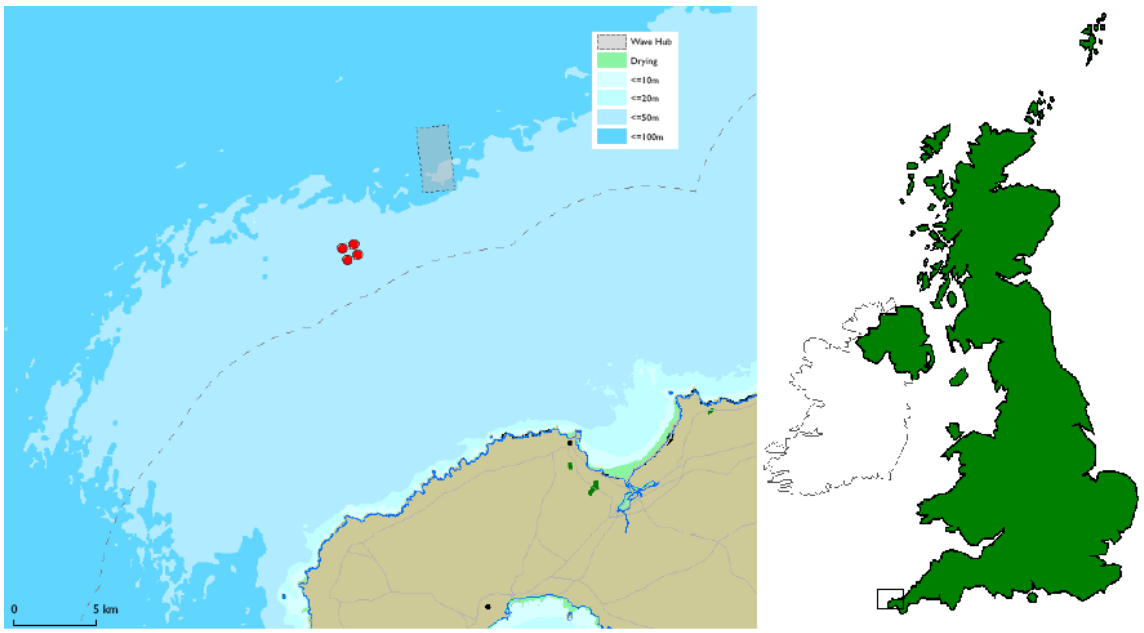

Figure 3.1: Location of the Wave Hub and the PRIMaRE buoy array, marked as red circles.

3.1.1 Geographical location

The deployment site is approximately $15km$ off the north coast of Cornwall, in the south west UK, situated $5km$ to the south west of the Wave Hub site (fig. 3.1). Water depth within the array is between $41m$ and $32m$ and the changing depth is due to a rocky reef situated in the centre of the array which extends east to west (fig. 3.2).

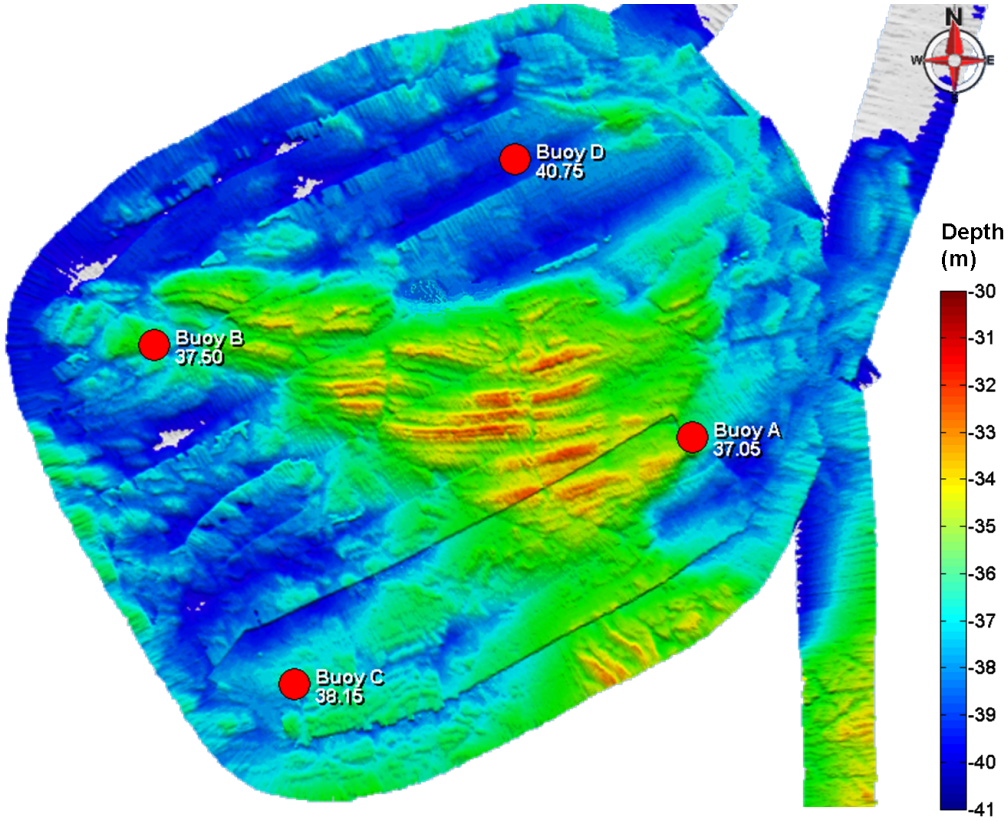

Figure 3.2: The bathymetry at the wave buoy site, captured using a multi-beam sonar.

The primary aim was to deploy the buoys at or close to the Wave Hub site, although it was necessary to consider the dangers the buoys posed to marine traffic and the danger marine traffic posed to the safe operation of the buoys. The Wave Hub sits close to a busy shipping lane, although it benefits from some protection due to a traffic separation scheme in operation around the western tip of Cornwall which forces transport vessels further north. It follows that the further south, towards the coast a buoy is located the less the risk posed by transport

The area around the Wave Hub site is situated in a mixed fishery and at certain times of the year is frequented by beam trawlers. These vessels would be highly destructive to a wave buoy and its mooring which, in turn has the potential to cause entanglement difficulties

for the fishing vessel. Despite the safety systems installed on the buoys, consultation with the local fishing groups indicated that prior to a permanent exclusion zone being set-up around the Wave Hub site, any buoy is vulnerable to collision. Furthermore, the Wave hub permitting process encountered significant opposition from the local fishing industry who saw the project as an unfair exclusion from their fishing grounds. The wave sensors deployed represented the first manifestation of the project at the site and it is possible that they became targets for opponents who felt that the administrative and legal processes were not adequately representing their views.

At the time of deployment, there were no installations at the Wave Hub site and no legal exclusion zone to marine vessels in this area. The considerable dangers that arose from fishing activities forced the buoys to be deployed outside the Wave Hub site. Consultation with the fleet active in the area identified a site with a rough sea bed of rocky outcrops (fig. 3.1). In this area, the fishing method of trawling is not possible and activities are limited to potting for cetaceans. The presence of the wave buoys does not inhibit this fishing method to the extent it would a trawling operation, reducing collision danger and the inconvenience to the fishermen. Although no guarantees of safety were offered, this site had the support of the majority of the fishing community.

3.1.2 Layout of the array

The buoys were deployed in a square $500m \times 500m$, with sides aligned broadly WSW - ENE and SSE - NNW (fig. 3.1). The northerly buoy was deployed on flat ground comprising sand or gravel in $41m$ depth, whilst the other buoys were deployed on rough ground in $37m$ (E and W) and $38m$ (S) (fig. 3.2).

In order to examine the spatial variability of wave fields on spatial scales relevant to wave energy test sites, it was necessary that the spatial separation between the buoys was on a representative scale. At an operating site, it is likely that safety will dictate that wave measurement devices and zones reserved for operating devices will be separated by at least $200m$. Where limited measurement resources are available, in larger installations this separation may increase to kilometers. The proposed size of a berth at the Wave Hub is a $1km$ square. Therefore, a distance between the buoys between $200m$ and $2km$ was considered suitable.

Measurements captured very close together will be measuring highly correlated wave fields, and the coherence will diminish with distance separation to a limit where the theoretical coherence between wave fields is 0. The limit of coherence varies with the directional

spread of the incident waves (see section 2.4). This is a potentially important factor on the accuracy of spatially separated wave measurements and it was considered advantageous that the separation of the buoys was within the range of this limit so that both data which would be expected to exhibit coherence, and data that would not, were present in the data set. Therefore, the proposed separation of $500m$ allows investigation of the effect of coherence on the measurements across the array and supports the requirements of this research.

3.1.3 Wave climate at the site

The deployment site lies in an exposed location on the continental shelf of the UK, in the North Eastern Atlantic. This area is at the eastern end of 3000km of fetch exposed to the predominantly westerly winds associated with low pressure systems that track west across the North Atlantic. The predominant winds at the site are from the western sector, also associated with these low-pressure systems. Fetch from the northerly sector is limited to approximately 300km by Ireland, which blocks waves generated in the seas surrounding Iceland and Greenland. The maximum fetch is 6000km from the south-west, which stretches into tropical seas in the Carribean and north of Brazil. Storms in this area are limited to hurricanes during the autumn.

The South West Regional Development Agency (SWRDA) commissioned a series of measurements at the Wave Hub site, comprising buoy deployments between February and March 2008, between December 2008 and March 2009, and during July and August 2009. These data were analysed alongside a 15yr hindcast of the UK Met-office global wave model to perform a resource assessment and estimate extreme conditions at the nearby Wave Hub site (Halcrow, 2006a). Prediction of the wave climate at the buoy site can draw heavily on these measurements and analysis for the nearby Wave Hub site, although the Wave Hub is situated $5km$ to the north east of the wave buoys and the wave climate will differ between these two sites due to a difference in water depth (WH 55m, Buoys 37-41m). A wave modelling study described in Pitt et al. (2006) demonstrated that an area of shallow ground to the west of the Wave Hub site can be expected to affect the wave climate at the site. The spatial extent of these effects will vary with wave conditions and has the potential to increase the differences between the buoy array and the Wave Hub site. Therefore, whilst the wave conditions can be considered broadly similar, further research would be required to quantify the extent to which estimates of extreme levels or long-term averages for the buoy site could be considered representative of the wider area.

Analysis of the data captured for the Wave Hub site indicates that based on the mean zero-crossing period, T_{m02}, a water depth of 37m would be considered as deep water for 80% of records captured, whereas for the others, it is intermediate water. This indicates that the waves propagating in this area are frequently interacting with the sea-floor. This process is wavelength dependent, and therefore, lower frequencies present in the sea state will be affected by interactions with the sea floor more commonly than suggested by this analysis of mean conditions.

Where the waves are interacting with the sea floor, processes such as shoaling, refraction and diffraction will alter the properties of waves incident on the buoy array site. The non-uniform shape of the sea bed (fig. 3.2) has the potential to induce spatial differences in the wave field through the variation of these effects across the site. Such effects would be dependent on the wavelength and propagation direction of the incident wave field, and modified by the height of the tide.

3.1.4 Tidal conditions at the site

Tidal flow conditions at the site were investigated using the national tidal model, Polpred (POL, 2010). The available model data spans a period from the 1st Jan 2008 until 31st Dec 2010, and provides estimates of tidal level, flow speed (integrated for depth) and direction on a grid of resolution 500m every 30 seconds. A grid output is situated close to the centre of the buoy array, and this is assumed to be representative of tidal conditions at all four buoys. From these data, the tidal range is 5.8m, inducing maximum currents of 1.5m/s. The flow direction is generally from the south west to the north east during a flood tide, and from the north east to the south west during an ebb tide.

3.2 The Seawatch mini II (SW)

The SW is a directional floating wave buoy of the surface following type, which operate on the principle that when submitted to forces by waves in a range of frequencies, they will follow exactly the motion of surface water particles. Thus, measuring the motion of the buoy is equivalent to measuring the exact motion of the waters surface. The buoy cannot respond to all frequencies of motion. In short, high frequency waves, the forces at the water surface and at the deepest point of the buoy will be significantly different and the buoy will not follow wave motion, which introduces a systematic error. The high frequency cut-off that should be applied to the data can be predicted for a linear wave as

$f_c = 2D$ where D is the diameter of the buoy (Joosten, 2006). The low frequency limit is governed by the mooring, which is discussed below. These limits are accounted for in processing through the application of a spectral filter, which is discussed in more detail in section 4.8.

Wave measurement and processing is performed on-board the buoy using the 'Wavesense' unit, a computer that collects raw data from the sensors and performs processing to output wave data. Data collection, processing, storage and communication are all controlled by this computer, and are programmed by the user during setup. The user-defined specifications that were established with Oceanor are defined in section 3.2.1, and the data analysis procedures applied on-board are examined in detail in chapter 4. During operation, changes to the setup can be applied by providing the buoys with different setup files, which program the on-board computer. Changes must be carefully considered to ensure continuity in the data sets.

Two key upgrades were developed by Oceanor in response to the requirements of this facility. The first was that the buoys are phase-locked so that measurements at each buoy commence at the same time. The second was an upgrade of the software that controls the storage and transmission of data on-board to enable the user to capture raw data from all sensors.

There is a global positioning system (GPS) on-board from which the instantaneous position of the buoy is recorded at the end of each record. This unit also provides the correct time, which governs the time synchronisation between buoys. A radio and an iridium satellite system provide communications, which are described in section 3.2.2. All of these systems are controlled by the on-board computer, and draw power from the central buoy system. This comprises a bank of rechargeable lead-acid batteries, which are charged via solar panels on the buoy. These are supported by a bank of non-rechargeable lithium batteries, which must be replaced annually.

To improve radio transmission to shore, a mast was added to the buoy design which raised the radio transmission aerial. Deployment of the buoys requires consent from the relevant marine authorities. In order to meet consent requirements, the following safety systems were added to the design of each buoy, which were also mounted on the mast:

- navigation lights with a defined flash pattern,
- active radar reflectors which return a large radar pulse to any received pulse giving each buoy a radar signal equivalent to an object of length $20m$.

3.2.1 Specifications of the buoy system

Software controlling the operation of the 'Wavesense' computer is programmed prior to deployment to set the specifications of the data capture, processing, communication and storage. The setup described here was developed by Fugro Oceanor after consultation, and represents the standard setup for the buoy array.

- Raw data - the raw accelerations sampled on three separate axes at $4Hz$, with simultaneous heading and tilt readings. These can be transmitted when software is prompted remotely. It is recommended that this is constrained to a short period of time, as this process requires continuous operation of the radio, which will present a significant drain on the batteries.

- Time-series data - a $17min$ $4sec$ time-series of buoy positions in the vertical, east and north directions, sampled at $2Hz$, is captured once every 30 minutes. measurement commences on the hour and half hour based on coordinated universal time (UTC). Data are digitised for storage at 8 bits, which translates to a resolution of approximately $4cm$.

- Spectral data - A directional spectrum is produced on-board the buoy for each $17min$ $5sec$ record, and used to calculate summary parameters. Spectra to be transmitted were digitised at 7bits, which translates to a resolution in $S(f)$ values of $0.4m^2/Hz$.

- Summary parameters - A spectral filter is applied prior to both spectral and time-domain parameter estimation which limits the contribution to frequencies above $0.04Hz$ and below $0.5Hz$. A range of time domain parameters are calculated from a zero-upcrossing analysis of the measured displacement time series (tab. 4.1). Spectral parameters are estimated from the moments of raw spectral estimates (tab. 4.2)

The procedures used for data processing are described in more detail in section 4.

3.2.2 Data storage and transmission

The buoys communicate directly with a shore-station, situated $16km$ away, via an ultra high frequency (UHF) radio link. The shore-station radios are synchronised with those on-board the buoys so that they transmit and receive only once every 30 mins, which reduces drain on the batteries. The distance of the buoys from shore, and interference

with meteorological and sea conditions, limits the amount of data that can be transmitted. The power of the radios is limited by the operation license, and the drain on the battery.

The displacement time-series files were considered too large for real-time transmission via the radio link, and are stored on-board the buoy to be transferred via a direct cable link, either at sea, or on retrieval of the buoy. A procedure has been established for the regular service and maintenance checks, and data transferral from buoys operating at site. Establishing a hard-wire connection to a floating buoy for an extended period of time requires very benign sea conditions. However, the nature of a wave energy site is that it is in a region of high energy. During winter, this makes the periodic retrieval of data (and maintenance) difficult to schedule. A successful cruise allows the retrieval of a full data set and offers an opportunity to communicate other data or commands to the buoy, such as sending new setup files.

Transmitted data are stored at the shore-station and transferred via the internet to a central file deposit at the University campus. This computer benefits from an automated back-up service via the university server. Therefore, at any time, four copies of these data are stored in separate locations (buoy, shore-station, central deposit and backup). Time-series data are stored on-board the buoy and retrieved during three-monthly maintenance visits where they are transferred to a portable computer on the service vessel. On return, they are transferred to the central file deposit. Therefore, copies of these data are routinely stored in three separate places, the buoy, the central file deposit and backup.

3.3 Moorings and station keeping

Previous wave buoys deployed at the Wave Hub site have experienced problems and have broken their moorings or ceased to function. No information is available regarding the nature of these failures, or what caused them. Pitt (2006) describes floating wave buoys as particularly vulnerable to collisions or malicious damage. An interesting account of the deployment of a single wave-buoy in an exposed location is given in Holmes and Barrett (2007), where various delays, radio faults and weather preventing the recovery led to an intermittent data set and culminated in the loss of the sensor. In this project, the wave buoys were purchased as a complete system, to the specifications of the site, and it is the responsibility of Fugro Oceanor to provide a system capable of surviving in the expected hydrodynamic conditions.

3.3.1 Mooring a surface-following directional wave buoy

Periodic changes in pressure change the buoyancy force and induce vertical motion (Datawell, 2005). The orbital motion of water particles due to wave action will exert impulse and drag forces on the buoy, setting it in motion, and an unhindered buoy can be expected to follow surface motion exactly when the wavelength is long enough. However, the water particles are also in motion with residual currents induced by tidal flow, surface wind and second order drift due to waves. Therefore, a buoy will not remain on station without a mooring. The requirements of a surface following directional wave buoy mooring are twofold, to allow unhindered response to wave motion across the range of frequencies corresponding to surface gravity waves, and to keep the buoy on station. To ensure the buoy still has the freedom to follow wave motion, mooring lines incorporate at least one section of elastic material (Joosten, 2006). The properties of this line is not discussed explicitly in Sanmuganathan (2009), although the mooring is deemed sufficiently compliant that the transfer function of the buoy-mooring system is considered unity across the active wave frequencies ($0.04Hz - 0.5Hz$).

The Datawell Waverider also uses elasticated sections in its moorings, and (Datawell, 2005) discusses the mooring response and its effect on measuring waves of different frequencies. Using spring-damper theory for the elastic section of the buoy mooring,

$$F_{moor} = CA \tag{3.3.1}$$

where F_{moor} is the force applied by the mooring, and C is the spring constant of the rubber cord and A is the amplitude of the motion. The mass spring resonance frequency, f_0, for the system can be calculated as,

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{C}{m}} \tag{3.3.2}$$

where m is the mass of the buoy, including added mass. Added mass for the buoy may change through an oscillation. No indication of how this is calculated, its value, or the combined mass, are given in Datawell (2005). The wave forces, F_{wave}, are,

$$F_{wave} = m(2\pi f)_2 A \tag{3.3.3}$$

$$\frac{F_{moor}}{F_{wave}} = \frac{CA}{m(2\pi f)^2 A} = \left(\frac{f_0}{f}\right)^2 \tag{3.3.4}$$

where f is the exciting wave frequency, and m is the mass (including added mass) of the buoy. f_0 represents the lowest frequency that a buoy can follow accurately without being impeded by its mooring.

The spring-constant of the elasticated section of the mooring provided by Oceanor has not been made avialable. However, a load-cell was attached to a spare unused mooring line in the laboratory, and the extension under various forces was measured. This experiment estimated the spring-constant to be $35Nm^{-1}$. This value can be substituted into equation 3.3.2, along with the mass of the buoy ($320kg$) to calculate f_0.

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{35}{320}} \tag{3.3.5}$$

$$= 0.0526Hz \tag{3.3.6}$$

A value that is close to the frequency cut-off applied by Oceanor during processing. Notably, this formulation neglects added mass, which would increase m, and reduce f_0. It is not a valid assumption that this is negligible, therefore, the result can be considered an over-estimate of f_0.

The elasticated section, of length L is supported by a back-up line which has length $1.6L$ and therefore becomes taught when $A > 0.6L$. When this occurs, the formulation for f_0 becomes invalid. Without a detailed analysis of the forces on the mooring components, it is not clear how often the back-up line is taught. During long-term deployments, a large amount of biological growth was observed on the back-up line, with algal fronds in excess of $2m$. These will increase drag on the mooring components further affecting the pre-tension and mooring extension. The changing nature of this growth through time has the potential to affect measurements, although without further analysis, this cannot be quantified.

3.3.2 Moorings supplied for the PRIMaRE buoy array

The moorings supplied comprise a primary and secondary anchor, connected together using a $16mm$ ground chain. This chain is attached to a fibre-wire combination rope which in turn connects to $30m$ of rubber cord complete with a back-up line. This elasticated section attaches to a cubic surface float of dimensions 1m x 1m, with a floated horizontal poly propylene rope providing a surface connection to the instrumentation buoy (fig. 3.3).

Sanmuganathan (2009) describes a single mooring riser with elastic section, attaching directly to the instrumentation buoy as the standard mooring for the SW. However, previous deployments of wave buoys (SW and Datawell waverider) at the WH site have experienced failures in their moorings and although the cause is unknown, the mooring in use in the current deployment was developed to account for the physical conditions at site including the tidal conditions and the energetic wave climate.

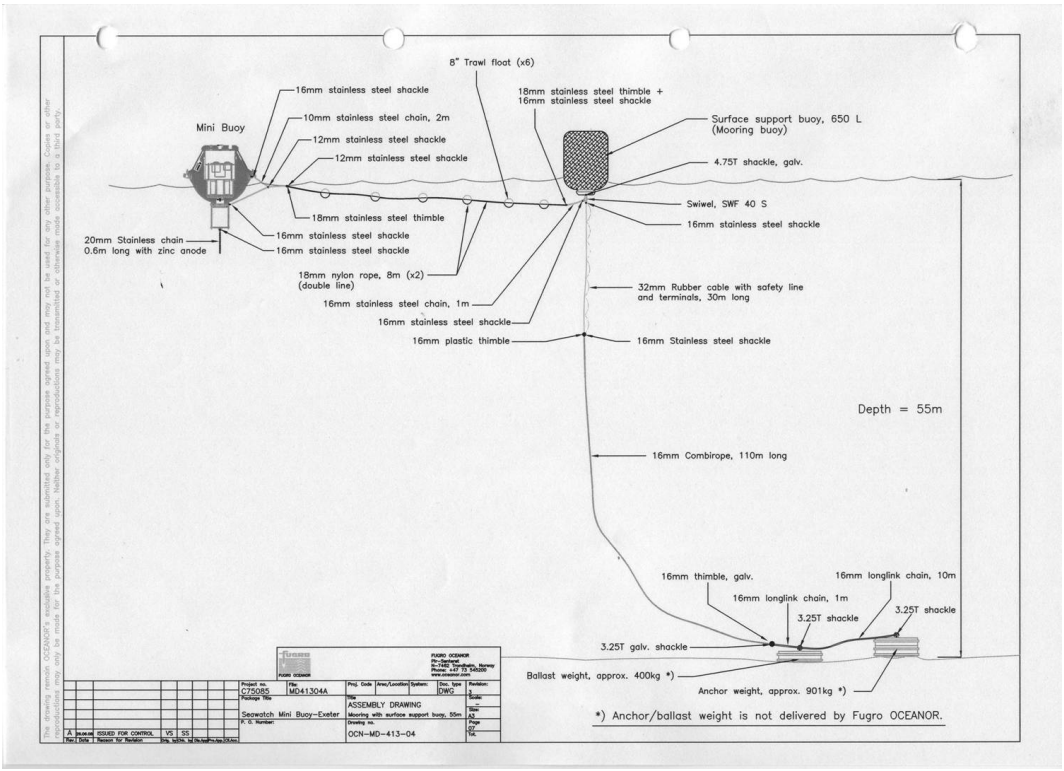

Figure 3.3: The mooring setup provided with the SeaWatch mini II wave buoy (Sanmuganathan, 2009). Copyright Fugro Oceanor.

Fugro Oceanor operate a range of moorings on their wave buoy systems, demonstrated in figure 3.4 (Oceanor, 2010) for their SeaWatch buoy [1]. This image indicates that the mooring in use for the PRIMaRE buoys is an offshore mooring system for deeper waters, provided to cope with the combined strong tidal flows and energetic wave climate. Using these moorings can be expected to alter the response of the buoy-mooring coupled system to incident waves when compared to the standard system.

The connection between the float at the top of the riser and the instrumentation buoy is not elasticated, and the float has the potential to apply forces that will divert the buoy from surface following. This effect will be frequency dependent, with minimum effect from long, low-frequency waves where there will be very little phase difference between the buoys, and maximised when the two buoys are π radians out of phase. In the latter case, motions will be in opposite directions. The lower weight of the surface float, gives it less inertia and it will yield to the instrumentation buoy, although some effect on the measurements would be expected.

The total length of the moorings supplied for the deployment depth of $37m$, was $124m$,

[1] A larger buoy used as a platform for multiple sensors including directional waves.

more than 3 times the mean water depth[2]. With the possible exception of slack tide, drift forces will ensure that the buoy is held at the full horizontal extent of the mooring. The length of the mooring ensures that the angle between the position of the mooring and the horizontal plane will be 18^o and therefore the majority of the restoring force from the mooring will be acting in the horizontal direction, reducing the effect on the vertical motions. The addition of the surface float into the design further reduces the angle at which the restoring force acts, and the vertical component of the restoring force applied by the mooring. This alignment also serves to increase the component of the restoration force that is acting in a horizontal direction.

In Oceanor core processing software, the transfer function of the system is specified as unity across all frequencies (although only those frequencies corresponding to surface gravity waves are used in further processing). This treatment is based on the assumption that the instrumentation buoy follows motions induced by these frequencies exactly, despite the restorative forces applied by the mooring.

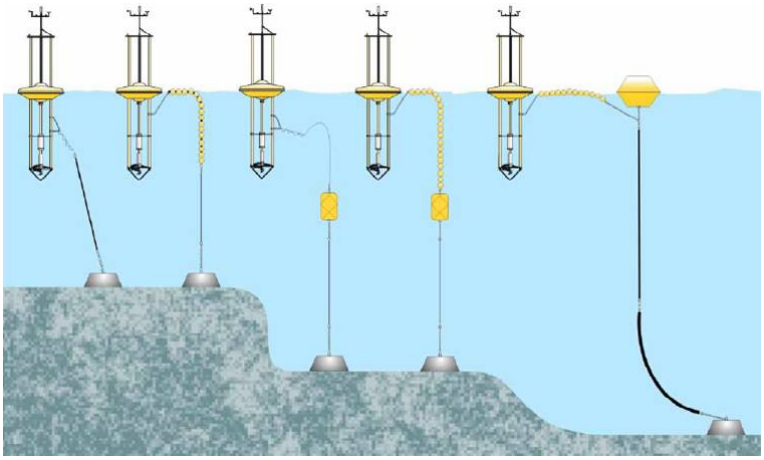

Figure 3.4: The range of moorings supplied by Fugro Oceanor for their SeaWatch measurement system (Oceanor, 2010).

3.4 Deployment

All four buoys were initially deployed on the 14th October 2009 under the direction of representatives from Fugro Oceanor. They were re-deployed following failures on the 16th December. The procedure for deployment is detailed in Sanmuganathan (2009). During the deployment, careful logging of all activities and photographing of all components was

[2]These proportions are different for the mooring specified for 55m depth, as shown in figure 3.3.

undertaken to contribute to supporting information for the facility.

The buoys were deployed at locations in a square, $500m$ by $500m$. Each buoy was positioned within $10m$ of its target location. The instantaneous position of each buoy is logged every $30mins$ to an accuracy of $+/-10m$ in both longitude and latitude. Excursion from the mean position of up to $56m$ was observed during operation, due to the influence of tide and wave drift forces. The buoys were named A, B, C and D, which corresponds to the radio that they are using, and is used primarily for identification of the data transmission. Individual buoys also have a serial number, which is fixed for the instrumentation buoy.

Chapter 4

Data processing

At the time of writing the wave buoy array facility has been operational for over 1 year, and 12 months of data have been collected from four buoys operating simultaneously. The author holds responsibility for the ongoing operation of the facility, which aims to provide a unique data set of value to the research community and industry users of the Wave Hub site. Both of these applications will value accuracy of the data very highly, and benefit from standardised measurement and processing procedures, aligned with published standards. This latter point may be particularly true for industrial activities because standardised procedures have been a formal requirement for public funding schemes such as the UK governments Marine Renewables Development Fund (MRDF).

Data processing, is split into stages. The initial processing of the raw sensor data was performed solely on-board the buoy and gives time-series files. The subsequent processing of these data was also performed on-board the buoy using Oceanor 'standard' procedures, details of which are provided in their manual (Sanmuganathan, 2009). In addition, a suite of processing functions were written using the programming environment Matlab ™, which processed time-series of surface elevation on-shore, to give a second estimation of non-directional parameters. These provide an independent verification of processed data outputs from the proprietary buoy software. This chapter describes the data processing applied, and provides results of comparisons. The processes are distinguished as 'on-board' processing, that were performed by the 'Wavesense' module on-board, and 'post-processing', performed using Matlab functions on-shore.

4.1 Processing raw sensor data

The sensors used to measure the buoy motion are physically situated within the 'Wavesense' unit. They comprise three accelerometers, situated on three separate axes, a gyroscope measuring the angle of inclination of the accelerometers (from the buoy position in still water), and a compass measuring the buoys heading. Accelerations on each of the three axes are measured at $4Hz$. However, the sensitive axis of each accelerometer will change continuously with respect to a fixed axis. The compass and gyro allow these motions to be resolved to accelerations in the heave (vertical), north and west directions. Raw accelerations are then double-integrated to give a time-series of positions in these directions. An example time-series is shown in figure 4.1. For each axis they consist of 2048 values at a frequency of $2Hz$, making the duration $1024s$, or $17mins\ 4secs$. The digitisation of the data is defined in the setup file at 10 bits per data point, between a lower limit of

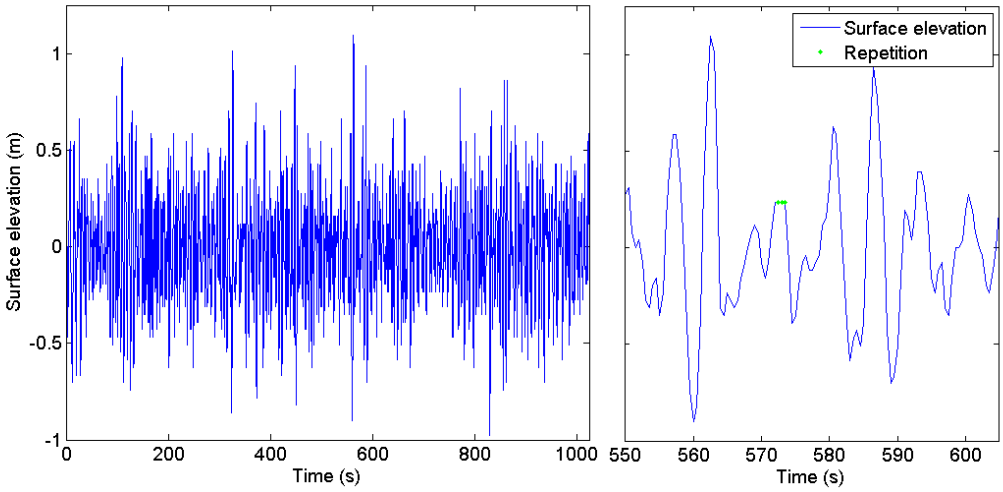

Figure 4.1: A sample time-series of surface elevation, recorded on 26-Jul-2010 11:30:00 ($H_s = 1.19m$), including a zoom showing a section of repeated data.

-20 and a high limit of 20. This gives a resolution of $0.0390625m$, or approximately $4cm$. Where the position of the sea surface moves slowly and does not change by more than $4cm$ between measurements, this resolution may lead to a number of points with the same value, examples of which can be seen in figure 4.1. This effect is most common at the peaks and troughs of the waves. Longer repetitions may be indicative of errors and are discussed below in section 5.5.

4.2 Direct time-series analysis

The methodology of direct time-series analysis has been described in general terms in section 2.1.1. The on-board processing of direct time-series parameters first applies an FFT function to the time-series. Subsequently, an inverse FFT is applied to data within the frequency band $0.04Hz$ to $0.5Hz$. This has the effect of filtering the data from low and high frequency waves. The requirement for, and the effect of this process, are examined in more detail in section 4.8. The parameters in table 4.1 are those that were calculated on-board the buoy when operating with the standard setup. The finite resolution of time-series data means that direct analysis will routinely underestimate crest height, h_c, and over-estimate trough height, h_t, (for definition see figure 2.1). Tucker and Pitt (2001) suggest that the resolution of the surface elevation data be a maximum of $1cm$ to minimise inaccuracies in direct-analysis. The resolution of $4cm$, supplied by the SeaWatch buoys, has the potential to reduce the accuracy of direct analysis, especially in small sea states.

Quality control tests were applied to highlight this effect, see section 5.5.

4.2.1 Post-processing direct analysis

Post-processing zero-crossing analysis of the time-series files uses functions provided in WAFO, an open source wave analysis package for Matlab, developed at the University of Lund, Sweden (WAFO, 2000). This package has become very widely used in the wave analysis community and is generally believed to be robust and reliable. The function *dat2steep* identifies zero-crossings in a time-series record of vertical positions to identify individual waves. It then identifies troughs and crests to measure wave height, and the time elapsed between up-crossings (or down-crossings) to derive the wave period. This analysis was performed on a smooth surface elevation time-series, estimated by applying a cubic spline interpolation across 5 consecutive data values in the time-series files. An example interpolated signal can be seen in figure 4.2. Using a smoothed signal will increase h_c and decrease h_t relative to the raw data. Therefore, it can be argued that the interpolation leads to a more accurate analysis than that based on a time-series with finite resolution. A comparison between the technique of filtering the time-series prior to analysis (on-board) and using an interpolated signal (post-processing) is examined in more detail in section 4.5.

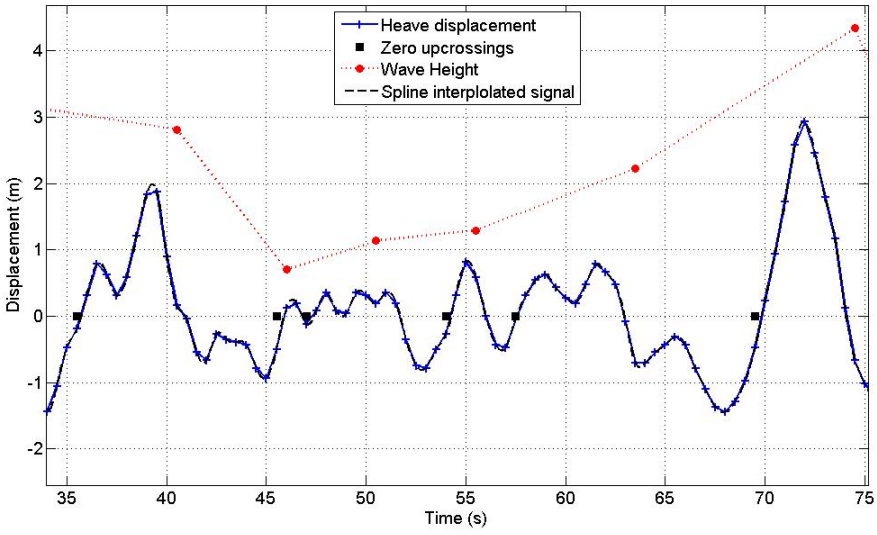

Figure 4.2: A sample time-series recorded on 21-Oct-2009 19:30:00. Also shown are the interpolated signal used to find the zero upcrossings, the index of zero upcrossings, and subsequently the heights associated with each wave.

Table 4.1: Time-domain parameters calculated on-board the SW for a standard setup at the PRIMaRE buoy array (Sanmuganathan, 2009).

Parameter	Definition
H_{max}	Height of the highest zero upcross wave in the record
H_{mean}	Mean height of zero upcross waves
H_s	Mean height of highest 1/3 zero upcross waves
H_{s1max}	Height of steepest wave
S_{1hmax}	Steepness of highest wave
S_{1max}	Steepness of steepest wave
S_{1mean}	Mean wave steepness
S_s	Significant steepness (time-domain), $S_s = H_s/(1.56T_z^2)$
σ_S	Standard deviation of individual steepness
T_s	Significant wave period (average period of 1/3 highest waves)

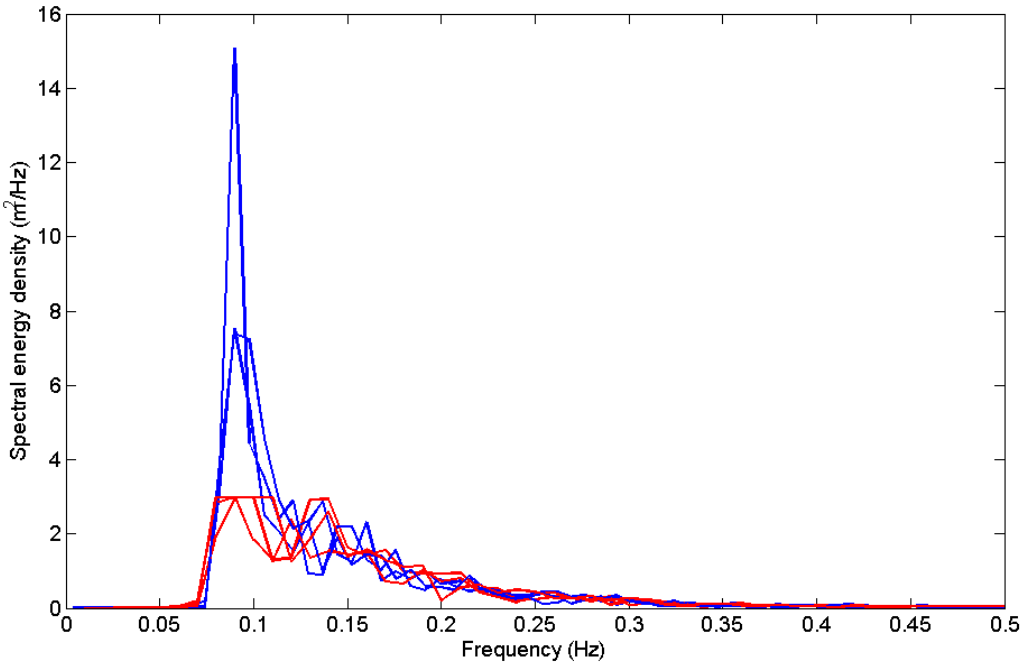

Figure 4.3: Variance density spectra for a large sea, estimated using post-processing of a surface elevation time series, and on-board by the 'Wavesense' package for transmission.

4.3 Spectral analysis

Spectral analysis is performed on-board the buoys for the calculation of parameters, and this process is described in Oceanor (2009). The averaging applied to the raw spectrum is an 8 point moving average giving 16 degrees of freedom to each estimate and an expected frequency resolution of $0.078Hz$. There is no indication in the manual that this averaging can be altered during setup. However, the setup file can contain an option for the resolution of the spectrum and the standard setup is for a frequency resolution of $0.01Hz$, giving 100 frequencies between 0 and $1Hz$. It is not clear how this frequency resolution is achieved in relation to the 8 point average stated in the manual. This lack of clarity in the documentation makes an objective definition of the degrees of freedom associated with each spectral estimate difficult.

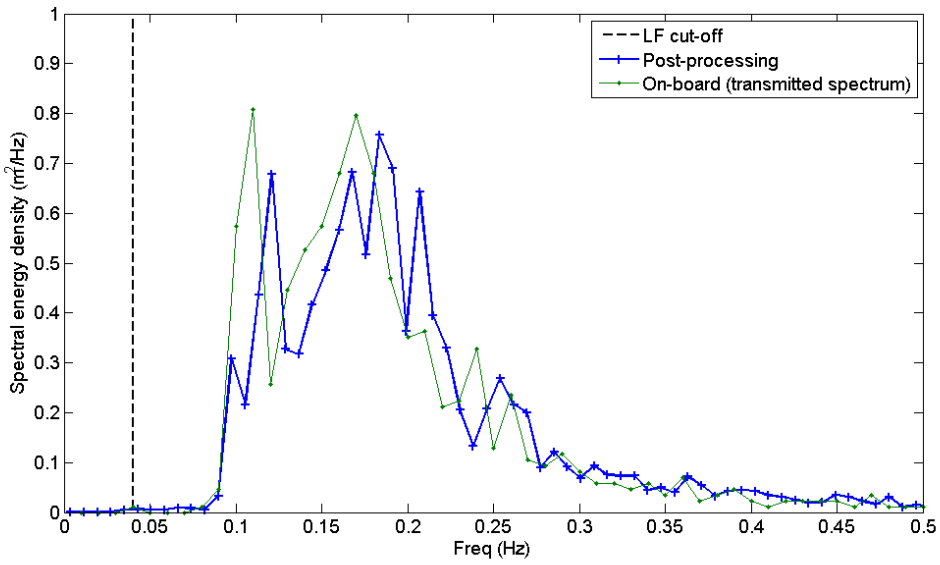

Figure 4.4: Sample variance density spectra, estimated using post-processing of the surface elevation time series in figure 4.1 , and on-board by the 'Wavesense' package.

In the initial storage and transmission setup for the buoys, concerns regarding the capacity of the transmission setup to send this larger (than summary parameters alone) data package meant that the transmitted spectra were stored as 7 bits. This compression meant that the resolution of the spectral energy densities was $0.78125m^2/Hz$, which is too coarse and did not define the spectral variations. In order to mitigate this, an optimisation scheme was initiated to test the transmission of spectra at a finer resolution. This gave positive results, with spectra successfully transmitted across a range of conditions.

However, to increase the resolution of the spectrum whilst keeping the data package size small meant that the upper limit of the spectral energy recorded was reduced. Analysis of the spectra from large wave events has shown that this limit was set too low, resulting in a loss of information at the peak frequency (fig. 4.3). Testing continues, with the aim that the transmission of accurate and useful spectra will be incorporated into standard processing for all buoys. This does not affect parameter estimation on-board, which is performed prior to the digitisation of the spectrum. In order to derive useful spectra from data already captured, an automated post-processing spectral analysis function was written using Matlab and was routinely applied to all time series data.

In post-processing spectral analysis, the Matlab function **spectrum** (Mathworks, 2007) is used to apply a fast Fourier transform (FFT) using Welchs periodogram method to estimate the one dimensional wave spectrum. Each record is 1024 seconds, which is sufficient that spectral leakage can be expected to be negligible and no window is applied (Tucker and Pitt, 2001). The raw spectral energy densities, $S(f)$, are averaged across 8 consecutive values. This method gives a frequency resolution of $0.078Hz$ and corresponds to guidance given in published guidelines and standard Tucker (1993),Tucker and Pitt (2001), Smith et al. (2006b), Pitt (2009), and Ingram et al. (2011). An example spectrum is shown in figure 4.4, where it is compared to a corresponding spectrum calculated on-board the buoy.

These spectra show a good general agreement in terms of the frequencies present in the wave record. A difference in frequency resolution can be observed, and greater variability (spikiness) is also observed in the post-processing spectrum. This is potentially indicative of greater variability, suggesting increased averaging for the on-board spectrum to achieve the resolution defined in the setup.

4.3.1 Calculating spectral moments

Table 4.2 shows the spectral parameters that are calculated on-board the buoy. Many of these are based on the spectral moments, which are calculated from the estimated spectrum using equation (2.1.12). The $S(f)$ values used are those of the raw spectral estimates (prior to averaging) between $0.04Hz$ and $0.5Hz$, corresponding to wave periods of $25s$ and $2s$ respectively. The high-frequency cut-off is imposed by the dimensions of the buoy (sec. 3.2), and will remove valid wave energy at the higher frequencies. Although consistent with the published recommendations (Smith et al., 2006b; Ingram et al., 2011), the lower cut-off frequency for the spectral estimates is higher than the expected frequencies of some

waves that are encountered in UK waters. Further discussion of this effect can be found below in section 4.8.

Spectral moments were also calculated from the post-processed spectra. This followed the same methodology of integration of the raw spectrum prior to averaging, although the frequency cut-offs applied on-board were not initially applied. Comparison between the data sets allowed a full discussion of the effect of the frequency cut-offs, which can be found below in section 4.8.

Table 4.2: Frequency-domain parameters calculated on-board the SW for a standard setup at the PRIMaRE buoy array.

Parameter	Definition
H_{m0}	Significant wave height, $4\sqrt(m_0)$
m_0	0th moment of the spectrum
m_1	1st moment of the spectrum
m_2	2nd moment of the spectrum
m_4	4th moment of the spectrum
Q_p	Spectrum Bandwidth parameter (2.1)
R_b	Width of spectral peak, $R_b = f_1 - f_2$ where $S(f_1) = S(f_2) = 0.1S(f_p)$
$S(f_p)$	Maximum spectral density
S_{m02}	Significant steepness (frequency-domain), $S_s = H_{m0}/(1.56T^2_{m02})$
Tl	Lower wave period of significance: 5% of the energy is at lower periods
T_{m01}	Estimate of mean wave period: $T_{m01} = m_0/m_1$
T_{m02}	Estimate of T_z, $T_{m02} = \sqrt{m_0/m_2}$
T_{m24}	Estimate of crest period, T_c, $T_{m24} = \sqrt{m_2/m_4}$
T_{mm10}	Energy period, T_E, $T_{mm10} = m_{-1}/m_0$
T_{mm20}	Estimate of T_{s1}, $T_{mm20} = \sqrt{m_{-2}/m_0}$
T_p	Period at spectral peak, $1/f_p$
T_u	Upper wave period of significance: 5% of the energy is at higher periods
$H_{m0}lf$	Low frequency significant wave height for a frequency band between $0.04Hz$ and $0.1Hz$
Directional analysis	
th_{lf}	Low frequency mean wave direction, for a frequency band between $0.04Hz$ and $0.1Hz$
m_{dir}	Mean direction across spectrum
spr_{tp}	Directional spreading at the spectral peak, $S(f_p)$
th_{hf}	High freq mean wave dir for δf between $0.8f_s/2Hz$ and $0.9f_s/2Hz$ where f_s is the sampling frequency
th_{tp}	Mean wave direction at the spectral peak period

4.4 Data archive

A data archive was developed to store raw and processed data as well as all available information that may be required for future analysis of the raw data. It is made larger and more complex by incorporating four buoys, and has been designed to accommodate data from more buoys and deep water ADCP's in the future. A sensible and traceable archive structure is vital in ensuring that a data set can be efficiently extracted, in which the researcher can have confidence. Bad archiving will at best result in an increase in the effort required to prepare the data for analysis, and at worst render the data useless due to the omission of some important metadata. The steps of processing to produce the files for the archive are detailed in figure 4.5. These consist of three separate Matlab^{TM} files, raw time-series, wave spectra (calculated during post-processing) and parameters (calculated on-board) for each buoy, each month. In addition, monthly text files are also stored for access using software other than Matlab^{TM}.

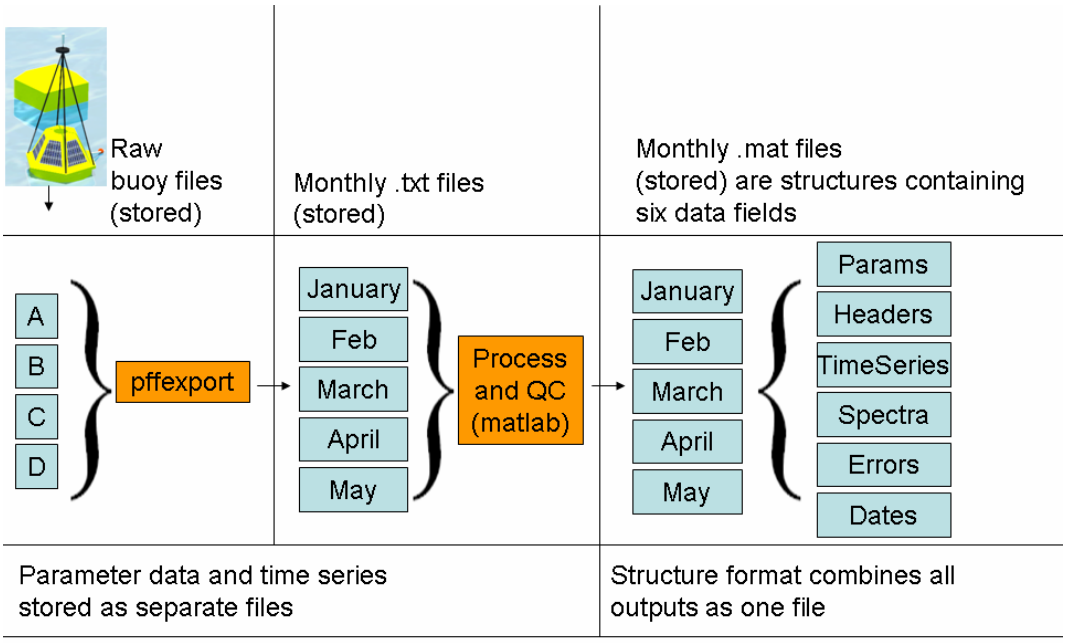

Figure 4.5: Representation of the steps in processing the PRIMaRE wave buoy data.

The working group QARTOD addressed the issue of standardising data collection for wave measurements and within their report suggested the metadata that should be kept with a file as follows (NOAA, 2005):

1. Data set title, person responsible for the data.

2. Instrument information (manufacturer, model, serial number)

3. Location of sensor. The standard convention for locations within PRIMaRE is to use latitude and longitude in degrees and decimal minutes
4. Water depth (m)
5. Date, time of each measurement
6. Data attributes
    - Parameters calculated and terminology
    - Record length and sample rate
    - Analysis technique, software
    - QC flags and additional QC info (tests applied)
    - Spectral parameters and spectral info (frequencies, dof, . . .)
7. Legal restrictions on using the data and disclaimers.

Where possible, the author has introduced the collection of these metadata as part of the routine data collection and processing, which are automatically stored with each file. In addition, a digital log-book detailing all operations undertaken with the buoys is stored alongside the archive. This is updated when required, and used to identify records where the sensor was not in routine operation and remove them from the data set.

4.5 Comparing results from different processing methods

This section describes the comparisons between the data derived from on-board processing and those data derived from post-processing the time-series files on-shore. This analysis is seen to be a powerful tool for identifying irregular characteristics of a data set and the results within this section highlight discrepancies in data derived from different processing methods. Irregularities are investigated, and conclusions are drawn as to the optimum processing methods for this data.

A full description of the processing applied on-board and on-shore can be found in sections 4.2 and 4.3. The principal difference between the processing methods is the frequency filters that were applied on-board, but not during post-processing. For frequency domain parameters, the filter is applied when calculating the spectral moments. For direct time-series analysis, each time series is subject to an FFT, and the spectral filter is applied before an inverse FFT (IFFT) is used to re-generate the time-series. On-board time-series

parameters are then estimated from this re-generated time-series, whereas those estimated during post-processing are based on the raw time-series.

Figure 4.6 compares the raw time-series to a filtered time series[1] and distinct differences between them can be seen. The effect of the frequency filter to the crest and trough levels is not consistent, increasing some wave heights whilst decreasing others. The high frequency filter means that the filtered data set is notably smoother than both the raw data and the interpolated data, missing some small variations in the data set. The effect of these differences on the calculated summary parameters that are investigated in this section, with a view to validating the processing techniques and providing a standardised data processing methodology to be adopted in further research.

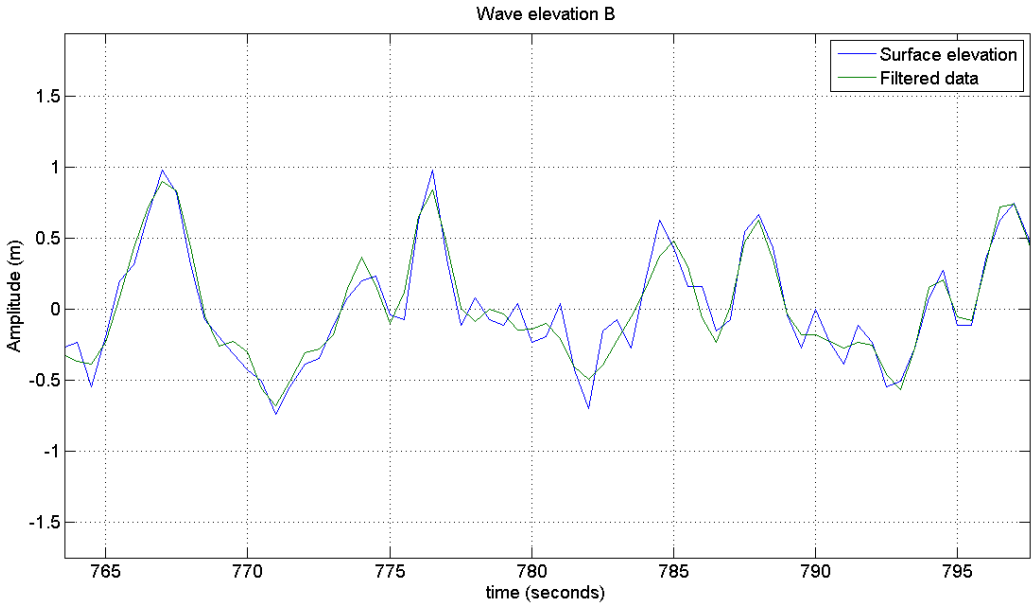

Figure 4.6: A subset of a surface elevation time-series captured on 23rd December 2009 between 06:30 and 07:00. Shown are the raw data and data that have been filtered using an FFT/IFFT frequency filter.

4.5.1 Definition of the data set used to evaluate processing methods

The sample data set was defined using data collected from buoy B via a cable connection, comprising raw time-series and summary parameters. This was carried out upon retrieval after mooring failure (November 2009), and subsequent to re-deployment during a routine service visit (April 2010). For the second deployment, the anchor positions were not

[1]On-board, filtered time-series are not stored and the filtered time-series shown was generated by replication of the FFT - filter - IFFT process on-shore.

moved, although the buoys were deployed to different positions within the array. Buoy B moved from the Northerly position to the Westerly position 3.1.

At the time of analysis, records measured over 6 months from 15^{th}-Oct-2009 until 30^{th}-Apr-2010 represented the longest available data set of both time-series files and summary parameters. The processing described in section 4, and summarised in figure 4.5 was applied to this sample data set from buoy B. Quality control (QC) procedures were also applied and an examination of data quality based on the results of QC tests can be found in chapter (5) for this sample data set. Consultation with the log-book removed 77 records (1%), and the error flags removed a further 178 (1.8%) of records leaving a data set comprising 8196 individual records.

Time-domain parameters of interest were the significant wave height, H_s and wave period, defined as the significant period[2], T_s. In addition, parameters that are used in QC procedures, H_{max} and S_{max} were examined. Spectral parameters were investigated through examination of the spectral moments as well as parameters P_t, H_{m0}, T_{m02} and T_p.

4.5.2 Methods of comparison

For each pair of parameters to be compared, a time-series plot of the sample data sets was plotted, and the following summary statistics calculated for each parameter for the sample data set,

- mean value
- maximum value
- standard deviation of values
- z statistic of the difference between the mean values

A scatter diagram was plotted for each pair of parameters, an ordinary least squares (OLS) regression was used to produce a line of best fit, and a corresponding R^2 value was computed using the methodology described in section 2.3. The limitations of OLS methods for two variables with similar errors means that the slope, β_1, should be viewed with caution (section 2.3). To quantify this bias, the OLS line of best fit was also computed regressing x on y, and where a significant difference between results was identified, both linear equations are displayed on the relevant scatter plot.

[2] T_s is not commonly used however it is the only zero-crossing period parameter that is calculated in the standard on-board processing.

The frequency-density function of the pairs of parameters were calculated using a kernel density plot, as estimates of the probability density functions of both parameters. These were plotted onto the same axis allowing direct comparison between the distribution of the pair of parameters. In the case of the significant wave height parameters, the frequency density functions were plotted onto a Weibull distribution axis, where data from a Weibull distribution would be a straight line. Similarly, wave period parameters were plotted on a Gaussian axis.

In addition, a z statistic was used to define whether the difference in the mean of the two parameter sets is statistically significant. It was defined as follows,

$$z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\bar{\sigma}_A^2}{n_A} + \frac{\bar{\sigma}_B^2}{n_B}}} \tag{4.5.1}$$

where $\bar{\sigma}$ is the mean sampling variability and n is the number of values in the sample. Where the two samples are drawn from the same population, with Gaussian error distributions, z would be a random variable from a standard Gaussian distribution (mean = 0 and std deviation = 1). Thus, a null hypothesis that the two data sets represent different samples from the same population can be tested based on the probability density function of the standard Gaussian distribution. The probability of a computed value of z arising from a standard Gaussian distribution can be computed from the probability density function,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2 2\sigma^2} \tag{4.5.2}$$

where μ represents the mean and σ the standard deviation (0 and 1 respectively for a standard Gaussian. Where $|z| > 2.96$, $p(z) < 0.01$, and it is concluded that the difference between the data sets is deemed to be significant to a critical level of 99%.

4.5.3 Sampling variability

Sampling variability has been reviewed in section 2.2, and section 2.3 established its importance when comparing data sets. Here, sampling variability for $H_{1/3}$ values was assumed to be equal to that for H_{m0} (Tucker and Pitt, 2001). Goda (1977) estimates the sampling variability of the parameter T_s, using multiple simulations of wave profiles. This provides an estimate for a record of duration 500s of 4.8%. However, this value will be dependent on the wave state within the record, and may not provide an accurate estimation for the data examined here.

An estimation of the sampling variability for T_s, S_{max} and H_{max} was calculated directly from the data. This process is described in detail in appendix A, and engenders the

assumptions that the sea state at two separate buoys in the PRIMaRE buoy array is equal, that the error, ϵ is Gaussian distributed and a constant proportion of the true value of the parameter, $\epsilon = cX$, where c is a constant and X is the true value of the parameter in the underlying sea state. See chapter 8 for further discussion of sampling variability.

4.6 Results of comparisons between processing methods

This section provides the results of comparisons between data processed on-board the buoy, and data calculated during post-processing the time-series files on-shore. In this chapter, the latter are defined using a *pp* subscript. For example, H_{m0} represents the spectral significant wave height value calculated on-board, whilst H_{m0pp} represents the value calculated during post-processing.

4.6.1 Comparisons between zero-crossing wave height parameters

The z statistic comparing the mean of H_s calculated on-board and that calculated on-shore during post-processing, H_{spp}, indicates that H_{spp} values are, on average less than the corresponding H_s. The z value ($p(z) < 0.01$) indicates that this difference is significant to a critical level of 99%. The regression between the two data sets shows a good agreement with a small spread reflected in an R^2 value of 0.997 (fig. 4.7). There is no indication of larger proportional differences at larger sea states, suggesting that the errors in this measurement are not dependent on wave height[3]. The probability distribution highlights the relatively coarse resolution of the on-board measurements. The larger waves of each data set are seen to follow a Weibull distribution closely. For lower wave heights, there is a lower probability of occurrence in the on-board measurements.

Maximum wave height, H_{max}

For the purpose of studying H_{max}, data with the error flag 10, which flags very large waves, were included in the analysis (see chapter 5) . The values of H_{max} are interspersed with large spikes, which are seen below to be flagged for a large wave during quality control (fig. 4.8).

The temporal structure and location of maxima in H_{max} and H_{maxpp} parameters are closely matched in the time-series. Frequent spikes are correlated between the two data

[3]The errors in these measurements will be related to the sea-state and therefore highly correlated. That the error between them doesn't increase with sea-state does not suggest that the absolute error in either measurement does not increase with sea-state, as would be expected.

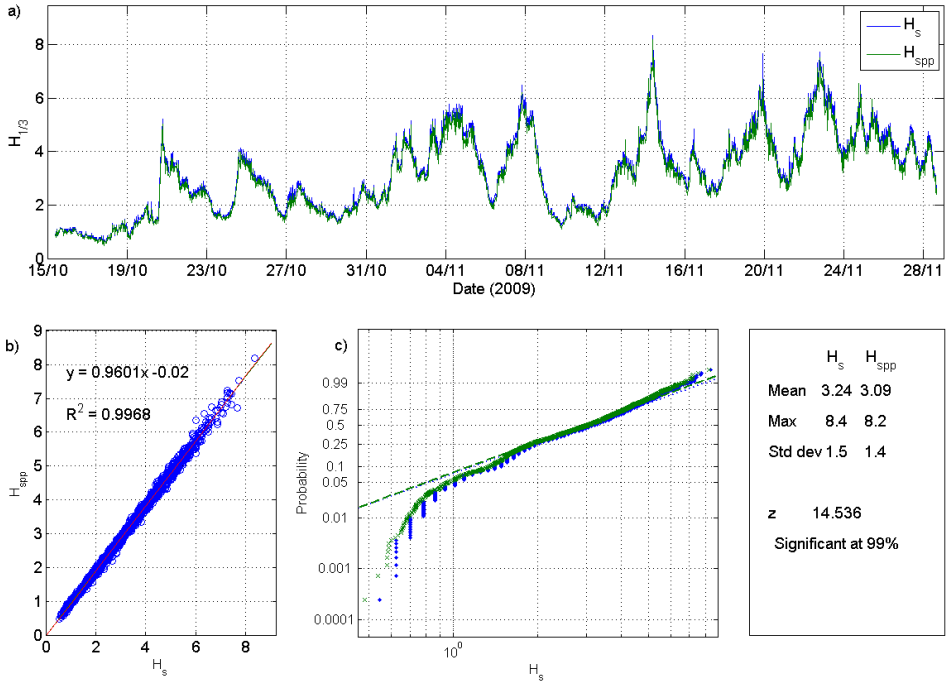

Figure 4.7: a) A time-series of significant wave height calculated as the highest 1/3 of zero-upcross waves in a filtered time-series record on-board the buoy (H_s) and from the raw time-series during post-processing (H_{spp}). b) A scatter diagram of H_s vs H_{spp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of values of H_s and H_{spp}, plotted on a Weibull-scale where a straight line indicates that the data follow a Weibull distribution.

sets, and in some cases have an error flag value, 10. The similarities between the data lead to a z statistic of -1.86, for which $p = 0.07$, and it cannot be concluded that H_{maxpp} is significantly different to H_{max}. The two OLS lines of best fit that were computed support the conclusion from the z statistic, that there is no significant difference between the two data sets. However, the scatter diagram shows significantly more spread than that for significant wave height, with a reduced R^2 value of 0.957. The spread in the scatter diagram increases with an increase in H_{max}, and contributes to a difference of $2.3m$ in maximum H_{max} between the two processing methods. Differences between the maximum wave measured of this magnitude have the potential to affect the prediction of extreme wave heights using these data.

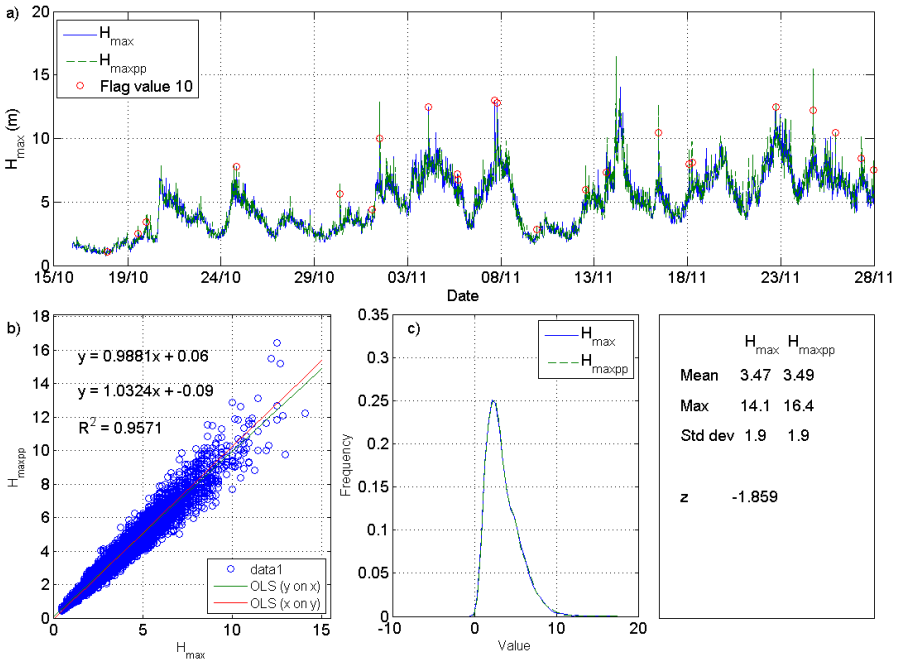

Figure 4.8: a) A time-series of maximum wave heights calculated as the highest wave in a filtered time-series record on-board the buoy (H_{max}) and from the raw time-series during post-processing (H_{maxpp}). b) A scatter diagram of H_{max} vs H_{maxpp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of different values of H_{max} and H_{maxpp} during the data set, calculated using a kernel density estimate.

4.6.2 Comparisons between zero-crossing wave period parameters

The significant period, T_s is the only wave period parameter output by the on-board processing and was compared to T_{spp} values (figs. 4.9a-c). The sampling variability was estimated using the procedure outlined above, which gave a standard error of 0.05, very close to the value of 0.048 calculated by Goda (1977) for a record of duratoin 500s. T_{spp} values are consistently larger than T_s leading to a strongly negative z value, for which $p \approx 0$, indicating a significant difference between the means. In the time-series, there is at least one period where the general trend is reversed, and $T_s > T_{spp}$. The occurrence of such records is also evident in the scatter diagram, with a group of points exhibiting low T_s values relative to T_{spp} (fig. 4.9b).

The parameter $\Delta T_s = (T_s - T_{spp})T_s$ was calculated to represent the proportional difference between concurrent values. This shows a strong dependence on sea state, with values representing small wave heights and a short period showing $T_s > T_{spp}$, whereas for other sea states, $T_s < T_{spp}$ (fig. 4.10).

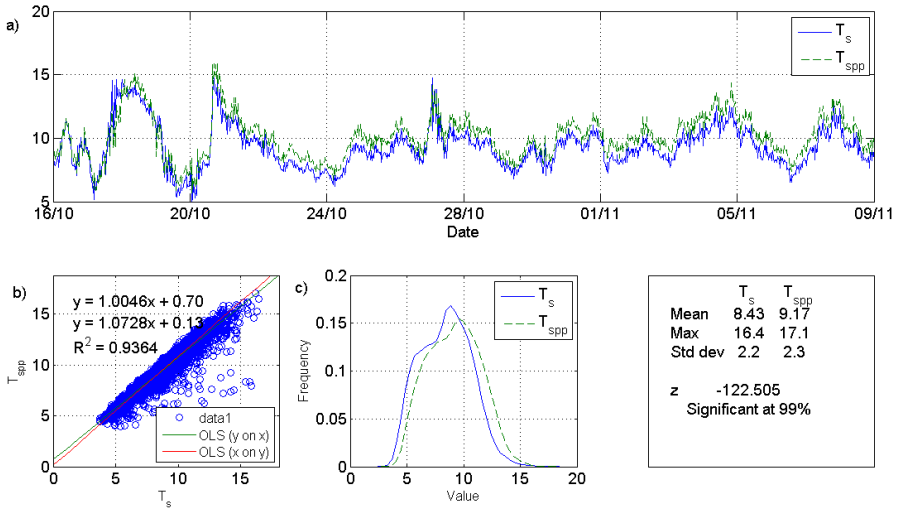

Figure 4.9: a) A time-series of significant period calculated as the mean of the longest 1/3 of zero-upcross waves in a filtered time-series record on-board the buoy (T_s) and from the raw time-series during post-processing (T_{spp}). b) A scatter diagram of T_s vs T_{spp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of different values of T_s and T_{spp} during the data set, calculated using a kernel density estimate.

Figure 4.11a is a plot comparing the time-series of a raw data set, and a filtered data set along with the wave periods identified by zero-crossing analysis. This record was measured during a period where, $T_s = 12.6s$, calculated on-shore from a filtered time-series was

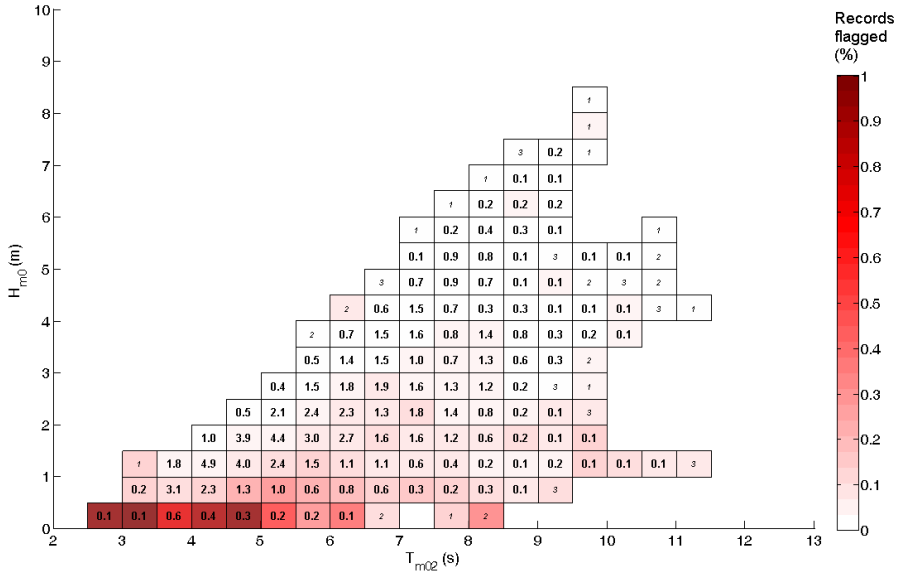

Figure 4.10: An H-T scatter diagram computed with H_{m0} and T_{m02} from data for buoy B between October 2009 and April 2010. The text values represent the percentage of records occurring in each cell, values in italics represent the total number of records occurring within that cell, whilst the colour overlay represents the mean value of ΔT for records in each cell.

greater than $T_{spp} = 10.2s$, calculated from the raw time-series. The filter removes high-frequency oscillations, and reduces the number of zero-crossings identified. Conversely, figure 4.11b where $T_s = 8.18s$ and $T_{spp} = 10.1$ shows an example of how the removal of low-frequencies in the time-series means that the values are more consistently close to the mean level and zero-crossings are increased.

It can be concluded with reasonable certainty that the periods where the significant period calculated on-board exceeds that calculated during post-processing correspond with low wave heights and short periods, and are caused by the high-frequency filter affecting the zero-crossing analysis. Outside of these periods, the low-frequency filter increases the frequency of zero crossings and the significant period calculated from the raw spectra exceed that calculated from a filtered time-series.

4.6.3 Comparisons between zero-crossing steepness estimations

The variability of S_{max} values was estimated using the two concurrent data sets from the PRIMaRE buoy array, which estimated $\sigma(S_{max}) = 0.14\bar{S}_{max}$, a large standard error when compared to other parameters. Maximum steepness values are consistently reduced when the filter is applied and despite the large standard error, the mean values were deemed

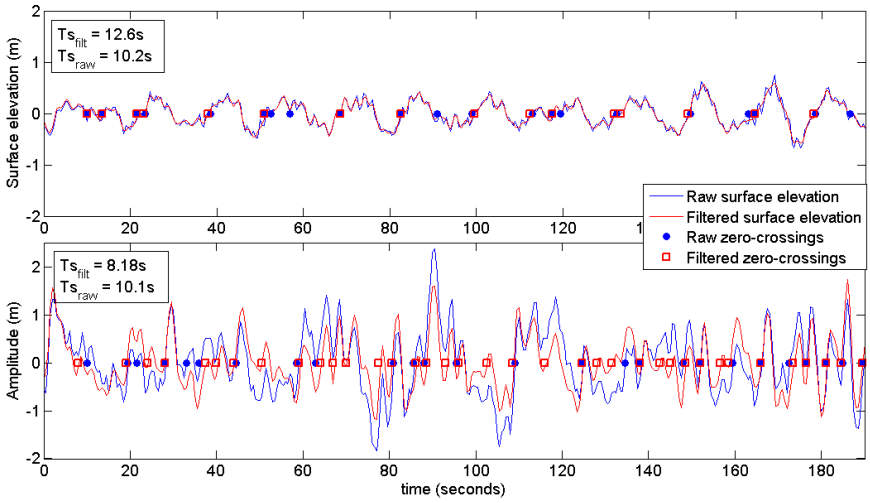

Figure 4.11: A raw time-series of surface elevation, and a filtered time-series of surface elevations for a) a record where the filter is seen to increase the significant wave period and b) a record where the filter decreases the significant period.

significantly different by the z statistic of -2.98, for which, $p(z) = 0.047$ (fig. 4.12). This result indicates that the steepest wave in a raw time-series is consistently of frequency higher than $0.5Hz$, which will be excluded by the frequency filter applied on-board.

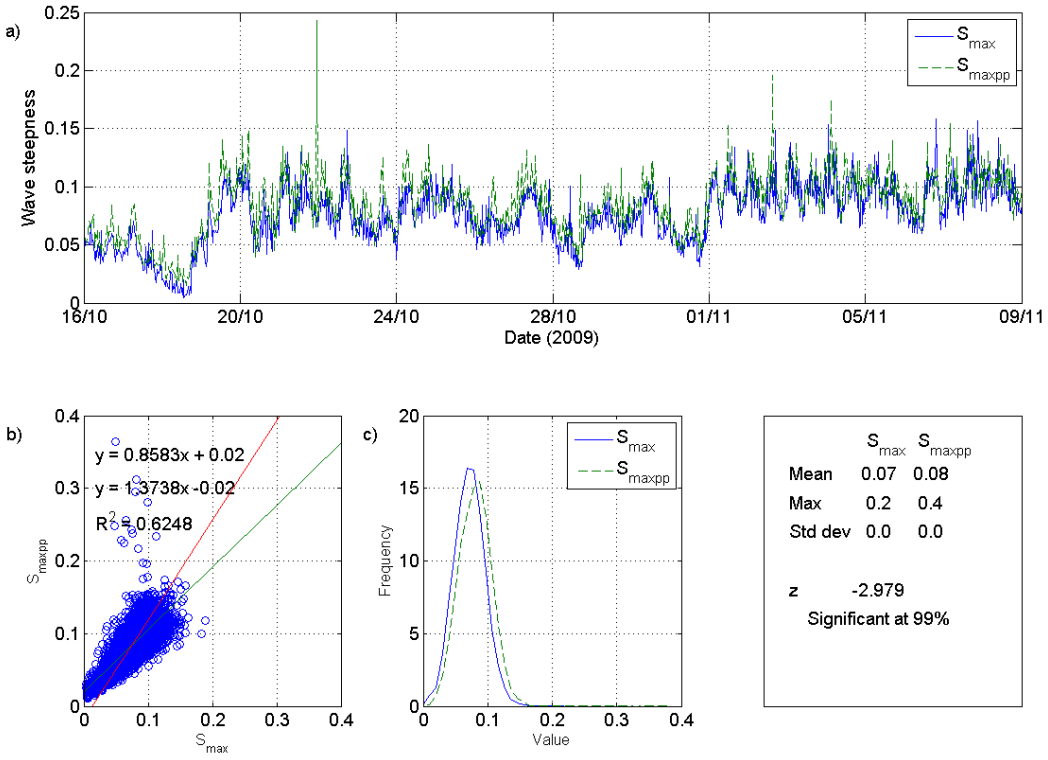

Figure 4.12: a) A time-series of maximum steepness calculated from zero-upcross waves in a filtered time-series record on-board the buoy (S_{max}) and from the raw time-series during post-processing (S_{maxpp}). b) A scatter diagram of S_{max} vs S_{maxpp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of different values of S_{max} and S_{maxpp} during the data set, calculated using a kernel density estimate.

4.6.4 Comparisons between spectral moments

The effect of the time-series filters on many spectral parameters will be governed by the effect on the spectral moments. This was examined for spectral moments, m_{-1}, m_0, m_2, m_4 and P_t. A time-series of differences reveal a consistent difference between raw and filtered values in m_2 and m_4 values (fig. 4.13). For m_4, This bias means that filtered values are less than the equivalent raw values, by a factor of approximately 2 (fig. 4.14).
Table 4.3 shows the differences caused by the spectral filter for mean values of spectral moments and wave power where,

$$\Delta(raw, filt)(\%) = \frac{m_{n,raw} - m_{n,filt}}{m_{n,raw}} \times 100 \tag{4.6.1}$$

The effect of the filters on the mean values is dominated by the high frequencies and is seen to increase with the order of the spectral moment.

Table 4.3: Mean values of spectral moments calculated from raw spectra, and from spectra with spectral filters applied.

Parameter	Raw	Filtered	$\Delta(raw, filt)(\%)$
$P_t(kW)$	16.89	16.47	2.41
m_{-1}	2.14	2.09	2.25
m_0	0.25	0.24	3.34
m_1	0.04	0.03	10.12
m_2	0.01	0.01	29.61
m_4	0	0	167.52

Scatter diagrams of raw (x axis) against filtered (y axis) m_{-1} and m_0 values demonstrate a number of outlying values where the filtered value is less than the equivalent raw value (fig. 4.14). These differences in lower order moments are attributed to the effect of the low-frequency filter, whose effect is not consistent. The time-series of lower order moments also show short periods with large raw values, which are not replicated by the filtered data (fig. 4.13).

4.6.5 Comparisons between spectral parameters

Significant wave height values derived from the spectrum compare well between methods of processing, both in the structure of the time-series and in their values (fig. 4.15). Although $p(z) = 0.011$, it is not sufficiently small that the difference in the means be

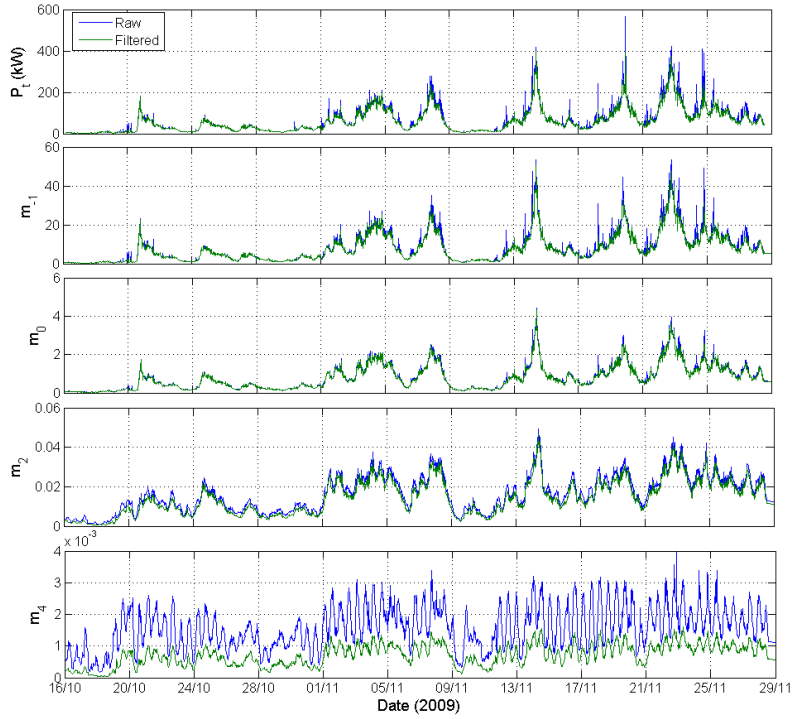

Figure 4.13: Time-series of spectral moments for records from buoy B captured during October and November 2009.

deemed significant at the 99% confidence limits. There is also good agreement between T_{m02} and T_{m02pp} on all plots in figure 4.16. A slight bias is present in the scatter diagram, which is also notable in the statistics and leads to a difference in the means (and associated z statistic that is deemed as significant at 99%.

Peak spectral period, T_p

The peak spectral period was not estimated using the spectral moments, rather it is the frequency at which the maximum spectral energy density occurs. The scatter diagram, or mean values of these parameters, demonstrate that there is no consistent difference between raw and filtered values. In both data sets, a number of large spikes can be identified, which are also evident in the frequency distribution, leading to a small maxima in the extreme values (fig. 4.17b). However, these maxima are at different values for the two data sets causing a large difference in the maximum values, $T_p = 24.9s$ and $T_{p,pp} = 28.44s$. The value of $24.9s$ corresponds to the low-frequency cut-off applied on-board, and is therefore the maximum value that this parameter can take when calculated from a filtered spectrum.

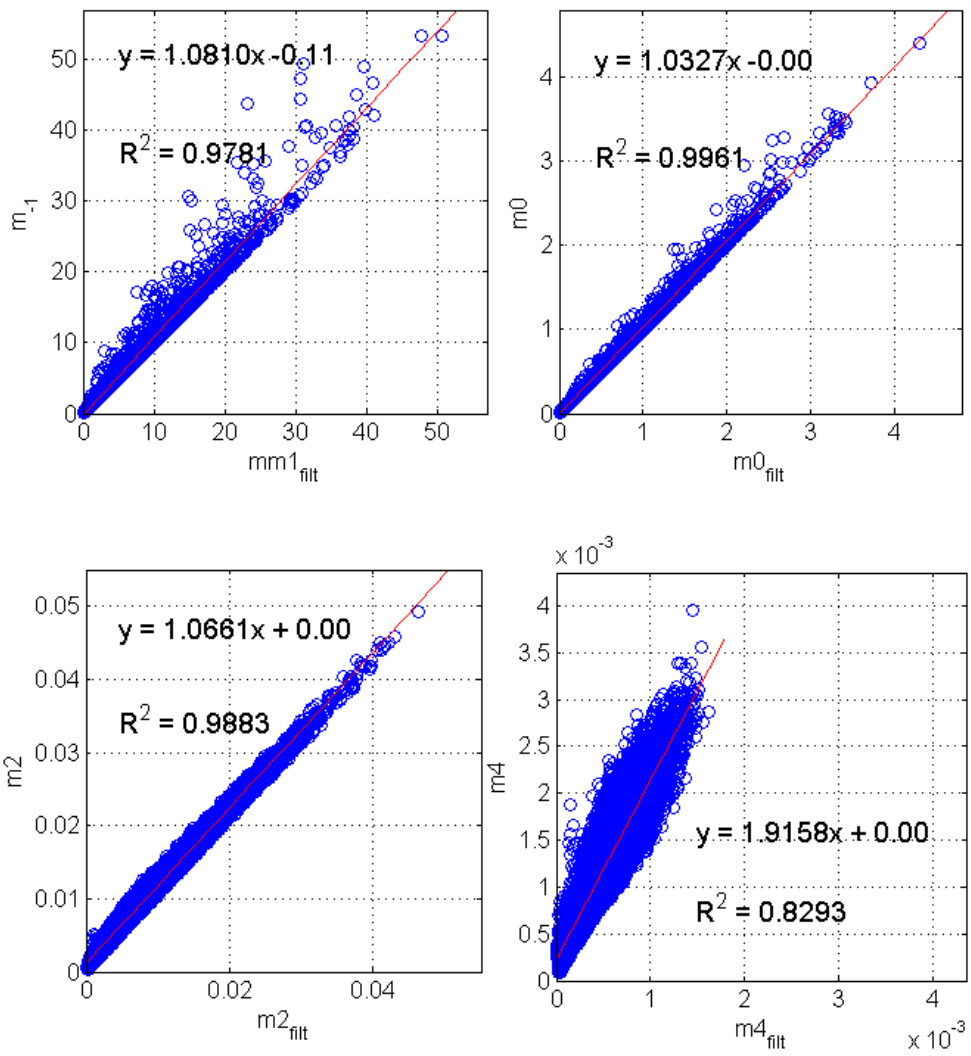

Figure 4.14: Scatter diagrams of spectral moments estimated from a raw time-series (x axis) and from a filtered time-series (y axis), for records from buoy B captured during October and November 2009.

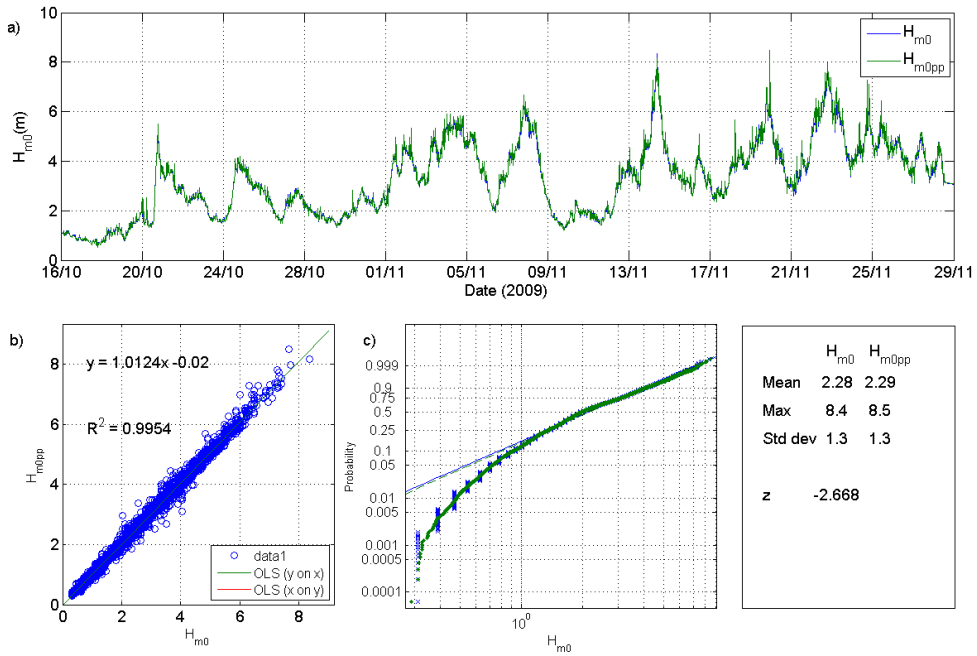

Figure 4.15: a) A time-series of significant wave height calculated from spectral analysis on-board the buoy (H_{m0}) and during post-processing (H_{m0pp}). b) A scatter diagram of H_{m0} vs H_{m0pp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of H_{m0} and H_{m0pp} values, plotted on a Weibull-scale where a straight line indicates that the data follow a Weibull distribution.

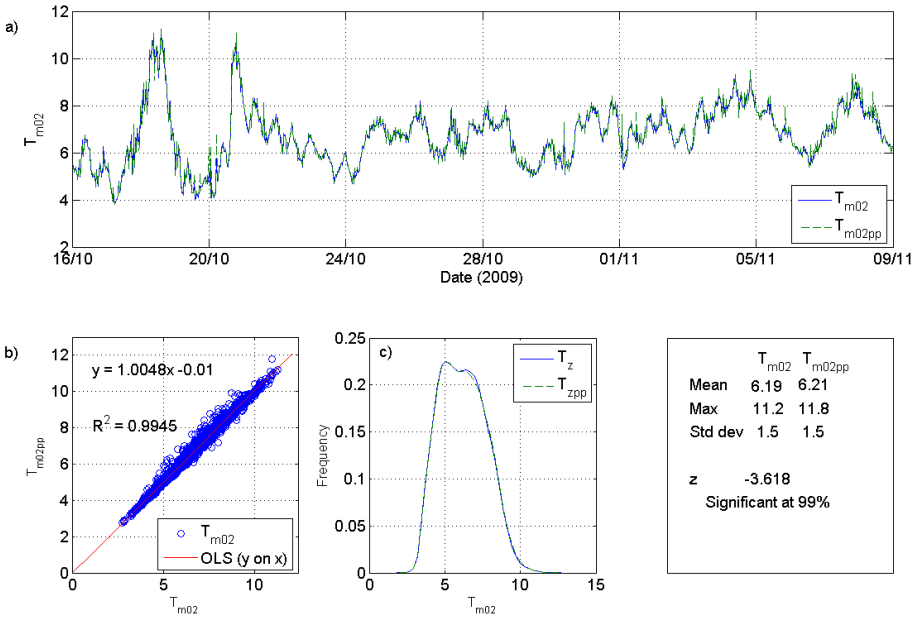

Figure 4.16: a) A time-series of mean zero-crossing period, T_{m02}, calculated from spectral analysis on-board the buoy and during post-processing (T_{m02pp}). b) A scatter diagram of T_{m02} vs T_{m02pp} with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of different values of T_{m02} and T_{m02pp} during the data set, calculated using a kernel density estimate.

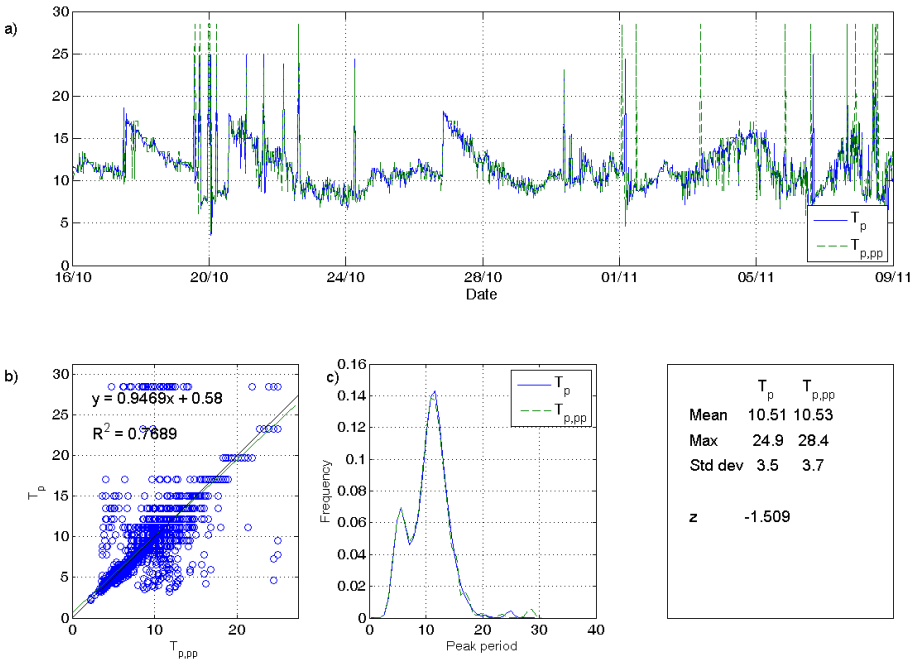

Figure 4.17: a) A time-series of peak period, T_p, calculated from spectral analysis on-board the buoy and during post-processing ($T_{p,pp}$). b) A scatter diagram of T_p vs $T_{p,pp}$ with a least-squares line of best fit, the associated equation and R^2 value. c) The frequency of occurrence of different values of T_p and $T_{p,pp}$ during the data set, calculated using a kernel density estimate.

4.7 Summary of the differences between processing methods

Differences between the two methods of processing have been seen to cause differences in the values for all parameters derived from zero-crossing analysis. These differences can be attributed to the application of spectral filters for both time-domain processing and spectral processing within the 'standard' processing.

Time-domain analysis for the majority of records found that the low-frequency filter increases the number of zero crossings by removing long oscillations and keeping the time-series closer to the mean water level. In contrast, removal of high-frequencies reduces the number of zero-crossings by smoothing the time-record and removing small oscillations that may cross the mean water level. In general, the low-frequency filter has the dominant effect, and for the majority of records, more zero crossings are identified because the removal of low frequencies keeps the time-series closer to the mean water level. As a consequence, the filtered wave period is less than that derived from a raw time series. However, when wave heights are small, the effect of the high-frequency filter becomes dominant, reducing the number of zero-crossings, and increasing the wave period estimated from the filtered data when compared to the raw time-series.

In all records, the high frequency filter had a consistent effect on the maximum steepness parameter, and it can be concluded that the steepest waves have frequencies above $0.5Hz$.

Comparison of the spectral moments derived from a filtered spectrum and those from a raw spectrum demonstrate that the filters are affecting all spectral moments. However, the effect on derived parameters was in general less than that observed in parameters derived from zero-crossing analysis. In contrast to the time-series analysis, a consistent effect was observed from the high frequency filter, whereas the effects due to the low frequency filter were sporadic.

High frequency energy was consistently removed by the filter causing differences between values for higher order moments in most records. These results demonstrate the sensitivity of the higher order moments to the spectral cut-off, and the use of parameters estimated using moments of order above 2 is not recommended.

Differences between the values of low order moments calculated with and without filter occur when a large amount of spectral energy is present at very low frequencies ($< 0.04Hz$). This effect caused values of T_p and $T_{p,pp}$ in excess of $20s$, with T_p commonly taking the highest possible value given the frequency filter $24.9s$. High values correspond with spikes

in the lower-order moments and occur when low-frequency spectral energy is present in the measured time-series, leading to a spectral peak centred at a frequency between $0.03Hz$ and $0.05Hz$. This is a sporadic occurrence, which does not affect all records, but it is observed to have a dramatic effect on values of T_p and $T_{p,pp}$, and other low order spectral moments.

In summary, the differences between data derived from the two methods of processing can be attributed principally to the application of spectral filters. Parameters derived from spectral moments of order 2 or below are seen to be more stable than those estimated from zero crossing analysis. However, for certain records in the data, the values of T_p indicate spurious low frequency spectral energy affecting both the filtered and the raw spectrum. In response, further work was undertaken to examine the validity of the application of this low frequency filter, including a detailed discussion of spectral energy that can be expected at these frequencies (sec. 4.8).

4.8 Review of spectral filters

The previous section highlighted the effect of frequency filters applied in standard processing on the derived wave parameters. This effect can be attributed to instances where significant spectral energy is present below $0.04Hz$, and above $0.5Hz$. These filter levels are not applied arbitrarily, rather they refer to physical limitations of the system.

The wave frequency operating limits of a surface following buoy have been discussed in section 3.2. The high frequency limit is clearly defined in the documentation supporting the buoys as twice the diameter of a buoy (Sanmuganathan, 2009). The $0.5Hz$ filter applied during Oceanor on-board processing corresponds to a deep water wavelength of $3.1m$, which is approximately twice the diameter of the buoy. As this is based on a physical constraint of the measurement system, it is removing data that cannot be verified, and its application is not discussed further.

The definition of the low-frequency filter was not as clearly defined in the documentation, although it has been suggested in section 3.2 that the mooring response governs the low-frequency operating limit. Validation of this statement would require a detailed hydrodynamic response analysis of the coupled buoy-mooring system, a procedure that is beyond the scope of the research presented here. Furthermore, this would likely duplicate previous work carried out by the manufacturer[4] that led to the definition of the standard filter as a cut-off set at $0.04Hz$

Energy is regularly measured below $0.04Hz$ which has been seen to affect parameter estimation from filtered spectra (sec. 4.5). The research described in this section was designed to verify that the filter was not excluding valid spectral energy, leading to errors in the parameters. As such, it contains a review of low frequency measurements in the literature and considers the wave conditions and geography of the sensor location to examine the potential spectral energy for this site. Closer examination of the captured wave spectra led to the conclusion that the measured low frequency energy was erroneous, and that the filter applied in standard processing was not fully effective at removing these errors. Examples of measured data are presented, which were used to develop alternative strategies for removing spurious data from the affected records. The outcome is a better understanding of this phenomenon and a robust strategy for mitigating errors in the measured data.

The analysis concentrates on the sample data set for buoy B used in the previous section.

[4]The author has not been provided details of or access to any such research. However, in discussions with the manufacturer, hydrodynamic modelling of the moorings systems have been alluded to

However, where necessary, data from all four buoys is used in order to highlight potential differences between sensor buoys, mooring sets and installation locations. The deployments were all carried out over the same period, although buoys A C and D failed prior to buoy B in November 2009 and their data sets are shorter (see chapter 6 for details).

4.8.1 Review of low-frequency wave measurements

Snodgrass et al. (1966) tracked the propagation of swell across the Pacific from storms in the southern regions to the far north using direct wave measurements spread out from the Southern Ocean to the far north of the Pacific. During this experiment, waves of frequencies below $0.04Hz$ were measured as fore-runners of swell events. However these contained a small amount of energy, typically waves with height less than $50cm$, and the duration for which the peak wave frequency was below $0.04Hz$ was less than $4hrs$. Within the Atlantic, long-range swell propagation is also possible. Vassie et al. (2004) observed swells from hurricanes in the northern hemisphere in wave measurements from islands in the southern hemisphere, although spectral data was not published. Hamilton (1992) identified swell propagating from the southern ocean in data collected from 19 floating wave buoys situated near the north-west Atlantic coast (US eastern seabord). The dispersion relationship dictates that waves with lower frequencies will travel faster than those with shorter frequencies. It follows that from a distant storm, the lowest frequency waves will arrive at a measurement point first. Hamilton (1992) measured swell with a peak frequency of $0.04Hz$, seen to progress to higher frequencies as the swell developed, with f_p reaching $0.05Hz$ in 4 hrs (fig. 4.18). An increase in the peak frequency with time was also observed by Snodgrass et al. (1966). This is in contrast with Work (2008), who observed low-frequency spectral energy with no temporal progression in directional wave buoy[5] measurements taken off the Georgia coast, US eastern seabord. Because no temporal progression was evident, the conclusion drawn was that these were not valid measurements and were attributed to noise in the accelerometer measurements.

4.8.2 Low frequencies at the PRIMaRE buoy array

As wave systems propagate, spectral energy migrates towards lower frequencies (Hasselman et al., 1973). Therefore, as well as the initial intensity of the storm, the frequency of the waves generated will also depend on the distance travelled. The lowest frequency waves at the buoy location are likely to arise from swell systems propagating from a distant

[5]Work (2008) used a TRIAXYS directional buoy manufactured by AXYS measurements.

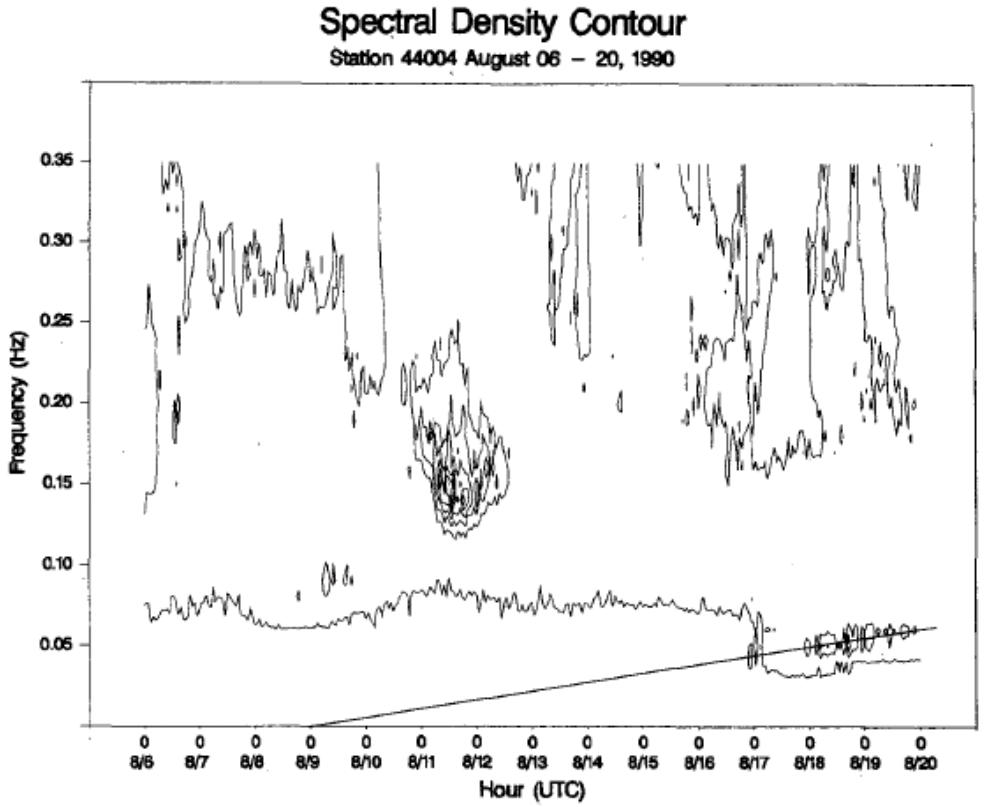

Figure 4.18: An example of temporal progression of swell waves in the literature (Hamilton, 1992).

storm which has exhibited very strong winds, as was seen with measurements in the north Pacific (Snodgrass et al., 1966). The geography and available fetch to the measurement site have been discussed in section 3.1 and the lowest frequency spectral energy, due directly to incident waves, will originate from hurricanes and tropical storms in the western Atlantic.

Four examples of records with significant low-frequency spectral energy are shown in figures 4.19, all of which show a peak of low frequency spectral energy centred below, and extending above the spectral cut-off. Accurate interpretation of these results requires a decision on whether the energy associated with this spectral peak is valid, or erroneous. If the low frequency system is deemed to be erroneous, it must be excluded from the spectrum in its entirety. Otherwise, if it is considered valid it must be included. This figure indicates that the fixed $0.04Hz$ spectral filter is not a valid treatment of the measured data because it excludes only part of the spectral peak.

Because the peaks span the spectral cut-off, they have an effect on parameters estimated from a filtered spectrum. A variable spectral filter was defined by eye for these four records, to eliminate the whole of the spurious low frequency peaks. Table 4.4 quantifies the effect of

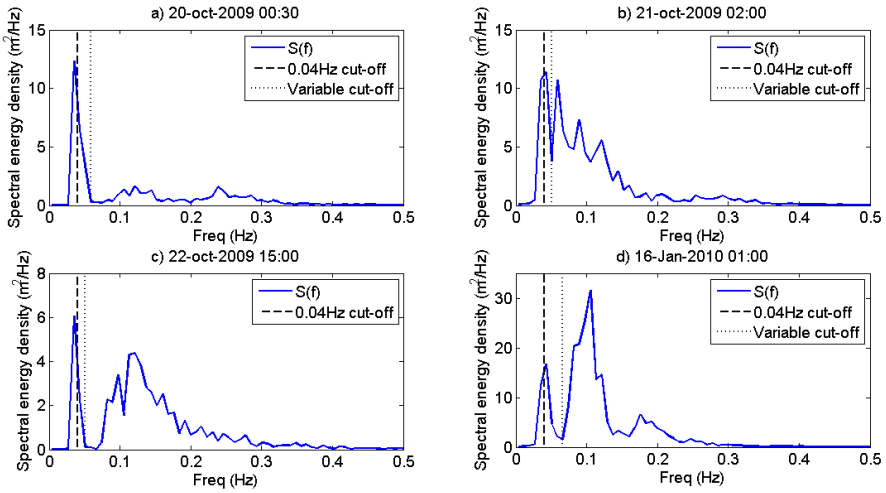

Figure 4.19: The omnidirectional variance spectrum for four records exhibiting anomalous low frequency spectral energy. Also shown is the Oceanor standard $0.04Hz$ low frequency cut-off, and a variable cut-off, estimated by eye as the limit of the influence of measured low frequency energy on the spectrum.

Table 4.4: The values of H_{m0} calculated for the spectra in figure 4.19 using raw spectral values, spectral values excluding those below $0.04Hz$, and spectral values using a variable low frequency cut-off, estimated by eye.

Date and time	20/10/09 00:30	21/10/09 2:00	22/10/09 15:00	16/01/10 1:00
Conf limits (90%)	+/- 0.07	+/- 0.17	+/- 0.14	+/- 0.30
H_{m0} raw	1.87	3.64	2.66	5.39
H_{m0} $0.04Hz$ cut-off	1.73	3.34	2.49	5.13
H_{m0} variable cut-off	1.65	3.23	2.45	4.95

the filters on the records in figure 4.19. Confidence limits were estimated using equation (2.2.5), and assuming a Gaussian error distribution of H_{m0}. The maximum difference between H_{m0} derived from the raw spectra and those derived using the variable cut-off for these records is $45cm$, and in all four cases exceed predicted sampling error. The standard filter does not remove all the energy associated with the spurious spectral peak, which affects to the parameter estimations. Furthermore, where these peaks are not present in all buoys simultaneously, they have the potential to be highly significant for comparisons between data sets. Therefore, it is essential for this research to establish whether these are valid measurements which should be included in analysis, and if not, to implement a filter process that fully removes their influence on the active spectrum.

4.8.3 The prevalence of low-frequency spectral energy through the data

The effect of the low-frequency filter on parameter estimation will be governed by the relative magnitude of energy below $0.04Hz$ compared to the energy in the rest of the spectrum. This was quantified using the ratio of spectral energy measured above $0.04Hz$ to that below as follows,

$$r_{lf} = \frac{m_{0lf}}{m_{0filt}} \tag{4.8.1}$$

where $m0_{lf} = \int_0^{0.04} S(f)df$ and $m0_{filt} = \int_{0.04}^{0.5} S(f)df$.

r_{lf} was calculated for each buoy for records between October 2009 and April 2010, a period that comprises two deployments. Subsequent to surface mooring failure during November 2009, the mooring system up to the top-mark float were left in place and the buoys were re-deployed during December 2009. For the second deployment, the buoys were re-deployed to a different mooring set than they had occupied previously. The position of buoys B and D were exchanged, as were buoys A and C. Results for these two deployments were considered seperately. In order to quantify the proportion of records that were affected, all records were categorised acording to their r_{lf} value (as shown in table 4.5).

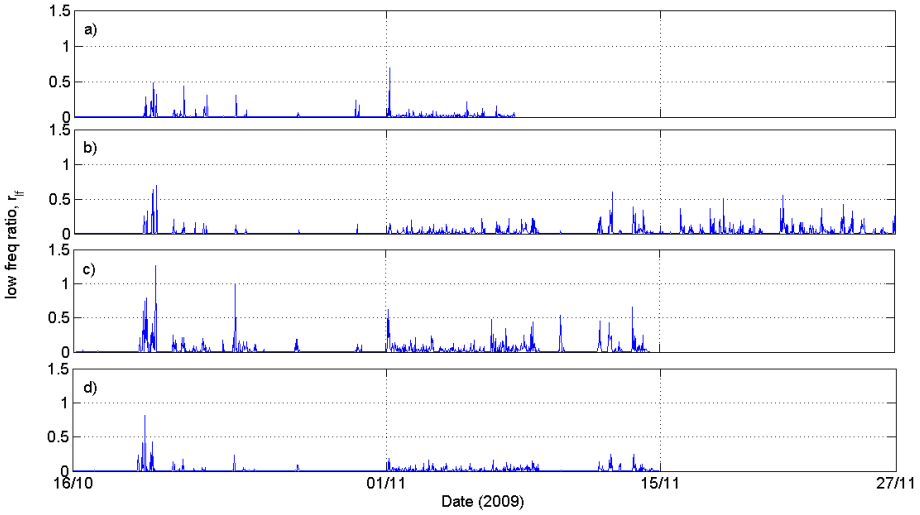

Figure 4.20: A time-series of the values of r_{lf} for deployment 1, a) data collected by buoy A, b) buoy B, c) buoy C and d) buoy D.

High values of r_{lf}, which indicate a large influence of low frequency energy, occur throughout the measurement period in all buoys (figs. 4.20 and 4.21), although they were more common during the first deployment (tab. 4.5). r_{lf} is seen to be well correlated in time between the buoys with the exception of a period during March 2010 where large values of r_{lf} in data collected by buoy D are not replicated in data from the other buoys (fig.

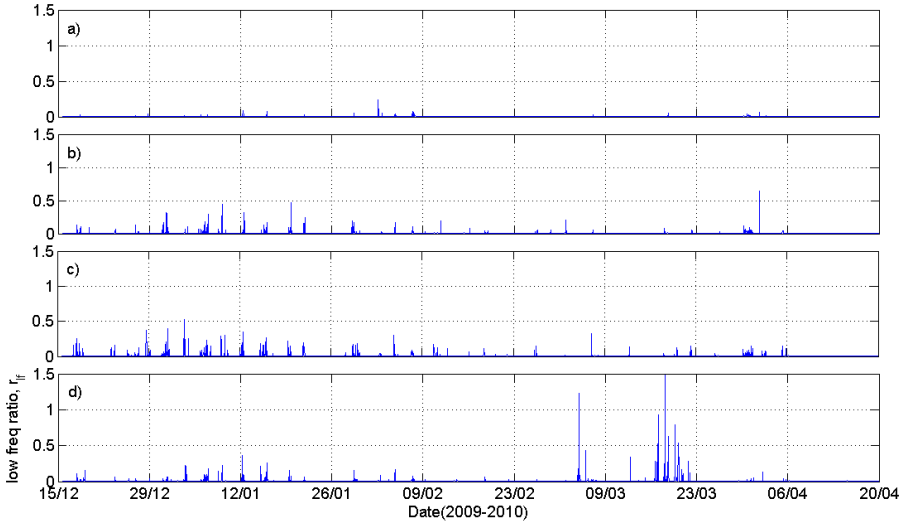

Figure 4.21: A time-series of the values of r_{lf} for deployment 2, a) data collected by buoy A, b) buoy B, c) buoy C and d) buoy D.

4.21d).

The proportion of data that fall into each category is given in table 4.5. For all buoys, more records were categorised as having negligible r_{lf} values during the second deployment. The relative number of records classified with negligible r_{lf} shows it to be more closely linked to the buoy itself rather than position. There is no evidence that the location of the buoy within the array, or the specific mooring set used has a dominant effect on the frequency of occurrence of low frequency energy.

Figure 4.22 shows the temporal progression of the omnidirectional spectrum for one week during October 2009 during which three distinct wave systems can be identified. The first commences during the 20th October with low-frequency waves arriving first, and the peak frequency increasing as the high frequency waves with lower phase velocity follow. This shape is typical of a swell event. The second commences during the 24th October, with high frequency waves, and the peak frequency decreases before stabilising. This shape is typical of a system generated by local winds. Finally, a smaller swell system can be identified commencing on the 27th October. In contrast, the occurrence of low frequency ($< 0.05Hz$) peaks are seen to be sporadic throughout the period, and cannot be attributed directly to wave systems incident on the buoys, Occurrences are observed prior to and throughout the duration of the first two wave systems measured at the buoys, with fewer occurences were measured during the final event, which is of smaller magnitude in terms of $S(f)$. In all cases, the spurious peaks can be seen to extend across the frequency

Table 4.5: The fraction of records for each buoy that occur in categories defined by r_{lf} values for both deployments.

Deployment 1, Oct - Nov 2009				
	A	B	C	D
$r_{lf} < 0.013$	0.84	0.76	0.70	0.81
$0.013 < r_{lf} < 0.06$	0.13	0.14	0.16	0.14
$0.06 < r_{lf} < 0.1$	0.02	0.05	0.05	0.03
$0.1 < r_{lf} < 0.3$	0.02	0.05	0.07	0.02
$r_{lf} > 0.3$	0.01	0.01	0.02	0.00
Deployment 2, Dec 2009 - Apr 2010				
$r_{lf} > 0.013$	0.99	0.97	0.94	0.96
$0.013 < r_{lf} < 0.06$	0.01	0.02	0.03	0.03
$0.06 < r_{lf} < 0.1$	0.00	0.00	0.01	0.01
$0.1 < r_{lf} < 0.3$	0.00	0.01	0.01	0.01
$r_{lf} > 0.3$	0	0.00	0.00	0.00

cut-off.

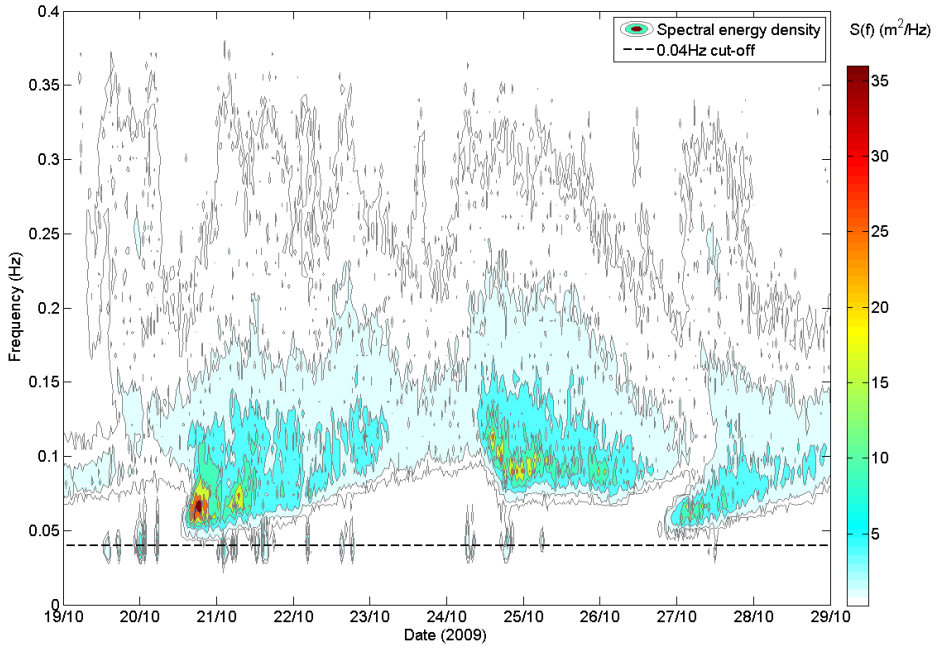

Figure 4.22: The temporal progression of spectral energy for a period during October 2009, as measured by buoy B.

4.8.4 Effect of low-frequency energy on the active spectrum

Prior to implementation of a process to eliminate low-frequency energy from individual records, a critical question that arises is whether the occurrence of the spurious low-frequency oscillations will affect the measurement of other frequencies in the spectrum. If this is the case, it would be necessary to re-evaluate the validity of the measurements entirely, as opposed to applying a filter to remove the erroneous low-frequencies.

Using the classification of records according to their r_{lf} value, the spectral shape of all categories were normalised according to the spectral energy density at the peak frequency, $S(f_p)$, and plotted on a logarithmic scale (fig. 4.23). f_p was calculated as the frequency above $0.05Hz$ that corresponds to the maximum value of $S(f)$. The mean spectra for each category show the effect of the spurious low frequency noise but converge at around $0.06Hz$. Above this frequency, the spectra are similar between the categories.

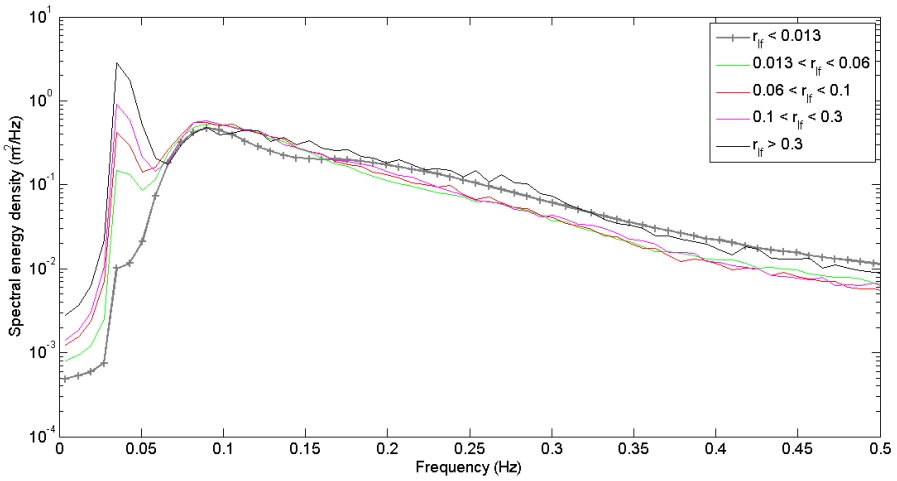

Figure 4.23: Mean of normalised spectra for each category in table C.3, plotted on a logarithmic scale.

A further examination, of whether active frequencies in the spectrum are affected by the erroneous low frequency energy, compared simultaneous records where low-frequency noise was observed in one of the buoys, but not in others. 54 records of this type were identified for buoy C, 48 for buoy D, 10 for buoy B and only 2 for buoy A.

Individual cases show a good agreement between spectra containing an erroneous low frequency system and those measured at the same time by the other buoys in the array (fig. 4.24). However, other examples showed some deviation in the spectrum in the low-frequencies (fig. 4.25).

The mean $S(f)$ values from all 54 records where low-frequency energy was observed only

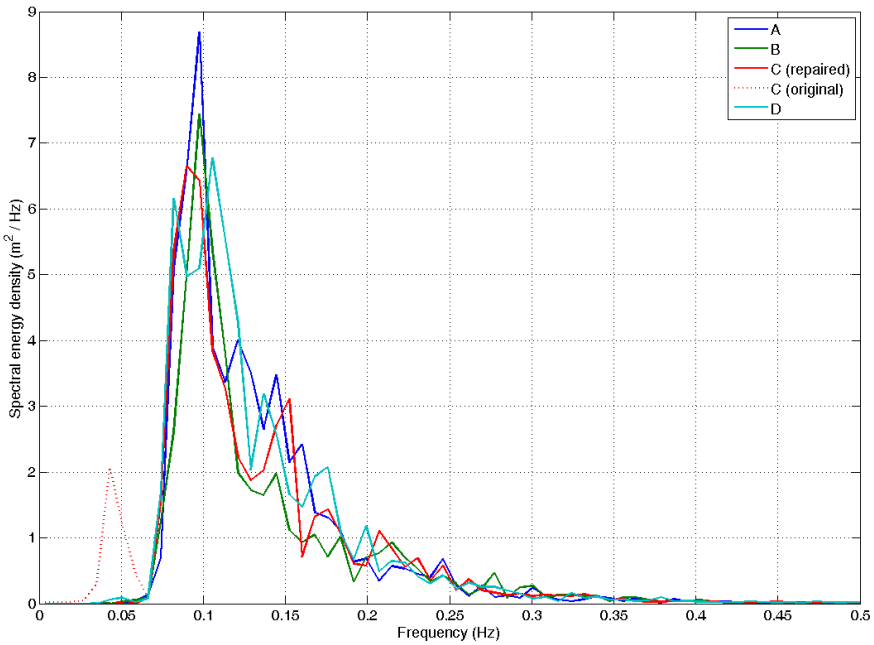

Figure 4.24: The original and repaired spectra for buoy C and the original spectra for all buoys from the same record. In this case, the repaired spectrum for buoy C is closely matched to those from the other buoys.

in data from buoy C were plotted against the mean $S(f)$ values for the same 54 records from all the other buoys combined (fig. 4.26a). The result demonstrates the erroneous low-frequency energy and a difference in the spectral peak value for data from buoy C which exceeds the confidence limits. The same process was carried out using 48 records where low-frequency energy was observed only in data from buoy D (fig. 4.26b). The agreement over the spectrum is good, although there appears to be a discrepancy in the location of the higher frequency spectral peak, which is outside of the confidence limits of the mean spectrum.

4.8.5 Discussion

Spectral energy occurring below $0.04Hz$ has been observed in data sets from all 4 buoys. The lowest frequency waves expected from a local storm can not be expected to produce energy below $0.04Hz$ at the scales that have been observed in the data from all buoys. Therefore, the spectral energy at low frequencies could only arise from the propagation of distant storms. Examples of this type of measurement in the literature display temporal progression of the wave frequencies, caused by the dispersion equation, which shows that

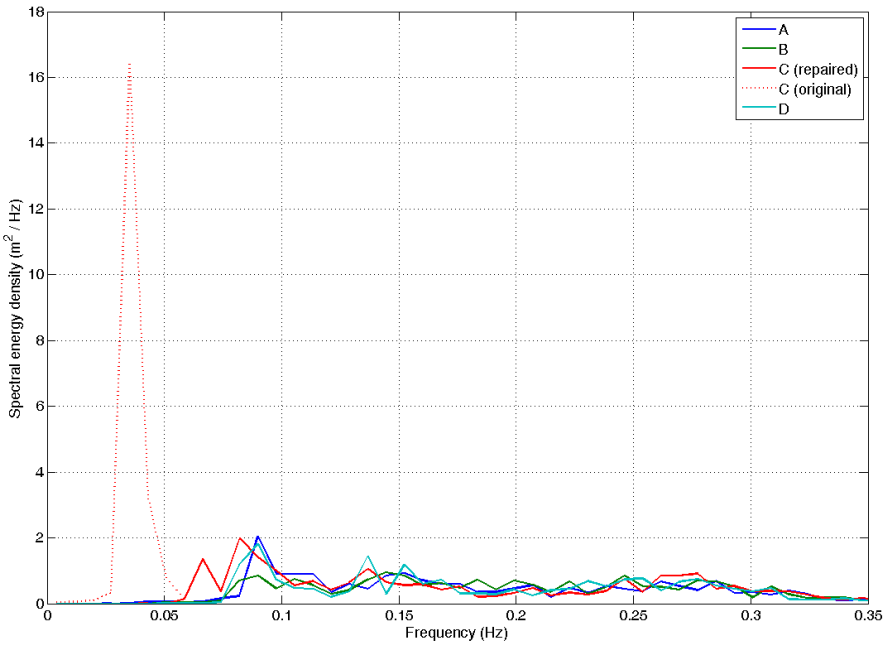

Figure 4.25: The original and repaired spectra for buoy C and the original spectra for all buoys from the same record. In this case, the repaired spectrum for buoy C and spectra from the other buoys are not well matched in the low frequencies.

waves of the lowest frequencies will travel fastest. Therefore, the lowest frequencies in a wave record represent fore-runners of swell wave systems. However, this was not observed for spectral energy below $0.04Hz$ in the data studied here (fig. 4.22), which indicates that the low-frequency energy measured by the wave buoys cannot be caused directly by naturally occurring ocean waves.

A marked difference in the frequency of occurrence of low-frequency energy and the magnitude of r_{lf} values have been observed between the data from different buoys (figs. 4.20 and 4.21). However, comparison of data between the buoys also showed no clear dependence of the occurrence of low frequency spectral energy with the position of the buoy within the array, or the mooring set with which it was tethered. There is some evidence that it is the individual sensor buoy that governs the frequency of occurrence of r_{lf}. Therefore, it must be considered that internal systems affect the measurement of low-frequencies.

Further investigation into the potential causes of the erroneous low frequency measurements was conducted, to aid definition of a suitable mitigation strategy. This work is provided in appendix C and concluded that the observed errors could not be accounted for solely by noise in the accelerometers. Erroneous low frequency energy was seen to oc-

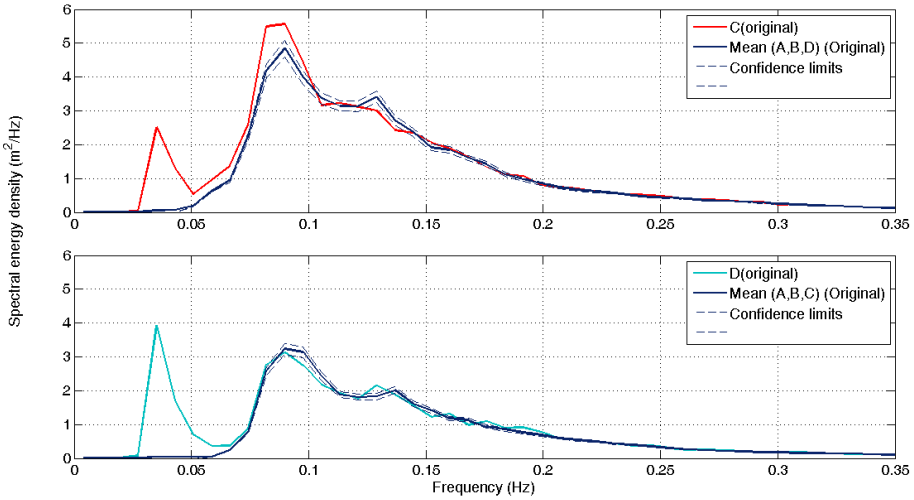

Figure 4.26: a) The mean spectrum for buoy C for all records where buoy C was seen to measure spurious low-frequency energy but other buoys were not, and the mean spectrum for buoys A,B and D from the same records b) The mean spectrum for buoy D for all records where buoy D was seen to measure spurious low-frequency energy but other buoys were not and the mean spectrum for buoys A,B and C from the same records.

cur more frequently in developing seas and large sea states, conditions in which the force applied to the surface mooring and the instrumentation buoy by the waves will be large. Furthermore, the occurrence of low-frequency spectral energy is more frequent when the tidal flow is broadly in alignment with the waves and wind. Therefore, the occurrence and magnitude of erroneous low frequency energy increases with larger drift forces.

Examination of the spectral data has demonstrated the effect of the erroneous energy affecting spectra at frequencies below $0.06Hz$. When the low frequency energy is removed, the active spectra have been observed to match those captured at the same time from other sensors in the array (fig. 4.24 and 4.26b)). However, certain cases have demonstrated differences in the active spectral frequencies between spectra. Figure 4.26a) suggests that there may be an associated effect on the rest of the spectrum, although at this stage, deterministic differences, caused by spatial variability in wave conditions between the measurement sites, cannot be discounted. The conclusion drawn here is that there was no definitive evidence that the presence of the erroneous low frequency energy invalidates measurements of the active spectrum. Therefore, a filter acting in the frequency domain would produce a valid spectrum and associated parameters.

4.9 A variable low frequency filter

Figure 4.23 demonstrates that the low-frequency energy to be removed, commonly has an influence in frequency bands up to $0.06Hz$. However, a filter at this level would exclude valid data. Therefore, an updated filter was designed and automated using Matlab. The process interrogates the spectrum to identify the entire peak that corresponds to the low frequency energy and remove it. This results in a variable filter level and prevents spurious low frequency energy from affecting parameter estimation.

A spectral peak was identified as the maximum spectral value between two minima in $S(f)$ values. Where the maximum of a system occurs below $0.05Hz$, the entire system was discarded as erroneous. f_1 was set as the frequency associated with the second minima, and was considered the first valid datum. Where no maxima occurs below $0.05Hz$, f_1 was set at $0.043Hz$. All data at frequencies below f_1 were then set to zero, $S(f < f_1) = 0$.

In certain cases, $S(f_1)$ was not close to 0, and after filtering, the shape of the low frequency part of the spectrum was not representative of naturally occurring conditions. To overcome this, the Pierson-Moskowitz spectral shape for a fully-developed sea ((Pierson and Moskowitz, 1964)), was used to estimate $S(f)$ values at frequencies below f_1. This was based on the standard formula,

$$S(f) = \alpha g^2 (2\pi)^{-4} f^{-5} e^{-\beta(f/f_p)^{-4}} \tag{4.9.1}$$

Constant values used were $\alpha = 0.0081$ and $\beta = (5/4)$. The shape of this spectrum is defined solely by f_p. The value of f_p that would mean the spectrum passes through a given value of $S(f)$ can be found by re-arranging equation (4.9.1) to give,

$$f_p = f_i \left(\frac{log\left(\frac{S(f_i)}{cf^{-5}}\right)}{-\beta} \right)^{1/4} \tag{4.9.2}$$

where $c = \alpha g^2 (2\pi)^{-4}$. Setting $f_i = f_1$, the corresponding value of f_p is found. The Pierson Moskowitz spectral values for all frequencies below f_1 were calculated using equation (4.9.1), and used to replace the zero values imposed by the filter. This ensures that the low-frequency part of the generated spectrum will pass through the first valid datum, and values below this are representative of naturally occurring wave fields.

Although this spectral shape is designed for a fully developed sea, it is used here only for the lowest frequencies in the wave record. For frequencies above $0.06Hz$, the shape of the spectrum will depend increasingly on sea conditions, and the assumption that it follows a Pierson Moskowitz spectrum will become innaccurate. Therefore where the spectral peak

to be removed is seen to extend beyond $0.06Hz$, an error flag is raised and the repair is visually checked.

4.9.1 The result of applying a variable filter

Application of the automated variable filter and spectral repair process was tested on the four example records given above in figure 4.19, and results are shown in figure 4.27, where the scales have been altered from the original figure for clarity. In all cases, the filter removes all of the erroneous low frequency energy in the spectrum, whereas not all of this energy would be removed by the original $0.04Hz$ filter. When applied to the record captured on the 20-oct-2009 at 00:30 (Fig. 4.27a), $S(f)$ values were altered at frequencies above $0.06Hz$. As such, these values were flagged for visual inspection. The repair is seen to be a reasonable estimation of the spectral shape at the low frequencies, and has succesfully excluded the erroneous measurements.

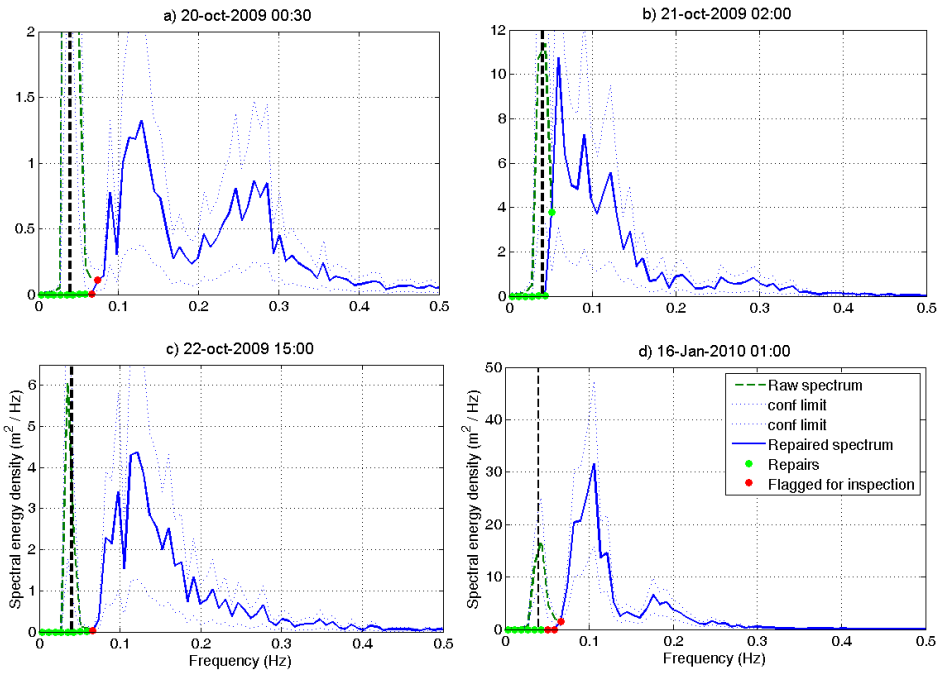

Figure 4.27: The adjustment applied by the repair function to the four records with low-frequency noise shown in figure 4.19.

The effect of the variable filter on spectral parameters will be governed by the change in spectral moments, which are displayed for the four test records in table 4.6. Because the filter affects the low-frequency measurements, the largest differences are observed in the lower spectral moments. A maximum difference of 49.53% between m_{-1} values using the

original filter and the variable filter was seen for the record captured on the 20-oct-2009 at 00:30, (fig. 4.27a) which exhibits proportionally the largest amount of low frequency energy out of the test cases. The variable filter demonstrates a significant difference on these test cases for both m_{-1} and m_0 values. The percentage differences are much lower for m_2, which is more strongly influenced by spectral energy at high frequencies in the spectrum.

Table 4.6: Spectral moments calculated for the 4 spectra in figure 4.19 with the $0.04Hz$ filter, and using the automated variable filter function.

	$m_{-1}(0.04Hz)$	m_{-1}(var)	$\Delta(m_{-1})(\%)$
20-oct-2009 00:30	1.88	0.95	49.43
21-oct-2009 02:00	7.87	6.84	13.08
22-oct-2009 15:00	2.96	2.72	8.08
24-oct-2009 19:30	10.00	8.88	11.18
	$m_0(0.04Hz)$	m_0(var)	$\Delta(m_0)(\%)$
20-oct-2009 00:30	0.21	0.16	20.71
21-oct-2009 02:00	0.70	0.65	6.35
22-oct-2009 15:00	0.39	0.38	2.74
24-oct-2009 19:30	1.05	1.00	4.85
	$m_2(0.04Hz)$	m_2(var)	$\Delta(m_2)(\%)$
20-oct-2009 00:30	0.01	0.01	1.14
21-oct-2009 02:00	0.01	0.01	0.71
22-oct-2009 15:00	0.01	0.01	0.84
24-oct-2009 19:30	0.02	0.02	0.51
	$P_t(0.04Hz)$	P_t(var)	$\Delta(P_t)(\%)$
20-oct-2009 00:30	14.83	7.50	97.74
21-oct-2009 02:00	61.95	53.85	15.04
22-oct-2009 15:00	23.27	21.39	8.79
24-oct-2009 19:30	78.72	69.92	12.59

The variable filter function was applied to data for buoy B captured between Oct 2009 and Apr 2010. Figure 4.28 shows the temporal progression of repaired spectral values using the same period as figure 4.22. This demonstrates that the low-frequencies have been removed while the naturally occurring structure of the spectral values is not affected.

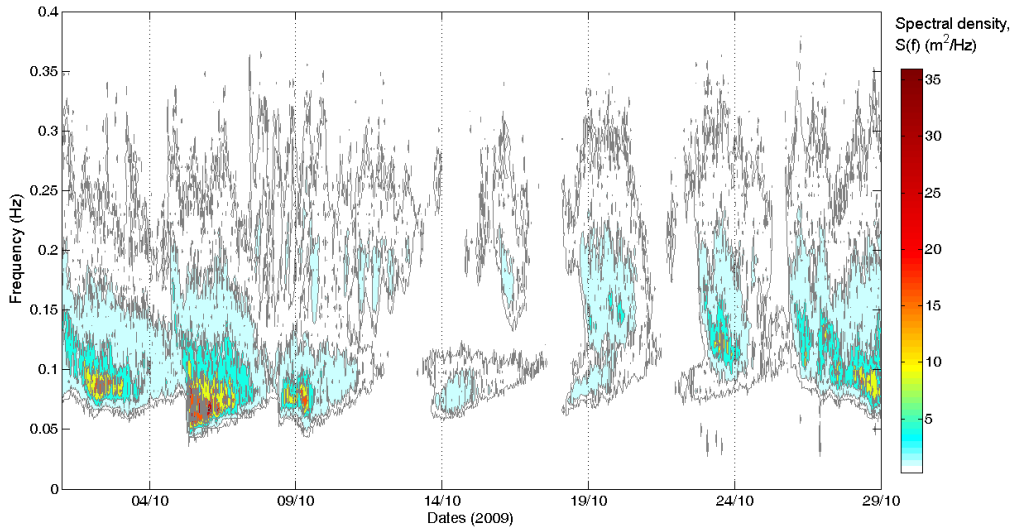

Figure 4.28: The temporal progression of spectral energy measured by buoy B for the period during October 2009 corresponding to figure 4.22, using spectra adjusted by the repair function.

The spectral moments, and total spectral power, P_t, were calculated from spectra measured by buoy B during Oct-Nov 2009 which had a variable filter applied; the high-frequency $0.4Hz$ filter was also applied. The differences between moments calculated from the variable filter with those calculated from the standard filter was less than 1% for all parameters (tab. 4.7). This demonstrates that the effect of the low frequency errors is limited in the long term data set.

Data from a single storm event were selected, occurring between $17th$ and $21st$ November, with maximum $H_{m0} = 6.5m$ and mean $H_{m0} = 4m$. Low frequency errors were observed to affect the data during this period, although it was not chosen because the effect was particularly prevalent, and is considered representative of medium storm events at the site. During this period, the spectral filters are seen to have a significant effect on the mean parameters (fig. 4.29 and tab. 4.8). The four examples presented demonstrate that the erroneous low frequency energy can have a significant effect on individual records. However, the mean values of spectral moments demonstrate that the effect of the erroneous low frequency energy is limited when analysing long-term data sets. Furthermore, the $0.04Hz$ spectral filter applied in standard processing does go some way to mitigate the effect. Where shorter data sets are used, the errors caused by these measurements will become more influential, and are seen to be significant for a selected period of 4 days. Analysis examining individual records must account for these errors, and the variable

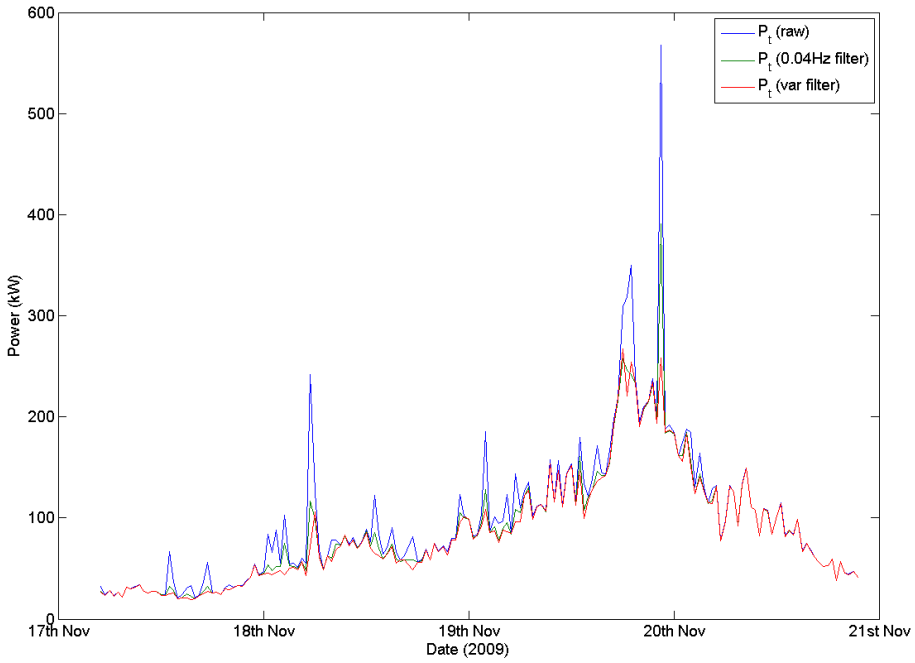

Figure 4.29: A time series of incident wave power, P_t for a single storm during November 2009. Shown are the raw, and filtered values.

Table 4.7: The mean values for spectral moments and total power, as calculated from raw spectra, from spectra with a $0.04Hz$ filter, and with a variable filter.

Parameter	Raw	$0.04Hz$ filter	var filter
m_0	0.24	0.24	0.24
m_1	0.03	0.03	0.03
m_{-1}	2.12	2.08	2.08
m_2	0.01	0.01	0.01
P_t (kW)	16.73	16.40	16.34

Table 4.8: The mean values during a single storm for total power, as calculated from raw spectra, from spectra with a $0.04Hz$ filter, and with a variable filter.

Parameter	Raw	$0.04Hz$ filter	var filter
P_t (kW)	97.11	86.47	88.98

filter has been seen to improve treatment when compared to the stationary filter. For records where the errors are not mirrored in simultaneous data, they will cause large differences between the data sets, which will significantly affect the analysis of differences. For such analysis, including that proposed for this thesis, application of the variable filter is essential.

4.10 Discussion

A comparison between parameters derived from the 'standard' wave data processing performed on-board the buoy, and that performed on-shore using functions developed in Matlab by the author, demonstrated that the definition of wave data processing can have a significant effect on the derived parameters. The principal difference between the methods was the application of spectral filters to the data prior to parameter estimation, which were applied on-board, but not during post-processing.

The application of the low frequency spectral filter to time-series prior to parameter estimation was seen to increase the wave heights measured, whilst reducing the wave period. This can be attributed to the low frequency filter increasing the number of zero-crossings identified in direct time series analysis. However, in certain conditions, the high frequency filter was seen to reduce the number of zero crossings and increase the wave period derived from a filtered time-series. The latter effect was limited to sea states with wave heights below $1m$, and the mean value of T_s was significantly less than that of T_{spp} due to the consistent effect of the low frequency filter.

No significant effect on the maximum wave height was noted, whilst the maximum steepness was significantly larger when estimated from a raw time-series. The effect on this latter parameter was taken into consideration when analysing the results of quality control tests, as discussed in chapter 5.

The application of a high frequency filter was seen to affect the definition of higher order wave parameters, particularly m_4, where regression analysis demonstrates that values from a filtered spectrum were, on average, half the magnitude of those derived from raw spectra (fig. 4.14). The effect of the high frequency filter reduced for lower order moments, and the difference in mean values of m_1 from a filtered spectrum and a raw spectrum was 10% (tab. 4.3). The high frequency limit is set by the dimensions of the buoy, and therefore the application of this filter is considered necessary, although this result indicates that any use of spectral moments of an order greater than 1, should consider its effect on the data.

Although differences between the mean values of lower order moments were relatively small, for individual records, m_{-1} values derived from a raw data set were seen to be 50% greater than the equivalent derived from a raw data set. These differences were attributed to distinct peaks of low frequency spectral energy partially removed by the spectral filter. These peaks occur throughout the data set for all buoys and have a significant effect on individual records, particularly the lower order spectral moments, wave power, P_t and the

peak period, T_p. Occurrence is sporadic and does not conform to the temporal progression of frequencies exhibited by naturally occuring wave systems. Therefore, the spectral energy associated with these peaks was deemed erroneous.

Further analysis ensured that the observed errors were not affecting other frequencies in the spectra, before a function was developed for their identification and removal using a variable spectral filter. For test cases, the variable filter was seen to have a significant effect on the values of spectral parameters when compared to spectra subject to a stationary filter, with differences up to 97% in P_t and 21% in m_0 for individual records.

For long-term averages, the difference caused by the filters was less than 0.1% in all parameters, which will not have a significant effect on analysis based on these data. Thus, the fixed $0.04Hz$ filter is sufficiently robust for high-level analysis, such as the estimation of the annual resource for a WEC. However, as the duration of the data set is reduced, the potential effect of these errors is increased, and a 12% difference in incident wave power, due to the filters, was observed in data captured for a single storm event of duration 4 days. Furthermore, the driving force behind measurement regimes is commonly to allow a more detailed analysis, such as a the estimation of the response of a WEC or other floating body using a frequency defined power transfer function (PTF). Such a procedure may be particularly sensitive to energy in the low frequencies, and spectral outputs from the buoys are not routinely filtered.

It is therefore recommended that for analysis based on a reduced data set, or analysis based on time-series and spectral data, records are subject to the variable filter function in lieu of the $0.04Hz$ filter (also the case for time-domain parameters). This is particularly relevant for research described in the subsequent chapters of this thesis, where comparisons between simultaneous data provide another example of where results from detailed research will be significantly affected by the erroneous low frequency energy.

For all data used in analysis going forward, unless otherwise stated, the following procedure was followed.

- All data recorded during a time corresponding to operations at the site that may interfere with normal operation of the buoys were excluded, regardless of error flags. This includes deployment, retrieval, service checks and data downloads, when the buoy will be tethered by the vessel.

- Spectra were estimated from the heave displacement time series using an automated function developed by the author using Matlab. This used a Welchs periodogram

method to perform an FFT and raw $S(f)$ values were averaged over 8 consecutive frequencies. This gives individual spectral estimates with 16 degrees of freedom, and a co-efficient of variation of 35%.

- The spectral repair function was applied to all data in order to remove low frequency energy prior to parameter estimation. Where this raised an error flag to indicate that the repair acted at frequencies above $0.06Hz$, the repair was checked visually. If no adequate repair could be acheived, the record was discarded.

- Parameters were estimated from the repaired spectra, and from time-series filtered using a variable low-frequency filter, to the definitions provided in tables 4.1 and 4.2, using Matlab functions developed by the author.

Chapter 5

Quality Control

Differences between measurements across the array are expected to be of the same order of magnitude as errors in wave measurements. In order to draw robust conclusions from the research described in this thesis, a detailed understanding of the base data sets, especially of errors or inaccuracies, is critical. Therefore, a significant research effort was applied to develop quality control (QC) procedures leading to the most accurate data possible, and a thorough understanding of factors that affect the data and data quality.

The quality control procedures developed here have been based on guidance found in technical reports on the use of wave data (Pitt, 2009), published standards on the treatment of wave measurements, (Tucker, 1993; CDIP, 2003; Ingram et al., 2011), and discussion of quality control tests in Tucker and Pitt (2001). Common practice is to predict certain properties of the measurements, and to define numerical values that represent the expected limits of these properties. Quality control tests then compare the same properties in the measurements to the limit, and where the measurements are not as predicted, a flag is raised that indicates the datum may be affected by an error. Flags are usually a number that is stored with the datum, with each number representing a particular test.

Predictions can be made based on theoretical models for the properties of wave fields, such as linear theory, which are not valid for all situations. Thus, the validity of certain tests relies on the accurate representation of the theory to the measured wave field. Accurate application of a test, or subsequent use of the error flags, requires an understanding of the limitations of the theory with which the test was designed. Other tests will be defined relative to the limitations of a particular sensor, or the processing methods applied. A detailed understanding of sensor operation is required for the definition of these tests.

For this research, QC tests described in the literature have been adapted for the particular processing used for this data set, and to suit the sensors used, which are summarised in

tables 5.1 and 5.2. A quality control function was developed, using Matlab ™, called *QC.mat* (see appendix B), to apply these tests to the data as part of the standard processing and storage regime for the PRIMaRE wave buoy array data, and the results of their application to the data are given in section 5.5. The tests form an integral part of the data processing and checks are made at each stage, summarised as follows,

- **Raw measurements** - On-board checks of raw data.
- **Time series files** - Check for flat periods, spikes and trends.
- **Wave spectra** - Statistical tests based on empirical observations of wave spectra.
- **Wave parameters** - Checks for data integrity, such as zero values. Comparison with site-specific wave-climate statistics to highlight departures from the expected range or variability of parameters.

Many of the tests are statistical in their nature and highlight records which have properties that can be considered unlikely but not impossible. Just as the validity of data will be affected by the presence of errors, the systematic removal of extremes will alter the statistical properties of the data sets. Furthermore, certain research applications may be specifically interested in the extremes or the unusual wave records, which are more likely to be flagged by QC tests. Therefore, flagged data will not be removed or altered prior to archiving. Which data are to be excluded, and potential effects of such exclusions on the results, form part of the research to which it is applied.

The author believes that this approach suits a research group with a focus on fundamental wave research. The integrity of this process depends upon detailed documentation of the quality control procedures applied, enabling an informed decision to be made based on the error flags. To complement the suite of QC functions, the results presented in this chapter provide a detailed description of the properties of the measured data, and aid researchers to decide the most appropriate strategy for the treatment of errors in their analysis. Furthemore, the analysis within this chapter demonstrates that QC tests are not directly transferable between measurement instruments, or locations, and highlights the detailed understanding required to design an effective QC regime.

5.1 Quality control tests applied to time-series files

Out of range, flag value 10

The height of individual wave crests (or troughs) in a time-series file was examined using a statistical test based on the assumption that individual wave crest heights follow a Rayleigh distribution, which is valid, assuming a narrow band Gaussian sea (Tucker and Pitt, 2001). The cumulative probability distribution that crest height (h_c), normalised by the root mean square (rms) of the surface elevation of the record, $h_c/\sqrt{m_0}$ will be below a given value, x, is defined as,

$$P_N(h_c/\sqrt{m_0} < x) = e^{-Ne^{(-1/2x^2)}} \tag{5.1.1}$$

where N is the number of zero-upcross waves in a record. The sensitivity of the test is defined by the choice of exceedance level, x, whose probability distribution is shown in figure 5.1s. N will be dependent on the wave period of waves within the record, and will vary with sea state. As an example, assuming an average wave period of $10s$, approximately 100 waves will pass the sensor during the $1024s$ in each record. Setting N to 100 and x to 5 gives,

$$h_c > 5\sqrt{m_0} \tag{5.1.2}$$

$$P_{100}(5) = 1 - 0.00037 \tag{5.1.3}$$

Thus, with a co-efficient of 5, the average frequency at which the error flag would be raised for legitimate time-series files is once in every 2684 legitimate records, or 6 to 7 times per year. This was deemed a suitable balance between sufficient sensitivity of the test and ensuring few valid records will be flagged (and potentially discarded).
Application of this test, and analysis of positive results must consider the potential causes of positive results:

- Statistical variability in h_c - As described above, a small number of valid exceedances would be expected in a linear sea (equation (5.1.1)),
- Narrow band assumption - The probability distribution in equation (5.1.1) assumes a narrow band sea with a Gaussian distribution of the water surface elevation. This assumption is less valid when the sea has broad-banded characteristics, or when waves become steep due to tidal currents or shallow water, and non-linear interactions become more important. In this instance, the exceedance may be accurately

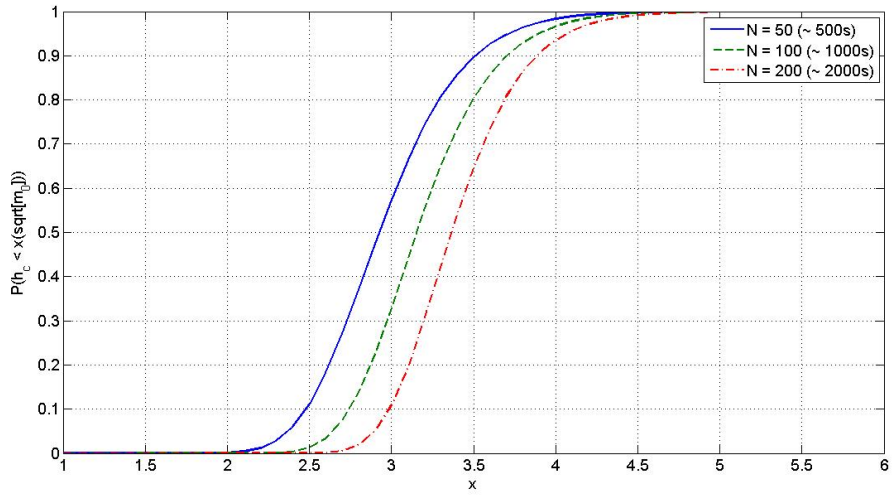

Figure 5.1: The cumulative probability function of a given crest height, h_c, being less than $x(\sqrt{m_0})$ for $N = 50$, $N = 100$ and $N = 200$, where N is the number of waves in the record.

measured by the sensor, but not fit the statistics. This must be considered when analysing the number of exceedances in a data set,

- Sensor error - Sensors may register spurious waves due to instrument failure, causing spikes in the accelerometer record. Raw accelerometer data are subject to high and low value filters to prevent this. However, Mackay et al. (2010) identified erroneous large waves in time-series files from a Datawell Waverider. These were caused in processing by the effect of the analogue filter on spikes in the accelerometer data. It is stressed that this error was identified for a different system and is not expected on the SW. However it serves as an example of errors arising from processing,

- Physical response of the buoy-mooring system - A departure from surface-following caused by the hydrodynamic response of the buoy-mooring system to physical conditions may result in an exceedance. This is avoided through the design of an appropriate mooring system, although should be taken into consideration when analysing exceedances,

- Site operations - Site operations where the buoy is interfered with (lifted or moved), will cause an erroneous record. Consultation with the log-book is routine practice when examining errors.

Wave steepness test, flag value 20

For the steepness test with flag value 20, it was necessary to use zero-crossing analysis to identify individual waves, extract their height and period and subsequently calculate the deep water wavelength, λ, and wave steepness, S, for each wave in the record. For $i = 1, 2, n...N$, where N is the number of waves identified in a record,

$$S_i = \frac{h_i}{\lambda_{0i}} \tag{5.1.4}$$

$$\lambda_{0i} = \frac{gT_i^2}{2\pi} \tag{5.1.5}$$

where h_i is the trough-to crest height of zero-upcrossing wave i with period T_i, both identified during direct time-series analysis (see section 2.1.1). The statistic, S_{max} is then the maximum value of S_i for each record, and this is tested against a threshold value to indicate spurious values in the time-series.

A discussion of wave steepness can be found in Ochi (1998), who recommends the use of a wave breaking criteria based on the limiting steepness for a Stokes wave of,

$$h \geq 0.020gT^2 \tag{5.1.6}$$

$$S \approx 1/8 \tag{5.1.7}$$

Wave-current interaction will affect the calculation of wavelength based on the wave period measured by the wave buoys, and subsequently, estimations of wave steepness will be rendered innaccurate.

Repetition test, flag value 30

Time-series data from the SW commonly exhibit a series of points with the same value, which can be identified using the repetition test in table 5.1. It is initially assumed that these are caused by the digitisation rate of the elevation data meaning that small differences in consecutive points are not resolved. A critical value when applying this test of 8 consecutive data points is suggested by Tucker (1993). However, the physical meaning of this value, in terms of the surface elevation, will be dependent on the sampling frequency and the digitisation rate of the time-series file. For data from the PRIMaRE buoy array, 8 consecutive values relates to a change of less than $+/-2cm$ during 4 seconds. The frequency of occurrence of this flag for varying numbers of consecutive repetitions is examined in section 5.5.

Acceleration, flag value 40

This test returns a positive value if the data indicates that the ocean surface was accelerating greater than $g/3$ where g is the acceleration due to gravity ($9.81ms^{-1}$),

$$\frac{\delta^2\eta}{\delta t^2} > \frac{g}{3} \tag{5.1.8}$$

It is used routinely by CDIP as a test for anomalies in time-series files from the Datawell Waverider (CDIP, 2003). However, the basis for this test is not instrument-specific and it is transferable to the SW data. Similar to test value 20, this test is based on the analysis of wave steepness and an associated wave breaking criterion derived from the acceleration of the wave crest. A review of the relevant research can be found in Ochi (1998).

Table 5.1: Time series quality control tests.

Flag value	Numerical test	Test description
10	$-5\sigma(\eta) < \eta(t) < 5\sigma(\eta)$	A test for very high crests, or low troughs in the displacement time-series. Limits are set statistically based on the standard deviation of surface elevation, assuming a Rayleigh distribution of wave heights.
20	$S > 1/8$	Testing the steepness of individual waves (calculated as $S = h/\lambda$), against a theoretical breaking limit for waves in deep water.
30	$\eta_i = \eta_{i+1:n}$	Tests for a number, of consecutive repeated values in a vertical displacement time-series. The limit, n, is set according to the vertical resolution of the heave data.
40	$\frac{\delta^2\eta}{\delta t^2}_i > g/3$	Tests that the acceleration of the waters surface does not exceed a critical limit, defined by a breaking limit of waves in deep water as approximately $g/3$ (CDIP, 2003)

5.2 Quality control tests applied to spectra and summary parameters

These tests are separated from those performed on the time-series as they are based on statistical representations of a time-series file, and error flags are applied to the whole record. The tests are sumarised in table, 5.2.

Comparing time-series and spectral analysis

This test compares the mean zero-crossing period estimated from the spectrum, $T_{m02} = \sqrt{m_0/m_2}$, to the mean zero-crossing period estimated from time-domain analysis of the wave record, T_z (Table 5.2). It is designed to highlight records where time domain analysis suffers because the limiting resolution of heave motions means that a significant number of zero crossings are missed. This affects the classification of individual waves and therefore, the definition of time-domain parameters. This test will also highlight records where errors, or a trend in the data, cause the record not to cross the mean level for extended periods of time. This factor is specifically tested as part of the quality control regime published in (CDIP, 2003). This latter effect is more prevalent in data where the entire depth of the water column is measured, such as pressure sensor or ADCP data.

Low frequency noise

Large low-frequency values have been seen to be indicative of sensor noise, particularly deterioration of the sensor stabilisation system, such as that in the Datawell Waverider (Datawell, 2005). The SW does not use a stabilisation system, although results in the previous chapter have highlighted that it still suffers from errors in measurements at frequencies below $0.5Hz$. The spectral filters developed render this test somewhat obselete. However, it was included in QC regimes in order to analyse its capability at identifying errors in the low frequencies.

The test itsef is based on the ratio of spectral energy density at frequencies below $0.035Hz$ (below which no spectral energy due to waves would be expected), to the spectral energy at the peak frequency, $S(f_p)$. Where this is in excess of 2%, an error flag is raised.

Exceedance of the Phillips spectrum

At high frequencies, the energy input from the wind is balanced by dissipation, and Phillips (1957) derived a maximum limit for spectral energy density,

$$S(f) = \frac{\alpha g^2}{(2\pi)^4} f^{-5} \tag{5.2.1}$$

This is known as the Phillips spectrum, in the form it is found in the Jonswap formulation (Hasselman et al., 1973), and α is the Phillips constant, whos value is related to the maturity of a sea.

$$\tilde{x} = gx/U_{10}^2 \tag{5.2.2}$$

$$\alpha = 0.076\tilde{x}^{-0.22} \tag{5.2.3}$$

where $\tilde{x}$ is known as the non-dimensional fetch, and is dependent on the actual fetch, x, and the wind speed, U_{10}. From equation (5.2.3), it can be seen that α is largest for high winds that have been blowing over a short fetch. Thus, as a sea becomes more mature, α reduces. Phillips defined α for a 'fully developed sea' as 0.0081 and Tucker and Pitt (2001) suggests that α rarely exceeds 0.015 even in developing seas. In the QC test, records were flagged if spectral values between the peak frequency and the cut-off frequency of $0.5Hz$ exceed the Phillips spectrum when α is set to 0.015.

The Phillips spectrum is only applicable to the saturated high-frequency part of the spectrum. Identification of the high frequencies was based on the identification of peaks in the spectrum and a threshold for application of this test was set at $1.5f_{pmax}$ where f_{pmax} is the peak frequency of the highest frequency spectral peak identified in the record (identified using the methodology described in section 4.9)

Testing for spikes in the time series

This ratio compares the time-series derived rms wave height, $H_\sigma = 4\sqrt{\overline{\eta^2}}$, [1] to $H_{1/3}$, defined in section 2.1.1). Tucker (1993) explains that in narrow band swell, this ratio approaches 1, whilst for very broad band swell it is less than 0.9. Flat spots or spikes in the data, will increase the value of $H_{1/3}$ relative to h_σ, and a critical value of 1.05 is suggested.

[1] equivalent to H_{m0}, but using the rms value of the time-series in place of $\sqrt{m_0}$.

Linear trend test

This test uses an ordinary least squares technique (Draper and Smith, 1981; Montgomery, 1982), to estimate a line of best fit for the surface elevation positions, when plotted with time. A hypothesis test then evaluates whether the gradient co-efficient of this line is significantly different from 0. A failure in this test suggests that there is no significant linear trend in the data set, to a 95% probability.

5.3 Quality control tests applied to monthly data sets

In addition to tests on the individual records, tests were applied to the time-series of parameters. These were designed to compare the parameter values of each record with those around it in the data set timeline, and to the overall statistics of the data. Tests applied are shown in table 5.3.

Range tests

These are tests applied to values of key parameters to ensure that they are within a sensible range. As this is a new facility, definition of a sensible range for parameters drew on previous measurements and modelling from the region, such as the analysis undertaken for the commissioning of the Wave Hub test site (Halcrow, 2006a). However, the use of these data must take into account possible differences between measurement locations, measurement technologies, as well as the accuracy of wave model data. More detailed prior knowledge of the wave conditions at the site would allow a more accurate definition of the range for key parameters, leading to more sensitive tests. This can be developed as more data are collected from this facility. Nevertheless, based on regional data, the tests will highlight serious errors, and will serve to alert the researcher to cases that would be considered unusual. In all cases it is recommended that the records are checked to ensure that they are not valid, but extreme measurements.

5.3.1 Parameter comparison tests

These tests compare T_z and H_{m0} values calculated from the spectrum estimated during post-processing to those calculated on-board the buoy. Some deviation should be expected because although the key parameters in spectral analysis are the same, differences may exist in the specific methods used to perform the FFT.

Table 5.2: Spectral quality control tests.

Flag value	Numerical test	Test description
1	$Tz_{TS}/Tz_S \geq 1.1$	This indicates errors in low sea states, where the resolution in vertical displacement data means that small waves do not cause zero crossings.
2	$R_s(lf) = \frac{S(0.0146)+S(0.0195)+S(0.0244)}{3S(f_{peak})}$ $R_s(lf) < 0.02$	An exceedance of the critical limit set for $R_s(lf)$ indicates excessive low frequency noise. The suitable value of the critical limit will be site specific. However, Tucker (1993) recommends 2% for time-series of surface elevation.
3	$S(f)_p = \alpha g^2 (2\pi)^{-4} f^{-5}$ $S(f) \leq 2S(f)_p$	The Phillips spectrum, $S(f)_p$ predicts the limiting value for the high-frequency parts of the spectrum. Tucker (1993) recommends that the critical exceedance be set at twice the value estimated by the Phillips spectrum, which indicates erroneous noise in the high frequencies.
4	ratio, $H_{1/3}/H_\sigma > 1.05$	Where $H_\sigma = 4 * \sqrt{rms(\eta)}$, and is an estimate of significant wave height. A value of more than 1.05 is suggested in Tucker (1993) to indicate large spikes or flat spots in the data
5	Number of data in the record that were flagged by direct time-series analysis	Identifies records where time-series analysis indicates a certain number of errors. The value required to raise a flag can be altered but provisionally was set to 1 error.
6	$P(\beta \neq 0)$	Identifies records where a significant linear trend is evident in the record of the time series. β is the gradient co-efficient calculated from a linear least squares analysis of η.

Table 5.3: Quality control tests applied to the time-series of summary parameters.

Flag value	Numerical test	Test description
100	$0.1 < H_{m0} < H_{50}$	Identifies zero records, and records where the wave height is larger than H_{50}, the 1 in $50yr$ wave height as estimated for the nearby Wave Hub site (Halcrow, 2006b).
200	$1 < T_z < 20$	Identifies zero records, and records where T_Z is large. No estimate of the return values of this parameter were made for the Wave Hub site. With data spanning a longer time, the upper limit of this test could be revised.
300	$2 < T_p < 25$	Identifies zero records. The cut-off limits are defined by the high and low-frequency cut-offs applied by the SW to the wave spectrum during processing.
400	$0 < m_{dir} < 360$	Identifies records where the mean direction parameter lies outside of the compass values.
101	$H_{pp} - 3\sigma_H < H_{m0} < H_{pp} + 3\sigma_H$	This compares the H_{m0} value calculated on-board the buoy to that calculated during post-processing as described in section 4. σ_H is the theoretical standard deviation of H_{m0} value, as calculated from the spectrum using equation (2.2.8).
201	$T_zpp - 3\sigma_T < T_z < T_{pp} + 3\sigma_T$	This compares the T_z value calculated on-board the buoy to that calculated during post-processing as described in section 4. σ_T is the theoretical standard deviation of the relevant T_z value, as calculated from the spectrum using equation (2.2.6).

5.4 Visual examination

An important aspect of the quality control procedure is the visual check. Looking at a suitable plot of any aspect of wave data is an invaluable tool for identifying errors, spikes, flat spots and other unusual values are quickly recognized. The author has developed, within the QC function, a multi-level plot function to examine errors (5), which includes the following plots,

- plots of significant wave height, H_{m0} and wave period, T_z on which the error flags are superimposed,
- an H_{m0} and T_z scatter diagram on which the fraction of records within each cell that have a particular error flag are superimposed as a colour overlay,
- plots of the time-series and one-dimensional spectrum for all buoys that were monitoring simultaneously, for each record which has an error flag.

Using this function, assuming the number of records flagged is not too large, each record can be visually checked prior to its exclusion (or inclusion) from analysis. In the first instance, these plots were used to support a detailed examination of the results of applying the tests described above to the measured data.

5.5 The results of applying quality control tests

The QC tests described in the previous sections were applied to the sample data set described in section 4.5.1. 77 records within the data set are highlighted in the log-book as being captured during deployment and site operations. These records contain examples of what would be serious errors in normal operation, collected when the buoy was on-board a vessel, or when the buoy was being lifted for deployment. Although they would usually be removed during processing, these records have been left in the data set to help analyse the effectiveness of the QC tests at identifying serious errors. The position of the buoy within the array was changed when re-deployed during December 2009. Therefore, results relate to two separate deployment locations, and these periods are separated in analysis. Wave conditions encountered during this period are described in more detail in section 6.

5.5.1 Time-domain tests

The time-domain tests detailed in table 5.1 were applied to the time-series of surface elevations, comprising 2048 data sampled at $2Hz$. Error flags are raised for each individual

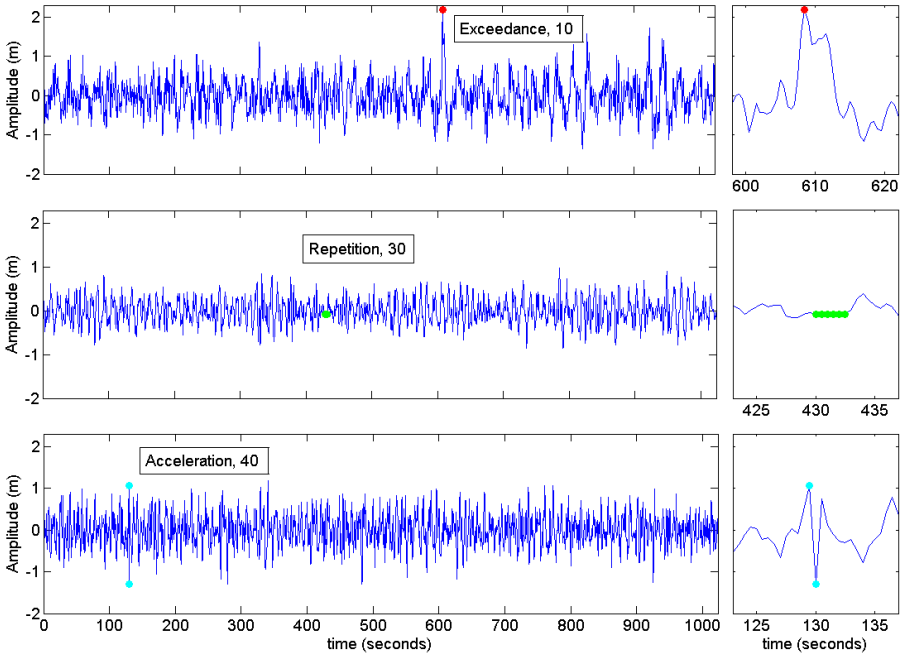

Figure 5.2: Examples of time-series where error flags were raised by time-domain tests.

datum, with a zero indicating that no error was identified. Examples of positive results are plotted in figure 5.2. This section details the results of the application of these tests and where appropriate, shows how test parameters were chosen.

Exceedance test, value 10

This test was applied to the sample data set, and 64 time-series files returned a positive test, whilst equation (5.1.3) predicts approximately 3 valid records would return a positive test during that period. 4 of the 64 errors were from records during site operations. Figure 5.3 suggests that exceedances are not constant through the year, being more common during the first deployment (October - Nov 2009).

The $H - T$ scatter diagram in figure 5.4 shows the combined frequency of occurrence of records in each cell, and the overlay colours show the proportion of records within that cell that have flags for an exceedance. Exceedances rarely occur in swell-dominated seas, and there is a trend for more exceedances in larger seas with relatively low periods, conditions which were more prevalent during the first deployment. It is possible that records which are an exception to this trend may be influenced by bi-modal conditions where a dominant swell occurs simultaneously with a low-period sea.

Low-period with high wave height is indicative of a developing sea. In these conditions waves are steeper and the bandwidth can be expected to be large, which will reduce the validity of the narrow-band assumption. However, when the ratio $h_c/\sqrt{m_0}$ was plotted against the bandwidth parameter Q_p in a scatter plot (fig. 5.5), no clear relationship occurs in the data between the exceedances and bandwidth of the sea state. Nevertheless, that the distribution of exceedances is dependent on the sea state, suggests that either the wave conditions cause more exceedances than expected (which are measured accurately), or that the buoy measurements become inaccurate in developing sea-states. The relationship between tidal conditions and the occurrence of an exceedance was also investigated using a scatter plot with both the tidal flow velocity, and the water depth. No relationship could be identified.

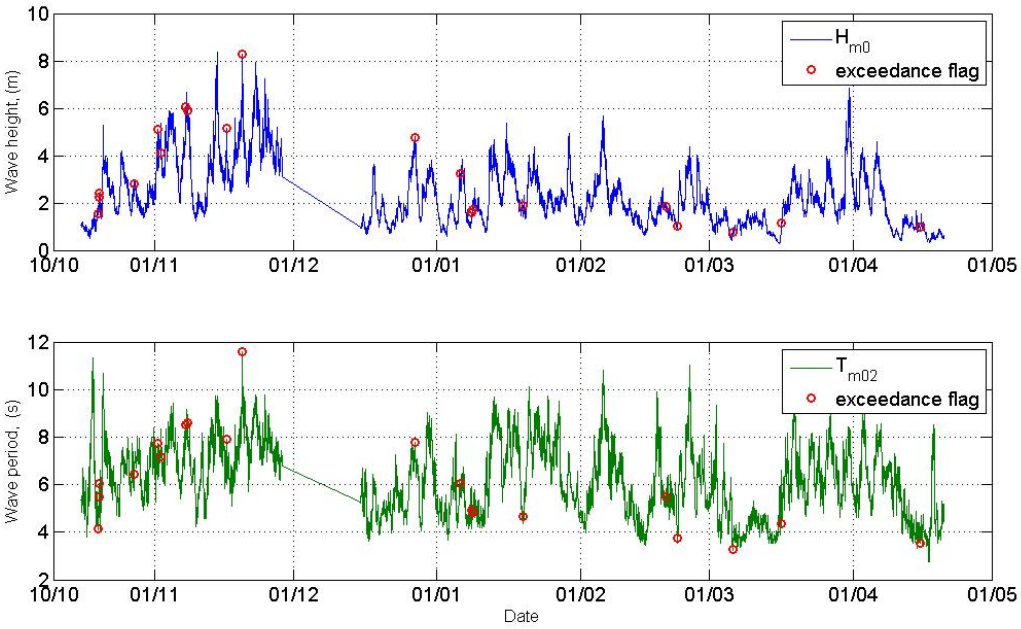

Figure 5.3: The time series of significant wave height, $H_{m0}(m)$ and period, $T_{m02}(s)$ calculated during on-shore post-processing of the time-series for buoy B from October 2009 until April 2010, overlaid with the occurrence of large waves, flagged where a value of $h_c > 5\sigma(\xi)$.

Steepness test, value 20

A tidal signal can be observed in a time-series of the S_{max} parameter (fig. 5.6). The mean tidal flow speed with depth is also plotted on this graph, as derived from Polpred[2] for

[2]Polpred is computational software for PCs that uses Proudman Oceanographic Library tidal models to predict the tidal levels and currents around the UK (POL, 2010).

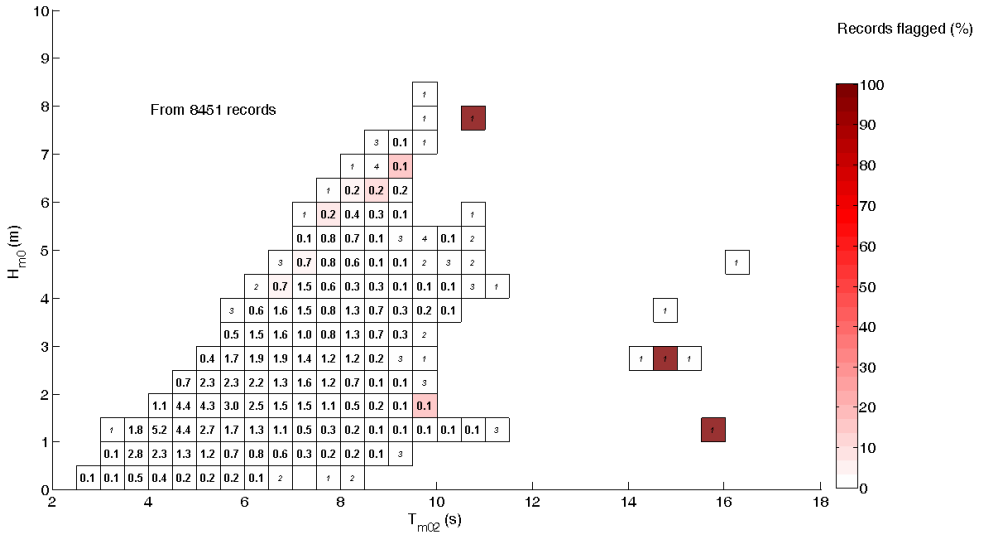

Figure 5.4: A $H_{m0}, Tm02$ scatter diagram for buoy B, where text shows the percentage of measured records within each cell, or the total number of records (values in italics), occurring in each cell. The colour overlay demonstrates the number of occurrences of large waves, flagged where a value of $h_c > 5\sigma(\eta)$.

a position $350m$ from the measurement point. This suggests that current modulation is affecting the steepness of the waves, and the validity of this test. The critical value was set at 1/8 and 16% of records in this data set returned a positive flag.

Repetition test, value 30

The frequency of occurrence of repetitions of varying lengths was tested on the sample data set. An example of a positive result is shown in figure 5.2, and the frequency of positive results for varying critical lengths are given in table 5.4. When the critical level is set to 8 consecutive repeated surface elevation data, 22 records are flagged, with 12 of these being identified in the log-book as periods of site-operations. All of the records flagged during normal operation were measured during low amplitude swell-dominated conditions, demonstrating a dependence of this test on sea-state (fig. 5.7). This effect is accentuated when the critical level is reduced.

Acceleration test, value 40

This flag was raised 102 times, approximately 1% of records, none of which relate to periods of site operations. Examination of the records where an acceleration is flagged shows that in all cases, a negative spike occurs as a wave crest passes the sensor (fig. 5.2).

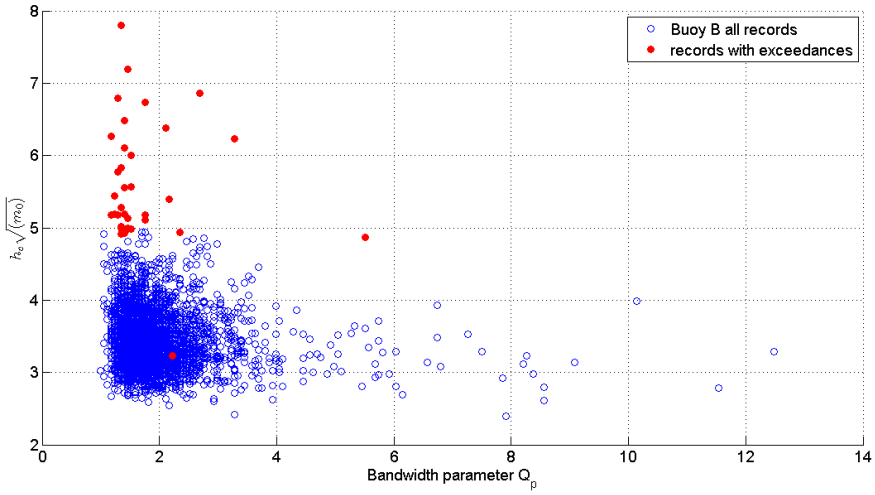

Figure 5.5: A scatter diagram showing the ratio $h_c/\sqrt{m0}$ vs Q_p of the time-series for each record. Highlighted in red are those which have been flagged for the test $h_c > 5\sqrt{m_0}$. The reason that some are flagged when $h_c/\sqrt{m0} < 5$is because the test is also applied to troughs, $h_t > 5\sqrt{m_0}$.

Table 5.4: The proportion of records flagged with different critical values for the length of consecutive data that constitute a repetition (N).

N	Flagged records (%)
3	96
4	33
5	6
6	1.60
7	0.50
8	0.27

A bi-modal distribution of errors with wave height can be identified on the scatter diagram (fig. 5.8), although there is no obvious distribution with period. Figure 5.9 shows the proportion of records flagged for this test that have an H_{m0} values within bins, ΔH_i of width $0.2m$ where i $= (1, 2, 3, i...n)$. Also shown is the distribution of all wave-heights for the same bins, which does not have a bi-modal shape. There are two peaks to the distribution of errored wave heights, centred at $1.5m$ and $3m$. No satisfactory explanation has been found for this distribution. Examination of the time-series does not reveal a temporal pattern in the positive results, and no relationship with tidal conditions can be identified.

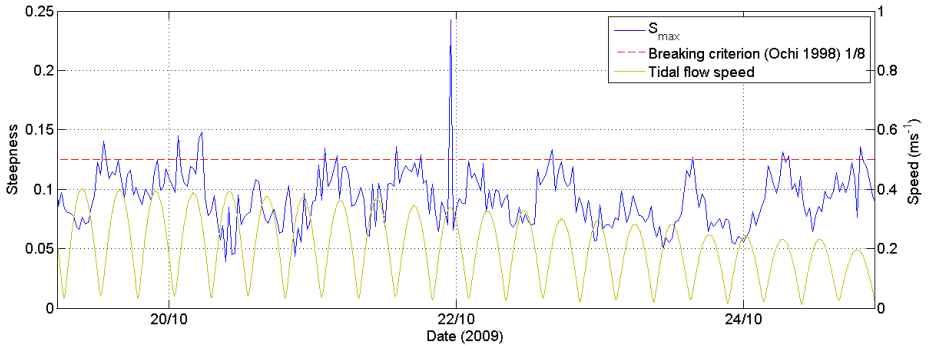

Figure 5.6: A time-series of the parameter S_{max}, the maximum steepness for an individual wave measured within a record (equation (5.1.4)), calculated for data from buoy B during October 2009.

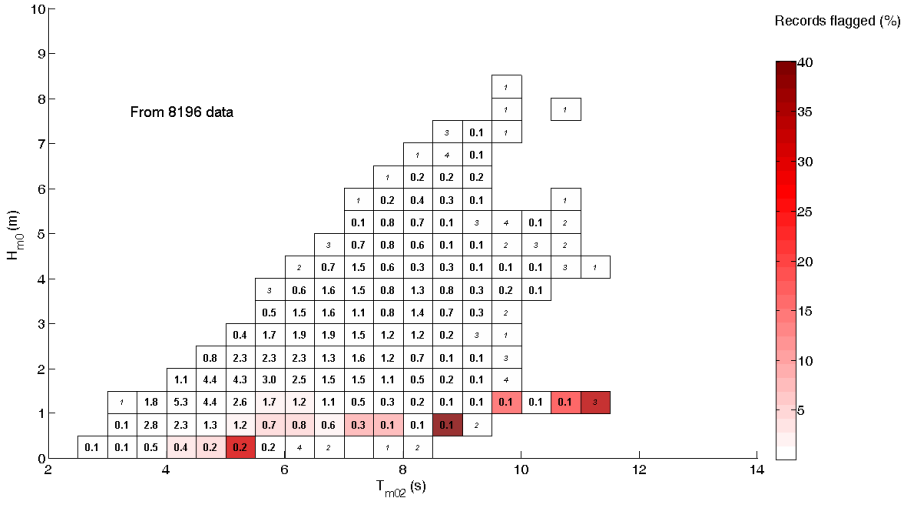

Figure 5.7: A $H_{m0}, Tm02$ scatter diagram for buoy B, where standard text shows the percentage of records, or the total number of records (values in italics), occurring in each cell. The colour overlay shows the percentage of records flagged in each cell because of repetition of more than 6 consecutive surface elevation data. For clarity, records coinciding with site operations have been removed before plotting this scatter diagram.

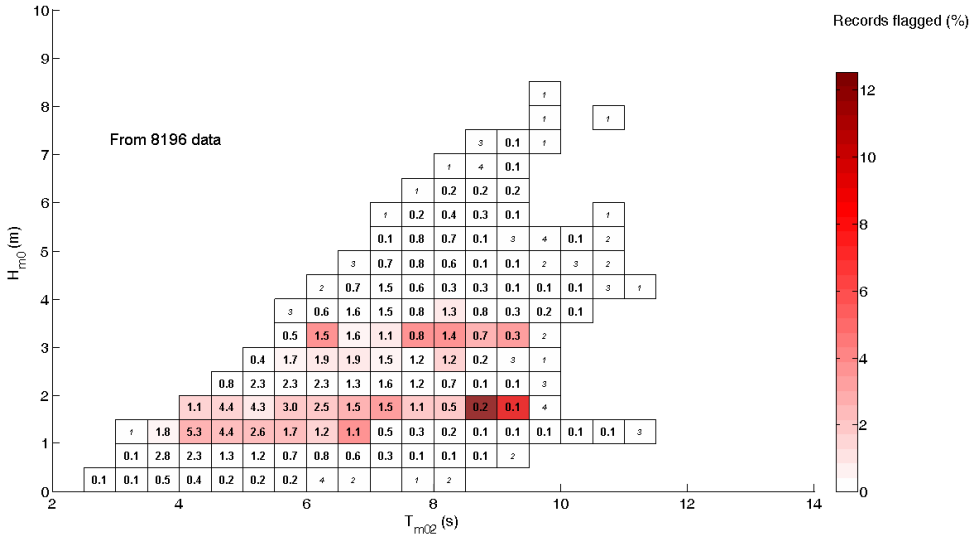

Figure 5.8: A $H_{m0}, Tm02$ scatter diagram for buoy B, where text shows the percentage of records, or the total number of records (values in italics), occurring in each cell, and colour overlay shows the proportion of records flagged in each cell because the waters surface was seen to accelerate faster than $g/3$. For clarity, records coinciding with site operations have been removed before plotting this scatter diagram.

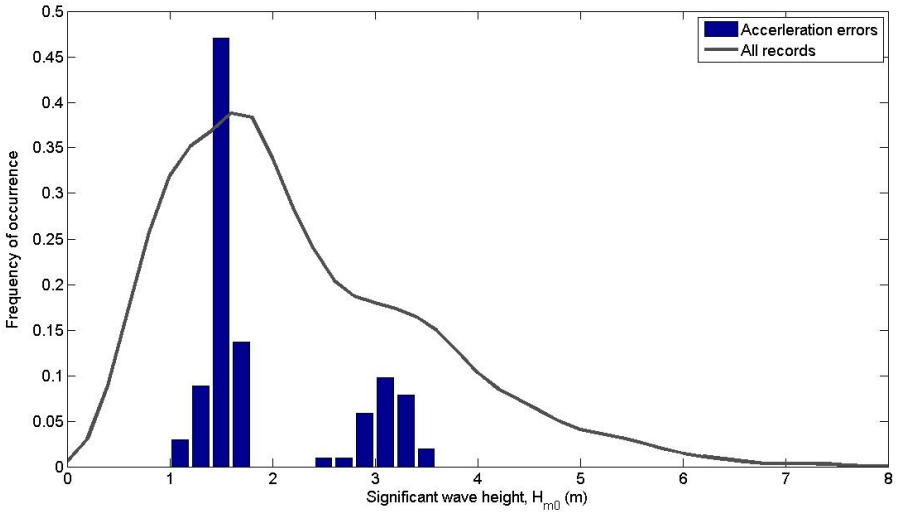

Figure 5.9: Wave height frequency distribution of records with error flag 40. The bars show proportion of records with error flag 40 that have an H_{m0} values within bins, ΔH_i of width $0.2m$ where i = $(1, 2, 3, i...n)$. Also shown is the distribution of all wave-heights for the same bins, plotted as a grey line.

5.5.2 Wave spectrum and parameter tests

The tests detailed in table 5.2 were applied to the omnidirectional wave spectrum and to summary parameters for each record. Error flags are raised for an entire record, where appropriate.

Comparing time domain and spectral wave periods, flag value 1

Flags for this test were raised on 113 of records, of which 15 were from records captured during periods of site operations (fig. 5.10). In the remaining cases, this error flag was unique, and was raised in low H_{m0}, swell-dominated sea states. In these conditions, the sea surface height will be varying relatively slowly and resolution will affect zero-crossings, causing errors in time domain derived parameters.

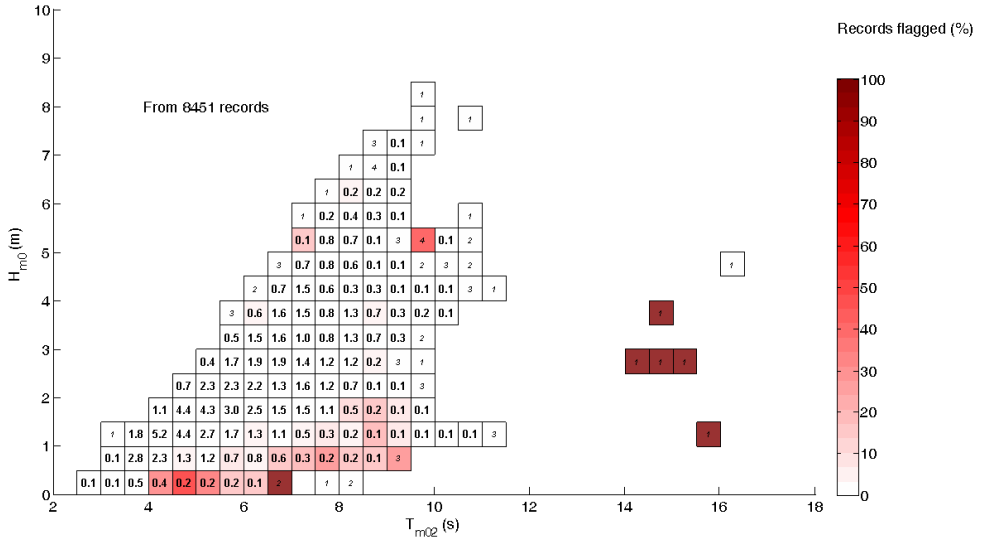

Figure 5.10: A $H_{m0}, Tm02$ scatter diagram for buoy B, where text shows the percentage of records, or the total number of records (values in italics), occurring in each cell, and colour overlay shows the proportion of records flagged in each cell because the ratio T_Z(Time-domain)/T_{m02}(spectral moments) > 1.1.

The effect that the resolution of the surface elevations will have on the accuracy of spectral processing was examined using a spline interpolation function to estimate the true variation in surface elevation during the time-series, as described in chapter 4. In order to establish whether the difference between parameters from the raw data and those from the interpolated signal are significant, they were compared to theoretical sampling variability. Sampling variability for each parameter was assumed to be Gaussian distributed with a standard deviation, σ, calculated using equation (2.2.7) for the spectral moments.

Where the difference between parameters exceeds 1.98σ, they are considered significantly different to a critical level of 95%. No differences between spectral moments were deemed significant. Therefore, records flagged by this test should be discarded only when time-series derived parameters are being used. A more general consideration would be that at the given sampling frequency ($2Hz$) and resolution of surface elevation ($4cm$), records derived from sea states with $H_{m0} < 1.5m$, and $T_z > 6s$ should all be viewed with caution.

Low frequency noise test, flag value 2

The critical level was set to 0.02, and exceedances occur 3 times in over 8000 records.

High frequencies exceed the Phillips spectrum, flag value 3

When applied to data from buoy B, 4% of records are flagged. Figure (5.11) demonstrates that the flags occur more commonly during small, wind-dominated seas. These are immature seas which would be expected to exhibit high values of α, suggesting that the flags are not due to sensor error, rather due to the sea state at time of measurement. Visual analysis of spectra with positive flags for this test does not indicate serious errors in the spectra. These positive flags could be avoided by setting the α value higher than recommended in the literature, to allow for temporarily high values during developing seas. Setting alpha to 0.3 reduces the number of occurrences to 1%, although a clear dependence on the sea-state remains.

Tucker (1993) suggests this test be used for identification of spikes in the time-series. However, it was found that when this test was applied, positive flags do not correlate well with exceedances flagged from direct time-series tests (flag value 10). Furthermore, it appears to be flagging immature seas which do not necessarily contain an erroneous reading. Therefore, it is not recommended that the flags from this test be used to exclude data from a data set.

Wave height test for spikes in the time series, flag value 4

Values of the test statistic (defined in table 5.2) in excess of a suggested critical threshold of 1.05 are potentially errors, caused by a large spike in the data, which will increase H_s relative to H_{sigma}.

When the test was applied to the test data, approximately 1% of records were flagged. Temporally, the flagged records were clustered, with periods of very frequent flags within

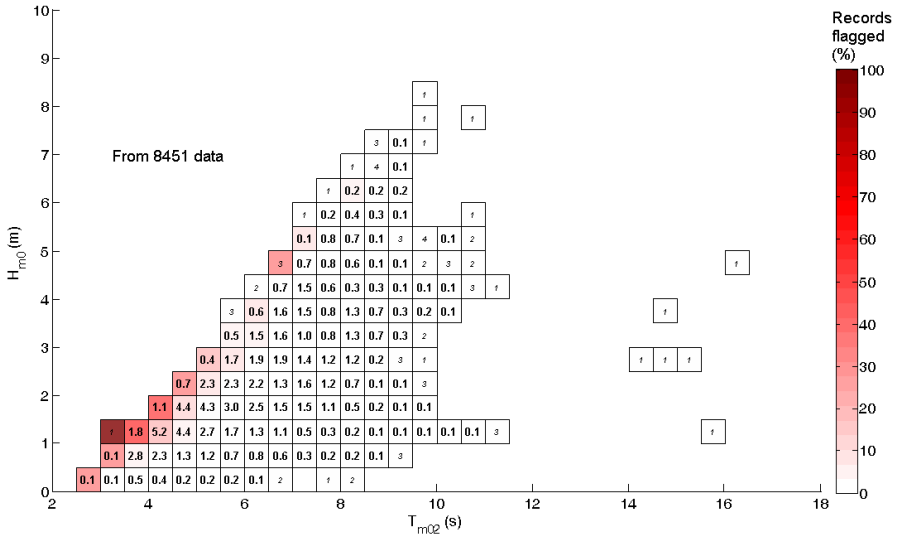

Figure 5.11: A $H_{m0}, Tm02$ scatter diagram for buoy B, where text shows the percentage of records, or the total number of records (values in italics), occurring in each cell, and colour overlay shows the number of records flagged in each cell for high frequencies that exceed the Phillips spectrum.

the data (fig. 5.12). The flagged records coincide with periods of low wave height, swell-dominated conditions (fig. 5.13). Flagged records were subject to visual examination of the time-series of surface motions, which often showed the presence of groups of waves with relatively large height and period within the records that had been flagged, with no indication of erroneous spikes or other reasons to discard the record. This suggests that groupiness associated with swell-dominated conditions are causing positive results in this test.

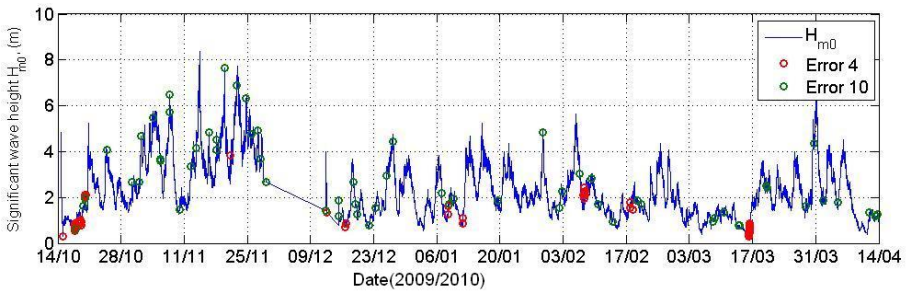

Figure 5.12: The time series of significant wave height, $H_{m0}(m)$ for buoy B from October 2009 until April 2010. Red circles indicate the records for each buoy that have been flagged for a ratio h_{RMS}/H_{σ} that exceeds 1.05. Green circles indicate records that have been flagged for an exceedance in the time-series.

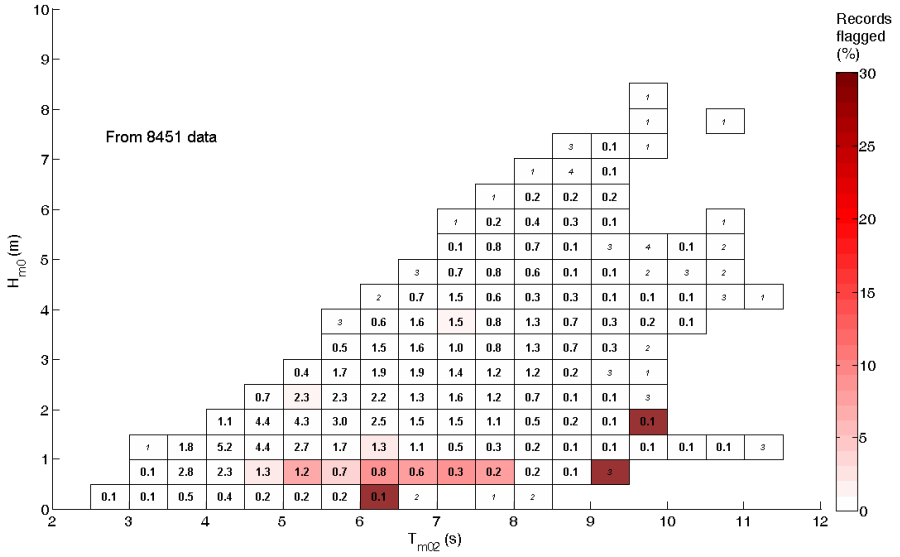

Figure 5.13: A H_{m0}, T_{m02} scatter diagram for buoy B, where text shows the percentage of records, or the total number of records (values in italics), occurring in each cell, and colour overlay shows the number of records flagged in each cell for $H_{1/3}/H_{\sigma} > 0$. For clarity, records coinciding with site operations have been removed before plotting this scatter diagram.

Trend test

The test for a linear trend does not flag any records.

5.5.3 Tests applied to a time-series of parameters

This represents the third application of quality controls to the data and it is applied once all the parameters have been collated.

Range tests

When these tests were applied to the data, they were effective at highlighting records that contain serious errors, experienced during deployment. However, care should be taken before excluding all records flagged by the range tests, as the return values do not represent absolute maxima and they can be expected to be exceeded. Therefore, it is recommended that the summary parameters which fail these tests should be examined on a time-series of that parameter. This will identify whether the flagged record occurred during extreme conditions, during which an exceedance of return values may be expected. Failure to perform this analysis may result in exclusion of perhaps the most interesting records.

5.6 Discussion

The principal of each QC test have been described in section 5, and the results of these tests have been described above. These results allow discussion of the performance of individual tests, leading to advice on the use of the error flags in future research.

Time-series tests

The number of records highlighted for an exceedance (test value 10) was significantly larger than predicted for an error-free data set consisting of narrow-band Gaussian seas (as described in section 2.1). Although exceedances were observed to occur more frequently during developing seas, no relationship with bandwidth could be identified. Further work to validate these records will be required prior to the analysis of large waves, extreme water levels, or use of the parameter H_{max}.

The steepness test (flag value 20) and the vertical acceleration test (value 40) are designed to identify steep waves. Both are based on a breaking limit although these are derived differently. The first, uses the measured steepness of each individual wave, S, which has been observed in chapter 4 to be sensitive to high frequencies, which were not filtered prior to analysis. As a result, this test is not recommended for data selection at this site. The acceleration test measures only the vertical acceleration of the sea, and has a threshold based on a breaking limit derived from vertical motion only. In operation, it is seen to highlight records exhibiting an inverse spike, which is an error, although the cause is unknown.

A high proportion of records captured during periods of site operations were flagged for repetitions, which indicates the suitability of this test for highlighting serious errors. However, all records flagged during routine operations related to small swell-dominated conditions. These are caused by the resolution of the surface elevation data, which is dependent upon the sensor configuration. For each sensor, the number of consecutive values that constitutes an error should be decided with reference to the vertical resolution and aim to avoid frequent flagging of valid conditions, whilst retaining sensitivity of the test to serious errors.

Spectral and parameter tests

The test using the ratio of wave period estimated using time-domain analysis to that estimated from the spectral moments (test value 1), is also affected by the resolution of

the time-series. Accordingly, it returns more frequent positive flags during low wave height sea states, and 53 of 113 records flagged with value 1 were also flagged by the repetition test (value 30). Again, the strong dependence of positive results on sea-state means that removing these records may introduce a bias into the resultant data set. Nevertheless, a positive result in this test indicates that zero-crossing analysis will be affected by the resolution at which the time-series is digitised and records flagged by this test should be excluded from analysis when time-series derived parameters are being used, although parameters estimated from spectral moments are not strongly affected. Importantly, this test is not as dependent on the sensor configuration as the test for repetition in the time-series, and as such is more robust for application to different data sets.

The test for low frequency noise, as defined by Tucker (1993) flagged only 3 records despite the high frequency of errors highlighted in section 4.8. This test was defined to highlight low-frequency noise in data from Datawell waverider buoys, and is not effective for the SeaWatch mini II. The ratio could be re-defined to relate frequencies greater than $0.35Hz$ to the peak frequency. However, at these frequencies, some spectral energy may occur naturally and care must be taken not to flag valid records.

The test for high-frequency noise returns positive values for valid sea states. This is related to the value of α, which varies with the wind and wave conditions over the site, as well as the tidal flow. The values recommended in the literature appear insufficient for measurements at this site and it is possible to reduce the frequency of flags by increasing α. However, exceeding the recommended limits for this parameter in the literature should only be undertaken alongside a detailed examination of high frequency measurements to ensure flagged data are valid measurements.

Test value 4, $H_{1/3}/H_{\sigma} > 1.05$ is designed to highlight records containing erroneous spikes in the time series records. However, results from this data set suggest that this test is returning positive flags during swell-dominated conditions. Visual analysis does not reveal a spike or other obvious error, rather a large wave group. The conclusion is that this test may be flagging valid data and it should not be used for excluding data from analysis.

It is noted that Tucker (1993) recommends spectral tests as more sensitive than direct time-series analysis for identifying erroneous spikes in the time-series and not identifying valid data. Specifically, tests 3, exceedances of the Phillips spectrum, and 4, $H_{1/3}/H_{\sigma} > 1.05$. However, here it has been recommended that the flags of value 3 and 4 be ignored as there is a danger that these will exclude valid data. Therefore, it is necessary to rely on the acceleration test and the exceedance test for identifying spikes, despite the drawbacks

noted in (Tucker, 1993). The spectral filters discussed in section 4.8 will affect certain tests, and may render certain errors obsolete. For this sensor, spectral filters should be applied prior to QC tests, which may improve the results from tests 3 and 4. However, test 2 uses low frequency noise to highlight a problem with the Datawell sensor. Were the filters applied prior to QC, this test would not be effective, which further supports the conclusion that QC tests and their application must be designed with careful consideration of the sensor.

5.7 Standardised Quality Control

Table 5.5 demonstrates the level at which each error flag is raised for data captured during 6 months from 15^{th}-Oct-2009 until 30^{th}-Apr-2010 from all buoys. Also shown is the recommended action to be taken for each flag. *No action* means that it is not recommended to act upon the error flags associated with this test, as it is not certain that the flagged records are erroneous. *Caution* means that caution should be exercised when using these data. The discussion above should be referred to before acting upon these error flags. *Excluded* means that it is recommended that these data are excluded from research as they are almost certainly errors. As described above, no data are removed from the archive, this decision is left to the individual researcher accessing the data.
The observed difference occurring in the frequency of flags in the heave resolution test is attributed to rounding to two decimal places. Absolute values of these frequencies do not differ significantly. Subsequent to the removal of records identified in the log-book as occurring during site operations, error flags $100 - 400$ are not raised in any records.
Thus, the procedure followed for all further analysis in this research can be summarised as:

- All data recorded during a time corresponding to operations at the site that may interfere with normal operation of the buoys were excluded, regardless of error flags. This includes deployment, retrieval, service checks and data downloads, when the buoy will be tethered by the vessel,

- Records with error flags $40, 100, 200, 300$ and 400 (see section5) were excluded from the analysis data set. Direct time-series, or parameters derived from direct time-series of records with error flag 1 were not used, whilst it was assumed that spectral parameters are not affected. The time-series of records with an error flag 30 were visually checked,

Table 5.5: The proportion of records flagged for each quality control test.

			Frequency			
Error	Action	Flag	A	B	C	D
Out of range	Caution	10	0.01	0.01	0.01	0.01
Steepness	No action	20	0.17	0.21	0.16	0.18
Repetitions	Caution	30	0.03	0.03	0.03	0.03
Accelerations	Exclude	40	0.01	0.01	0.01	0.01
Heave res	Caution	1	0.02	0.01	0.02	0.02
Low freq noise	No action	2	0	0	0	0
High freq noise	No action	3	0.03	0.04	0.03	0.03
$H_{1/3}/H_{\sigma}$ ratio	No action	4	0	0.01	0	0
Trend test	No action	6	0	0	0	0
Out of H_{m0} range	Exclude	100	0	0	0	0
Out of T_z range	Exclude	200	0	0	0	0
Out of T_p range	Exclude	300	0	0	0	0
Out of Dir range	Exclude	400	0	0	0	0
$H_{buoy} \neq H_{pp}$	Caution	101	0.01	0.01	0.01	0.01
$T_{buoy} \neq T_{pp}$	Caution	201	0.00	0.01	0.02	0.01

- Records with an error flag 10 were not removed systematically, but were submitted to visual examination and discarded if the large value was a single point or spike in the time-series. As discussed above, further investigation of records flagged by this test should be carried out prior to examination of extreme wave heights or H_{max}.

Chapter 6

Measured wave conditions

The previous chapters have established a strategy for data processing and quality control that ensure a high quality data set and these procedures were applied to all available data. This chapter uses data from the northerly position within the array, which comprises data from both buoy B (Oct-Nov 2009) and buoy D (Dec 2009 - Oct 2010). This is used to demonstrate some fundamental properties of the measured data and describe the wave conditions during the deployment. Subsequently, data captured by all 4 buoys is presented including a preliminary analysis of the differences between the data sets. For this section, data captured from the two deployments are treated separately to account for the changing positions within the array.

All records that coincide with deployment, retrieval or other site-operations have been removed through consultation with the log-book. The data were subject to quality control described in section 5.7 and all parameters were estimated from spectra that had low-frequency energy removed by the variable filter function described in section 4.8. Notably, a break from recording occurred after mooring failure during the winter period, which is usually the most active period during the year. Therefore, although data are captured during 12 months, annual statistics such as the average power may not be representative of the annual wave climate.

For parameters estimated using direct time-series analysis, a variable low frequency filter was used with a high-frequency cut-off of $0.5Hz$. These were applied to time-series using the spectral filter technique applied by Oceanor processing (see section 4), which involves a FFT of the time-series, application of the filter and the subsequent IFFT to re-generate the time-series.

6.1 Time-series data files

Time-series data comprise a series of the surface displacement in vertical, East and North directions. These are captured at $2Hz$, with a resolution of $0.0391m$.

As described in section 2.1, a stochastic approach to the analysis of wave records assumes that the surface profile is a random variable that exhibits a Gaussian probability distribution. Where water is shallow (relative to the wavelength), or in the presence of strong currents, non-linearities become increasingly important and the surface elevation is seen to deviate from a Gaussian distribution. The buoy array site is situated in a depth which will be considered deep or intermediate depending on the wavelength of the incident waves. Furthermore, tidal flow in the area have the potential to affect the linearity of the wave field (sec. 3.1).

The collection, quality control and archiving of time-series data described in the previous chapters has generated a high quality data set. Figure 6.1 displays a probability density function from a wave record which demonstrates a distribution that is broadly Gaussian, although a detailed examination of the time-series specifically was beyond the scope of this research.

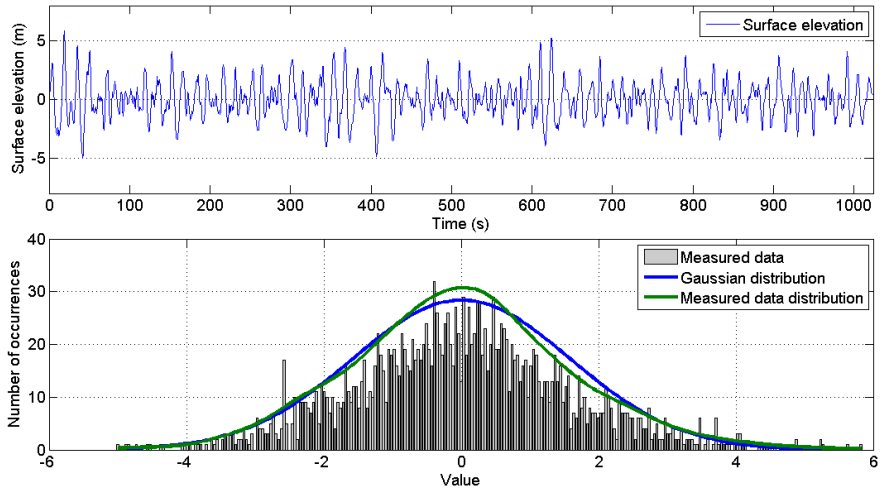

Figure 6.1: The frequency distribution of a measured time-series showing a) a time-series of surface elevation around the mean water surface and, b) the associated frequency density function. Also shown is a Gaussian distribution with mean and standard deviation the same as the measured data.

6.2 Properties of measured spectra

The low frequencies and high frequencies of the measured spectra have already been discussed in the context of quality control, spectral filters and spectral repair. This section examines the shape of the one dimensional spectra for all records in order to characterise the measured sea states. In particular, each spectrum was analysed to determine the number of distinct peaks using the Matlab function, *findpeaks*, which uses a rainflow filter to identify global maxima and minima in the spectrum. Following the methodology of Guedes-Soares (1984), a peak was only considered where, the peak $S(f)$ value was at least 15% of the maximum of the largest peak, the 90% lower confidence limit of each peak was higher than the 90% upper confidence limit of the corresponding troughs, and each peak was separated by at least $0.03Hz$.

This formulation was used to identify component systems within each spectrum, identified by their peak frequency, f_p and corresponding maximum spectral energy density, $S(f_p)$. 36% of measured spectra exhibit 2 or more simultaneous wave systems, and an example of a bi-modal spectrum is given in figure 6.2.

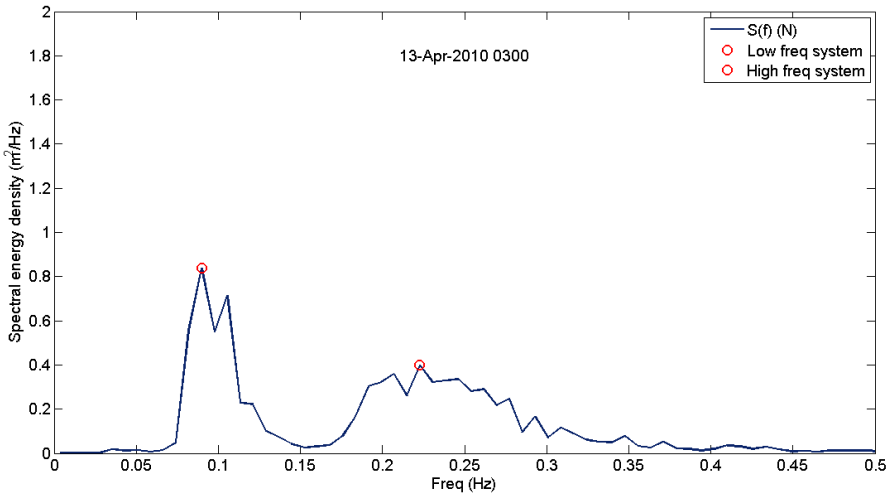

Figure 6.2: Wave spectrum measured during 13th April 2010 at the northerly position within the array. Two peaks in this spectrum were found to be significant using the methodology of Guedes-Soares (1984), which are marked with red circles.

Guedes-Soares (1984) examined the prevalence of twin-peaked spectra at stations across the North Atlantic, and observed that 22% of records were multi-modal, and that this is more common at smaller wave heights. Table 6.1 shows that the total proportion of records that are classified as multi-modal is greater in the PRIMaRE buoy array. However,

where the statistics were broken down into wave height categories, the frequency of multi-modal spectra are consistently less in the buoy data, and it is the difference in the relative frequency of wave states occurring in each category that causes the difference in the total values (tab. 6.1). Guedes-Soares (1984) also found that the prevalence of multi-modal records is reduced close to coastlines due to fetch-limitation. The location of the buoy array means that fetch is severely limited to the south and east, which may explain the less frequent measurement of multi-modal spectra observed in the buoy data when compared to results from Guedes-Soares (1984) (tab. 6.1).

Table 6.1: The number of records for each buoy that have multi-modal spectra, in categories defined by H_{m0} values for both deployments.

	PRIMaRE buoys		Guedes Soares (1984)	
Category	N		N	
$0 < H_{m0} \leq 1m$	8342	0.59	16	0.75
$1 < H_{m0} \leq 2m$	23130	0.37	91	0.37
$2 < H_{m0} \leq 3m$	9226	0.21	143	0.32
$3 < H_{m0} \leq 4m$	4901	0.11	158	0.22
$4 < H_{m0} \leq 5m$	2451	0.07	156	0.23
$5 < H_{m0} \leq 6m$	1084	0.08	121	0.24
$6 < H_{m0} \leq 7m$	4129	0.11	101	0.17
$7 < H_{m0} \leq 8m$	1300	0.08	72	0.11
$8 < H_{m0} \leq 9m$	534	0.07	55	0.09
$9 < H_{m0} \leq 10m$	0	-	34	0.09
$10 < H_{m0} \leq 11m$	0	-	30	0.03
$11 < H_{m0} \leq 12m$	0	-	20	-
$12 < H_{m0} \leq 13m$	0	-	6	-
$13 < H_{m0} \leq 14m$	0	-	5	0.2
Total	55097	0.31	1008	0.22

The process of partitioning using the one-dimensional spectrum will not separate wave systems with similar frequencies and different directions. Methods for this type of analysis using a directional spectrum have been developed and their application to the directional measurements taken at the array provide a more detailed examination of the extent of multi-modality (Saulnier et al., 2011a,b).

6.3 Wave height and period

An extended gap in measurements during November and December 2009 occurred due to mooring failure as described in section 3.2. Notably, the period prior to failure has markably higher wave heights than any other period in the data set, with 8 identifiable storms in less than 2 months exceeding a wave height of $5m$, compared to 4 during the 4 months after re-instatement and the maximum significant wave height measured was $8.36m$ during November 2009 (fig. 6.3). Halcrow (2006b) used a 15yr hindcast of the UK Met-office global wave model to estimate extreme conditions at the nearby Wave Hub site. The 1 in $1yr$ wave height estimated was $10.4m$, which is $2m$ larger than the largest measured Hs in this data set. This may indicate an over-estimation from the wave model, or reflect inter-annual variability. In addition, comparisons will be influenced by a difference in wave climate between the buoy site and the Wave Hub site, and the large gap in this data set, which covers approximately one month during the most active winter season. A comparison of wave period parameters with Halcrow (2006b) is not possible as no definition was given for the period parameters used in estimation of extreme events in the report.

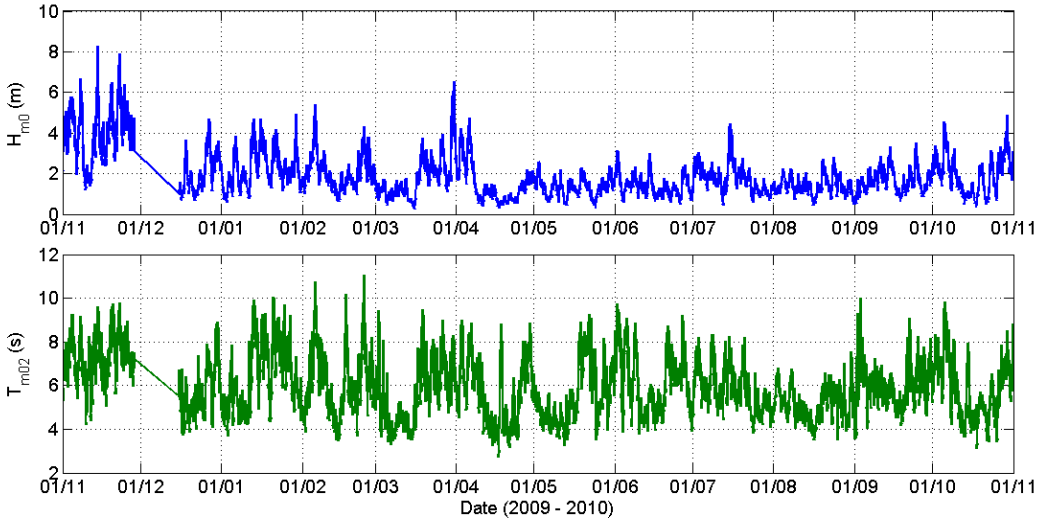

Figure 6.3: Time-series of a) significant wave height, H_{m0} and b) mean zero-crossing period, T_{m02} recorded at the northerly position in the array, estimated during post-processing using a variable low-frequency filter.

The most frequent wave conditions are categorised as $4s < T_{m02} < 5.5s$, and $1m < H_{m0} < 2m$, and occur close to or below the line of constant steepness, 1/19.7, which

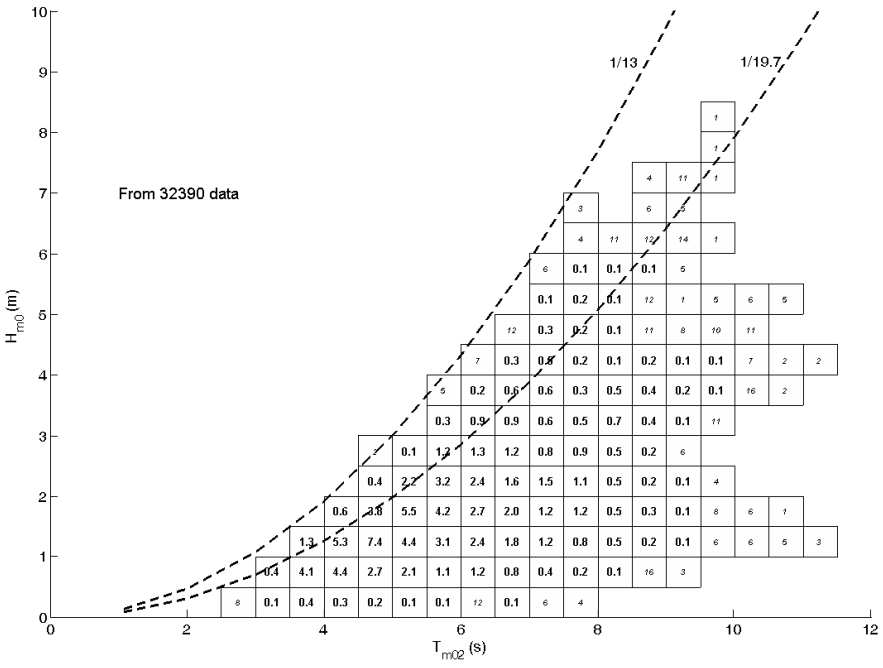

Figure 6.4: A H_{m0}, T_m scatter diagram for data from October 2009 - April 2010 from buoy B. Text displays the percentage of all records, or the total number of records (values in italics), with H_{m0} and T_z values corresponding to that cell. The lines represents constant significant steepness of 1/13, and 1/19.7, the latter of which corresponds to a Pierson Moskowitz spectrum.

represents a fully-developed sea (fig. 6.4). The data are bound by the line of constant significant steepness, 1/13, which is typical of an exposed site (Tucker and Pitt, 2001), and is consistent with other data sets captured in the Eastern Atlantic (Holmes and Barrett, 2007).

In an infinitely narrow band spectrum, wave heights will follow a Rayleigh distribution and H_{m0} will be equal to $H_{1/3}$. In naturally occurring sea states, this is not the case and a bias can be observed in figure 6.5, with H_{m0} consistently larger. Tucker and Pitt (2001) state that for most sea states, $0.9H_{m0} < H_{1/3} < H_{m0}$ which is in agreement with values in the recorded data. Notably, these differences were smaller than those observed between H_{m0} values derived from different methods of processing (sec 4.5).

The long-term distribution of the significant wave heights forms the basis of the prediction of the extreme conditions expected at a site, commonly quantified by the return values of H_{m0} (Tucker and Pitt, 2001; Carter and Challenor, 1981). Here, H_{m0} values from the northerly position were compared to a Fisher Tippett type 1 distribution (FTI) (fig. 6.6) and to a Weibull two parameter (Wb2) distribution (fig. 6.7). The latter provides the

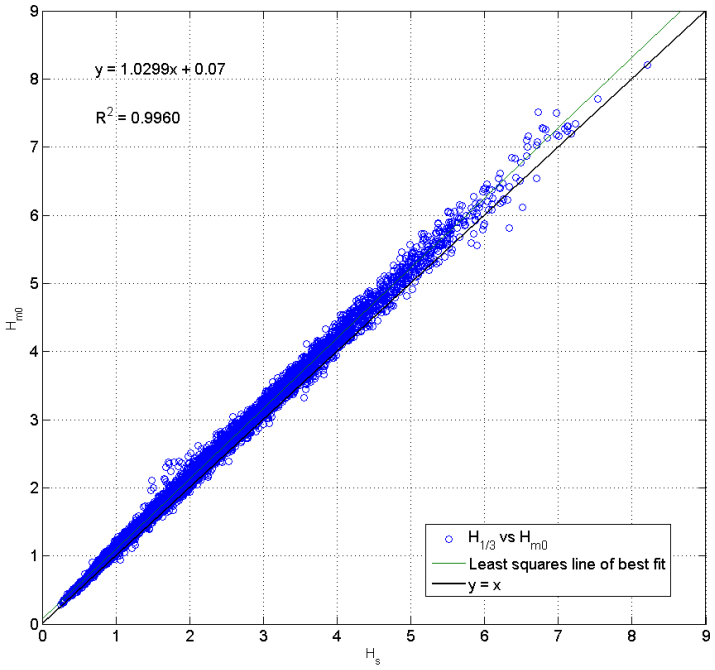

Figure 6.5: A scatter diagram of $H_{1/3}$ vs H_{m0} from the sample data set. Also shown is a lines of best fit estimated from an ordinary least squares regression.

better fit across the full range of values, whereas the Weibull distribution provides a better fit at the tail, which is in agreement with the literature (Ferreira and Guedes-Soares, 2000).

Return values estimated from these data are smaller than those predicted for the Wave Hub site (tab. 6.2). However, these values cannot be considered accurate because 12 months of data is not sufficient to account for inter-annual variability. Furthermore, the gap in the data during the most active period of the year is likely to reduce the estimated return value. Differences between the wave climates at the sites may also contribute.

6.4 Wave power

The measured spectra were used to calculate the total wave power incident on the buoy using equation (2.1.15) (fig. 6.8). The mean power measured at the site was $19.65kW/m$, with monthly mean values between $6.17kW/m$ during May 2010 and $81.3kW/m$ during November 2009 (fig. 6.9). The large seas measured during November 2009 which led to failure of the mooring system, can be seen to have a large effect on the monthly mean. The low return during December occurred because the re-deployment was delayed by a

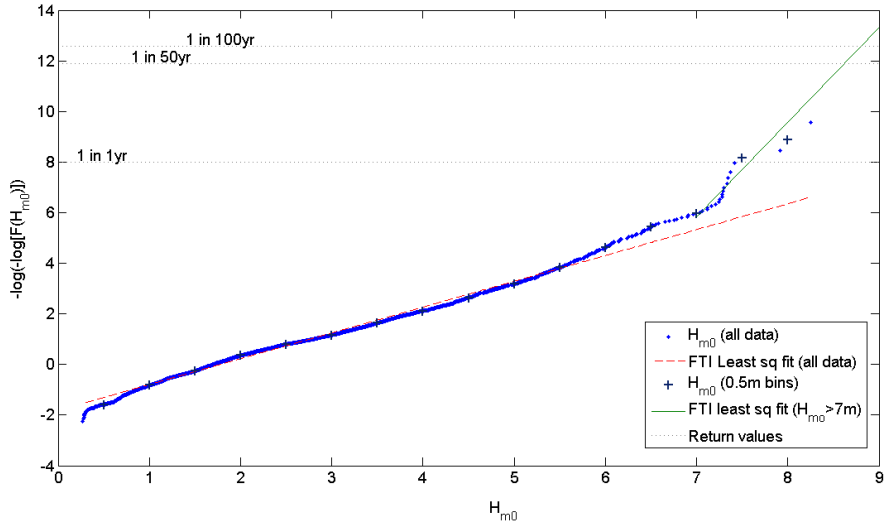

Figure 6.6: The frequency with which data did not exceed values of H_{m0}, plotted on axes scaled that data from a FTI distribution will be a straight line. Also shown are the results of fitting of these data to a FTI distribution, and the probabilities associated with return values.

Table 6.2: Return values for H_{m0}, estimated for the Wave Hub site, using a long-term data set from a global wave model (Halcrow, 2006a), and using data from the northerly position within the array. These are for comparison purposes only and estimates from the measured data cannot be considered accurate due to the limited data set.

Return period	Return values	
	Wave Hub	Northerly position
1yr	10.4	8.6
50yr	13.8	10.8
100yr	14.4	11.2

period of bad weather. Therefore, in this case, the missing data certainly includes large sea states and the mean shown here is inaccurate.

The distribution of wave power for a site can be displayed using a power matrix, as defined in Pitt (2009). To calculate a power matrix, the total power for each record is calculated from the spectrum, P_t, and the total power in the data set is calculated as $Ptot = \sum P_t$. Subsequently, each record is categorised into bins according to H_{m0} and T_{m02} values, as with the H-T scatter diagram, and the total wave power attributed to each cell is calculated. The plot shows the proportion of the total power for the data set that

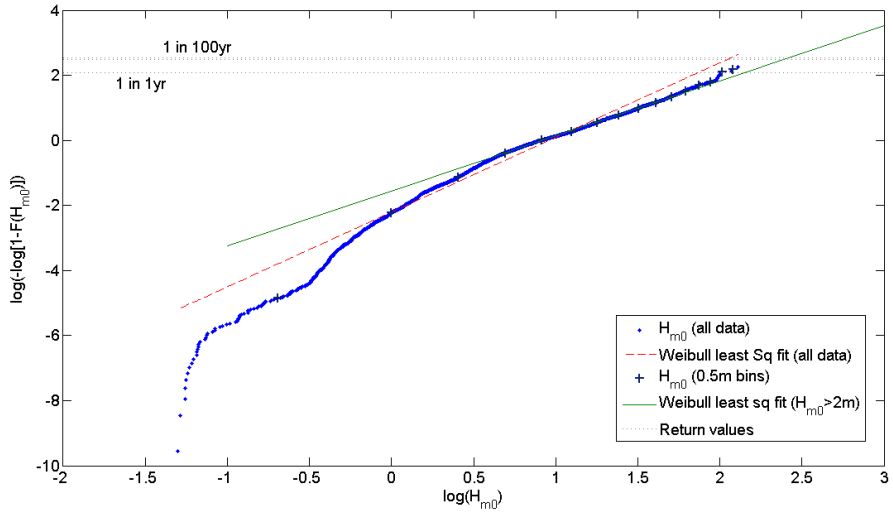

Figure 6.7: The frequency with which data did not exceed values of H_{m0}, plotted on axes scaled that data from a Wb2 distribution will be a straight line. Also shown are the results of fitting these data to a Wb2 distribution, and the probabilities associated with return values.

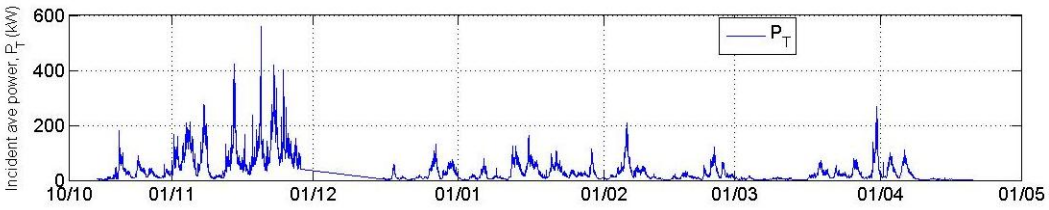

Figure 6.8: A time-series of incident wave power, P_T, estimated during post-processing using a variable filter for the months October 2009 until April 2010.

occurred in each bin. These figures are useful as they provides an indication of the sea states to which the incident wave power can be attributed. Because wave energy devices will convert energy at different efficiencies, according to their varying response to different sea states, a power matrix aids device-specific resource assessment.

Figure 6.10 plots the power matrix for the data set from the northerly position in the array. When compared to figure 6.4, the sea states that occur the most frequently do not provide the highest proportion of the incident power. Rather, the highest poroportion of incident wave power occurs in seas where $4m < H_{m0} < 5m$ and $7s < T_{m02} < 8s$. Furthermore, these cells are associated with seas close to, and above, the 1/19.7 line of constant steepness, which relates to the Pierson-Moskowitz spectrum for fully-developed seas. This suggests that the most power is available in local storms and developing seas, although a significant contribution is observed from seas that are more swell-dominated.

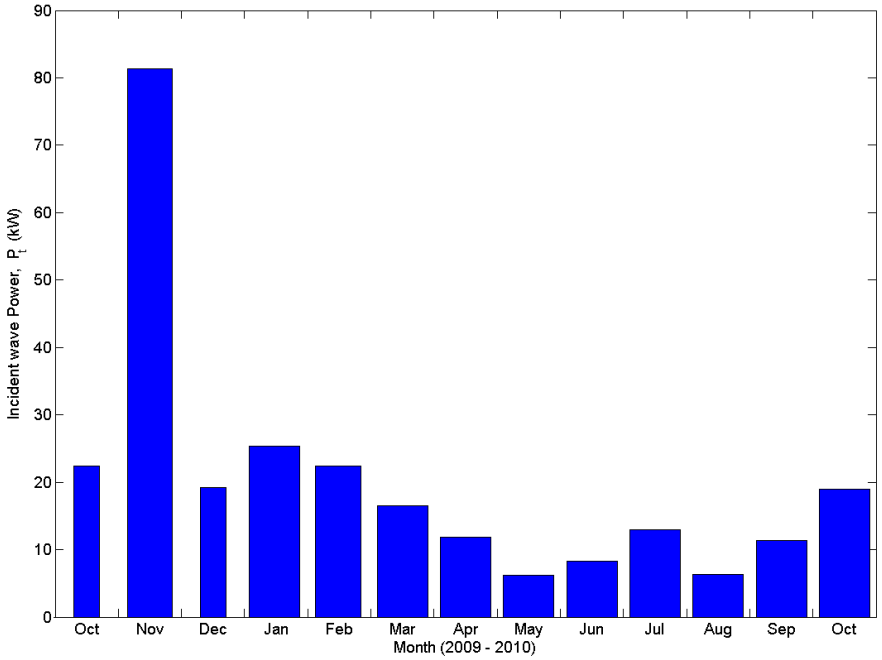

Figure 6.9: The monthly mean power, P_t, calculated from the spectra of records captured from the Northerly position within the array between October 2009 and October 2010. The width of the bars represents the percentage of data returns for each month.

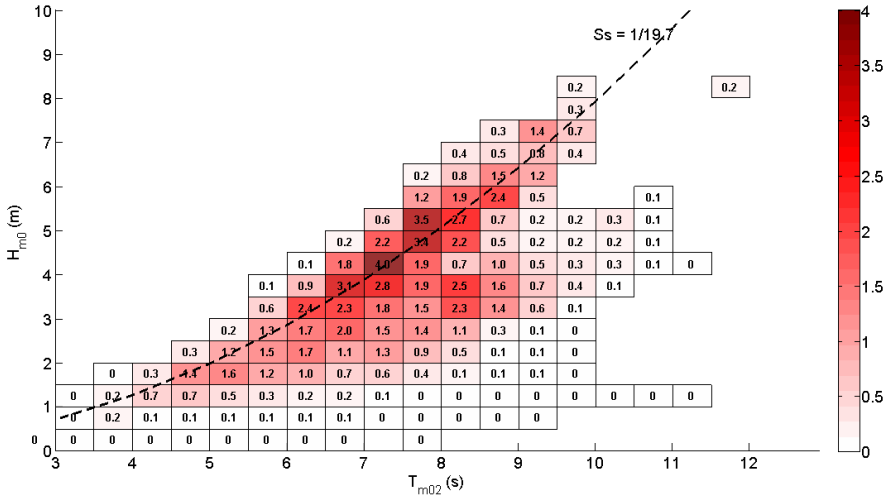

Figure 6.10: A H_{m0}, T_m power matrix for buoy B data from October 2009 - April 2010. Text and colours represent the proportion of all power that occurs in that cell, expressed as a percentage, or the total power (values in italics), expressed in kW/m. For a full definition see (Smith et al., 2006a).

6.5 Directional parameters

The primary consideration of the research presented has been non-directional parameters. However, directional spectral analysis was performed on-board the buoy, and these were examined for the sample data set. Directional properties of the wave field can be categorised by the mean direction, m_{dir}, calculated as,

$$m_{dir} = arctan(b/a) \tag{6.5.1}$$

$$\text{where, } a = \frac{\int_{0.04}^{0.5} S(f)cos(\theta_0(f))df}{\int_{0.04}^{0.5} S(f)df} \tag{6.5.2}$$

$$b = \frac{\int_{0.04}^{0.5} S(f)sin(\theta_0(f))df}{\int_{0.04}^{0.5} S(f)df} \tag{6.5.3}$$

$$\text{and, } \theta_0 = arctan(\frac{b1}{a1}) \tag{6.5.4}$$

where a_1 and b_1 are the first order Fourier co-efficients of the directional distribution. As such, m_{dir} represents the direction of the vector sum of all components of a directional spectrum (Tucker and Pitt, 2001). A full definition of the directional analysis performed on-board can be found in the relevant manual (Oceanor, 2009). Notably, these data were processed using the standard frequency filter and, therefore, may contain errors due to the erroneous low frequency energy. The directional properties of these errors has not yet been investigated.

The majority of records show waves propagating from the westerly and south westerly sectors, with a small proportion from the north (fig. 6.11). This is as expected with the geography of the site, which has a long Atlantic fetch to the west and south west, very little fetch from the south and east, and limited fetch from the north (fig. 3.1).

During certain periods, a periodic oscillation was observed in the time-series of m_{dir} (figs. 6.12a, and 6.13b), and it was hypothesised that this represents a tidal signal in the data. In order to examine this oscillation, a moving average was applied to the m_{dir} parameter data, with a window size of 12 hrs, in order to smooth tidal oscillation. $\varsigma(m_{dir})$ was then calculated as the difference between each individual point and the moving average,

$$\varsigma(m_{dir,i}) = \bar{m}_{dir,i} - m_{dir,i} \tag{6.5.5}$$

where $i = 1, 2, 3...n$ and n is the number of records in the data set. Thus, $\varsigma(m_{dir,i})$ represents the deviation of mean direction from the moving average for each record. The time series of $\varsigma(m_{dir,i})$ highlights the observed oscillation, which is of the same frequency as that of the tidal level (figs. 6.12b and 6.13b). The observed oscillations are 90^o out

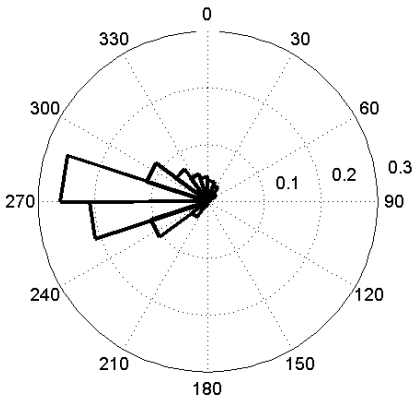

Figure 6.11: A wave rose plotted for all data from the northerly position, using the mean wave direction, m_{dir}. The size of the bars show the proportion of records where the mean direction of wave propagation is from within each directional bin, of width 18^o. 0^o relates to magnetic north.

of phase with the tidal level suggesting that it is driven by tidal flow. When the tide is in flood, the currents will flow north east to south west, and the direction of the waves becomes more northerly. When the tide is in ebb, the direction of the waves becomes more southerly. This suggests that the waves are showing a tendency to oppose the direction of the current. A detailed directional analysis, is described by (Saulnier et al., 2011a), who also found directional parameters to be highly correlated to tidal conditions, and attributed this predominantly to variability in the higher frequencies of the spectrum.

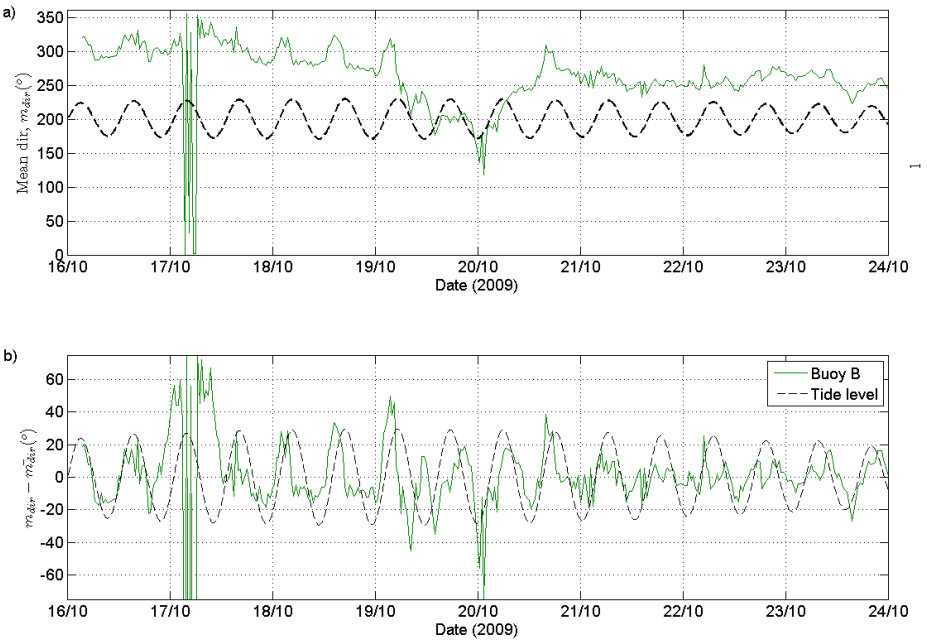

Figure 6.12: Measurements of the mean direction, m_{dir} made by buoy B during October 2009, compared to the tidal signal, with figure a) showing a time series of m_{dir} values, and b) showing the variation from a moving average, $\varsigma(m_{dir})$.

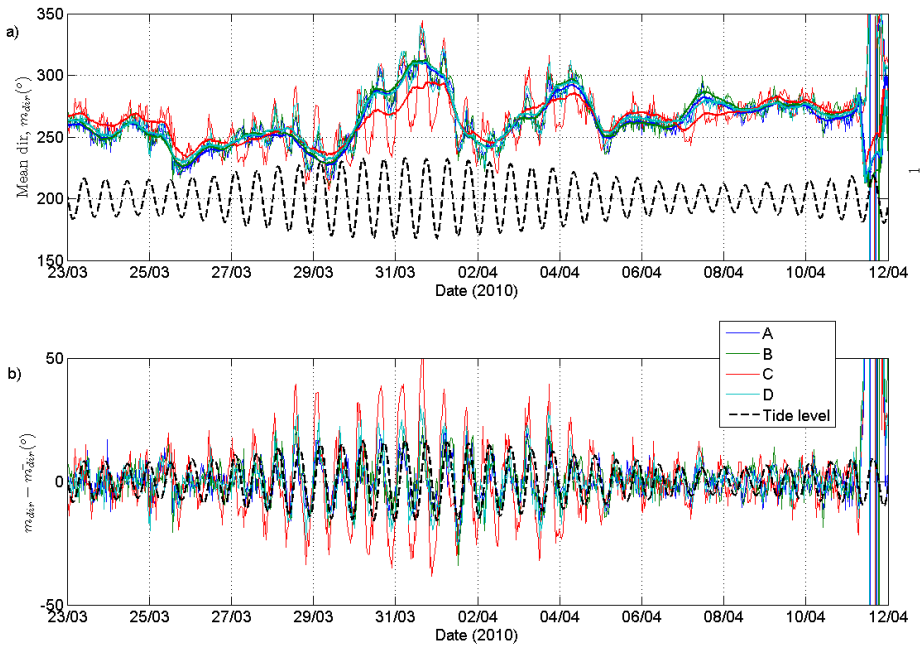

Figure 6.13: Measurements of the mean direction, m_{dir} made by buoys A,B,C and D between January and April 2010, compared to the tidal signal, with figure a) showing a time series of m_{dir} values, and b) showing the variation from a moving average, $\varsigma(m_{dir})$.

6.6 Data from 4 buoys

Data from the 1st deployment is limited by the failure of buoy A on November 7th 2009, and the subsequent failure of buoys C and D later in the month. The gap in the data captured by buoy B during the second deployment, which occurs during April 2010, was caused by a data recording failure (fig. 6.14).

Data from the 4 buoys can be seen to agree in terms of the general wave conditions, and the time-series of key parameters are closely matched. Closer inspection of a subset of the data, indicates the periodic oscillation, most strongly observed in T_{m02} values, and occurring at the frequency of the tides, is present in all buoys (fig. 6.15).

Mean values vary between the buoys, in particular the wave height and power statistics from the first deployment, and the return value estimations for H_{m0} (tab. 6.3). These differences are attributed to the sea conditions that contribute to the data sets from the buoys. Buoy B measured for longer during the more active winter period, subsequent to the other buoys failures and this difference in deployment length causes the observed differences in the statistics. This demonstrates that for analysis of the spatial differences in the wave field within the array, it is essential to limit data to records captured simultaneously.

Table 6.3: Mean values for key parameters from all four buoys during both deployments.

Deployment 1			
Buoy	$H_{m0}(m)$	$T_{m02}(s)$	$P_t(kW)$
A	2.74	6.84	43.65
B	3.25	6.94	60.25
C	2.87	6.91	49.47
D	2.82	6.76	46.93
Deployment 2			
A	1.7	5.83	15.32
B	1.69	5.78	14.68
C	1.7	5.84	15.3
D	1.66	5.75	14.15

6.6.1 Data availability

To ensure comparability between data sets for the analysis in chapters 7 and 8, data were constrained to records where all 4 buoys were measuring simultaneously. Availability is

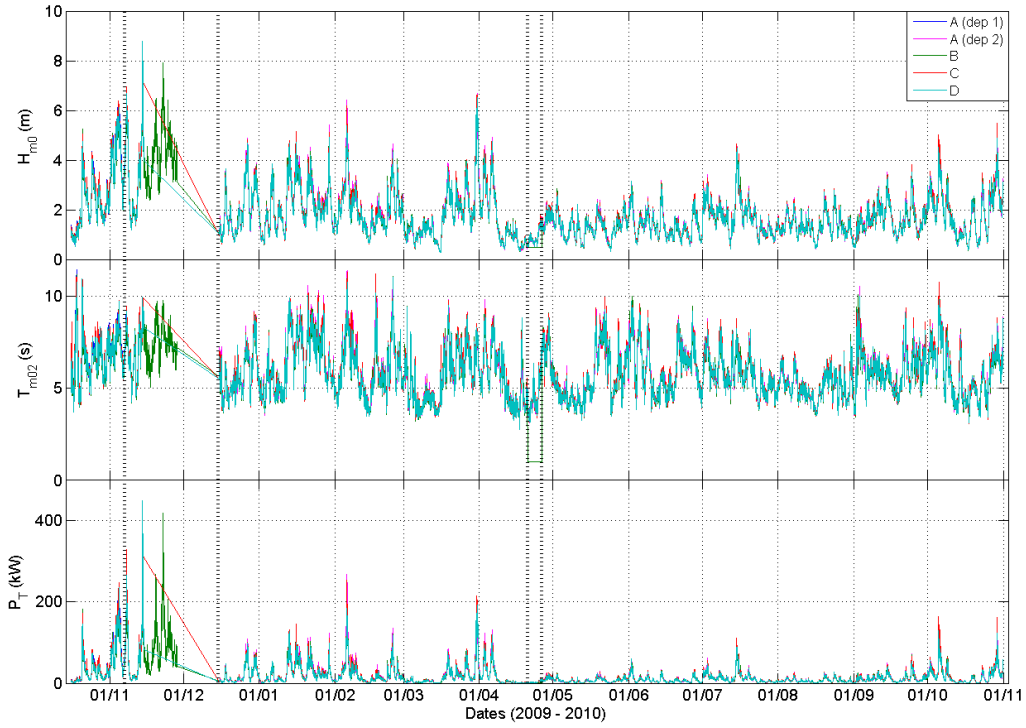

Figure 6.14: Time-series of a) Significant wave height, H_{m0} b) Mean zero-crossing period, T_{m02} and d) Total power, P_T for all records from all four buoys. Dotted lines indicate the two extended periods of missing data from one or more buoys. Data from these periods do not contribute to a combined data set limited to simultaneous data.

therefore limited by the combined failure rates of all sensors; an error in a single record from one buoy results in the exclusion of simultaneous data from the other 3 buoys.

Table 6.5 demonstrates the number of records recorded during normal operation (after removal of records identified in the log-book) and the number of records available after QC procedures for each buoy and for the combined data set. The final data set comprises 14673 individual records from each buoy, measured between 16-Oct-2009 and 31-Oct-2010. This is an 80% return compared to an average return of 89% from individual buoys (91% prior to QC).

Notably, the period following the failure of buoy A during November 2009 was very active, and the buoys that remained operational recorded the largest significant wave height in either deployment. It follows that the mean statistics of the simultaneous data set are not representative of the annual wave climate.

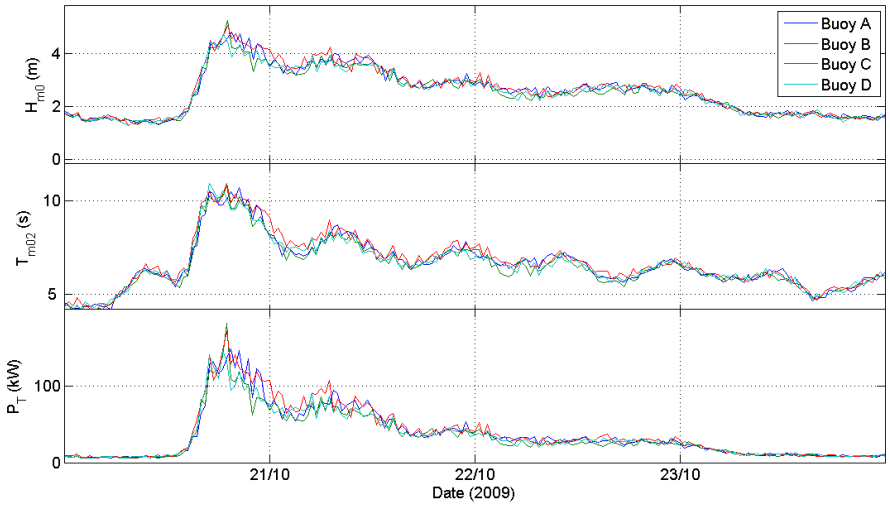

Figure 6.15: Time-series of a) Significant wave height, H_{m0} b) Mean zero-crossing period, T_{m02} and d) Total power, P_T for records from all four buoys captured during October 2009.

Table 6.4: Return values for H_{m0}, estimated for the Wave Hub site, using a long-term data set from a global wave model (Halcrow, 2006a), and using data from all buoys within the array. These are for comparison purposes only and estimates from the measured data cannot be considered accurate due to the limited data set.

Return period	Return values				
	Wave Hub	A	B	C	D
1yr	10.4	6.8	8.1	7.2	7.1
50yr	13.8	8.9	10.7	9.5	9.3
100yr	14.4	9.2	11.1	9.9	9.7

Table 6.5: The number of individual records available from each of the 4 buoys, and from the combined data set where data were limited to instances where error-free records were available from 4 buoys simultaneously.

Buoy	Recorded		After QC	
A	16342	89%	16031	88%
B	17012	93%	16677	91%
C	16645	91%	16308	89%
D	16768	92%	16436	90%
Combined	15834	81%	14673	80%

6.7 Prediction of correlation between simultaneous records

Section 2.4 reviewed correlation between spatially separated wave measurements and provided a theoretical method for quantifying the limit of correlation (eqns. (2.4.3) and (2.4.4)). It was hypothesised that correlation between surface elevation records will cause coherence in their error terms, and therefore alter statistical properties of the differences between simultaneous data. This has important connotations for those methods described in section 2.4, designed to quantify the error term in wave measurements using multiple measurements. Measurements taken by the buoys are used in this section to discuss the potential for correlation between simultaneous data captured within the array.

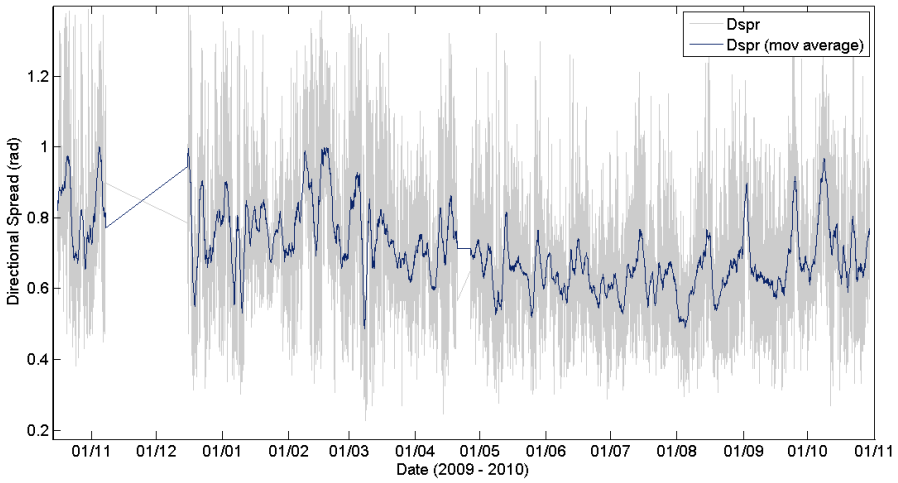

Figure 6.16: Directional spreading values measured at the Northerly position in the array for the entire data set.

Equations (2.4.3) and (2.4.4) show that the limit of correlation depends upon the directional spreading of the incident wave field, d_{spr}, and figure 6.16 shows the time-series for this parameter. A moving average over 12 consecutive hrs is also shown, which removes short-term variability and gives a clearer indication of changes on scales longer than a day. High directional spreading is associated with wind seas and conversely, low directional spreading occurs during swell conditions. It can be seen that directional spreading is higher during the winter months when storms are more active over the measurement site. Within the data, a minimum of $0.2rad$ occurred during March 2010 and a maximum of $1.39rad$ during December 2009.

For all data, the limit of correlation both perpendicular, D_{perp}, and parallel, D_{par} to the direction of propagation of an incident wave train were calculated using the measured

directional spreading, the peak frequency of the spectrum, f_p (estimated from a repaired spectrum), and the deep water wavelength[1], calculated as $\lambda_0 = \frac{g}{2\pi f_p^2}$. 66% of records can be expected to exhibit correlation in the time-series over $500m$, or 47% over the longer separation between buoys of $700m$, in a direction parallel to the propagating waves. Conversely, no records are expected to exhibit correlation at $500m$ in a direction perpendicular to wave propagation (fig 6.17). The stratification of the data observed in this figure can be attributed to the resolution of the input data for the parameter d_{spr}, as derived from standard processing on-board.

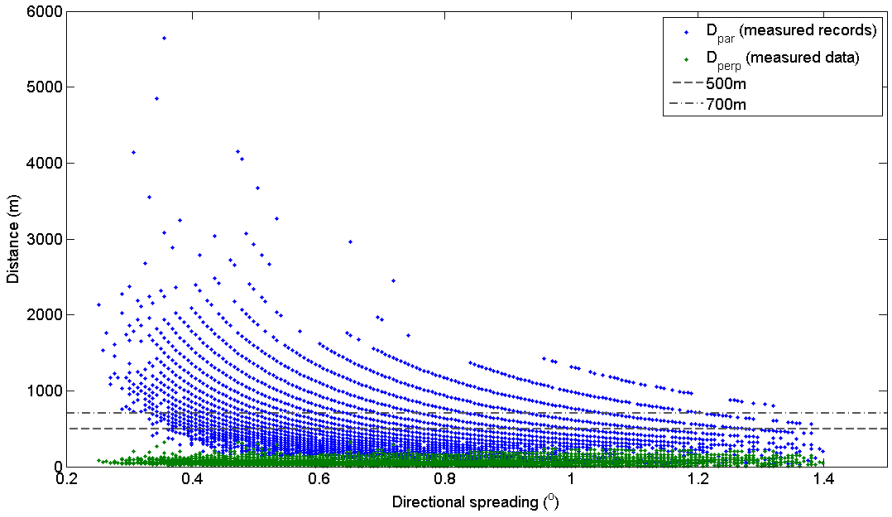

Figure 6.17: The theoretical limit of coherence calculated using equations (2.4.3) and (2.4.4) from measured data, plotted against measured directional spreading values.

The difference between the results for D_{par} and D_{perp} observed in figure 6.17 demonstrates the significance of the direction of propagation to the likelihood of correlation between the time-series. For example, where a wave field is travelling WSW - ENE, measurements taken at the southerly (S) and easterly (E) positions are likely to be correlated (dependent on conditions), whereas measurements taken at the northerly (N) and E position will not. To reflect this dependence on direction, the data were categorised according to the direction from which the incident waves were propagating (defined using the parameter m_{dir}) into directional sectors covering 45^o each. For each category, pairs that were aligned parallel to the direction of propagation were identified, and the proportion of records for which $D_{par} > L$ was calculated, where L is the distance separation (either $500m$ or $700m$).
The results for each sector are shown in table 6.6, and demonstrate that correlation can

[1]The validity of deep water formulations at this site have been discussed in section 3.1.

Table 6.6: The proportion of records where the limit of correlation is longer than the separation between wave buoys, $D_{par} > L$ for waves propagating from a range of directional sectors.

Sector	Classification	Pairs parallel	Deployment 1 (%)		Deployment 2 (%)	
			sector	all data	sector	all data
S-SW	$180 < m_{dir} < 215$	S:N ($700m$)	74	35	67	18
SW-W	$215 < m_{dir} < 270$	W:N and S:E ($500m$)	44	19	56	22
W-NW	$270 < m_{dir} < 315$	W:E ($700m$)	30	< 1	62	7.5
NW-N	$315 < m_{dir} < 360$	W:S and N:E ($500m$)	33	< 1	60	4
N-NE	$0 < m_{dir} < 45$	N:S ($700m$)	-	-	63	< 1
NE-E	$45 < m_{dir} < 90$	N:W and E:S ($500m$)	42	1	76	< 1
			Total:	56	Total:	51.5

be expected to occur for all pairs, although the most common occurrence is for waves propagating from the SW-W sector which affect 54% of the first deployment and 40% of the second deployment. From these results, it can be concluded that correlation would be expected to affect the random variability of differences between the S and N wave buoys, as well as between the W:N and S:E pairs. Correlation will also affect other pairs of buoys although the wave conditions measured during this experiment mean that the effect on the overall data set will be limited or negligible.

6.8 Discussion

Once the standardised processing had been applied, the analysis of data from 12 months of operation revealed a data set of high quality, with data returns of up to 99% during normal operation. These data demonstrate an energetic site with a mean wave power of $19.6kW/m$, which is comparable to results published in DTI (2004). Monthly mean values of wave power demonstrate a high variability between the months, and the potential effect of missing data on the values of long term averages. These monthly values compare well to those published for the Wave Hub site (Halcrow, 2006b), although the mean wave power measured for the month of November, 2009, is seen to be larger than any measurements or model predictions taken during 2005-2006 (Pitt et al., 2006).

The directional properties of the data set were as expected for the geographical location of the site, and the climate of the North Atlantic. However, directional parameters were seen to exhibit a strong tidal signal.

Data from the 4 buoys can be seen to agree in terms of the general wave conditions,

with no immediate evidence of severe differences. The difference between mean values demonstrates that in order to derive meaningful conclusions about the differences between the wave conditions across the array, it is essential to limit data to records captured simultaneously.

The analysis in section 6.7 indicates that correlation between simultaneous time series captured by the the S and N wave buoys, as well as between the W:N and S:E pairs. The effect of correlation on the other pairs of buoys will be limited or negligible.

Chapter 7

Comparing data from 4 buoys

This chapter quantifies the measured differences between simultaneous measurements, using key parameters to represent the wave conditions. Initial examination found a persistent difference between values measured at separate locations in the array, that cannot be attributed to random statistical variability, arising from sampling error. Further analysis aims to identify a dependence, or causal, relationship with the physical conditions at the site. The result is a direct measurement of the spatial homogeneity of the wave field at the measurement site, and how this can be affected by local conditions.

7.1 Methodology

The identification of persistent or deterministic differences requires quantification of whether a measurement from one position is greater or smaller than that at another. This was achieved by defining dX_i as the proportional difference between two measurements (given previously in section 2.3),

$$dX_i = \frac{(X_{1i} - X_{2i})}{\frac{1}{2}(X_{1i} + X_{2i})} \tag{7.1.1}$$

where X_{1i} and X_{2i} are simultaneous mesaurements of a given parameter, X. dX_i represents the difference between values as a fraction of their overall value and will take a positive value when $X_{1i} > X_{2i}$, and vice-versa. For a long-term data set, a non-zero mean for dX_i indicates a bias between X_{1i} and X_{2i}. As described in section 2.3, the coefficient of variation can be calculated using the values of dX as, $c.o.v. = \sqrt{2}dX$ (Mackay, 2009; Goda, 1977).

The proportional differences, dX_i, in key wave parameters (H_{m0}, T_{m02} and P_t) were calculated, and the properties were examined to identify whether differences are consistent, if they vary in time, or if they are affected by infrequent large values. Each of these conditions are indicative of a source of differences that cannot be accounted for by random sampling variability. dX was also compared with factors governing wave propagation and interaction with local bathymetry, specifically tidal conditions, wave direction, and the wavelength (represented by the wave period).

7.1.1 Multivariate least squares regression

A multivariate least squares regression methodology was applied to measured wave data with the aim of identifying whether the physical conditions at the measurement array were affecting the wave conditions, and to quantify this effect. Through regression of significant wave height measurements from one buoy on simultaneous measurements from another,

the outcome is a set of linear equations that predict the significant wave height measured at a location within the array, based on the value measured at another, and the value of covariates. Covariates were chosen to represent physical processes affecting the wave field within the array.

As described in section 2.3.1, linear regression techniques derive a linear formula to represent the relationship between two sets of measured data, where the data are paired. For the case where multiple factors affect the relationship, these can be included in the analysis, to give an equation of the form,

$$\hat{y}_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} ... \beta_k x_{i,k} + \epsilon \tag{7.1.2}$$

for $k = 1 : N$ covariates and $i = 1 : n$ records. β co-efficients are optimised such that the sum of the squares of the deviances of each observation, y_i, from the estimator, $\hat{y}_i$ are minimised. Thus, for each sample, the least squares criterion to be minimised is,

$$S(\beta_{1:k}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7.1.3}$$

With multiple regressors (or covariates), β_k represents the expected change in y_i with changing x_k, when all other x_i values are kept the same (Draper and Smith, 1981; Montgomery, 1982).

Analysing the model

The methodology used in analysing the model is based on the assumption that the error terms, $y_i - \hat{y}_i$, are Gaussian distributed. Drawing on central limit theory, this assumption will be correct if the errors in the measurements are from a combination of a large number of sources of error, regardless of their individual distributions. However, in the case where errors are principally caused by one or few sources, their error distribution may cause that of the final measurements to deviate from Gaussianity. A serious deviation from Gaussianity in the errors will lead to inaccurate levels of confidence placed in the results of a least squares analysis. To test this assumption, a Kolmogorov-Smirnov test was applied to the residuals, with the null hypothesis that the residuals, $y_i - \hat{y}_i$, follow a Gaussian distribution.

Before drawing conclusions based on the derived model, or making predictions about the system in question, it is important to establish both its precision and stability. The key statistic used to analyse the regression model is the deviation of the data (SST), which can be split into the deviation accounted for by the regression (SSR), and that caused by

unexplained variability (SSE). These are defined as the sum of squares as follows,

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{7.1.4}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7.1.5}$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{7.1.6}$$

These statistics were applied to evaluate the significance of the regression model, by establishing whether the change in the data that is accounted for by the model, SSR, is significant when compared to unexplained variability, SSE. To achieve this, the sum of square values are converted to mean square values by dividing by the appropriate degrees of freedom and a statistic, F, was derived as the ratio of the mean square regression, MSR to the mean square errors, MSE,

$$F = \frac{MSR}{MSE} \tag{7.1.7}$$

Statistics that are calculated as the sum of the squares of a number of independent standard Gaussian random variables, will follow a Chi-squared distribution. The ratio of two chi-squared distributed variables is known to follow an F distribution. The appropriate degrees of freedom can be calculated for each sum of squares,

$$dof_R = N \tag{7.1.8}$$

$$dof_E = n - N - 1 \tag{7.1.9}$$

where N is the number of X values (covariates), and n is the length of the data set. This methodology for the comparison of two measures of variance (which will be chi-squared distributed) is commonly known as analysis of variance (ANOVA).

This formulation can be applied to a hypothesis test, with null hypothesis that the variability in the data that is unexplained is equivalent to the variation accounted for by the regression model,

$$H_0 : MSR \approx MSE$$

$$H_1 : MSR > MSE$$

Thus, the probability of the statistic F being drawn from the distribution $F(N, (n-2))$ is calculated using look-up tables, and where this is less than 0.01, the null hypothesis is rejected and the regression is deemed significant.

For multivariate analysis, it is critical to establish whether each covariate is having a significant effect on the model. Analysis of the β_k coefficients was also based on a hypothesis test, which establishes whether each coefficient is significantly different from 0. To perform such a test, the variability of each parameter must be calculated and Draper and Smith (1981) define the variance of each parameter as,

$$var(\beta_k) = \frac{\sigma^2}{SS_{xx}(k)} \tag{7.1.10}$$

where, σ^2 can be estimated by the mean square error term, MSE, and SS_{xx} is the sum of squares of the regressor values, x_k,

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{7.1.11}$$

Retaining the assumption that the error term is Gaussian distributed, the observations, y_i, are also Gaussian distributed as $N(\hat{y}_i, \sigma^2)$. Because β_k is a linear combination of the observations, it too will follow a Gaussian distribution, with standard deviation calculated from equation (7.1.10), $N\left(\beta_k, \sqrt{\frac{\sigma^2}{SS_{xx}}}\right)$. Thus, the statistic,

$$t_k = \frac{\hat{\beta}_k - \beta_{test}}{\sqrt{\frac{\sigma^2}{SS_{xx}}}} \tag{7.1.12}$$

will be distributed $N(0,1)$ in the instance where $\beta_k = \beta_{test}$. When using MSE as an estimator of σ^2, t_k will be t distributed with $(n-2)$ degrees of freedom, when $\beta_k = \beta_{test}$. This formulation can be applied to a hypothesis test, with null hypothesis that the value of a coefficient, β_k is not significant,

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Thus, $\beta_{test} = 0$ and equation (7.1.12) becomes,

$$t_k = \frac{\hat{\beta}_k}{\sqrt{MSE/SS_{xx}}} \tag{7.1.13}$$

The probability that t_0 would be exceeded by a random variable drawn from a t distribution with degrees of freedom, $n-2$, was then calculated from look-up tables, $p = P(t > t_k)$. If p is smaller than a chosen critical probability level, then the null hypothesis was rejected. In this case, where $p < 0.01$, it was concluded that the co-efficient is significantly different from 0, and therefore the parameter, x_k, is having a significant effect on the model, and on the wave heights measured across the array.

7.1.2 Covariates used in multivariate regression

Covariates were chosen to represent the physical processes that may affect waves within the array, as follows,

- Mean zero-crossing wave period, T_z - In deep water the wave period is proportional to wavelength[1], which will govern the interaction of waves with the bathymetry. The time-series derived parameter was used because the use of spectral moments in the estimation of T_{m02} will introduce direct correlation between this parameter and the dependent and independent variables.

- Tidal conditions, V - Depth integrated tidal data was supplied by the Polpred (POL, 2010) model as three separate parameters, tidal level above mean water level, tidal flow speed and tidal flow direction. The mean axis of flow was estimated from these data using an OLS line of best fit, which gave the equation $N = 0.83E$, which broadly corresponds to a line running SW to NE. The tidal flow speed and the tidal flow direction were combined to give tidal flow vector for each record along the mean axis of flow. A positive value represents a flood tide and a negative value represents an ebb tide.

- Mean wave direction, m_{dir} - The wave direction will affect wave-current interactions as well as the interaction of waves with the bathymetry.

- Directional spreading, spr_{Tp} - The directional spreading will govern the coherence between spatially separated waves.

Thus, where significant wave heights are to be analysed, the model to be estimated will take the form,

$$\hat{H}_{m01} = \beta_1 + \beta_2 H_{m02} + \beta_3 T_{z2} + \beta_4 V + \beta_5 m_{dir2} + \beta_6 spr_{Tp2} \tag{7.1.14}$$

where H_{m01} and H_{m02} are simultaneous estimates of H_{m0} from two separate buoys within the array. This will yield 6 β coefficients, which describe the observed relationship between physical conditions and the wave heights measured at one of the buoys. These can be used to estimate the data set for the dependent variable, H_{m01}, which is equivalent to the line of best fit. Notably, despite the linear function, when covariates are included in the model, estimates of the dependent variable will not produce a linear plot. The model

[1]The deep water assumption is retained here, although for the wave conditions observed at this site, it is regularly defined as intermediate water.

was designed such that it could be used to predict H_{m0} at a separate point using all the available information from a wave buoy measurement. In this way, it replicates a possible application to a wave energy site with limited measurement resource. Inclusion of data from the other buoys in a single model would allow a higher sensitivity.

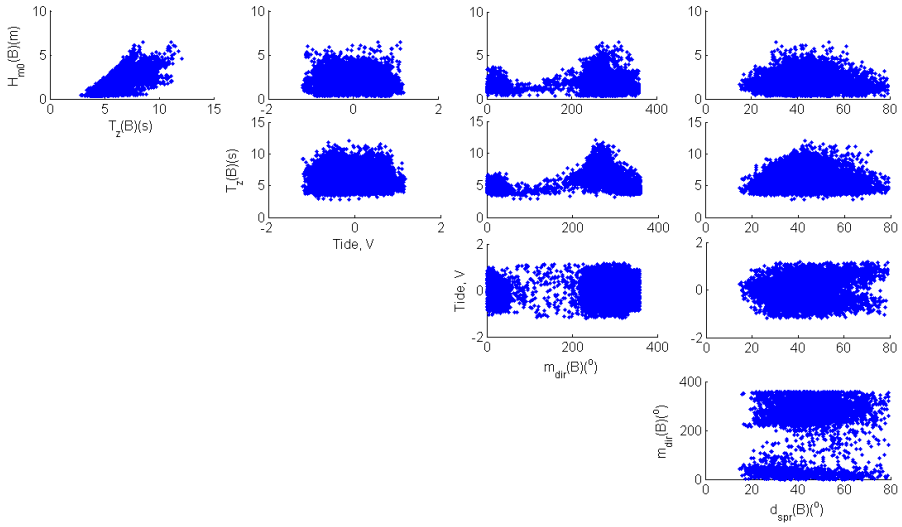

Figure 7.1: A set of figures showing the covariates plotted against each other as scatter plots. This is a diagnostic plot used to identify dependency in covariates.

This methodology represents a simplification of a complex system and there are certain considerations in its application. Correlation between co-efficients will lead to inaccuracies in the estimation of β_k values. The characteristics of the site mean that swell wave systems only arrive from the western sector. This was seen to cause correlation between the wave characteristics, T_{m02}, and spr_{Tp}, and the wave direction, m_{dir} (fig. 7.1). Because a large amount of variability of these parameters will also occur independently of the wave direction, this correlation was initially assumed negligible for the analysis. Correlation between T_z and H_{m0} has been extensively studied (Longuet-Higgins, 1975), and means that some variation in the significant wave height could be attributed to the wave period. This must be considered when drawing conclusions.

The coefficients estimated for this model quantify the relationship between each covariate and the input parameter. Hypothesis testing is applied to test whether each co-efficient is significantly different from 0, which indicates that its effect (and therefore the effect of the relevant physical parameter) on the correlation between dependent and independent variables is significant. A positive result here suggests that a physical process is likely to be causing a difference to the wave fields across the array. The combination of the sig-

nificant co-efficients provides a statistical model which describes the relationship between the dependent variable and covariates. This could potentially be applied to the prediction of wave fields across the array, based on measurements taken from a separate location (or locations).

7.1.3 Weighting for a variable error term

Section 2.3 has discussed the effect of a heteroscedastic error term in regression analysis. To overcome this variable error term in wave data, regression was performed using a weighting vector that governs the level of significance afforded each datum in the dependent data set (see section 2.3 and equation (2.3.8)). Weightings are commonly applied using the inverse of the estimated error variance for each datum. In this case, the weightings were defined using the inverse of the theoretical standard deviation, $wX = 1/\sigma_X$ calculated using equation (2.2.8). The suitability of the weight function applied was analysed using the weighted residuals and results are provided below (sec. 7.3).

7.2 Results

The mean values for spectral moments and parameters over the whole data set are shown in table 7.1, which have been calculated from 14673 records (see table 6.5). The large number of records means that variability associated with sampling error would be expected to be negligible and the differences between the measurements at the buoys represent a significant bias.

Table 7.1: Mean values for wave parameters (simultaneous data).

Parameter	buoy A	buoy B	buoy C	buoy D	Max proportional difference, dX_μ
H_{m0}	1.78	1.75	1.78	1.74	2.3%
T_{m02}	5.93	5.85	5.94	5.85	1.5%
$P_t(kW/m)$	17.45	16.34	17.42	16.05	8.2%
m_{-1}	2.22	2.08	2.21	2.04	8.8%
m_0	0.25	0.24	0.25	0.24	4%
$m_1 \times 10^3$	34.94	34.36	34.95	33.78	3.4%
$m_2 \times 10^3$	6.15	6.16	6.15	6.05	1.6%

These differences can be related to the time-series of key parameters shown in figure 6.14.

There are periods where values from buoy C and to a lesser extent buoy A, appear larger than the other buoys, which can be observed more clearly when the time-series is shown separately for each month (appendix E). These differences cause similarity between mean parameter values from buoys A and C, which are different from B and D (tab. 7.1). This pattern is mirrored in data for each of the descriptive wave parameters except for higher order spectral moments. The maximum difference in average power observed was $1.4kW/m$ or 8% between the mean incident wave power, $\bar{P}_t$, measured at buoys A and D. The proportional differences between buoy A and the other three buoys were calculated using equation (7.1.1) for H_{m0}, T_{m02} and P_t (fig. 7.2). When examined against time, the pattern observed in the mean differences can be identified. The proportional differences between buoys A and C are commonly closer to 0 than A and B or A and D. However, this pattern is not consistent throughout the data set, and differences are seen to change sign during certain periods. This pattern is not indicative of a calibration error, or other source of constant bias. As with the mean values, the differences are largest for the parameter P_t.

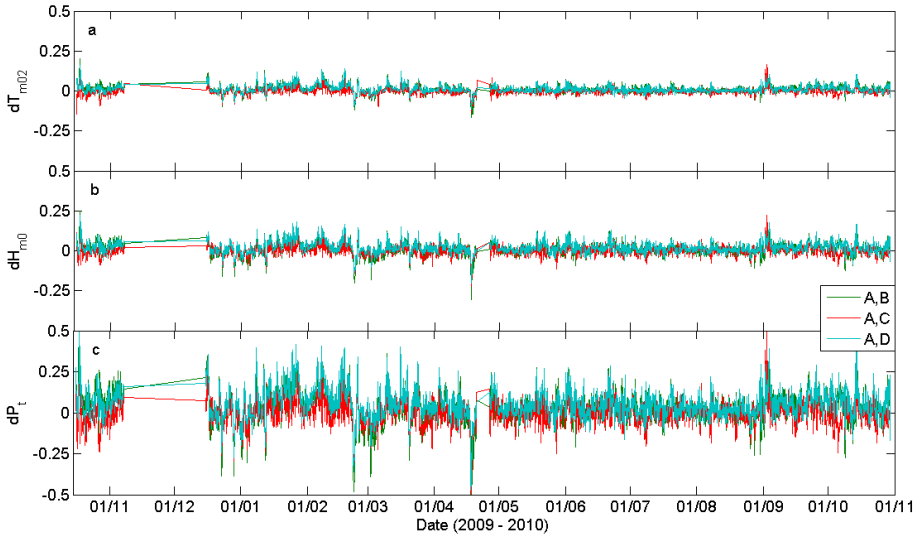

Figure 7.2: The proportional difference between a) T_{m02}, b) H_{m0} and c)P_t values measured by buoy A and the three other buoys in the array, for the entire 12 month data set.

During October and November 2009, a consistent difference is evident between parameter values measured at buoys A and C and those measured at buoys B and D. Differences are greatest in P_t, which leads to the mean measured wave power at buoy A being 11% greater than that at buoy B for the 22 days of simultaneous data capture during the first deployment. There is also a 4% difference in mean H_{m0} and a 3% difference in mean T_{m02}

measured during this time. This trend is repeated during late January and February 2010, to a lesser extent during March and April 2010 and again during October 2010. These periods have been seen to co-incide with the more active sea states measured (see section 6, fig. 6.3).

For the period between 17th Dec 2009 and 5th Jan 2010, the pattern of differences changed, and the mean power measured at buoy A was 5% smaller than that measured at buoy B for the same period. This pattern was repeated for a period of 10 days during late February and early March 2010, and again for 4 days immediately prior to instrument failure during April 2010. During these periods, the pattern of buoys A and C measuring similar values, larger than those at buoys B and D, is not evident, and buoys A and D measure smaller wave power.

7.2.1 The dependence of observed differences on physical conditions

When the differences in incident power, dP_t are separated according to the joint distribution of H_{m0} and T_{m02}, and overlain onto the $H - T$ scatter diagram (fig. 7.3), cells that represent more than 1% of records support both the pattern of the mean values shown in table 7.1, and the trend observed in figure 7.2. For example, for the majority of cells on figure 7.3, data at buoy A is greater than that measured at buoy B. Cells for which these trends were not followed, were limited to small swell seas, and very large seas, which represent less than 1% of contributing records.

The proportional differences between simultaneous parameters captured by the array demonstrate a dependence on the wave conditions at the time of measurement. Examination of figure 7.3 reveals a general trend of P_t at buoy A increasing relative to buoy B as wave period increases relative to wave height. This causes the majority of cells representing a wave period greater than 8s to exhibit a mean P_t buoy A more than 10% greater than the mean P_t at buoy B.

A comparison of the time series of dP_t and T_{m02} indicates correlation between them, with larger values of T_{m02} co-inciding with an increase in P_t measured by buoy A, relative to the other three buoys (fig. 7.4). However, this effect is not consistent, and periods of time where P_t at buoy B is small relative to buoy A do not consistently co-incide with very small values of T_{m02} (fig. 7.4). This observation supports the increase in DP_t observed in figure 7.3. A least squares regression of $dP_t(A, B)$ on wave period, T_{m02}, also reveals a correlation (fig. 7.5), with a linear relationship,

$$dP_t(A, B) = 0.03T_{m02} - 0.16 \tag{7.2.1}$$

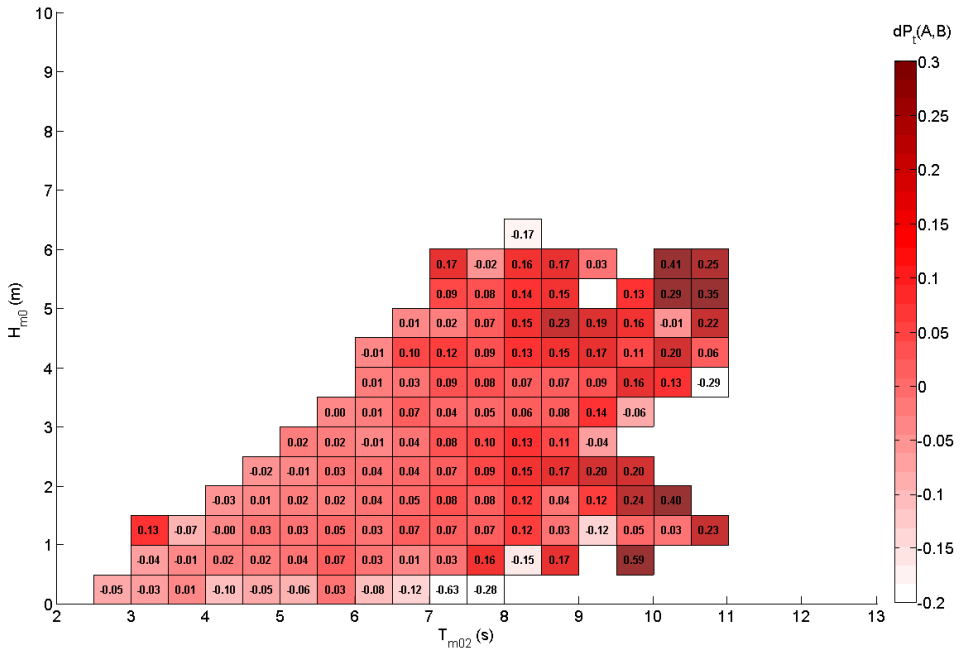

Figure 7.3: The H-T scatter diagram for all records where error-free records were available from all 4 buoys. The text values and the colour overlay represent the average value of $dP_t(A, B)$ for records within that cell.

ANOVA analysis finds this correlation to be significant ($F = 558, p = 0.0018$), although there is a wide spread of points about the OLS line of best fit, with R^2 value of 0.07. These results indicate that the wave period (and therefore, the wavelength) is correlated with differences in the wave field across the array. However, the overall affect will be due to a combination of physical processes, which causes significant spread in the scatter plot.

Directional properties of the wave field measured by the buoy array have been described in section 6.5. The parameter m_{dir} (also described in section 6.5) was used to quantify the direction of propagation of the incident wave field. When m_{dir} was plotted in a scatter diagram against the proportional differences in measured wave power, $dP_t(A, B)$ a directional distribution was observed (fig 7.6). An increased frequency of measurements where $P_{tA} > P_{tB}$ can be observed for wave fields travelling from the west-north-western sector. Conversley, $P_{tA} < P_{tB}$ occurs more frequently when the wave field is travelling from the south western sector (fig. 7.6).

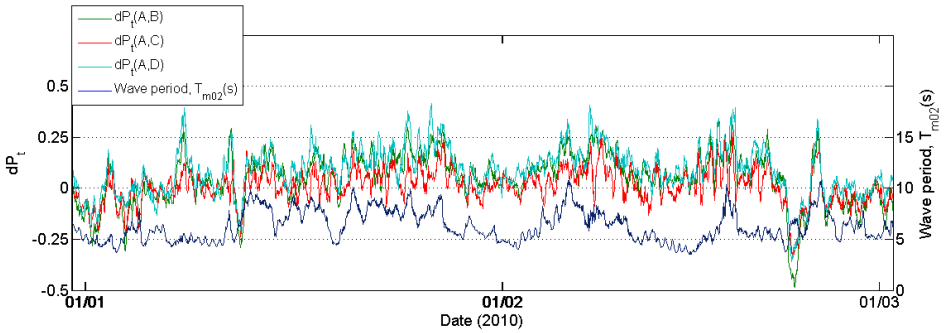

Figure 7.4: The proportional difference between P_t values measured by buoy A and the three other buoys in the array, and the zero-crossing wave period, T_{m02}, measured by buoy A for a subset of the 12 month data set.

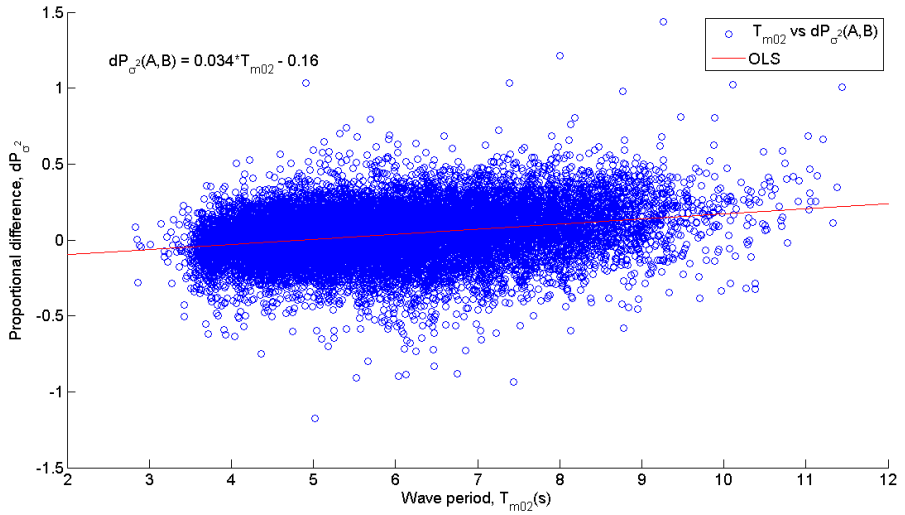

Figure 7.5: Proportional difference in simultaneous measurements of P_t at buoys A and B, compared to the wave period, T_{m02}, measured at buoy A. Also shown is the line of best fit and associated linear model resulting from an OLS procedure.

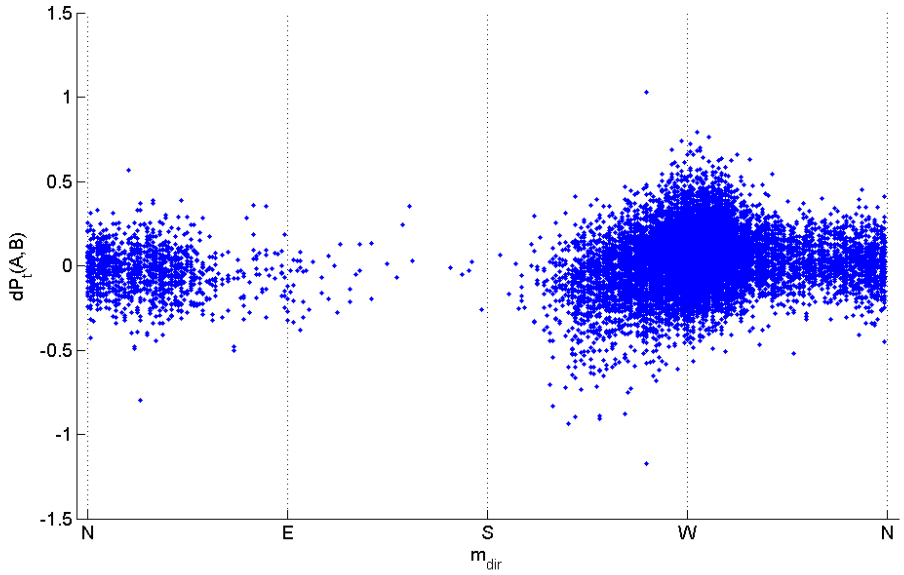

Figure 7.6: A scatter diagram of m_{dir} against the proportional difference between power measurements at buoys A and B, $dP_t(A, B)$.

7.3 Multivariate least squares analysis (MVLS) - results

The multivariate least squares analysis described in section 7.1.1 was used to compare H_{m0} values measured across the array, estimated every 30 minutes using the standard processing methods described in section 6.8. Data from the first deployment were analysed separately to that from the second to account for the change in positions. All results are presented with the buoy identification followed by its position within the array in brackets. Thus, B(W) relates to data captured by buoy B when operating in the westerly position within the array.

The first deployment comprises 984 data from the more active winter period (fig. 6.3), and was predominantly characterised by waves from a westerly sector, with larger average wave period, T_{m02} and a larger average directional spreading than the second deployment (fig. 7.7 and tab. 7.2). Results from the second deployment are based on 13689 data and cover a wide range of conditions. Due to the larger data set, and the variety of conditions measured, results from the second deployment are therefore considered more robust and are presented first in this section.

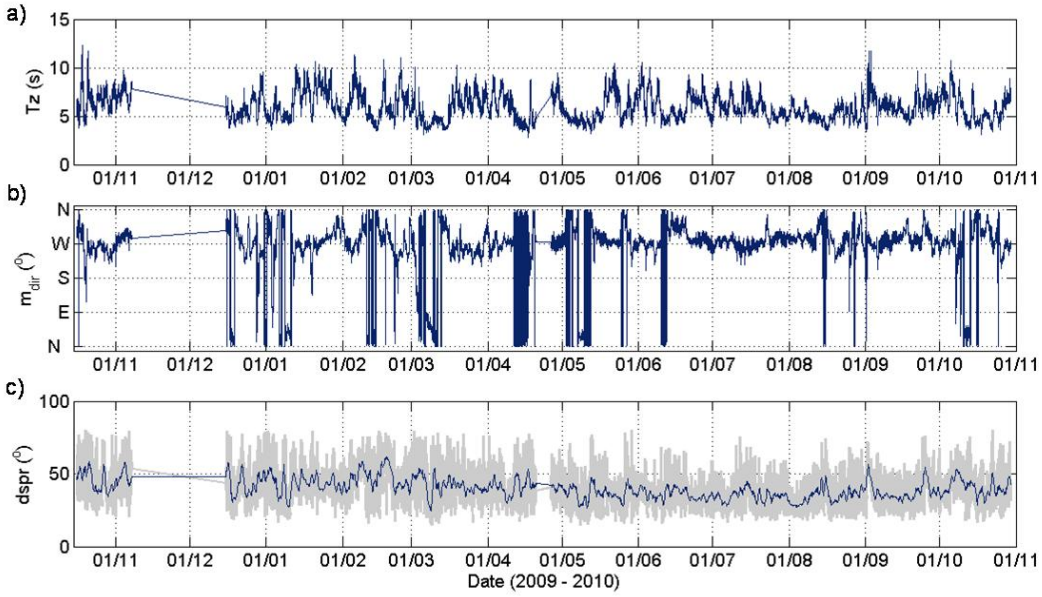

Figure 7.7: A time-series of co-variates used in the MVLS analysis. a) Mean zero-crossing period (calculated using zero-crossing analysis), T_z b) Mean direction, m_{dir} and c) Directional spreading (also shown is a moving average over 24 consecutive values).

Table 7.2: Mean values for co-variates measured during the two deployments at the northerly position in the array.

Parameter	Oct - Nov 2009	Dec 2009 - Oct 2010
$T_z(s)$	6.83	5.91
$Dspr(^o)$	46.09	39.58

7.3.1 MVLS using data from the second deployment

The model calculated for buoys A(E) and B(W) between December 2009 and October 2010 was,

$$\hat{H}_{m0A} = 0.158 + 0.983H_{m0B} - 0.020T_z - 0.026V - 0.0001m_{dir} + 0.0002spr_{Tp} \quad (7.3.1)$$

The sum of squares and appropriate degrees of freedom for this regression are given in table 7.3, and the R^2 value, which was 0.986. In order to support the assumption that the error term in the model will be Gaussian distributed (required for the hypothesis testing described in section 7.1.1), a Kolmogorov-Smirnov test applied to the empirical distribution of the weighted residuals. This test found a KS statistic of 0.0323, with associated probability, 0.26. This result supports the hypothesis that the error term in the model follows a Gaussian distribution (fig. 7.8).

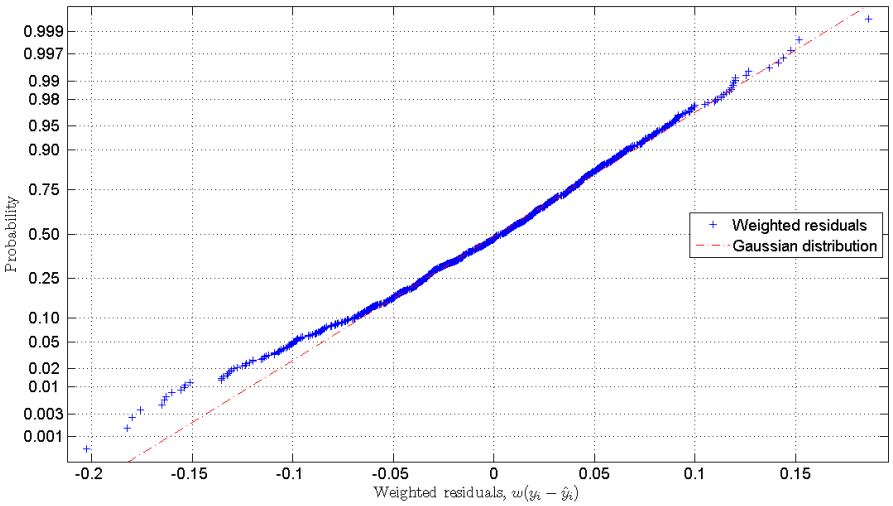

Figure 7.8: The empirical cumulative distribution of the weighted residuals from the multivariate least-squares analysis of significant wave heights measured at buoy A(E) and buoy B(W) between Dec 2009 and Oct 2010. The axes are defined such that a Gaussian distribution will fall as a straight line, as shown by the red line.

The F statistic, as defined in equation (7.1.7), was calculated for this regression as 3110345.76, for which the probability of being drawn from the F distribution with the relevant degrees of freedom, $p < 0.01$, therefore the correlation between the dependent variable and co-variates was deemed to be significant.

The t statistics were calculated using equation (7.1.12) for all coefficients, and these are also given in table 7.3, along with the probability that each was drawn from the t distribution with appropriate degrees of freedom, $p(t > t_k)$. Despite the small values for

some β coefficients, these probabilities deemed the tidal conditions, wave period (chosen to represent wave-length) and the wave direction coefficients to be significantly different from 0, and therefore, have an effect on the differences between H_{m0} measured at the buoys A(E) and B(W), that is significant to a 99% confidence interval. The β coefficient representing directional spreading was not found to be significantly different from 0 ($t = -1.086, p = 0.277$), which suggests that correlation between measurements at the two buoys is not having a significant effect on the differences.

Table 7.3: Results of multivariate least-squares analysis of significant wave heights measured at buoy A(E) and buoy B(W) between Dec 2009 and Oct 2010.

ANOVA	SS	dof	MSE	F_{reg} Ratio	$P(F > F_{reg})$
Regression	2300.91	6	460.18	3110345.76	0
Error	2.02	13683	0.00		
Total	2302.94	13689			
R^2	0.986				
Coefficient	$\hat{\beta}$	se of$\hat{\beta}$	t-value	$P(t > t_k)$	
Mean	0.158	0.010	15.739	**0.000**	
$H_{m0}(A)$	0.983	0.003	344.092	**0.000**	
T_z	-0.020	0.002	-11.404	**0.000**	
V	-0.026	0.003	-7.576	**0.000**	
m_{dir}	-0.00010	0.000	-5.941	**0.000**	
$Dspr$	-0.00020	0.000	-1.086	0.277	

The results of the MVLS model show some spread about an ordinary least squares (OLS) line of best fit between the H_{m0} values at buoy A and those at buoy B (fig. 7.9) There is also some deviation between these two models evident in the larger wave heights, with smaller wave heights predicted at buoy A by the MVLS, than the OLS. These effects are due to the influence of the covariates in the model. Analysis of the residuals shows that the spread of weighted residuals about the mean is approximately constant for varying values of H_{m0} (fig. 7.10). This demonstrates that each point is having an equal significance in the model and is a key indicator of a suitable weighting term.

The magnitude of the coefficient β_1 suggests that wave heights measured at buoy A(E) are less than those measured at buoy B(W), which contradicts results from the previous section. This result may be related to bias caused by equal error terms, as discussed in

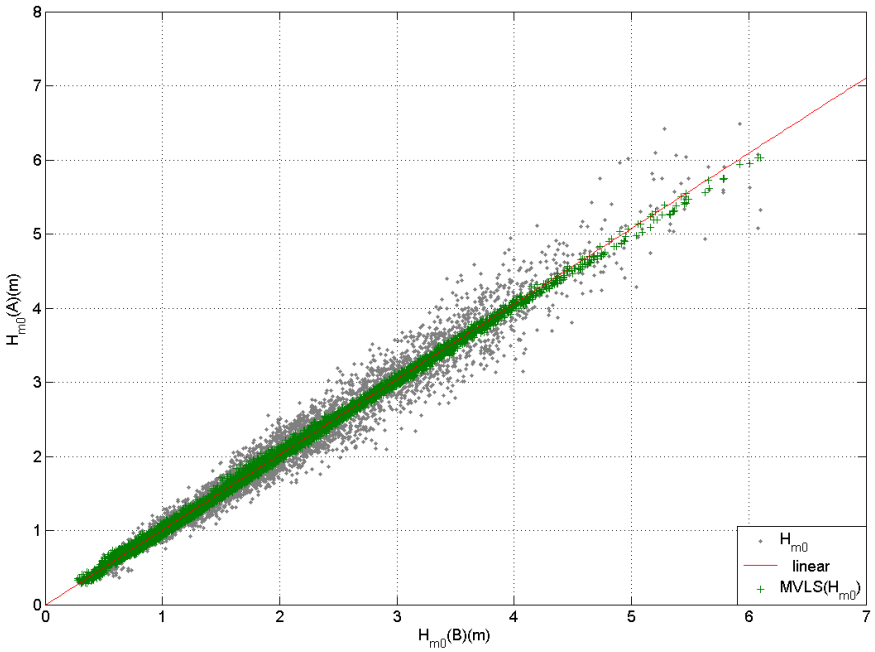

Figure 7.9: Wave heights at buoy A against wave heights at buoy B, with the results of the MVLS analysis shown as green crosses, and a simple OLS linear model for comparison.

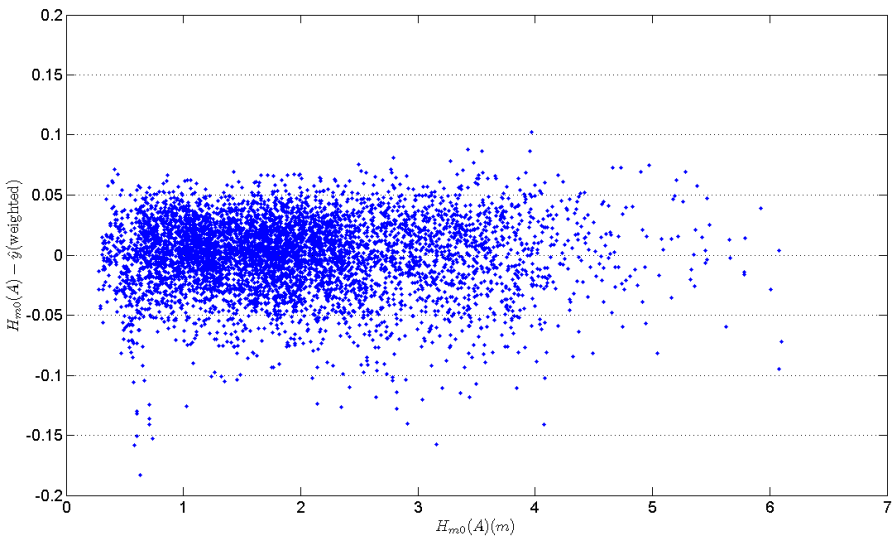

Figure 7.10: The weighted residuals for the multivariate regression, where the weights are defined as the inverse of the theoretical standard deviation, $wX = 1/\sigma_x$.

section 2.3. Michael (1983) demonstrates that the magnitude of this bias can be examined by exchanging which parameter is defined on the x axis and which on the y axis. With this aim, the regression for buoys A(E) and B(W) was reversed, to regress wave heights from buoy A(E) onto those from buoy B(W) and the results are compared in table 7.4. The reversed model indicates that the same parameters are having a significant effect on the model, which supports the initial result. The β co-efficients of the principal variable, (H_{m0}) were 0.98 and 0.97 for the initial and the reversed models respectively. This indicates in both cases that the wave height at the independent variable is less than that at the dependent variable. This contradiction is representative of variability in the β coefficients caused by the equal error terms, which means that it is not possible to predict whether H_{m0} will be larger at either buoy.

Table 7.4: The results of a weighted multivariate least-squares comparison of significant wave heights measured at buoy B(W) and buoy A(E) between Dec 2009 and Oct 2010.

	A:B		B:A	
Coefficient	β	p	β	p
Mean	0.158	**0.000**	-0.156	**0.000**
$H_{m0}(B, D)$	0.983	**0.000**	0.968	**0.000**
T_z	-0.020	**0.000**	0.036	**0.000**
V	-0.026	**0.000**	0.028	**0.000**
m_{dir}	-0.00010	**0.000**	0.000	**0.000**
$Dspr$	-0.00020	0.277	0.000	0.431

To investigate the correlation of measurements at buoy B(W) to the other buoys within the array, the MVLS analysis was repeated. This can be considered equivalent to estimating a model that predicts the wave height at buoy B using data captured by the other buoys in the array, and covariate data sets. In both cases, the probabilities associated with the F statistic were below 0.01 and therefore, the regression models are considered significant. β coefficients for each parameter and the associated probability that they are equal to zero, are given in table 7.5.

The wave period, and the mean direction are both found to have a significant effect on all three pairs of data. This supports the result that interaction of waves with bathymetry is causing differences between wave heights measured by the buoys. The results for tidal conditions, V, also indicate an effect, although the β coefficient was significant for the

Table 7.5: The results of weighted multivariate least-squares comparisons of significant wave heights between Dec 2009 and Oct 2010.

	A(E):B(W)		C(S):B(W)		D(N):B(W)	
Coefficient	$\hat{\beta}$	$p(t > t_k)$	$\hat{\beta}$	$p(t > t_k)$	$\hat{\beta}$	$p(t > t_k)$
Mean	0.158	**0.000**	0.09	**0.000**	0.023	**0.010**
$H_{m0}(B, D)$	0.983	**0.000**	0.97	**0.000**	0.991	**0.000**
T_z	-0.020	**0.000**	-0.01	**0.003**	0.012	**0.000**
V	-0.026	**0.000**	-0.01	0.026	-0.010	**0.001**
m_{dir}	-0.00010	**0.000**	0.00	**0.000**	0.000	**0.001**
$Dspr$	-0.00020	0.277	0.00	**0.002**	0.000	0.324

pairing D(N) and B(W) ($t = -9.6, p = 0.001$), but not significant for the pairing of B(W) and C(S) ($t_k = 0.4, p = 0.026$). Directional spreading was found to have a significant effect only on the pairing of B(W) and C(S) ($t_k = -3.04, p = 0.0024$). Because this parameter governs the level of correlation between time-series, this result implies that correlation between time-series is affecting the differences between measured parameters for these locations.

MVLS using data from the first deployment

The position of the buoys was changed upon re-deployment, and in order to represent this, different pairs of buoys were used when analysing data from the first deployments, to ensure that the geographical positions in the array were the same. In this case, the model was estimated as,

$$\hat{H}_{m0C} = 0.241 + 0.959H_{m0D} - 0.019T_z - 0.025V + 0.002spr_{Tp} \tag{7.3.2}$$

The full results of the regression of H_{m0} values between buoys C(E) and D(W) during the first deployment (which is equivalent to A and B in the second deployment) are provided in table 7.6. The F_{reg} ratio and associated probability (0) indicates that the model is describing the variability observed in the data set and the R^2 value indicates a smaller spread from the model of 0.9%. Again, the MVLS model points are spread around the OLS linear regression model for H_{m0} values between buoys C(E) and D(W), due to the co-variates, although no consistent deviation of the model is observed.

As with the results from the second deployment, both the mean zero-crossing wave period,

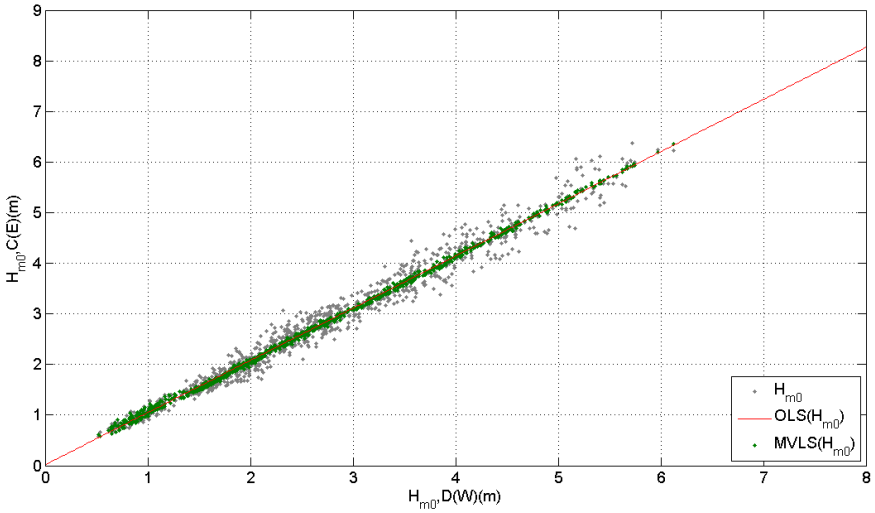

Figure 7.11: Wave heights at buoy C against wave heights at buoy D, with the results of the MVLS analysis shown as green crosses, and a simple OLS linear model for comparison.

Table 7.6: Results of a weighted multivariate least-squares comparison of significant wave heights measured at buoy C(E) and buoy D(W) between Oct 2009 and Nov 2009.

ANOVA	SS	$d.o.f.$	MSE	F Ratio	$P(F > F_{reg})$
Regression	655.98	6	131.2	147139.01	0
Error	0.87	978	0		
Total	656.86	984			
R^2	0.991				
Coefficient	$\hat{\beta}$	$Seof\hat{\beta}$	t-value	$p(t > t_k)$	
Mean	0.241	0.035	6.896	**0.000**	
$H_{m0}(B, D)$	0.959	0.005	197.508	**0.000**	
T_z	-0.019	0.004	-4.210	**0.000**	
V	-0.025	0.009	-2.730	**0.007**	
m_{dir}	-0.0003	0.000	-1.928	0.054	
$Dspr$	-0.0016	0.000	-3.737	**0.000**	

T_z, and the tidal conditions, V, were seen to have a significant effect. However, the mean direction (m_{dir}), was not identified as a significant factor, whilst for this deployment, directional spreading, d_{spr}, was deemed significant for this pair of locations (table 7.6).

Table 7.7: Results of weighted multivariate least-squares comparisons of significant wave heights measured between Oct 2009 and Nov 2009.

	C(E):D(W)		A(S):D(W)		B(N):D(W)	
Coefficient	$\hat{\beta}$	$p(t > t_k)$	$\hat{\beta}$	$p(t > t_k)$	$\hat{\beta}$	$p(t > t_k)$
Mean	0.241	**0.000**	0.285	**0.000**	0.007	0.852
$H_{m0}(B, D)$	0.959	**0.000**	0.959	**0.000**	1.001	**0.000**
T_z	-0.019	**0.000**	-0.030	**0.000**	0.003	0.511
V	-0.025	**0.007**	-0.013	0.239	-0.011	0.266
m_{dir}	-0.0003	0.054	-0.001	0.027	0.000	0.532
$Dspr$	-0.002	**0.000**	-0.001	0.323	0.000	0.832

When the analysis was repeated comparing data at buoy D(W) to the remaining two other buoys in the array (tab. 7.7), fewer parameters were observed to be significant. In particular, the pair of B(N) and D(W) did not show any co-variates to be significant. This suggests that all the differences between these two buoys are described by the theoretical error term, and is perhaps indicative of the lower sensitivity of the method for this shorter data set.

Wave period (T_z) was found to be significant for the pair C(E) and D(W) ($t = -4.21, p = 0$), as well as A(S) and D(W) ($t = -5.899, p = 0.00$), whilst tidal conditions (V) was only found to be significant for one pair (C(E):D(W)), and mean direction ,m_{dir}, was not deemed significant for any pair. These results provide limited support to the conclusion that tidal flow and interaction with the local bathymetry are affecting the measured wave heights around the array although the results are less conclusive than those from the longer second deployment.

Directional spreading was only seen to be significant for buoys C(E) and D(W) ($t = -3.737, p = 0$), which suggests that correlation between the time-series was only having an effect across the longest separation in the array. This separation is parallel to the principal direction of propagation during this deployment (fig. 7.7), which would increase correlations (see sec. 2.4), although it is not the same pair for which this parameter was deemed significant in the second deployment.

When the regression for buoys D and C was reversed, the β coefficients for H_{m0} from both models predict wave heights at buoy C(E) to be larger than those at buoy D(W) (tab. 7.8). This result is in agreement with results from section 7.2 and a contradiction between models such as that observed in results from the second deployment was not evident. The reversed model finds the same parameters to be having a significant effect as the initial model (tab. 7.6) which again builds confidence in this result.

Table 7.8: Results of a weighted multivariate least-squares comparison of significant wave heights measured at buoy B(W) and buoy A(E) between Oct 2009 and Nov 2009.

	C(E):D(W)		D(W):C(E)	
Coefficient	β	$p(t > t_k)$	β	$p(t > t_k)$
Mean	0.241	**0.000**	-0.260	0.000
$H_{m0}(B, D)$	0.959	**0.000**	1.015	**0.000**
T_z	-0.019	**0.000**	0.029	**0.000**
V	-0.025	**0.007**	0.027	**0.005**
m_{dir}	-0.0003	0.054	0.000	0.436
$Dspr$	-0.002	**0.000**	0.002	**0.000**

7.4 Discussion

Initial analysis of the time-series of proportional differences (or bias) (fig. 7.4) demonstrated a weak correlation with the wave period, with longer wave periods attributed to greater differences between measurements. Also, differences were seen to vary with the direction of propagation (fig. 7.6). These findings are indicative of a spatial distribution of waves caused by interaction with the local bathymetry.

Multivariate regression (MVLS), described in section 7.1.1 was used to identify physical conditions that were having a significant effect on the relationship between measurements across the array. Co-variates used were described and shown in figure 7.7. The analysis produced a linear regression model for each pair of data sets analysed, with a coefficient to represent each covariate used. Weighting was applied to account for the varying sampling error in H_{m0} values, and was defined as the inverse error variance, as described in section 2.2, and calculated using equation (2.2.8). This formulation was effective, and a constant error variance with increasing wave height was observed for all models (fig. 7.10), with errors exhibiting a Gaussian distribution (fig. 7.8). For all models, the F test deemed the regression model to be significant, indicating a correlation between wave measurements separated by $500m$, as would be expected.

The β co-efficients calculated to represent the mean zero-crossing wave period, T_z, were found to be significant in the regression model for all but one of the pairs of data analysed. Wave propagation direction, m_{dir}, was also shown to have a significant influence in all models for the second deployment, which exhibits a greater variation of directions than the first deployment. These results contribute to the evidence that spatial properties of the wave field are governed by the interaction of incident waves with the local bathymetry, and support the observations from the direct analysis of proportional differences (sec. 7.2).

Tidal conditions were included in the model, and results from both deployments showed them to have a significant effect, albeit not consistent for all pairs. This presents limited evidence that tidal conditions affect the spatial distribution of wave heights across the array. However, the direct effect of the tidal conditions may not be on the wave heights, rather, it changes other co-variates, T_z and m_{dir} (see section 6). Furthermore, the rising and falling of the tide, which has a range of 3 - 7m at this site will affect the interaction of surface waves and the local bathymetry, which is the principal effect highlighted by T_z and m_{dir} on the correlation between wave heights across the array. These indirect effects of the tide, or co-linearity between covariates, will not be properly resolved by the MVLS

technique, and results may not represent the full effect of the tidal conditions.

Section 2.4 examined correlation between wave records for the site and demonstrated that some level of correlation would be expected for separations parallel to the direction of wave propagation. With the longest fetch to the West-South West sector, it would be predicted that directional spreading would have a greater effect on regression models for pairs of buoys in the east and west positions, in the south and east positions and in the north and west positions (fig. 3.1).

The results from the MVLS analysis do not reflect the predicted correlations. In the second deployment, directional spreading is shown to have a significant influence for the S:W separation, which would arise principally from wave systems propagating from the north-northwest. In results from the first deployment, only the pair of data collected at the western and eastern points on the array demonstrated a significant influence of directional spreading. The wave propagation direction during this period can be seen to be predominantly East-West. Thus, these results may be indicative of correlation between the measured time-series causing coherence in the error terms between buoys and therefore affecting the measured differences (as discussed in section 2.4).

This analysis has demonstrated the application of a multivariate regression technique to the comparison of spatially separated wave data. Increased sensitivity was achieved through using a variable error term, which was seen to effectively account for the heteroscedasticity of wave parameters. The results have indicated that the wave period (and therefore the wave length), and the wave direction are causing deterministic differences between the wave heights measured across the array. Further analysis would be required to validate the MVLS methodology for developing predictive tools, as in certain cases, values of the β parameters appear to contradict each other (tabs. 7.3 and 7.4).

Chapter 8

Quantifying variability in wave measurements

8.1 Analysis of differences as variability

The previous chapter has examined differences between simultaneous measurements in terms of bias and deterministic differences. In doing so, variability, or the error term, in the wave parameters was assumed to be equal to that predicted by sampling theory, as described in section 2.2. This chapter uses the same data to compare the magnitude of observed differences in key wave parameters measured across the array, to errors predicted by sampling theory. The outcome quantifies the combined contribution of deterministic differences, sampling errors and other potential sources of error affecting the measured wave data, and demonstrates the suitability of using sampling theory alone to predict errors in key wave parameters measured by wave buoys.

A range of statistical methods for estimating the errors in an empirical data set can be found in the literature, and these have been reviewed in section 2.3. Mackay (2009) estimates the coefficient of variation (*c.o.v.*) for significant wave height values from the proportional differences, dX, using a coefficient derived from numerical simulation. This methodology can be applied to any paired data set, and is used in this research to allow direct comparison with the published results.

Sampling variability in spectral parameters will vary with the spectral shape, and not simply the magnitude of a parameter (sec. 2.2). This is not accounted for when calculating the mean *c.o.v.* for a data set, and a swell-dominated site would be expected to exhibit a larger *c.o.v.* value than a site dominated by wind-seas. The Z statistic, as applied to the comparison of simultaneous wave data in Sova and Wyatt (1995), compares measured differences with the predicted sampling variability. As such, it provides a measure of accuracy that is not dependent on the sea conditions during the data set, but is only applicable to pairs of data.

All of the reviewed methodologies assume no coherence between error terms, and section 2.3 demonstrated that some coherence can be expected over these separations, dependent on the directional spreading, spectral composition of the wave field, and alignment of the measurement buoys to the direction of propagation of incident waves. Therefore, where these methodologies are applied, it should be with consideration of the effect of dependence between error terms.

8.2 Methodology

The availability of four data sets captured from identical sensors operating simultaneously (and temporally synchronised) offers an opportunity to examine the variability of a given parameter directly from the data. A technique was developed, based on the same principals developed in Sova and Wyatt (1995), described in section 2.3, which compares instantaneous differences to the expected sampling distribution.

For each time step, four values of a given parameter, x are available, X_A, X_B, X_C and X_D. The standard deviation of these values can be used as an estimator of σ for each record, $\hat{\sigma} = \sigma(X_A, X_B, X_C, X_D)$. When calculated from a sample of 4 values from a population, $\hat{\sigma}$ is a biased estimator and $E[\hat{\sigma}] < E[\sigma]$[1], which will affect the direct calculation of the *c.o.v.*

However, the variance of a small sample, $\hat{\sigma}^2$, is an unbiased estimation of σ^2 where,

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum \left(X - \bar{X}\right)^2 \tag{8.2.1}$$

Therefore, for direct comparison with theoretical sampling error, comparisons of variance, using $\hat{\sigma}^2(X)$ are preferable to using the standard deviation.

Analysis of variance (ANOVA) techniques are widely used in statistics to evaluate the significance of observed differences by comparing the variance within each sample to the total variance across all[2]. If they are deemed significantly different, it can be concluded that the samples are drawn from populations with different means and/or variances (Fowler et al., 1998).

Following classic ANOVA methodology, the null hypothesis is that the four samples are drawn from the same population, and a test statistic,F, was calculated as,

$$F = \frac{\text{Variance between samples}}{\text{Variance within samples}} \tag{8.2.2}$$

$$F_{\sigma^2}(X) = \hat{\sigma}^2(X)/\sigma^2(X) \tag{8.2.3}$$

where X is a single value of a given wave parameter, which is taken in this case to represent the mean of a sample. Variance between the samples, $\hat{\sigma}^2$, represents the variance between the data sets, and was estimated using equation (8.2.1). Variance within the samples, σ^2, is the mean theoretical sampling variance, and was calculated using the methodology described in section 2.2.1. A hypothesis test was defined with the null hypothesis that the statistic, F_{σ^2} was drawn from an F distribution with the correct degrees of freedom.

[1] $E[x]$ represents the expected value of a given parameter, x.

[2] An example fo this form of analysis can be found in the previous chapter.

Where the null hypothesis was rejected, it was concluded that the observed differences cannot be described by sampling theory.

Degrees of freedom define the F distribution and can be calculated as,

$$dof = N - N_p \tag{8.2.4}$$

where N is the number of samples used in the estimation and N_P is the number of parameters estimated.

$dofB$ is the degrees of freedom in $\hat{\sigma}^2(X)$ (variance between samples). For this parameter, $N = 4$ and $N_P = 1$, giving $dofB = 3$. $dofW$ represents the degrees of freedom in $\sigma^2(X)$ (variance within samples). In this case, the degrees of freedom will be the sum of those in each parameter, which can be calculated using equations (2.2.12) and (2.2.13) (see section 2.2.1). This gives a value of $dofW$ for every record, which are dependent on the spectral shape.

For each record, the test F distribution was defined as $F(3, dofW)$. Where F_{σ^2} exceeds this distribution at a given critical level, α, noted as $F_{\sigma^2} > F_{(\alpha,3,dofW)}$, the null hypothesis was rejected and the values of X, for that record, were deemed significantly different.

To draw conclusions about the entire data set (not just an individual record), it would be advantageous to compare the frequency distribution of all values of F_{σ^2} to the F probability distribution. This would provide a probability that values of F_{σ^2} are drawn from a particular F distribution. However, the varying degrees of freedom mean that each F_{σ^2} value is drawn from a different distribution, which is dependent on the degrees of freedom of the spectrum, $dofW$. Therefore, results are stated as the proportion of records within a data set for which the null hypothesis is rejected. Taking the example $\alpha = 0.9$, 10% of records within a data set (where sampling error is equal to measured differences) would be expected to reject the null hypothesis. If the null hypothesis is rejected for more records than predicted, it is an indication that observed differences cannot be explained by sampling theory alone. In order to present statistical significance of these results, confidence intervals of the proportion of F_{σ^2} values that exceed the threshold set by the relevant F distribution were derived using simulation.

8.2.1 Confidence intervals for the number of F_{σ^2} values exceeding a critical probability limit

Confidence intervals for the fraction of records for which the null hypothesis was rejected, were derived using simulation. Random numbers were simulated from an F distribution,

and the proportion of them that exceeded a given threshold, r, was noted. The amount of random numbers generated was chosen to match the data set under consideration, and the degrees of freedom were chosen as 3 and 63, to reflect the data under consideration (although their value does not affect the outcome of the analysis). This procedure was repeated 10000 times, to give 10000 separate values representing the proportion of records for which the null hypothesis would be rejected despite it being true, R, where $R = r_1, r_2, ..., r_n$. 99% confidence limits are set according to the standard deviation of R, assuming that is follows a Gaussian distribution, $R = N(\mu(R), \sigma(R))$, as,

$$R - 2.96\sigma(R) < R < R + 2.96\sigma(R) \tag{8.2.5}$$

8.3 Results

The coefficient of variation was calculated for the measured data, following the methodology described in Mackay (2009) (tab. 8.1). In addition to the results for the EMEC site published in Mackay (2009), Barrett et al. (2009) calculated the RMS differences between concurrent measurements for three separate data sets and the results were used to estimate the *c.o.v.* values. These results are also shown in table 8.1. Notably, *c.o.v.* values estimated for the PRIMaRE buoy array closely match those calculated for the EMEC site separated over a similar distance.

Table 8.1: Coefficient of variation (*c.o.v.*) values estimated from the whole data set.

	c.o.v.(%)		
Data set	H_{m0}	T_{m02}	P_t
EMEC 1500m Mackay (2009)	6.2	4.5	-
EMEC 1500m Barrett et al. (2009)	7.8	-	-
EMEC 500m Barrett et al. (2009)	5.2	-	-
Galway 200m Barrett et al. (2009)	3.4	-	-
PRIMaRE A-B 700m	5.3	3.2	12.5
PRIMaRE A-C 500m	5.2	3.3	12.5
PRIMaRE A-D 500m	5.4	3.3	13.0

8.3.1 Theoretical variability in key wave parameters

The variability, or the random error term, associated to spectrally derived parameters due to sampling theory has been discussed above, in section 2.2.1. Equation (2.2.8) defines variability in terms of the error variance, $\sigma^2(H_{m0})$, which is the expected variance of a population of H_{m0} calculated from the spectra of an infinite number of wave records captured during the same sea state. Theoretical error variances of H_{m0}, T_{m02} and P_t data were calculated for the measured data using equations, (2.2.8), (2.2.9) and (2.2.11) respectively, to give the expected sampling variability in measured data (tab. 8.2). These were then used to calculate the theoretical *c.o.v.* for comparison with values for the measured data, given in table 8.1. The results of this analysis demonstrate that measured variability is greater than predicted by sampling theory, with the largest proportional discrepancy between measured and predicted *c.o.v.* values occuring for the parameter T_{m02} ($c.o.v._{\cdot theory} = 1.8\%$, $c.o.v._{\cdot meas} = 3.3\%$) (tab. 8.2).

Table 8.2: *c.o.v.* values calculated from measured and theoretical variability, using data from buoys measuring in the Northerly position.

	H_{m0}	T_{m0}	P_t
$c.o.v._{\cdot theory}$	3.3%	1.8%	7.9%
$\frac{c.o.v._{\cdot meas}}{c.o.v._{\cdot theory}}$	1.57	1.83	1.58

The variance of spectral parameters depends on the integral of the spectrum squared (eqn. (2.2.7)); parameters derived from a peaked spectrum characterised by values over a small range of frequencies will exhibit higher variability than those from a wide spectrum, with the same area[3]. Thus, narrow band swell systems exhibit a greater proportional sampling variability term than wide band systems and it is at a maximum for a monochromatic wave.

The error variances for parameters H_{m0}, T_{m02} and P_T calculated above, were compared to the variance of measured data captured simultaneously within the array. Figure 8.1 plots the measured data against the associated theoretical error variance for each parameter. Pierson Moskowitz spectra were calculated using equation (4.9.1) with $f_p = 0.05, 0.1, 0.15...0.5$ and $\alpha = 0.0081$. Parameters and associated theoretical variance were then estimated from the spectra to demonstrate the expected variance associated with a fully-developed sea, which are also shown on figure 8.1.

[3]Two spectra with the same area will have the same significant wave height (H_{m0}).

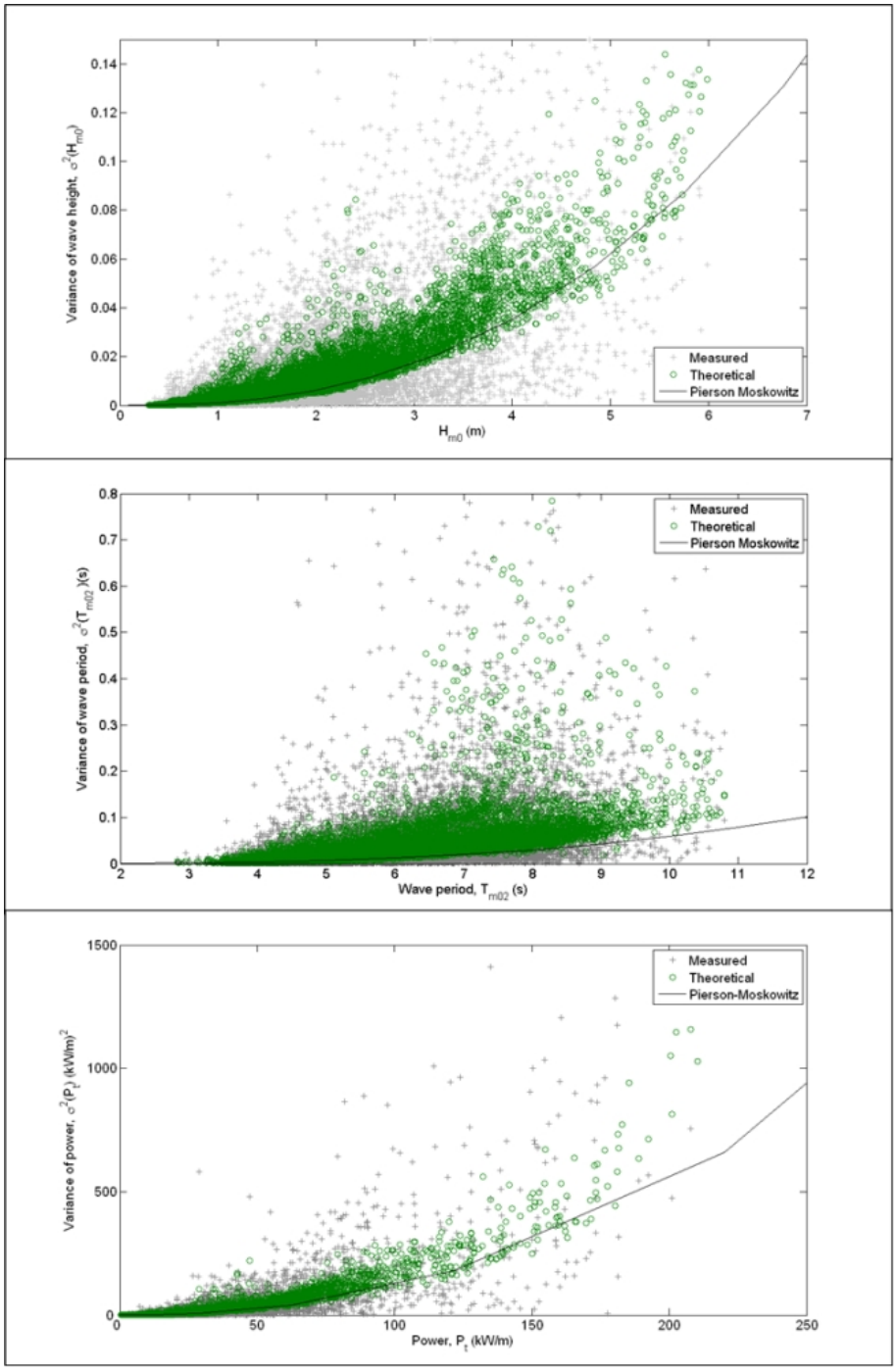

Figure 8.1: A comparison of measured parameters, H_{m0}, T_z and P_t, and the associated error variance, ($\sigma^2(X)$, predicted by sampling theory for the measured spectra, and for Pierson-Moskowitz spectra. Also shown is the measured variance, $\hat{\sigma}^2(X)$.

The lowest sampling variabilities relative to parameter value in the measured data are aligned with the values for Pierson-Moskowitz spectra, and therefore associated with wide-band fully developed seas, which is in agreement with the theory. For each parameter, σ^2 values greater than those for a fully developed sea are observed, which indicates the influence of peaked spectra. The range of $\sigma^2(T_{m02})$ for a given value of T_{m02} is larger than the range observed for $\sigma^2(H_{m0})$ and $\sigma^2(P_t)$ showing that a higher variability of this parameter may be expected under certain wave conditions.

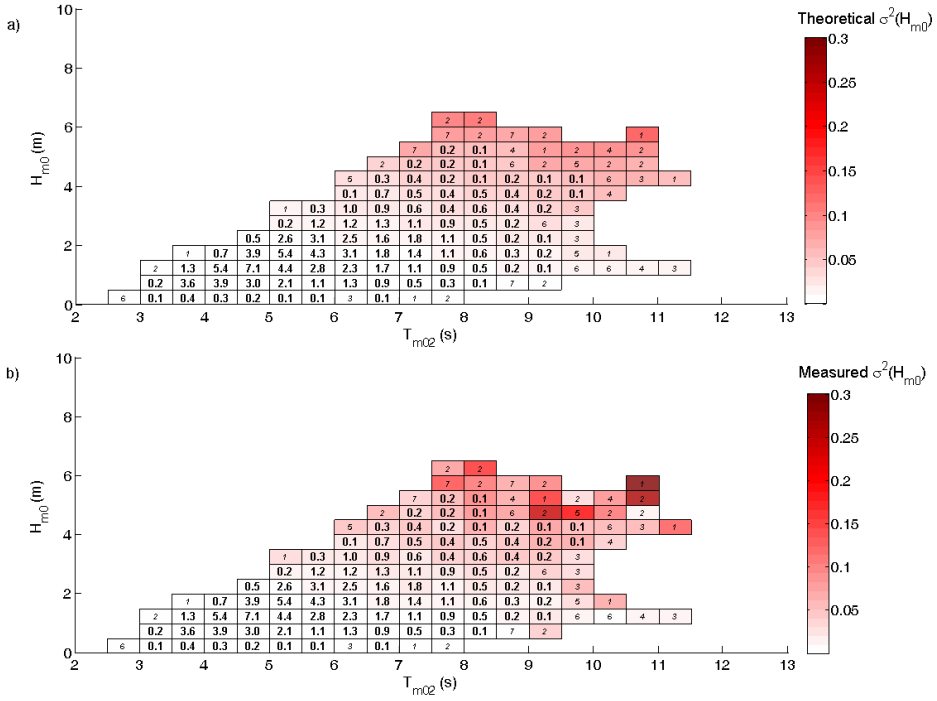

Figure 8.2: H-T scatter diagrams for all records where error-free records were available from all 4 buoys. The text values represent the percentage of records, or the total number of records (values in italics), occurring with H_{m0} and T_{m02} values corresponding to each cell, and the colour overlay represents a) the average theoretical sampling variance of H_{m0} values for records within that cell, and b) the average measured variance of simultaneous H_{m0} values within each cell.

8.3.2 Variability in the measured data

The variance in measured values was calculated for all records where simultaneous measurements were available from all 4 buoys, using equation (8.2.1). These values are also plotted on figures 8.1a-c, although for clarity, the y axis were limited, and the largest values are not shown, (max $\hat{\sigma}^2(H_{m0}) = 0.4$m, max $\hat{\sigma}^2(T_{m02}) = 2.8$s and max $\hat{\sigma}^2(P_t) = 2.3 \times 10^6$kW). What can be observed is the large amount of spread in the mea-

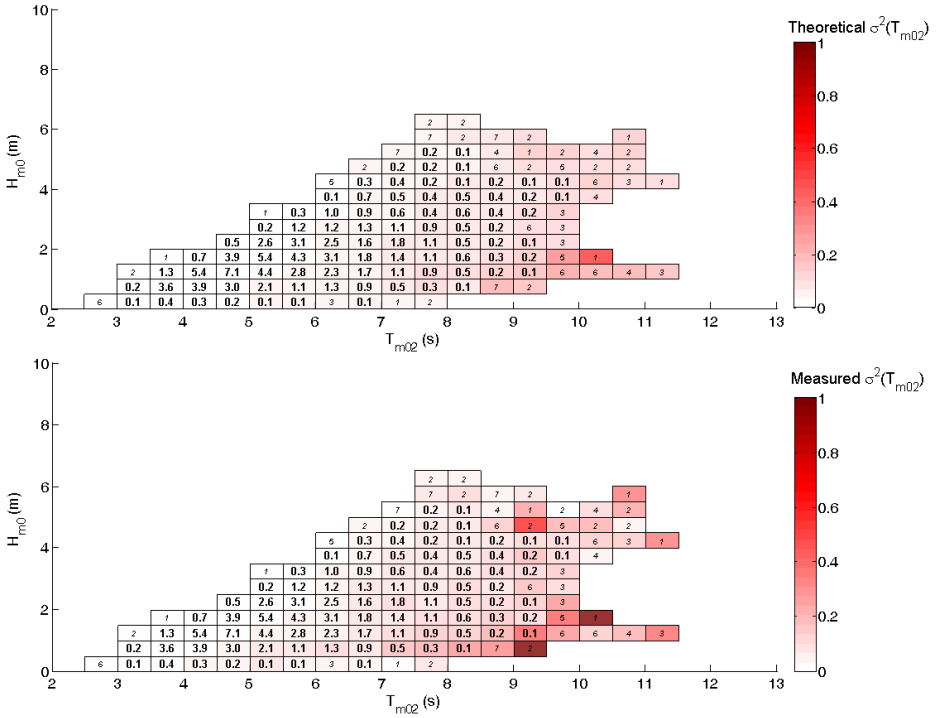

Figure 8.3: H-T scatter diagrams for all records where error-free records were available from all 4 buoys. The text values represent the percentage of records, or the total number of records (values in italics), occurring with H_{m0} and T_{m02} values corresponding to each cell, and the colour overlay represents a) the average theoretical sampling variance of T_{m02} values for records within that cell (eqn. (2.2.9)), and b) the average measured variance of simultaneous T_{m02} values (eqn. (8.2.1)) within each cell.

sured values due to the high variability in $\hat{\sigma}^2$, as each value is derived from a sample of only 4 data.

When displayed as a colour overlay on an $H - T$ scatter diagram, the average variance of parameter values for records within each cell show a distribution with wave conditions (figs. 8.2b, 8.3b, and 8.4b). Theoretical variability is seen to increase for larger wave heights, and when wave period is large relative to wave height. For all parameters, the distribution of measured variance with wave height and period reflects that predicted by the theoretical values.

Were the difference between simultaneous parameters explained by sampling theory, F_{σ^2} values in 10% of records would be expected to exceed the 90% critical limit of the distribution $F(3, dofW)$. For all months in the study period, the percentage of F_{σ^2} values that exceed the tabulated values for this critical limit was greater than 10% (tabs. 8.3 - 8.5 and fig. 8.5), and in all cases, was outside the 99% confidence limits calcuated in section 8.2.1. Therefore, calculated sampling variability is not sufficient to explain observed differences

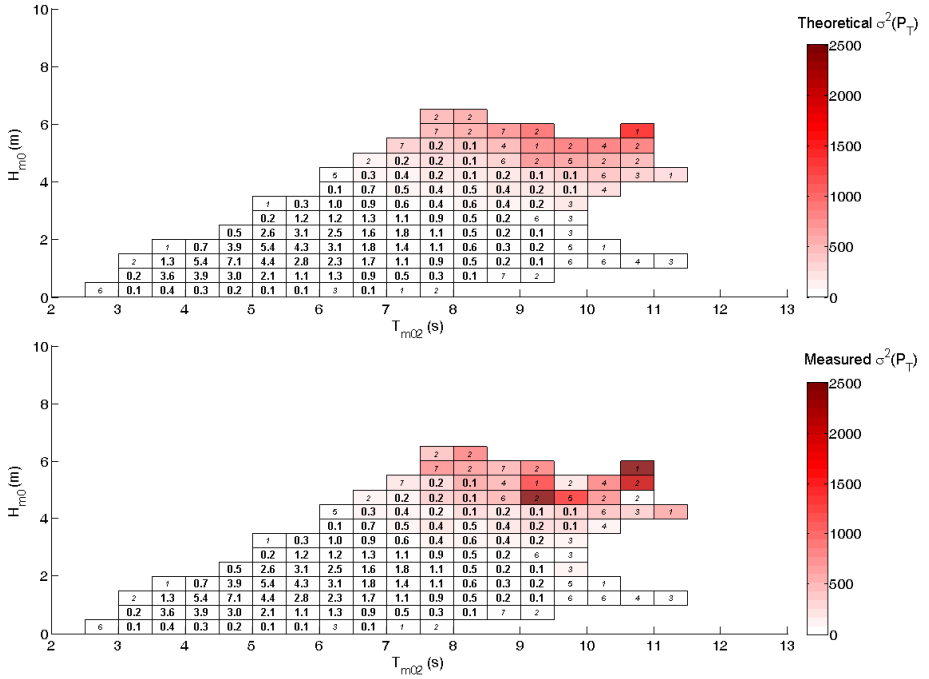

Figure 8.4: H-T scatter diagrams for all records where error-free records were available from all 4 buoys. The text values represent the percentage of records, or the total number of records (values in italics), occurring with H_{m0} and T_{m02} values corresponding to each cell, and the colour overlay represents a) the average theoretical sampling variance of P_t values for records within that cell, and b) the average measured variance of simultaneous P_t values within each cell.

between simultaneous parameters. For H_{m0} values, the highest proportion of records exceeding the threshold was 0.19 (19%) during December 2009, where data from only part of the month is available, and the minimum of 0.12 (12%) for the month of August 2010. The monthly distribution is similar in all parameters, with greater variability during the more active winter months, and the minimum occuring between June and August.

Variability in the spectral moments

Section 4.5 demonstrated how the analysis of spectral moments can be used to highlight a frequency dependency when comparing data sets. This methodology was applied to this analysis, and the variability between simultaneous measurements of spectral moments of the order -1 to 2 was examined (tab. 8.6). Mean variability increases for lower spectral moments, which suggests that variability between the measurements is greatest in the lower frequencies. The lowest value for $F_{\sigma^2}(m_n)$ is not however for the highest spectral moment. This indicates that there is also increased variability in the higher frequencies.

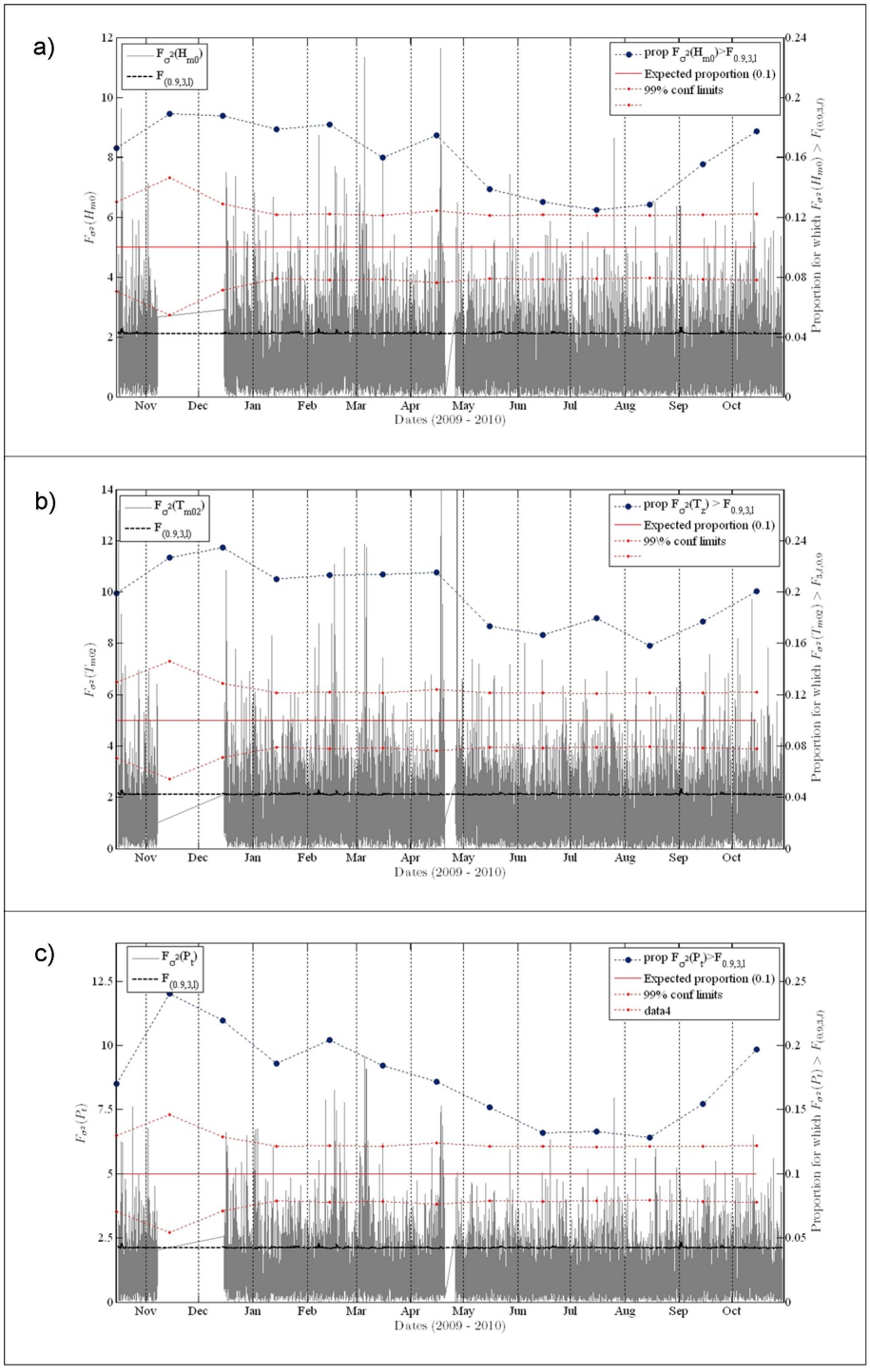

Figure 8.5: A time-series of F_{σ^2} for H_{m0}, T_{m02} and P_t between October 2009 and October 2010, and the $\alpha = 0.9$ critical limit for each record (RH axis). Also shown are the proportion of records that exceed the critical limit during each month (LH axis), the expected proportion of records that would exceed the limit $(1 - \alpha)$, and the 99% confidence limits about $1 - \alpha$ (see section 8.2.1).

Table 8.3: The mean value of $F_{\sigma^2}(H_{m0})$, and the proportion of records for which $F_{\sigma^2}(H_{m0})$ was deemed significant for each month, at a critical level of 90%. Also shown is the upper confidence interval for the proportions, estimated through simulation (see section 8.2.1).

Month	N	$\mu(F_{\sigma^2}(H_{m0}))$	$F_{\sigma^2}(H_{m0}) > F_{(0.9,3,dofW)}$	Upper 99% conf limit
Oct	693	1.29	0.17	0.13
Nov	291	1.30	0.19	0.15
Dec	725	1.34	0.19	0.13
Jan	1354	1.27	0.18	0.12
Feb	1228	1.31	0.18	0.12
Mar	1358	1.28	0.16	0.12
Apr	1077	1.32	0.17	0.12
May	1358	1.15	0.14	0.12
Jun	1320	1.13	0.13	0.12
Jul	1371	1.12	0.12	0.12
Aug	1374	1.10	0.13	0.12
Sep	1306	1.22	0.16	0.12
Oct	1218	1.28	0.18	0.12

Table 8.4: The mean value of $F_{\sigma^2}(T_z)$, and the proportion of records for which $F_{\sigma^2}(T_z)$ was deemed significant for each month, at a critical level of 90%. Also shown is the upper confidence interval for the proportions, estimated through simulation (see section 8.2.1).

Month	N	$\mu(F_{\sigma^2}(T_z))$	$F_{\sigma^2}(T_z) > F_{(0.9,3,dofW)}$	Upper 99% conf limit
Oct	693	1.46	0.20	0.13
Nov	291	1.39	0.23	0.15
Dec	725	1.50	0.23	0.13
Jan	1354	1.38	0.21	0.12
Feb	1228	1.44	0.21	0.12
Mar	1358	1.44	0.21	0.12
Apr	1077	1.50	0.22	0.12
May	1358	1.27	0.17	0.12
Jun	1320	1.27	0.17	0.12
Jul	1371	1.27	0.18	0.12
Aug	1374	1.19	0.16	0.12
Sep	1306	1.30	0.18	0.12
Oct	1218	1.35	0.20	0.12

Table 8.5: The mean value of $F_{\sigma^2}(P_t)$, and the proportion of records for which $F_{\sigma^2}(P_t)$ was deemed significant for each month, at a critical level of 90%. Also shown is the upper confidence interval for the proportions, estimated through simulation (see section 8.2.1).

Month	N	$\mu(F_{\sigma^2}(P_t))$	$F_{\sigma^2}(P_t) > F_{(0.9,3,dofW)}$	Upper 99% conf limit
Oct	693	1.16	0.17	0.13
Nov	291	1.26	0.24	0.15
Dec	725	1.29	0.22	0.13
Jan	1354	1.17	0.19	0.12
Feb	1228	1.24	0.20	0.12
Mar	1358	1.22	0.18	0.12
Apr	1077	1.20	0.17	0.12
May	1358	1.06	0.15	0.12
Jun	1320	1.01	0.13	0.12
Jul	1371	1.01	0.13	0.12
Aug	1374	0.97	0.13	0.12
Sep	1306	1.09	0.15	0.12
Oct	1218	1.17	0.20	0.12

Table 8.6: Mean values for spectral moments and the mean value of the statistic $F_{\sigma^2}(m_n)$.

Moment	buoy A	buoy B	buoy C	buoy D	$\mu(F_{\sigma^2}(m_n))$
m_{-1}	2.22	2.08	2.21	2.04	1.12
m_0	0.25	0.24	0.25	0.24	1.06
m_1	0.03	0.03	0.03	0.03	1.03
m_2	0.01	0.01	0.01	0.01	1.06

8.3.3 The dependence of variability on physical conditions

When the the observed differences between simultaneous measurements are greater than the predicted sampling variability, F_{σ^2} will be greater than 1. The colour overlay on figure 8.6 demonstrates the mean value of $F_{\sigma^2}(H_{m0})$ for each cell on the $H-T$ scatter diagram. $F_{\sigma^2}(H_{m0})$ does not demonstrate a clear dependence on sea state, and large values occur in cells with few contributing records, where higher variability of this parameter would be expected due to the relatively small number of contributing records.

Table 8.7: The proportion of records for which $F_{\sigma^2}(H_{m0})$ is significant for different wave heights and at different critical levels.

$H_{m0}(m)$	N	$\mu(F_{\sigma^2}(H_{m0}))$	$F_{\sigma^2}(H_{m0}) > F_{(0.9,3,dofW)}$	Upper 99% conf limit
0 - 1	2677	1.37	0.18	0.12
1 - 2	7434	1.19	0.14	0.11
2 - 3	3072	1.20	0.13	0.11
3 - 4	1078	1.19	0.15	0.12
4 - 5	319	1.19	0.15	0.14
5+	93	1.26	0.18	0.18

In order to examine the relationship between wave heights and the variability, records were categorised according to their H_{m0} values, to the categories shown in table 8.7. As with the monthly data, the proportion of records where $F_{\sigma^2}(H_{m0})$ exceeds the critical limit of 0.9 ($F_{(0.9,3,63)}$) in each category is greater than predicted by theory (tab. 8.7), which supports the conclusion that calculated sampling variability is not sufficient to explain observed differences between simultaneous parameters. Exceedances are measured more frequently during sea states with low wave heights, which leads to $\mu(F_{\sigma^2}(H_{m0})) = 1.37$ for the smallest seas[4]. This is also the case for very large sea states, although these are represented by a relatively small sample ($N = 93$), and the proportion of records exceeding the critical limit is less than the 99% confidence interval about 0.1, and therefore, is not deemed significant.

The time-series of $F_{\sigma^2}(H_{m0})$ values were compared to the measured wave spectra for each month. Figure 8.7 is an example of the analysis plots used, showing the spectral energy density against time alongside $F_{\sigma^2}(H_{m0})$ values recorded. The full range of these plots can be found in appendix F.

[4]The expected value for $\mu(F_{\sigma^2}(H_{m0}))$ is 1 when the differences are explained by sampling theory.

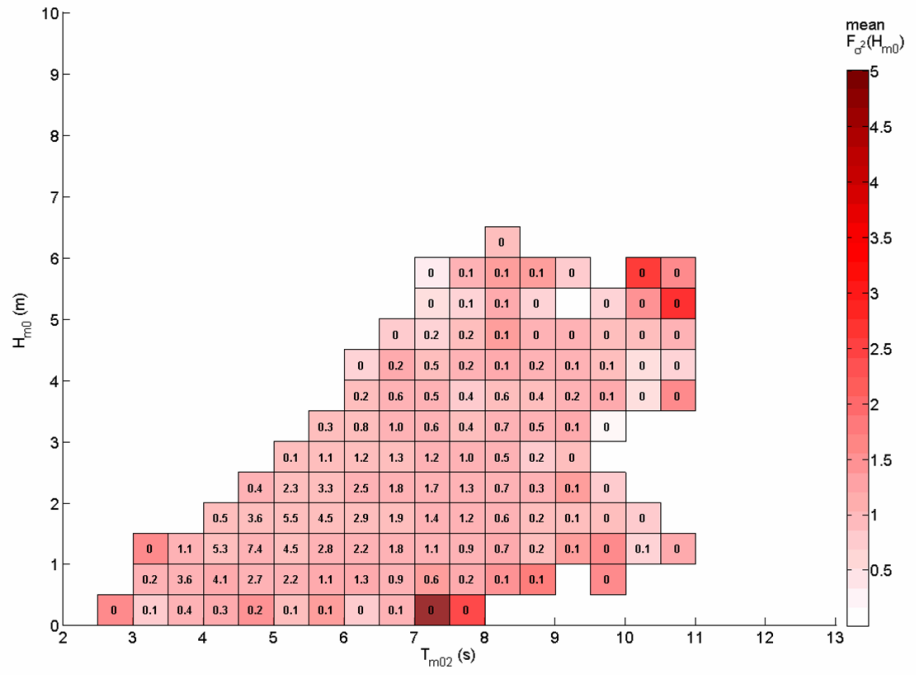

Figure 8.6: The H-T scatter diagram for all records where error-free records were available from all 4 buoys. The text values represent the percentage of records occurring with H_{m0} and T_{m02} values corresponding to each cell, and the colour overlay represents the average value of $F_{\sigma^2}(H_{m0})$ for records within that cell.

In addition, a moving average over 12 consecutive records (6Hrs) of $F_{\sigma^2}(H_{m0})$ was plotted, which is used to aid visual analysis. This average will highlight periods where variability in $F_{\sigma^2}(H_{m0})$ is persistently different from 1 more effectively than the raw figures. However, averaging will remove all variability at shorter time-scales and spread the influence of very large or small values. When drawing conclusions, care was taken to relate observations in the moving average to the raw data to ensure robust conclusions. Further analysis and examination of tidal influences used the raw data as opposed to the averages.

The time-series of $F_{\sigma^2}(H_{m0})$ (fig. 8.7b) is punctuated with periods of large values with duration greater than 10Hrs and usually less than 24Hrs, which are not representative of random variability. During these periods, the $6Hr$ moving average values of $F_{\sigma^2}(H_{m0})$ are consistently higher than 2, and in some cases exceed 4.5. Periods of elevated values commonly co-incide with low-sea states, particularly those immediately prior to the arrival of a swell, or the development of a new wave system. This corroborates the assertation that high variability is associated with small swell dominated sea states from figure 8.6, and the results in table 8.7.

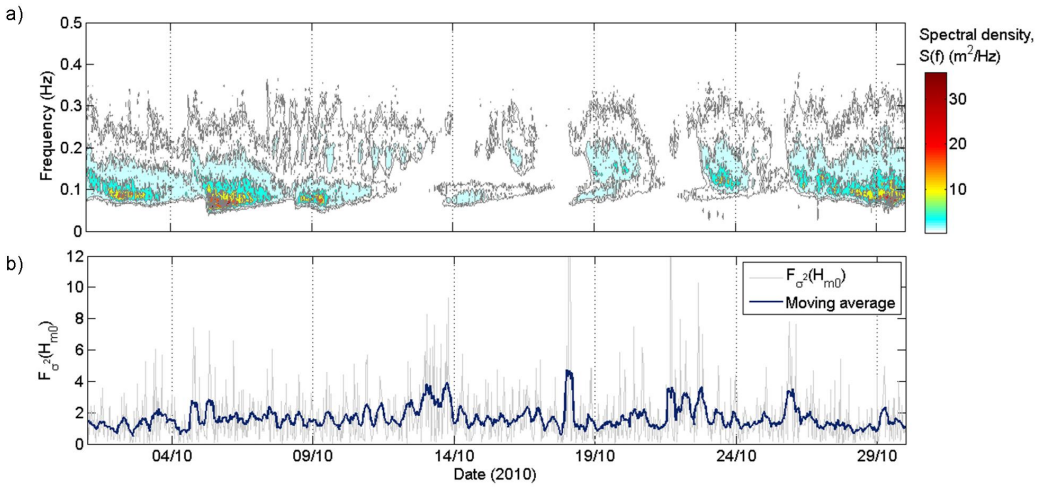

Figure 8.7: A comparison of spectral energy density and the variability in significant wave heights measured. a) Spectral energy recorded at buoy A during October 2010. b) The variability between significant wave height values measured at the four buoys during October 2010, quantified in terms of theoretical sampling variability using the parameter $F_{\sigma^2}(H_{m0})$. Also shown is a moving average of this parameter over 12 consecutive records (6Hrs).

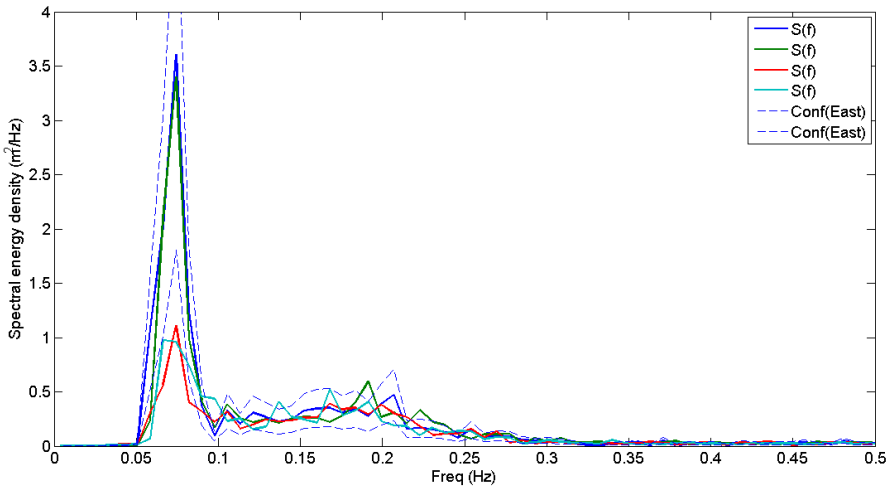

Figure 8.8: Wave spectra measured at the four buoys 17th Feb 2010 00:30. Also shown are the confidence limits for buoy A, situated at the eastern location in the array.

A visual examination showed that the records containing a peaked wave system, with peak frequency below $0.1Hz$, often exhibit large differences in the strength with which this peak is measured by the buoys, an example of which is given in figure 8.8.

The dependence of variability in T_{m02} on physical conditions

Analysis of $F_{\sigma^2}(T_{m02})$ values found that the frequency of exceedances of the 90% threshold is relatively constant for changing wave periods (tab. 8.8). However, the greatest value of $\mu(F_{\sigma^2}(T_{m02}))$ being calculated for the shortest wave periods. For all categories, the proportion of records where $F_{\sigma^2}(T_{m02})$ exceeds the critical limits are greater than would be expected were differences due to theoretical sampling error alone, and are outside the 99% confidence limits.

As with H_{m0}, the time-series of $F_{\sigma^2}(T_{m02})$ is punctuated with periods of large values with duration greater than 10Hrs and usually less than 24Hrs, which appear outside of the background random variability. During these periods, the $6Hr$ moving average is seen to consistently exceed 2, and for isolated instances, exceed 4. Visual analysis revealed that these periods of elevated values of $F_{\sigma^2}(T_{m02})$ coincide with swell conditions or periods directly preceding swell development, which corroborates the high $F_{\sigma^2}(T_{m02})$ values observed for categories of very short and very long wave periods (tab. 8.8). Visual analysis of affected records again indicates that differences between the measurements can be attributed to the magnitude of a sharply peaked swell component.

Table 8.8: The proportion of records for which $F_{\sigma^2}(T_{m02})$ is significant for different wave periods and at different critical levels.

$T_{m0}(m)$	N	$\mu(F_{\sigma^2}(T_{m02}))$	$F_{\sigma^2}(T_{m02}) > F_{(0.9,3,l)}$	Upper 99% conf limit
0 - 4	784	1.4488	0.22	0.13
4 - 6	13888	1.3445	0.19	0.11
6 - 8	6217	1.3776	0.20	0.11
8 - 10	1182	1.3869	0.20	0.12
10+	47	1.3904	0.21	0.21

The dependence of variability in P_t on physical conditions

For analysis of variability in the incident wave power, the parameter $F_{\sigma^2}(P_t)$ was calculated, and the data were again categorised by H_{m0} and T_{m02} values (tabs. 8.9 and 8.10).

For most categories, the proportion of records where $F_{\sigma^2}(P_t)$ exceeds the critical limits is greater than would be expected were differences due to theoretical sampling error alone. As with variability in H_{m0} and T_{m02}, categories of small wave heights and short periods exhibit the largest F_{σ^2} values. For P_t, the proportion of records exceeding the critical limit is only significant for wave height categories below $4m$. A high proportion of records in the largest category for wave heights ($H_{m0} > 5m$) exceed the 90% threshold, although this is less than the 99% confidence interval for the expected proportion, and was not deemeds significant.

Table 8.9: The proportion of records for which $F_{\sigma^2}(P_t)$ is significant for different wave heights and at different critical levels.

$H_{m0}(m)$	N	$\mu(F_{\sigma^2}(P_t))$	$F_{\sigma^2}(P_t) > F_{(0.9,3,l)}$	Upper 99% conf limit
0 - 1	2677	1.2722	0.17	0.12
1 - 2	7434	1.0955	0.12	0.11
2 - 3	3072	1.0864	0.11	0.11
3 - 4	1078	1.0731	0.12	0.12
4 - 5	319	1.1102	0.13	0.14
5+	93	1.1519	0.16	0.18

Table 8.10: The proportion of records for which $F_{\sigma^2}(P_t)$ is significant for different wave periods and at different critical levels.

$T_{m0}(m)$	N	$\mu(F_{\sigma^2}(P_t))$	$F_{\sigma^2}(P_t) > F_{(0.9,3,l)}$	Upper 99% conf limit
0 - 4	784	1.3639	0.20	0.13
4 - 6	13888	1.113	0.13	0.11
6 - 8	6217	1.1085	0.12	0.11
8 - 10	1182	1.117	0.13	0.12
10+	47	1.0166	0.11	0.21

Periods of elevated values of $F_{\sigma^2}(P_t)$ were again seen to coincide with swell conditions or periods directly preceding swell development. During these periods, the $6hr$ moving average values of $F_{\sigma^2}(P_t)$ are seen to consistently exceed 1.5, and in isolated cases exceed 3. The spectra of affected records again demonstrate that the principal differences are between the amount of energy measured within a swell component.

8.3.4 Tidal influence on variability

A tidal influence on the data captured by the buoy array has been observed, principally for directional parameters, but also in T_{m02} and spectral moments of an order above 0 (sec. 6). Where this influence was not the same across all 4 buoys, it would be expected to cause a tidal signal in the measured variance F_{σ^2}. No periodic signal can be observed in the time-series of F_{σ^2} values for all three parameters, although the variability hampers examination. A frequency analysis was applied using an FFT procedure applied to segments of data. This did not identify a periodic signal at the tidal frequencies in any of the data sets of F_{σ^2} (fig. 8.9).

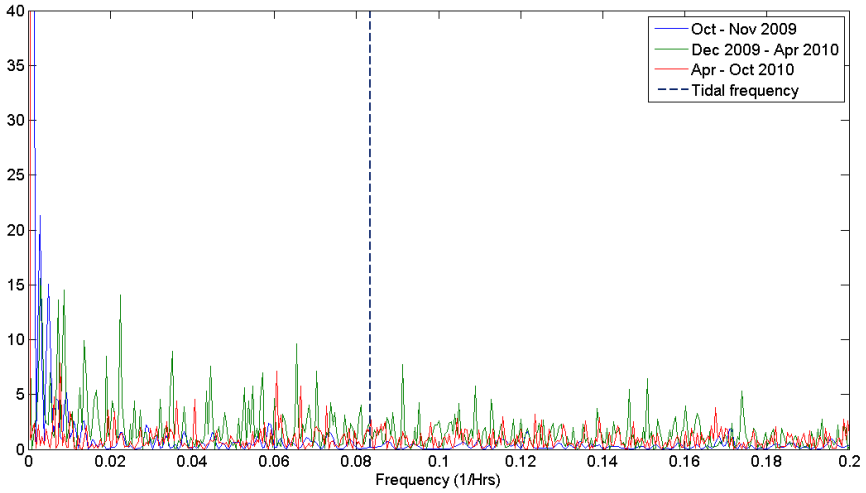

Figure 8.9: Spectral density of $F_{\sigma^2}(H_{m0})$ against frequency as $1/Hrs$ for three subsets of data.

8.4 Disucssion

The calculation of theoretical variability in key parameters, from the measured spectra, demonstrated its distribution with sea conditions (figs. 8.2,8.3,8.4). Measured variability was also seen to follow this distribution, implicating sampling variability as a primary factor in observed differences between simultaneous data. This result highlights how an estimation of variability that does not take into account theoretical variability, such as the *c.o.v.* parameter, will reflect the sea conditions in the data set from which it is derived. A swell-dominated site with a prevalence of peaked spectra would be expected to exhibit higher *c.o.v.* values than an equivalent site where conditions are dominated by wind seas.

The F_{σ^2} parameter was defined as the ratio of measured variability to theoretical variability, and was designed to provide a measure of variability relative to theoretical sampling errors, that can be applied where multiple data are measured simultaneously. As with the Z statistic, derived in Sova and Wyatt (1995), this parameter avoids a dependency on the spectral shape, or on the processing methods used. It follows that only errors in the measurement systems, or deterministic differences will influence these statistics and intercomparison between data sets of measurement sites will not be influenced by sea conditions during the measurement period. Thus, this statistic (or the Z statistic when only paired data are available) is preferable when analysing the errors in a measurement system.

Coefficient of variation (*c.o.v.*) values for the PRIMaRE buoy array, estimated using the methodology described in Mackay (2009), compare well with those values calculated for the EMEC site separated over a similar distance. The variation of *c.o.v.* with distance separation observed in Barrett et al. (2009) is not observed in results from the buoy array (tab. 8.1).

Despite displaying a similar distribution with H_{m0} and T_{m02} (figs. 8.2,8.3,8.4), the variability in the measured data is greater than that predicted by sampling theory. Measured *c.o.v.* values were consistently larger than those derived from theoretical predictions (tab. 8.2), with the greatest measured *c.o.v.*, relative to theoretical *c.o.v.* for the wave period parameter. This result suggests that the mean variability is largest in the high frequencies.

The F_{σ^2} values were also greater than predicted by sampling theory, for all 3 key parameters. However, the difference between measured and theoretical variability was more frequently deemed significant for the P_t parameter. The dependence of P_t on low frequencies in the spectrum indicates a frequent source of significant variability in the low

frequencies, a result that was corroborated with larger F_{σ^2} values for low order spectral moments. This result demonstrates how the more sensitive analysis that is possible when theoretical sampling variability is considered for each record individually, can highlight sources of variability that are potentially overlooked using the *c.o.v.* term for the whole data set.

Increased variability was seen to be dependent on the wave conditions and for all parameters, F_{σ^2} values were greatest for short periods and low wave heights. Despite this, F_{σ^2} values were greatest during winter months when the wave heights and wave periods are largest. This apparent contradiction can be explained by the combination of two factors. Firstly, the formulation of F_{σ^2} is such that, as theoretical variability decreases for smaller and wide-banded wave states, the relative effect of other sources of differences (errors or bias) will increase. Therefore, differences will have a larger effect on the magnitude of F_{σ^2} during smaller or more wide-banded sea states, categorised by smaller wave heights and shorter wave periods.

Secondly, as discussed above, low frequency wave measurements were identified as a source of increased variability. Increased storm activity in the North Atlantic means that long-range swells, and energetic local weather systems that create low frequency waves at this site, are more common during winter months and the autumnal hurricane season. When these wave systems co-incide with low sea-states, the greatest departures from theoretical sampling variability are observed.

Attributing the increased variability observed in the measurements must take into account the differences identified in section 7.2, that cannot be attributed to random variability, and are indicative of deterministic differences across the array. These differences will contribute to both the *c.o.v.* and F_{σ^2} parameters. Removal of deterministic differences prior to analysis of variance would reduce F_{σ^2} values, potentially aligning them more closely with theoretical estimations.

Chapter 9

Conclusions

Field work was successfully undertaken to deploy and maintain an array of four floating wave buoys, situated in a square formation, with sides $500m$ in length, at a location close to the Wave Hub test site, Cornwall UK. The buoys were operated for a 13 month period, measuring displacement time-series in three dimensions, complemented with a full data set of wave spectra and parameters.

This extensive field campaign was conceived and implemented with the aim of providing a unique data set for fundamental wave research, and to contribute to the data available for the Wave Hub wave energy test site. These applications demand data of the highest quality, and considerable research effort has been applied to ensure accurate data processing and quality control. This process was essential to support the detailed analysis described in this thesis, as well as developing this unique facility into a valuable resource for original research, and for supporting marine operations for the Wave Hub site.

Data processing and quality control

Marine renewable energy, as a developing field, has specific requirements from wave data that are not the same as those of traditional offshore industry. As the sector develops, so does the urgent requirement for standardisation of wave data analysis methods, to establish accepted practice. Accordingly, this has been the subject of recent research activity. UK organisations such as EMEC have contributed draft standards (Pitt, 2009), and the UK department of trade and industry (DTI) contributed draft protocols (Smith and Taylor, 2007). More recently the EQUIMAR project has worked to develop protocols for the European wave energy industry (Ingram et al., 2011), that will lead into the formation of international standards by the International Electrotechnical Commission (IEC).

A key output of the work described in this thesis is a set of bespoke functions that provide an automated processing regime based on the guidelines in the literature, along with robust algorithms for the identification of errors in the measured data. The research undertaken here to verify the processing (including quality control (QC)) methods used to derive descriptive parameters has highlighted that within the guidelines laid out by these sources, significant variability in the results of analysis can occur according to the precise treatment of errors in the raw data. These results have highlighted the difficulty of initial verification for a new data set and demonstrated that data processing methods, and the results of quality control, can be highly dependent on the properties of the measurement site and the instrumentation used.

A specific issue affecting data accuracy was found to be erroneous spectral energy caus-

ing errors in the processed data outputs. Spectral energy at frequencies below $0.04Hz$ and above $0.5Hz$ are routinely removed during parameter estimation on-board the buoys. This is in agreement with recommendations in Ingram et al. (2011), and is governed by the physical limitations of the buoy system. However, detailed analysis of the low frequency spectral filter applied during on-board processing revealed that it does not adequately remove the erroneous low frequency signal, which results in errors in the summary parameters. Compensating for these errors, an automated variable filter function was developed, which identifies and removes all spectral energy associated with an erroneous peak.

The occurrence of low frequency errors that are overcome with the variable filter was seen to be a sporadic occurrence, and the effect on long term calculated parameters averaged over many records is limited ($< 0.05\%$ of mean incident wave power, P_t). However, P_t estimated from individual records was reduced by up to 50% when the errors were removed, whilst for a single storm event lasting 4 days, the difference was 12%. The effect of these errors is not limited to short term statistics. The errors were seen to cause differences between simultaneous data, which had the potential to affect the detailed analysis described in chapters 7 and 8. The development of the variable filter was vital to drawing robust conclusions about the spatial differences in wave conditions, and is a valuable output from this research, which improves the quality of the data set.

Results of the application of QC tests to the measured data in chapter 5 demonstrated two key factors that make this process specific to a given deployment. The first is that certain tests are defined with reference to a particular property of the sensor; for floating wave buoys, this is most frequently the Datawell wave rider directional wave buoy. The second is that many of the QC tests are defined in consideration of the stochastic model of a random sea comprising the linear superposition of harmonic, regular waves. This site is in intermediate water ($35 - 42m$), with strong tidal flows. Under certain conditions non-linearities become important, reducing the applicability of linear theory to these data. For these reasons, the rate of positive flags returned by this data set were found to be specific to the conditions in which they were measured, and to the instrument used for measurement. The detailed discussion of these results in section 5.6 established certain tests to be unsuitable for the measurements captured by the array. This research, and subsequent analysis, led to specific recommendations for the treatment of error flags for the wave measurement array (sec. 5.7), providing a key contribution to the supporting documentation available to inform further research using data from this facility.

The process of error identification and subsequent removal is critical to the validity of

conclusions that are drawn. The findings presented in this thesis have shown that for high level analysis, more general tests based on the summary parameters will be sufficient to minimise errors and published guidelines provide sound advice on this topic, to support decisions (NOAA, 2005; Ingram et al., 2011). However, wave measurements are commonly undertaken to provide a data set for more detailed analysis, of which the research described in this thesis provides an interesting example. Detailed analysis will be critical in supporting the development of wave energy conversion devices and wave energy sites. The findings presented in this thesis demonstrate that QC procedures must be developed to reflect the level of detail in the analysis. Furthermore, the correct interpretation of QC tests requires a detailed knowledge of wave theory and the wave measurement technology used, as well as a thorough understanding of the measurement site. The specificity of QC procedures to a site and measurement technology cannot be fully represented in global standards, and should be established on a case by case basis using relevant previous experiences.

The identification of errors in the data set provide valuable feedback to the sensor manufacturer on the performance of their technology, and are particularly relevant to users of this sensor, and this facility specifically. A specific recommendation to the wave buoy manufacturer will be that default processing should remove erroneous spectral energy from all data outputs, or information should be provided describing these errors within the supporting documentation. Previous research has also highlighted errors in wave data captured by sensors from other manufacturers (Mackay, 2010), showing that the identification and reporting of errors in measurement technologies is a critical role for research in supporting instrument manufacturers.

Wave climate at the measurement site

The processed wave data demonstrated an energetic wave climate at the site, with some influence of tide on the wave measurements, particularly the directional parameters. The maximum significant wave height of $8.36m$ was recorded during November 2009, and the most frequent sea-states are associated with developing seas, although a significant contribution from swell seas was also observed.

A strong seasonal variability was observed, with monthly mean values for the incident wave power varying between $6.17kW$ during May 2010 and $81.30kW$ during November 2009. It follows that the absence of data for a month during the more active period of the year has a significant effect on the estimation of the long-term statistics. This is highlighted by differences of up to 17% ($1.3m$) in the estimated 1 in 1yr return values

between the buoy that was monitoring the longest during winter, and that which failed after only 2 weeks. Because all data sets have a significant gap during the winter period, the long-term statistics are not considered representative of this site. Failures are not uncommon for wave measurement technologies in energetic locations. Smith et al. (2011) highlighted how intermittent measurements at the Wave Hub site inhibit resource assessments based on measured data, and concluded that modelling must be used for all long term assessments. In this case, direct measurements can be used to validate model output to a particular site. Differences between the data returns from the 4 buoys were overcome in subsequent comparisons through limiting all data to records when all 4 buoys were recording simultaneously, that passed quality control procedures; a stipulation that was seen to reduce the data availability by 10% (sec. 6.6.1).

Correlation between wave records

The data set established offers a unique opportunity to study the effect of correlation on the combined statistics of separate wave measurements. Theoretical predictions of correlation have been discussed in this thesis (section 2.4), and applied to the measured data (section 6.7). The results indicates that correlation between wave records would be expected within the array for approximately 66% of records when the separation is parallel to the direction of propagation of a wave system. This prediction that was supported by results in chapter 7, where d_{spr} was found to have a significant effect on the regression model.

The goal of this part of the research was to quantify the effect of correlation in the displacement time series between two points in order to adjust for this dependence in their relative random error term. However, the theory applied here can only predict whether correlation occurs. Further work to predict the level of correlation between two displacement time series, and quantify the effect on the combined statistics of the data sets would be required to achieve this goal. Furthermore, correlation in the time series is dependent on the direction of propagation of the waves relative to the separation of the measurement sites. The analysis undertaken here assumed that the mean direction paramter, m_{dir}, was representative for the whole spectrum, which is not valid for multi-modal sea states. Therefore, decomposition of each measured record, into the component wave systems, would allow each system to be represented by its direction of propagation and peak frequency. This would allow a more accurate analysis of the expected correlation between measured records.

Differences between simultaneous data sets

Significant differences in simultaneous measurements of key parameters, including 8.2% of the mean value (average over 12 months), and 11% (average for one month) in the incident wave power, P_t, were identified. These figures would be of particular concern to the operation wave energy converters at this site, particularly the operation of arrays of devices, or performance assessments that engender the assumption that wave fields are stationary over a distance of $500m$. In fact, were the sea-trials of a new technology based on wave measurements taken $500m$ away from the device during October and November 2009 at this measurement site, then the observed difference would introduce a systematic bias of 11% to performance estimates based on the incident power. The observed differences were most pronounced in the low-order spectral moments, which indicates that it is variability in the measurement of low frequencies that causes differences in summary parameters. This leads to a greater difference in values of P_t across the array than H_{m0} or T_{m02}.

A multivariate least squares (MVLS) regression method was developed and applied to pairs of simultaneous measurements of the significant wave height, H_{m0}. Theoretical estimations for sampling variability were used to define a weighting for each measurement that was applied within the regression methods, to reflect the heteroscedasticity of wave parameter estimations. The analysis of residuals showed a constant error, which demonstrated the suitability of the weighting term, and revealed a Gaussian distribution for the errors, which validated the subsequent analysis of the model using hypothesis tests.

Results of the MVLS analysis demonstrated a statistically significant relationship between the relative value of simultaneous wave measurements within the array and the wave period and wave direction of the incident waves. These results provide compelling evidence that the physical processes governing wave propagation, such as the interaction with local bathymetry, are causing a significant spatial distribution of wave conditions across the array, and the measured differences can, in part, be attributed to deterministic differences. At this stage, an influence of the sensors caused by a varying physical response of the wave buoys to wave and tidal conditions cannot be ruled out entirely. However, it is a distinct advantage of this deployment that all the sensors and mooring sets were the same, which will minimise differences between their measurements. Nevertheless, follow up experiments where the full mooring and sensor systems are exchanged would add confidence to these conclusions.

Smith and Taylor (2007) recommended that in areas of complex bathymetry or tidal flows, multiple wave sensors be deployed to support wave energy installations, particularly during

testing. The research described in this thesis demonstrates methodologies that could be applied to this form of analysis in the future. The results offer a unique quantification of physical processes affecting wave fields over a spatial scale that has direct relevance to the monitoring of WEC during the test phase, and the magnitude of the differences observed over these distances are sufficient to be of concern, particularly to the wave energy community.

However, the cost implications of the field work undertaken for this thesis are significant and the physical processes under consideration are well understood. The conclusion that wave interaction with the bathymetry and the tidal flows are influencing the spatial distribution of wave conditions across the measurement array is in itself an indication of the occurrence of known physical processes. Therefore, despite the recommendations in Smith and Taylor (2007), for commercial wave energy installations, site-scale spatial variability may be predicted using computational wave models. The results from this thesis, alongside the data set, provide an excellent resource for developing the research to support these activities. In particular, it will be possible to validate model processes at this scale, without which, sufficient accuracy could not be guaranteed.

Variability in wave field parameters

Variability was initially quantified in terms of the coefficient of variation (*c.o.v.*). Values estimated for the entire data set were found to be 5.3% for H_{m0} and 3.3% for T_{m02}, which are comparable to those calculated for the EMEC site separated over a similar distance (Mackay, 2009; Barrett et al., 2009). Values predicted by sampling variability were 3.3% and 1.8% for H_{m0} and T_{m02} respectively, which are less than those derived from the measurements. For general estimations using the mean parameter values of the data sets for a buoy separated by $500m$, this implies that the theoretical variability term underestimates the errors in the measurements.

This *c.o.v* term has two principal shortcomings when estimated from spatially separate measurements. The first is that it is based on the assumption that there is no bias between the measurements from which it is derived, which results in chapter 7 have demonstrated is innaccurate for these measurements. Secondly, the dependency of sampling variability on spectral shape means that the error variance, and subsequently the *c.o.v.*, is dependent on the wave conditions that contribute to the data set. As such, it is site-specific and will exhibit seasonality, and variability on other temporal scales that is observed in wave conditions.

A new statistic, F_{σ^2}, was defined to provide a measure of variability relative to theoretical sampling errors, using simultaneous data from 4 buoys. This statistic was defined following the approach of the Z statistic derived in Sova and Wyatt (1995), and quantifies the observed variability as a ratio against the expected sampling variability. This removes the dependency of the statistics on the wave conditions that contribute to the data set, which is preferable when analysing the accuracy of wave sensors, and for the intercomparison of variability between data sets. The more sensitive analysis made possible using this statistic demonstrated that measured variability in wave parameters follows closely the predicted theoretical sampling variability. However, F_{σ^2} values for the measured data indicated that the observed variance of H_{m0}, T_{m02} and P_t were greater than predicted by sampling theory.

Increased variability (large F_{σ^2} values) was observed during the winter months, and for low order spectral moments, which implies that it is low frequencies that exhibit the greatest differences relative to sampling variability. However, in the key parameters, sea states with low wave heights and short periods were highlighted as having the largest variability relative to sampling theory. These results indicate that F_{σ^2} are at a maximum during calm conditions, when low frequency wave systems are incident on the array, but are not equally measured by all 4 buoys.

This thesis has described the considerable research effort required to deploy 4 directional wave buoys at a site and demonstrated how this unique facility was designed to support both fundamental research, and wave energy activities at the Wave Hub site. The subsequent chapters demonstrate the extensive research required to validate data collected by a new facility, highlight specific errors, and provides detailed guidelines for the treatment of errors in a wave data set. The high quality data set that resulted from this work was then applied to examine the spatial varibility of waves in an area measuring $500m \times 500m$ of depth $36-41m$. The observed differences in simultaneous mesurements were not explained by theoretical random variability, and the interaction of waves with the bathymetry and tidal flow were seen to cause deterministic differences across the area. This led to significant differences in the measured incident power between measurement buoys, which would affect the assessment of the resource, and monitoring for wave energy installations.

9.1 Further work

The work undertaken in this research project has developed a unique facility for examining spatial properties of ocean waves, which has been validated to ensure a high quality data set. It is envisaged that this work will provide critical information for a wide range of analysis and further work. Certain specific errors have been identified in the data, including the filtering of low frequency noise, and inverse spikes in the vertical displacement time-series. Discussion is underway with the instrument manufacturer into how these errors could be prevented, or whether the proposed solutions could be integrated into their standard systems.

Discussion of the correlation between time-series, and the associated coherence in the error term between separate measurements in this thesis has been based on a theoretical method for estimating correlation (Tucker, 1993). This method predicts the maximum distance between measurement points at which correlation can be expected. The data set available from the wave buoy array allows a direct assessment of correlation between wave fields using the simultaneous displacement time-series. Results from such an analysis would allow an assessment of the method of Tucker (1993), and could be used to calculate the coherence between error terms, and to quantify the effect on the differences between simultaneous measurements of summary parameters. The layout of the buoys in this facility provides an ideal data set for this analysis, which will benefit from the extensive research undertaken in this thesis to identify errors in the data.

It has been discussed in this thesis that point wave measurements captured at a wave energy site will be applied to assessments at locations separated from the measurement device. Analysis has highlighted that deterministic differences can affect wave fields over a scale of a small wave energy site. This result indicates that deterministic differences will affect the outcome of resource and performance assessments, as well as monitoring transmitted wave energy for environmental impact assessments. These effects will be site specific, but can be expected to increase for larger sites, and are likely to govern site selection and device placement. It follows that the development of an accurate methodology for the prediction of spatial differences in the wave field will be of direct benefit to the development of wave energy sites. The data captured by the wave buoy array represents a unique opportunity to develop and validate such methods.

The deterministic differences that have been identified in this thesis are caused by known oceanographic processes, that are well understood. A statistical method has been defined

to quantify deterministic differences and develop a model to predict the wave conditions across the array. However, the co-efficients which quantify the contribution of different covariates to the measured wave field could not be verified to sufficient accuracy to allow robust predictions of the spatial wave field. This area of research can be advanced through development and refinement of the methodology that was applied, or through the exploration of alternative methodologies.

An alternative method would be to develop a computational wave model for the measurement region that predicts differences across the array. Based on the physical properties of the system, these models could be considered more adaptable to changing conditions at the site. Given the expense of operating multiple wave measurements, computational models are likely to be in use, for the wider area for the purpose of resource assessments. It may be feasible to reduce the grid size (unstructured grid), or to nest a high resolution grid within the regional model in order to predict deterministic differences of wave fields within a site. In order to have confidence in a robust model, validation with direct measurements is required. Therefore, it is beneficial to pursue a statistical analysis of the measured data, alongside computational wave models, as this will offer a vital opportunity to validate results.

The research described in chapter 7, demonstrates the successful development and application of a specific regression method to examine the relationship between simultaneous wave measurements. Development of this model is recommended, to improve the quality of the results. In particular, application of a principal component analysis (PCA) prior to regression, would reduce the co-dependent variables into independent variables, for use in the regression analysis. Equivalent error variances in each variable could also be addressed by using some form of average for the independent variable; perhaps taking a moving average over a number of consecutive records. Such preparation of the variables would have to take into account potential differences within the records averaged. Another approach to this problem is to change the method of regression to account for equivalent errors in both variables. A review of regression methods can be found in the literature (Krogstad et al., 1999), although the author is not aware of any regression methodology applied to wave data that accounts for both variable errors and equivalent errors in both variables.

Results presented in this thesis have indicated that differences across the array are greatest in low frequencies. A more sensitive analysis of both deterministic differences, and errors in wave measurement, could be achieved through direct comparison of wave spectra at individual frequency bands. This would allow further conclusions to be drawn regarding

the sources of variability, such as variable wind fields affecting high frequencies, or correlation reducing differences in low frequencies. Another option that could be considered would be partitioning the spectrum into individual component wave systems and examining the difference between them. Saulnier et al. (2011a) successfully applied a partitioning algorithm to a subset of data from the measurement array, demonstrating that tidal flow primarily affects wind seas. This provides a particularly relevant example of the success of this approach to analysis of data from this facility.

The research undertaken here has provided original results for the spatial distribution of ocean waves, and identified issues regarding data handling and quality control. Specific recommendations have been made, which it is envisaged will inform the global standardisation of these procedures. The combined contribution of the thesis is a high quality data set, particularly suited to the examination of the spatial characteristics of ocean waves, and supported by results of such analysis. This represents an initial contribution of this facility to the development of the understanding of spatial characteristics of ocean waves on the scale of wave energy test site. Application of these recommendations for further work would continue to develop the potential of this unique measurement facility, to support a deeper understanding of wave energy site-scale variability. The contribution to this field of research is timely as it provides valuable information for the monitoring of wave energy devices, particularly as the industry moves towards larger sites and array deployments.

Appendix A

Sampling variability from direct measurements

Where sampling variability could not be derived theoretically, a suitable value was estimated by comparing the data from other buoys in the array. This procedure assumes that the sampling errors are Gaussian distributed, with a standard deviation that is a constant proportion of the parameter value. An estimate of X, denoted using $\hat{X}$ can be considered as,

$$X = X + \epsilon \tag{A.0.1}$$

$$\frac{X}{\hat{X}} = 1 + \frac{\epsilon}{X} \tag{A.0.2}$$

where ϵ is Gaussian distributed with mean X and standard deviation cX, where c is a constant. Therefore,

$$\sigma\left(\frac{\hat{X}}{X}\right) = \sigma\left(\frac{\epsilon}{X}\right) \tag{A.0.3}$$

making $\left(\hat{X}/X\right)$ Gaussian distributed, with mean 1, and standard deviation, $\epsilon/X = c$. The value for c estimated in this way can then be used to describe the error term for the parameter, X. To calculate $\sigma\left(\hat{X}/X\right)$, two sets of concurrent data captured from two seperate buoys in the PRIMaRE wave buoy array are used. The procedure also assumes that there are no significant differences between concurrent sea states measured by the buoys, an assumption that is tested later in this thesis, and that the sampling errors are equal. With these assumptions, the expected value of the variance of the proportional difference between them will be the sum of the proportional error variance in each parameter,

$$E\left[var\left(\frac{\hat{X_A} - \hat{X_B}}{1/2\left(\hat{X_A} + \hat{X_B}\right)}\right)\right] = var(\hat{X_A}/X) + var(\hat{X_B}/X) \tag{A.0.4}$$

$$= 2var\left(\frac{\hat{X}}{X}\right) \tag{A.0.5}$$

where X is the true value of the parameter for a given sea state.

Appendix B

Matlab source code for QC function

```
function[errflag_record errflags_TS trendout] = QC(TS, spec, plotflag, nowarn)
%IGA v3.3 27/07/2010
% Function applies Quality control procedures to wave data
%
%% Inputs
%TS is a time series with both Time and surface elevation as a matrix. %Time should be supplied in seconds. [t,
Z]
%
%Spec is a spectrum, as a matrix with frequency in the first column and
%S(f) in the second. [f,S]
%
%plotflag allows the option to plot the errors where appropriate and if
%required. Any value greater than 0 will activate plotting
%
%nowarn disables the warnings if it is set to 1
%
%% Outputs
%errflag_record is a vector stating the errors found in the record. It may contain any number of the following:
%0 - no errors
%1 - discrepancy between Tz(TS) and Tz(spec). This
%indicates errors in low sea states due to the heave resolution  causing
% small waves not to cause zero crossings.
%2 - low frequency noise, tested with a spectral ratio of
%low frequencies (below 0.03Hz) to peak freq, 3*f(peak).
%3 - High frequencies exceed the Phillips spectrum
%4 - ratio, Hsig to hsigma exceeds 1.05 - indicates
%problems with the time-series
%5 - Direct time series analysis indicates a problem with
%the time-series
%6 - A significant trend was identified
%
%errflags_TS will be a vector the same length as TS containing index values for the points flagged for errors.
%
%10 - out of range, value > 5*std(eta)
%20 - Too steep, S > 1/12 %This is not working as flagging too often
%30 - Repetition, three identical values
%40 - Acceleration test, acc>g/3
%
%trendout is a matrix with the co-efficients calculated from a least squares regression of the time series in
the first column and the probability that each co-efficient would arise if its expected value was 0 in the 2nd
column (used for hypothesis testing of whether the trend is significant)
%
%% Notes + updates
%Periods are defined as the Down crossing period
%Requires WAFO to be initialised
%
%IGA - 23rd Feb
%A high frequency cut-off has been introduced and errors corrected in the
%calculation of the phillips spectrum. The high frequency cut-off has been chosen as 0.5Hz -> T = 2s -> lambda =
6.25m. This is close to the lowest theoretical operating frequency of the buoy (twice its diameter).
%
%IGA - 23rd Feb
%The out of range test in the direct time series analysis was updated to
%identify any excursions from the absolute range of the buoy
%
%IGA 18th may
%If no spectral checks are required, enter the spectrum as an empty
%variable [] for example,
%[errflag_record errflags_TS] = QC(TS, [], 1, 0);
%this will only perform the TS checks.
%
% -> 2.0
%IGA changed 4\sigma(eta) to 5\sigma(\eta) to match details in Tucker and Pitt (2001).
%->2.1 changed value of the Phillips parameter (and the coding of the
%Phillips spectrum)
%->2.2 changed the threshold for steepness from 5*std(S) to 1/12 as
%suggested in Tucker and Pitt (2001)
%->2.3 Changed the Phillips tail to (-4) to reduce flags
%->2.4 Changed back to the original formulation for the Phillips spectrum
%to attempt to reduce flags (v2.3 was not successful) - it is seen that the (-5) tail gives more lee-way
%->v2.5 Defining the Phillips spectrum according to peaks in the wind
%swells rather than according to swell peaks. This is done by limiting the
%spectrum to the higher frequencies (set by f_lim to >0.1Hz). The Philips
%spectrum is generated from the maximum in this zone
%v2.6 Previous had no significant effect. flag value was changed to double
%the Phillips spectrum and only for values greater than 1/20 S(fpeak). The aim
%is to only flag serious violations.
```

```
%v2.7 - removed one of the low frequency ration values as the poor
%resolution means that the upper is within the range of where actual waves could be expected.
%v2.9 changed the error flag 5 to the percentage of values in the time
%series that have been flagged
%v2.10 changed to dat2steep for direct TS analysis rather than dat2steepI
%V3.0 -> added acceleration test which successfully identifies inverse
%spikes in the data.
%v3.1 updated the out of range test so that m0 is calculated from the
%spectrum rather than from the time-series.
%update 3.2 - attempted to isolate the largest spectral peak by smoothing
%the spectrum and identifying highest freq maximum
%3.3 identify spectral peaks using spec_peaks function and then fit the
%Phillips spectrum only on the highest frequency wave system.

%% set-up variables
specplot = 0;
errflag_record = zeros(10,1);
hfc = 0.5;
lfc = 0;
Nave = 16; %(change according to spectral estimation)

%% ensure the time series is valid
if length(find(TS(:,2))) > length(TS)/2 && length(find(diff(TS(:,2))))>length(TS)/2
    % Parameters required for analysis
    [S, H,Ac,At,Tcf,Tcb,z_ind]=dat2steep(TS,10,(-3));

%% De-trend time-series
     %least squares linear fit
    x = [ones(length(TS),1)  TS(:,1)];
    [b stdb mse] = lscov(x,TS(:,2));%LSquares the time series, x is time, y is the heave motions
    dof = length(TS)-length(b);
    t = b./stdb;
    p1=2*(1-tcdf(abs(t),dof));%finds the probability that T is larger than calculated t values
    %when T from a distribution where b = 0. If T is unlikely to be larger than the
    %calculated t (value of p1 below critical level), then we conclude b~=0
    trendout = [b,p1];
    if p1<0.1
        errflag_record(6) = 6;
        if nowarn ~=1
            trend_warn = ['A trend was identified in the time series. Coefficents are ' num2str(b,2),];
        warning(trend_warn)
        end
    end

%% check whether a spectral input has been registered
    if isempty(spec) == 1
        Fs = 1/(TS(3)-TS(2));
        Nave = round(length(TS)*0.004);
        [spec] = get_spec_nave(TS,Fs,Nave,hfc,lfc);
    end
        spec = spec_repair(spec,0);
        spec = spec{1};
        [m0 m1 mm1 m2] = moms_anydata(spec(:,1),spec(:,2),0.5,0);
        %% High freq cut-off
        %
        %f0 = g/(2*pi) %This is the theoretical cut-off of the Phillips spectrum
        %however, for wind speeds less than 55km/h this will be higher than the
        %Nyquist frequency. The high freq cut-off has been defined above (hfc)
        cut_ind = find(spec(:,1) >=hfc,1,'first');%finds the index of the cut-off
        %frequency
        cspec = spec(1:cut_ind,:);%cuts the spectrum down to the frequencies lower
        %than the cut-off (assumes the spec data are in frequency order)

%% Discrepancy between Tz(TS) and Tz(Spec)

        %Time series Tz
        TzTSs = dat2wa(TS,0,'d2d');
        TzTS = mean(TzTSs);

        %spectral Tz
        TzS = sqrt(m0/m2);

        if TzTS/TzS >= 1.1
            errflag_record(1) = 1;
            if nowarn ~=1
            warning('The ratio of Tz(Time series) to Tz(Spec) indicates errors in Tz(spec)')
            specplot = 1;
```

```
            end
        end

%% Low frequency noise
        %S(0.0146)S(0.0195)+S(0.0244)/3fpeak > 2%
        [val indmax] = max(cspec(:,2));
        ind1 = find(cspec(:,1)>=0.0146,1,'first');
        ind2 = find(cspec(:,1)>=0.0195,1,'first');
        %ind3 = find(cspec(:,1)>=0.0244,1,'first');%removed because of poorer freq res in get_specv2.0

        %Sratio = (cspec(ind1,2)+cspec(ind2,2)+cspec(ind3,2))/(3*val);%v2.6
        Sratio = (cspec(ind1,2)+cspec(ind2,2))/(2*val);

        if Sratio >= 0.02
            errflag_record(2) = 2;
            if nowarn ~=1
            warning('The ratio of low frequency to peak frequency indicates excessive low frequency noise')
            specplot = 1;
            end
        end

%% Exceedance of the phillips spectrum
    lf_lim = find(cspec(:,1)>= 0.1,1,'First');
    hf_spec = cspec(lf_lim:end,:);

    [npeaks peaks] = spec_peaks(hf_spec,Nave,1,0);
    if isempty(peaks)==0
        lim_ind = peaks(end);
    else
        lim_ind = lf_lim;
    end
    hf_spec = hf_spec(lim_ind:end,:);%seperate the portion of the spectrum higher than the spectral peak
    assignin('base','hf_spec',hf_spec);

    %From Tucker and Pitt p65 and 102
    %where alpha rarely exceeds 0.015 (no indication of how rarely)
    alpha = 0.015;
    pspec = alpha*(9.81^2)*((2*pi)^(-4)).*((hf_spec(:,1)).^(-5));
    pspec2 = alpha*(9.81^2)*((2*pi)^(-4)).*((hf_spec(:,1)).^(-4));
    pinds = find(2.*pspec-hf_spec(:,2) < 0 & hf_spec(:,2) > val/20);
    %where val is the maximum S(f)

    if isempty(pinds) == 0

        errflag_record(3) = 3;
        if nowarn ~=1
         warning('The threshold of 2*Phillips spectrum is exceeded')
         specplot = 1;
        end
    end

%% Ratio of height(sig) to height(rms)
    Hsort=sort(H,'descend');
    Hsig=mean(Hsort(1:(floor(1/3*length(H)))));%gives H1/3

    zrms = sqrt(sum(TS(:,2).^2)/length(TS));
    Hrms = 4*zrms;

    if Hsig/Hrms > 1.05
        errflag_record(4) = 4;
        if nowarn~=1
        warning('Height ratio indicates errors')
        specplot = 1;
        end
    end

%% Direct time series analysis for errors
    dc = length(TS);
    errflags_TS = zeros(dc,1);

    % 1. out of range
    mA = mean(TS(:,2));

    %Out of range for time series
    for k = 1:dc
        if TS(k,2)>(mA+(5*sqrt(m0)))
            errflags_TS(k) = 10;
```

```
            errflag_record(7) = 10;
        end
        if TS(k,2)<(mA-(5*sqrt(m0)))
            errflags_TS(k) = 10;
            errflag_record(7) = 10;
        end
    end

    %Out of range for absolute maximum of the buoy (+/-20m)
    for k = 1:dc
        if abs(TS(k,2))>(20)
            errflags_TS(k) = 10;
            errflag_record(7) = 10;
        end
    end

    % 2. Too steep
    %threshold for steepness
     %stddif = std(S);difthresh = 5*stddif;
    difthresh = 1/12;
    for d = 1:length(S)
        if S(d) > difthresh
            u_ind = z_ind(2:2:end);
            twave = u_ind(d):u_ind(d+1)-1;
            errflags_TS(twave)=20;
            errflag_record(8) = 20;
        end
    end

    % 3. Repetition
    for k = 1:(dc-6);
        if TS(k,2)== TS((k+1),2) && TS(k,2) == TS((k+2),2) && TS(k,2) == TS((k+3),2)&& TS(k,2) == TS((k+4),2) &&
TS(k,2) == TS((k+5),2)&& TS(k,2) == TS((k+6),2)&& TS(k,2) == TS((k+7),2);
        %if TS(k,2)== TS((k+1),2) && TS(k,2) == TS((k+2),2)
           errflags_TS(k+1:k+7) = 30;
           errflag_record(9) = 30;
        end
    end

    % 4. Acceleration
        sp = diff(TS(:,2));
        acc = diff(sp);
        [val ind] = max(acc);
        if  val > 9.81/3
            errflags_TS(ind:ind+1) = 40;
            errflag_record(10) = 40;
        end

    if isempty(find(errflags_TS))== 0 %If errors are noted
        errflag_record(5) = (length(find(errflags_TS))/length(TS))*100;
        if nowarn ~= 1
            warning('Direct analysis of the time-series indicates errors')
        end

%% plot TS
        if plotflag >0
            range_inds = find(errflags_TS == 10);
            steep_inds = find(errflags_TS == 20);
            rep_inds = find(errflags_TS == 30);

            figure1 = figure('PaperSize',[20.98 29.68]);

            % Create axes
            axes1 = axes('Parent',figure1,'YGrid','on','XGrid','on','FontSize',14);
            box('on');
            hold('all');
             % Create plot
            plot(TS(:,1),TS(:,2),'DisplayName','Surface elevation','Parent',axes1);
             % Create xlabel
            xlabel('Time (s)','FontSize',14);
             % Create ylabel
            ylabel('Vertical Elevation (m)','FontSize',14);
             % Create plot
            plot(TS(range_inds,1),TS(range_inds,2),'DisplayName','Out of range >
4\sigma_\eta','Parent',axes1,...
                'MarkerFaceColor',[1 0 0],...
                'Marker','o',...
```

```
                'LineStyle','none',...
                'Color',[1 0 0]);
             % Create plot
            %plot(TS(steep_inds,1),TS(steep_inds,2),'DisplayName','Too steep > 4\sigma_S','Parent',axes1,...
             %    'MarkerFaceColor',[0 0 0],...
              %   'Marker','o',...
               % 'LineStyle','none',...
                %'Color',[0 0 0]);
             % Create plot
            plot(TS(rep_inds,1),TS(rep_inds,2),'DisplayName','Repetition (3 in a row)','Parent',axes1,...
                'MarkerFaceColor',[0 1 0],...
                'Marker','o',...
                'LineStyle','none',...
                'Color',[0 1 0]);

            % Create legend
            legend(axes1,'show');
            axis('tight');

        end
    end
%% plot spec
    if plotflag>0 && specplot == 1;
        figure2 = figure('PaperSize',[20.98 29.68]);

        % Create axes
        axes2 = axes('Parent',figure2,'YGrid','on','XGrid','on','FontSize',14);
        box('on');
        hold('all');
         % Create plot
        plot(cspec(:,1),cspec(:,2),'DisplayName','S(f)','Parent',axes2,'LineWidth',2);
         % Create plot
        plot(hf_spec(pinds,1),hf_spec(pinds,2),'DisplayName','Error values?','Parent',axes2,...
            'MarkerFaceColor',[1 0 0],...
            'Marker','o',...
            'LineStyle','none',...
            'Color',[1 0 0]);

        plot(hf_spec(:,1),pspec,'DisplayName','Phillips Spectrum(-5)','Parent',axes2,'LineWidth',2,'Color',[0 1
0]);
        plot(hf_spec(:,1),pspec2,'DisplayName','Phillips Spectrum(-4)','Parent',axes2,'LineWidth',2,'Color',[1 1
0]);

        ylim([0 max(cspec(:,2))*1.1]);

        % Create xlabel
        xlabel('Frequency (Hz)','FontSize',14);
         % Create ylabel
        ylabel('Spectral Energy density (m^2/Hz)','FontSize',14);
         % Create legend
        legend(axes2,'show');
    end
 else
    errflag_record =99;
    errflags_TS = ones(length(TS),1)+99;
    trendout = 99;
    if nowarn ~= 1
        warning('Record is mainly zeros')
    end
end
```

Appendix C

Investigation of the cause of spurious low frequency energy

An effort was made to establish the underlying cause of the spurious low frequency energy discussed in section 4.8. Low frequency energy could not be directly attributed to incident wave systems, however work was undertaken to establish whether wave conditions have an effect on low frequency measurements.

The distribution of the magnitude of r_{lf} with wave conditions was examined by plotting a scatter diagram using H_{m0} and T_{m02} values estimated from a filtered spectrum ($0.04Hz$ filter), and calculating the mean value of r_{lf} for all records within each cell. The results for buoy B are shown in figure C.1, which demonstrates that the majority of low-frequency energy is measured during steep, developing seas where the wave period is low relative to the wave height.

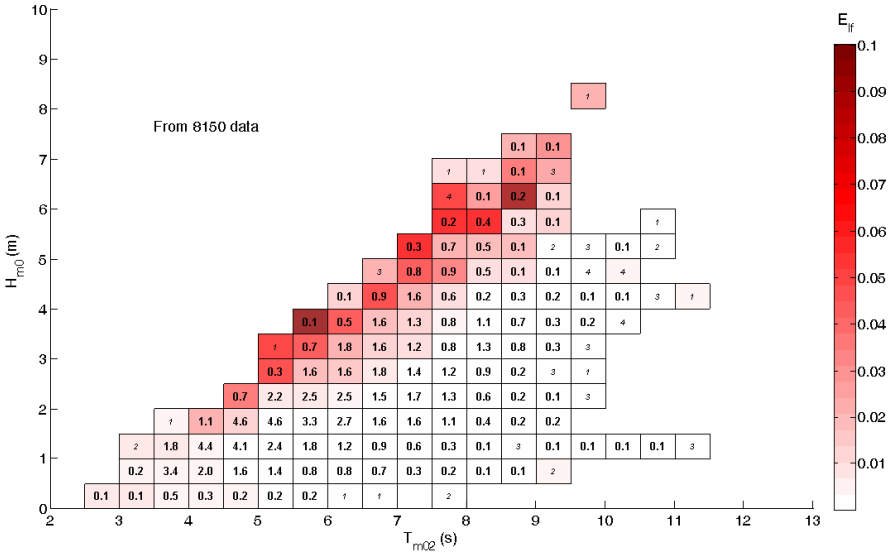

Figure C.1: A $H_{m0}, Tm02$ scatter diagram for all data collected by buoy B between Oct 2009 and Apr 2010, where text shows the percentage of records, or the total number of records (values in italics), occurring in each cell, and colour overlay shows the mean value of r_{lf} for records occurring within that cell.

The mean spectra corresponding to cells in the scatter diagram were plotted in figure C.2 (For clarity, the size of each cell was increased). An influence of spectral energy below the $0,04Hz$ threshold can again be identified in spectra related to cells where wave heights are large relative to the mean zero-crossing period. Low-frequency energy cannot be seen in the spectrum for zero-crossing periods between 6 and $8s$ and wave heights above $6m$. However, such sea states are rare, the spectrum for this cell was estimated from only one record and may not be representative of mean conditions.

For records with $H_{m0} < 3m$, and $T_{m02} < 6s$, two peaks can be observed in the mean spectra on figure C.2. This is not due to spurious low-frequency energy, but true bi-modality in the mean conditions for these cells. However, in some instances a further peak can be identified close to the $0.04Hz$ cut-off, which is attributed to spurious low frequency energy.

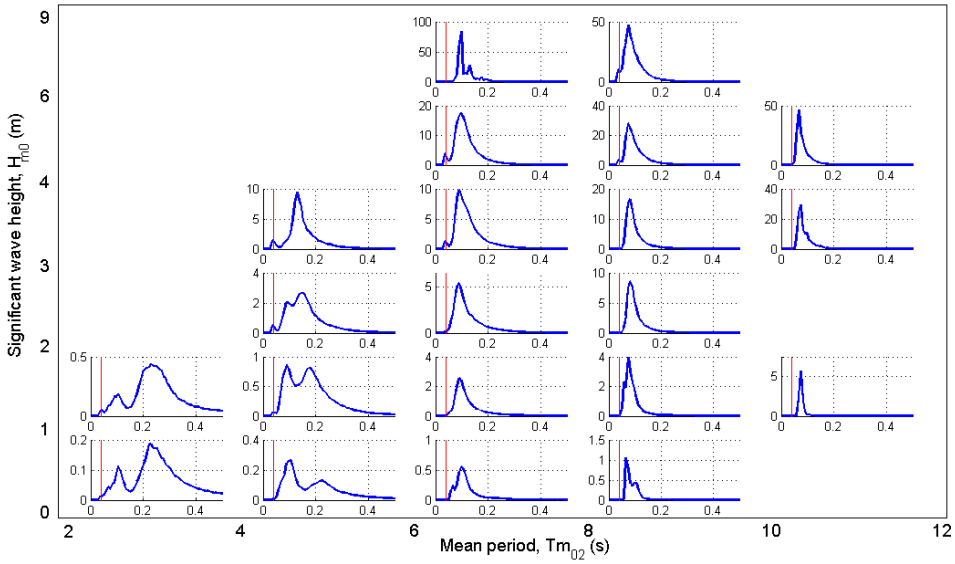

Figure C.2: An $H_{m0}, Tm02$ scatter diagram for buoy B, where each sub-figure represents the mean spectrum for all records occuring in that cell. Shown on each sub-diagram is the $0.04Hz$ low-frequency cut-off.

C.1 The observed distribution of low frequency spectral energy with tidal flow conditions

Tidal flow will affect the forces on the mooring, which may change the response characteristics and alter the physical response of the buoy to incident waves. Furthermore, the properties of the waves themselves will be altered through wave-current interactions. It was considered possible that along with the wave conditions, tidal currents contribute to inaccuracies in buoy response or data processing, and the distribution of low frequency spectral energy with tidal conditions was examined. This research used data from the national tidal model, Polpred (POL, 2010), described in section 3.1.4. Available data were a depth integrated flow speed, tidal flow direction and water level above mean sea level.
A hypothesis test was designed to examine the dependence of the records in each category from table 4.5 with the state of the tide. The null hypothesis was defined that the oc-

currence of low-frequency noise is independent of the tidal conditions. Therefore, because approximately half of all records will be recorded during flood tides, the number of records recorded during a flood tide in each category would be expected to follow the Binomial distribution, $BIN(N, p)$,

$$p(x) = \frac{N!}{x!(N-x)!} p^x q^{(k-x)} \tag{C.1.1}$$

where N is the number of records recorded in each category, p is the probability of an individual record being captured during a flood tide, $q = 1 - p$, and $P(x > X)$ is the probability that the observed number of records from a flood tide, X would be less than a given value of x. X was set to the number of records in each category that were captured during a flood tide, and p set at 0.5. Where $P(x < X) > 0.95$ or $P(x > X) < 0.05$, the null hypothesis was rejected, and it was concluded that the occurrence of low-frequency noise was not independent of the state of the tide.

Table C.1: The categories of r_{lf} values for records captured during October and November 2009 from all buoys, the corresponding number of records in each category, N, the proportion which correspond to a flood tide, N_{flood}/N, and the probability of the observed value of N_{flood}/N from a binomial distribution, $P(X > x)$. Values in bold are statistically significant with a 95% confidence limit.

	Buoy A			Buoy B		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	887	0.49	0.75	1545	0.45	**1.00**
$0.013 < r_{lf} < 0.06$	133	0.59	**0.02**	278	0.64	**0.00**
$0.06 < r_{lf} < 0.1$	19	0.58	0.18	101	0.68	**0.00**
$0.1 < r_{lf} < 0.3$	17	0.65	0.07	101	0.78	**0.00**
$r_{lf} > 0.3$	7	1	**0.00**	17	0.88	**0.00**
	Buoy C			Buoy D		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	961	0.47	**0.96**	1140	0.47	**0.97**
$0.013 < r_{lf} < 0.06$	217	0.55	0.09	194	0.6	**0.00**
$0.06 < r_{lf} < 0.1$	73	0.53	0.24	40	0.68	**0.01**
$0.1 < r_{lf} < 0.3$	102	0.61	**0.02**	31	0.71	**0.01**
$r_{lf} > 0.3$	21	0.86	**0.00**	3	1	**0**

For the first deployment, the proportion of records in each category that were recorded during flood tides was seen to increase for larger values of r_{lf} (tab. C.1). For buoys B

Table C.2: The categories of r_{lf} values for records captured between December 2009 and April 2010 from all buoys, the corresponding number of records in each category, N, and proportion which correspond to a flood tide, N_{flood}/N.Values in bold are statistically significant with a 95% confidence limit.

	Buoy A			Buoy B		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	6327	0.5086	0.0833	5917	0.51	0.06
$0.013 < r_{lf} < 0.06$	59	0.4746	0.6026	162	0.52	0.38
$0.06 < r_{lf} < 0.1$	6	1	**0**	33	0.55	0.28
$0.1 < r_{lf} < 0.3$	1	1	**0**	38	0.58	0.16
$r_{lf} > 0.3$	0	0	N/A	22	0.77	**0.01**
	Buoy C			Buoy D		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	5577	0.5116	**0.0409**	5938	0.51	0.5
$0.013 < r_{lf} < 0.06$	203	0.5222	0.2414	146	0.43	0.94
$0.06 < r_{lf} < 0.1$	57	0.5263	0.2983	30	0.6	0.1
$0.1 < r_{lf} < 0.3$	71	0.493	0.5	31	0.58	0.14
$r_{lf} > 0.3$	7	0.1429	0.9375	6	0.83	**0.02**

and D, the null hypothesis can be rejected for each category, whereas the differences are not as definitive for buoys A and C. It can be concluded that the tidal conditions have a significant effect on the frequency of occurrence of low-frequency spectral energy during this deployment. Occurrences of $r_{lf} > 0.013$ were less frequent during the second deployment when wave conditions were generally less energetic (tab. C.2). The dependence of r_{lf} on the tide that was observed during the first deployment is only in evidence in data from buoys B, D and to a lesser extent buoy A, and in all cases the distribution is weaker. It is only for the largest values that it can be considered significant. Data from buoy C demonstrates a reversal of the trend, with large values of r_{lf} less frequently measured during flood tides than low values.

C.2 The combined effect of incident waves and tidal conditions

Data from buoy B were categorised according to the mean direction of the incident waves, m_{dir}. This designates the direction from where the waves are travelling. 0^o represents magnetic north and the value increases clockwise making 90^o East and 270^o West. The categories were defined as four quadrants, shown in table C.3. Values in the 1st quadrant mean that the waves are broadly aligned with ebb tidal flow, whist values in the third quadrant mean that they are aligned with flood tidal flow. The number of records in each category demonstrates the dominance of Westerly and South westerly waves during the first deployment, compared to the variability during the second deployment (tab. C.3). The predominant east-west storm track across the North Atlantic combined with the largest fetch available to this site from the west and south west, mean that large wave states such as those seen during the first deployment most commonly arrive from these directions.

Table C.3: The number of records from buoy B that occur in categories defined by wave direction, m_{dir}, for both deployments.

Quadrant	Direction	Oct-Nov 09	Dec 09-Apr 10
1	$0 < m_{dir} \leq 90$	4	965
2	$90 < m_{dir} \leq 180$	7	136
3	$180 < m_{dir} \leq 270$	1248	2496
4	$270 < m_{dir} \leq 360$	778	2628

To test for a dependence of low-frequency occurrence on direction, the data were categorised by the mean propagation direction. Tables C.4 and C.5 demonstrate that where the tidal flow is aligned with the wave propagation direction, a high proportion of records demonstrate low frequency noise.

This result is displayed graphically on figure C.3, which plots the tidal flow for records in each category on a vector plot, where the displacement of each point from the origin is equivalent to the speed of flow at the beginning of the record in ms^{-1}, and the displacement is in the direction that the tidal current is travelling towards. Records relating to different r_{lf} categories were plotted with different colours in order to identify a distribution of low-frequency noise with tidal flow conditions.

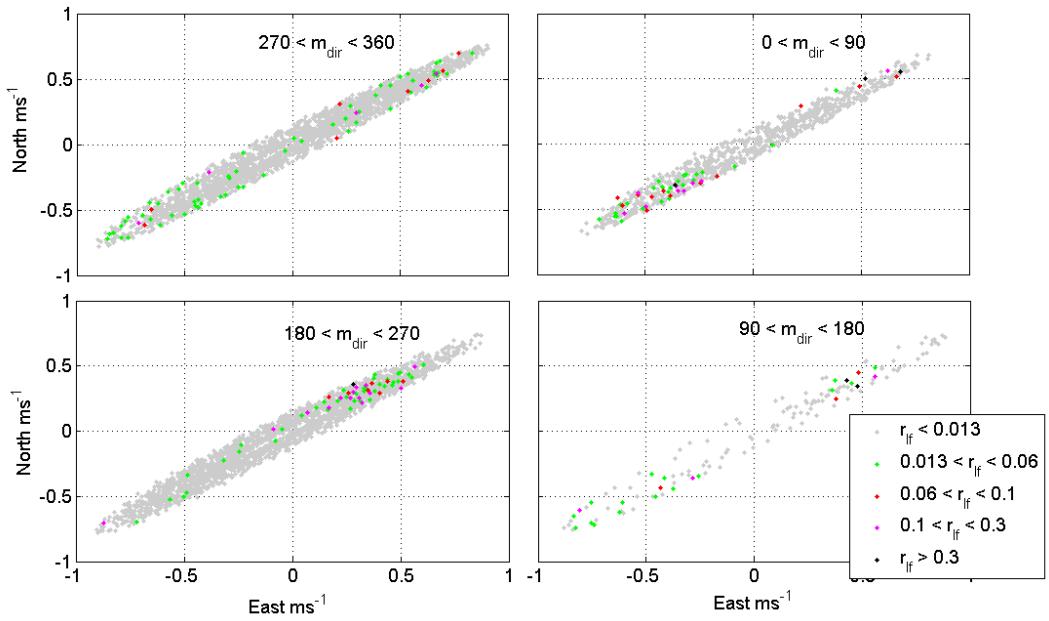

Figure C.3: Vector diagrams of tidal flow conditions for records in the sample data set. a) records in directional category 1, b) category 2, c) category 3 and d) category 4. The position of each point is governed by the tidal flow speed and direction predicted by the PolPred tidal model. Coloured points represent those with large values of r_{lf}.

Where tidal flows are aligned with the mean direction of wave propagation, m_{dir}, low-frequency spectral energy occurs more frequently. Occurrences are not limited to extreme tidal flows, nor do they only occur when tide and waves are aligned. Also evident is that occurrences are more frequent during the first deployment, even when tidal flows are not aligned with wave directions.

Table C.4: The categories of m_{dir} values for records captured during October and November 2009 from buoy B, the corresponding number of records in each category, N, the proportion which correspond to a flood tide, N_{flood}/N, and the probability of the observed value of N_{flood}/N from a binomial distribution, $P(X > x)$. Values in bold are statistically significant with a 95% confidence limit.

	$m_{dir} < 90$			$90 < m_{dir} < 180$		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	3	0.33	0.50	1	1.00	**0.00**
$0.013 < r_{lf} < 0.06$	1	1.00	**0.00**	0	N/A	N/A
$0.06 < r_{lf} < 0.1$	0	N/A	N/A	1	1.00	**0.00**
$0.1 < r_{lf} < 0.3$	0	N/A	N/A	1	1.00	**0.00**
$r_{lf} > 0.3$	0	N/A	N/A	4	0.50	0.31
	$180 < m_{dir} < 270$			$m_{dir} > 270$		
$r_{lf} < 0.013$	956	0.42	**1.00**	585	0.50	0.53
$0.013 < r_{lf} < 0.06$	149	0.70	**0.00**	128	0.55	0.09
$0.06 < r_{lf} < 0.1$	61	0.87	**0.00**	39	0.38	0.90
$0.1 < r_{lf} < 0.3$	75	0.88	**0.00**	25	0.48	0.50
$r_{lf} > 0.3$	11	1.00	**0.00**	2	1.00	**0.00**

Table C.5: The categories of m_{dir} values for records captured between December 2009 and November 2010 from buoy B, the corresponding number of records in each category, N, the proportion which correspond to a flood tide, N_{flood}/N, and the probability of the observed value of N_{flood}/N from a binomial distribution, $P(X > x)$. Values in bold are statistically significant with a 95% confidence limit.

	$m_{dir} < 90$			$90 < m_{dir} < 180$		
r_{lf}	N	N_{flood}/N	$P(x)$	N	N_{flood}/N	$P(x)$
$r_{lf} < 0.013$	915	0.48	0.91	112	0.65	**0.00**
$0.013 < r_{lf} < 0.06$	30	0.07	**1.00**	16	0.25	**0.9616**
$0.06 < r_{lf} < 0.1$	12	0.25	0.927	3	0.67	0.13
$0.1 < r_{lf} < 0.3$	9	0.1111	**0.9805**	3	0.33	0.50
$r_{lf} > 0.3$	3	0.6667	0.125	2	1.00	**0.00**
	$180 < m_{dir} < 270$			$m_{dir} > 270$		
$r_{lf} < 0.013$	2334	0.48	**0.99**	2556	0.55	**0.00**
$0.013 < r_{lf} < 0.06$	40	0.78	**0.00**	60	0.43	0.82
$0.06 < r_{lf} < 0.1$	7	1.00	**0.00**	8	0.75	**0.04**
$0.1 < r_{lf} < 0.3$	14	0.93	**0.00**	5	0.60	0.19
$r_{lf} > 0.3$	1	1.00	**0.00**	0	N/A	N/A

Appendix D

Variable spectral filter function for matlab

```
function[spec_out err] = spec_repair(spec,plotflag,L)
%IGA
%Function will repair a spectrum with a large amount of low-frequency
%energy that is considered spurious.
%IMPORTANT - that sufficient analysis has been performed prior to this
%moment to ensure correct removal, and not removal of valid data.
%
%% Input
%
%spec is a spectrum as either [f S] or a cell array where each cell
%contains a spectrum [f S].
%
%Nave is set at the number of averages used in processing
%cut-off is the frequency below which a spectral peak is considered
%invalid.
err = 0;
if iscell(spec) == 0
    spec_c = cell(1);spec_c{1} = spec;
    spec = spec_c;
end
%make spec a cell array
%Set L - the limit at which peaks are considered potentially erroneous
if isempty(L)
    L = 7;
end
for i = 1:length(spec)
    spec_peaks(spec{i},16,1,plotflag);
    spec_out{i} = spec{i};
    spec_out{i}(1:5,2) = zeros(5,1);%Cut-off filter
    rs = spec{i}(1:L,:);
    %% ID peaks
    dif = diff(rs(:,2));
    max = find(dif<0,1,'Last');%find highest freq peak ID'd (last peak below 0.045Hz)
    if isempty(max)==0
    %% ID next min
        rs = spec{i}(max:end,:);
        dif = diff(rs(:,2));
        min = (find(dif>0,1,'First'))+(max-1);

    %% Remove peak
        spec_out{i}(1:min,2) = zeros(length(spec_out{i}(1:min,2)),1);
    %% Repair spec using linear interpolation

        if spec{i}(min,1)> 0.06
            err = 1;
        end
    else
        min = 5;%if no peaks are identified, spec is tapered betw 0.045Hz and 0.04Hz
    end
    cS = spec{i}(min,2);%Value at which spec is cut-off
    cf = spec{i}(min,1);
    [f0] = PM_inv(0.0081,cS,cf);%fitting PM spec to LF values according to first valid freq
    [S] = PM(spec{i}(1:min,1),f0,0.0081);%Generating PM spec for removed section
    spec_out{i}(1:min,2) = S;
end
if plotflag ~= 0
    plot(spec_out{i}(:,1),spec_out{i}(:,2),'DisplayName','Repaired spectrum','LineWidth',2,'Color',[0
0 1]);

plot(spec_out{i}(1:min,1),spec_out{i}(1:min,2),'DisplayName','Repairs','Marker','o','LineStyle','none
','Color',[0 1 0]);
    if err ==1
        plot(spec_out{i}(L:min,1),spec_out{i}(L:min,2),'DisplayName','Possible
error','Marker','o','LineStyle','none','Color',[1 0 0]);
    end
end
end
```

Appendix E

Time-series of wave parameters for 4 buoys

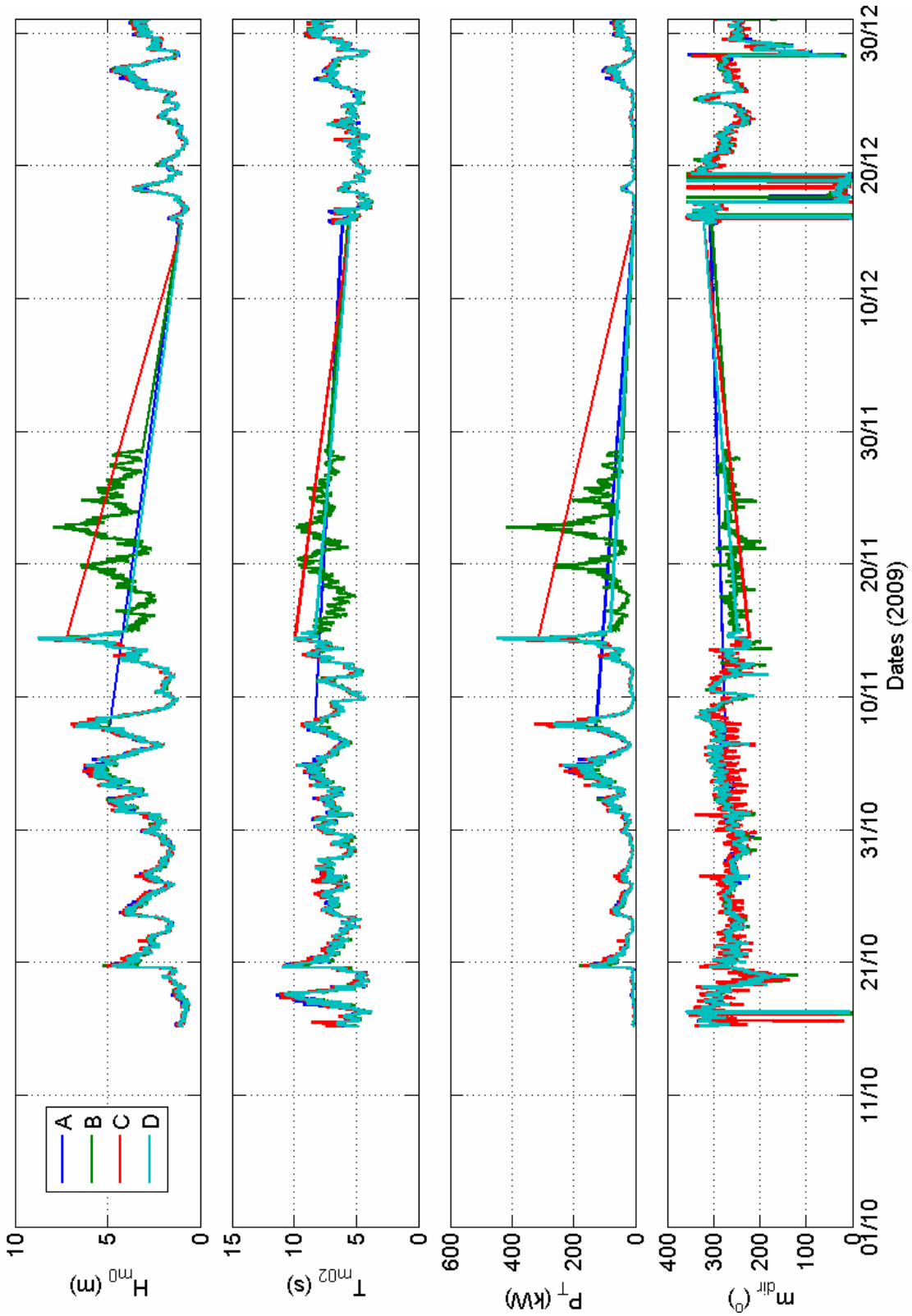

Figure E.1: Time-series of H_{m0}, T_{m02}, P_T and m_{dir} for Oct - Dec 2009, estimated using the standardised processing methods.

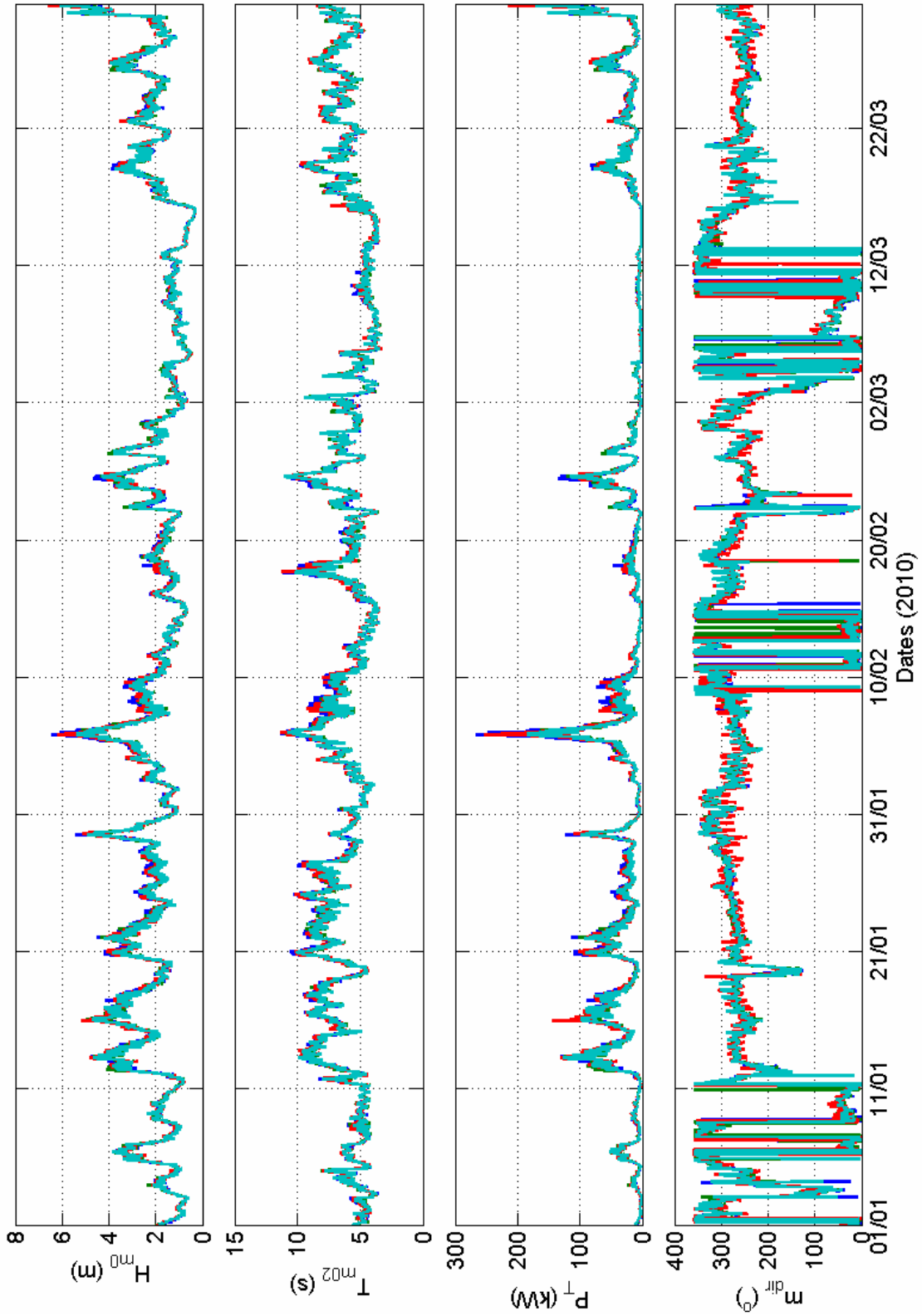

Figure E.2: Time-series of H_{m0}, T_{m02}, P_T and m_{dir} for Jan - Mar 2010, estimated using the standardised processing methods.

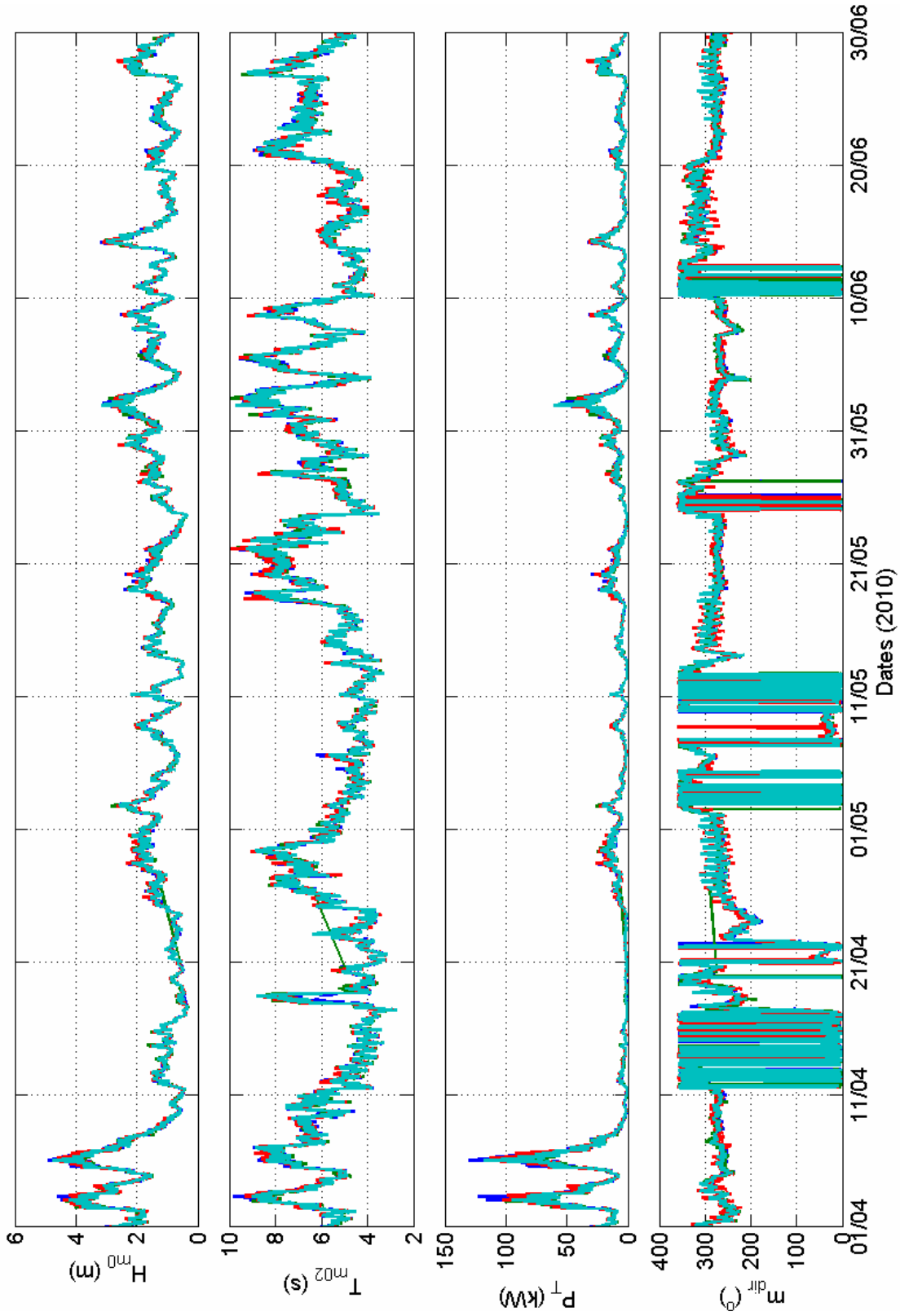

Figure E.3: Time-series of H_{m0}, T_{m02}, P_T and m_{dir} for Apr - Jun 2010, estimated using the standardised processing methods.

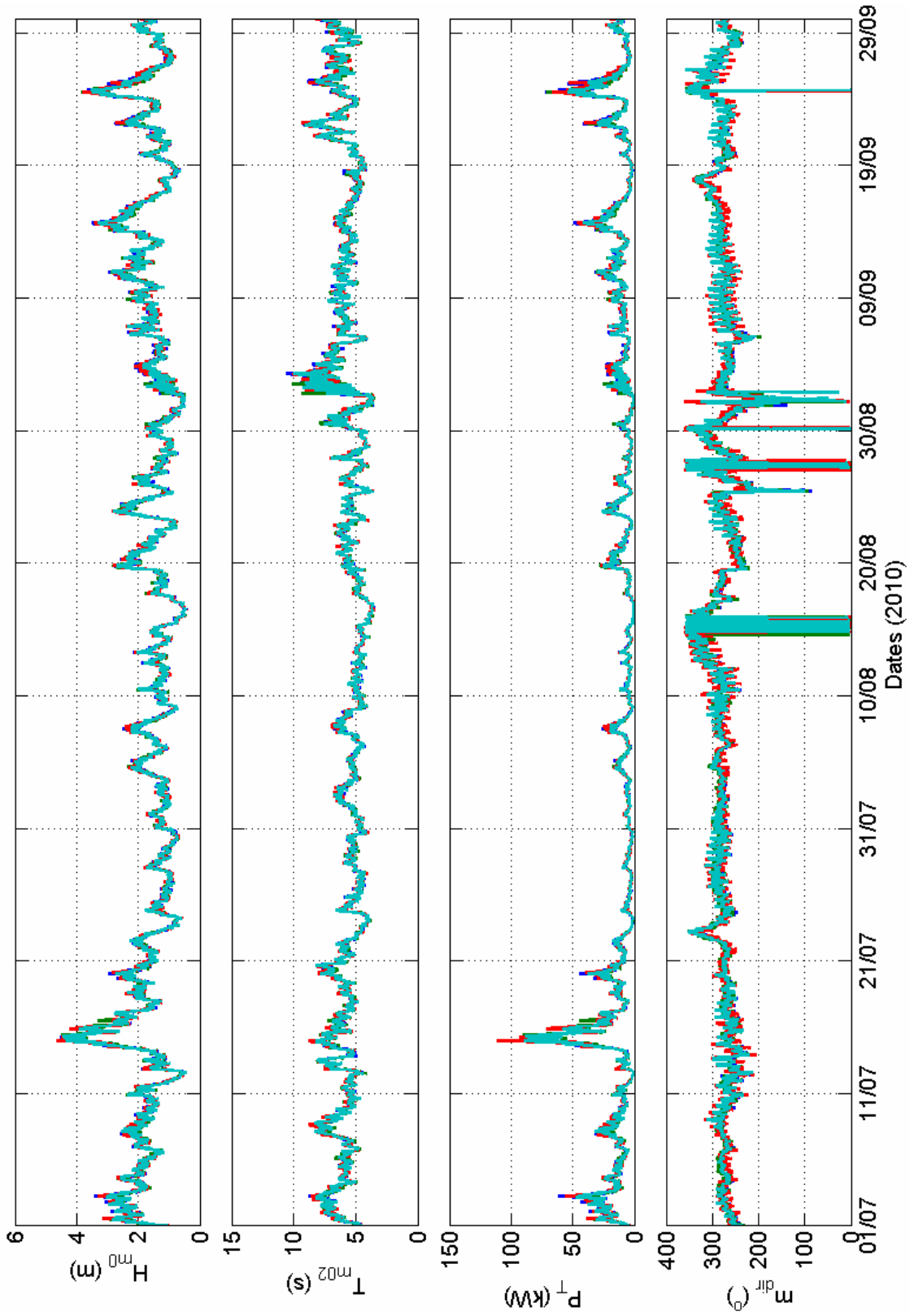

Figure E.4: Time-series of H_{m0}, T_{m02}, P_T and m_{dir} for Jul - Sep 2010, estimated using the standardised processing methods.

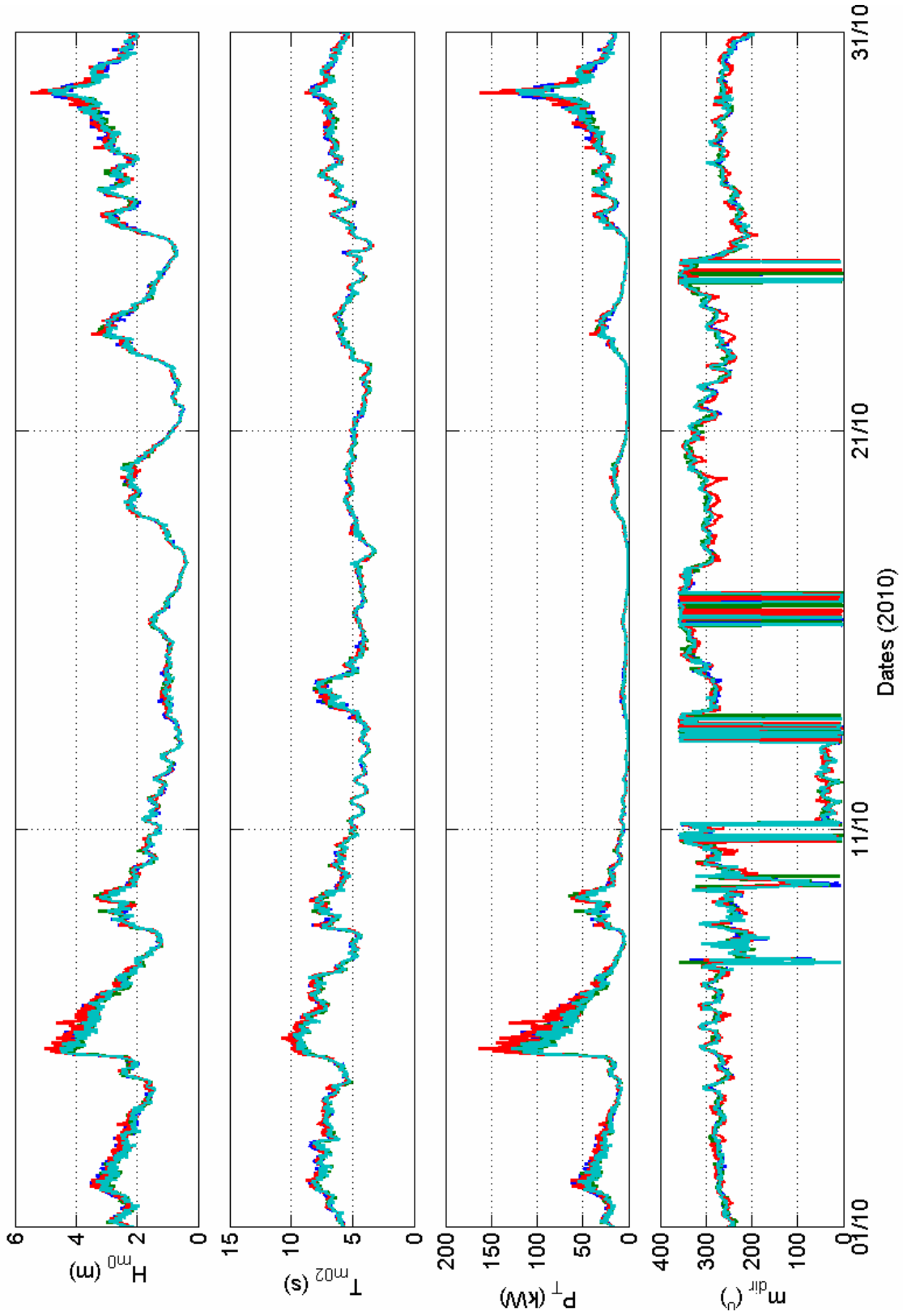

Figure E.5: Time-series of H_{m0}, T_{m02}, P_T and m_{dir} for Oct 2010, estimated using the standardised processing methods.

Appendix F

Measured spectra and variability in associated parameters, monthly plots

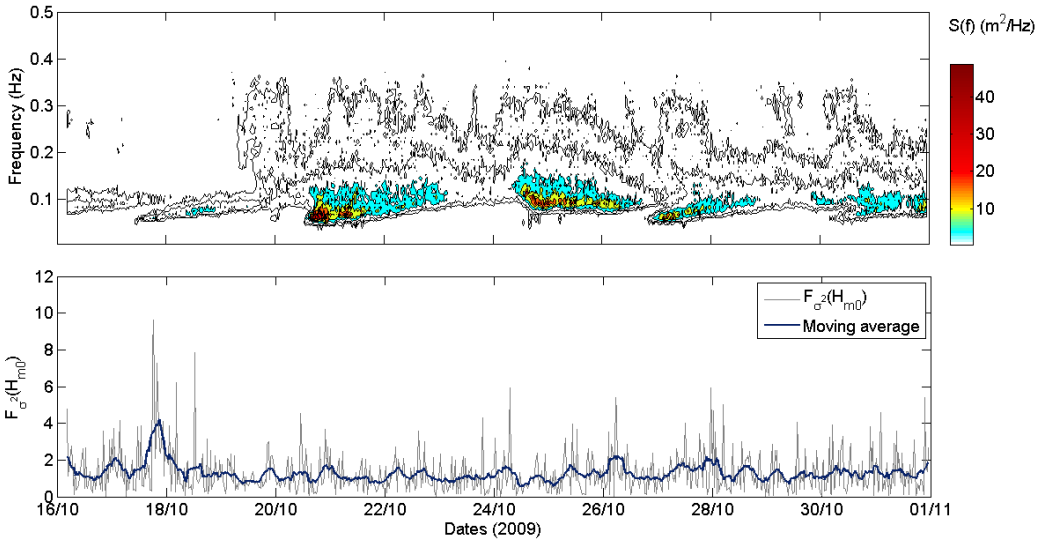

Figure F.1: Spectral energy density measured at each frequency during October 2009, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

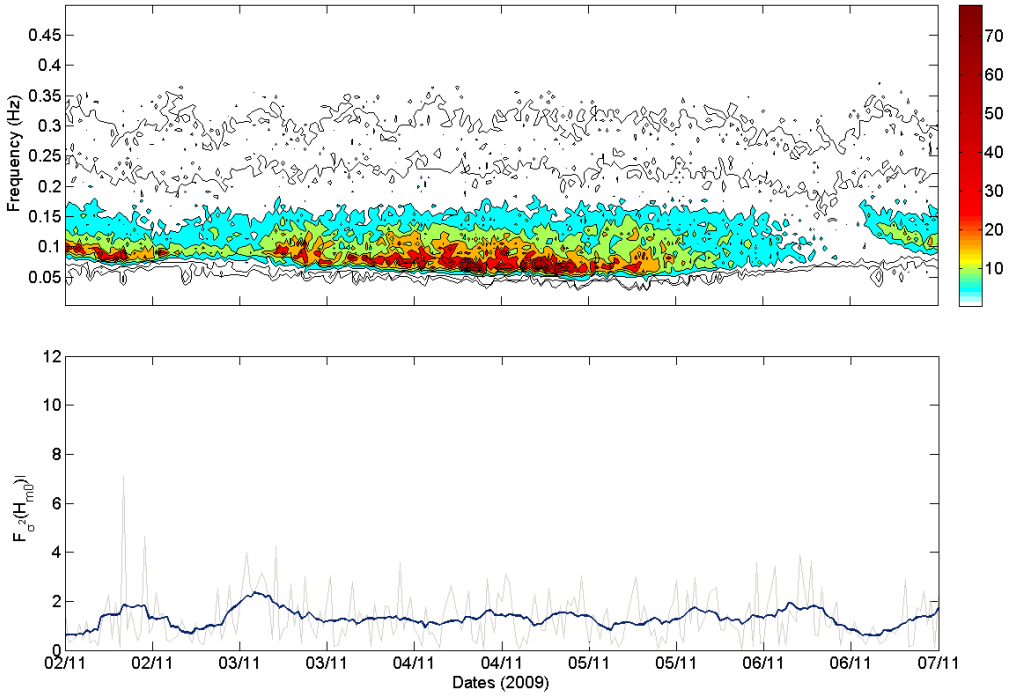

Figure F.2: Spectral energy density measured at each frequency during November 2009, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

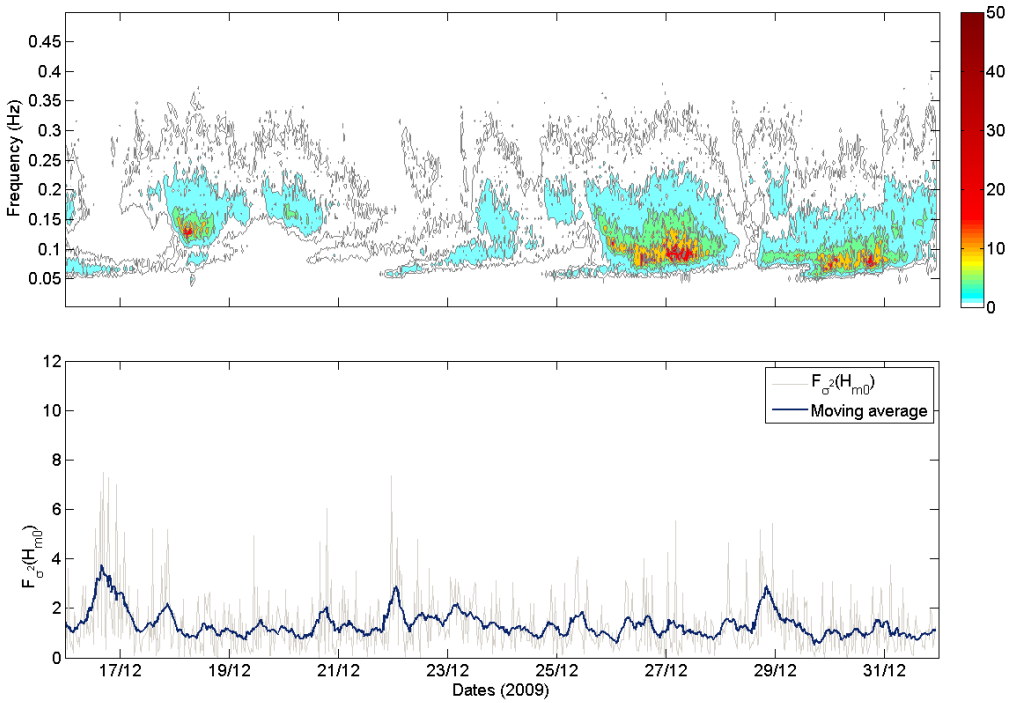

Figure F.3: Spectral energy density measured at each frequency during December 2009, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

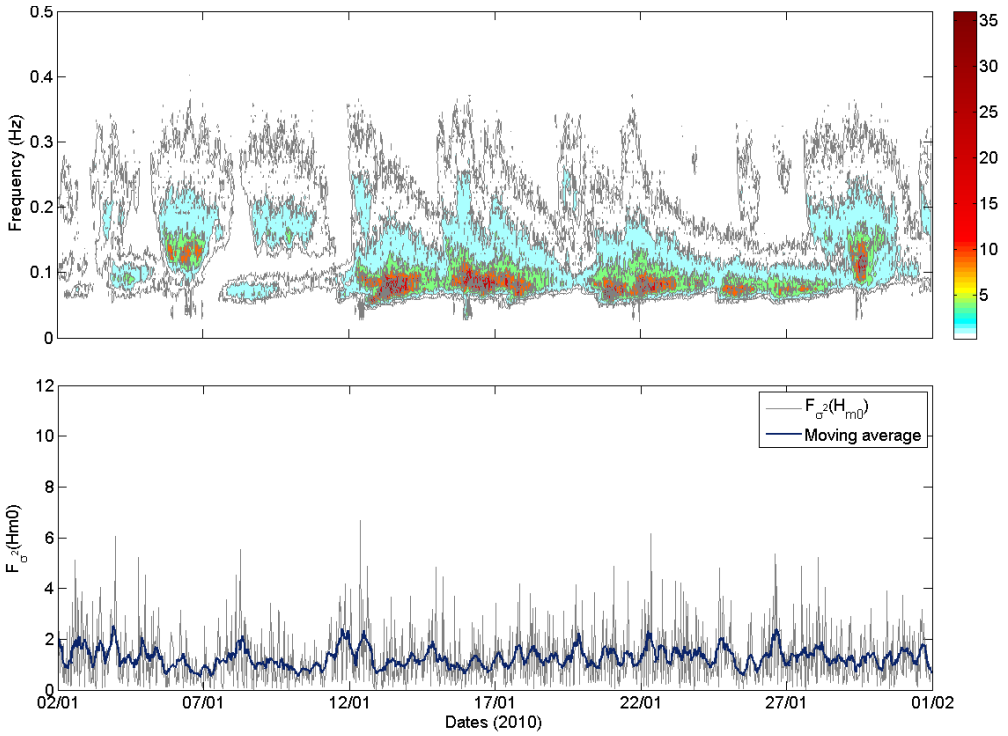

Figure F.4: Spectral energy density measured at each frequency during January 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

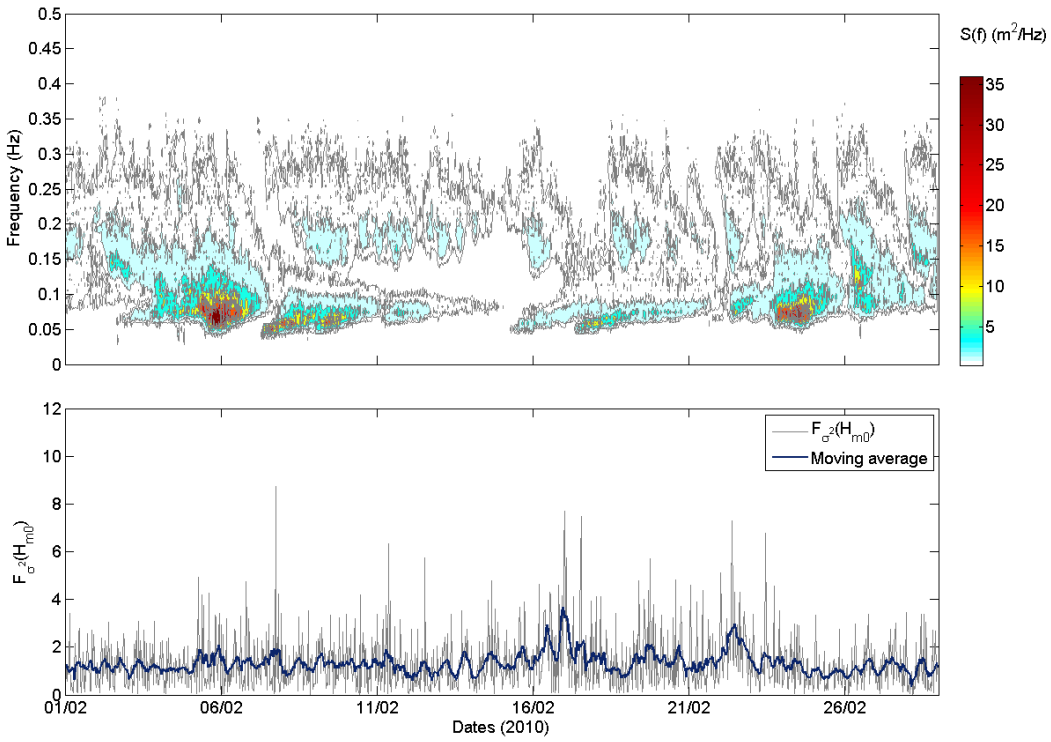

Figure F.5: Spectral energy density measured at each frequency during February 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

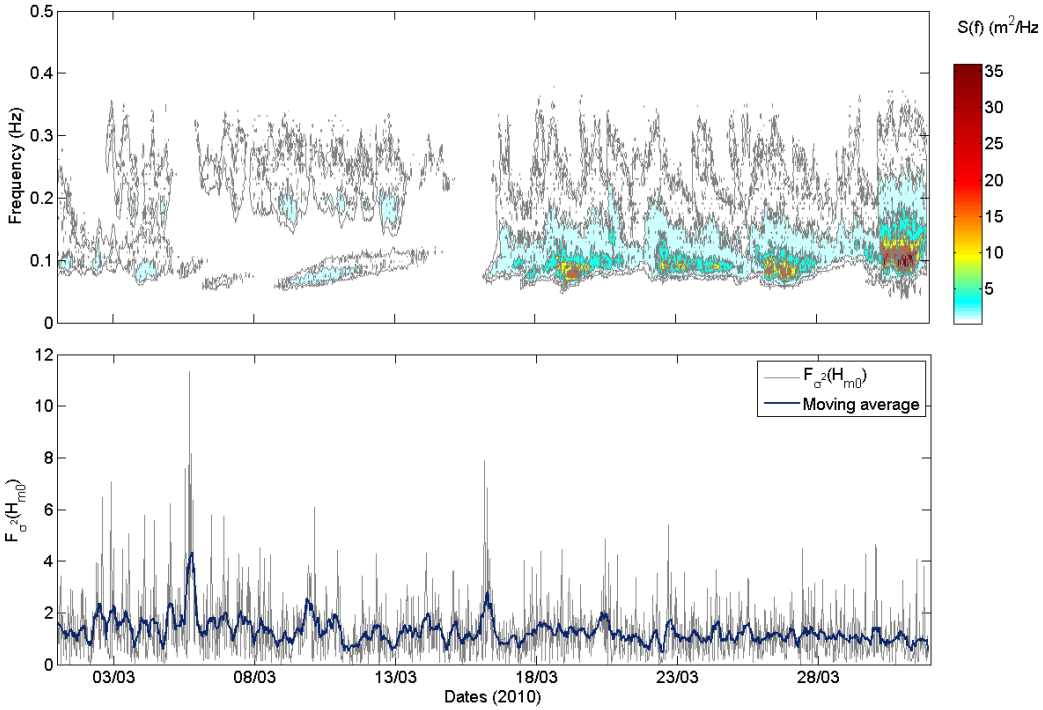

Figure F.6: Spectral energy density measured at each frequency during March 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

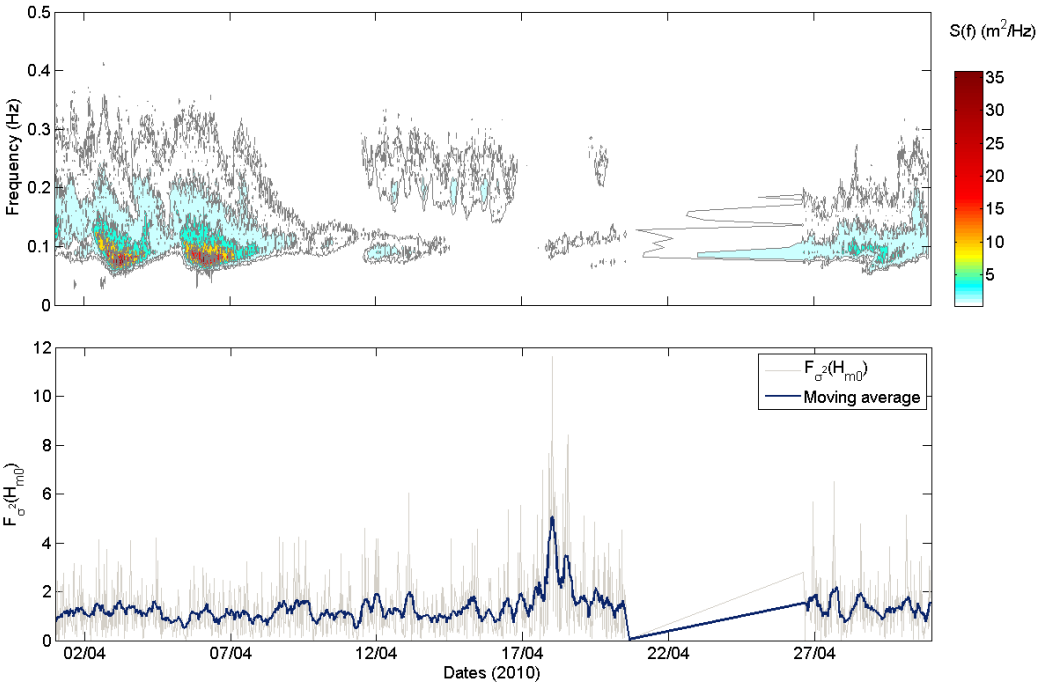

Figure F.7: Spectral energy density measured at each frequency during April 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

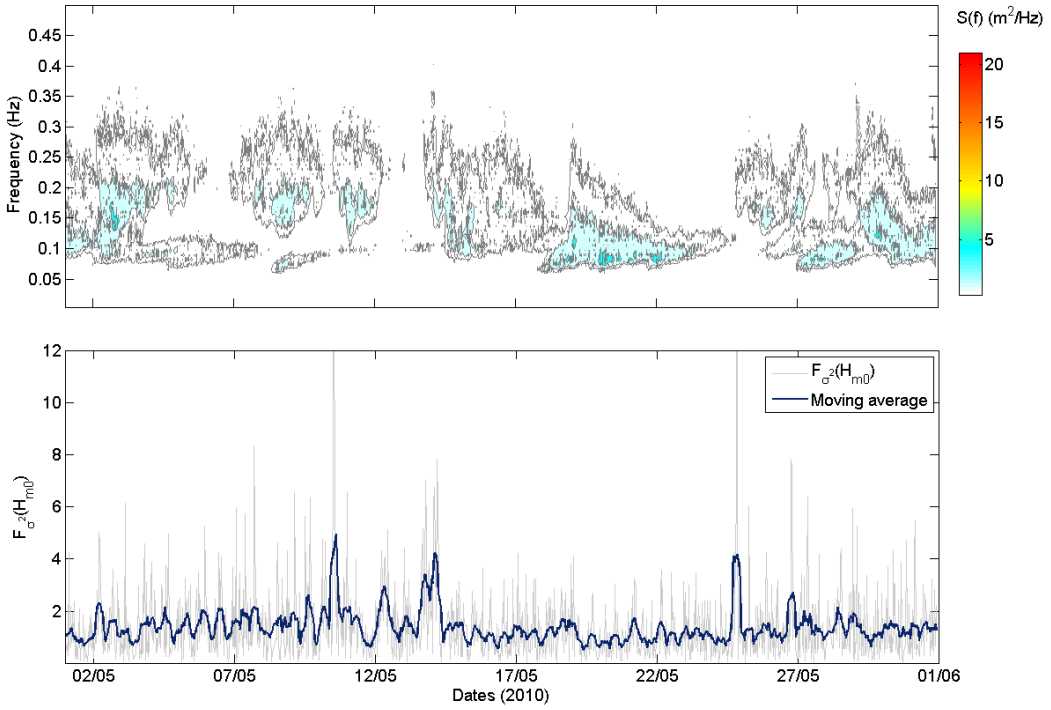

Figure F.8: Spectral energy density measured at each frequency during May 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

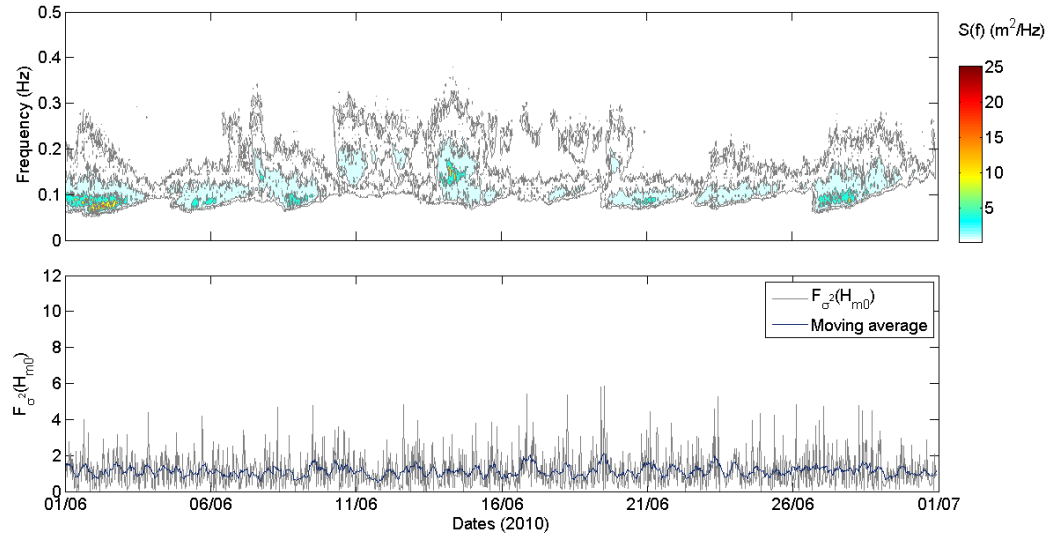

Figure F.9: Spectral energy density measured at each frequency during June 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

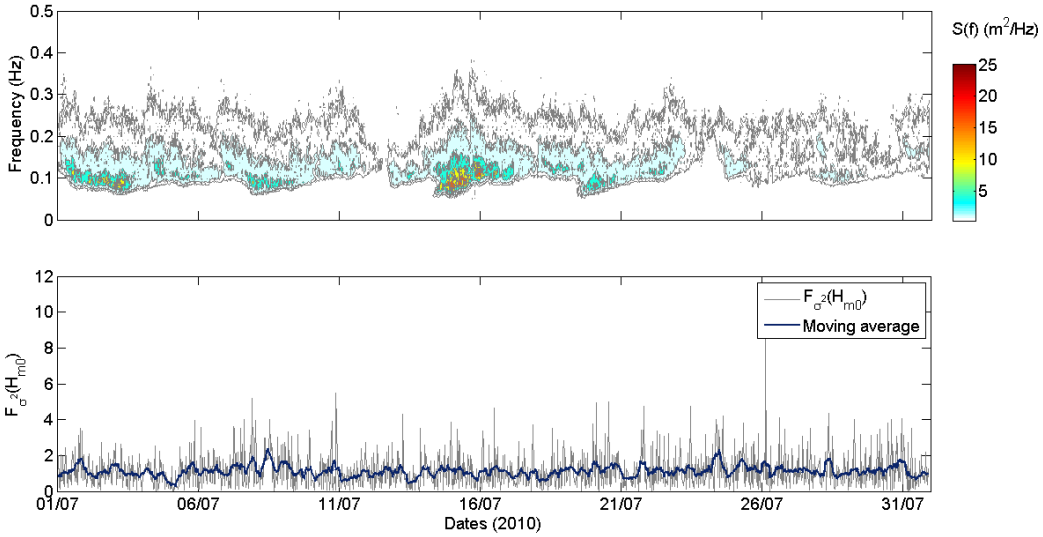

Figure F.10: Spectral energy density measured at each frequency during July 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

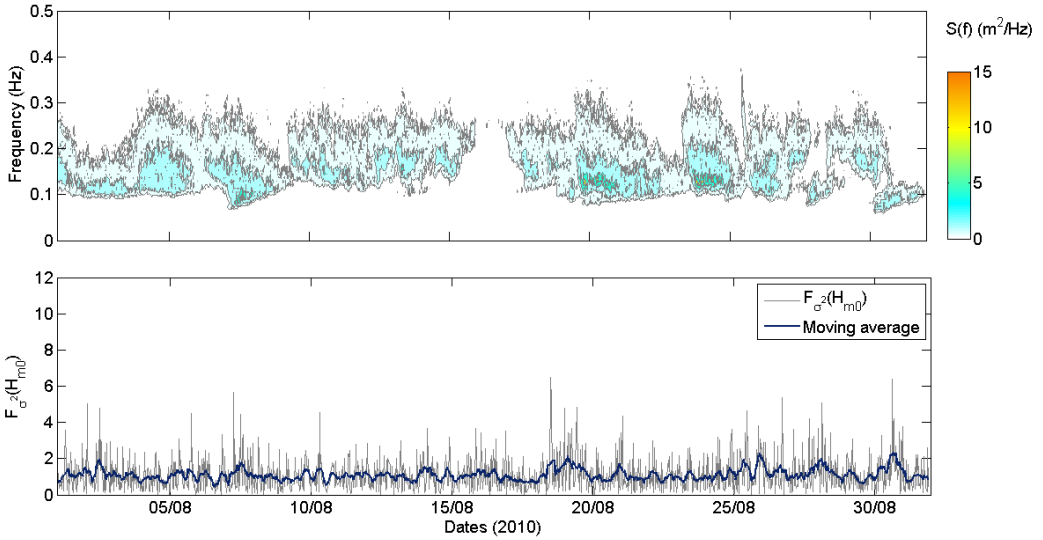

Figure F.11: Spectral energy density measured at each frequency during August 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

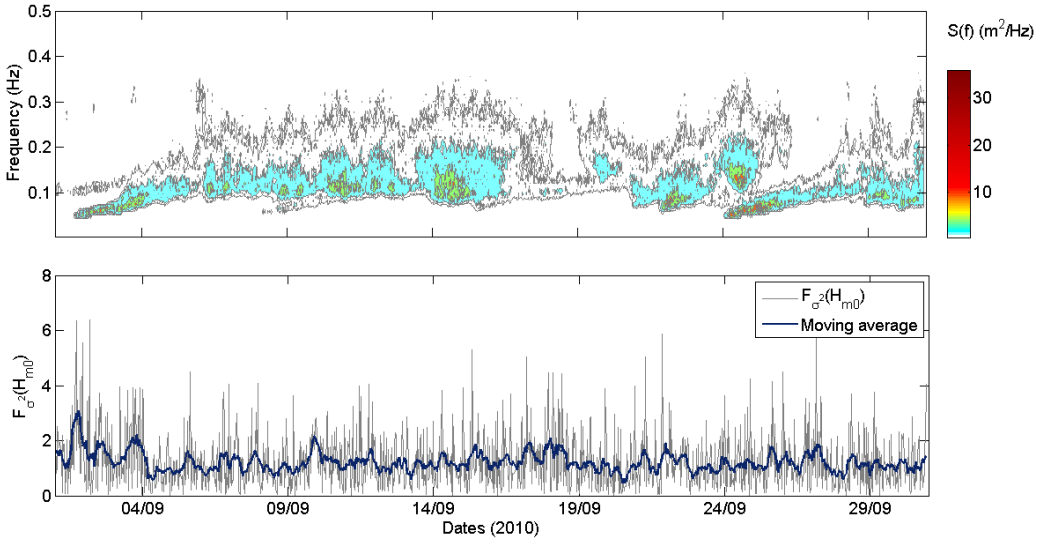

Figure F.12: Spectral energy density measured at each frequency during September 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

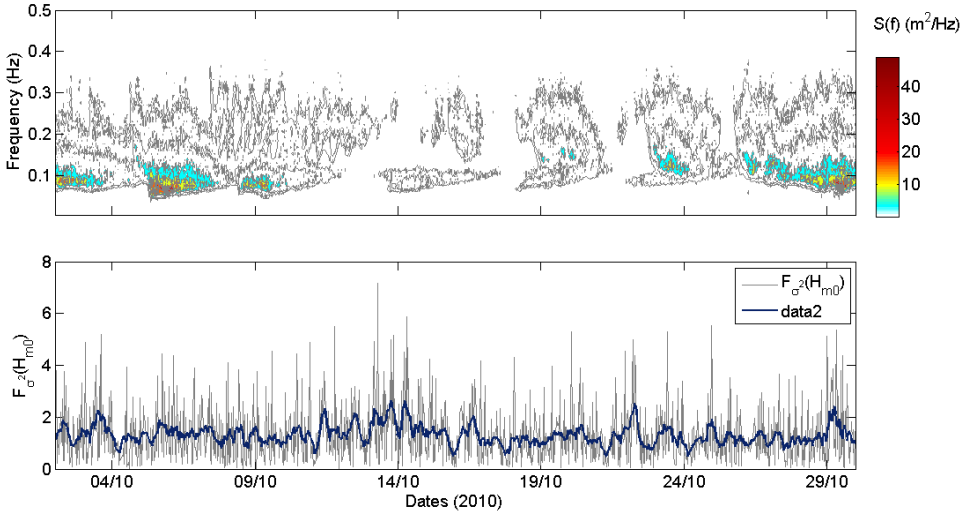

Figure F.13: Spectral energy density measured at each frequency during October 2010, and a time series of the measured variability in significant wave height, $F_{\sigma^2}(H_{m0})$.

Bibliography

Allender, J., Audunson, T., Barstow, S. F., Bjerken, S., Krogstad, H. E., Steinbakke, P., Vartdal, L., Borgman, L. E., Graham, C., 1989. The WADIC project: A comprehensive field evaluation of directional wave instrumentation. Ocean Engineering 16 (5-6), 505 – 536.

Barrett, S., Ashton, I., Lewis, T., Smith, G., 2009. Spatial & spectral variation of seaways. Proceedings of the 8th European Wave and Tidal Energy Conference, Uppsala, Sweden.

Beels, C., Troch, P., De Visch, K., Kofoed, J., De Backer, G., 2010. Application of the time-dependent mild-slope equations for the simulation of wake effects in the lee of a farm of wave dragon wave energy converters. Renewable Energy 35, 1644–1661.

Bendat, J. S., Piersol, A. G., 2000. Random data: analysis and measurement procedures. Wiley.

BSI, 1984. British standard code of practice for maritime structures, Part 1. General criteria. Tech. rep., British Standards Institution.

Carroll, R., Ruppert, D., 1988. Transformation and weighting in regression. Chapman and Hall.

Carter, D. J. T., Challenor, P. G., 1981. Estimating return values of environmental parameters. Quarterly Journal of the Royal Meteorological Society 107, 259–266.

CDIP, November 2003. Coastal data information program. online at http://cdip.ucsd.edu.

Clement, A., McMullen, P., Falcao, A., Fiorentino, A., Gardner, F., Hammarlund, K., Lemonis, G., Lewis, T., Neilsen, K., Petroncini, S., Pontes, M.-T., Schild, P., Sjostrom, B.-O., Sorensen, H. C., Thorpe, T., 2002. Wave energy in Europe: Current status and perspectives. Renewable and Sustainable Energy Reviews 6, 405–431.

Cotton, P., Carter, D., 1994. Cross calibration of TOPEX, ERS1, and Geosat wave heights. Journal of Geophysical Research 99, 25025–25033.

Cruz, J. (Ed.), 2008. Ocean Wave Energy. Springer.

Datawell, April 2005. Waverider Reference Manual. Datawell BV Laboratory for instrumentation.

Dean, R., Dalrymple, R., 1984. Water wave mechanics for engineers and scientists. Prentice Hall.

Draper, N. R., Smith, H., 1981. Applied regression analysis. Wiley.

DTI, 2004. Atlas of UK marine renewable energy resources. Tech. rep., DTI.

Elgar, S., Guza, R., Seymour, R., 1985. Wave group statistics from numerical simulations of a random sea. Applied Ocean Research 7, 2, 93–96.

EMEC, Jun 2010. http://www.emec.org.uk/.

Ferreira, J. A., Guedes-Soares, C., 2000. Modelling distributions of significant wave height. Coastal Engineering 40, 361–374.

Fowler, J., Cohen, L., Jarvis, P., 1998. Practical statistics for field biology. Wiley.

Goda, Y., June 1977. Numerical experiments on statistical variability of ocean waves. Report of the Port and Harbour Research Institute 16, 3:26.

Guedes-Soares, C., 1984. Representation of double-peaked sea wave spectra. Ocean Engineering 11, 185–207.

Halcrow, June 2006a. Wave Hub development and design phase coastal processes study report. www.wavehub.co.uk.

Halcrow, June 2006b. Wave Hub environmental statement. www.wavehub.co.uk.

Hamilton, G. D., 1992. Measurement of long period, low amplitude swell in the western North Atlantic Ocean. Journal of Atmospheric and Oceanic Technology 9, 645 – 658.

Hasselman, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselman, D., Kruesman, P., Meerburg, A., Muller, P., Olbers, D., Richter, K., Sell, W., Walden, H., 1973. Measurements of wind wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Tech. rep., Deutches Hydrographisches Institut, Hamburg.

Holmes, B., Barrett, S., 2007. Sea and swell spectra. In: Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal.

Holthuijsen, L. H., 2007. Waves in oceanic and coastal waters. Cambridge University Press.

Ingram, D., Smith, G., Bittencourt-Ferreira, C., Smith, H., 2011. Protocols for the equitable assessment of marine energy converters. Tech. rep., University of Edinburgh, School of Engineering.

Joosten, H., May 2006. Wavebuoys and their elastic mooring. International Ocean Systems.

Kalantar, A. H., 1990. Weighted least squares evaluation of slope from data having errors in both axes. Trends in Analytical Chemistry 9, 5, 149 – 151.

Krogstad, H. E., Wolf, J., Thompson, S. P., Wyatt, L. R., 1999. Methods for intercomparison of wave measurements. Coastal Engineeering 37, 235–257.

Little, A. D., 2005. Wave Hub: Summary business case. Tech. rep., SWRDA.

Longuet-Higgins, M. S., 1952. On the statistical distribution of the heights of sea waves. Journal of Maritime Research 11, 1245–1266.

Longuet-Higgins, M. S., 1975. On the joint distribution of wave periods and amplitudes of sea waves. J. Geophys. Res. 80, 2688–2694.

Mackay, E., March 2010. Filter noise in datawell waverider data. Pers-comm.

Mackay, E. B. L., 2009. Wave energy resource assessment. Ph.D. thesis, University of Southampton.

Mackay, E. B. L., Bahaj, A. S., Challenor, P. G., 2010. Uncertainty in wave energy resource assessment. Part 2: Variability and predictability. Renewable Energy 35, 1792–1808.

Mathworks, 2007. Matlab help pages. http://www.mathworks.com.

Michael, J. R., 1983. The stabilized probability plot. Biometrika 70, 11–17.

Millar, D. L., Smith, H. C. M., Reeve, D. E., April 2007. Modelling analysis of the sensitvity of shoreline change to a wave farm. Ocean Engineering 34, 884–901.

Montgomery, D., 1982. Introduction to linear regression analysis. Wiley.

Mueller, M. A., August 2008. UKERC energy research landscape: Marine renewable energy. Retrieved 08/10/2008, from http://ukerc.rl.ac.uk.

NOAA, July 2005. Second workshop report on the quality assurance of real-time ocean data,. on-line at :http://nautilus.baruch.sc.edu/twiki/bin/view, CCPO Technical Report Series No. 05-01.

Oceanor, 2009. Wavesense manual. Fugro, Oceanor.

Oceanor, 2010. Company website. http://www.oceanor.no/.

Ochi, M. K., 1998. Ocean waves: The stochastic approach. Cambridge University Press.

Phillips, O., 1957. On the generation of waves by turbulent wind. Journal of Fluid Mechanics 2, 417–445.

Pierson, W. J., Moskowitz, L., 1964. A proposed spectral form for fully developed wind seas based on the similarity theory of Kitaigorodskii. Journal of Geophysical Research 69, 5181–5203.

Pitt, E. G., 2006. Review of wave measurement technology. Applied Wave Research.Available via www.wavehub.co.uk.

Pitt, E. G., 2009. Assessment of wave energy resource. Tech. rep., EMEC.

Pitt, E. G., Saulter, A., Smith, H., 2006. The wave power climate at the Wave Hub site. Tech. rep., Applied Wave Research.

POL, 2010. Polpred website. http://www.pol.ac.uk/appl/polpred.html.

Previsic, M., Bedard, R., Hagerman, G., June 2004. E2I EPRI assessment, offshore wave energy conversion devices. Report, EPRI.

Rye, H., Svee, R., July 1976. Parametric representation of a wind-wave field. In: Proceedings of the Fifteenth Coastal Engineering Conference, Honolulu, HI. ASCE, pp. 183–201.

Sanmuganathan, V., September 2009. Seawatch Mini II buoy, User Manual. Fugro Oceanor.

Saulnier, J. B., Ricci, P., Pontes, M. T., de O Falcao, A. F., 2007. Spectral bandwidth and WEC performance assessment. In: Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal.

Saulnier, J.-B. M. G., Maisondieu, C., Ashton, I., Smith, G., 2011a. Sea state directional analysis near the wave hub test facility. Applied Ocean Research In press.

Saulnier, J.-B. M. G., Maisondieu, C., Ashton, I., Smith, G., 2011b. Wave Hub test facility: Sea state directional analysis from an array of 4 measurement buoys. In: Proceedings of 9th European Wave and Tidal Energy Conference, Southampton UK.

Smith, G., Venugopal, V., Bullen, C., Norris, J., Fasham, J., 2006a. A wave resource assessment for the the European Marine Energy Centre, Orkney. Tech. rep., EMEC.

Smith, G. H., Taylor, J., 2007. Wave energy device performance protocol. Tech. rep., Heriot Watt.

Smith, G. H., Venugopal, V., Fasham, J., 2006b. Wave spectral bandwidth as a measure of available wave power. In: Proceedings of 25th International Conference on Offshore Mechanics and Arctic Engineering.

Smith, H., Haverson, D., Smith, G., Cornish, C., Baldock, D., January 2011. Assessment of the wave and current resource at the wave hub site. Tech. rep., University of Exeter and Marine Energy Matters.

Smith, H. C. M., Millar, D., Reeve, D. E., 2007. Generalisation of wave farm impact assessment on inshore wave climate. In: Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal. CSM.

Smith, R. J., 2009. Use and misuse of the reduced major axis for linefitting. American Journal of Physical Anthropology 140, 3, 476 – 486.

Snodgrass, F. E., Groves, G. W., Hasselmann, K. F., Miller, G. R., Munk, W. H., Powers, W. H., 1966. Propagation of ocean swell across the pacific. Philosophical Transactions of the Royal Society of London 259 - 1103, 431–497.

Sova, M., 1995. The sampling variability and the validation of high frequency wave measurements of the sea surface. Ph.D. thesis, Univ. Sheffield.

Sova, M., Wyatt, L. R., 1995. Spatial and temporal variability in ocean wave measurement. Tech. rep., Applied mathematics section, University of Sheffield.

Tucker, M. J., 1993. Recommended standard for wave data sampling and near real-time processing. Ocean Engineering 20, 459–474.

Tucker, M. J., Pitt, E. G., 2001. Waves in ocean engineering. Vol. 5 of Elsevier ocean engineering book series. Elsevier.

Vassie, J. M., Woodworth, P. L., Holt, M. W., 2004. An example of north Atlantic deep ocean swell impacting Ascension and St. Helena islands in the central south Atlantic. Journal of Atmospheric and Oceanic Technology 21, 1095–1103.

Venugopal, V., Smith, G., 2007. Wave climate investigation for an array of wave power devices. In: Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal.

WAFO, August 2000. WAFO, A Matlab Toolbox for analysis of random waves and loads. Lund University, centre for mathematical studies, 2nd Edition.

WavePlam, 2009. A cautiously optimistic review of the technical status of wave energy technology. Tech. rep., Ente Vasco de la Energia.

Woolf, D. K., 2002. Sensitivity of power output to wave spectral distribution. Tech. rep., Seapower Europe.

Work, P., 2008. Nearshore directional wave measurements by surface-following buoy and acoustic doppler current profiler. Ocean Engineering 35, 727–737.

Wyatt, L., Thompson, S., Burton, R., 1999. Evaluation of high frequency radar wave measurement. Coastal Engineering 37, 259–282.

Young, I. R., 1986. Probability distribution of spectral integrals. Journal of Waterway, Port, Coastal and Ocean Engineering 112, 338–341.