University of Exeter
Department of Computer Science

Sea state from monoscopic ocean video in real environments

Antonis Loizou

Thesis supervisor: Dr Jacqueline Christmas

Submitted by Antonis Loizou, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science, March, 2021.

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.

Signed: .....

March 2021

Abstract

Video of the ocean surface is used as a means for estimating useful information about the scene. A methodology is first introduced for approximating the pixel to metre scale from high-scale videos of the ocean, such as from an aeroplane. Radar images are used for testing. The temporal and spatial domains are associated through the phase modulation of waves, and a process is introduced that selects the waves with the highest energy to be used for estimating the pixel scale. The spatial information is then used with the calculated pixel scale for approximating the sea state.

Due to the difficulty of obtaining high-scale videos, a methodology is then introduced that uses the temporal variation from video, and specifically time series of pixel intensities. It aims to isolate and utilise the temporal variation of the wave field from all other video elements, such as environmental brightness flunctuations. The methodology utilises the Kalman filter and the least squares approximate solution for providing an uncalibrated video amplitude spectrum. A method is proposed for scaling this spectrum to metres with the use of an empirical model of the ocean. The significant wave height is estimated from the calibrated video amplitude spectrum. Videos of the ocean in real environments from a shipborne camera and a tower are used for testing. In both sets of data, *in situ* buoy information is used solely for validation.

The next technique aims to approximate the sea state from the same kind of data, namely videos of the ocean in real environments, without calibrating a video amplitude spectrum. The proposed methodology tracks the principal component of the movement of water in the video, which is speculated to be associated with the dominant frequency of the ocean. To accomplish this, the singular spectrum analysis algorithm and the extended Kalman filter are used. Then, the shape of an empirical spectrum is utilised in order to translate the dominant frequency output into a significant wave height estimation.

The problem of not using ocean theory associated with a particular empirical energy spectrum for calibration is examined in the next methodology. A secondary

oscillatory component from the singular spectrum analysis algorithm is identified with the incorporation of the extended Kalman filter. Ocean theory involving the equilibrium range of oceans is used for calibration.

The shipborne videos are used for testing the behaviour of the techniques for approximately the same sea state of 3.1 m to 3.4 m of significant wave height. The tower videos are used for testing the techniques for a variety of sea states ranging between 0.5 m and 3.6 m of significant wave height. From all methodologies, the maximum observed values of root mean square error 0.37 m and of mean absolute percentage error 18 % suggest that the work is promising at estimating these states.

Contents

List of Figures

List of Tables

Acronyms

KE	Kinetic energy
MAE	Mean absolute error
MAPE	Mean absolute percentage error
PC	Principal component
PE	Potential energy
RC	Reconstructed component
RMSE	Root mean square error
SSA	Singular spectrum analysis

Nomenclature

Ocean theory

A	Amplitude
c_p	Phase speed
E	Energy
f	Temporal frequency
f_m	Dominant or modal temporal frequency
f_s	Spatial frequency
g	Numeric constant for the acceleration of gravity
h_s	Significant wave height
k	Wavenumber
S	Spectral energy
T	Wave period
U	Wind speed
γ	Peak-shape parameter
θ	Direction of the wind
λ	Wavelength
λ_p	Wavelength in pixels
λ_m	Wavelength in metres
ρ	Water density
ϕ	Phase
ω	Angular frequency
$\bar{\omega}$	Mean angular frequency
ω_h	Angular frequency associated with a wave of finite depth h
ω_m	Dominant or modal angular frequency

Kalman filter

B	Control input model
F	Transition model
H	Observation model
K	Gain matrix
Res	Residual
u	Input variable
V	Measurement noise matrix
W	Process noise matrix
x	True state
$\bar{x}$	Projected state
$\hat{x}$	Predicted state
z	Observed state
ϵ	Process noise
ζ	Observation noise
Φ	Fundamental matrix

Video

a_{er}	Amplitude from the equilibrium range theory
a_p	Amplitude of the peak from the video amplitude spectrum
a_{pm}	Amplitude from the Pierson-Moskowitz spectrum
a_u	Amplitude from uncalibrated video amplitude spectrum
c	Column number
L	Frame width in pixels
N	Number of frames
r	Row number
t_{max}	Length of video in seconds
v	Frame rate
α	Amplitude scale
Δt	Time between two successive frames
η	Pixel to metre scale
ξ	Intermediate value for calibration of the video amplitude spectrum

Chapter 1

Introduction

In this work, video of the ocean surface in real environments is investigated as a means for getting useful information about the scene and the sea state. The study and modelling of the sea surface in deep and shallow water is of great importance for the construction of offshore structures and the execution of maritime operations [73][42]. Any system design is an optimisation process [24]. In the case of designing and building harbours and other offshore structures, the effects of waves are a primary constraint [42]. This is why knowledge of the state of the ocean in an area over preferrably long periods of time is important for the design of these structures.

In the case of maritime operations, in some cases operating limits associated with the sea state are clearly defined. For example, helicopter operations on ships are considered high risk operations, where clearly established procedures are established. These procedures include restrictions associated with the state of the sea, wind speed and direction [96]. This information is usually passed to the helicopter providing service before any type of operation is executed. The sea state modelling for deep water can also be used for the efficient construction of sea vessels and platforms [19] and for improving the efficiency of wave energy converters [4][173]. Sea vessels are vulnerable to damage due to the unpredictable nature of the seas. This brings the need for instruments that can measure the ocean surface and techniques and methods that can use this data in order to provide accurate and reliable information about the state of the sea.

Operative limits in the form of maximum allowable sea state are defined in the oil and gas industry [43]. These operative limits are defined in terms of maximum allowable significant wave height or maximum allowable vessel motion. For the execution of operations such as berthing, vessel manoeuvring, loading and offloading, storage etc., the operational teams in oil and gas platforms require reliable sea state information [44]. The vulnerability of these terminals and their operations is reduced with the definition of decision-making processes that take information of the sea state as input [43]. If not appropriate procedures are designed for interruption of activities during adverse weather conditions, infrastructure damage and human life risk are possible. Even in periods of 'good' season, severe weather conditions could be encountered in the locations of gas and oil platforms in the North Sea, Brazilian offshore and South China sea [43].

Reliable sea state information is also essential for the construction of oil and gas platforms. The building operations are defined in terms of sea state. Specifically, the operations for the construction of these infrastructures are defined in terms of maximum allowable sea state [43]. Companies in the oil and gas industries are especially interested in minimising the stand-by time of interruption for the construction of these platforms due to the weather conditions [43]. In these cases, a careful study of state of the ocean area over long periods of time is required.

Operations of the coast guard of various countries could be improved with the use of information of the state of the sea. The coast guard has a mission of providing timely, useful and actionable information for maritime operations to decision makers and maritime partners [33]. Accurate and recent knowledge of the state of the sea is an essential part of such information.

Sea state information could also enable the coast guard to make better decisions in regards to asset allocation for their missions. For example, the U.S. coast guard deploys major cutters (their commissioned vessels) and sometimes aircrafts as allocated assets for living marine resources (LMR) missions due to geography, weather and sea state challenges in some offshore fishing grounds [32]. These major cutters have greater endurance capabilities than patrol boats. Knowledge of the sea state in

these grounds could help the coast guard provide more efficient asset allocation for the LMR missions. For example, in lower sea states (and other favourable conditions in terms of geography and weather), patrol boats could be used as a surface asset of choice, which are smaller, easier to maintain and require a smaller workforce. In such a scenario, the coast guard's major cutters and aircrafts could be allocated for more critical operations.

The coast guard's search and rescue (SAR) missions could also be improved in terms of asset allocation when sea state information is available. Operational commanders evaluate numerous factors for the determination of the proper resource allocation for missions, so that the possibility of these missions being executed with success is maximised [32]. Sea state information could be used by the operational commanders for making better decisions in terms of asset allocation. In this way, the possibility of success could be improved, while at the same time the asset allocation could be more efficient. For example, in lower sea states the commanders may not decide to send the guard's major cutters for a SAR operation and decide to send smaller vessels, such as station small boats or patrol boats. Major cutters are the preferred allocated assets for long transit distances, large search areas and severe weather conditions [32].

The term sea state refers to a set of statistics describing the conditions of the sea, such as the significant wave height, dominant period, energy spectrum and wind speed. The sea state can be estimated with *in situ* devices, such as wave buoys and tidal gauges [20][103]. The ownership and maintenance of wave buoys is expensive. Severe weather conditions may cause damages to these devices. With the advancement of technology and computational algorithms, the remote sensing of the ocean is made possible. For example, stereophotography has been widely used for acquiring a highly detailed representation of the ocean [17][167][39]. The information from radars on ships or satellites are also used for getting detailed sea state information [105][10]. Remote sensing is also done with images or videos of the ocean [39][34]. In some cases, *in situ* devices are used for calibration, for example [31]. The use of images and video has the advantage of being inexpensive, widely available, and requiring little or no

calibration.

This work investigates whether and to what extent it is possible to get the sea state and other useful information about the scene with the use of video from a single uncalibrated camera in real environments. Using images and videos of the ocean for getting the sea state is not a new concept. Dating back to the early 1970s [112], the percentage of whitecaps from photographs was used as an indication of the wind speed. The geometric configuration of the scene and known physics of light reflection and refraction have also been used in the literature in order to get the sea state [152][154]. In [34], the authors used the sun's glitter pattern from aerial photographs. Research also includes the use of the polarisation property of light for estimating the slope field (for example [176][97]).

With the advancement of computational algorithms, attention has been given into acquiring the sea state with the emphasis being made on the methodology used and not the geometry of the scene. One example is with the use of the PIV algorithm in order to track high intensity values from video of the ocean in the nearshore associated with water crests [51]. Some research involves the use of single uncalibrated cameras for acquiring the sea state. For example, the work in [178] uses video data in shallow water in order to get analysis of the directional wave spectrum. The methodology involves the extended maximum likelihood method and the Bayesian directional method. Wave gauge measurements are used for calibration. In [175], the dominant period is estimated from video in the surf zone by using the information of the wave crests. The work in [146] utilises the phase speed information from pixel time series and the Phillips ocean energy spectrum for the same purpose. An extension to this work is presented in [102], where large scale images from radar and the spatial variation are used for getting the sea state. The work in [133] uses large scale airborne images of the ocean and the dominant frequency is determined as the centroid of a selection of frequencies near the peak frequency obtained with an iterative thresholding process.

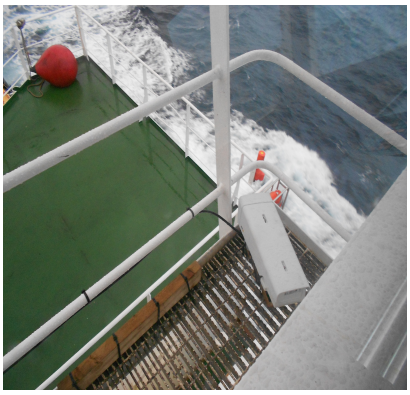
Additionally to techniques that use machine learning algorithms for the estimation of the state of the ocean surface, machine learning is also used in research for

underwater image analysis. For example, deep learning is applied for object detection and classification, such as for fish [145] and seagrass [156]. Another field of research in underwater image processing is the automation of image annotation [180]. In this case, annotation is done in a semi-automated manner, where the system generates candidates of labels that can then be manually selected. One main problem in underwater image annotation is the lack of sufficiently annotated data to be used for learning. An approach proposed to overcome this is with the task of transfer learning [122], where the weights from the learning of one dataset are used for training another dataset. Knowledge transfer is also used in research for improving the image annotation task. In [142], the knowledge obtained from a simulated environment is used for learning in the real environment. The performance of deep learning can also be improved with the use of physics based normalisation and data augmentation [166] and georeference regularisation, physics based colour correlation and image rescaling [174].

In regards to underwater research, statistical approaches have been proposed for the analysis of underwater measurements, such as temperature and salinity. This data can be obtained with the use of autonomous underwater vehicles, which travel the underwater at various depths and provide this information through sensors. For example, in [72] a process-convolution approach is presented for modelling temperature data in the North Atlantic. In [135] space-time modelling is performed on the joint distribution of temperature and salinity levels from data acquired with floats. In [111] a statistical approach is proposed for tracking an unknown three-dimensional oceanic feature with an underwater glider device.

This work presents techniques that utilise video from a single uncalibrated camera in real environments as input. The aim is to estimate the state of the ocean surface. The techniques are introduced in the main chapters. Machine learning algorithms such as the Kalman filter (linear and extended) and the singular spectrum analysis (SSA) are utilised with the aim of isolating and utilising the information from video that is associated with the ocean's movement and for the representation of the information from video in a manner that enables the sea state estimation.

(a)



(b)

(c)

Figure 1-1 Shipborne experimental results. (a) The ship used in the experiments (b) outdoor measuring camera (c) indoor measuring camera.

The structure of this work is as follows:

- Chapter 2 provides background theory of ocean waves that are utilised throughout the work.
- Chapter 3 provides theory of the Kalman filter.
- Chapter 4 provides a literature review of the field of research associated with the present work.
- Chapter 5 describes the experimental data that are utilised and topics associated with the accuracy of the experimental results.
- In Chapter 6, the pixel to metre scale is estimated from large scale video with a technique that iteratively selects only the waves with the highest energy from

Figure 1-2 Frying Pan tower.

the phase modulation of waves from video. Then, the sea state is estimated with the spatial variation from video and the pixel to metre scale estimation of the previous step.

- In Chapter 7, a methodology is proposed that aims to estimate the video amplitude spectrum with the use of the temporal variation of the wave field. This enables the approximation of parameters associated with the state of the ocean.

- In Chapter 8, a methodology is proposed that aims to estimate the significant wave height with the use of the singular spectrum analysis (SSA) algorithm, the nonlinear Kalman filter and ocean theory of an empirical energy spectrum.

- In Chapter 9, a methodology is presented that aims to identify a signal with a frequency higher than the dominant frequency from video. The equilibrium range theory is then used for calibration in order to approximate the significant wave height.

- In Chapter 10, a comparison is given of the experimental results of chapters 7, 8 and 9.

The main chapters of this work are supported by the following publications:

- Chapter 6: A. Loizou and J. Christmas, 'Estimating Pixel to Metre Scale and Sea State from Remote Observations of the Ocean Surface,' IGARSS 2018 -

2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 3513-3516.

- Chapter 7: A. Loizou and J. Christmas, 'Sea state estimation from uncalibrated, monoscopic video,' SN COMPUT. SCI. 2, 328 (2021). https://doi.org/10.1007/s42979-021-00727-0

- Chapter 8: A. Loizou and J. Christmas, 'Sea state from ocean video with singular spectrum analysis and extended Kalman filter,' SIViP (2021), Accepted subject to revisions

For testing the techniques proposed in this work, radar images as well as two sets of video data are used, one from a moving ship and one from a stationary tower. For the shipborne videos, cameras were placed on-board the ship in figure 1-1a, recording the ocean surface in experiments done in 2014 in the North Atlantic ocean. The cameras used for the recording are shown in figures 1-1b and 1-1c.

For the tower videos, video footage from the Frying Pan tower [47], shown in figure 1-2 are used. This tower is a former lighthouse located 34 miles off the coast of Cape Fear, North Carolina. Cameras are placed on the tower showing a panoramic view of the Atlantic ocean. At some time instances, the camera is not rotating, allowing for the acquisition of short duration videos of the ocean surface. Both sets of video have the advantage of having accurate concurrent measurements of the sea state from wave buoys. The wave buoy data are used solely for validation. Calibration is achieved with the incorporation of ocean theory.

Chapter 2

Ocean theory

This work introduces techniques that utilise machine learning and ocean theory concepts in order to provide useful information from video of the ocean. This chapter provides an introduction to the general theory from oceans that are utilised throughout this work. Any additional ocean theory that is unique to each technique is presented in each associated chapter. Topics that are covered include main concepts in oceans (Sect. 2.1), empirical energy spectrum (Sect. 2.2), ocean wave energy and amplitude (Sect. 2.3), dispersion relation (Sect. 2.4), significant wave height (Sect. 2.5) and ocean measurement devices (Sect. 2.6).

2.1 Main concepts

This section provides a short introduction to the different concepts associated with ocean waves.

2.1.1 Wind and swell waves

Ocean waves are created mostly by the transfer of energy by wind [90][159]. The wind can be local or distant. In the case the wind is local, the waves are referred to as wind waves. In the case of distant wind, a constant wind causes waves that are able to travel well beyond their origin that are referred to as swell waves. Swell waves

usually have longer wavelengths than wind waves [155].

2.1.2 Dominant wave

The ocean dominant wave is the sinusoidal element with the highest energy in the ocean energy spectrum [155]. This spectrum is usually acquired with wave buoys.

2.1.3 Deep water

Deep water is defined as the water where the depth is larger than half the wavelength of the dominant wave [36].

2.1.4 Energy

Wave energy (or energy flux) is the energy transfer rate per unit area through the water surface [161][125]. Although it seems to be the case, the waves do not travel a long distance. What is happening is an energy transference between waves on the ocean surface. As a wave moves, there is a constant energy transference between potential and kinetic energy [41]. The kinetic energy of an ocean wave is the energy it possess due to its motion and the potential energy the energy it possess due to its vertical position relative to the water surface. The first law of thermodynamics states that energy is not created or destroyed, rather it is converted from one state to another [148]. This law in oceans is adapted as the total energy of a wave at a given moment being equal to the sum of its kinetic energy and potential energy.

2.1.5 Energy spectrum

If the assumption is taken that the ocean surface may be represented as a sum of sinusoidal waves with given frequencies, amplitudes and directions, the ocean wave energy spectrum can be used for its description [49]. The ocean energy spectrum $S(\omega)$ gives information of wave energy S as a function of wave angular frequency ω.

2.2 Empirical energy spectrum

Since the 1950s, many spectral formulations have been proposed as an answer to the need for wave energy spectra for a range of sea states, for the efficient design of marine systems and offshore structures. These spectra are formed from empirical measurements of the ocean surface. The shape is usually derived by averaging a set of energy spectra of the same sea state acquired with *in situ* measurements. Examples are the Pierson-Moskowitz spectrum, the Jonswap spectrum, the TMA spectrum, the Phillips spectrum and the two-parameter spectrum. In the following text an overview is given of each of these spectra. This work utilises the Pierson-Moskowitz spectrum (Sect. 2.2.1). The rest of the energy spectra are provided as additional informative material, they are not utilised in this work. Information on the reason the Pierson-Moskowitz spectrum is selected instead of the other empirical energy spectra is provided in the Discussion chapter 11 (Sect. 11.1).

2.2.1 Pierson-Moskowitz spectrum

The Pierson-Moskowitz spectrum [127] gives an estimation of the energy distribution across frequencies for various sea states, given the wind speed or dominant angular frequency as input. This empirical ocean energy spectrum was formulated with data acquired from accelerometers installed on weather ships in the North Atlantic ocean. It is applicable on fully developed seas. In fully developed seas, the energy from a wind with a given speed is supplied at its maximum capacity to the ocean as energy transference between waves and wave breaking in a manner that the wave heights do not exceed a set limit for that speed.

The spectral energy in terms of angular frequency is expressed as [127]:

$$S(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left(-\beta\left(\frac{\omega_0}{\omega}\right)^4\right) \tag{2.1}$$

where α and β are constants, $\alpha = 8.1 \times 10^{-3}$, $\beta = 0.74$ and $\omega_0 = g/U$, where U is the wind speed at a height of 19.5 m above the ocean surface. The area under the

spectrum is equal to the integral of the function, that is:

$$\int_0^\infty S(\omega)d\omega = \frac{\alpha U^4}{4\beta g^2} \tag{2.2}$$

It is very useful to have a relationship between the dominant angular frequency and the wind speed. From equation (2.1), the dominant or modal angular frequency ω_m can be expressed as:

$$\omega_m = 0.87\omega_0 \tag{2.3}$$

Since the angular frequency is connected to the frequency with $\omega = 2\pi f$, from equation (2.3) the dominant or modal frequency is found to be equal to:

$$f_m = \frac{0.87\omega_0}{2\pi} \tag{2.4}$$

This work utilises the Pierson-Moskowitz spectrum where an empirical energy spectrum is required for the calibration of the information from video to real units. The choice of this particular spectrum is based on the facts that it is a widely accepted and used customised energy spectrum and that it is relatively simple, since only the parameter of wind speed or dominant frequency is sufficient for generating the shape of the spectrum.

2.2.2 Jonswap spectrum

In 1973, [70] introduced an empirical energy spectrum based on observations from the Joint North Sea Wave Project (JONSWAP). This spectrum introduces the concept of the peakedness of the spectrum, which is the ratio of the base maximum of the spectrum to its approximation value 1 at the same frequency. The energy in terms of angular frequency is given by:

$$S(\omega) = a\frac{g^2}{\omega^5}\exp\left\{-1.25\left(\frac{\omega_m}{\omega}\right)^4 \gamma^{exp\frac{-(\omega-\omega_m)^2}{2(\sigma\omega_m)^2}}\right\} \tag{2.5}$$

where $a = 0.076\overline{x}^{-0.22}$, γ is the peak-shape parameter, $\sigma = 0.07$ for $f \leq f_m$ and 0.09

for $f > f_m$, $f_m = 3.5\left(\frac{g}{\overline{U}}\right)\overline{x}^{-0.33}$, $\overline{x}$ is the dimensionless fetch and is equal to $\overline{x} = \frac{gx}{\overline{U}^2}$, x is the fetch length and $\overline{U}$ is the mean wind speed.

2.2.3 TMA spectrum

The TMA spectrum [77] was designed as an extension to the Jonswap spectrum for finite water depth based on the similarity law shown by [91]. The work in [91] introduces the concept of a transformation factor to extend the work by Phillips [106] in order to include finite water depth. This factor is defined as:

$$\phi(\omega_h) = \left[\frac{k(\omega,h)^{-3}\frac{\partial k(\omega,h)}{\partial\omega}}{k(\omega,\infty)^{-3}\frac{\partial k(\omega,\infty)}{\partial\omega}}\right] \tag{2.6}$$

where ω_h is the angular frequency associated with a wave of finite depth and is equal to $\omega_h = \omega\sqrt{h/g}$, h is the water depth and $k(\omega,h)$ is the wave number associated with a wave of finite depth. This transformation factor is used by [77] and defines the TMA spectrum as:

$$S(\omega) = S_J(\omega)\phi(\omega_h) \tag{2.7}$$

where S_J is the Jonswap energy spectrum equation.

2.2.4 Phillips spectrum

The Phillips spectrum [106] was introduced based on the concepts of wave breaking and the equilibrium range. The energy spectrum is given by:

$$S(k) = A\frac{\exp(-1/(kL)^2)}{k^4}\left|\hat{k}.\theta\right|^2 \tag{2.8}$$

where A is a constant, $L = U^2/g$ is the largest possible waves arising from a continuous wind speed U and θ is the direction of the wind.

2.2.5 Two-parameter spectrum

This energy spectrum [25] represents both fully and partially developed seas. The spectral formulation is given by:

$$S(\omega) = 0.278\frac{\overline{\omega}^4}{\omega^5}\overline{H}^2 \mathrm{e}^{-0.437(\overline{\omega}/\omega)^4} \tag{2.9}$$

where $\overline{\omega}$ is the mean frequency and $\overline{H}$ is the average wave height. This energy spectrum is reduced to the Pierson-Moskowitz spectrum by letting $\omega_m = 0.4\sqrt{g/h_s}$.

2.3 Energy and amplitude

In general wave theory, it is a challenge to get amplitude values from energy values in a wavelength, as the energy of a wave is directly proportional to the square of the amplitude of the wave [52]. The direct correspondence between energy and amplitude is not known.

In the case of ocean waves, the amplitude can be estimated from the energy with the use of the first law of thermodynamics (see Sect. 2.1.4) and the linearized velocity potential for waves. In [29] the author shows that with the use of this theory, the total kinetic energy in deep water for a given wavelength λ can be expressed as:

$$KE = \frac{\lambda}{4}\rho g A^2 \tag{2.10}$$

where ρ is the water density, g the gravitational acceleration and A the amplitude of the wave. Based on the same theory, the total potential energy in deep water in a wavelength can be expressed as:

$$PE = \frac{\lambda}{4}\rho g A^2 \tag{2.11}$$

This means that for a developed two-dimensional wave the kinetic energy density is constant and stationary, and that the potential and its averaged value is equal to half of the total space-averaged wave energy [29]. From the first law of thermodynamics

interpreted as the total energy is the sum of kinetic and potential energy and equations (2.10) and (2.11), the total energy in deep water in a wavelength is equal to:

$$E = \frac{\lambda}{2}\rho g A^2 \tag{2.12}$$

2.4 Dispersion relation

The dispersion relation is one of the most significant relations in ocean waves. It provides a link between the wavelength or wavenumber of a wave to its frequency. Gravity waves are waves that propagate on the ocean surface through the effects of gravity. In water waves, if the assumption is taken that gravity is the only restoring force for the propagation of the waves, as the value of the wavelength increases, so does the phase speed of that wave. That is, for gravity waves the phase speed of each wave can be expressed as a function of only its wavelength or period. In deep water, the phase speed is found to be [94]:

$$c_p = \sqrt{\frac{g\lambda}{2\pi}} = \frac{gT}{2\pi} \tag{2.13}$$

where λ is the wavelength and T is the period of the wave. With the same assumption and linear theory, it can be found that the dispersion relation of water waves expressed in terms of angular frequency ω and wavenumber λ is equal to [110]:

$$\omega^2 = gk \tanh(kh) \tag{2.14}$$

$$\lambda = \frac{gT^2}{2\pi} \tanh\left(\frac{2\pi h}{\lambda}\right) \tag{2.15}$$

where k is the wavenumber and h is the water depth. As the phase velocity in deep water increases with the increase of wavelength or period and $c_p = \lambda/T$, the wavelength and period of a wave are related. For gravity waves on deep water with $kh \gg 1$ and $\tanh(kh) \to 1$, the dispersion relation is expressed as:

$$\omega = \sqrt{gk} \tag{2.16}$$

$$\lambda = \frac{gT^2}{2\pi} \tag{2.17}$$

2.5 Significant wave height

The significant wave height is one of the most important parameters describing the state of a measured sea. This metric is defined as the average wave height of the top one third of the highest waves in the ocean. If the assumption is taken that the shape of the ocean energy spectrum is narrow, the significant wave height can be found to be equal to four times the square root of the area under the spectral density function or the zeroth moment [150]. That is:

$$\int_0^\infty S(\omega)d\omega = \left(\frac{h_s}{4}\right)^2 \tag{2.18}$$

Equivalently, the significant wave height is equal to four times the standard deviation of the water elevation.

It is useful to have an equation that connects the significant wave height with the wind speed. This can be achieved with the introduction of theory of an empirical energy spectrum. From equation (2.2) of the Pierson-Moskowitz spectrum and equation (2.18), the significant wave height can be expressed as a function of wind speed as:

$$h_s = 2\sqrt{\frac{\alpha}{\beta}}\left(\frac{U^2}{g}\right) \tag{2.19}$$

2.6 Measurement devices

The true ocean sea state and shape of the waves is more reliably measured with *in situ* devices, such as wave buoys and tidal gauges [158][123]. These devices are positioned on the measured ocean and have embedded sensors for measuring the wave elevation. In this work measurements from buoys are considered as the 'true' ocean

state. This section provides a small introduction to the measurement devices that are commonly used for approximating the true state of the ocean. This work utilises wave buoys (Sect. 2.6.1) for testing the presented video techniques. The other devices are provided here as additional informative material, they are not utilised in this work.

2.6.1 Wave buoys

Wave buoys are moored or drifting devices that are positioned floating on the ocean surface with the purpose of measuring sea conditions [158][100]. These devices measure surface weather observations, such as temperature, dew point, wind speed and sea level pressure. Accelerometers are embedded on these devices, enabling the measurement of the wave elevation, which is used for acquiring sea state parameters, such as the significant wave height and the dominant period. Buoys are connected to satellites for transmittance of the measurement data.

2.6.2 Tidal gauges

These devices are positioned on the water surface and are used for measuring the sea level [158][129]. They work on the principle of measuring a current flowing between two parallel stainless steel rods. The signal is given to a conditioning box, which converts it into varying voltage relative to wave height. These devices are also installed in wave tanks for the sea level measurement. They need to be calibrated before use, and their calibration sensitivity can last several days. In comparison to wave buoys, these devices have the advantage of minimal distortion of the wave shape.

2.6.3 Optical wave gauges

These devices, like tidal gauges, are used for the measurement of the sea level. In this case, lasers are used for the non-intrusive measurement [78]. One application is the light detection and ranging LIDAR system, which consists of a laser source and a photo sensor. The laser source is located above the water surface beaming downwards. Some of the light is scattered back to the device and measured by the photo sensor.

The distance between the LIDAR system and the water surface is derived from the time the laser beam takes to travel from the source to water surface and back, and is used for the measurement of the sea level. These devices are under developmental stage.

2.6.4 Video cameras

A typical way of obtaining wave elevation in wave tanks is with the use of video cameras positioned at the mean water elevation [164][168]. These cameras are used for tracking floating objects in the wave tank, providing information about the water elevation. Calibration with the floating marker object is required before the use of the system.

2.6.5 Acoustic devices

Devices such as Doppler profilers are used for acquiring water velocity information at a point [48]. They transmit short sound pulses of known frequency and measure the Doppler shift of the scattered sound signal. They are mounted on the bottom in an upward facing position. Ocean theory is used for translating the velocity information to water elevation.

Chapter 3

Kalman filter

This work introduces techniques that utilise machine learning concepts in order to provide useful information from video of the ocean. This chapter provides an introduction to the Kalman filter theory that is utilised in some of the techniques of this work. Specifically, the methodology in chapter 6 does not utilise the Kalman filter. The methodology in chapter 7 utilises the linear Kalman filter and the methodologies in chapters 8 and 9 utilise the extended Kalman filter. Any additional machine learning theory that is unique to each technique is presented in each associated chapter.

The Kalman filter is performed in the underlying environment of linear dynamical systems with discretised time [134]. In the following text, an introduction is given to hidden Markov models (Sect. 3.1), linear dynamical systems (Sect. 3.2) and the Kalman filter algorithm (Sect. 3.3). After that, an introduction is given to the Kalman filter algorithms for nonlinear systems. The algorithms presented are the linearised Kalman filter (Sect. 3.4), the extended Kalman filter (Sect. 3.5) and the unscented Kalman filter (Sect. 3.6).

Apart from the theory of linear dynamical systems (Sect. 3.2), the linear Kalman filter (Sect. 3.3) and the extended Kalman filter (Sect. 3.5), all other material in this chapter is provided as additional informative material and is not utilised in this work. All information here is provided as introductory material, the specific utilisation and mathematics behind each algorithm is given in detail in the chapters that introduce the video methodologies.

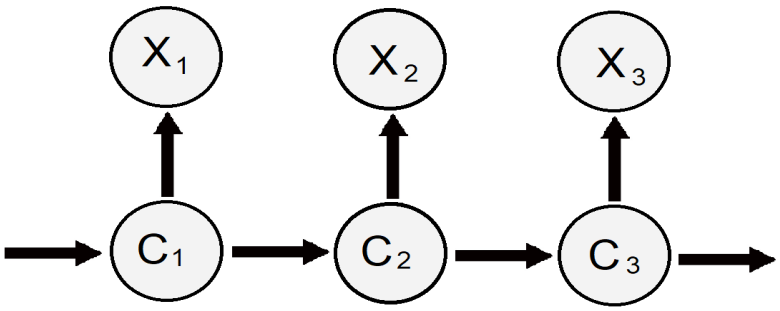


Figure 3-1 Illustration of the structure of the hidden Markov model, where the distribution of X_t in the observed process depends only on the current state of the hidden process C_t and the hidden process satisfies the Markov property.

3.1 Hidden Markov model

A hidden Markov model is a model where the distribution that generates an observation depends on the state of an underlying and unobserved Markov process [130][162]. A simple form of this model is defined by:

$$\Pr\left(C_t|C^{(t-1)}\right) = \Pr(C_t|C_{t-1}) \qquad t = 2, 3, ..$$

$$\Pr\left(X_t|X^{(t-1)}, C^{(t)}\right) = \Pr(X_t|C_t) \qquad t \in \mathbb{N}$$

where $C^{(t-1)} = C_1, C_2, .., C_{t-1}$, $C^{(t)} = C_1, C_2, .., C_t$ and $X^{(t-1)} = X_1, X_2, .., X_{t-1}$. The model consists of two processes, the unobserved or hidden process $\{C_t : t = 1, 2, ..\}$ which satisfies the Markov property and the observed or state-dependent process $\{X_t : t = 1, 2, ..\}$. The Markov property is satisfied when the probability of each event in the process depends only on the state of the previous event. In the observed process, the distribution of X_t depends only on the current state C_t.

The graph in figure illustrates the two processes of the hidden Markov model and their relationship. One main aim of algorithms executed in this environment is the estimation of C_t from the observations in X_t. Linear dynamical systems, introduced in the following section, are a continuous version of the hidden Markov model representation.

3.2 Linear dynamical systems

A dynamical system is a system that describes the transition between different states of a system over time [86]. The state of the system is assumed to expressed by a set of variables. At a given time, the values of the system's variables represent the current state of the system. The system can be expressed in the state-space representation, where the state variables are usually related by first-order differential equations. The transition from one state to the next is done for a short time step ahead. If the time is discrete, the system is usually transitioned from one state to the next in time through a function, which takes as input the state at a time step and gives as output the state of the system at the next time step. If the state of the system is known at a particular time, the state for a future time can be found by integrating the transition function until the desired time step is reached.

In a continuous first order linear dynamical system [76], the variations over time of the N-dimensional state vector $\mathbf{x}_t$ are represented in flow terms as :

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{F}\mathbf{x_t} \tag{3.1}$$

where $\mathbf{F}$ is the transition matrix. Linear dynamical systems are a continuous version of the hidden Markov model representation. In the case of linear dynamical systems, the variables of the hidden states are continuous. In this representation, the state of the system changes over time and what is observed is not the actual state, but a representation of the state. In many cases the observed state is the actual state with the addition of noise. The aim is to track the underlying state of the system at any given time step.

3.3 Kalman filter algorithm

In 1960, Kalman published his work introducing the Kalman filter [84], a recursive algorithm for the discrete linear filtering problem. The algorithm takes data with the addition of noise in the input and aims to provide a better representation of the state

in the output, where the difference between the actual state and the observed state is smaller.

The Kalman filter takes as input a series of measurements, which are representations of the true states of the system, but may differ from the true states. The algorithm receives the observations as a representation of the true states with the addition of Gaussian noise. In order to provide estimations of the true states, the algorithm utilises the state transition model. This model is used for the transition from one state of the system to the next. If input control is present, the control input model is additionally defined.

The transition from one state to the next is expressed as:

$$x_t = \mathbf{F}x_{t-1} + \mathbf{B}u_t + \epsilon_{t-1} \tag{3.2}$$

where $\mathbf{F}$ is the state transition model, x_t is the true state at time t, $\mathbf{B}$ is the control input model, u_t is the input variable and ϵ is the process noise, which is zero mean Gaussian. The relationship between true state and observed state is:

$$z_t = \mathbf{H}x_t + \zeta_t \tag{3.3}$$

where z is the observed state, $\mathbf{H}$ is the observation model, which provides a relationship between the true and the observed state and ζ is the observation noise, which is also zero mean Gaussian. The process and measurement noise are statistically independent.

The estimate of the true state at a given time step is found with the use of the estimate of the state in the previous time step, the transition model and the control model, if input control is present. Then the observation of the state is received, which the algorithm utilises in order to provide an updated version of the state estimate. The difference between estimated state and observed state is used with the aim of providing more accurate estimations in future time steps, by adjusting the algorithm's internal variables.

3.4 Linearised Kalman filter

The discussion in the previous section involves the Kalman filter for linear systems. However, many systems can be nonlinear. Although the linear Kalman filter is useful for systems that are close to linear, versions of the Kalman filter exist for nonlinear systems. In the linearised version of the Kalman filter, a nonlinear system is linearised and then the linear estimation techniques are applied [88][98]. This nonlinear version of the Kalman filter is based on linearising the nonlinear system around a nominal state trajectory. After the system is linearised, the theory of the Kalman filter is used in order to estimate the deviations of the state from a nominal state value. This provides the estimation of the state of the nonlinear system.

In a nonlinear system where the system equation and the measurement equation are nonlinear functions, Taylor series is used for expanding these equations around a nominal state, control state, output and noise. The nominal values are based on *a priori* knowledge of the system. This *a priori* knowledge is used for forming the equations of the nominal trajectory. The partial derivative matrices are then evaluated at the nominal trajectory values. Then, the linear Kalman filter is applied for solving the system.

3.5 Extended Kalman filter

In the linearised Kalman filter, the states of the nonlinear system are linearised around a nominal state trajectory, which is formed with *a priori* knowledge of the system. It is not always straightforward to form the nominal state trajectory. This is where the extended Kalman filter can be used. In this version of the Kalman filter, the Kalman filter estimates are used for the nominal state trajectory [132][157]. The nonlinear system is linearised around the Kalman filter estimate, and the estimate is based on the linearised system. The partial derivative matrices are evaluated at the current state estimate and the Kalman filter equations are executed based on these formations.

If the dynamics of the system and the measurements are discrete, the discrete extended Kalman filter is performed. In this case, the predicted state is found from the estimate of the previous step and the predicted measurement from the predicted state with the use of partial derivative matrices.

3.6 Unscented Kalman filter

The unscented Kalman filter [82] is an extension to the extended Kalman filter that aims to reduce the errors from the linearisation of the extended Kalman filter. The algorithm uses sigma points, which are deterministic vectors whose ensemble mean and covariance are equal to the original mean and covariance. The nonlinear function is then applied to each deterministic vector in order to provide the transformed vectors. The ensemble mean and covariance of the transformed vectors then are used for the approximation of the true mean and variance of the function.

Chapter 4

Literature review

The approximation and prediction of the parameters of ocean waves, such as the significant wave height, wind speed, dominant wavelength and period has a wide range of practical applications, such as to provide prior knowledge for the performance of critical maritime operations [177] and building of harbours and other ocean related structures [67]. It can also enhance the performance of ocean wave energy converters [4][173]. These parameters can be measured with a wave buoy, a usually anchored device floating on the ocean surface with an embedded accelerometre. Known ocean theory is applied on the acceleration data in order to retrieve the sea state parameters.

Remote sensing of the ocean surface has gained popularity for the estimation of the characteristics of the ocean surface, as it provides useful information about the ocean in a non-intrusive way and has relatively low operational costs. Two main advantages of remote sensing methods against *in situ* measurements are lower expenses (owning, maintaining and installing the *in situ* equipment) and the avoidance of exposing personnel and equipment to extreme weather conditions. Popular categories for remote sensing of the ocean surface are the use of stereo photography and radar data obtained from satellites or ships.

Remote sensing of the ocean can also be done with photographs and videos. A method of this category was first introduced by [112], where a small number of photographs were manually processed for associating the percentage of whitecap coverage in the ocean with the wind speed. With the progress of technology, a variety of meth-

ods have been proposed for obtaining information about the ocean.

Computer vision is a field of science concerned with algorithms that enable computers to extract, analyse and understand useful information from photographs or video footage. In this review, computer vision methods for obtaining parameters of the ocean surface are described. These methods are based on concepts of stereo photography, the brightness intensity of images and videos, the properties of reflection, refraction and polarisation of light, the nature of radar measurements and optical flow.

Some of the ocean surface parameters that are approximated with the following techniques are: the two dimensional slope field, the two-dimensional wave elevation maps, the significant wave height, the energy spectrum and the dominant frequency, period and wavelength. Camera calibration, assumptions of idealised environments, the complexity of experimental setup, operational costs and complexity of camera equipment are some of the drawbacks of the methods described. The literature of computer vision techniques for measuring the ocean of the following text comprises of the following categories: stereo photography (Sect. 4.1), spatial information from images (Sect. 4.2), temporal variation from video (Sect. 4.3), shape from refraction (Sect. 4.4), polarisation imaging (Sect. 4.5), radar (Sect. 4.6), shape from shading and optical flow (Sect. 4.7).

4.1 Stereo photography

Stereo photography has been widely studied for the measurement of the ocean's surface parameters. The techniques proposed in this remote sensing of the ocean category involve the use of two or more calibrated cameras and the correlation of concurrent images representing the scene from different views. This configuration enables the perception of the three-dimensional depth from the two-dimensional images. For the correlation procedure, matching features are identified in concurrent images from different cameras in a pixel level. The pixel pairs are then triangulated to correspond to the three-dimensional real world, thus solving the depth problem of images.

The main task of the image correlation task is to convert the two-dimensional information from images to the information of the three-dimensional world. The spatial resolution of these methods is constrained, making the measurement of short waves a challenging task. The cross-correlation of images is a computationally intensive process, which translates as many techniques not running in real-time.

If objects are not present in the scene, measuring long gravity waves has the difficulty of identifying markers for matching the pairs of images from different cameras. This is in most cases solved by using short waves as markers for longer waves. As the scale of waves to be measured with stereo photography require waves of smaller scale as markers, many techniques are not applicable for the study of shorter waves. The sun's reflection may decrease the accuracy of the estimations. One disadvantage of techniques that use stereo photography for measuring ocean surface is that the rays of the light source are reflected and refracted, and the pattern of the received light changes with a change of the location, making it difficult to perform image correlation [160].

In [17] the authors propose a three-dimensional stereo system, the wave acquisition stereo system (WASS) to obtain topographic information of the surface of ocean waves from video. The structure information is then used in order to reconstruct the three-dimensional shape of the waves. After that, spectral analysis is performed on the ocean waves' shape information to acquire the direction of the waves. In this system, two calibrated cameras are used and two experiments are performed for evaluating their performance. The results are compared with *in situ* sensor data obtained during the experiments and are found to be consistent.

Similarly to other techniques that use stereo photography, the work requires a prior calibration of cameras and cross-correlation of images at a pixel level. The calibration is a challenging task. The cross-correlation problem, usually referred to as the correspondence problem, is computationally intensive. Specular reflection may additionally be a source of errors in the cross-correlation task.

The automated trinocular stereo imaging system (ATSIS) is introduced in [167] for targeting correspondence problems between images and thus estimating the ocean

surface with more accuracy. The system comprises three cameras, leading to a more complex calibration process. This process involves the sub-processes of interior and exterior calibration, which are performed in the laboratory. The measurement setup is thus more complicated than the setup that uses two cameras. The field of view of the measurement setup is relatively small, restricting the method's applicability. For image matching, the authors propose a two-step process, in which the input data from left and right cameras are correlated and then a search window is used in order to correlate data from the three cameras with the use of a normalised cross-correlation method on the intensity values. The experimental results were found promising.

Short ocean waves are waves with periods less than 20 s that are generated by local or distant wind fields [90]. Gravity waves are waves generated by distant wind fields. In [141] and [12], the authors investigate the application of stereo photography for obtaining the directional wave height spectra of short ocean gravity waves. During the tower and ocean water wave and radar dependence experiment - TOWARD, two cameras were placed on a tower and an oil platform. The authors noticed that the estimations were more accurate when the mean surface of long waves was subtracted from the measurements. Although the results in both methods are promising, drawbacks of many stereo photography methods, such as camera calibration and computationally intensive correlation of pairs of images limit the method's applicability. Concerning the correlation of pairs of images, illumination conditions are required, which further limit the method's applicability. Only a limited number of images were assessed, as the image processing task was done manually.

The surf zone is the ocean offshore area of approximately 5 m-10 m depth, where wave breaking usually occurs. In [40] the ocean wave height was estimated using stereo imaging in the surf zone. The authors conducted experiments in a wave gauge and in an outdoor environment using two overlapping cameras. The significant contribution of the study is the large area of investigation, which covers 1800 m^2 in the outdoor experiment. Image matching is initially done with large window sizes. Then, the results from the initial matching are used for constraining the search region for higher resolution passes. The height estimation results are compared with the pres-

sure sensor data for the wave gauge experiment and indirectly through wave station, tide gauge and GPS data for the outdoor experiment. Although results were found relatively consistent, the authors do not validate their technique adequately for the outdoor experiment. Environmental assumptions have to be taken, which restrict the method's applicability, such as ovecast sky and adequate lighting. The complexity of the experimental setup is another drawback, as well as the computational complexity of the method due to the correlation task for paired images covering large areas of ocean surface.

A two-step methodology for estimating the spectrum of short ocean waves is proposed in [93]. The methodology uses information from both image brightness and stereo photography. As a first step, fast Fourier transform is applied to ocean surface photographs. The intermediate results lack in providing real world scale measurements, as only relative measurements are retrieved. Then as a second step, stereo photography is applied for the determination of the wave spectrum, based on the previous results. Combining the two methods comes with advantages, such as the measurement of both small and long ocean waves. Disadvantages of both methods are however present. These include the requirement of an overcast sky and uniform illumination, camera calibration and image correlation. These restrict the method's applicability.

In [160] the authors make efforts in optimising the image correlation task by staining the water in wave tanks with white dye, making the diffused light more visible. The correlation of pairs of images is done by dividing the area of the first image into segments, and a line passing through the centre of each segment is used for acquiring candidate matching points for the second image through cross-correlation with parabolic sub-pixel fitting. With the methodology, the authors show that it is possible to capture wave breaking. The complexity of the experimental setup, which includes an LCD lamp placed above the water surface, limits the method's applicability. The method can be executed only experimentally in a wave tank, and not in real environments.

4.2 Spatial information from images

The brightness intensity from single images is also used in research for measuring ocean parameters. In this case, the property of reflection of the sun's rays or the envrionment's illumination on the ocean surface according to the experimental geometry is utilised for estimating the slope field. Specifically, techniques are proposed for obtaining information about the ocean's surface from single photographs, modelling the relationship between the radiance of the sea surface received by the camera lens and the grayscale intensity of images. The methods of this category are based on the nature of the ocean water being a highly reflective, specular surface.

Compared to the techniques of stereo photography (see Sect. 4.1), one image in this case is sufficient, instead of at least two concurrent in the case of stereo photographic techniques, and thus no image cross-correlation is required. The techniques of this category model the way the wave slopes on the ocean surface are mapped into a radiance distribution received by the camera. The reflection of the sun's rays on the ocean surface on each measured by the camera point changes as the slope of the surface changes. This modulation of reflection is recorded by the camera and into the images. If the viewing angle of the camera to the surface is adjusted in a particular manner in accordance to the angle of the sun to the surface, an image of the ocean surface is assumed to provide an almost linear relationship between the surface slope and amplitude transmittance of the film.

The work by [13] is one of the first works that uses the intensity values from single photographs in order to estimate directional information. In [34] the authors use the sun's reflection on the water surface for getting slope information. Specifically, the authors use the direct reflection of the sun's rays, or sun's glitter, to get information about the slope. A two-step methodology is proposed, which first consists of identifying the points in the photographs from the geometry of the scene that directly reflect the sun's light on the camera lens. Then, the methodology uses the brightness of this point in order to get the frequency of the slope. Two relationships are formed, one that correlates the slope probability to the surface radiance, and one that correlates

the surface radiance to the film density. Additional sources of surface radiance that are studied are the reflection caused by the skylight and the scattered sunlight from beneath the water surface.

The restriction of taking aerial photographs with direct sunlight is required, in addition to the requirement of avoiding multiple reflections and hidden steep wave valleys. Thus, the position of the camera, sea surface and sun must be configured before the experiments. The presence of whitecaps, slicks and cloud shadows introduce errors in the estimations. The study has been a benchmark for future studies of the ocean surface using the sun's glitter or the environment brightness. The method does not estimate the ocean wave steepness or energy distribution with respect to scale and thus is not capable of measuring wind wave directional spectrum information or the profile of wind waves. Furthermore, significant illumination of the environment is required.

The authors in [151] propose a methodology that utilises the relationship between the glitter pattern and the slope field of the ocean described in [34]. In this work, a linear relationship is formed between the radiance of the water surface and the slope of the surface, where the unified luminance of the environment is studied in the place of the glitter. From the slope information, the author proposes a method for obtaining the two-dimensional Fourier spectrum of the surface. The environment illumination is required to be of an overcast or clear sky.

According to the authors, phenomena such as light emanating from the scattering under the water surface, polarisation of the reflection coefficient and white capping are not modelled and violate the assumptions taken by the model. Oil slicks and other localised reflective material on the surface introduce spectral components not existing in the wave structure. As the camera approaches the grazing incidence, the scene gets highly distorted and the highest amplitude waves shadow some of the ocean surface, an effect not modelled by the proposed work. Only relative measurements are provided, and additional measurements are required to be concurrently taken for normalising the resulting spectrum. The validation of the methodology is not straightforward.

In further work [152], the authors introduce a methodology for estimating the directional spectrum of the ocean from single photographs. The variation of surface slopes and the diffraction pattern are utilised as the directional spectrum of the ocean, since the transfer function from the front to the back frontal plane of the lens was considered identical to the Fourier transform at that time. Seas with whitecapping are not applicable. A complex mathematical model is proposed that depends on the specific geometry of the scene and a relationship is introduced between the photographic intensity at a point and its sources of illumination, specifically reflected, refracted and scattered light.

The system requires a polariser to be included in the camera system for excluding vertically polarised light. Several environmental assumptions are taken for more accurate results, which could limit its applicability. These include a diffuse sky illumination, small wave slopes, negligible presence of white capping in the photographs and insubstantial presence of particles in the atmosphere. The validation of the method is not straightforward and the work did not receive any scientific following. Furthermore, direct reflection of the sun may provide challenges in the estimations. This work is based on the dependence of the reflection coefficient on the incidence angle of light. The reflectance of the ocean surface is mostly non-linear. Specifically, for angles of 0 to 40 degrees it does not change in a high degree, but for angles above 53 degrees (Brewster's angle), the reflectance increases rapidly.

The holographic method is introduced in [154] in order to get information about the slope spectrum from the negative of photographs. The authors investigate the slope from the negative of the photographs of the ocean and compare the spectrum against the energy spectrum from images with the stereo photographic method. The spatial variation of the ocean from a single photograph is used. The technique makes the assumption that the camera is facing the ocean vertically and as the photographs are taken from an angle a process is proposed for fixing the errors from the computation. The authors show that the slope spectrum of the ocean is proportional to k^{-2}, where k is the wave number. As the technique utilises physics of the reflection of the sun's rays on the camera lens, factors such as mean angle of incidence, small-scale

wave slope, colour of water and brightness intensity have to be taken into consideration, as well as the gradient of the colour and intensity in zenith and azimuth angle.

The authors in [113] extend the holographic method by introducing more variables in the model, such as the upwelling radiance as an additional source of radiance received by the camera. The authors suggest that the geometry plays a significant role in minimising the nonlinearities in the transfer function between radiance and slope. The work proposes a second order model and a new form of the transfer function from the surface wave slope to the incident radiance on the camera sensor. The authors suggest that if the relationship between the wave slope and surface radiance received by the camera is not linear, Fourier transform may not be meaningful. The requirement of clear sky is to be taken for the method to be valid, in which case a polariser is required to be placed in front of the camera sensor in order to avoid the non-linear effects of the vertical polarisation and thus take only into account the horizontally polarised radiance.

The work by [34] is extended by [53], which uses hyperspectral images from the airborne visible/infrared imaging spectrometer (AVIRIS) system. The transfer function between the wave slope field and the radiance received by the system takes into account additionally to the model introduced in [34] the slope information of small scale waves and the atmospheric transmittance. For the transfer function, the information of atmosperic transmittance, the AVIRIS and solar geometries and Fresnel reflectivity have to be known. This, in addition to the requirement of the atmosperic conditions being homogeneous restrict the method's applicability. According to the authors, phenomena not modelled such as aerosol phase function and multiple scattering may affect the estimations. Buoy measurements are used for validating the methods and the measured results are found consistent in a satisfactory level. The method is limited by its environmental requirements, which include homogeneous illumination, significant sun glitter and clear sky. The plane in this setup is required to follow a linear path.

One significant challenge with techniques of this category is the presence of white-

caps in high wind speeds, which are not taken into consideration for the modelling of the environment, due to the complexity of such phenomena. Thus, these techniques are applicable for only moderate or low wind speeds. Additionally, the requirement of uniformly bright sky limit their application. The polarisation of the sky is an important factor that affects the radiance of the surface as it is received by the camera. This factor is dependent on the experimental geometry and is difficult to model. Other effects not modelled are foam and the shadowing effect of longer waves on smaller scale waves.

Additional main drawback of techniques of this category is the variation of the relationship between wave slope and radiance across the image. Even if this relationship is completely linear, it is different for different parts of the image [113]. The adjustment of the viewing camera angle to the ocean surface according to the sun's angle to the surface in a way that the relationship between wave slopes and radiance of the surface is almost linear limits the applicability of the methods. Such configurations that solely use the theory of the reflection of the sun's rays on the surface and the Fourier transform are limited by the adjustment and are not applicable when the angle of the camera is positioned in such a way that the relationship between radiance and slope field are not linear. In addition to the nonlinearity of the wave slope and the surface radiance, the nonlinearity of the transfer function between the surface radiance and the sensor output is an additional source of errors.

4.3 Temporal variation from video

In the case of video, the change of brightness intensity in time is utilised for acquiring useful information from the temporal variation of the wave field. Some of the proposed techniques have less dependence on the experimental geometry.

The extension of the research of [13][152][154] to work on spatial and temporal variations of the wave field is done by [66]. The same environment illumination restrictions apply as in previous work that uses only the spatial information from images. The slope field and the phase speed for different wavelengths are studied.

A polarising filter is used for minimizing the upwelling radiance. The assumption is taken that the measured radiance is the surface slope component. A two-dimensional Fourier transform is performed on sequences of images of the ocean after correcting for perspective. The mean surface slope component spectrum is calculated by means of the fast Fourier transform with Hanning window of the image radiance in a direction parallel to the dominant wave. The authors observe that linear theory for the phase speed is valid for distant wind sources and high phase speed values are observed for high frequencies when local wind speed is present, which makes the relationship non-linear.

The study of the temporal variation of the wave field has been done in research with the use of time series of grayscale intensity values at a pixel level. An example of such work is the work of [31], where the image sequences are acquired from a wave basin. In this case, non-uniform illumination of the measured area has an effect on the results, something the authors try to overcome by forming a relationship for the distance of the light source. A charged couple device (CCD) camera is used, and a transfer function is formed with the use of *in situ* wave gauges. A polarising filter is required for minimising the effects of scattered light from the surface.

The work uses points in the image that correspond to locations of wave gauges previously positioned inside the wave basin. Time series of pixel intensity values are then formed from the video at these points. Efforts are made to minimise the effects of noise with spatial and temporal smoothing techniques. The authors observe an association between periods from grayscale time series and periods from the wave gauges. Drawbacks of the method are the complexity of the experimental setup and the requirement of additional *in situ* measurements for the transfer function.

The work by [178] utilises videos of the ocean in real environments. The authors propose a methodology that uses the extended maximum likelihood method (EMLM) [80] and the Bayesian directional method (BDM) [68] in order to estimate the directional energy specturm of the ocean in shallow water from a video camera. With the EMLM method, the directional spectrum is obtained as a linear summation of the respective cross-power spectra obtained from the measured waves. With the BDM

method, the estimation of the directional wave spectrum is done with a regression analysis to find the most suitable model from the data. Configurations of arrays of pixel intensity values are used. When the authors performed Fast Fourier transform with Hanning window on pixel intensity time series, only the most appropriate images were visually inspected and selected, avoiding effects such as foam. Sea states of significant wave height above 1 m were not studied because of the non-linearity of the water surface and presence of foam. The authors utilise a windowing process, which can rearrange the peaks of the energy spectra. The authors do not ellaborate on the selection of the windowing function parameters.

Compared with wave gauge data, the authors observe a correspondence between the peaks of the intensity values of images and the *in situ* wave gauge data, but not in all cases. The validation for the results of the directional energy spectrum is limited, as only a numerical model, and specifically the SWAN model, is used. Additionally, the manual inspection process gives more accurate results, but a more effective way would be the design of automated methodolologies without inspection of the input video data and without the loss in accuracy. The work is also limited to sea states with significant wave height lower than 1 m.

In another published work the same authors [179] use time series of pixel intensities and the BDM method to get the directional wave spectrum from shallow water video. Bathymetry information is estimated with the utilisation of the dispersion relation and the non-linear inversion method Levenberg-Marquardt (LM). As with [178], wave gauges are used for matching the peaks between video energy spectrum and *in situ* energy spectrum and the validation of the directional energy spectrum results are limited with only the customised ocean model SWAN.

Time series of pixel intensity values from coastal surveillance systems in the nearshore are utilised by [121] for approximating the dominant period with the use of the fast Fourier transform algorithm. The technique first through a thresholding value identifies the areas of non-breaking waves, which are then passed through a filter. The challenge is the identification of the low cut-off frequencies of the filter. The authors observe that the time series contains information from both the surface

variation and the environmental brightness flunctuation. Isolating these is a big challenge of the work. Problems are present in the spectrum from the Fourier transform, such as multiple peaks in nearby low frequencies and high energy values in very low frequencies, making the identification of the peak frequency a challenging task.

The authors tackle the problem of identifying the peak frequency with an arbitrary process of selecting the cut-off frequency as 30% increment of the peak frequency, without providing the intuition of such heuristic. The cut-off frequency also differs for different time series of pixel intensities. After that, a windowing function is performed on the spectrum, a process that can modify the position of the peak and thus the estimation. When compared to the available data, the peak period estimations from video were not in all cases favourable. Additionally, the authors observe that the applicability of the technique is not valid for low and high sea states, but rather only medium states.

In [153] the authors utilise a cross-shore timestack, which is a set of time series at an array of pixel locations, in order to get bathymetry information from near-shore locations. In this case, the foam in nearshore areas is used for getting the propagation of the waves. The complex empirical orthogonal function (CEOF) is performed in order to get the phase speed, which is then used with the dispersion relation to get the bathymetry of the scene. When this method is tested in [121], the relative error of the estimations is found to be 56%, which questions the effectiveness of the method.

The information of wave crests is used in [175] in order to estimate the dominant period from video in the surf zone. First, a low-pass filter is performed and backward frame differencing to remove noise caused by foam, and then a thresholding process with arbitrary thresholding values is performed for isolating the wave crests from the rest of the video. The methodology is called linear feature extraction. The estimation of wave properties is done with the use of temporal and space information of wave crest locations instead of pixel time series. The dominant period is found from the time between successive crests. The problem faced was that, unlike in the case of a single sinusoid, in the ocean multiple crests appear, not always representative of the dominant wave. An arbitrary solution is formed, in which minimum and maximum

values for duration between two successive crests is set, and the dominant period is found as the time between crests that is most frequent from the observations. The phase speed is calculated from the linear feature images as the velocity of the crests. Then the dominant wavelength is found from the dominant period and the phase speed. The thresholding process in this work is arbitrary and may introduce errors in the estimations in cases the wave crests have lower intensity values than the thresholds. Although the estimation of various sea state parameters is more straightforward from the wave crest information instead of pixel time series, the authors do not elaborate on the selection of the thresholds used for distingishing the wave crests. Additionally, the authors do not demonstrate that their methodology successfully distinguishes foam from wave crest information.

The method of applying discrete Fourier transform on photographs is utilised in [147] in order to get the pixel to metre scale and sea state from video of the ocean surface in real environments. The spatial energy spectrum of single images is represented as the summation of energy for each radius, which represents a spatial frequency. For the temporal energy spectrum, each frame of the video is represented as a set of coefficients of each principal component of the principal components analysis algorithm (PCA), and Fourier transform is performed on each one. Then, the summation of the energy values gives the energy value for the specific temporal frequency. The pixel to metre scale is acquired by matching the peaks of the temporal and spatial spectrum. However, due to the Nyquist criterion theory, it is challenging to make this correlation since the spatial spectrum does not capture the dominant wave in all sea states. Due to this, incorrect estimations are possible, as the peak matching task can correlate frequencies in space and time domains that do not both represent the dominant wave. The sea state estimations are not specific enough, providing with estimation on where the state is related to the Beaufort scale, and not actual values of wind speed or significant wave height.

The authors in a future work [146] present a different technique for obtaining the pixel to metre scale and sea state from real video. One slice in the dominant direction is selected from each frame, and the phase modulation of each spatial frequency is

used with the Fourier transform to get a correspondence between temporal and spatial frequencies. All waves present in the frequency domain are used for the estimation of the pixel to metre scale, by fitting a parabola that provides an association betweeen spatial and temporal frequencies in the frequency dispersion.

To get the wind speed, the temporal variation of the intensity at a pixel level is used with the Fourier transform in order to get the frequency of the dominant wave. As the peak can easily be distorted, the method fits each energy value from the frequency domain to the Phillips energy spectrum shape of various wind speeds in order to get the best match through a least squares error metric critirion. However, this requires that the shape of the specific ocean measured is very similar to the Phillips spectrum shape. As two parameters of the customised spectrum are to be estimated, amplitude and wind speed, two or more sets of values of the variables may provide fitting with very near error metrics to the measured shape of the spectrum. The results of sea state are not validated sufficiently, as the synthetic scene in the experimental results is rendered with the same Phillips spectrum that is used for the fitting, and no *in situ* devices are utilised for testing the wind speed estimation.

The work of [133] extends the work of [146] in order to be applicable to large scale airborne images of the ocean. In the case of airborne video, the authors notice that effects such as sub-surface, above-surface radiance and sun's glitter are not taken into account and may affect the estimations. The authors identify sequencially the frequencies with the highest energy values instead of using all the frequencies. A threshold is set, after which the iteration stops, and the sea state is identified from the dominant frequency which is found as the centroid of the selected frequency bands that have the highest energy.

4.4 Shape from refraction

This image processing category of measuring the ocean surface is based on the refractive property of light. In an inverse way of the function of laser slope gauges, where a laser beam is passed through the surface from above the surface and a camera is

receiving the input underwater, the techniques for estimating the two-dimensional shape from refraction have an underwater illumination source and a camera placed above the surface receiving the input. The presence of wave breaking, foam and bubbles produce problems in the reconstruction of the slope field.

Although utilising the refractive property of light has a more complicated experimental setup, there are some advantages over techniques based on reflection, such as [152][154]. Due to the fact of only 2% of the light gets reflected from the water surface, high luminance in the environment is required for techniques based on reflection. A larger light source is additionally required, as the angles are doubled at the point of reflection. Additionally, comparing the reflective property of light in terms of slope with the refractive property, the refractive property is less non-linear.

The refractive property of the water surface is studied in [81], who utilise an underwater illumination system in an indoor wave facility and a CCD camera in order to estimate the two-dimensional wave-number spectrum of short ocean waves. The wave slope estimation is based on light refraction at the water surface. A relation is formed between the intensity modulation and wave slope by ray tracing with the assumption that the light source radiates light isotropically in all directions. Theory of the refraction law is utilised for the formation of this relation. Calibration is achieved with an *in situ* device. In this experimental setup, the illumination source is underwater and the intensity change is considered to be related to the wave slope. The nonlinearities related to the refraction law are not taken into consideration and may have an effect to the spectral densities. The geometry of the scene is required for the calculations.

In [137][138] the authors use data from two CCD cameras in a wave tank and calibrate the transfer function with wave gauge measurements in order to get the energy spectrum of the ocean surface, which is found from the three-dimensional energy spectrum of the videos with a process that uses spectral decomposition and local wavenumber regression. Disadvantages of the work are the complicated experimental setup, which includes the use of near infrared filters on the cameras, projection foils for reflection which are illuminated by spotlights, radiometers and the use of ground

control points and tachymeters for coordinate correlation.

A system that estimates the two-dimensional slope field of the ocean surface based on refraction is introduced in [87]. This system is capable of measuring small scale waves. The setup of downward looking camera with telecentric lens and underwater illumination is used. The light source is complex and consists of rows and columns of LED lights that can have two brightness settings. This lighting setup provides four gradient images, which are utilised in order to provide more detailed estimations of the slope field. The conversion from slope to height is done with an algorithm described in [50], which reconstructs the surface shape and the slopes from the shading information. The image intensity in terms of the surface orientation is used for estimating the slopes. Specifically, the surface slopes are estimated by means of a finite set of basis functions and integrability is enforced by calculating the orthogonal projection onto a vector subspace spanning the set of integrable slopes.

4.5 Polarisation imaging

The sun is an electric charge that vibrates in many directions. A light wave that originates from the sun thus vibrates in more than one planes and is called an unpolarised light wave. An unpolarised light wave can be transformed into a wave that vibrates in only one plane through various conditions, one of which being the reflection from a non-metallic surface. This transformation is called polarisation [139]. Thus, light that reflects from the ocean surface is in some degree polarised light. In the case of unpolarised light in the illumination of the environment, the orientation of the surface determines the polarisation parameters, making it possible to estimate the two-dimensional slope field of the ocean surface, provided that a polarisation camera is measuring at least three different polarisation states. The polarisation state of the light from images contains information that cannot be acquired from the image brightness alone [136]. Challenges in methods that use images or video of the ocean in real environments and utilise properties of the reflection and polarisation of light such as [152] and [176] are phenomena not considered into the modelling of the world,

such as sky radiance polarisation, upwelling radiance and the direct rays of the sun's light.

In [176] the polarimetric slope sensing (PSS) method is introduced for the generation of the two-dimensional slope field with the aim of studying the short scale waves of the ocean. A polarisation camera is utilised in overcast sky conditions and a methodology is presented that uses the relationship between surface orientation and the change in polarisation of reflected light in order to get the slope field, taking into consideration only the reflected and upwelling skylight. Specifically, the polarisation camera is configured to simultaneously measure 0, 45 and 90 polarisation orientations to determine a complete state of partial linear polarisation. A relationship is then formed between incident light with a known polarisation and the response of the polarisation intensities in order to get the wave slope information. In cases of non overcast sky conditions, such as sunny or partly cloudy days, where partially polarised light is present, sky and upwelling radiance must be taken into consideration, and this method is not applicable. In this case, the authors claim that the use of two cameras is required for the estimations. The environmental condition restrictions is one of the main drawbacks of the methodology.

The scene geometry is required and the setup is restricted in specific incidence angles not exceeding the Brewster angle. Compared to stereophotographic methods, the method has the advantage of a more simple measurement setup, as one polarisation measuring camera is required compared to at least two calibrated cameras in the case of stereophotographic methods. The requirement of unpolarised light in the environment in this case however restricts the proposed method's applicability. Additionally, the method does not provide real world scale of the ocean surface measurements.

An alteration of the PSS method is used in [97], using a shipborne camera for the measurements. The authors measure capillary and gravity-capillary waves and conclude on their contribution to the sea surface roughness and their sensitivity to wind speed increase. They further validate the use of the PSS method for measuring fine-scale wind waves. Similarly to [176], the assumptions of unpolarised light in the environment and no upwelling radiance restrict the method's applicability. The

results are only tested in the controlled environment of a wave tank and are not validated with *in situ* measurements, but with a numerical empirical model.

An airborne twelve camera system with three polarisation measurements for each of the system's colour, red, green, blue and near-infrared, called airborne remote optical spotlight system with a multispectral polarimeter is introduced in [74] and further studied in [75]. The information from these twelve measurements is combined with two image registration algorithms, a feature-based correlation algorithm and a mutual information-based algorithm. The system requires individual and inter-camera calibrations and is capable of capturing surface orientation, surface roughness and index of refraction in an area of several square kilometres. The authors do not measure wave slope and wave height. Additionally, inter-camera alignment and the feature-based correlation techniques make the system complex and computationally restrictive.

The same system as in [74] is investigated in [16] in order to get the degree and angle of linear polarisation, and the measurements are compared with an imaging performance model, where they are found to agree for look angles close to 90 degrees from the sun, whereas a large deviation is found for angles close to 180 degrees. However, the authors are not able to retrieve the two-dimensional slope field or the significant wave height, due to the incapability of the method to isolate the reflected signature, which can be used for the estimations, to the full signal measured by the system.

Similarly with [176] and [97], the work in [92] uses a single polarisation camera to get the two dimensional slope field of the ocean in a similar methodology to the other works. Knowledge of the geometry of the scene is required and unpolarised light in the environment, which restrict the method's applicability. In this work, three images with different polarisaitons from the CCD camera are aligned with the use of floats as control points. The resultant images have polarisation components substituting the RGB components. From the images, all pixels are used in order to get the parameters such as viewing angle, reflecting angle and slope projections. These parameters are then used for the construction of the slope field of the measured scene.

The wave slopes are found through measurements as slopes of triangles formed by the float center. Then, the shape of the sea surface is obtained by integration of the slope projections from the Stokes parameters. As the scale is unknown, the authors are unable to derive the wave heights in real units from the slope field. The slope field is tested in specific points with wave elevation from nearby wave buoys and the estimations are found to be consistent.

4.6 Radar

The measurement of ocean waves with marine radars is based on the temporal and spatial information of the radar's backscatter [105]. Specifically, the modulation of the backscatter of the radar's electromagnetic waves on the sea surface ripples or roughness is used for making the ocean measurements. Such type of data require that the sea surface has a particular roughness, which is the result of local wind. Thus, if only distant wind is present, such as in the case of swell gravity waves, the surface might be smooth and the backscatter might not yield sufficient wave field information for accurate estimations.

The WaMoS X-band radar-based ocean surface monitoring system is first introduced in [131]. The radar systems in [22][21][38] use radars in the X-band that are equipped with a horizontal polariser. They emit microwave pulses at grazing incidence and receive backscattered microwave energy from the ocean surface in the form of intensity images. These intensity images do not depend directly to the wave heights of the ocean surface, but to other factors of the digitalisation of the radar system [117].

In [22] the WaMoS II radar system is used with the aim of getting information about the two-dimensional energy spectrum and the significant wave height from radar image sequences. The information about the current is estimated, which is used for filtering in order to get the energy spectrum. The significant wave height is found as a function of the signal to noise ratio, which is the ratio of the true wave signal to the background noise. Two calibration coefficients are calculated with the use of nearby

buoy measurements. Although plots are used for presenting the comparison of peak period and significant wave height between radar measurements and nearby buoys, the relative errors in the estimations are not presented. Furthermore, the validation of the significant wave height is done with the same buoy data that are used for the calibration.

In a future work [21], the authors use empirical spectra to simulate radar images, taking into account the effects of linear dispersion relation and current speed. Specifically, the authors present the inversion method in order to obtain the spectral representation of the wave field. The spectrum is obtained with a three-dimensional discrete Fourier transform of the radar images. After a low-pass filter is applied, the current is estimated with the use of the dispersion relation. In order to isolate the spectral energy points associated with the ocean's movement from all other effects due to the radar imaging, the authors use simulated ocean images and a 20 % threshold from the spectral peak for filtering. These simulated radar images are used as filters in order to identify the energy spectrum of the wave field from the image energy spectrum. The complexity of the design of the simulated model, which consists of a number of user defined parameters, restrict the method's automaticity and the reliability of the estimations, as different parameter configurations of the simulated images can result in different estimations of the method. Additionally, effects in a given sea that are not modelled by the empirical spectrum, such as wave breaking and foam, might introduce errors in the estimations.

In another future work, the authors [117] use a method to get information about the significant wave height based on the assumption that this metric is proportional to the square root of the signal to noise ratio. In this case, the signal is taken to be equal to the image spectrum after a filtering process similar to [22], and the noise is derived from the speckle background noise components. The significant wave height in real units is then given with two calibration constants, which are obtained empirically.

In [38], the authors use a different empirical inversion method for estimating a time series of ocean surface elevation maps. This method is based on the ocean surface tilt angle in a pixel level in antenna look direction and does not require the presence of *in*

situ sensors, such as buoys. The parameterisation of the mean radar cross section is done by fitting a third-order polynomial curve for each look direction. The tilt angle is found from the difference of the depression angles and the ocean surface elevation is determined from the tilt angle information by integration. The 3D wave spectrum of the ocean surface elevation field is estimated with the transformation of the tilt angle sequence to the wave number frequency domain with a 3D FFT. Then, this spectrum is integrated by multiplying with a transfer function.

In a future work [37], the authors use a neural networks algorithm in order to get a relationship between the radar backscatter to the wave field characteristics, which is used for getting the wind speed. The wind streaks are identified by integrating radar image sequences over time, and their orientation is determined to be the normal to the local gradient. Specifically, the wind directions are estimated form the local gradients of the mean radar cross section images. The wind speeds are estimated by training a neural network with the parameters of mean radar cross sections, the radar retrieved wind direction and the antenna look direction. The authors observe an improvement in the wind speed retrieval by adding the air-sea temperature difference in the neural networks parameters. The difficulty in verifying the detailed wind fields provides uncertainty on the accuracy of the method.

4.7 Shape from shading and optical flow

The shape from shading technique aims to estimate the three-dimensional shape of an object from a two-dimensional monochrome image of the object [15]. The estimation uses the brightness variability of the image. Optical flow in image processing refers to the change in the brightness patterns of a scene after the motion between camera and scene. The task of motion estimation based on optical flow is a correspondence problem, in which the motion of the pixels in a sequence of images is tracked so that there is a dense correspondence of pixels [171]. In the context of ocean waves, the intensity values of grayscale images of the ocean from a video can be given as input to a method based on optical flow for the extraction of useful information.

The particle image velocimetry (PIV) algorithm is used in [51] in order to get the phase speed information from videos of the nearshore ocean. The foam noise is removed with the same method as in [175], the backward frame differencing method (described in section 4.3). Then the PIV algorithm is used for tracking the increment of high intensity values, which are considered to be temporal increments of wave crests. The surface velocity vectors are used in order to get the phase speed estimations, which are found to be consistent with the estimations of the method by [175]. An adaptive multi pass algorithm is introduced that initially performs the cross-correlation interrogation in a relatively large window. Then the calculated vector field is used as a reference for higher resolution levels and the interrogation window size is refined after each iteration. This iteration is repeated until the final size window is reached. However, the authors do not validate the results against real *in situ* information for phase speed or with bathymetry information. The authors do not provide evidence that the frame differencing technique efficiently isolates wave crests from wave breaking. If this process is not done in an efficient manner, tracking the wave crests could not yield estimations corresponding to the true ocean conditions in all cases.

In the field of computer graphics, the authors in [99] introduce a method that takes as input a video file of the water's surface and renders the surface in real time. The method estimates the water's surface shape based on the shape from shading algorithm described in [128], which is given as a prior to an optical flow algorithm and provides constraints to the horizontal velocity tracking. The authors investigate whether the relative height of the water surface can be estimated through the horizontal flow of the surface. Thus, the use of an optical flow algorithm that estimates the horizontal velocities is enough for estimating the relative height motion of the water surface. The most realistic rendering is produced when the camera is positioned vertically to the water surface.

The method works in water surfaces where water is a suspension, and not completely opaque. Shadows and reflections from sunlight compromise the results. The authors do not provide real world scale estimates of the water surface and do not

validate the estimated shape and motion of the water surface, as they are interested in rendering realistic water scenes. However, this work provides a premise for possible applicability of methods such as shape from shading and optical flow for getting information about the sea state from video of the ocean.

4.8 Conclusion

Computer vision is a field of science with a vast number of applications, including the measurement of the ocean surface parameters. Methodologies of this field have the advantage of measuring in a non-intrusive manner. Stereo photography provides the depth information by using two-dimensional images as input and involves the use of two or more cameras and a correlation process of pairs of images. In the context of ocean waves, it is used for providing approximations of ocean wave parameters such as spectral information, wave direction, shape and height. Stereo photography methods require camera calibration and cross-correlation of images. The cross-correlation task is a challenge, in addition to the environmental illumination restrictions for providing more accurate results.

The property of the polarisation of light is also used in the literature in order to provide slope information of the ocean's surface. Unpolarised light in the experimental setup is necessary and the real world measurements of parameters of ocean waves cannot be approximated. Methods based on refraction are often only suitable for indoor environments, such as laboratories, as they require controlled conditions and in many cases the incorporation of an underwater illumination system.

The reflection of the light is studied in the relevant research as a means for providing slope information from photographs of the ocean. The glitter of the sunlight is also studied for approximating this information. Only relative measurements are estimated and environmental assumptions have to be considered. In a large number of the presented work, sufficient validation for the methods is not provided with nearby *in situ* measurements. In the case of the work that estimates the two-dimensional slope field or wave heights, it is a challenge to know that the estimated values at each

point are the true values of the ocean, as such information about the true values is difficult to obtain.

In this work the interest is focused on computer vision and image processing techniques that utilise video of the ocean in real environments for estimating ocean wave parameters and other information about the scene. Although some parts of this work utilise the spatial information from frames, this work is more closely related to the category of techniques that use temporal variation from video for the estimations.

Chapter 5

Experimental data

This chapter introduces the experimental data utilised for testing the methodologies of chapters 7, 8 and 9. The methodologies introduced in chapter 6 are tested with radar data, due to the requirement of the input video data covering larger area of ocean surface.

The experimental data used for testing the methodologies in chapters 7, 8 and 9 comprise of two video sets, captured in different dates, locations and settings. The first set is recorded video of the ocean from a ship (Sect. 5.1) and the second set from a tower (Sect. 5.2). Both sets of video have corresponding buoy measurements indicating the true ocean sea state. The buoy data are used solely for testing.

In this introductory chapter a discussion is then given about limitations of the data sets and how an ideal data set would look like (Sect. 5.3) and possible sources of error from video of the ocean in real environments (Sect. 5.4). The accuracy metrics utilised for the evaluation of the methodologies of this work are introduced in Section 5.5.

5.1 Shipborne data

Experiments were conducted in 2014 in the North Atlantic sea. A single camera was placed onboard a ship, and during the times the ship was not moving towards a destination, that is the movement of the ship is the rotational (pitch, yaw, roll),

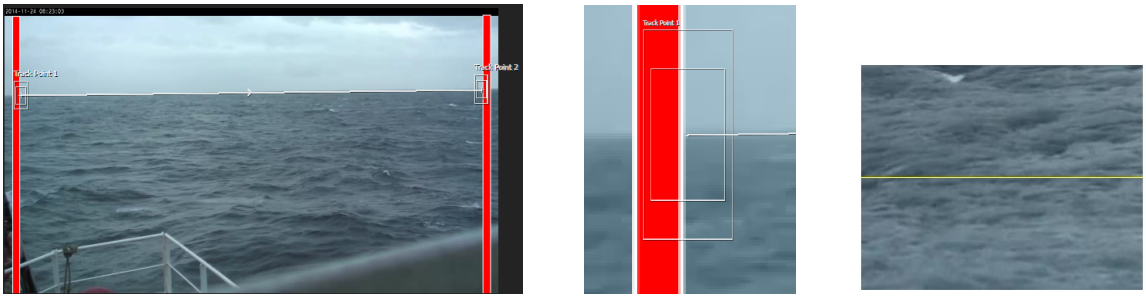
the video is isolated and split into one minute videos. During the experiments, two buoys recorded concurrently the wave elevation at a nearby location. The videos were captured on the 24th of November 2014. The buoys were on the sea surface after 9:15 a.m., and thus the sea state is not validated for videos of previous times. The sea state in this set of videos is approximately the same, as the state is not expected to change in a large degree in the time span of a few hours. The ship video are captured in a sea state of approximately 3.1 m-3.4 m.

A preprocessing step is performed for the ship data in order to overcome the ship's rotational movement. To stabilise the ship video, the software Adobe After Effects [5] is used. The rotational movement (pitch, roll, yaw) of the shipborne videos is overcome by tracking and stabilising the horizon. The rotational tracker is utilised with two tracking points, which form a line that in ideal conditions is exactly the line of the horizon. To achieve this, two rectangles are drawn as in figure 5-1a in order to stop the tracking points from moving horizontally while tracking the horizon. Figure 5-1a shows a typical frame from ship video and figure 5-1b presents how each tracking point is selected.

From a video, image sequences in the form of image 5-1c are given as output from the preprocessing step. A single row of pixel locations is used in the experimental results for computational efficiency. The videos have a duration of approximately one minute. The shipborne video frame rate is 15 frames per second.

5.2 Tower data

The second set of videos is captured from the Frying Pan Ocean tower, a former lighthouse located 85' above the Atlantic ocean. This tower is located on the Frying Pan Shoals approximately 39 miles southeast of Southport, North Carolina. A 24 hour live video footage is available online from that tower [47] that captures the ocean surface. A buoy device 'Station 41013' owned and maintained by the National Data Buoy Center [116] provides hourly updated significant wave height and other sea state information. This buoy station is located in a near location to the Frying



(a) (b) (c)

Figure 5-1 Shipborne data. (a) video with horizon stabilisation (b) left tracking point (c) video after preprocessing. Line depicts the pixel locations utilised in the experimental results.

(a) (b) (c)

Figure 5-2 Tower video data. (a) sample frame (b) buoy station 41013 (c) tower video after preprocessing. Line depicts the pixel locations used in the experimental results.

Pan Ocean tower.

Although the camera is moving showing a panoramic view, in some time instances the camera is stable at fixed positions, and during these short videos were captured for the experiments. In cases of high local wind being present, the shaking of the camera is fixed with the stabilisation features of Adobe After Effects. A single row of pixel locations is used in the experimental results for computational efficiency as in figure 5-2c.

As the tower videos are captured in different dates, the ocean is captured in a variety of sea states. This enables the testing of the techniques of this work with videos of lower and higher sea states. Figure 5-2a shows a typical frame from tower video and 5-2c the result after preprocessing. The tower video has a frame rate of 30 frames per second.

5.3 Limitations of the data sets

The three data sets used in this work are the radar images (presented in chapter 6) and two sets of video data in real environments (described in Sect. 5.1 and 5.2). The radar data are used for testing the methodology in chapter 6. The video data in real environments are used for testing the methodologies in chapters 7, 8 and 9.

In the case of the methodology in chapter 6, radar data are used in the place high-scale videos of the ocean. High-scale videos are videos that have a large area of coverage. There are however some limitations in this data set. The radar images have a lower frame rate than real video, as the time between successive frames is the time that the radar instrument takes in order to cover a complete cycle of measurement, which is equal to approximately 2 s. Another limitation for that particular data set is the mechanism that is followed for the formation of the intensity values. Specifically, in order to acquire the intensity values for the radar images, the backscatter of the radar's electromagnetic waves on the sea surface ripples or roughness is used [22][38]. In real video, the intensity is considered proportional to the lights reflected from the water surface and depends on wave slope [31].

There are however some advantages in the use of the radar images. They enable the validation of the pixel to metre technique because the value of this metric is known *a priori*. They also have a large area of coverage, enabling the estimation of the sea state with the use of the spatial variation from video.

Because the methodology in chapter 6 utilises the spatial variation from video, it would be required that the ocean area coverage includes at least two times the dominant wavelength, if the estimation is based on the dominant frequency. For example, a sea state of significant wave height 2.5 m, which is a moderate to high sea state according to the Beaufort scale [1], corresponds to a dominant wavelength of approximately 102 m. According to the Nyquist theorem [118], if we were to capture the dominant wave, we would need to capture at least two of its cycles. Therefore, if we were to use the spatial variation from video, we would need an area coverage of at least 204 m. In particular, it would be required that the line in the dominant direction

(chapter 6, Sect. 6.1.1 explains how the dominant direction is approximated) has a length of at least 204 m. For higher sea states, the dominant wavelength increases. Thus, the area coverage requirement also increases. This is the reason an ideal data set to be used as input for validating the methodology in chapter 6 would need a large area of coverage. For example, the data set could be acquired with video from an aeroplane.

In an ideal scenario, we would use a video with a high coverage that has the same frame rate as a regular video (e.g. 25 frames per second). However, realistically this is challenging to obtain because the aeroplane is in motion. The same area would need to be captured for the duration of the video. In [133] successive images with time between them equal to 1 s are acquired from an aeroplane while it is following a continuous straight line flight path across the water surface. Each individual measurement is considered independent. In an ideal scenario, the aeroplane would follow a path where it would be possible to capture the same area. That is, it would capture the area and then return back to the same area to capture it again. In this context, the aeroplane would have a small size, so that its movements are more flexible. For example, in [133] the sea state is estimated with images of the ocean from a small fixed wing airframe with wingspan 2.2 m. A drone could also be used for the acquisition of the video. This would make things easier, as drones will potentially be able to capture the same area over a time period of approximately one minute.

An ideal video data set would also have corresponding measurements of the pixel to metre scale and sea state from reliable sources, to be used as the 'true' values in order to calculate the error metrics. In the case of the pixel to metre scale, some camera equipment is able to calculate this metric based on the distance of the camera and the ocean surface. As a note on why we estimate this metric in chapter 6, the methodology estimates this metric for videos captured in the past that do not have this information and for utilising this metric in order to estimate the sea state. In the case of the sea state estimation, reliable measurements would be in the form of *in situ* devices located in a nearby location to the captured ocean surface. *In situ* devices such as wave buoys and tidal gauges are considered reliable instruments for

measuring the ocean surface, although they are still providing an approximation.

The video data in real environments are used for validating the methodologies of chapter 7, 8 and 9. Because these methodologies utilise the temporal variation from video, due to the Nyquist theorem [118], we would need to capture at least two cycles of the dominant wave in time, if we were to measure the dominant frequency. For example, a sea state of significant wave height equal to 2.5 m corresponds to a dominant period of approximately 8 s. Thus, the video duration would need to be at least 16 s. The videos in this work have a duration of approximately one minute, which is sufficient for measuring even higher sea states.

Limitations of these data sets include the video quality, consistency of measurements, movement of the ship (for the shipborne video), movement of the camera in cases of high local wind and brightness fluctuations. In terms of video quality, the cameras of both data sets are not recording in high definition. An ideal data set would be acquired with a high definition camera. This would enable the video techniques to work on more high quality video data, which could improve the estimation accuracy. In terms of consistency of measurements, both data sets lack recording the exact same ocean surface for prolonged periods of time (e.g. several weeks or months). This would enable a better validation of the video techniques. In the case of the shipborne video, this is due to the short duration of the experiments (less than a week). In the case of the tower video, the camera is panning in some time instances (i.e. it is not measuring the exact same area for prolonged time periods).

An ideal data set would comprise of video from a camera positioned on a stable location (e.g. a pier) capturing the ocean surface for prolonged periods of time (ideally 24/7). This would allow a better verification of the video techniques. It could also help provide theoretical error bounds of confidence in the estimations. This is because we would have multiple videos of the ocean surface with the same corresponding values of significant wave height. Similarly to both shipborne and tower data sets, an ideal data set would also have corresponding reliable sea state measurements from *in situ* devices located in a nearby location, so that the validation process is more accurate.

In terms of movement of the camera, from the shipborne data the camera move-

ment is mainly due to the rotational movement of the ship and the heave. This movement is stabilised in order to capture the same area of ocean for the duration of the video by stabilising the horizon (process described in more detail in Sect. 5.1). This however requires some additional pre-processing, which limits the execution of the technique in real time. In the case of the tower video, the position of the camera is stable. There is however some 'shaking' or vibration of the camera in cases of high local wind. The stabilisation effects of a computer software (See Sect. 5.2) are used for stabilising. This could also limit the execution of the video techniques in real time.

There is however a reason for using these specific video data sets in these specific circumstances. Mariners are interested in obtaining estimations of the sea state from ships (see introduction chapter 1 for more details). In terms of the vibration of the camera in cases of high local wind, an ideal video data set would have protection mechanisms in place so that the camera is not affected by the weather conditions. For example, the camera could be positioned indoors and in areas where the local wind has less influence on the camera.

In terms of brightness fluctuations, both video data sets capture the ocean in a variety of weather conditions. These include weather of overcast (mostly cloudy) sky and clear sky with high reflections of the sun's rays. Some pixel locations include the sun's glitter (direct reflection of the sun's rays), while others do not. This variation of brightness that depends on the weather conditions could affect negatively the video estimations. For example, there is a difference in pixel intensities between videos capturing the sun's glitter and videos captured during an overcast sky. Additionally, for the same video data, some pixel locations are affected by the sun's glitter, while others are not. This difference in brightness is a possible source of errors.

An ideal data set would measure the ocean surface in a mostly overcast sky, in order to avoid the presence of the sun's glitter on the ocean surface. However, if a video technique was limited to only overcast skies, its applicability to the real world would be limited. A more desirable scenario would be the design of techniques that provide sea state estimations even during the presence of environmental brightness

fluctuations. This is the case with the techniques of this work.

5.4 Sources of error

Modelling the ocean surface in real environments from video is not a straightforward task. The nature of ocean water includes phenomena which in some cases may not be included in the model. These phenomena may not be included in the modelling of ocean water due to their complex nature, which is unpredictable in nature. However, their presence in the measured scene may decrease the accuracy of estimations of some techniques. Examples of such phenomena are ripples, rain, fog and nonlinear wave interactions, such as wave breaking and foam. Foam in particular can be easily interpreted as wave crests, as they both have high intensity pixel values.

Environmental brightness fluctuations is an additional source of possible loss in accuracy. For example, the presence of the sun's direct reflection or sun's glitter on some areas of the measured ocean may provide different estimation values for these areas compared to areas that do not directly reflect the sun's rays on the camera. If the geometry of the scene and the environmental brightness are not considered in the model, the presence of high pixel intensity values in some areas of ocean due to the sun's glitter may introduce inaccurate estimations.

Errors may also be introduced due to the nature of the camera equipment capturing the ocean surface. For example, electronic noise, low quality video, out-of-focus and distortions are some possible sources of error. If the camera is placed on a non-stable location, such as a ship, the movement due to the ship's rotational motion (pitch, roll, yaw) is an additional challenge to overcome. This is solved in this work by tracking and stabilising the horizon. If this preprocessing step is not done in an efficient manner (for example if in some time instances the horizon tracking is not correct), errors in the estimations are likely to be introduced.

As mentioned, ripples on the surface of the ocean may additionally be a source of noise in the video. In [81] the authors become more specific about this type of noise, and how it affects the quality of the estimations. When the ocean surface is

flat, small particles become apparent, which are not visible when the ocean surface is rough, introducing a new source of noise that is highly unpredictable and depends on the nature of water. As mentioned in [81], this source of noise becomes especially challenging in low sea states. The authors describe an additional challenging source of noise in video data of the ocean, specifically fixed-pattern noise, which usually occurs in long captures and is characterised by brighter or darker pixel intensities than what is expected. Although in some cases it may not be noticable, these effects can have a significant impact on the sea state estimation, as they may appear as spikes in the energy spectrum.

5.5 Estimation accuracy measures

These measures are used in the experimental results for the evaluation of the accuracy of the proposed methodologies against the sea state measurements of buoy devices. The buoy data are considered to indicate the true sea state. The accuracy measures used are the mean absolute error MAE (Sect. 5.5.1), root mean square error RMSE (Sect. 5.5.2), mean absolute percentage error MAPE (Sect. 5.5.3) and mean percentage error MPE (Sect. 5.5.4). A comparison of these metrics in the context of this work is given in Section 5.5.5. In the following text, predicted or estimated value y_p indicates the estimation from video and true value y_t indicates the measurement from buoys or other *apriori* knowledge about the scene.

5.5.1 Mean absolute error (MAE)

This metric is the average of the absolute difference of pairs of estimated and true values. It is calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_t - y_p| \tag{5.1}$$

where n is the number of data points, y_t is the true value and y_p is the predicted or estimated value. This metric has the advantage of having the output in the same scale as the input values.

5.5.2 Root mean square error (RMSE)

This metric indicates the square root of the average of the squared differences of pairs of estimated and true values. Compared to the MAE, this metric has a larger increase in the presence of outliers due to the squared terms. Additionally, presence of larger differences causes more increase to this metric compared to the MAE. Compared to the mean square error (MSE), which is the average of the squared differences of pairs of estimated and true values, this metric has the advantage of having the output in the same scale as the input values. It is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - y_p)^2} \tag{5.2}$$

5.5.3 Mean absolute percentage error (MAPE)

This metric provides as output a percentage of accuracy between estimated and true values based on the formula:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_t - y_p}{y_t}\right| \tag{5.3}$$

The fact that the output is given as a percentage makes the metric more easily interpretable in some cases. However, the metric has some disadvantages. The output is higher for pairs of true and estimated values with lower values. Additionally, it favours estimation values that are lower than the true values. That is, for the same difference between estimated and true values, the output is smaller if there is an underestimation instead of an overestimation.

5.5.4 Mean percentage error (MPE)

The equation of the MPE is similar to the equation of the MAPE. Here, the absolute value operation is not used [23]. This is useful for showing whether the approximations are overestimating (negative error) or underestimating (positive error). The equation

is:

$$MPE = \frac{100}{n} \sum_{i=1}^{n} \frac{y_t - y_p}{y_t} \tag{5.4}$$

Since the absolute value operation is not used, positive and negative values of MPE can offset each other. Thus, it is useful for showing overestimation or underestimation rather than the absolute percentage error value.

5.5.5 Comparison

Both MAE and RMSE metrics provide values in the same scale as the input values. In the context of this work, this is advantageous. For example, it provides the error in metres when the input are values of significant wave height. Comparing MAE and RMSE, RMSE has a larger increase in the presence of outliers. This is useful in cases where the estimations are required to be more accurate across all input data.

The MAPE metric provides the output as a percentage. This is more easily interpretable in some instances. For example, the output from this metric can be more easily understood by someone not familiar with ocean theory when evaluating significant wave height estimations compared to MAE and RMSE. Compared to the MAPE, the MPE metric gives information of overestimation or underestimation of the model, whereas the MAPE provides an absolute percentage value of the error.

Chapter 6

Pixel scale and sea state from ocean large scale remote observations

This chapter is based on the previously published research paper:
A. Loizou and J. Christmas, 'Estimating Pixel to Metre Scale and Sea State from Remote Observations of the Ocean Surface,' IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 3513-3516.

The problem that is discussed in this chapter is the estimation of the pixel to metre scale or ground range resolution and the sea state from high-scale videos of the ocean surface, such as from an aeroplane. This work uses radar images for testing the proposed methodology. Although radar is not the same as real video of the ocean, this chapter will provide an investigation of the nature of the data, how they can be used for acquiring useful information and will propose methodologies for estimating the aforementioned metrics. The radar image is used in the place of a video image because the pixel to metre scale of the radar image is known. This enables testing the proposed methodology with the true pixel to metre scale. An example of a radar image used as input for the proposed methodologies is given in figure 6-1a.

After the related research paper of this chapter was published [102] (the first author in [102] is the author of this work), researchers were able to use heuristics

similar to the one discussed here (see Sect. 6.1.4, iterative thresholding process for the selection of frequencies) for estimating the sea state from airborne real video [133]. This suggests that the ideas and methodologies discussed in this chapter are not limited to radar images, but can be modified and used for real video. This is the reason that in the related research paper [102], it was expressed as 'radar as a surrogate to real video'.

The pixel to metre scale or pixel resolution metric gives a correspondence to metres of the pixel size of an image. For example, if the pixel scale of an image is equal to 6 m per pixel, this means that the length of one pixel is equal to 6 m. The usefulness of the pixel to metre scale lies in the calculation of the length of objects in the scene, if this information is not available in the measuring camera equipment. Video captured in the past without such available information can also be used. Additionally, in this work, the pixel to metre scale is used directly in order to get the sea state with a methodology that aims to capture and use the spatial variation of the wave field from ocean images.

The methodology of this chapter utilises the spatial variation from video for the approximation of the pixel to metre scale and the sea state. When we attempt to track the spatial variation of the wave field from video, we are required to use a video of a large area of coverage. In the following text, we refer to this kind of video as *high-scale video* of the ocean. This is why radar images are used for validating the methodology. Apart from the fact that we know the true pixel scale *a priori* for validation, these images cover large areas of water, sufficient for tracking the dominant wave from the spatial variation.

The Nyquist theorem [118] is the answer to why we need high-scale video when tracking the dominant wave from the spatial variation of the video. For example, a sea state of significant wave height equal to 2.5 m (moderate to high sea state according to the Beaufort scale [1]) corresponds to approximately a dominant wavelength of 102 m. In order to track the dominant wave from the video's spatial variation, according to the Nyquist theorem, we would need the video to cover at least two times the dominant wave in space. That is, we would be required to have a coverage of at least

204 m. In particular, it would be required that the line in the dominant direction (Sect. 6.1.1 explains how the dominant direction is approximated) has a length of at least 204 m. As the sea state (or significant wave height) increases, so does the dominant wavelength. This means that the coverage area requirement also increases.

As mentioned in the introduction chapter, the sea state is useful for maritime operations and for the design and implementation of sea vessels, platforms and offshore structures. The work in [146] proposes a methodology for estimating the same metrics from real video. However, the validation of the pixel to metre scale is limited. Specifically, the authors validate their results on the approximate length of a surf board, which is not measured before the experiments. Information about true scale is not available. For the sea state, online meteorological data is used for validation in an approximate manner. However, their methodology has been influential for the development of the work of this chapter. The authors in [146] associate temporal and spatial frequencies from the modulation of the phase speed, which is used in this chapter as the phase modulation matrix (see Sect. 6.1.1).

The novelty of this work involves the proposal of a different process for acquiring the pixel to metre scale from the phase modulation matrix. This method is based on the heuristic of selecting the waves from the phase modulation with the highest energy in an iterative thresholding manner. In [146], the authors propose fitting a parabola considering all frequencies in order to get the scale. In the experimental results (see Sect. 6.2.1), it is shown that considering all frequencies might not always be a good idea, and that the process proposed here provides more accurate results. In sections 6.1.3 and 6.1.4 it is discussed why the proposed heuristic of this work is useful.

The novelty of this work also includes the whole methodology for estimating the sea state in section 6.1.6. In [146], the authors fit the shape of the video energy spectrum acquired by applying a Fourier transform to the Phillips spectrum. They utilise the temporal variation of the wave field for the sea state estimation. In this work, the spatial variation of the wave field, the pixel to metre scale and the waves with the highest energy found from the proposed iterative process of the pixel to

metre scale methodology are used.

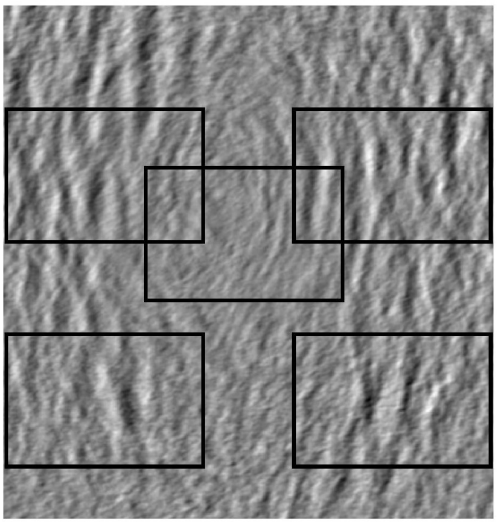
6.1 Methodology

The methodology involves using video of high scale remote observations of the ocean surface for getting the pixel to metre scale. The sea state is then estimated with the pixel to metre scale of the previous step and spatial information, specifically the wavelength of the dominant wave and a few other secondary waves close in terms of energy and frequency to the dominant wave. A Fourier transform is performed [14] in the space and time domain for getting the phase speed matrix, a matrix where columns represent temporal information (periods) and rows spatial information (wavelengths).

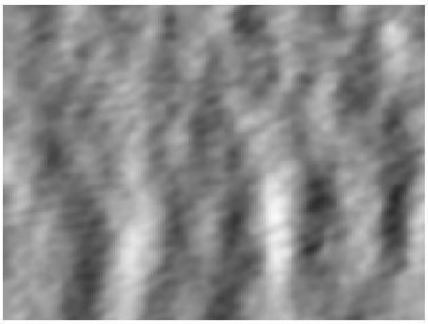
6.1.1 Phase speed matrix

Information about the phase speed is used in order to estimate the pixel to metre scale and the sea state. First, each frame of the video is rotated so that the dominant direction is parallel to the horizontal plane. The dominant direction in this case is found by applying a Fourier transform on the radar images. An example of a two-dimensional Fourier transform of a radar image is given in figure 6-1c. Specifically, from a single frame of the video, a Fourier transform [14] is performed and the dominant direction is determined to be the direction of the line that passes through the center of the image in Fourier space that maximises the summation of the intensity of the points through which it passes. In other words, it is the radius with the highest sum of pixel intensity. Then, each frame is rotated in the dominant direction and a matrix is built from each rotated frame, where each row captures the spatial variation in the dominant direction. In other words, we reposition the frames so that the direction of the rows is the same as the direction of the dominant wave. We do this because we are attempting to track the dominant wave in space. In order to accomplish this, we need to work on the spatial variation in the dominant direction.

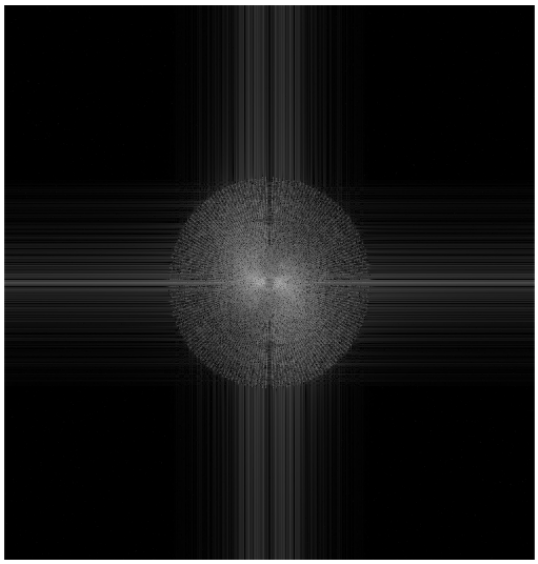
By performing a Fourier transform on each row it is possible to get the magnitude and the phase values for each spatial frequency. On each frame a Fourier transform



(a) Example of a radar image and selection of 5 different areas.

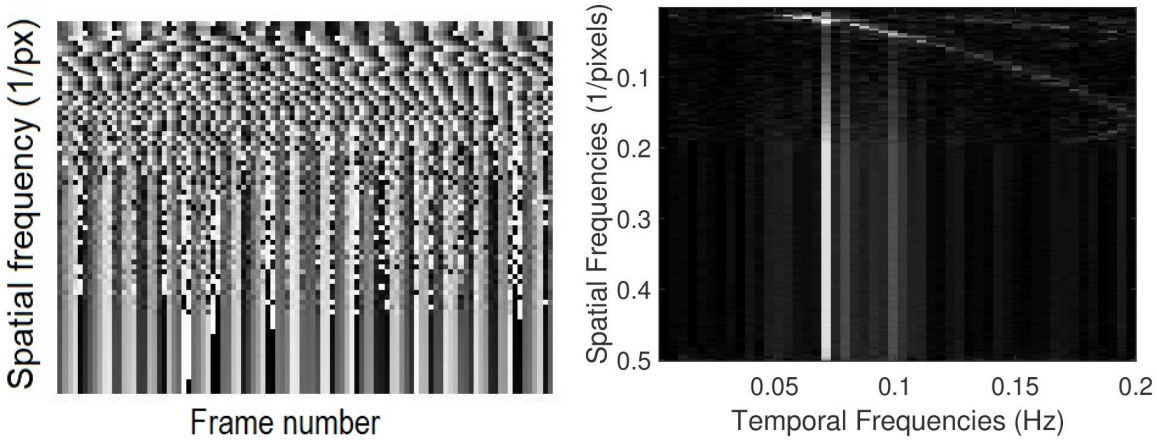


(b) Example of a single area from the radar image, given in the input for running the methodology.



(c) A two-dimensional Fourier transform of the area of the radar image.

Figure 6-1 Example of a radar image and its two-dimensional Fourier transform.



(a) The phase modulation matrix

(b) The phase speed matrix

Figure 6-2 The matrices formed from the spatial variation of the wave field.

is performed on each row. Then, the phase values are selected and then averaged for the same spatial frequencies. The output from each frame thus far is a set of phase values, each one corresponding to a given spatial frequency. By stacking the phase values of each frame, the phase modulation matrix is formed (see fig. 6-2a), where each column contains the phase values that correspond to a particular frame and each row represents a spatial frequency. Specifically, each row presents the phase values of a specific spatial frequency across time. It demonstrates how the phase changes for each spatial frequency.

In the next step, a Fourier transform is again performed on each row, giving the rate at which the phase changes. This rate is thus equal to the phase speed. The phase speed metric, according to ocean theory, is equal to wavelength divided by period of a wave, or equally temporal frequency divided by spatial frequency [2]. An example of the phase speed matrix is given in figure 6-2b.

6.1.2 Pixel to metre scale from phase speed matrix

By using the phase modulation in time, a wave is described in both temporal and spatial terms, and specifically as spatial and temporal frequencies. At each column of the phase modulation matrix, the peak frequency would in ideal circumstances (with

only the effects of the frequency dispersion) give the temporal frequency of that wave, and all other entries would have minimal energy values. As can be seen in figure 6-2b of a typical video, practically this isn't the case. Due to the nature of the Fourier transform on discrete measurements and the presence of additional phenomena and components not taken into account in the model of the world (See Sect. 6.1.3), other frequencies additionally have given energy values.

As a summary to what is discussed in Sect. 6.1.3, possible reasons that we do not have only one peak value for each pair of spatial - temporal frequency are: spectral leakage, presence of capillary or small amplitude waves and the amplitude dispersion property of water waves. In the case of capillary or small amplitude waves, they present an increase of phase speed with the increase of frequency [110], which differs from the effect of the frequency dispersion. This is the possible reason we can observe high intensity vertical bars (e.g. between temporal frequencies 0.05 Hz and 0.1 Hz) in figure 6-2b. In the case of the amplitude dispersion, which is a nonlinear effect, we have an increase of phase speed with an increase of amplitude [172]. Thus, the phase speed is higher for waves with higher amplitude (and energy). In other words, the phase speed with the amplitude dispersion is higher for the dominant wave and waves near the dominant wave. This is a possible reason the intensity values in the area between temporal frequencies 0.05 Hz and 0.1 Hz in figure 6-2b are higher. It is very possible that the dominant frequency lies within that frequency range.

As mentioned, if only the effects of the frequency dispersion for the frequencies in the model are taken into account, the energy value of the peak frequency at each row is associated with the phase speed of that particular wave. In deep water, the phase speed c_p is equal to [144][107]:

$$c_p = \frac{\lambda}{T} = \frac{\omega}{k} \tag{6.1}$$

where λ is the wavelength, T is the period, ω is the angular frequency and k is the wavenumber. For a fixed water depth or deep water, the phase speed increases with the increase of wavelength [144]. From the phase speed matrix the energy values at the points connecting temporal and spatial frequencies would be expected to descrease as

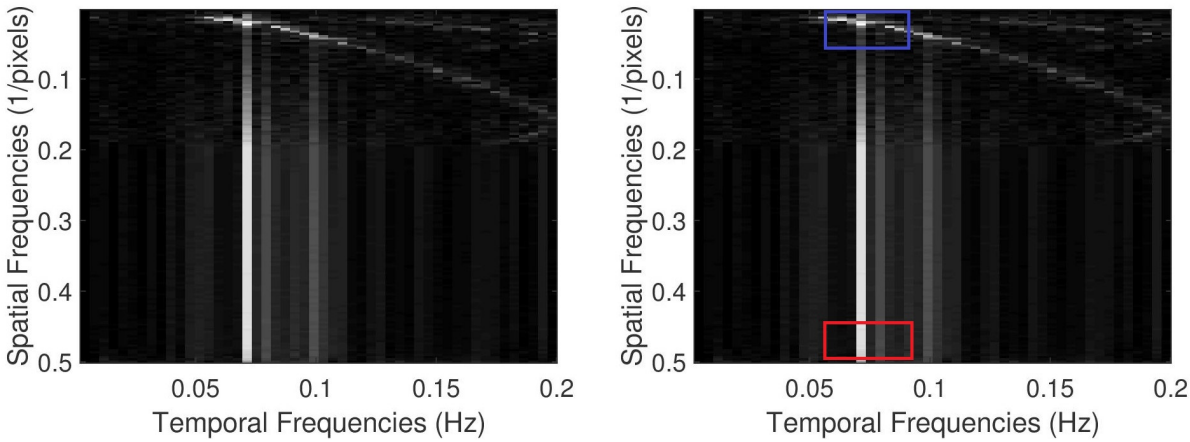
the values of the axis of spatial frequencies increase, something that can be observed to be indeed the case in figure 6-2b. This effect of increasing phase speed with the increase of wavelength is due to the frequency dispersion property of water waves. However, other properties of water waves and noise additionally have an effect on the value of the phase speed and consequently the energy values in the phase speed matrix, effects that will be discussed in the following paragraphs.

6.1.3 Unexpected effects and noise

From the application of the discrete Fourier transform algorithm on the phase speed matrix for each spatial frequency, the phenomenon of spectral leakage [104] introduces high energy values in frequency positions other than those to be identified. This phenomenon is visually observed as if the energy corresponding to phase speed is leaked into other frequencies. This effect is present due to the nature of the discrete algorithm, when the measured signal is not an integer number of periods.

A big source of noise are ripples on the water surface, or capillary waves, which are not taken into consideration in the modelling of the world and exhibit a behaviour in the frequency dispersion exactly opposite of that of gravity waves, that is they propagate with higher phase speeds as the values of the spatial frequencies increase [110], appearing as high energy values in high spatial frequencies.

An additional source of error is the amplitude dispersion of water waves, which is a non-linear effect and is characterised by higher phase speed values for waves with higher amplitude [172]. Specifically, waves of larger amplitude have a different phase speed from small-amplitude waves. This effect is not modelled in the environment definition and could introduce errors in the estimations. This, in addition to the presence of capillary waves and the spectral leakage phenomenon are possible reasons that in the phase speed matrix high energy values are present in places not expected, see figure 6-2b.



(a) Phase speed matrix.

(b) Example of selected entries (blue rectangle) and when the iteration stops (red rectangle).

Figure 6-3 Example for the selection of the entries in the phase speed matrix with iterative thresholding process.

6.1.4 Solution

In all these phenomena, the aim is to distinguish those values on the phase speed matrix that correctly provide a link between temporal and spatial frequencies. Due to the frequency and amplitude dispersion, it is expected that the highest energy values in the phase speed matrix are apparent in lower frequencies (frequency dispersion) with high amplitude (amplitude dispersion). In this work, the heuristic of selecting the first neighbouring entries in terms of frequency with the highest energy values is used. Specifically, the algorithm iteratively selects the coordinates of the entries in the phase speed matrix that have the highest energy values and stops the selection once an entry has been found that is not close in terms of frequency (spatial or temporal) to the other entries. This final entry is interpreted as energy value due to the effect of unexpected effects discussed in the previous section. An example for the selection of the entries in the phase speed matrix with this iterative thresholding process is given in figure 6-3.

In the following text, the intuition behind this heuristic is explained. As can be seen in figure 6-4, which shows a typical energy spectrum of the ocean (the Pierson-Moskowitz spectrum), the waves with the highest energy (or amplitude) have the

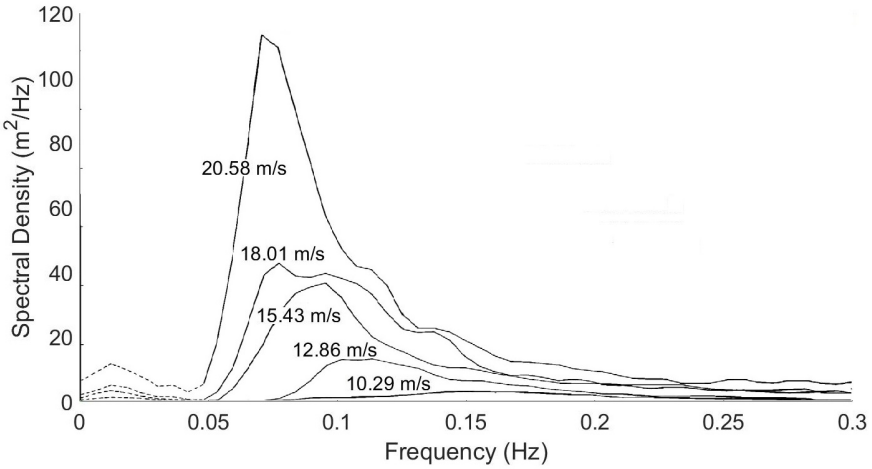


Figure 6-4 The shape of the Pierson-Moskowitz energy spectrum of fully developed seas for varying wind speeds adapted from [114].

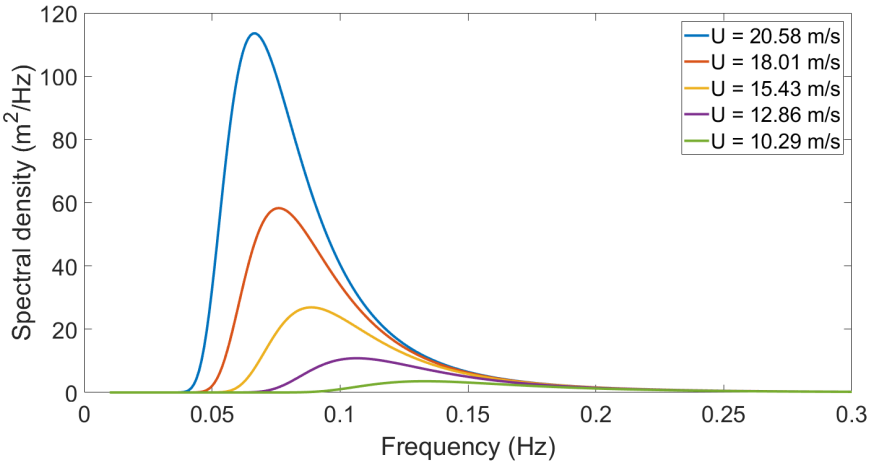


Figure 6-5 The Pierson-Moskowitz spectrum for varying wind speeds derived with the use of the spectral density equation (See equation 2.1 in Chapter 2).

characteristic of also being in close proximity to each other in terms of frequency. Additionally, the energy values of the peak and those waves close to the peak in terms of frequency have significantly higher energy values than the energy values of waves of higher frequencies. In figure 6-5 the energy spectra for the same sea states as in figure 6-4 are formed with the Pierson-Moskowitz equation (See equation 2.1 in Chapter 2). With the amplitude dispersion property, waves with higher amplitude (and energy) have higher phase speed values [172]. In essence, we are trying to select the entries from the phase speed matrix where the high values from frequency dis-

persion and amplitude dispersion converge. This is the intuition behind the heuristic described in the previous paragraph in order to solve the temporal and spatial frequency association problem, and thus solve the pixel to metre scale and sea state problem.

It is conjectured that these dominant waves are less affected by all types of present noise and unmodelled effects in the data due to their high amplitude. It should be noted that in general the energy of a given wave is directly proportional to the square of its amplitude. Thus the amplitude of waves with high energy would overcome the amplitude of noise and be more easily distinguished rather than waves with amplitudes near or lower than the amplitude of noise in the data. From the selected entries, only their position is used for calculating the pixel to metre scale, and not their energy values.

6.1.5 Pixel to metre scale calculation

Each selected entry in the phase speed matrix of the previous step (c_i, r_i), where c_i is the column number and r_i is the row number, corresponds to a set of temporal and spatial frequencies, or inversely, period and wavelength. The temporal frequency f in this case is equal to:

$$f = \frac{vc_i}{N} \tag{6.2}$$

where v is the frame rate and N is the number of frames. The spatial frequency f_s is equal to:

$$f_s = \frac{r_i}{L} \tag{6.3}$$

where L is the width size of the frames in terms of number of pixels. Having found the coordinates of the waves of interest with the heuristic described above, for each one the pixel to metre scale metric ζ is calculated as:

$$\zeta = \frac{2\pi u^2 c_i^2 L}{g N^2 r_i} \tag{6.4}$$

where g is the gravitational constant. In order to derive this equation, first the pixel to metre scale metric is defined as wavelength in pixels λ_p by wavelength in metres λ_m as:

$$\zeta = \frac{\lambda_p}{\lambda_m} \tag{6.5}$$

The wavelength in pixels can be found from equation (6.3) as the inverse of the spatial frequency and is equal to:

$$\lambda_p = \frac{L}{r_i} \tag{6.6}$$

In order to get the wavelength in metres, the period is found from equation (6.2) as the inverse of temporal frequency, and the dispersion relation is used as in equation (2.17) in order to get the wavelength in metres λ_m which is found to be equal to:

$$\lambda_m = \frac{gN^2}{2\pi u^2 c_i^2} \tag{6.7}$$

Having a pixel to metre scale value for each selected entry of the phase speed matrix, the final pixel to metre scale is estimated as the average of these values.

6.1.6 Sea state

Having calculated the pixel to metre scale, it is possible to work with the temporal information or spatial information in order to provide estimations of the sea state. With the use of remote observations of the ocean with radar or an aeroplane at a high altitude, the spatial information is sufficient for capturing the dominant wave in space. For example, from the Beaufort scale, a sea state 3 has a typical wavelength of 41.46 m (from the dispersion relation equation (2.17) using the period found from the Pierson-Moskowitz spectrum equation (2.3) with known wind speed), and at least two cycles of the dominant wave in space are needed in order to work with the spatial information [118]. As the shipborne radar captures an area of the scale of a few hundred metres, the spatial variance of the wave field is wide enough to make it possible to extract the spatial information of the dominant wave. The same applies for airborne video.

From the phase speed matrix, it is possible to work with temporal frequencies or with spatial frequencies in order to estimate the sea state. The input data set (radar images) have a lower frame rate compared to regular video, as the time between successive frames is equal to approximately 2 s. Because of the larger time between successive frames, in the case of the input data, it is conjectured that the accuracy measures would be less accurate with the use of the temporal frequencies directly. It is an intuition that a wave is captured more accurately when the time between measurements is smaller. With this type of data in the input, we hypothesise that using the spatial frequencies is more meaningful, as the radar data (and aeroplane data in the case of video in real environments) capture a wide ocean area, enough to track the dominant wave from the spatial variation. Even if radar images are not used, in the case of aeroplane images, the time between successive frames is still higher than regular video (e.g. with real airborne ocean images the time between successive frames is 1 s in [133]).

From the steps followed for finding the pixel to metre scale, the entries in the phase speed matrix that have the highest energy values and are at a close proximity to each other in terms of frequency (See Sect. 6.1.4) are again used for determining the sea state. That is because the dominant wave is the wave with the highest energy value in the ocean wave spectrum, and its information is contained inside the selected entries of the phase speed matrix. These entries represent the waves of the ocean spectrum with the highest energies or amplitudes. The coordinates of these entries are used in order to get the information of the frequency of the dominant wave.

For each selected entry, its spatial wavelength in pixels is calculated from its row coordinate as in equation (6.3). The wavelength in metres is then found by multiplying the inverse of the pixel to metre scale ζ of the previous step with the wavelength in pixels. Specifically, for each selected entry of the phase speed matrix, its wavelength in metres is equal to:

$$\lambda_m = \frac{1}{\zeta}\lambda_p \tag{6.8}$$

The dispersion relation and equation (2.17) are then used in order to get the period

in seconds. The frequency of that wave is then the reciprocal of the period. The final frequency of the dominant wave is calculated by averaging the temporal frequencies of the selected entries from the previous step.

Ocean theory, and specifically theory of customised energy spectra, is used for getting additional information about the sea state. Theory from the Pierson-Moskowitz spectrum [127] is used for getting the wind speed of the dominant wave. From this customised energy spectrum, equation (2.4) provides a wind speed estimation from the frequency of the dominant wave. The dominant wavelength is found as the average of the wavelength in metres of each selected entry of the phase speed matrix found in equation (6.8). The dominant period is the average of the period of each selected entry or the reciprocal of the final dominant frequency.

6.2 Experimental results

The experimental data used come from experiments performed in 2014 in the Atlantic sea (see Sect. 5.1). The marine radar X-band Sigma S6 WaMoS II is used for the remote sensing measurements. Two buoys are concurrently measuring the wave elevation. The radar system covers large areas of ocean surface of up to 4 kilometres, enough to capture the dominant wave from the video spatial variation, even in higher sea states. First, the ground range resolution or pixel to metre scale technique is tested.

6.2.1 Pixel to metre scale

For testing the pixel to metre scale technique, videos of various lengths are used as input (first column of table 6.1). The second column from the table presents the scale estimation from the methodology proposed in this chapter and the third column the scale estimation from the methodology proposed in [146]. The true scale is equal to 1/6 px/m, as shown in the following column. The two last columns of table 6.1 contain the mean absolute percentage error metric (See Sect. 5.5.3) for the methodology presented here and the work by [146] respectively.

Table 6.1 Experimental results for pixel to metre scale with varying parameter of number of frames. Columns 1: number of frames, 2: pixel scale from present work, 3: pixel scale from [146], 4: radar pixel scale (considered the true value), 5: MAPE from present work in comparison to the radar pixel scale, 6: MAPE from [146] in comparison to the radar pixel scale.

Frames	**Scale (m/px)**	**Scale [146] (m/px)**	**Scale radar (m/px)**	**MAPE columns 2-4 %**	**MAPE columns 3-4 %**
400	6.737	5.213	6.00	12.28	13.11
450	6.83	6.598	6.00	13.83	9.83
500	6.769	8.145	6.00	12.66	35.66
550	6.72	68.992	6.00	12.00	1.04×10^3
600	6.838	70.376	6.00	13.96	1.07×10^3
650	6.665	68.829	6.00	11.08	1.04×10^3
700	6.988	63.86	6.00	16.33	964.33
750	6.731	73.309	6.00	12.16	1.12×10^3
800	6.973	62.557	6.00	16.16	942.61

The results for the scale estimation are also presented in figures 6-6a and 6-6b. Figure 6-6a presents the results for different video durations varying from 50 up to 1000 with a step of 50 frames. Figure 6-6b presents the results for different times, with varying start frame from 100 to 1000 and increment of 50 frames. The sea state is not expected to have significant change during that time, enabling the testing for different sets of videos of the same state.

In the case the number of frames is below 500, both techniques estimate the true pixel to metre scale to a good degree. As the number of frames increases, the present technique appears to be resilient to the effect of frame number increase, providing accurate estimations for videos of higher number of frames. The estimations from the technique in [146] diverge in a large degree from the true value for videos with higher number of frames.

The speculation behind such effect is that the technique in [146] fits a parabola considering all waves, even the waves of higher frequencies, but does not take into account the effects of amplitude dispersion and spectral leakage (See Sect. 6.1.3). The present technique utilises the iterative process introduced in section 6.1.4 that aims to overcome the estimation problems associated with the aforementioned effects by selecting only the waves with the highest energy from the phase speed matrix.

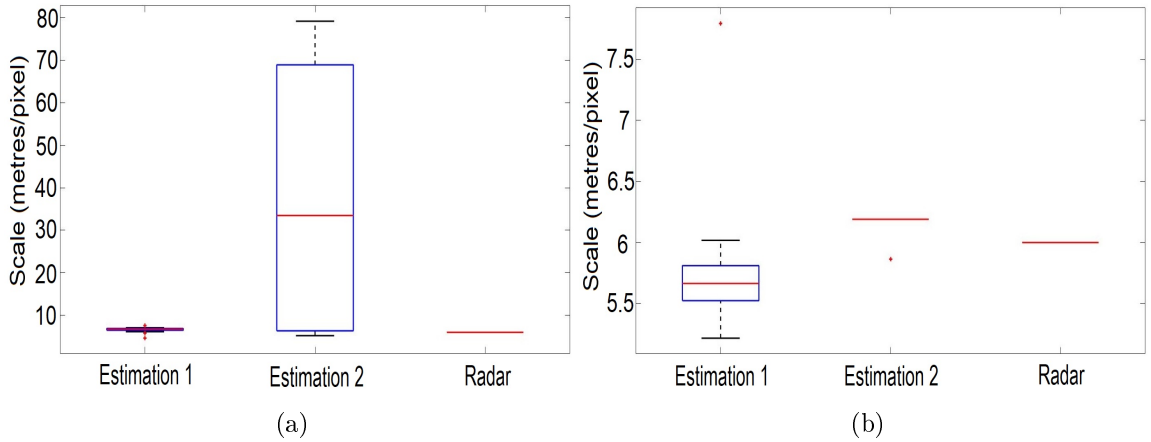


Figure 6-6 Experimental results for pixel to metre scale. (a) Varying parameter of number of frames, (b) varying parameter of start frame (i.e. videos at different times). Columns: 'Estimation 1': metholology in present work, 'Estimation 2': methodology in [146], 'Radar': pixel scale from radar data, considered to be the true value.

6.2.2 Sea state

Two areas of ocean are selected and utilised from the radar data for testing the sea state technique. The experimental results for the wind speed are compared to the values of wind speed from the 2 buoys that concurrently measure the ocean surface in a nearby location. The estimations from the present technique are additionally compared with the estimations from applying directly Fourier transform on the spatial variation of the data.

In a similar manner as with getting the phase modulation matrix in section 6.1.1, a Fourier transform is performed on each frame in the dominant direction. Then the magnitude values for each same spatial frequency are averaged. The result is a magnitude value for each spatial frequency or wavelength in pixels. The coordinate of the peak is identified and its wavelength in pixels is found to be equal to y_i/N, where y_i is the position of the peak and N is the length of each row in pixels. Then, in a similar manner as before in section 6.1.6 (see equations (6.8) and (2.17)) the dominant frequency is acquired and with equation (2.4) the wind speed. This process is done solely for testing whether the technique of applying directly Fourier transform on the spatial variation provides accurate results, and how these estimations compare

Table 6.2 Sea state experimental results. Two different areas of ocean are utilised as input. In the first column, the wind speed from the present technique is compared to the wind speed from the buoys. In the second column, the technique of selecting the peak from the Fourier transform of the spatial variation is compated to the wind speed from the buoys.

Area	**Present MAPE (%)**	**DFT MAPE (%)**
Ocean area 1	20.08	39.58
Ocean area 2	20.06	39.45

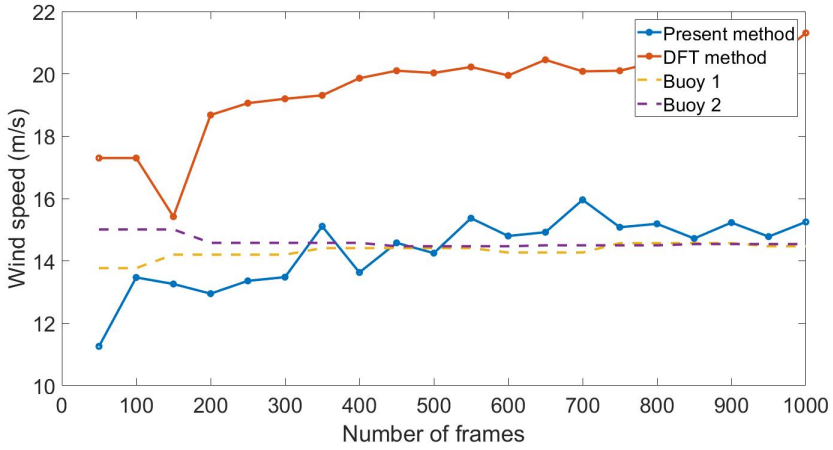
to the present technique for estimating the sea state.

The results of directly applying a Fourier transform on the spatial variation of the data described in the previous paragraph are included and compared to the results of the present technique and the sea state from the 2 buoys in figures 6-7a and 6-7b. Table 6.2 presents the accuracy metrics from the experimental results.

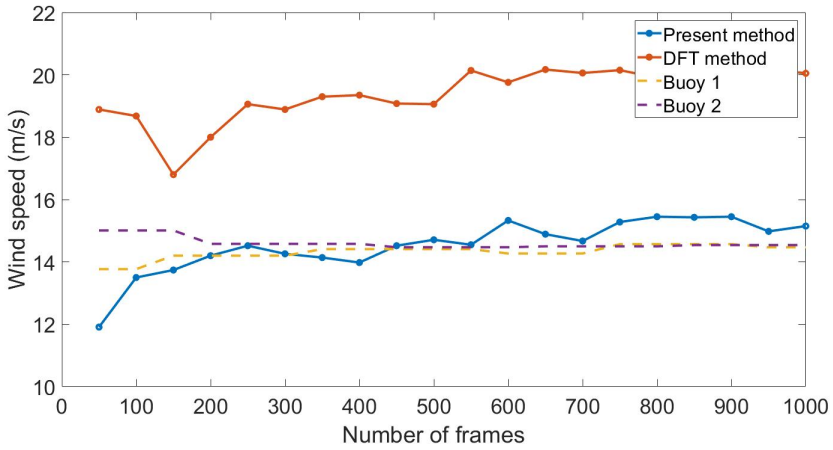
The present methodology finds a wind speed value that is closer to the wind speed of the 2 buoys. The method of directly applying a Fourier transform on the spatial variation overestimates what the 2 buoys show to be the true state. Regarding the present methodology, it can be observed that as the number of frames increases, so does the accuracy of the estimation of wind speed. Additionally, even smaller sizes of radar video, such as 100 frames, are sufficient for getting estimations of the wind speed.

It can be concluded that simply performing a Fourier transform on the spatial variation of the data is not sufficient for identifying the dominant wave, even if the data are of a large spatial scale to include the necessary information about the dominant wave, such as the data used in this chapter. The heuristic of using the waves from the phase speed matrix with the highest energy values that are in close proximity to each other in terms of frequency provides wind speed results that are more close to the buoy wind speed.

The higher accuracy of the present methodology is attributed to the fact that the waves with the highest energy are more resilient to the effects of noise. This noise can be caused by the nature of the video data, the discretisation of the continuous wave field or the presence of non-modelled phenomena, such as the effects of capillary



(a) Ocean surface area 1.



(b) Ocean surface area 2.

Figure 6-7 Sea state experimental results. Two different areas of ocean are utilised as input. The wind speed estimation from the present technique is compared with the technique of directly applying DFT on the spatial variation and the wind speed from the 2 buoys.

waves and the amplitude dispersion.

6.3 Conclusion

Remote observations of the ocean are utilised for acquiring information about the pixel to metre scale and sea state of the measured ocean. The introduced methodology aims to use the spatial variation of the wave field from ocean video for the sea state estimation. Because the spatial variation from video is utilised, the ocean videos

are required to capture wide areas of ocean. Examples of practical applications of the proposed methodology for estimating the sea state are videos captured from an airplane, drone or a tower. Practical applications for the method for the pixel to metre scale include videos recorded in the past that can be used for providing information about the size of objects in the scene, such as ships or platforms, or videos where this information is not readily available.

Possible main sources of error are non-uniform illumination of the environment (brightness fluctuations), which includes the direct reflection of the sun's rays on the camera lens. Such effects might have an effect on the recording of the spatial variation in real video. Future work could include the development of metholodogies that are based on the concepts discussed in this chapter that estimate the same metrics from real high-scale video of the ocean with the use of the spatial variation of the wave field.

This chapter provides only an introduction to the use of the spatial variation of the wave field and ocean theory for getting useful information from remote observations of the ocean surface. The difficulty of obtaining spatial information of the ocean in large scales enables the search for the use of the temporal information of the wave field (i.e. pixel variation in time from video). In such cases, as the period of the dominant wave is studied, even small duration videos are sufficient to capture the dominant wave in time. This is the reason further elaboration on the use of the spatial variation is not done in this work. In the following chapters, the temporal information of the wave field from video is studied in detail.

Chapter 7

Significant wave height from ocean video amplitude spectrum

This chapter is based on the research paper:

A. Loizou and J. Christmas, 'Sea state estimation from uncalibrated, monoscopic video,' SN COMPUT. SCI. 2, 328 (2021). https://doi.org/10.1007/s42979-021-00727-0

In the previous chapter, the pixel to metre scale is estimated from ocean video by identifying the waves with the highest energy from the phase modulation matrix, a structure that associates temporal and spatial information from video. The spatial variation of the wave field is then used in order to estimate the sea state. The methodology for the sea state of the previous chapter is applicable to high-scale ocean video, such as airborne video. This is because at least two times the ocean dominant wavelength needs to be captured in space [140].

If the temporal variation of the wave field is used instead, the video duration needs to be at least two times the dominant period of the ocean. Since even higher sea states have dominant periods of a few seconds, short duration videos can be used without any limitations in the space coverage. In the previous chapter, radar data is used for validating the proposed methodologies. However, such data are much more refined than video of the ocean in real environments. In the case of real environments,

brightness fluctuations, movement of the camera, effects due to camera function and other effects make videos more challenging to use for obtaining useful information about the state of the sea.

Although pixel intensity from video is not equal to ocean wave elevation, by following the pixel intensities in time from video is conjectured to provide useful information about the state of the ocean. Although the data has large amounts of noise (noise being all effects not associated with the movement of the ocean), in this chapter efforts are made to use the temporal variation of the wave field for estimating the significant wave height.

The methodology described in the following text attempts to track the movement of the ocean despite the amount of non-modelled elements in the video (not associated with the ocean movement). The Kalman filter is used for this purpose. The environment is configured in such a way that the filter aims to track the sinusoidal component from the video time series of pixel intensity that is related with the ocean's movement. This is executed for a set of predefined frequencies, providing an uncalibrated spectrum of amplitudes, here defined as the *ocean video amplitude spectrum.*

The methodology starts by incorporating the Kalman filter (Sect. 7.1.1) and the least squares approximate solution (Sect. 7.1.2) in order to provide the uncalibrated video amplitude spectrum (Sect. 7.1.3). The next problem that the present methodology addresses is the scaling or calibration problem. This is not a new problem, as it is used widely in research that uses radar data (e.g. [22]) or video data (e.g. [31]). In some cases, such as in [31], *in situ* devices (e.g. wave gauges) are used for this calibration. This requires the presence of these devices concurrently measuring the ocean as the video. This work incorporates ocean theory for the calibration (e.g. [146]).

In the calibration step (Sect. 7.1.4), the present methodology uses a process that is based on the shape of the video amplitude spectrum found in the previous steps and theory of an empirical ocean spectrum. This enables the transfer of the uncalibrated amplitude spectrum to real world units (i.e. metres). Although the video amplitude spectrum is still not the same as the true ocean energy spectrum (which can be

reliably found with the use of *in situ* devices), its shape does have some similarities to the true ocean video (See Sect. 7.2).

This amplitude shape is unique to each specific measured ocean and is not dependent on any generalised version of an energy spectrum, such as in the work of [146]. The generalised spectrum in this chapter is only used for calibration of the existing shape. With the scaled video amplitude spectrum, it is a straightforward next step to get an estimation of the significant wave height (Sect. 7.1.5). Although the mean of the one third of the highest waves from the video wave height spectrum (wave height is two times the amplitude) is still not the same as the mean of the one third of the highest waves of the true ocean wave height spectrum, the present work investigates whether these two values are close.

7.1 Methodology

The input of the proposed methodology of this chapter is video of the ocean surface from a single camera. For each pixel time series, an uncalibrated video amplitude shape is given in the output, with a methodology that uses the Kalman filter with the environment definition as described in section 7.1.1 and the least squares approximate solution (Sect. 7.1.2). Then, an averaged shape is formed by using the results of multiple time series (Sect. 7.1.3), and scaling is performed in order to transfer this shape to metres (Sect. 7.1.4). From this, the significant wave height is estimated (Sect. 7.1.5). The final significant wave height estimation is found as the average of each pixel time series estimation.

7.1.1 Kalman filter

The frequency domain is first determined. The maximum period or minimum frequency is the basic frequency and is equal to $1/t_{max}$, where t_{max} is the length of the video in seconds. The minimum period or maximum frequency depends on the sampling rate and is equal to $1/\Delta t$ where Δt is the time between two successive frames [140]. The Kalman filter as specified in the following text is performed for each fre-

quency of the frequency domain and for the same time series of pixel intensity in the input.

Here, the true state of the signal is defined as the element from video that represents the ocean movement, and the observation is the actual pixel intensity measurement. This is not to imply that the algorithm correctly captures the true state. The enviroment definition is used so that a distinction is made between what it is received from video and what the required or useful part is.

For a given frequency, the sinusoidal form after the observation is:

$$x_t^* = A\sin(\omega t + \phi) + \epsilon_t \tag{7.1}$$

where x_t^* is the pixel intensity at time t, A is the amplitude of the wave, ω is the angular frequency, ϕ is the phase and ϵ is the noise (the Kalman filter is defined for zero-mean Gaussian). The true signal is defined as:

$$x_t = A\sin(\omega t + \phi) \tag{7.2}$$

where x_t is the element from pixel intensity caused by the ocean movement. No input model is used in the Kalman filter environment definition. The derivative of the true signal is equal to:

$$\dot{x_t} = A\omega\cos(\omega t + \phi) \tag{7.3}$$

The second derivative of the signal is equal to:

$$\ddot{x_t} = -A\omega^2\sin(\omega t + \phi) \tag{7.4}$$

Comparing the form of the signal and its second derivative, the second derivative can be expressed in regards to the original signal:

$$\ddot{x} = -\omega^2 x \tag{7.5}$$

This is very useful in the environment definition, as it removes the amplitude and

phase, whose variance across the algorithm iterations might introduce errors, and focuses on the true signal x. Additionally, the Kalman filter makes use of the sinusoidal form of the input signal, providing more accurate estimations.

In essence, the Kalman filter is defined to track a sine wave with a particular frequency that moves across the pixel time series, with the presence of noise. It is defined to isolate the sinusoidal part that is moving across the time series from all other elements present in the video. And this is executed separately for each frequency.

The environment in state-space form is defined as:

$$\begin{pmatrix} \dot{x} \\ \ddot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} \tag{7.6}$$

where $\mathbf{F}$ is the system transition or dynamics matrix and is defined as:

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \tag{7.7}$$

The process noise could be used in order to reflect uncertainty in the frequency of the signal, or to reflect the change in the sea state with the passage of time. Here, with videos of approximately a minute, it is assumed that the sea is statistically stationary, and do not use initial values in the process noise matrix.

Practically, when running this algorithm with a mixture of multiple sinusoidal signals of different frequencies, amplitudes and phases, it was observed that the algorithm performed very well in estimating the true signal when only one frequency was given in the input for examination and the algorithm considered the rest as noise. This was the intuition for using this algorithm as it is configured here for the amplitude estimation from video of the ocean. Section 7.4 provides further information for solving the Kalman filter with the environment definition described here.

7.1.2 Least squares approximate solution

With the use of the Kalman filter, for one time series of pixel intensity from video and one frequency from the frequency domain, the output is the estimate of the position of the true signal with the given frequency. From this, the aim next is to get a value of amplitude (not in metres, in an uncalibrated metric in this step) of the estimated signal. This is achieved with the use of the least squares approximate solution.

Having the signal estimate that is given as output from the Kalman filter, each point x_i can be expressed:

$$x_i = A\sin(\omega t_i + \phi) \quad i = 0, 1, 2, ... \tag{7.8}$$

and the amplitude A is to be calculated. With three position estimations the vector form of their differences $\mathbf{\Delta x}$ can be expressed as: $\mathbf{\Delta x} = \mathbf{J}.\mathbf{y}$

$$\begin{pmatrix} x_1 - x_0 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} \sin\omega t_1 - \sin\omega t_0 & \cos\omega t_1 - \cos\omega t_0 \\ \sin\omega t_2 - \sin\omega t_1 & \cos\omega t_2 - \cos\omega t_1 \end{pmatrix} \begin{pmatrix} A\cos\phi \\ A\sin\phi \end{pmatrix} \tag{7.9}$$

where $y_0 = A\cos\phi$ and $y_1 = A\sin\phi$. The amplitude is then equal to $A = \sqrt{y_0^2 + y_1^2}$. In order to derive equation (7.9), the following trigonometric identity is used [3]:

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \tag{7.10}$$

With the use of equation (7.10), the difference between each pair of successive points (e.g. x_1 and x_0) can be expressed as:

$$\begin{aligned} x_1 - x_0 &= A\sin(\omega t_1 + \phi) - A\sin(\omega t_0 + \phi) \\ &= A(\sin\omega t_1\cos\phi + \cos\omega t_1\sin\phi - \sin\omega t_0\cos\phi - \cos\omega t_0\sin\phi) \\ &= A\cos\phi(\sin\omega t_1 - \sin\omega t_0) + A\sin\phi(\cos\omega t_1 - \cos\omega t_0) \end{aligned} \tag{7.11}$$

If all points are used, vectors $\mathbf{J}$ and $\mathbf{\Delta x}$ are extended to include the differences

between all points:

$$\begin{pmatrix} x_1 - x_0 \\ x_2 - x_1 \\ x_3 - x_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sin\omega t_1 - \sin\omega t_0 & \cos\omega t_1 - \cos\omega t_0 \\ \sin\omega t_2 - \sin\omega t_1 & \cos\omega t_2 - \cos\omega t_1 \\ \sin\omega t_3 - \sin\omega t_2 & \cos\omega t_3 - \cos\omega t_2 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} A\cos\phi \\ A\sin\phi \end{pmatrix} \tag{7.12}$$

and vector $\mathbf{y}$, which includes y_0 and y_1 is equal to:

$$\mathbf{y} = (\mathbf{J}^\mathsf{T}\mathbf{J})^{-1}\mathbf{J}^\mathsf{T}\boldsymbol{\Delta}\mathbf{x} \tag{7.13}$$

It should be noted that the signal amplitude estimation from the output of the Kalman filter is not performed for the first seconds of the video. This is to allow the algorithm to configure its internal variables.

7.1.3 Wave height shape

Until now, the Kalman filter is applied on each frequency of the defined frequency domain on one time series of pixel intensity, and the least squares approximate solution is used for getting a value of the amplitude of the signal estimation. Taking all the frequencies of the frequency domain, the shape of the amplitudes from video is formed. If multiple pixel intensity time series are selected from each video, multiple shapes are averaged in order to extract a final averaged video amplitude spectrum. Example of such shapes are given in figures 7-3b, 7-3e and 7-3h, with the difference that until this step, the shape is given in an uncalibrated metric.

Next, it is investigated if useful information about the sea state can be derived from this shape. In the next step, the shape is used in order to scale the results into metres (Sect. 7.1.4) and then get the significant wave height (Sect. 7.1.5).

7.1.4 Scaling to metres

With the uncalibrated averaged amplitude spectrum of the ocean video from the previous steps, the amplitude multiplier variable α is introduced here, and is used in order to scale this spectrum to metres. This involves the use of an empirical spectrum. No *in situ* devices are required to be present. The amplitude multiplier is defined as:

$$\alpha = \frac{a_{pm}}{a_u} \tag{7.14}$$

where a_{pm} is a value in metres from the Pierson – Moskowitz spectrum (Sect. 2.2) and a_u is a value from the uncalibrated spectrum.

The key in this process is the calculation of a_u. After this step, the calculation of a_{pm} is straightforward. The a_u variable represents the peak of the uncalibrated amplitude spectrum. Unlike the empirical energy spectrum, which was formed by averaging a set of spectra for the same sea state with *in situ* devices, the peak of the uncalibrated amplitude spectrum from video does not necessarily represent the ocean dominant frequency. The value of a_u is found as the average of the amplitude of a number of selected frequencies from the video uncalibrated amplitude spectrum. In the following text the process for acquiring the selected frequencies is described.

First, the amplitudes of the uncalibrated video spectrum are sorted in descending order and the frequency of the peak amplitude is selected. Then, for each next frequency, an intermediate variable ξ is calculated as:

$$\xi = \frac{a_p - a_c}{a_p} \tag{7.15}$$

where a_p is the peak amplitude and a_c is the current amplitude. Then a threshold of 30% is set, and if the value of ξ is lower than the threshold, the frequency associated with the current amplitude is selected. If it exceeds the threshold, the procedure ends. The amplitudes of the selected frequencies are then averaged to give the value of a_u in equation (7.14). The threshold value is determined empirically and the purpose is to include additional frequencies and not determine the scaling based solely on the

video peak frequency.

Before moving forward, a note about the intuition behind this method. While working with various sea states, it was observed that the amplitude spectrum would have bigger differences in amplitudes between its peak and the rest of the frequencies, whereas in lower sea states, the peak would have smaller amplitude differences than the rest. This is logical, as in higher sea states the amplitude of the peak is higher, and since the number of frequencies is fixed, the difference increases for higher sea states.

The presented scaling method will favour the inclusion of more frequencies in the averaging, and thus moving the average frequency to higher values for lower sea states, and will keep the average more concentrated to the peak in higher sea states. Additionally, the high amplitudes in the uncalibrated video amplitude spectrum are concentrated in lower frequencies for higher sea states and in higher frequencies for lower sea states.

The aim of the proposed scaling process is to not calibrate based solely on the position of the peak. In other words, the aim is to increase the bandwidth of frequencies considered for the calibration. The calibration in this case depends on the differences between the amplitudes of waves near the peak and the position of the peak.

In the experimental results (See Sect. 7.2.3) we perform a sensitivity analysis for the value of the threshold for variable ξ. We observe that the sea state estimation does not depend greatly on the value of the threshold, for threshold values between 10 % and 80 %. For very low values of threshold, the calibration depends strongly on the position of the peak, which is not always beneficial. For example, if there are multiple peaks present in the amplitude spectrum, it would be more beneficial to calibrate based on the position of these multiple peaks and not based solely on the position of the highest peak.

For very high values, a large number of waves are included for the calibration, and the calibration depends less on the position of the peak, which is not meaningful. From the sensitivity analysis, values of threshold for the variable ξ within the aforementioned range (10 % to 80 %) are not expected to provide significant changes

in the sea state estimation.

To continue the process, the average of the selected frequencies is then given as an input to equation (2.1) of the Pierson-Moskowitz spectrum. This will form an energy spectrum. The energy value of the peak of this spectrum is selected. Then the amplitude value of the peak of this spectrum in metres a_{pm} is found with the use of equation (2.12).

Having both a_{pm} and a_u of equation (7.14), the peak multiplier α is calculated, and is used for transferring the uncalibrated amplitude spectrum to metres. Examples of this final scaled video amplitude spectrum is given in figures 7-3b, 7-3e and 7-3h.

7.1.5 Significant wave height

The significant wave height is found from the video amplitude spectrum in metres after the scaling procedure as the one third of the highest waves of the ocean (see chapter 2, Sect. 2.5). Wave height is equal to two times the amplitude.

If the shape of the uncalibrated video amplitude spectrum is estimated correctly and the scaling process is executed in an efficient manner, it would be expected that this value from video is close to the true significant wave height, something that will investigated with two sets of experimental data in the following section.

7.2 Experimental results

The presented methodology is tested with real video and with the presence of *in situ* buoy devices. The first set of videos is taken from a moving shipborne camera (Sect. 7.2.1) and the second set from live video footage from a tower, i.e. fixed location (Sect. 7.2.2). Both sets of video have concurrent buoy measurements for validation. Details about the experimental data are included in chapter 4. A sample sensitivity analysis is provided in Sect. 7.2.3 in order to investigate the behaviour of the methodology for varying values of the threshold for the variable ξ. In Sect. 7.2.4 a discussion is given on the extension of the experimental results in order to include theoretical bounds of confidence. Because this extension of the sea state estimation

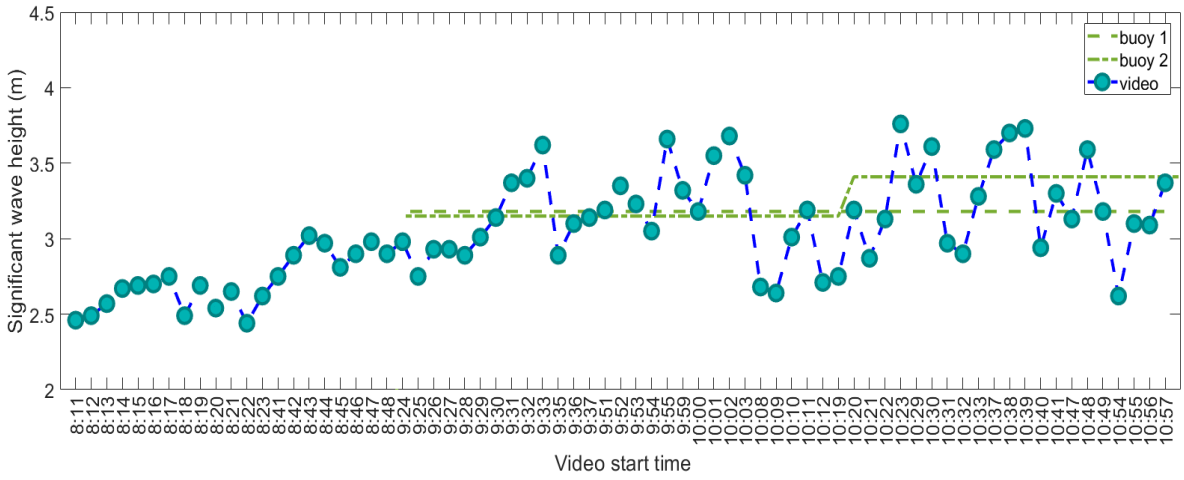


Figure 7-1 Shipborne video showing stability of video estimations of a similar sea state and correlation of the significant wave height estimation between video and buoys.

is also applicable to the methodologies of the following chapters 8 and 9, an extended discussion on this topic is given in the discussion chapter 11, Sect. 11.2.

7.2.1 Ship video

The shipborne data results are presented in figure 7-1. All videos were captured on the same day, which is 24/11/2014. The buoys were on the sea surface after 9:15 a.m., and thus the sea state is not validated for videos of previous times.

Table 7.1 presents the error metrics from the shipborne data, i.e. the difference between the proposed method's estimation of the significant wave height and that obtained from the buoys. It only includes the videos with corresponding buoy sea state. Although some fluctuations do exist from video estimations around the buoy sea state, the metrics of MAE of 0.31 m and RMSE of 0.37 m, as well as MAPE of 9.86 % indicate that the methodology is promising. The small positive value of MPE

Table 7.1 Error metrics with shipborne video.

Error metric	Value
Mean Absolute Error (MAE)	0.31
Root Mean Square Error (RMSE)	0.37
Mean Absolute Percentage Error (MAPE)	9.86
Mean Percentage Error (MPE)	4.26

indicate that the results are underestimating the buoy measurements, but not in a large degree.

With ship data, multiple videos with approximately the same sea state are examined. In the following text the performance of the methodology is examined for a variety of sea states.

7.2.2 Tower video

To examine the performance of the methodology for a variety of sea states, video captured from the Frying Pan Ocean tower located 85' above the Atlantic ocean is used (see Sect. 5.2) and the estimations are verified with data from the nearby buoy station.

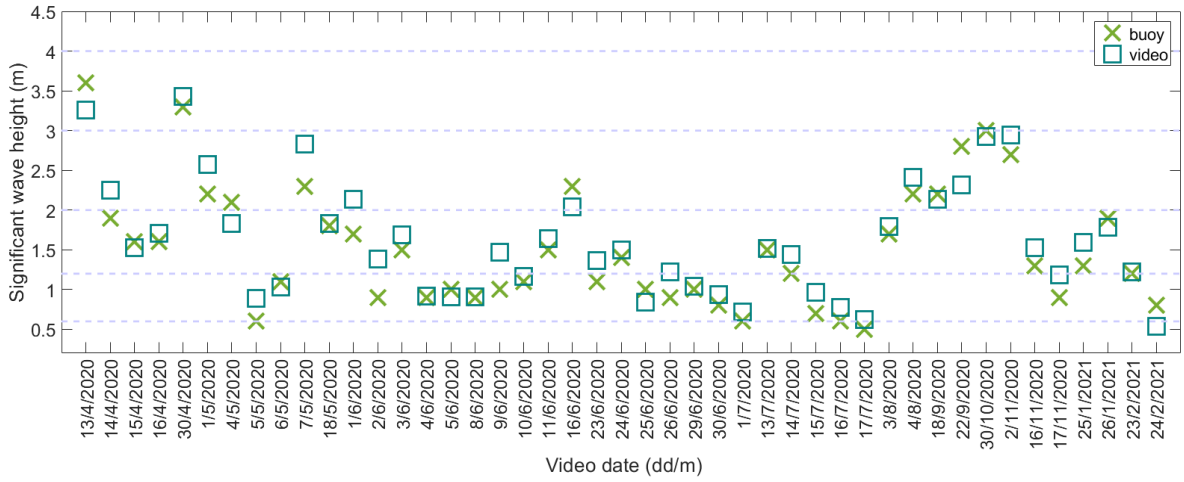
The tower video results are presented in figure 7-2. In each of these dates, one video is used, the significant wave height is estimated and is compared with the buoy's significant wave height of the same time. As the videos are captured in different dates, the proposed methodology is tested for a variety of sea states.

From figure 7-2b it can be observed how closely the video estimations are to the buoy measurements. This correlation between video estimations and buoy measurements is valid for both lower and higher sea states. From figure 7-2a it can be observed that the video methodology works across a significant time period in different sea states.

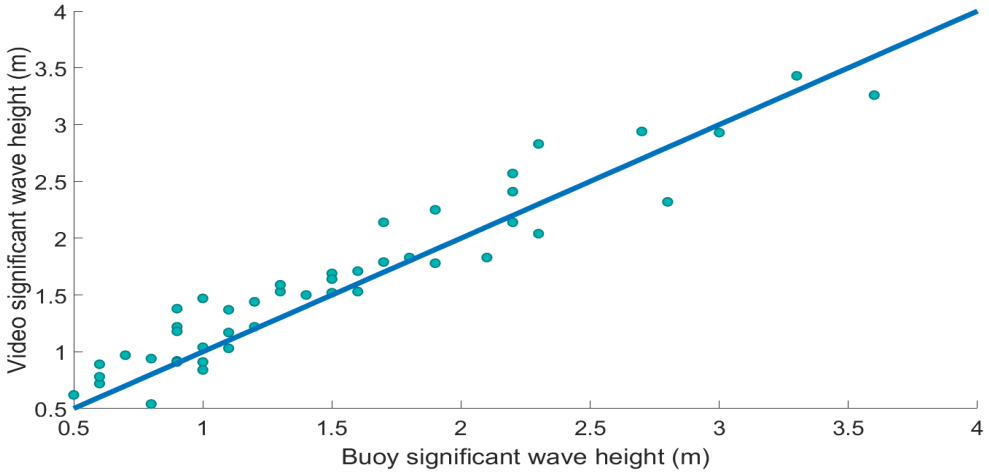
Figure 7-3 shows the buoy energy spectrum from three days with different sea states and the scaled video amplitude spectra. Table 7.2 presents the error metrics for the results of the tower video. The low values 0.20 m and 0.24 m of MAE and RMSE respectively indicate the relativeness between the video estimation and the

Table 7.2 Error metrics with tower video.

Error metric	Value
Mean Absolute Error (MAE)	0.20
Root Mean Square Error (RMSE)	0.24
Mean Absolute Percentage Error (MAPE)	16.14
Mean Percentage Error (MPE)	-10.35

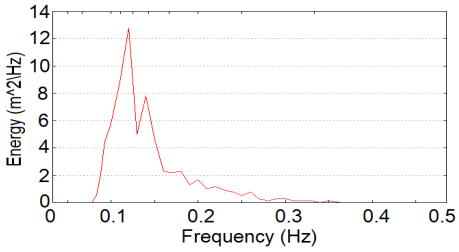


(a)

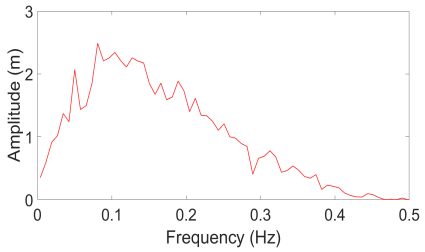


(b)

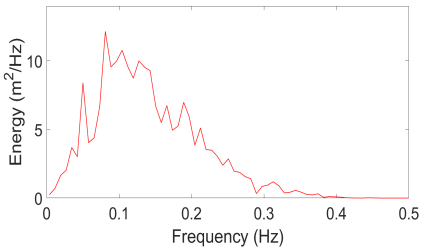
Figure 7-2 Tower video results showing correlation of the significant wave height estimation between video and buoy across a variety of sea states. The dashed lines in (a) present the upper and lower limits of wave height of Beaufort scale. Each space between dashed lines represents a different sea state according to the Beaufort scale. The blue line in (b) is the diagonal.



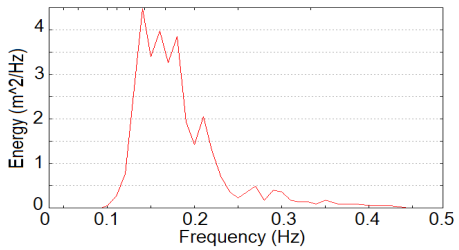
(a) Buoy energy spectrum 30/4/2020 at 8:40 a.m. ($h_s = 3.3m$).



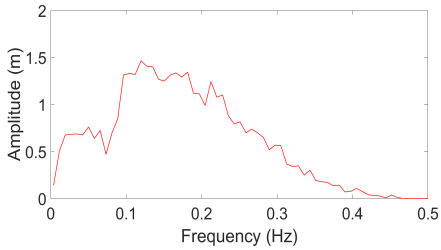
(b) Video amplitude spectrum 30/4/2020 at 8:41 a.m. ($h_s = 3.43m$).



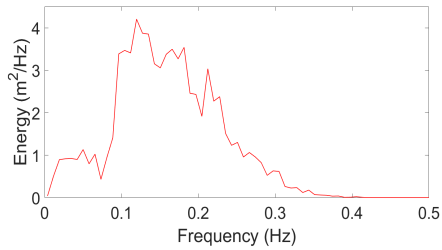
(c) Video energy spectrum ($h_s = 3.43m$).



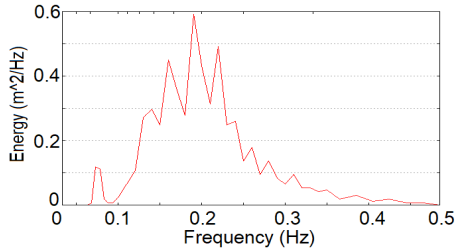
(d) Buoy energy spectrum 16/6/2020 at 8:27 a.m. ($h_s = 2.3m$).



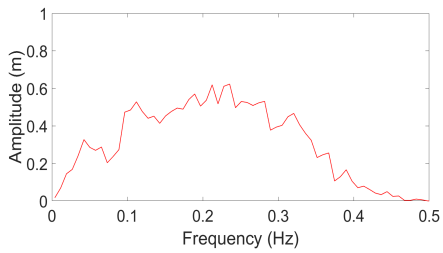
(e) Video amplitude spectrum 16/6/2020 at 8:27 a.m. ($h_s = 2.04m$).



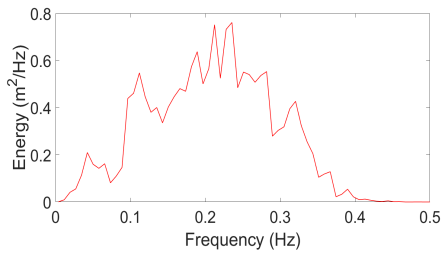
(f) Video energy spectrum ($h_s = 2.04m$).



(g) Buoy energy spectrum 9/6/2020 at 7:40 a.m. ($h_s = 1.0m$).



(h) Video amplitude spectrum 9/6/2020 at 8:17 a.m. ($h_s = 0.89m$).



(i) Video energy spectrum ($h_s = 0.89m$).

Figure 7-3 Tower video energy and amplitude spectra comparison with buoy energy spectra from [116], showing a correspondence between the peaks of video and buoy across a variety of sea states. The buoy energy spectra are taken from [116].

true sea state. The 16 % MAPE further validates the possible applicability of the video methodology. Contrarily to the results from the shipborne video, the negative value of MPE from the tower video indicate an overestimation. Possible reason for this difference between overestimation and underestimation could be the different environmental setup of these two data sets.

With values of significant wave height between 0.6 m and 3.6 m, the results are promising. The applicability for higher sea states is not known at this point. Although

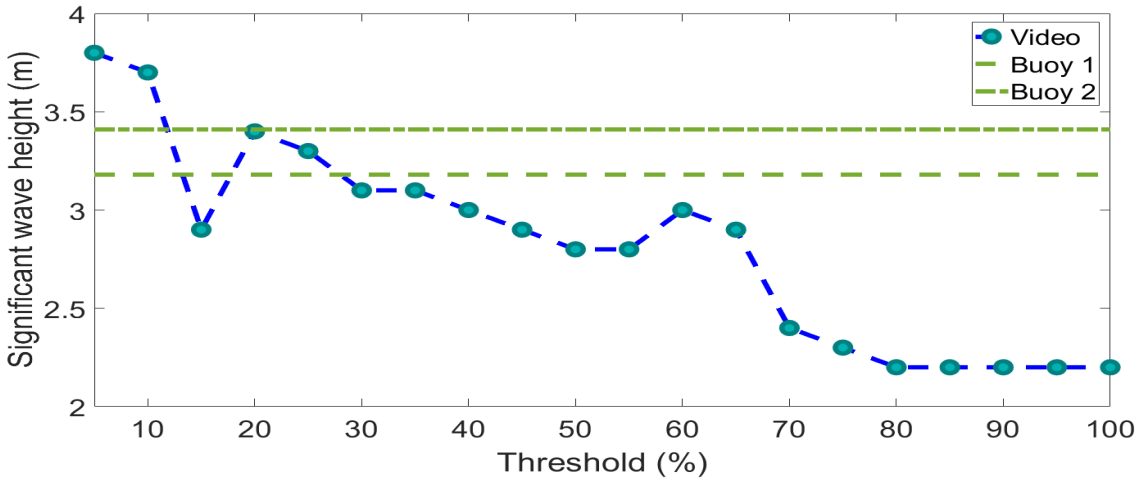


Figure 7-4 Sensitivity analysis of the methodology with shipborne video as input. The significant wave height estimation with a varying value of threshold for variable ξ. Significant wave height h_s Buoy 1: 3.18 m, Buoy 2: 3.41 m.

with ship data it is possible to check the variability of the estimations across multiple videos with approximately the same sea state, this is not available with the tower video due to the nature of the data.

7.2.3 Sensitivity analysis

It would be interesting to compare the sea state estimation with different values of the threshold for the variable ξ, introduced in the calibration process in Sect 7.1.4. The experimental results presented in Sect. 7.2.1 and 7.2.2 are derived with a threshold of 30 %.

The variable ξ (See Eq. 7.15) determines the distance in terms of amplitude between the peak and a given wave with a lower amplitude (the iteration process takes the waves in descending order of amplitude). If the threshold is too small and the distance between the two waves is larger, the iteration stops. Thus, as the threshold increases, more waves are included in the calibration process.

In figure 7-4 a sensitivity analysis is performed on a sample shipborne video, with varying values of threshold ranging between 5 % and 100 % with a step of 5 %. From the figure, we observe that for very low threshold values (5 %, 10 %), the sea state is overestimated. However, with maximum values of 0.4 m root mean square error

(RMSE) and 11 % mean absolute percentage error (MAPE), these estimations are within acceptable limits to be considered useful. This higher sea state estimation is expected, because as the threshold is decreased, so is the bandwidth of selected frequencies. The calibration is thus more dependent on the position of the peak. Because in this example we have a moderately high sea state of significant wave height 3.1 m-3.4 m, the peak is located in lower frequencies.

As we increase the threshold, we observe that the values for the sea state estimation decrease. For threshold values between 15 % and 65 %, the sea state estimation has a maximum error of 0.60 m RMSE and 17 % MAPE, which are in acceptable levels to show that the estimations are meaningful. For threshold values greater than or equal to 70 %, the sea state is underestimated in a larger degree, as higher frequencies are included in the calibration process.

As a summary, from this example, we observe that the sea state estimation does not vary in a large degree and remains in acceptable levels in comparison to the true values (here indicated from the measurements of the two buoys) for threshold values smaller than 70 %. Larger values of threshold do not seem to provide reliable sea state estimations. Thus, larger values of threshold should be avoided, as the frequency bandwidth becomes too wide, and the calibration is less dependent on the position of the peak.

We now perform a sensitivity analysis with a sample tower video that has a lower sea state of significant wave height equal to 1 m. In figure 7-5 we observe that for a very low value of threshold (5 %), the sea state is overestimated. This is possibly due to the fact that the calibration is more focused on the peak, and the peak is in a position not corresponding exactly to the sea state of the buoy. Thus, an inclusion of more frequencies is beneficial in this case.

For threshold values between 10 % and 80 %, the sea state estimation is within acceptable levels, with a maximum error of 0.20 m RMSE and 20 % MAPE. It should be noted that the relatively high value of MAPE in this case is due to the behaviour of the metric, which provides higher values of error when the pairs of estimated-true values are smaller. The low value of RMSE is a better indicator of the accuracy of

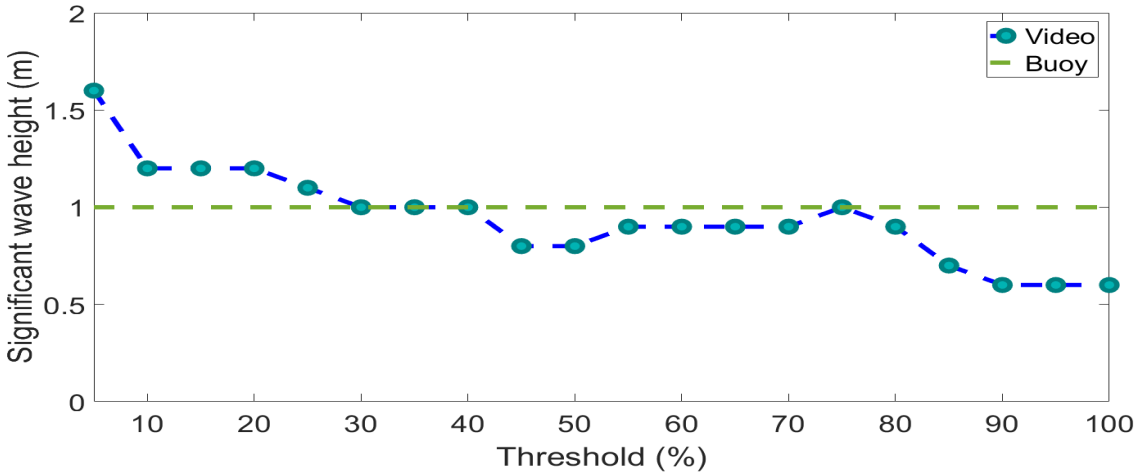


Figure 7-5 Sensitivity analysis of the methodology with tower video as input. The significant wave height estimation with a varying value of threshold for variable ξ. Significant wave height h_s Buoy: 1 m.

the methodology. For threshold values greater than or equal to 85 %, the sea state is underestimated, as more frequencies are included in the calibration process.

From the two examples, we observe that the sea state estimation is not too sensitive to the value of threshold for variable ξ. Only very low and very high values provide estimations with higher error. In the case of very low values, the calibration is based solely on the position of the peak, and thus could be more prone to errors. This is because the calibration algorithm will select the first highest peak, even if multiple peaks are available. Thus the model is 'overfitting'. In the case of very high values, the calibration process depends less on the position of the peak and considers a large number of frequencies. In this case the model is too general.

7.2.4 Theoretical bounds of confidence

It would be beneficial if the results in figure 7-2b were demonstrated in terms of theoretical bounds of confidence for the estimations. In such a case, each point in figure 7-2b would be instead represented as a set of mean and standard deviation of uncertainty with a coverage probability. For example, a given sea state estimation would be: significant wave height with mean μ equal to $\mu = 2\,\text{m} \pm 0.45\,\text{m}$ with a confidence interval of 95 %.

Because this extension of the experimental results can also be applied in chapters 8 and 9, an extended elaboration of how the confidence intervals can be formed is provided in the discussion chapter 11, Sect. 11.2. In chapter 11, Sect. 11.2.2 a discussion is given about possible sources of error in estimating the significant wave height and how they might contribute to the final error. As a summary of what is discussed in chapter 11, Sect. 11.2.2, possible sources of error such as insufficient quality of the video data, captures of the ocean at long distances, adverse weather conditions, presence of capillary waves and environmental brightness fluctuations can negatively affect the accuracy of estimation of the sea state, and thus increase the values of the error metrics.

It is also possible to use the bootstrap method [45] in order to calculate the theoretical bounds of confidence for the video estimations. This is discussed in chapter 11, Sect. 11.2.3 and 11.2.4. In Sect. 11.2.3, a discussion is given on possible ways to acquire multiple video estimations. This is required in order to generate the sample data, from which the bootstrap method will draw its samples in order to perform resampling with replacement. In Sect. 11.2.4 the functionality of the bootstrap method is described, and how it can be used in the context of this work.

As a summary to what is mentioned in the discussion chapter 11, Sect. 11.2.4, the bootstrap method selects different samples in the input and calculates the required metric (e.g. the mean). Thousands of resampling procedures are executed, recording the mean in each one. The algorithm simulates the process of running the same experiment under different conditions and thus obtaining different estimations thousands of times. This enables the calculation of the theoretical limits of confidence.

7.3 Conclusion

By testing with two sets of ocean video in real enviroments with corresponding buoy measurements, the methodology estimates the significant wave height in a good degree. However, the testing is limited to the range of the sea states of the videos. Higher sea states than 3.6 m of significant wave height are not tested. Future work

could focus on testing or improving the methodology for highest sea states. In higher sea states, the presence of foam is more apparent, which could introduce challenges in the estimations.

Future work will include improvements in the calibration process. One advantage of the methodology is the use of the video amplitude spectrum and not the shape of an empirical energy spectrum of the ocean for getting the sea state (the empirical energy spectrum is only used for calibration based on the video amplitude spectrum). The sea state estimation is thus more localised to each specific video in the input. One disadvantage from this however is the fact that the calibration depends on the shape of the ocean video amplitude spectrum. If this shape is not found in an efficient manner (e.g. due to effects in a given input video not present in the input videos of this work), the scaling process could yield inaccurate results.

In the following chapter, the video ocean energy spectrum is not used. Without it, no calibration is required. This makes things less complicated. Instead of focusing on estimating the significant wave height directly, the methodology in the following chapter focuses on estimating the dominant frequency. Theory of an empirical energy spectrum is then used for estimating the significant wave height.

7.4 Appendix

This section provides a further elaboration and solution to the Kalman filter algorithm with the environment definition descibed in section 7.1.1. The continuous fundamental matrix is found through the inverse Laplace transform and the transition matrix. The discrete fundamental matrix Φ_k is equal to the continuous by substituting time with the time between two successive measurements Δt and is equal to:

$$\Phi_k = \begin{pmatrix} \cos\omega\Delta t & \frac{\sin\omega\Delta t}{\omega} \\ -\omega\sin\omega\Delta t & \cos\omega\Delta t \end{pmatrix} \tag{7.16}$$

From the environment definition, the measurement matrix is found to be:

$$H = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{7.17}$$

The measurements are related to the states according to:

$$z_k = Hx_k + \zeta \tag{7.18}$$

where z is the measurement vector, and ζ is the measurement noise, which is zero-mean Gaussian.

The noise matrices are equal to:

$$W = E[\epsilon\epsilon^{\mathsf{T}}] \tag{7.19}$$

$$V = E[\zeta\zeta^{\mathsf{T}}] \tag{7.20}$$

where ϵ is the process noise and ζ is the measurement noise.

Very large values are set on the main diagonal of the initial covariance matrix to reflect the uncertainty of the initial state values (the variance of each variable). The process noise could be used in order to reflect uncertainty in the frequency of the signal, or to reflect the change in the sea state with the passage of time. Here, with videos of approximately a minute, the sea is approximately statistically stationary, and thus initial values are not used in the process noise matrix. Further information on how the Kalman filter described here can be solved is described in [9].

Chapter 8

Sea state from ocean video with singular spectrum analysis and extended Kalman filter

This chapter is based on the research paper:

A. Loizou and J. Christmas, 'Sea state from ocean video with singular spectrum analysis and extended Kalman filter,' SIViP (2021), Accepted subject to revisions

In chapter 6, the spatial variation of the wave field is used for getting the sea state in high-scale videos. The temporal and spatial domain are linked with the phase modulation matrix and the highest energy waves are used for getting the pixel to metre scale. In the previous chapter 7, the requirement for high-scale videos is no longer required, as the temporal variation from video is utilised from ocean video in real environments. The methodology calculates a video amplitude spectrum and ocean theory is used for calibration, which enables the estimation of a value for the significant wave height.

In this chapter, a methodology is presented for acquiring the significant wave height in a different manner. The methodology does not use the video amplitude spectrum. It thus does not require calibration. It focuses instead on estimating a

dominant frequency value from video that is close to the ocean dominant frequency. Then, ocean theory is used for transferring this into a value of the significant wave height.

Similarly to the previous chapter, the temporal variation from video is utilised in the form of time series of grayscale pixel intensities as input. The methodology uses the singular spectrum analysis (SSA) algorithm (Sect. 8.2.1) and the nonlinear Kalman filter (Sect. 8.2.2) for estimating a value of a dominant frequency, which is conjectured to be close to the ocean dominant frequency. Then, the Pierson-Moskowitz spectrum and theory of the ocean energy spectral moments are used for transfering this dominant frequency into an estimation of the significant wave height (Sect. 8.2.3). For testing the methodology, the same data sets are used as in the previous chapter. These are described in the experimental data chapter (chapter 5), sections 5.1 for shipborne video and 5.2 for tower video.

In the following text first an introduction is given to the general theory of the SSA algorithm (Sect. 8.1). Then, the methodology is presented in Sect. 8.2 followed by experimental results in Sect. 8.3.

8.1 Singular spectrum analysis (SSA) algorithm

Historically, the SSA algorithm is associated with work published in the 1980s, e.g. [26]. In the context of time series analysis, the SSA algorithm can be used for time series analysis [56], time series forecasting [57] and homogeneity analysis [58]. The algorithm aims to decompose the input signal into a set of additive components, each having a meaningful interpretation. These additive components can be labelled as either trend components, which are slowly varying components throughout the series, oscillatory components and noise components. In practice the SSA algorithm is applied to trend resolution [7][59], smoothing [8][60], extraction of seasonality components [61], extraction of periodicities [62] and noise identification and reduction [169][163][69]. In the following text, the execution of the algorithm is presented in section 8.1.1 and theory of the separability between reconstructed components is pro-

vided in section 8.1.2.

8.1.1 Execution steps of SSA algorithm

The execution of the SSA algorithm consists of the following steps: embedding, singular value decomposition, eigentriple grouping and diagonal averaging. In the embedding step, the algorithm takes a window length parameter L and expresses the input time series $\mathbb{X} = (x_1, \ldots, x_N)$ of length N as a set of $K = N - L + 1$ lagged vectors of size L:

$$X_i = (x_i, \ldots, x_{i+L-1})^\intercal \qquad (1 \leq i \leq K) \tag{8.1}$$

The trajectory matrix of the time series $\mathbb{X}$ is equal to [54]:

$$\mathbf{X} = [X_1 : \ldots : X_K] = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \ldots & x_N \end{pmatrix} \tag{8.2}$$

The rows and columns of the trajectory matrix $\mathbf{X}$ are subseries of the original time series. The trajectory matrix $\mathbf{X}$ is a Hankel matrix. This means that the entries on each parallel to the main anti-diagonal are equal.

In the next step, singular value decomposition (SVD) [149][83] is performed on the trajectory matrix $\mathbf{X}$. The advantage of this approach is the maximality of directionality of the configured eigenvectors. For example, the direction of the first eigenvector maximises the variation of the projections of the lagged vectors onto this direction. Each following eigenvector is orthogonal to the direction of all previous eigenvectors and its direction also maximises the variation of the projection of the lagged vectors onto this direction. The SVD of the trajectory matrix is equal to:

$$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^\intercal = \mathbf{X}_1 + \ldots + \mathbf{X}_d \tag{8.3}$$

where $d = \operatorname{rank} \mathbf{X} = \max \{i, \text{ such that } \lambda_i > 0\}$, $\lambda_1, \ldots, \lambda_d$ are the positive eigenvalues of $\mathbf{S} = \mathbf{X}\mathbf{X}^\intercal$ taken in decreasing order of magnitude, $U_1, \ldots, U_d$ is the orthonormal system of eigenvectors of $\mathbf{S}$ corresponding to these eigenvalues and $V_i = \mathbf{X}^\intercal U_i \sqrt{\lambda_i}$. For real-world time series, the value of d is typically equal to $\min(L, K)$ [54].

The eigentriples of eigenvalue $\sqrt{\lambda_i}$, left singular vectors U_i and right singular vectors V_i $(\sqrt{\lambda_i}, U_i, V_i)$ are grouped in the eigentriple grouping step so that the separation of the additive components is more optimal. Specifically, the set of indices of the eigentriples $\{1, \ldots, d\}$ is partitioned into a set of disjoint subsets of size m $I_1, \ldots, I_m$. If the set of the new indices is equal to $I = \{i_1, \ldots, i_p\} \subset \{1, \ldots, d\}$, the resultant matrix $\mathbf{X}_I$ corresponding to group I is equal to $\mathbf{X}_I = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_p}$ [54]. The decomposition of the grouping procedure is equal to:

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m} \tag{8.4}$$

According to the terminology in [63], in the following text the ith vector $z_i = \sqrt{\lambda_i} V_i$ is referred to as the ith *principal component* (PC).

In the diagonal averaging step, the result time series are reconstructed from the matrices of the grouping step by first performing hankelization to all matrix components and then obtaining the decomposition of the original time series by diagonal averaging to the resultant matrices. The initial time series is decomposed into a sum of m reconstructed time series:

$$x_n = \sum_{k=1}^{m} \tilde{x}_n^{(k)} \qquad (n = 1, 2, \ldots, N) \tag{8.5}$$

The output time series associated with each eigentriple is referred to in the following text as a *reconstructed component* (RC). Further information about the algorithm can be found in [55].

8.1.2 Separability between reconstructed components

The measure of separability between two reconstructed components (RCs) $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ is a measure of orthogonality between them [54]. This is due to the fact that the algorithm uses the biorthogonal SVD decomposition. Specifically, two given RCs $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are defined as *weakly separable* if their column trajectory spaces are orthogonal and their row trajectory spaces are also orthogonal [64]. The column trajectory spaces are the column spaces of the trajectory matrices and the row trajectory spaces are the row spaces of the trajectory matrices. The two RCs $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ are *strongly separable* if they are weakly separable and there is no intersection between the sets of singular values of their trajectory matrices [64].

The measure of separability is useful for grouping RCs that have weaker separability between them in the grouping step of the algorithm. This measure is also useful for categorising groups of RCs with stronger separability between the groups as trend, oscillatory or noise components. Specifically, slowly varying smooth components (trend) are asymptotically separable from oscillatory components and they both are asymptotically separable from noise. The tasks or grouping and categorising are important for the efficient execution of the SSA algorithm [64].

The orthogonality between two RCs $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ is equal to the orthogonality of their reconstructed trajectory matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. The orthogonality between $\mathbf{X}_1 = x_i^{(1)}$ and $\mathbf{X}_2 = x_i^{(2)}$ of length N can be found from their Frobenius inner product as [64]:

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_F = \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} \tag{8.6}$$

The normalised measure of orthogonality between $\mathbf{X}_1$ and $\mathbf{X}_2$ is equal to:

$$\rho\left(\mathbf{X}_1, \mathbf{X}_2\right) = \frac{\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_F}{\|\mathbf{X}_1\|_F \|\mathbf{X}_2\|_F} \tag{8.7}$$

The degree of approximate separability between RCs $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ is called weighted correlation or w-*correlation* and is found from the weighted normalised measure of orthogonality between $\mathbf{X}_1$ and $\mathbf{X}_2$. The w-*correlation* of RCs $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ is equal

to [64]:

$$\rho_w\left(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}\right) = \frac{\left(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}\right)_w}{\|\mathbb{X}^{(1)}\|_w \|\mathbb{X}^{(2)}\|_w} \tag{8.8}$$

where $\left(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}\right)_w = \sum_{i=1}^{N} w_i x_i^{(1)} x_i^{(2)}$ and the weights w_i are equal to:

$$w_i = \begin{cases} i, & \text{for } 0 \leq i < L^* \\ L^*, & \text{for } L^* \leq i \leq K^* \\ N - i + 1, & \text{for } K^* < i \leq N \end{cases} \tag{8.9}$$

where L is the window length, $K = N - L + 1$, $L^* = \min(L, K)$ and $K^* = \max(L, K)$.

The matrix of w-correlations contains the values of w-correlation between the elementary reconstructed components of a given time series in the input. This matrix can be used for detection of separability and identification of groups [65]. Components that are highly separable have a value of w-correlation near or equal to zero. Components that have weaker separability between them have higher values of w-correlation.

In the grouping step of the SSA algorithm (See Sect. 8.1.1), the matrix of w-correlations can be used as an indication in order to put RCs with strong correlation between them in the same group. The matrix of w-correlations can be represented graphically as a matrix of scaled values between 0 and 1, where a value of 0 indicates a strong separability and a value of 1 a high correlation. An example of such matrix with the use of a time series of pixel intensities from shipborne video as input is given in figure 8-1.

8.2 Methodology

The aim of this work is to identify the principal component of the movement of water from pixel time series of video of the ocean surface, which is conjectured to be related to the dominant wave of the ocean. This enables the estimation of the significant wave height. To achieve this, a methodology is introduced that combines the SSA algorithm and the nonlinear Kalman filter. It also incorporates ocean theory

presented in Sect. 2.2.

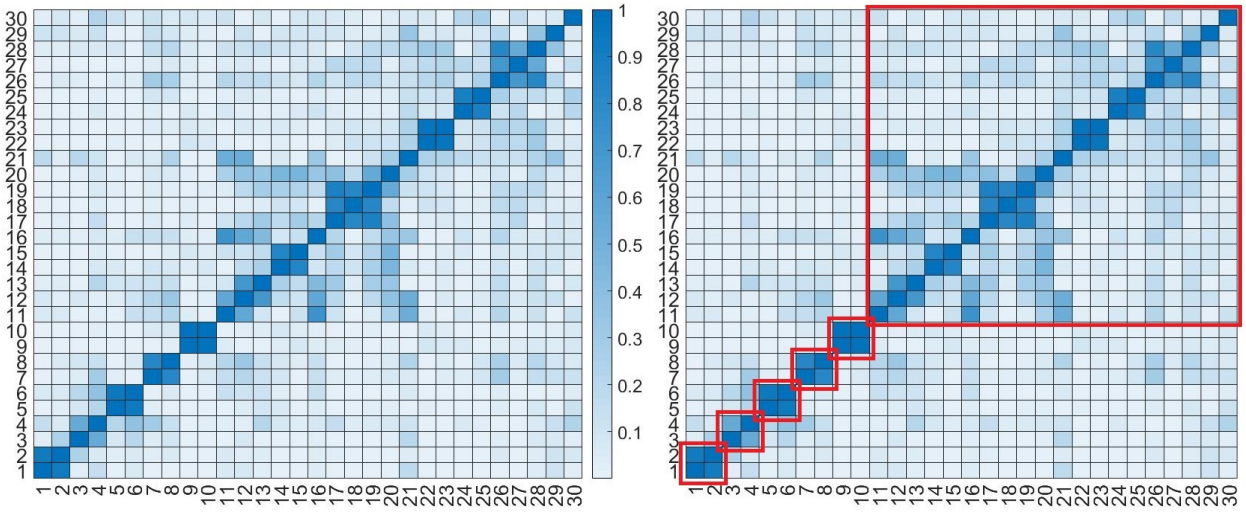
The input is time series of grayscale pixel intensity from ocean video. The methodology comprises of the following steps. First, the time series is given as input to the SSA algorithm (Sect. 8.2.1). Then, the output from the SSA is used by the nonlinear Kalman filter to get an estimation of the dominant frequency (Sect. 8.2.2). Finally, the shape of the Pierson-Moskowitz spectrum is used for getting the significant wave height (Sect. 8.2.3).

8.2.1 Singular Spectrum Analysis (SSA) algorithm

Information on how the SSA algorithm, introduced in section 8.1, is used in the context of this work is included in the following text. Due to its SVD nature, the SSA algorithm concentrates the information of the central component from the video, which is conjectured to be associated with the dominant wave of the ocean, in the first elementary reconstructed components (RCs). The rest of the RCs include additional video oscillatory components and then noise. Video time series of pixel intensity is given to the SSA algorithm with a given window length. Discussion about the selection of the window length is provided in the experimental results (See Sect. 8.3). The first four RCs are summed to provide a new time series, which is given to the extended Kalman filter in order to estimate the dominant frequency.

Although it would be expected that the dominant frequency is isolated in the first RC, practically this was not found to be the case. Instead, the dominant wave is concentrated in the first few RCs. Empirically, selecting the first four in all cases (See Sect. 8.3) was sufficient. This selection is based on observations from the matrix of w-correlations (See fig. 8-1). Specifically, in many cases it was observed that the first four RCs had a strong correlation with each other.

From the matrix of w-correlations it is observed that in many cases the first two or the first four elementary reconstructed components (RCs) have a strong correlation. It is also observed that components after RC 10 contain more noise. Additionally, there is a strong correlation between different components, such as RCs 3-4 and RCs 5-6 (See Fig. 8-1).



(a) Original. (b) Red boxes added for illustration purposes.

Figure 8-1 Example of matrix of w-correlations of SSA algorithm from ocean video. From this example, a correlation of certain RCs is visible, such as RCs 1-2, 3-4, 5-6, 7-8. Additionally, some higher RCs can be considered to contain more noise (RCs 11-max).

From empirical observations with multiple videos, the selection of only the first two RCs does not provide accurate estimations. The selection of RCs 1-4, 1-6 and 1-8 provides more accurate estimations. From RCs 1-10 and higher the estimations become less accurate. Possible reason for this effect is the inclusion of more noise as the component number increases.

The next step involves determining the dominant frequency from the sum of RCs 1-4 of the SSA algorithm. Practically, the Fourier transform of this time series will include more than one peak and selecting the highest peak does not in all cases correspond to the correct dominant frequency of the ocean (see for example figure 8-4a). Since this selection of RCs usually includes more than one waves, determining one frequency out of them is not a straightforward task. This is the reason the extended Kalman filter is used in the following step.

8.2.2 Extended Kalman filter algorithm

The role of the general version of the Kalman filter in the context of linear dynamical systems is described in the Kalman filter chapter (See Sect. 3.3). The application of the Kalman filter for nonlinear systems was first introduced in [109][143]. The environment definition is nonlinear and is described by differential equations. To compute the state transition matrix, the Jacobian is computed at each time step with the current estimates of the states. A more detailed introduction about the nonlinear version of the Kalman filter is available at [157].

In the context of this work, the algorithm with the environment definition described below is very efficient at distinguishing one main frequency and isolating the remaining video elements as noise. Again, the goal is to identify the principal component of water movement from the sum of RCs 1-4, which is hypothesised to be related to the dominant wave of the ocean. The true signal is a sinusoid:

$$x = A\sin(\omega t + \phi) \tag{8.10}$$

where A is the amplitude and ϕ the phase. The derivative of the signal is equal to:

$$\dot{x} = A\omega\cos(\omega t + \phi) \tag{8.11}$$

and the second derivative:

$$\ddot{x} = -A\omega^2\sin(\omega t + \phi) \tag{8.12}$$

The second derivative can be expressed as a function of the angular frequency and the true signal as:

$$\ddot{x} = -\omega^2 x \tag{8.13}$$

This is especially useful in the context of this work, because it does not include amplitude and phase, and instead focuses on the true signal. Additionally, the sinusoidal form of the signal is now included in the environment definition. The environment

definition is:

$$\mathbf{g} = \begin{pmatrix} \dot{x} \\ \ddot{x} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -\omega^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \omega \end{pmatrix} = \mathbf{q} \begin{pmatrix} x \\ \dot{x} \\ \omega \end{pmatrix} \tag{8.14}$$

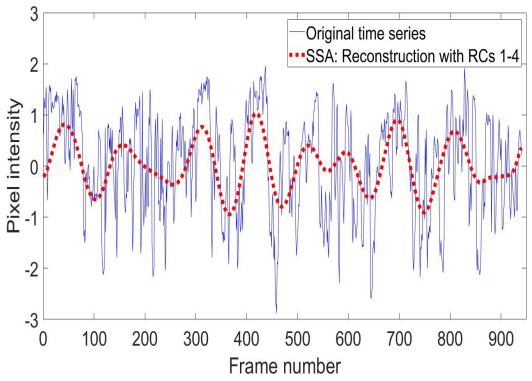
The matrix $\mathbf{q}$ contains a term that is also part of the states of the system, thus making the system nonlinear and this matrix cannot be applied in the covariance directly. In order to find the state transition matrix $\mathbf{F}$, the Jacobian of matrix $\mathbf{g}$ is computed at each time step with the current estimates of the states. By taking the partial derivatives of matrix $\mathbf{g}$, the systems dynamics matrix is found to be equal to:

$$\mathbf{F} = \frac{\partial \mathbf{g}}{\partial x} = \begin{pmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial \dot{x}} & \frac{\partial \dot{x}}{\partial \omega} \\ \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial \dot{x}} & \frac{\partial \ddot{x}}{\partial \omega} \\ \frac{\partial \dot{\omega}}{\partial x} & \frac{\partial \dot{\omega}}{\partial \dot{x}} & \frac{\partial \dot{\omega}}{\partial \omega} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -\hat{\omega}^2 & 0 & -2\hat{\omega}\hat{x} \\ 0 & 0 & 0 \end{pmatrix} \tag{8.15}$$

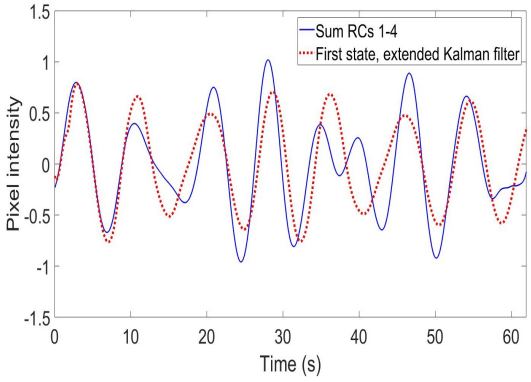
where $\hat{\omega}$ is the predicted value or estimate of ω and similarly $\hat{x}$ is the predicted value or estimate of x at the current time step. The fundamental matrix is not used for propagating the states, but rather only for the calculation of the gains, and is approximated with the first two terms of Taylor series. Detailed information on how the extended Kalman filter with the environment definition described here is solved can be found in Sect. 8.5.

The Kalman filter outputs a value of the unknown angular dominant frequency. Although this can be used directly, in the following text this value is given as input to the theory of the Pierson-Moskowitz spectrum (see Sect. 2.2) in order to get a value of the significant wave height. As a sidenote, the described methodology is performed on one pixel time series. For acquiring more accurate and reliable dominant frequency estimations, a set of pixels (or all pixels) can be used individually and the dominant frequency is found as the average. In the case of the experimental results of this work (Sec. 8.3), a set of pixels was selected from a horizontal line across the video frame.

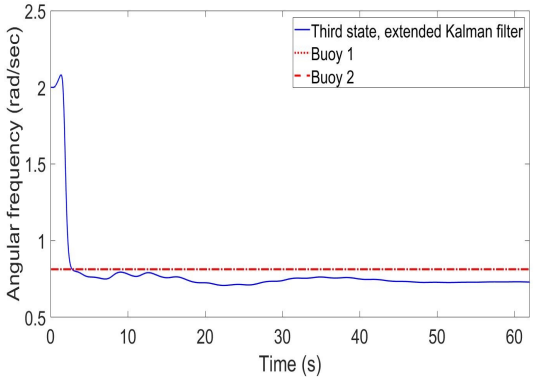
Figure 8-2 demonstrates how the methodology works thus far. The time series of pixel intensities is used as input to the SSA algorithm, and the principal component of the movement of water in the video is speculated to be included in the sum of



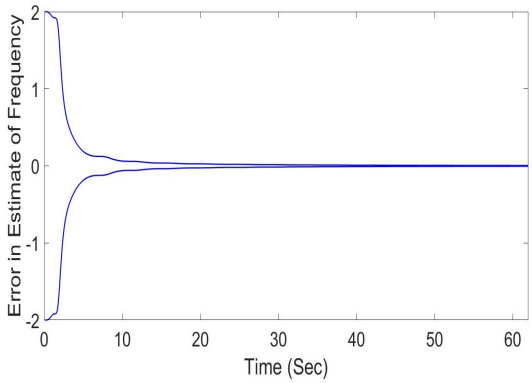
(a) SSA algorithm



(b) Extended Kalman filter



(c) Angular frequency estimation



(d) Theoretical errors in the estimation.

Figure 8-2 Example of running the methodology with shipborne video. (a) Time series of pixel intensities from video (solid blue line) and the sum of reconstructed components (RCs) 1-4 of the SSA algorithm (dashed red line) (b) The extended Kalman filter (first state estimate in dashed red line) attempts to establish one main frequency from the sum of RCs 1-4 (solid blue line). The extended Kalman filter gives in the third state the estimate of the unknown angular frequency. (c) Angular frequency estimation (third state) of the extended Kalman filter in regards to time compared to *in situ* buoy measurements. (d) Theoretical errors in the angular dominant frequency estimation in regards to time.

RCs 1-4 (fig. 8-4a). This central component of movement is isolated from all other video components with the result of the previous step and the extended Kalman filter (fig. 8-4b). The extended Kalman filter provides the estimate of the dominant angular frequency in each time step (fig. 8-2c), along with limits of certainty for that prediction from the covariance matrix (fig. 8-2d).

8.2.3 Significant wave height

The significant wave height can be found to be equal to four times the square root of the area under the spectral density [11]. Although from the methodology the spectral density is not available, the dominant frequency found in the previous steps can be given as input to the Pierson-Moskowitz spectrum (Sec. 2.2), and that shape can be used. From equations (2.2) and (2.3) the significant wave height h_s can be expressed in terms of dominant angular frequency ω_m as:

$$h_s = 2\left(\frac{0.87}{\omega_m}\right)^2 \sqrt{\frac{\alpha}{\beta}} \tag{8.16}$$

where $\alpha = 8.1 \times 10^{-3}$ and $\beta = 0.74$. The values of dominant frequency and significant wave height are the outputs of the methodology, which is validated in the following section with two sets of videos that have corresponding buoy measurements.

8.3 Experimental results

The two data sets of video in real environments with corresponding buoy sea state are described in the experimental data chapter sections 5.1 and 5.2. The first set of data comes from a shipborne camera (See Sect. 5.1) with two buoys measuring the ocean at the same time in a nearby location. The sea state in this set of videos is approximately the same, as the state is not expected to change in a large degree in the time span of a few hours. The ship videos are captured in a sea state of approximately 3.1 m-3.4 m. The second set of data comes from a tower (See Sect. 5.2) and a nearby buoy. This set of data allows the testing of the methodology for a variety of sea states.

Table 8.1 Error metrics with shipborne video (includes only videos with concurrent buoy sea state).

Error metric	Value
Mean Absolute Error (MAE)	0.15
Root Mean Square Error (RMSE)	0.19
Mean Absolute Percentage Error (MAPE)	4.83
Mean Percentage Error (MPE)	-0.05

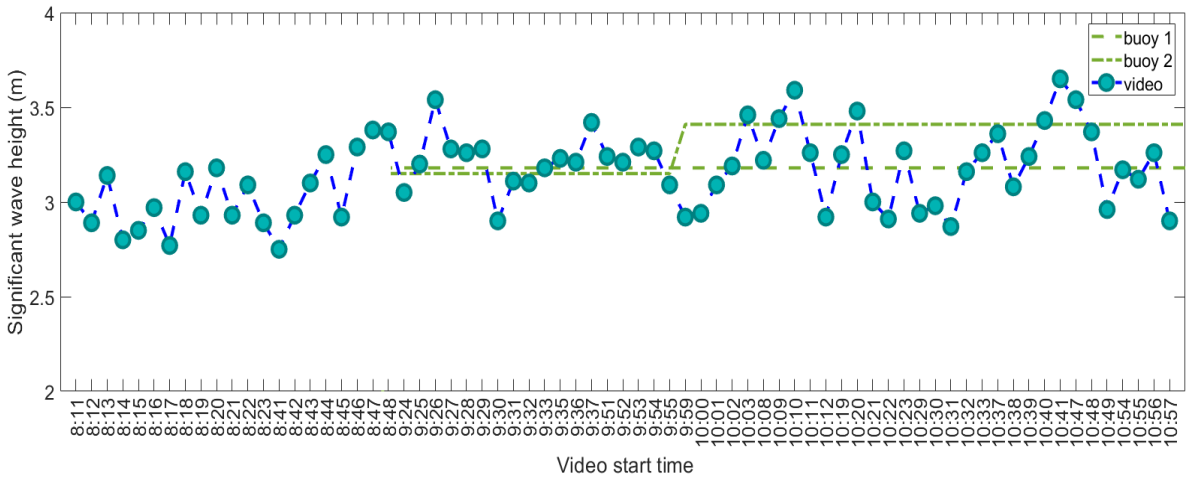


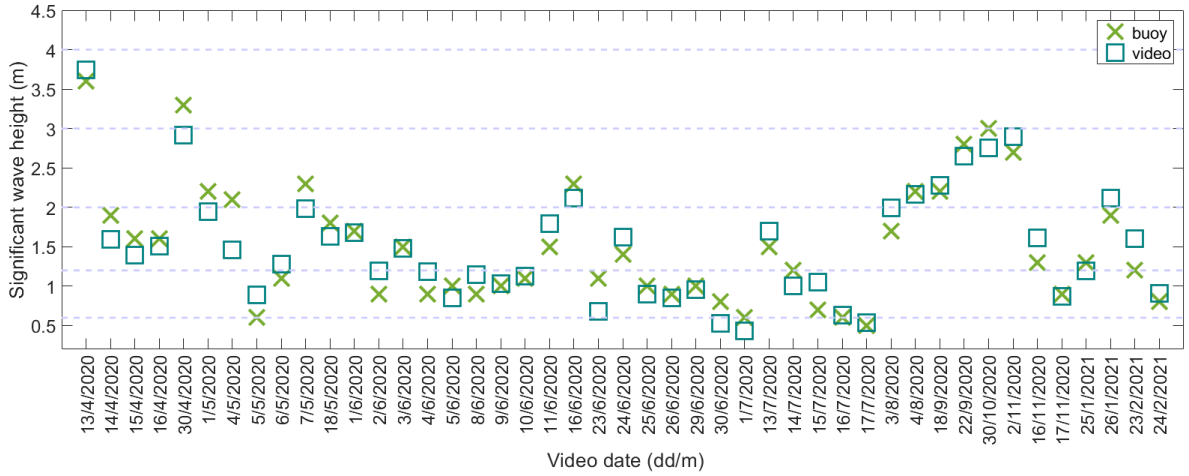
Figure 8-3 Shipborne video showing stability of video estimations of a similar sea state and correlation of the significant wave height (h_s) estimation between video and buoys. Buoy 1 $h_s = 3.18m$ (9:15-10:00 am), $h_s = 3.18m$ (10:00-11:00 am), Buoy 2: $h_s = 3.15m$ (9:15-10:00 am), $h_s = 3.41m$ (10:00-11:00 am).

The first set of data from the ship examine the behaviour of the method for an approximately stationary sea state. The second set of data from the tower examine the behaviour for a variety of sea states, as the videos were captured in different days. The SSA algortihm is run in all cases for a window length of 350 which is determined empirically. The matrix of w-correlations provides a good indication of whether the window is too small or too large, as in cases of smaller window the values are less concentrated to the main diagonal and the model is too general. With large windows, high values are concentrated in positions away from the main diagonal and thus the algorithm is overfitting. From empirical observations even if different window lengths are used (for example 250 or 450) the impact on the method estimations is minimal.

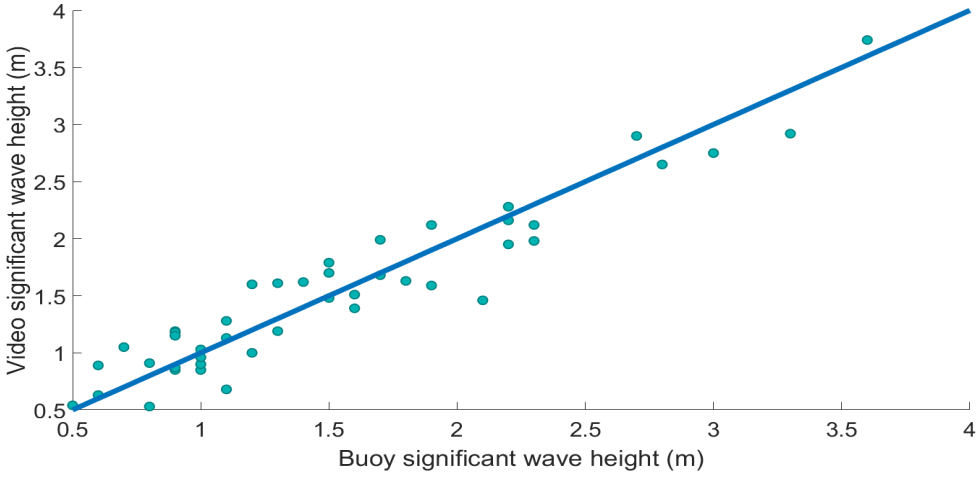
The shipborne results are presented in figure 8-3. Each point represents a one minute video captured in the same day. The buoys were deployed on the sea surface

Table 8.2 Error metrics with tower video.

Error metric	Value
Mean Absolute Error (MAE)	0.19
Root Mean Square Error (RMSE)	0.23
Mean Absolute Percentage Error (MAPE)	15.29
Mean Percentage Error (MPE)	-1.92



(a)



(b)

Figure 8-4 Tower video results showing correlation of the significant wave height (h_s) estimation between video and buoy across a variety of sea states. Buoy h_s: $min = 0.5m$, $max = 3.6m$. The dashed lines in (a) denote the significant wave height range of the sea states according to the Beaufort scale. The blue line in (b) is the diagonal.

after 9:15 a.m., any videos before then are presented here only to show the behaviour of the method. As mentioned, the sea state is not expected to change in a large degree in the span of one hour. The error metrics between the video estimations and buoy measurements are presented in table 8.1. The results of 0.19 m of RMSE and 4.83 % of MAPE look promising for the specific sea state. The small value of MPE indicate that the overestimations and underestimations offset each other.

Next, tower video are used for examining the behaviour of the technique across

a variety of sea states. The tower video results are presented in figure 8-4. Each point represents a video captured on a specific date. The technique seems to estimate lower dominant frequencies for higher sea states and higher dominant frequencies for lower sea states, as expected. This means that, as it was suspected, the dominant wave is concentrated in the first reconstructed components of the SSA algorithm. In higher sea states the peaks and the valleys are wider and the principal component from video has a lower frequency, whereas in lower sea states the peaks and valleys are more concentrated and the principal component has a higher frequency.

The error metrics from tower video are presented in table 8.2. Although the error metrics of 0.23 m RMSE and 15.29 % of MAPE are higher than the ones observed from shipborne video, they still remain in acceptable levels in showing that the estimations are meaningful. Similarly with the shipborne video, the small value of MPE indicate that the overestimations and underestimations offset each other. The methodology seems to provide a balanced set of results in terms of overestimation and underestimation for both sets of input video data. The significant wave height examined from the tower video is between 0.5 m and 3.6 m. Higher sea states were not made available and the efficiency of the method for higher sea states is not known.

As a final note, an example is given for demonstrating a sensitivity analysis of the SSA algorithm. As mentioned, this window size of 350 is empirically derived based on observations on the matrix of w-correlations. In the remainder of this section, two sample videos are used as input for a small sensitivity analysis.

The purpose is to investigate how sensitive the results are in the selection of the window size parameter. Figure 8-5 presents the significant wave height estimation with shipborne video in the input. The buoy measurements indicate a significant wave height of 3.1 m from the first buoy and 3.4 m from the second buoy. Here, the window size parameter of the SSA algorithm varies between 50 and 1000 with a step of 50 frames. The observation here is that the estimation results are not highly sensitive to the window size parameter. Although a window size of 350 is selected empirically based on the observation from the matrix of w-correlations, slightly lower or higher window size provides very similar results. Even if the window size is smaller

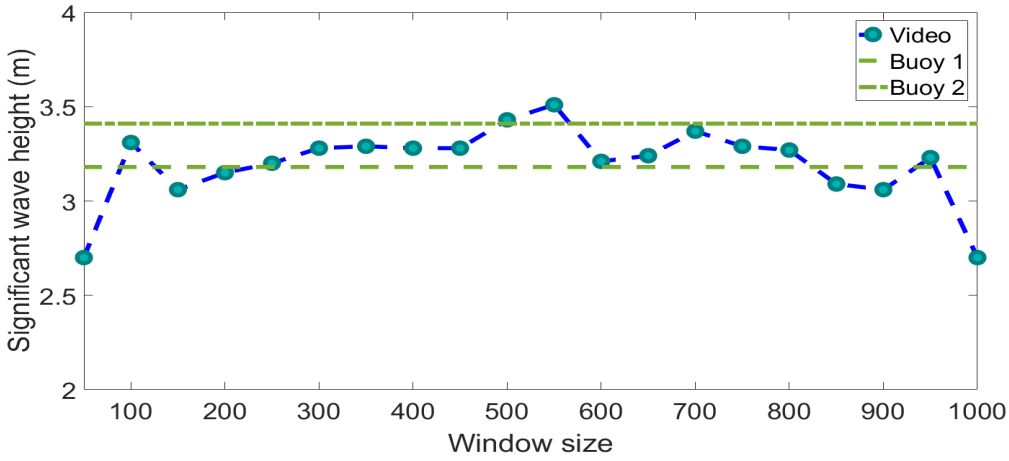


Figure 8-5 Sensitivity analysis of the methodology with shipborne video as input. The significant wave height estimation with a varying window size of the SSA algorithm. Significant wave height h_s Buoy 1: 3.18 m, Buoy 2: 3.41 m.

(e.g. 100) or larger (e.g. 900) the estimations are still accurate. For very small (e.g. 50) or very large (e.g. 1000) window size the algorithm provides results that are less accurate.

It would be interesting to see of this behaviour is similar with a different video (tower video) at a lower sea state. Figure 8-6 presents the sensitivity analysis for tower video as input. The buoy measurements indicate a significant wave height of 0.9 m. Similarly with the previous video, the change in window size does not greatly affect the significant wave height estimation. Only a very low (50, 100) window size presents some difference. From figures 8-5 and 8-6, the significant wave height estimations do not seem to be sensitive in a large degree to the window size parameter of the SSA algorithm. The speculation for the less accurate results with a small window size is that the model is too general, that is the association between the different elementary reconstructed components (RCs) is not clearly defined. Similarly, the speculation for the less accurate results with a very high window size is that the model is too specific, that is the association between the different RCs is overspecified.

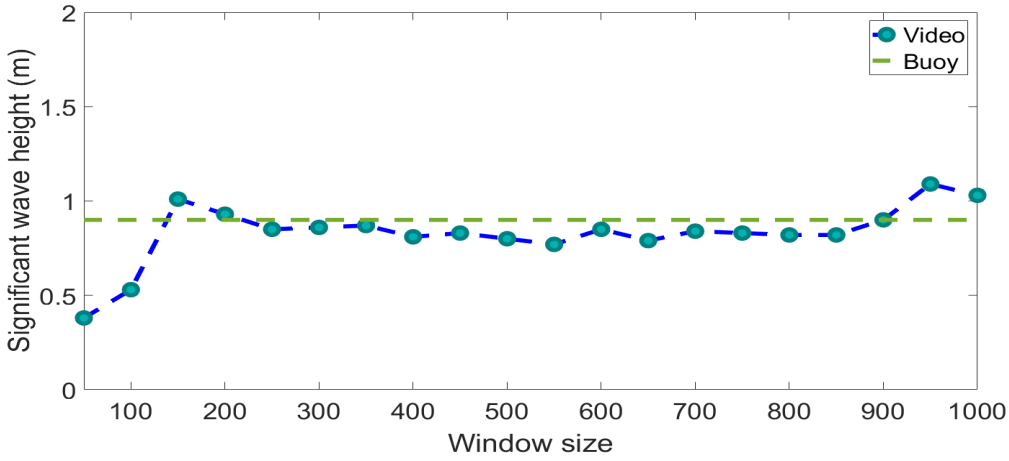


Figure 8-6 Sensitivity analysis of the methodology with tower video as input. The significant wave height estimation with a varying window size of the SSA algorithm. Significant wave height h_s Buoy: 0.9 m.

8.4 Conclusion

The results are promising, indicating that it is possible to get the sea state from video of the ocean in real environments. However, the results are preliminary and further experiments are required. Higher sea states are not examined ($h_s > 3.6m$) and thus the applicability in such instances is not known.

Compared to chapter 6, the methodology of this chapter has the advantage of not requiring high-scale videos in the input. Compared to chapter 7, the methodology of this chapter has the advantage of not relying on the shape of the video amplitude spectrum for calibration. A drawback of the present methodology however is the dependence on the shape of an empirical ocean energy spectrum for getting the significant wave height from the dominant frequency.

The following chapter examines if it is possible to overcome the drawback of dependence on the shape of an empirical energy spectrum for estimating the significant wave height. In this case, similarly to chapter 7, some means is necessary for calibration. This is again solved with ocean theory, which does not in the case of the following chapter include a specific empirical ocean energy spectrum.

8.5 Appendix

The following text provides a detailed presentation of the solution of the extended Kalman filter with the environment definition described in the methodology (Sect. 8.2.2). Having the system dynamics matrix, the function of the filter is structured in a way that uses the state variables for the propagation and not the error states. In this case, the fundamental matrix is not used for propagating the states, rather only for the calculation of the gains.

If the systems dynamics matrix is constant between successive iterations, the fundamental matrix can be estimated with the Taylor series approximation [9]. For simplicity, the first two terms are used for approximating the fundamental matrix, which is found to be equal to:

$$\mathbf{\Phi}(t) = \mathbf{I} + \mathbf{F}t = \begin{pmatrix} 1 & t & 0 \\ -\hat{\omega}^2 t & 1 & -2\hat{\omega}\hat{x}t \\ 0 & 0 & 1 \end{pmatrix} \tag{8.17}$$

The discrete fundamental matrix $\mathbf{\Phi}_k$ is defined by substituting t with the time between two successive frames $\Delta\tau$:

$$\mathbf{\Phi}_k = \begin{pmatrix} 1 & \Delta\tau & 0 \\ -\hat{\omega}_{k-1}^2\Delta\tau & 1 & -2\hat{\omega}_{k-1}\hat{x}_{k-1}\Delta\tau \\ 0 & 0 & 1 \end{pmatrix} \tag{8.18}$$

The observed state z_k is linearly correlated to the true state x_k, and z_k is equal to:

$$z_k = \mathbf{H}\begin{pmatrix} x \\ \dot{x} \\ \omega \end{pmatrix} + \epsilon_k \tag{8.19}$$

where $\mathbf{H}$ is the observation model, ϵ is the observation noise, which is assumed to be drawn from a zero mean multivariate normal distribution with covariance $\mathbf{R}_k$, where $\mathbf{R}_k$ is equal to:

$$\mathbf{R}_k = \sigma_x^2 \tag{8.20}$$

where σ_x^2 is the variance of the observation noise, and the observation model $\mathbf{H}$ is equal to:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \tag{8.21}$$

As mentioned, the fundamental matrix as found with the first two states of the Taylor series approximation is not used for propagating the states. The projected states of the true signal $\bar{x}$, the derivative of the true signal $\bar{\dot{x}}$ and angular frequency $\bar{\omega}$ are found by integration of the differential equations over the sampling interval. The extended Kalman filtering equations are equal to:

$$\hat{x}_k = \bar{x}_k + K_{1k} Res_k \tag{8.22}$$

$$\hat{\dot{x}}_k = \bar{\dot{x}}_k + K_{2k} Res_k \tag{8.23}$$

$$\hat{\omega}_k = \hat{\omega}_{k-1} + K_{3k} Res_k \tag{8.24}$$

where Res_k is the residual and is equal to $Res_k = z_k - \bar{x}_k$ and K_{1k}, K_{2k}, K_{3k} are the Kalman gains found through the recursive Riccati equations [9].

Chapter 9

Sea state from ocean video with equilibrium range theory

The methodology of the previous chapter works on identifying the principal component of movement from ocean video, which is conjectured to be associated with the dominant wave of the measured ocean. It then uses the Pierson-Moskowitz spectrum and ocean theory of the ocean spectral moments to convert the dominant frequency into a value of the significant wave height. In the final step of converting dominant frequency into wave height, the value of the wave height depends on the shape of the empirical spectrum used. However, not all oceans share the same shape of energy spectrum for the same sea state. The shape of the Pierson-Moskowitz spectrum for a particular sea state was derived by averaging a set of energy spectra of the same sea state [114].

In this chapter, a methodology is presented that aims to accomplish the same goal, namely to estimate the significant wave height, without the use of such an empirical spectrum. It would be expected that, since the estimations in this case do not depend on the shape of a particular empirical spectrum, the model would be more localised to the measured ocean and the estimations would be more accurate. Since *in situ* devices are still to be used solely for testing the results, ocean theory other than of an empirical energy spectrum is to be used for calibration.

In the present work, theory involving the equilibrium range of oceans is used for

calibration. The equilibrium range equation gives an energy value for waves with frequency higher than the frequency of the dominant wave. This is the first time in this work that a methodology uses information not involving the dominant wave. As a reminder, in chapter 6 the waves with the highest values from the phase modulation matrix are used with the spatial variation of the wave field. In chapter 7, the process for calibration of the video amplitude spectrum involves the use of the wave with the peak amplitude. In the previous chapter, the SSA algorithm and extended Kalman filter are used for estimating the dominant frequency.

The intuition behind the utilisation the dominant wave is that, in the case of very noisy data, such as in this case (video from a monoscopic camera in real environments), waves with higher frequencies are more likely to be affected by noise. Specifically, for higher frequencies it is more likely that the amplitude of the noise is higher than the amplitude of the true signal. In this context, the true signal is the element from video that is associated with the ocean's movement. Contrarily, it is expected that waves with frequency close to the dominant frequency are less affected by noise due to their higher amplitudes.

When choosing to work with waves with frequency higher than the dominant frequency, some challenges arise. Apart from the presence of noise, if the requirement of a frequency higher is given, such as in this case, there is a difficulty in selecting such a frequency without having information about the dominant wave. If the frequency of the dominant wave is not known, the problem arises in identifying a frequency higher than that. From these challenges, the use of waves other than the dominant wave for getting the sea state might not at first seem like a good idea.

This problem could be solved if the methodology of the previous chapter that estimates the dominant frequency is incorporated. Then the question arises of why would any further work be involved, since the sea state is already found. The only new thing the equilibrium range theory would bring in such a case would be to get the estimation of the sigificant wave height without the empirical spectrum shape. This is not the case in the methodology of this chapter, which is to be used as a standalone process. Specifically, the methodology of this chapter does not depend on

the use of any methodology from previous chapters. Instead, it focuses on identifying a frequency that is higher than the dominant frequency that is not too high as to contain mostly noise.

Since the sea state is not known, and high frequencies have higher noise (chapter 6 discusses more about this subject), the technique aims to identify a frequency near the unknown dominant frequency (less noise), but not the dominant frequency (otherwise selected ocean theory is not applicable). In this chapter, like in the previous chapter, the SSA algorithm is utilised. This particular algorithm is found to be promising in the previous chapter in concentrating the elements from video temporal variation that are associated with the dominant wave in the first PCs.

The SSA algorithm is used again here, but this time in order to identify a wave with the frequency of the aforementioned requirements. To do this, a different selection of PCs is done. Specifically, the sum of RCs 5-8 is isolated from time series of pixel intensities from video. This sum is derived empirically and is conjectured to include a frequency that is different from the dominant frequency. The hypothesis is, since the SSA algorithm concentrates the trend and very low frequency elements in the first PC [63], and in the previous chapter the sum of RCs 1-4 was found to be associated with the dominant frequency, it would be expected that the frequencies present in the RCs 5-8 are different than the dominant frequency.

The next challenge involves identifying one frequency from this selected summation. A Fourier transform of this summation could practically include multiple peaks of similar magnitudes. This is where the extended Kalman filter is involved, which with the environment definition described in the following text aims to distinguish one main frequency out of this summation. Once this is done, this frequency is given to the equilibrium range equation to get an energy value. This energy value can be converted into amplitude in metres with ocean theory and this amplitude in metres is used for calibration.

In the following text, first an introduction is given to the equilibrium range theory that is used for calibration (Sect. 9.1). Then, the methodology is described (Sect. 9.2) and tested in the experimental results (Sect. 9.3).

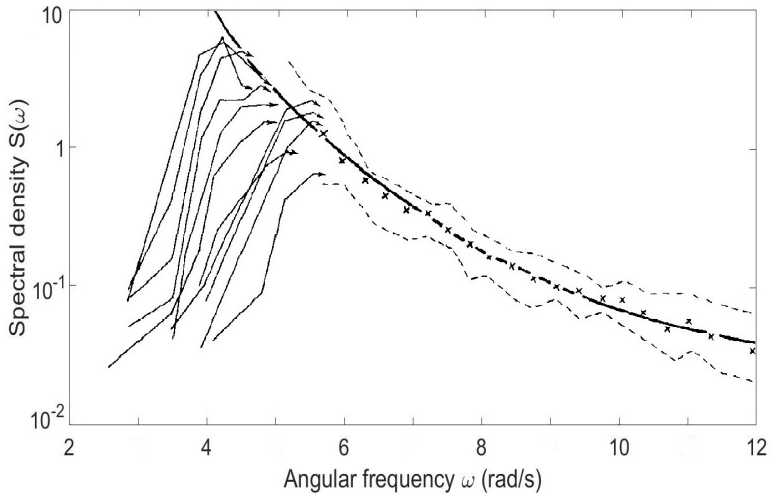


Figure 9-1 Energy spectra for different wind speeds from Phillip's 1958 paper [124] demonstrating the concept of the equilibrium range, where for higher frequencies the energy values converge for different wind speeds. Reproduced with permission from [124].

9.1 Equilibrium range theory

In 1958, Phillips introduced an equation for higher frequencies of the ocean energy spectrum in fully developed seas [124]. Based on observations made from Burling's wind generated energy spectra from a wave tank [28], it was observed that for higher frequencies of the spectrum the energy values are independent of wind speed and fetch. This range of frequencies is referred to as the equilibrium range (fig. 9-1). This is not the case for lower frequencies, as with the decrease of frequency, the energy values are observed to depend more on the wind speed and other conditions.

From [124], the energy for higher frequencies of the spectrum, that is for $\omega > \omega_m$, where ω_m is the dominant angular frequency, is expressed as:

$$S(\omega) = \beta g^2 \omega^{-5} \tag{9.1}$$

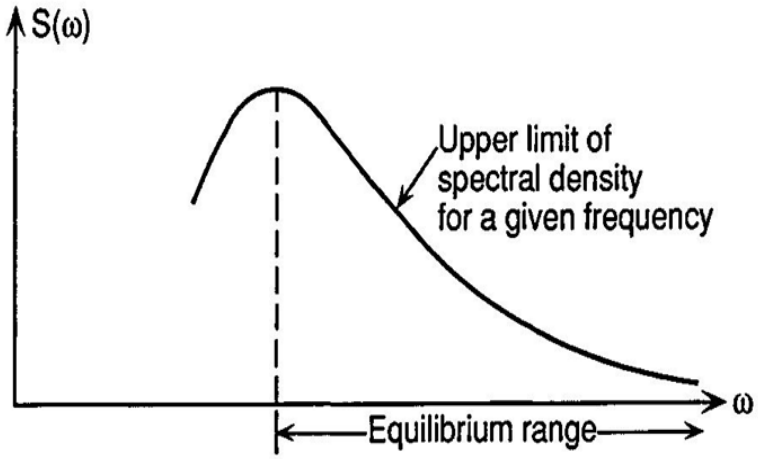
where β is an absolute constant, set to be equal to $\beta = 7.4 \times 10^{-3}$ in order to match Burling's observations and g is the gravitational constant.

It is important to note that there is no literature indicating a superiority of the Phillip's equilibrium range theory to the Pierson-Moskowitz spectrum theory (See

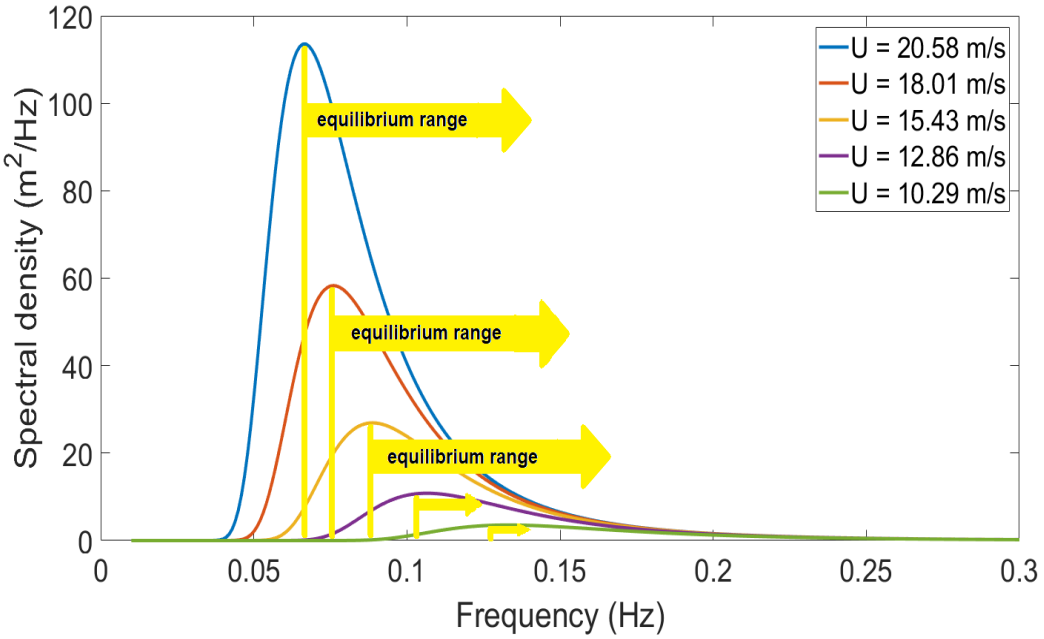
ocean theory chapter 2, Sect. 2.2.1). The latter is utilised in the methodologies of chapters 6, 7 and 8. The utilisation of the equilibrium range is just a different approach. There is also no literature indicating a superiority of the Phillips energy spectrum [106] compared to the Pierson-Moskowitz energy spectrum [127]. They both utilise similar assumptions, such as fully developed seas. The difference between the equilibrium range theory utilised in this chapter to the Pierson-Moskowitz spectrum is that the input in the equilibrium range equation is a frequency higher than the dominant frequency. In the Pierson-Moskowitz spectrum equation, the input is the dominant frequency.

According to the equilibrium range theory, the tail of the energy spectra in very high frequencies is approximately identical for all sea states (See Fig. 9-1). The equilibrium range equation gives an energy value in the output for a given frequency in the input that belongs in the equilibrium range. The energy value from the output of the equilibrium range can be converted to an amplitude value in metres with ocean theory (see ocean theory chapter 2, Sect. 2.3). Thus the equilibrium range equation enables us to acquire a measure in metres (real world units) when information of frequency from video is given in the input. In this sense, the equilibrium range equation enables us to translate the information from video to real world units. In other words, we are able to calibrate the video information so that it is represented in units of the real world and not pixel intensity units from video.

The methodology of this work makes efforts to identify a frequency belonging in the equilibrium range, so that it can be used as input to the equilibrium range equation. Although the tail is identical for very high frequencies, the equilibrium range differs for each sea state. Theoretically, the equilibrium range is the range of frequencies that are higher than the dominant frequency [119] (See Fig. 9-2a). Because the dominant frequency is different for different sea states, the equilibrium range 'begins' in lower frequencies for higher sea states and in higher frequencies for lower sea states (See Fig. 9-2b). In that sense, a given frequency can belong in the equilibrium range if the sea state is high, however, the same frequency may not belong in the equilibrium range if the sea state is lower. Although the equilibrium



(a) The equilibrium range is the range of frequencies that are higher than the dominant frequency of a particular ocean. Reproduced with permission from [119].



(b) The energy spectrum for different sea states. The yellow arrows indicate the equilibrium range for each sea state. Here it is illustrated that the equilibrium range theoretically starts at lower frequencies for higher sea states, and at higher frequencies for lower sea states.

Figure 9-2 Illustration of the equilibrium range.

range is the same for very high frequencies, these frequencies are too high to be used directly from video, as their amplitude and energy values are very low and can be easily affected by the high noise of the video. This 'noise' or sources of error from video is discussed in the experimental data chapter 5, Sect. 5.4.

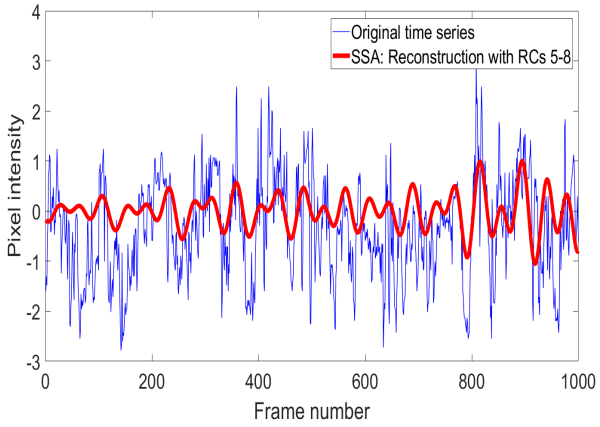
The goal of this methodology is thus to select a frequency that belongs in the equilibrium range for the specific video in the input, without prior information of the sea state or ocean dominant frequency. Because the equilibrium tail begins in different frequencies depending on the sea state, and because we want to avoid using very high frequencies (more affected by noise), the SSA algorithm is utilised in order to select a secondary oscillatory component from video. We hypothesise that this oscillatory component has a frequency higher than the dominant frequency, and not too high as to be largely affected by noise.

Once we have identified a frequency that we conjecture belongs in the equilibrium range for each specific video, we can use this frequency in order to get its corresponding amplitude in metres with the equilibrium range equation. We can also calculate its amplitude in pixel intensity from video. Because the amplitude scale is defined here as the amplitude in metres divided by the amplitude in pixel intensity (described in the following section), we are able to use the equilibrium range theory in order to get an estimate of the amplitude scale. This amplitude scale enables us to transfer our video information to real world units of metres, and thus calculate the significant wave height. In the next section, this methodology is fully described.

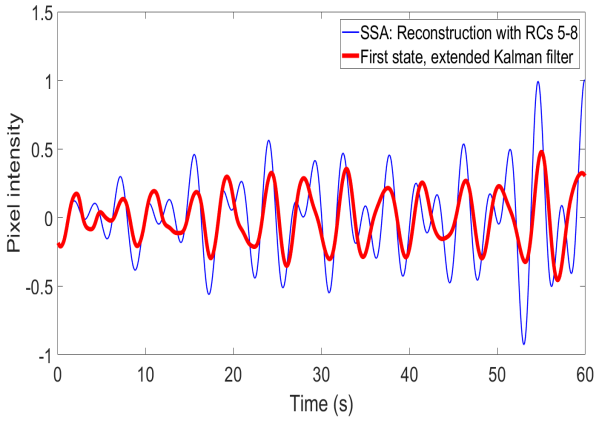
9.2 Methodology

Time series of pixel intensities from the ocean video is used as input. First, the SSA algorithm is applied on the data and a selection is made of a sum of RCs that represent a secondary oscillatory component (Sect. 9.2.1). Then, the extended Kalman filter is used in order to identify the main frequency from this selection (Sect. 9.2.2). An amplitude scaling process is developed that uses the linear Kalman filter, the least squares approximate solution, the equilibrium range equation and theory of the first

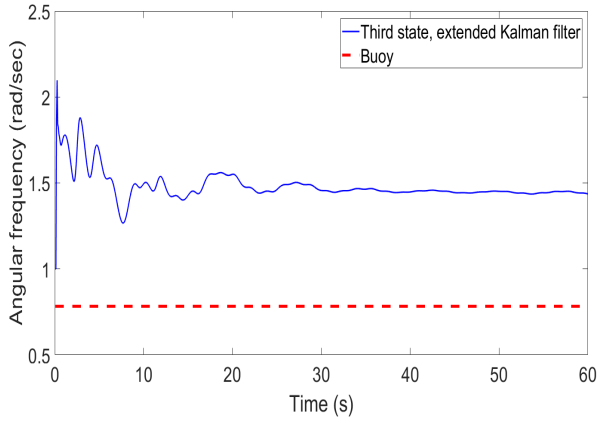
law of thermodynamics and the linearised velocity in oceans (Sect. 9.2.3). The sum of a selection of RCs from the SSA algorithm is multiplied with this factor in order to give a measure of the significant wave height (Sect. 9.2.4).



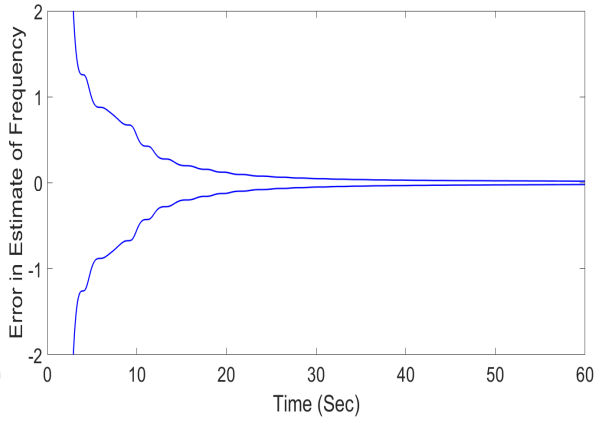
(a) SSA algorithm



(b) Extended Kalman filter



(c) Angular frequency estimation



(d) Theoretical errors in estimation

Figure 9-3 Example of running the methodology with shipborne video. The aim is to identify a frequency that is higher than the ocean dominant frequency. (a) Time series of pixel intensities from video (blue line) and the sum of reconstructed components (RCs) 5-8 of the SSA algorithm (red line) (b) The extended Kalman filter (first state estimate in red line) attempts to establish one frequency from the sum of RCs 5-8 (blue line). The extended Kalman filter gives in the third state the estimate of the unknown angular frequency. (c) Angular frequency estimation (third state) of the extended Kalman filter in regards to time compared to the ocean dominant angular frequency from *in situ* buoy measurements. (d) Theoretical errors in the angular frequency estimation in regards to time.

9.2.1 Singular spectrum analysis (SSA) algorithm

General information about the SSA algorithm is provided in Sect. 8.1 of the previous chapter. The SSA algorithm is first performed on the pixel time series. In this work the equilibrium range is used, which is valid for frequencies in the ocean energy spectrum that are higher than the dominant frequency. The SSA algorithm, in ideal conditions, given a time series of this nature would concentrate in the first principal components the oscillatory components of higher importance or principality (and the trend if available), and will sort the principal components in order of importance in a way that after a certain point, the rest can be considered noise.

Empirically, with the window selection used here, which is 350, upon testing with real video and corresponding buoy measurements, the dominant frequency was found to be residing in the first principal components. Testing with different window lengths further showed that the dominant frequency is concentrated in such a manner. In this work, since the equilibrium range theory is to be used, a frequency higher than the dominant frequency is to be selected. To do this, the RCs 5-8 were selected, empirically derived. The following text describes the logic behind this selection of RCs.

The matrix of w-correlations provides a good guide on which window length and RCs to be selected for grouping. If the window is too small and the model too general, components are less concentrated in the main diagonal. If the window is too large and the model too specific, the main diagonal has high values, however, high values additionally appear in random positions (the model is too specific). The main idea here is to isolate a secondary oscillatory component from the pixel intensity time series, which is different from the principal component which is hypothesised to be correlated to the ocean dominant frequency (RCs 1-4) and is not noise (RCs 8-max). Figure 9-4 gives an example of a matrix of w-correlations from ocean video, where it is relatively easy to group different PCs based on their intercorrelation and separate the main oscillatory components from noise. Figure 9-3a gives an example of how the SSA algorithm is used in order to track a secondary oscillatory component from the

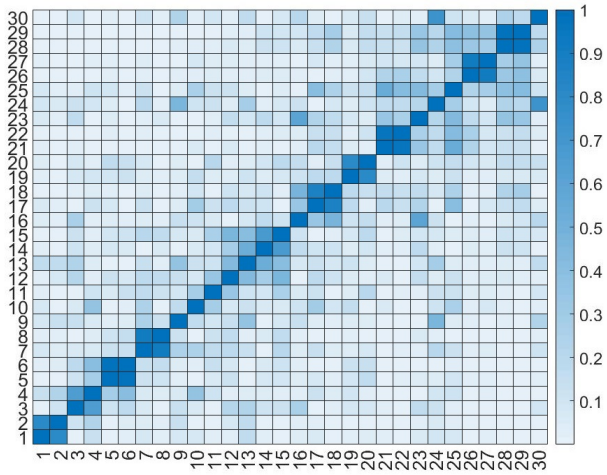


Figure 9-4 Example of matrix of w-correlations of SSA algorithm from ocean video. From this example, a correlation of certain PCs is visible, such as PCs 1-2, 3-4, 5-6, 7-8. Additionally, some higher PCs can be considered to contain more noise (PCs 11-max).

time series of pixel intensities.

9.2.2 Extended Kalman filter

From the SSA algorithm, the sum of RCs 5-8 are selected and it is conjectured that this sum contains a secondary oscillatory component of the ocean video. A Fourier transform of this time series will contain multiple peaks that in some cases are of similar magnitude. From the sum of RCs 5-8, the frequency of the main secondary oscillatory component is identified with the use of the extended Kalman filter.

The sinusoidal signal with angular frequency ω to be found from the sum of RCs 5-8 is:

$$x = a\sin(\omega t + \phi) \tag{9.2}$$

where a is the amplitude and ϕ the phase. The solution to the extended Kalman filter is the same as in chapter 8, Sect. 8.2.2.

From the extended Kalman filter, an estimation is given of the angular frequency ω of the signal given in the input. The question here arises, why go through all this trouble to get a frequency higher than the dominant frequency, and not just randomly pick a high frequency? The answer is, the dominant frequency is unknown, and so

a very high frequency might need to be selected in such a case. With the nature of this data, as mentioned in chapter 6 (see Sect. 6.1.3) and the relevant research paper [102], higher frequencies are more affected by noise, which makes any estimations more inaccurate and thus should be avoided.

9.2.3 Amplitude scale

With the use of what is hypothesised to be a secondary oscillatory component of the ocean video found in the previous steps, the frequency can be used as input to the equilibrium range equation (see Sect. 9.1) in order to provide an energy value of this wave in the ocean energy spectrum. The key here is that this energy value can be translated into a value in metres with the use of equation (2.12) (see Sect. 2.3).

Until this point, the frequency of the secondary component and its amplitude in metres are available. This information is used next in order to solve the amplitude scale problem. The amplitude scale enables the transition of the information of pixel intensity to real world units, such as metres. This is not to imply that the pixel intensity is the wave elevation in an uncalibrated form, which is not true. However, there is a correlation between the temporal variation of pixel intensity and the general temporal variation of the true ocean. This is the reason it is easy to distinguish a lower sea state (typical summer day) from a higher sea state (during a storm with high wind) simply by watching a video. The troughs and crests of waves are wider in higher sea states.

The amplitude scale is defined as:

$$\alpha = \frac{a_{er}}{a_u} \tag{9.3}$$

where a_{er} is the amplitude in metres found in the previous steps with the equilibrium range equation and a_u the amplitude in pixels which until this point is unknown. What is left until this point in order to calculate the amplitude scale is the amplitude in pixels for the specific frequency found in section 9.2.2. In order to achieve this, a methodology is followed described in chapter 7 and the associated research paper,

which uses the linear Kalman filter and the least squares approximate solution. This process can be found in chapter 7 in sections 7.1.1 and 7.1.2.

Although the process might seem complicated, the end goal here is worth the trouble, as once this amplitude scale is found, the information from video acquires a meaning in real world units and the video can now be used in order to get sea state approximations without using the shape of a particular empirical energy spectrum of the ocean, such as the Pierson-Moskowitz spectrum.

9.2.4 Significant wave height

With the original time series as input, the SSA algorithm is used again, this time in order to isolate the oscillatory components of the temporal variation from the noise. Empirically, from the matrix of w-correlations the first eight principal components were identified as main oscillatory components. In the matrix, principal components denoting noise have values spread in various locations, whereas the main components have high values in the main diagonal, as in figure 9-4.

The amplitude scale α (Sect. 9.2.3) is multiplied to the sum of RCs 1-8 in order to get a time series in metres that resembles wave elevation. Although it does not represent the ocean's true elevation, as this is usually found with *in situ* devices such as buoys, it does give a representation of the video temporal variation in real units. If typical ocean wave elevation from a buoy is used, the significant wave height can be found as four times the standard deviation of the wave elevation [73], which is based on the fact that the significant wave height can be found from the zeroth moment or area of the ocean energy spectrum.

From video, an approximation of the significant wave height is found in a similar manner, as four times the standard deviation of the scaled to metres sum of RCs 1-8. Since temporal variation associated with pixel intensity is not the same as ocean wave elevation, the value found in this manner is not exactly the same as the value of the significant wave height of the true ocean. The goal of this work is to approximate as much as possible the true value with the data that is available here. The experimental results in section 9.3 present the accuracy of the video estimations against the values

of the real ocean.

9.2.5 Summary

Time series of pixel intensity is given as input to the SSA algorithm, and the sum of RCs 5-8 is used as input to the extended Kalman filter in order to provide the frequency of a secondary oscillatory component. Its energy is found with the use of the equilibrium range equation and its amplitude in metres with theory based on ocean theory. For the same oscillatory component, its amplitude is found in pixel intensity metric with the linear Kalman filter and the least squares approximate solution methodology described in chapter 7. The amplitude scale is found as the amplitude in metres divided by the amplitude in pixels and is multiplied with the sum of RCs 1-8 of the SSA algorithm. The significant wave height is found as four times the standard deviation of the scaled to metres sum of RCs 1-8.

9.3 Experimental results

The methodology is tested against two sets of video-buoy data. Details about the experimental data can be found in the experimental data introductory chapter sections 5.1 and 5.2. The first set is from a shipborne camera capturing the ocean from a ship. The second set comes from a camera placed on a tower. In both cases buoy data are available for validation. From the videos after preprocessing, time series of pixel intensities are given in the input. The process is performed for a set of pixel locations individually according to the selection described in sections 5.1 and 5.2 and the final significant wave height is found as the average of the estimation from each time series. This is done for providing more accurate estimations. The video data have a duration of approximately one minute. The frame rate of the shipborne video is 15 frames per second and of the tower video 30 frames per second.

The shipborne experimental results are presented in figure 9-5. It should be noted that the buoys were deployed after 9:15 am. The significant wave height for the buoys vary between 3.1 m and 3.4 m for the duration of the video presented here.

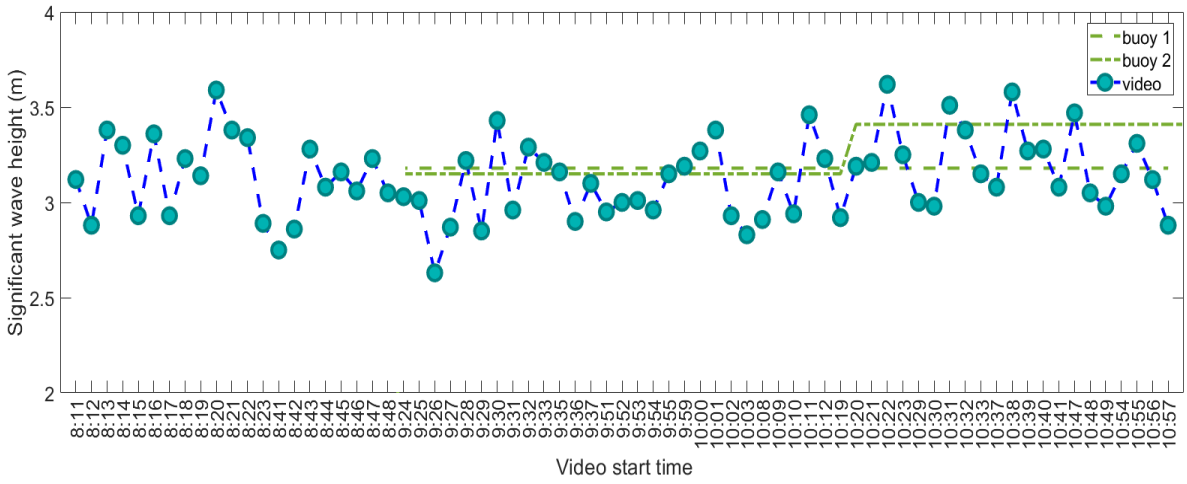


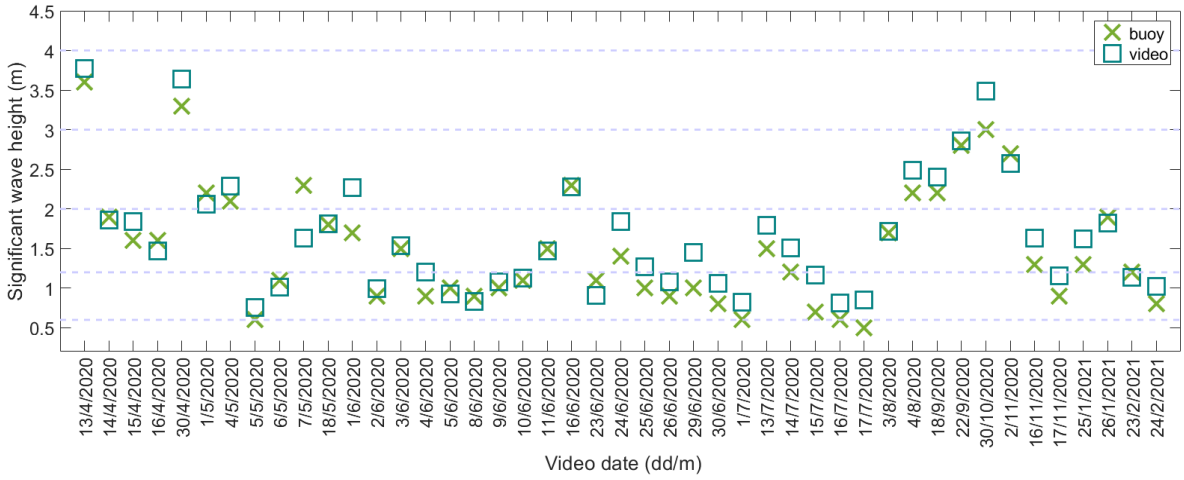
Figure 9-5 Shipborne video showing correlation of the significant wave height (h_s) estimation between video and buoys. Buoy 1 $h_s = 3.18m$ (9:15-10:00 am), $h_s = 3.18m$ (10:00-11:00 am), Buoy 2: $h_s = 3.15m$ (9:15-10:00 am), $h_s = 3.41m$ (10:00-11:00 am).

The video estimations flunctuate in value around the true values. A better picture of the accuracy of the technique for the shipborne video is presented in table 9.1. The RMSE of 0.22 m for a 3.1 m sea state shows that the difference between the video estimations and the buoy values (which are considered the true ocean sea state) is not very large. The 5.65 % of MAPE additionally verifies that the methodology is doing something meaningful. The small MPE value indicates that the overestimations and underestimations offset each other. However, one particular sea state has been investigated thus far.

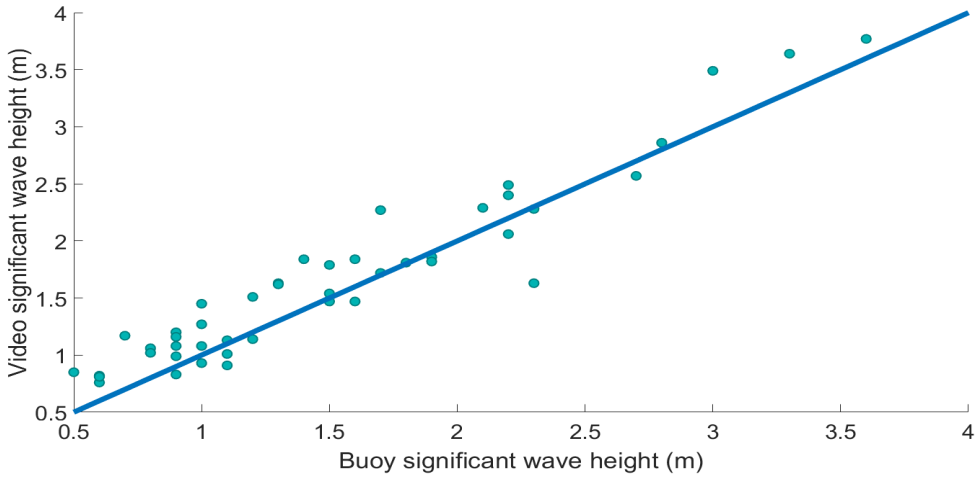
The tower video was captured across various days. This enables the testing of the accuracy of the video estimations across a variety of sea states. The tower video results are presented in figure 9-6. The significant wave height varies between 0.5 m and 3.6 m. The error metrics are presented in table 9.2. Here an increase can be

Table 9.1 Error metrics with shipborne video (includes only videos with concurrent buoy sea state).

Error metric	Value
Mean Absolute Error (MAE)	0.18
Root Mean Square Error (RMSE)	0.22
Mean Absolute Percentage Error (MAPE)	5.65
Mean Percentage Error (MPE)	2.19



(a)



(b)

Figure 9-6 Tower video results showing correlation of the significant wave height (h_s) estimation between video and buoy across a variety of sea states. Buoy h_s: $min = 0.5m$, $max = 3.6m$. The dashed lines in (a) denote the significant wave height range of the sea states according to the Beaufort scale. The blue line in (b) is the diagonal.

observed in both RMSE at 0.26 m and MAPE at 18.16 % compared to the ship data. However, from the figure and table, the results show that the methodology is doing something meaningful even when the sea state varies. The negative value of MPE indicates that the video estimations are overestimating the buoy measurements.

In comparison to the previous chapter (chapter 8) that uses the same data, the error metrics here are larger. However, in the previous chapter an empirical spectrum was used for facilitating the wave height scale estimation. Here, the scale estimation does not use such information, which is advantageous since the shape of the energy spectrum for a measured sea is not always the same as the shape of the empirical spectrum. Additionally, this methodology uses a secondary sinusoidal element for the scaling instead of approximately the peak as in the previous chapter. Since the peak has a higher wave height, it is less prone to errors caused by noise and thus the estimations are expected to be more accurate. This methodology however, as mentioned, has the advantage of not using an empirical spectrum. The theory of the equilibrium range is not restricted to a particular empirical energy spectrum of the ocean.

9.4 Conclusion

A methodology is proposed for acquiring sea state information from the temporal variation of ocean video in real environments. Ocean theory is used for calibration instead of *in situ* measurements. Testing with videos from two data sets in different oceans and sea states validate the applicability of the methodology. The results are limited, as the significant wave height acquired from buoys of the two data sets

Table 9.2 Error metrics with tower video.

Error metric	Value
Mean Absolute Error (MAE)	0.21
Root Mean Square Error (RMSE)	0.26
Mean Absolute Percentage Error (MAPE)	18.16
Mean Percentage Error (MPE)	-13.59

vary between 0.5 m and 3.6 m. Future work could examine the applicability of the methodology in higher sea states. Future work could also focus on improving the calibration process.

The methodology has the advantage that it does not depend on the specific shape of an empirical ocean energy spectrum for calibration. The results suggest that for approximately the same sea state the results are consistent from shipborne data and that the methodology is adaptive to the change in the sea state from the tower video. Compared to chapter 6, this methodology has the advantage of not being limited to large scale videos and is tested against video of real ocean and not radar images. Compared to chapter 7, this methodology has the advantage of not using the position of the peak from the video amplitude spectrum for calibration and not using an empirical energy spectrum for calibration.

Since the present methodology functions as a standalone process, the dominant frequency is not estimated directly. Selecting the sum of RCs 5-8 of the SSA algorithm is speculated to provide a frequency higher than the dominant frequency. Although the experimental results validate that the methodology is doing something meaningful, it is not certain whether the sum of RCs 5-8 will always provide a frequency higher than the dominant frequency. Future work could use additional input video for further testing and improving the technique.

Compared to chapter 8, testing the technique of this chapter provides higher error metrics of RMSE and MAPE. An explanation to this would be that the technique in chapter 8 focuses on finding the dominant wave, which is easier to be identified due to its higher energy, whereas here by using a secondary oscillatory component, noise could more easily affect the estimations.

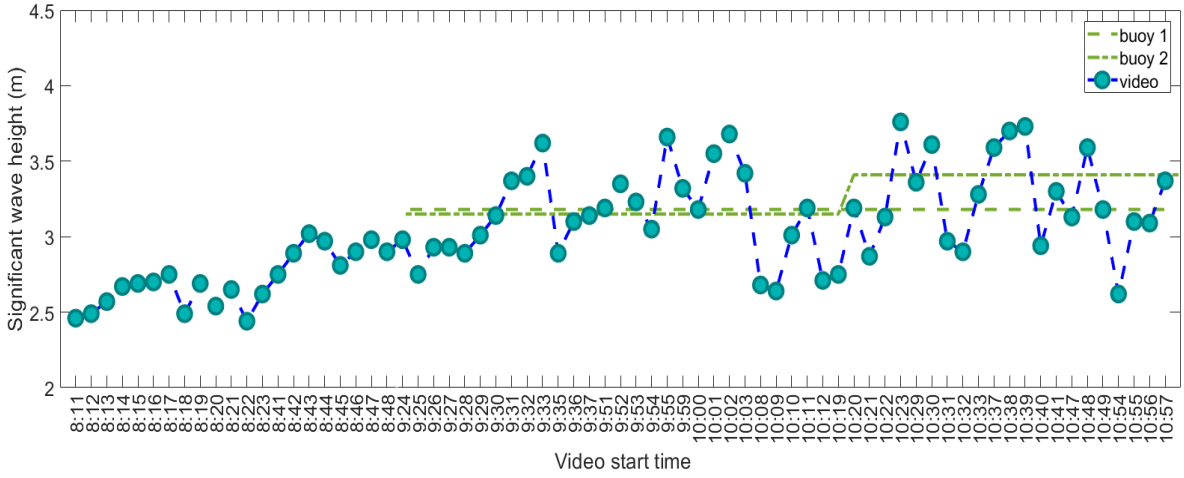
Chapter 10

Experimental results

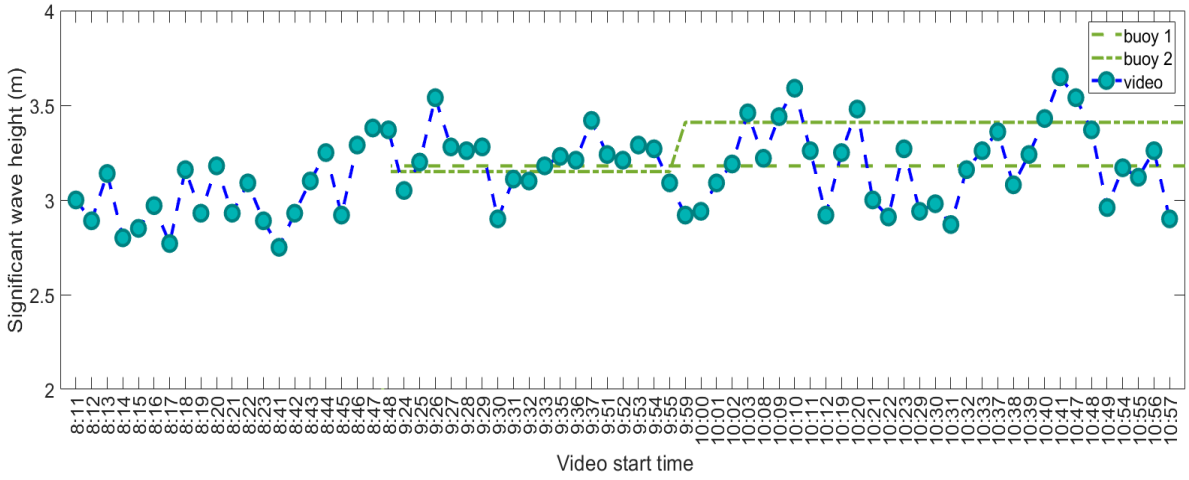
This chapter is provided for a comparison between the results of chapters 7, 8 and 9. As a reminder, the methodology in chapter 6 uses the spatial variation from high-scale images of the ocean and utilises radar images for testing. Chapters 7, 8 and 9 use the same video data as input, namely shipborne and tower data, introduced in chapter 5, sections 5.1 and 5.2. This enables the direct comparison of the experimental results of these chapters. Although the experimental results are presented in full in each chapter, here the same plots are reintroduced for the purpose of comparison. A description of the error metrics used in this chapter (and the previous chapters) is provided in chapter 5, section 5.5. In the following text, 'methodology 7' denotes the methodology in chapter 7, 'methodology 8' the methodology in chapter 8 and 'methodology 9' the methodology in chapter 9.

10.1 Shipborne video

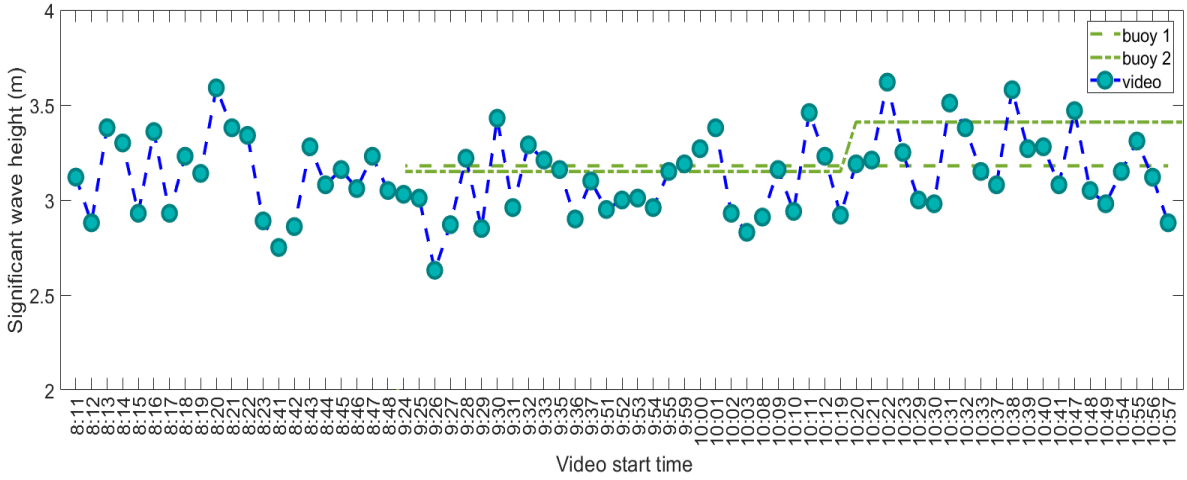
First, a comparison is given between the plots of the shipborne data in figure 10-1. For times between 8:11 to 9:24, methodology 7 gives a sea state between approximately 2.4 m to 3.1 m. For the same time instances, methodologies 8 and 9 give higher estimations. For the rest time instances, where buoy measurements are known, all three methodologies give results that are close to the buoy measurements. Methodology 9 seems to be performing very well, as the differences between buoys and video esti-



(a) Chapter 7



(b) Chapter 8



(c) Chapter 9

Figure 10-1 Comparison of experimental results of chapters 7, 8 and 9 for shipborne video. Buoy 1 $h_s = 3.18m$ (9:15-10:00 am), $h_s = 3.18m$ (10:00-11:00 am), Buoy 2: $h_s = 3.15m$ (9:15-10:00 am), $h_s = 3.41m$ (10:00-11:00 am).

Table 10.1 Error metrics with shipborne video. 'M 7' is the methodology introduced in chapter 7, 'M 8' the methodology of chapter 8 and 'M 9' the methodology of chapter 9.

Error metric	M 7	M 8	M 9
Mean Absolute Error (MAE)	0.31	0.15	0.18
Root Mean Square Error (RMSE)	0.37	0.19	0.22
Mean Absolute Percentage Error (MAPE)	9.86	4.83	5.65
Mean Percentage Error (MPE)	4.26	-0.05	2.19

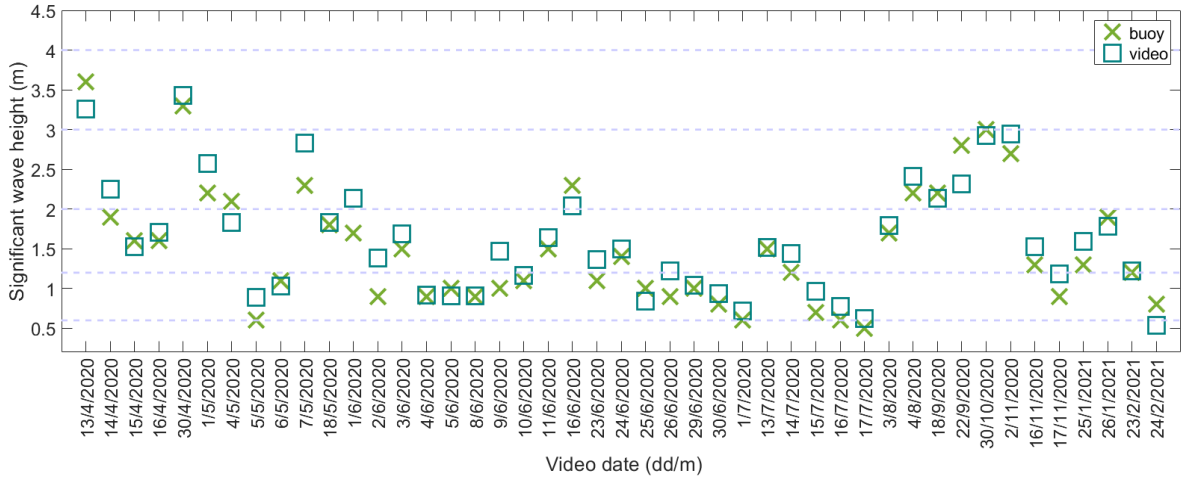
mations are not very large. Methodology 7 seems to have bigger differences between buoy measurements and video estimations.

A clear picture of the performance of each methodology for the shipborne video is given in table 10.1. The higher error metrics for methodology 7 validate the observation in the plots of bigger differences between buoy measurements and video estimations. Methodology 8 outperforms the other 2 methodologies, although the differences in errors between methodologies 8 and 9 are very small. In general, all 3 methodologies provide good estimations that are close to the buoy measurements. From the MPE values, the results seem balanced between overestimations and underestimations, with a small underestimation with methodology 7.

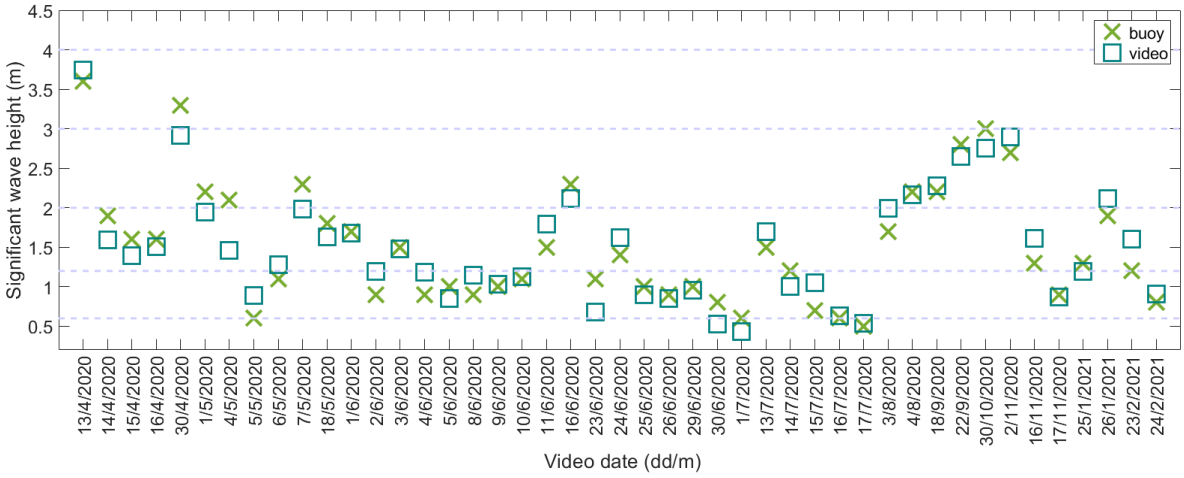
10.2 Tower video

For the tower video, the plots are presented in figures 10-2 and 10-3. All 3 methodologies provide estimations that are close to the buoy measurements. One observation from the plots is that one methodology might overestimate, whereas another might underestimate. There is no correlation between the estimations of one methodology with the estimations of another methodology. From the scatter plots in figure 10-3, the results for methodology 8 seem balanced between overestimations and underestimations. For methodologies 7 and 9 there seems to be an overestimation of the results.

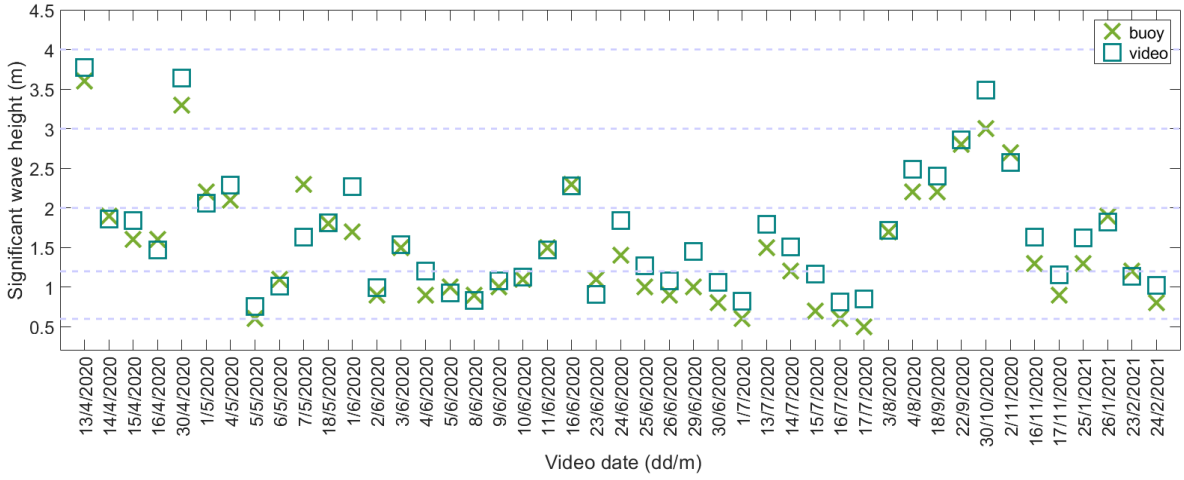
Table 10.2 gives a more clear picture of the performance of each methodology against the tower video. The same observation made with shipborne video is also



(a) Chapter 7

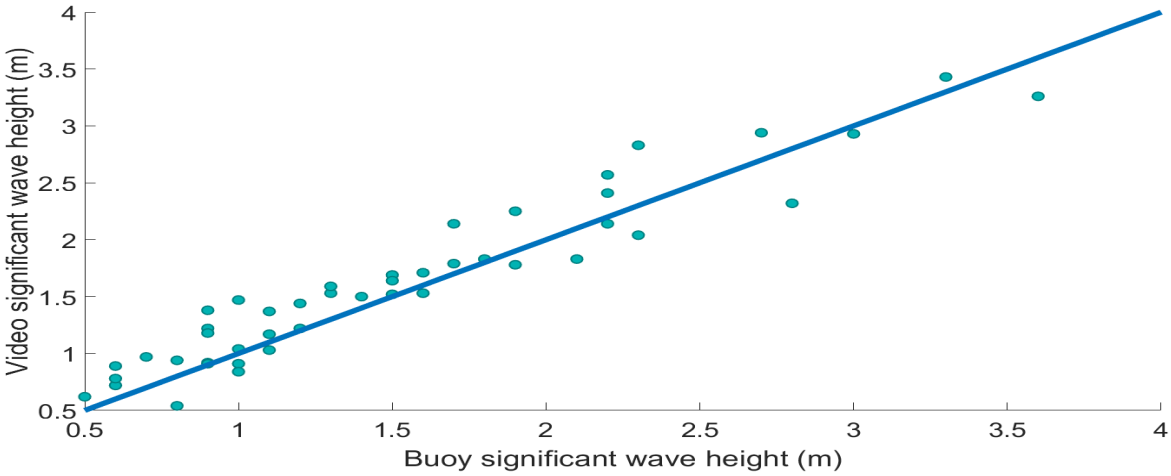


(b) Chapter 8

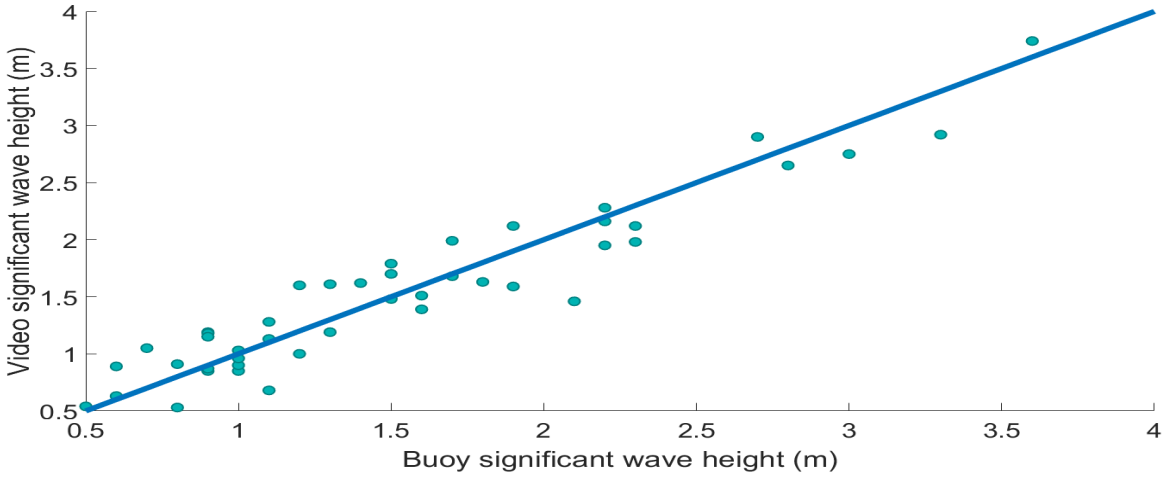


(c) Chapter 9

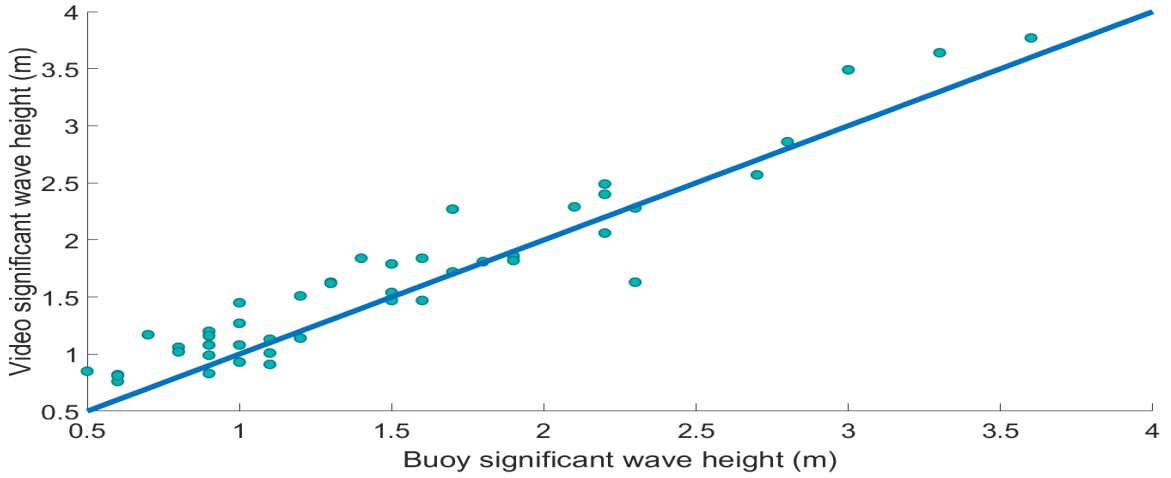
Figure 10-2 Comparison of experimental results of chapters 7, 8 and 9 for tower video (part 1).



(a) Chapter 7



(b) Chapter 8



(c) Chapter 9

Figure 10-3 Comparison of experimental results of chapters 7, 8 and 9 for tower video (part 2).

Table 10.2 Error metrics with tower video.

Error metric	M 7	M 8	M 9
Mean Absolute Error (MAE)	0.20	0.19	0.21
Root Mean Square Error (RMSE)	0.24	0.23	0.26
Mean Absolute Percentage Error (MAPE)	16.14	15.29	18.16
Mean Percentage Error (MPE)	-10.35	-1.92	-13.59

made with tower video, that methodology 8 outperforms the other 2 methodologies. The error metrics are higher for the tower video in comparison with the shipborne video. This is speculated to be the case due to the variety of sea states in the tower video, whereas the shipborne video are captured in a relatively stable sea state. Methodology 9 seems to have the highest errors for the tower video. The speculation is that this larger error might be due to the use of a secondary oscillatory component from video for the calibration. Methodologies 7 and 8 use the dominant frequency for calibration.

The MPE values indicate an overestimation of the results for methodologies 7 and 9. In the case of methodology 8, the overestimations and underestimations offset each other. Practically, in some instances, it is more desirable that a given methodology is overestimating rather than underestimating. For example, when used for providing information for the execution of maritime operations (e.g. landing of helicopter on ship), a methodology that underestimates might allow the execution of the operation when it shouldn't. A methodology that overestimates could thus be used more safely for these operations.

From both data sets, methodology 8 provides the most accurate estimations that are the most balanced between overestimations and underestimations. From the shipborne video, methodology 9 overcomes methodology 7 in terms of accuracy, but not in a large degree. From the tower video, methodology 7 overcomes methodology 9 in terms of accuracy, again not in a large degree. The estimations with the tower data set are less accurate from the estimations with the ship data set. Possible reason for this difference is that the tower data set is captured across a variety of sea states, whereas the ship data set is captured in an approximately stable sea state. The

difference in the experimental setup and environment could be an additional reason.

Chapter 11

Discussion

In this chapter, first information is provided on the selection of the Pierson-Moskowitz spectrum in comparison to the other empirical energy spectra (Sect. 11.1). Then, a discussion is given on possible ways the experimental results of this work could be extended and improved in terms of uncertainty of estimation (Sect. 11.2).

11.1 Selection of empirical energy spectrum

The empirical energy spectra introduced in the ocean theory chapter 2 (Sect. 2.2) are amongst the most popular ocean energy spectra in the oceanography literature. These are: the Pierson-Moskowitz spectrum (Sect. 2.2.1), the Jonswap spectrum (Sect. 2.2.2), the TMA spectrum (Sect. 2.2.3), the Phillips spectrum (Sect. 2.2.4) and the two-parameter spectrum (Sect. 2.2.5). All these spectra have the common input parameter of sea state in terms of wind speed or dominant angular frequency and provide as output the spectral density.

This work utilises the Pierson-Moskowitz spectrum in chapters 6, 7 and 8. The Pierson-Moskowitz spectrum and the Phillips spectrum have only the parameter of wind speed (or dominant angular frequency) as input. Choosing between the two spectra is based on the preference of the user. There is no previous research in the literature indicating a superiority between the two. These spectra are formed from different scientists based on different measurements of the sea, with common

assumptions (e.g. presence of fully developed sea).

The remaining empirical energy spectra mentioned in the ocean theory chapter 2 (Sect. 2.2) utilise additional parameters for the generation of the spectral density. They have the potential advantage of producing energy spectra that are more localised to a given ocean surface, due to the presence of more parameters. That is, these spectral formations could produce spectrum shapes that are more similar to the true spectrum shape of a given measured ocean. This could therefore have a positive effect on the estimation results of the video techniques. These additional parameters are: the fetch distance (and peak enhancement factor) for the Jonswap spectrum, the fetch distance (and peak enhancement factor) and water depth for the TMA spectrum and a parameter indicating fully or partially developed seas for the two-parameter spectrum. The fetch distance is the horizontal distance between wind source and measured ocean surface.

The two differences between the Pierson-Moskowitz spectrum and the Jonswap spectrum are the peak enhancement or peak-shape parameter and the fetch length parameter. Specifically, the Jonswap spectrum is the product of the Pierson-Moskowitz spectrum and the peak enhancement factor [18]. The peak-shape parameter is usually chosen to be equal to 3.30, although it may vary in terms of values between 1 and 10 in even constant sea states [101]. It is usually approximated normally distributed with a mean of 3.30 and a variance 0.63 [120]. This peak enhancement factor is multiplied in order to account for non-linear wave-wave interaction in partially developed seas. The Jonswap spectrum can thus be thought as an extension of the Pierson-Moskowitz spectrum with a more pronounced peak, an enhancement that takes into account non-linear wave-wave interactions and fetch distance.

The methodology of chapter 6 utilises the Pierson-Moskowitz spectrum in order to translate the value of the dominant frequency found from the phase speed matrix into a value of the wind speed. In other words, it is used as a 'conversion' mechanism in order to transfer the value of the dominant frequency into a value of a metric that is more meaningful. From empirical experiments, the use of the Phillips equation instead of the Pierson-Moskowitz equation provides values of wind speed that are

close to the values derived with the latter. This is to be expected because of the similar assumptions between the two spectral formations.

Choosing an empirical energy spectrum that accepts more parameters in the input will require the technique to estimate these parameters from video. Alternatively, this information (e.g. fetch distance) could be calculated or estimated during the acquisition of the video data. For example, with the Jonswap spectrum, the value of the fetch distance would need to be known *a priori* (and the peak-shape parameter be clearly defined). There is no previous research on estimating the peak enhancement or fetch distance from video, thus estimating these parameters from video could be challenging.

The most probable scenario would be that these parameters are known *a priori*. If this kind of information is available, then it would be advantageous for the user to choose an empirical energy spectrum that takes into account these conditions (i.e. takes more parameters as input). From the experimental data of this work (described in the experimental data chapter 5, Sect. 5.1, 5.2), the parameters of peak-shape and fetch distance are not known. Thus it is not possible in the context of this work to utilise the benefits of a more elaborate spectrum, such as the Jonswap spectrum.

In terms of chapter 8, the Pierson-Moskowitz spectrum is used, in a similar manner as in chapter 6, in order to convert the estimated from video dominant frequency into a value of a metric that is more meaningful (the significant wave height). The spectrum is not used in order to calculate the dominant frequency. Because the empirical energy spectrum is only used for conversion of what is already estimated from video, the user of a particular experiment may decide which spectrum to use based on the available *a priori* information of the measured ocean.

In conjunction to the methodologies of chapter 6 and 8, in chapter 7 the Pierson-Moskowitz spectrum is used in the calibration process (not for converting the estimated dominant frequency into a more meaningful metric). In this case, choosing a more elaborate spectrum could have a greater positive impact on the video estimations. As mentioned, research in the literature does not study the estimation of parameters such as peak-shape and fetch length from video. Future work could focus

on measuring these parameters during the video capture experiments and using these in order to incorporate a more elaborate spectrum (such as the Jonswap spectrum) in the technique.

It would be noteworthy to mention previous research comparing the Pierson-Moskoritz with the Jonswap spectrum. In [6] Monte Carlo simulations are used in order to compare the simulation results of offshore structural responses due to random wave loading with the Pierson-Moskowitz spectrum and the Jonswap spectrum. These responses are generated with the use of linear random wave theory. Comparing the two spectra with the same data in the input, the Jonswap spectrum is observed to have more sharp peaks. This is due to the presence of the peak enhancement factor. With the Pierson-Moskowitz spectrum, the spectral energy is more spread towards higher frequencies. The authors conclude that both spectra reproduce realistic energy spectral shapes.

In conclusion, this work selects the Pierson-Moskowitz spectrum. If the Phillips spectrum is used in chapter 6 in order to convert the estimated dominant frequency into a value of the wind speed, the results are similar. This is because both energy spectra have the common assumption of fully developed sea. We conjecture that selecting a more elaborate energy spectrum will produce a more localised shape to the specific measured ocean. It would thus return more accurate estimations. We are however limited by not having available information (e.g. fetch distance) from our specific datasets in order to utilise these elaborate spectra. Additionally, there is no research indicating a superiority of one spectrum over another. That is, it is not certain that the utilisation of another energy spectrum will provide more accurate results. Future work will focus on calculating parameters such as fetch distance during the acquisition of the video data and comparing the results between the Pierson-Moskowitz spectrum and other more elaborate spectra.

11.2 Uncertainty

The estimation results from the video methodologies are validated against buoy measurements, which are considered to indicate the true sea state. The video methodologies are evaluated with four error metrics, the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean percentage error (MPE). More information about the error metrics can be found in the experimental data chapter 5 (See Sect. 5.5). A description of the input video data for validating the presented methodologies is provided in chapter 5, sections 5.1 and 5.2. A description of the error metrics is provided in chapter 5, section 5.5. In this discussion section, information in regards to the uncertainty in the sea state estimation is provided, in addition to where future work will focus in order to provide this measure of uncertainty.

In terms of theory of statistical inference [35][30], it would be beneficial the techniques of this work to provide in addition to the value of sea state estimation a value for the theoretical error bounds in the form of a value of standard deviation. In this case, each estimated sea state value would be accompanied with theoretical bounds as a measure of how certain the methodology is in regards to the estimated value. Each sea state estimation would be expressed as an interval of values within which the true sea state value is expected to be contained, with a stated coverage probability. These theoretical error bounds provide a quantitative indication of the reliability of the estimation.

Each sea state estimation in this scenario is expressed as a set of values μ, σ and P, where μ is the mean value, σ is the standard deviation of uncertainty and P is the coverage probability [170]. For example, the sea state can be expressed as a value of significant wave height 1 m with a standard deviation of 0.3 m and a coverage probability of 68 %. This means that the true value of significant wave height is believed to lie within the interval 0.7 m-1.3 m (one standard deviation) with a probability of 68 %. Future work will focus on improving the methodologies of this work so that each estimated sea state value is accompanied with theoretical bounds

of uncertainty.

In the literature [46][108], two main categories of inferential uncertainty are the aleatory uncertainty and the epistemic uncertainty. The aleatory uncertainty follows a frequentist approach and is formed by specifying the experiment that will be repeated [71][85]. It involves the execution of experiments under different conditions in order to investigate the different outcomes. The sampling distribution quantifies the uncertainty and can be obtained by executing an infinite number of experiments. In order to derive the sampling distribution, a clear definition of the experiment that will be repeated is required. The epistemic uncertainty is formed by specifying the prior belief about the system [115]. It involves the investigation of missing knowledge that can be either true or false. A prior distribution is formed that describes the uncertainty in regards to a parameter based on prior knowledge, such as knowledge from a similar situation.

In the following text, in Sect. 11.2.1 a discussion is given on whether the values in the main diagonal of the covariance matrix of the Kalman filter can be used as measures of uncertainty for the sea state estimation. In Sect. 11.2.2, a discussion is given on possible sources of error in estimating the sea state and how they might contribute to the values of the error metrics. In Sect. 11.2.3 different ways to acquire multiple video estimations are considered. These can be multiple sea state estimations for the same video, or multiple video estimations for the same sea state. Section 11.2.4 provides information on how these multiple estimations can be used in order to derive the theoretical bounds of confidence with the bootstrap method.

11.2.1 Uncertainty from covariance matrix of Kalman filter

In the case of the methodologies in chapters 8 and 9, an initial idea would be to use the square root of the corresponding diagonal element from the covariance matrix of the extended Kalman filter as a measure of uncertainty. However, this value indicates how certain the extended Kalman filter is about the identification of a single sinusoid from the sum of elementary reconstructed components (RCs) of the SSA algorithm given in the input.

If the input of the extended Kalman filter was directly the time series of pixel intensities, the square root of the entries in the main diagonal of the covariance matrix could be used as values of the standard deviation indicating certainty of the methodology in regards to the estimation of the sea state. Specifically, if the extended Kalman filter receives directly the time series of pixel intensities and the environment is defined so that the algorithm tracks one main frequency (to be used as the ocean dominant frequency), the entry in the main diagonal of the covariance matrix indicates how certain the algorithm is about the specified frequency in the output being the 'main' frequency of the input time series.

Because the input is already a pre-processed time series, the theoretical bounds of error from the covariance matrix of the extended Kalman filter do not reflect the uncertainty of the sea state estimations. The meaning behind the square root value of the main diagonal entry of the covariance matrix is (for example third diagonal element corresponding to angular frequency): 'how certain the algorithm is that the main sinusoidal component of the time series given in the input has the angular frequency given in the output'.

11.2.2 Possible sources of error

One possible way to provide a measure of uncertainty is to first identify all potential sources of error and then create a mathematical model that combines them [170]. In the context of this work, possible sources of error could be: quality of the video capture [81], experimental setup (e.g. angle of camera), environmental brightness fluctuations [121], presence of rain or fog in the environment, presence of capillary waves or ripples on the ocean surface [81], presence of multiple wave fields (e.g. presence of swell and wind wave fields) and the general environmental brightness (e.g. cloudy or sunny day). Then, for each potential source of error, a standard deviation of uncertainty value is assigned from experimental observations or from other available information.

A mathematical model is created that combines all potential sources of error and represents their contribution to the overall uncertainty of the sea state estimation process. The final value of standard deviation of uncertainty for each video is then

calculated by using the mathematical model of uncertainty with the specific values of the underlying sources of error for that specific video. A challenge here is to quantify the measure of uncertainty for each underlying source.

The variability of the estimate of the mean is a measure of uncertainty in the estimate [165]. In the context of this work, this variability can be measured by capturing the same ocean surface under different conditions. For example, the same ocean surface can be measured from different camera angles. The variability in the sea state estimations is the measure of uncertainty for that underlying source of uncertainty (camera angle).

Each possible source of error has a given contribution to the values of the error metrics MAE, RMSE and MAPE, that are used in order to evaluate the methodologies of chapters 7, 8 and 9. In the following text, a small discussion is given on how each possible source of error might contribute to the values of the error metrics.

In terms of quality of the video capture, it is expected that the better the quality of the video, the smaller is the value of the error. This is because the pixel intensities are capturing the wave field more accurately, which enables the methodologies to track the important information from video more easily.

In terms of the angle of the camera, video captures that are more vertical to the ocean surface are expected to provide better accuracy. This is because, as the camera is facing more directly the ocean surface, the distance between camera and ocean surface gets smaller. Thus, the camera is able to capture more accurately, with more precision and clarity, the ocean because it is closer in distance. Each pixel corresponds to a smaller area of ocean. Thus, the video methodology receives in the input a more accurate representation of the temporal variation of the wave field.

Environmental brightness fluctuations are expected to negatively affect the values of the error metrics. This is because the sun's direct reflections (sun's glitter) on the ocean surface have very high intensities in the video. These high intensity values could potentially have a negative effect on the efforts of the methodology to track the temporal variation of the wave field from video.

The presence of the sun's glitter in certain pixel locations of the same video (while

other pixel locations do not reflect directly the sun's rays) might also introduce errors. The video methodologies are executed with a single time series of pixel intensities in the input, from a specific pixel location of the video. As discussed in the experimental data chapter 5 (See Sect. 5.1 and 5.2), from a single video multiple estimations are derived from multiple pixel locations and then the final sea state is their average. This is done in order to improve the estimations. However, if some pixel locations are affected by the sun's glitter and some are not, the results might not be as accurate. It is expected that videos that capture the ocean surface on a cloudy day will provide more accurate sea state estimations, due to the homogeneity of the brightness of the environment and thus homogeneity of the variation of the pixel intensities from multiple locations of the same video.

Adverse weather conditions can have an effect on the values of the error metrics for the evaluation of the sea state estimation. The presence of rain or fog in the environment is expected to negatively affect the estimations. Rain will introduce ripples on the water surface, which could have a negative influence on the estimations. More about the effect of ripples is discussed in the following paragraphs. Additionally, rain can affect the quality of the video capture if the camera is not protected in an indoor environment. Presence of adherent water droplets on camera lenses or enclosures can render a video unusable. The presence of fog also has a negative effect on the quality of the video and thus potentially the video estimations.

In the case of a storm, where high local wind is present, the camera could have a given 'shaking' or vibration due to local wind (this is the case with the tower video). Although stabilisation features of a computer software (See experimental data chapter 5, Sect. 5.2) are used, these features are not always perfect, and thus high local wind could also introduce higher values of error in the estimations.

The presence of very small amplitude waves, called capillary waves or ripples could have a negative influence on the sea state estimation. As discussed in Chapter 6 (See Sect. 6.1.3), these waves do not follow the typical frequency dispersion relation of water gravity waves [126][110]. Because the methodologies of this work do not model these waves, their presence could thus have a negative effect on the estimation results.

In Chapter 7, the scaling process (See Sect. 7.1.4) is designed in order to calibrate the video information even if multiple peaks are present in the video amplitude spectrum. Therefore, the presence of multiple wave fields (e.g. swell and wind) is not expected to have a great negative effect. In Chapter 8, it is not clear how the methodology will execute if multiple wave fields are present. As a reminder, the methodology tracks the main oscillatory component from video and uses the extended Kalman filter in order to get one main frequency and use it as the ocean's dominant frequency. Even if multiple wave fields are present, the methodology will potentially be able track the wave with the highest energy, and thus the estimations might not be greatly negatively affected.

In the case of the methodology in Chapter 9, the presence of multiple wave fields could have a greater negative effect in the estimations. This is because the algorithm aims at tracking a secondary oscillatory component from video with a frequency that belongs in the equilibrium range (theoretically frequencies higher than the dominant frequency). The methodology could potentially select a peak of a secondary wave field instead of a frequency in the equilibrium range. This could have a negative effect on the sea state estimations.

11.2.3 Acquiring multiple video estimations

In order to express the sea state estimation in terms of a standard deviation of uncertainty, the first step is to collect a set of estimations for the same video or for the same sea state. In the case of acquiring a set of estimations for the same video, multiple videos could be captured in a short time period (e.g. in the time frame of 10-20 min) under different conditions, such as different angles of the camera. With the assumption of statistically stationary and homogeneous ocean surface for that time period, it is possible to acquire multiple sea state estimations for the 'same' video.

Another approach to acquire multiple sea state estimations from the 'same' video would be to capture multiple consecutive videos of the ocean surface with the assumption that the sea state is exactly the same for that time period (e.g. one hour). This is almost the same as what is discussed in the previous paragraph, only this time

it is under the same experimental conditions. Only the video capture time varies. Although the assumption of statistically stationary sea is usually considered true in the time span of one hour, in practice it is not always certain that this assumption is valid. If this assumption is not valid for a given time span, the limits of uncertainty could potentially be less accurate.

Another approach to collect multiple sea state estimations to be used for generating confidence intervals is to capture multiple videos across many days that have the exact same corresponding value of significant wave height. In other words, videos captured in different dates are categorised based on their corresponding *in situ* sea state information. This is a completely different approach to what is discussed in the previous paragraphs. In this case, the experimental results are not presented for each individual video, but for each individual sea state.

In this scenario, videos are categorised into groups based on their corresponding sea state. For example, videos can be categorised into groups that have a 0.1 m step between them of difference in significant wave height. In this example, three different categories would comprise of videos with significant wave height of 1.1 m, 1.2 m and 1.3 m.

In the present work, each pair of estimated-true sea state from video represents a video captured on a specific date and time. In the scenario where videos are categorised into groups, each pair of estimated-true sea state represents a set of multiple videos that have the same corresponding sea state. The experimental results in this scenario are reported in terms of mean and standard deviation for each sea state and not for each video.

11.2.4 The bootstrap method for generating confidence intervals

In the previous section 11.2.3, we discussed possible ways to acquire multiple sea state estimations. In the following text we will consider the case where we collect multiple sea state estimations for the 'same' video, where we change the experimental

conditions or capture multiple videos in a short time period, and we consider the sea state statistically stationary for that period. Now that we have multiple values of sea state estimation for the same video, we can use the bootstrap method [45] in order to generate the confidence intervals for the video. The multiple values of sea state estimation for a given video (See Sect. 11.2.3) are the samples given as input to the bootstrap method.

The bootstrap method [45] is a popular process for acquiring the confidence intervals from data, with the advantage of not requiring any assumptions about the distribution of the estimations [89]. The basic idea is to replicate the same experiment thousands of times by selecting different samples in the input. In the context of this work, the samples are selected from the set of sea state estimations for the same video. In this case, the sample size is the size of the sea state estimations for one particular video.

The algorithm selects the samples in the input in a random manner. A sample can be selected multiple times in one selection process. The number of samples selected is equal to the sample size. This procedure is called *resampling with replacement* [79]. The selected data set is called the *resampled data set*. Then, the algorithm calculates the required metric from the resampled data set. In our case, this metric is the mean. This resampling procedure is executed thousands of times, simulating the procedure of executing the same experiment under different conditions. This enables the estimation of the theoretical bounds of uncertainty. Each time the resampling procedure is executed, the algorithm records the mean value. After the execution of thousands of resampling procedures, the algorithm provides in the output a set of thousands of mean values.

The process after the execution of the resampling procedures is straightforward. The standard deviation of uncertainty or standard deviation of confidence is found as the standard deviation of the thousands of values of the mean. The formula for the confidence interval of the mean value is equal to [79]:

$$\mu \pm Z\frac{\sigma}{\sqrt{n}} \tag{11.1}$$

where Z is the Z-value from the standard normal distribution of the given confidence interval, μ is the mean, σ is the standard deviation and n is the number of samples (how many times the resampling procedure was executed). For example, if we were to have a confidence level of 95 %, the corresponding Z-value is equal to 1.96 [95].

From the bootstrapping method we have already calculated the mean μ and standard deviation σ of the population and we know the population number n. We thus report the results of 95 % confidence interval for the sea state estimation as $\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$. If for example μ_- is equal to $\mu_- = \mu - 1.96\frac{\sigma}{\sqrt{n}}$ and μ_+ is equal to $\mu_+ = \mu + 1.96\frac{\sigma}{\sqrt{n}}$, the theoretical bounds of confidence for the sea state estimation are reported as: with 95 % confidence, the sea state estimation is between μ_- and μ_+.

In the context of this work, the performance of a video methodology is evaluated in terms of how narrow the confidence interval is and whether the true value lies within this interval. The true sea state is considered to be equal to the sea state from buoy measurements. If the confidence interval is too wide, the methodology is less confident about the estimation. If the interval is wide and the true value is not within the interval, the estimation is less meaningful. In a more ideal scenario, a methodology would estimate the sea state with narrow confidence limits and the true sea state would be within these limits. It is noteworthy to mention that the bootstrap method does not contain more information about the sample population than the original sample [27]. Thus, it might not be useful if the sample size is small.

Chapter 12

Conclusion

The use of video of the ocean surface in real environments for obtaining the sea state and other useful information is investigated in this work. Chapter 6 examines the estimation of the pixel to metre scale and sea state in large scale videos of the ocean with the utilisation of radar images. The spatial variation and the phase modulation are used for the estimations and an iterative process is proposed that selects only the waves with the highest energy. Chapter 7 proposes a methodology that takes as input time series of grayscale pixel intensity values and utilises the Kalman filter and the least squares approximate solution to get the uncalibrated video amplitude spectrum. A scaling method is then proposed for transferring the results into metres, enabling the calculation of the significant wave height. This scaling method uses ocean theory of an empirical energy spectrum.

In chapter 8, a methodology is presented for estimating the significant wave height with the same type of input. The methodology utilises the SSA algorithm and the extended Kalman filter in order to approximate the dominant frequency, and then uses the shape of an empirical energy spectrum in order to translate this value into a measure of the significant wave height. The technique in chapter 9 aims to estimate the sea state without the use of an empirical energy spectrum. Calibration is done with the use of the equilibrium range theory of oceans. In this case, the SSA algorithm and the extended Kalman filter are utilised in order to identify a secondary oscillatory component from video, which is used with the ocean theory for calibration. One

common characteristic of all proposed techniques is the incorporation of ocean theory for the acquisition of the sea state estimations.

Testing is done with two video data sets in different oceans and sea states that have corresponding buoy measurements indicating the true sea state. The range of significant wave height for which the techniques of this work are tested is between 0.5 m and 3.6 m. Higher sea states are not examined and thus the applicability in such instances is not known. The estimation of higher sea states might not be required for some practical applications. For example, in the case of the use of the aforementioned techniques for providing information for the execution of maritime operations (for example landing of helicopters on ships), higher sea states are distinguishable by mariners. Future work could however examine the applicability of the presented methodologies in higher sea states with the availability of videos.

The present work can be helpful for decision-making in marginal sea states. For example, if information is required for the execution of maritime operations, a mariner can observe a sea state 10 and know that the operation cannot take place due to the very rough surface of the ocean. Equally, simply by observation a mariner is able to verify the calmness of the sea in a sea state 1 and provide information for allowing the execution of the maritime operation. The challenging decisions are to be made where the sea state is marginal. In such a case, when it is not certain whether it is safe to execute the operation by simple observation, the presented techniques can be used as a useful guide.

One common characteristic of some of the techniques of this work is the scaling or calibration process, which provides a link between the digital information from video with real world units. In chapter 6, the estimation of the pixel to metre scale enables the direct estimation of the wind speed with the use of the spatial variation of the wave field in large scale videos of the ocean. In chapter 7, this is achieved with the use of the location of the peak in the calculated video amplitude spectrum and the Pierson-Moskowitz spectrum. In chapter 8, the area of the shape of the Pierson-Moskowitz spectrum is used for acquiring a value of the significant wave height from the estimated dominant frequency. In chapter 9, calibration is achieved with the

identification of a wave with a frequency higher than the dominant frequency and the equilibrium range theory.

Future work will focus on the expansion of the presented techniques in order to provide additional sea state information. From the theory of the empirical energy spectrum, a value of significant wave height can be transferred to a value of wind speed. From the dominant frequency estimation, the dominant wavelength can be found with the use of the dispersion relation. Expansion of the techniques could be done in order to provide directly information of these sea state parameters without the use of ocean theory, that is ocean theory would be used only for calibration. Expansion could also be done for providing wind direction, as well as for providing information of multiple wind fields. An expanded technique could potentially provide detailed information of combinations of wave fields, for example swell and wind wave fields.

Future work will also focus on expanding the evaluation of the model for higher sea states. This would involve the acquision of video footage in higher sea states. In such a case, *in situ* concurrent measurements are useful for indicating the true sea state. The direction of future projects could potentially be focused at using information of the pixel variation from single frames as additional information to the time series of pixel intensity with the purpose of providing more accurate estimations. Information of the phase speed from chapter 6 could also be incorporated in the techniques of chapters 7, 8 and 9. Future projects could combine two or more of the techniques of this work for the same purpose of providing more accurate estimations.

Future work will focus on extending the video methodologies in order to provide theoretical limits of confidence in the estimations. The concepts of this extension are elaborated in the discussion chapter 11, Sect. 11.2. The most possible scenario that future work will focus on is to capture multiple consecutive (e.g. 20-30) one-minute videos of the ocean. We will assume that the sea state is statistically stationary for that time period. Then, the bootstrap method (See chapter 11, Sect. 11.2.4) can be applied in order to get the confidence limits. These multiple consecutive video estimations are used as the input sample data set of the bootstrap method. In order

to achieve this, additional experimental results will be obtained, where video of the ocean surface is captured for extended periods of time across many days. Similarly to the data sets used in this work, concurrent *in situ* measurements are to be used for validating the estimations. Measuring continuously the ocean surface enables the estimation of the sea state in terms of limits of confidence. Measuring the ocean surface across many days enables the estimation in different sea states.

The accuracy error metrics for evaluating the experimental results are the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE). In regards to the MAPE, the technique described in chapter 7 has an error of 9 % and 16 % for shipborne and tower video respectively. The technique described in chapter 8, although more simplistic than the technique of chapter 7, has a MAPE of 4 % and 15 % for shipborne and tower data. The technique in chapter 9 has a MAPE error of 5 % and 18 % for shipborne and tower data. The larger error values for the technique of chapter 9 is attributed to the use of a secondary oscillatory component for calibration, which is conjectured to be more affected by noise than the dominant frequency. The remaining techniques use calibration processes dependent more on the dominant frequency. Thus, it can be concluded that when using ocean theory for calibration, the results could be more accurate if the dominant frequency is used.

In general, the proposed techniques seem to be estimating something meaningful. This observation is derived from the results of both the shipborne video data, which are of approximately the same high sea state, and the tower data, which have both low and high sea states. The proposed work, although preliminary, provides promising prospects on the use of video in real environments for the sea state estimation.

Compared to chapter 6, the techniques described in chapters 7, 8 and 9 and are not limited to large scale videos of the ocean. The technique of chapter 8 has the advantage of being more simplistic than the remaining techniques, in addition to providing more accurate estimations. It does however depend on the specific shape of an empirical energy spectrum, in conjuction to the technique of chapter 7 that utilises the video amplitude spectrum for the estimations. The technique of chapter 9

has the advantage of not using the shape of a specific customised energy spectrum for the calibration. Utilising however frequencies different from the dominant frequency might provide estimations with lower accuracy due to the presence of noise mixed with secondary waves in video.

Apart from chapter 6, which is provided as an introduction to the use of the spatial variation of the wave field from large scale video with radar images, the remaining chapters utilise the temporal variation of the wave field. This is due to the unavailability of large scale videos in real environments, such as airborne videos. Future work could focus on the utilisation of the spatial variation of the wave field from airborne videos of the ocean for estimating the sea state. Drones could also be used for this purpose.

Compared to the use of *in situ* devices, such as wave buoys, for estimating the sea state, utilising video of the ocean requires less resources, as buoys require funds for their acquisition and maintenance. Buoys also have the danger of being damaged in extreme weather conditions. Wave buoys however are more reliable at the estimation. The decision on the use of video, other means of remote sensing or *in situ* devices lies on the preference of the user, as well as the availability of funds for acquiring and maintaining the necessary equipment.

This work provides an introduction to what is possible with the use of video in real environments, and does not claim any superiority of the use of video to other remote sensing methods or *in situ* measurements. From the literature review chapter, it can be observed that the research utilising video of the ocean in real environments for the sea state estimation is limited. The techniques of this work provide an introduction to possible ways the information from videos can be incorporated with ocean theory for estimating the sea state.

Bibliography

[1] Beaufort wind force scale. https://www.metoffice.gov.uk/weather/guides/coast-and-sea/beaufort-scale. Accessed: 2019-11-04.

[2] Chapter 2 - Wave Theory. In Ian R. Young, editor, *Wind Generated Ocean Waves*, volume 2 of *Elsevier Ocean Engineering Series*, pages 3–23. Elsevier, 1999.

[3] Jay Abramson, Valeree Falduto, Rachael Grosss, David Lippman, Melonie Rasmussen, Rick Norwood, and Nicholas Belloit. *Precalculus*, chapter 7.2 Sum and difference identities. OpenStax, 2014.

[4] Tunde Aderinto and Hua Li. Ocean wave energy converters: Status and challenges. *Energies*, 11(5):1250, 2018.

[5] Adobe. After Effects CC 2019. Available at https://adobe.com/products/aftereffects.

[6] Ahmad, Sayyid Zainal Abidin Syed, Husain, Mohd Khairi Abu, Zaki, Noor Irza Mohd, Mohd, Mohd Hairil, and Najafian, Gholamhossein. Comparison of Various Spectral Models for the Prediction of the 100-Year Design Wave Height. *MATEC Web Conf.*, 203:01020, 2018.

[7] Theodore Alexandrov. A method of trend extraction using singular spectrum analysis. *arXiv preprint arXiv:0804.3367*, 2008.

[8] FJ Alonso, JM Del Castillo, and P Pintado. Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of biomechanics*, 38(5):1085–1092, 2005.

[9] Kelly Alonzo. A 3D state space formulation of a navigation Kalman filter for autonomous vehicles. *Carnegie Mellon University. Technical Report CMU-RI-TR-94-19-REV*, 2:105, 1994.

[10] Fabrice Ardhuin, Justin E Stopa, Bertrand Chapron, Fabrice Collard, Romain Husson, Robert E Jensen, Johnny Johannessen, Alexis Mouche, Marcello Passaro, Graham D Quartly, et al. Observing sea states. *Frontiers in Marine Science*, 6:124, 2019.

[11] Yong Bai. *Marine structural design*, chapter Wave Loads for Ship Design and Classification, pages 73–93. Elsevier, 2003.

[12] ML Banner, Ian SF Jones, and JC Trinder. Wavenumber spectra of short gravity waves. *Journal of Fluid Mechanics*, 198:321–344, 1989.

[13] NF Barber. Finding the direction of travel of sea waves. *Nature*, 174(4440):1048, 1954.

[14] Jean Baptiste Joseph Baron Fourier. *The analytical theory of heat*. The University Press, 1878.

[15] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.

[16] B Baxter, BA Hooper, JZ Williams, and JP Dugan. Polarimetric remote sensing of ocean waves. In *OCEANS 2009*, pages 1–5. IEEE, 2009.

[17] Alvise Benetazzo. Measurements of short water waves using stereo matched image sequences. *Coastal Engineering*, 53(12):1013 – 1032, 2006.

[18] M.A. Benitz, M.A. Lackner, and D.P. Schmidt. Hydrodynamics of offshore structures with specific focus on wind energy applications. *Renewable and Sustainable Energy Reviews*, 44:692–716, 2015.

[19] Elzbieta M. Bitner-Gregersen, Sheng Dong, Thomas Fu, Ning Ma, Christophe Maisondieu, Ryuji Miyake, and Igor Rychlik. Sea state conditions for marine structures' analysis and model tests. *Ocean Engineering*, 119:309 – 322, 2016.

[20] A. Bonaduce, N. Pinardi, P. Oddo, G. Spada, and G. Larnicol. Sea-level variability in the mediterranean sea from altimetry and tide gauges. *Climate Dynamics*, 47(9):2851–2866, Nov 2016.

[21] JC Nieto Borge and C Guedes Soares. Analysis of directional wave fields using X-band navigation radar. *Coastal Engineering*, 40(4):375–391, 2000.

[22] José Carlos Nieto Borge, Konstanze Reichert, and Jürgen Dittmer. Use of nautical radar as a wave monitoring instrument. *Coastal Engineering*, 37(3):331 – 342, 1999.

[23] Alexei Botchkarev. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14:045–076, 2019.

[24] Piotr Breitkopf and Rajan Filomeno Coelho. *Multidisciplinary design optimization in computational mechanics*. John Wiley & Sons, 2013.

[25] Charles L Bretschneider. *Wave variability and wave spectra for wind-generated gravity waves*. Number 118. The Board, 1959.

[26] David S Broomhead and Gregory P King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2-3):217–236, 1986.

[27] David Burgstahler. Discussion of an application of the bootstrap method to the analysis of squared, standardized market model prediction errors. *Journal of Accounting Research*, pages 55–58, 1984.

[28] RW Burling. The spectrum of waves at short fetches. *Deutsche Hydrografische Zeitschrift*, 12(2):45–64, 1959.

[29] S.M. Calisal. A note on the derivation of potential energy for two-dimensional water waves. *Ocean Engineering*, 10(2):133 – 138, 1983.

[30] Chris Chatfield. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158(3):419–466, 1995.

[31] CR Chou, JZ Yim, and WP Huang. Determining the hydrographic parameters of the surface of water from the image sequences of a CCD camera. *Experiments in fluids*, 36(4):515–527, 2004.

[32] U.S. Coast Guard. Coast Guard Concept of Operations for Offshore Assets. Technical report, Washington, D.C., 2017.

[33] U.S. Coast Guard. The United States coast guard's vision for enabling maritime commerse. Technical report, Washington, D.C., 2018.

[34] Charles Cox and Walter Munk. Measurement of the roughness of the sea surface from photographs of the sun's glitter. *Josa*, 44(11):838–850, 1954.

[35] Douglas Curran-Everett. Explorations in statistics: standard deviations and standard errors. *Advances in Physiology Education*, 32(3):203–208, 2008.

[36] Robert A Dalrymple and Robert G Dean. *Water wave mechanics for engineers and scientists*, volume 2. World Scientific Publishing Company, 1991.

[37] H. Dankert, J. Horstmann, and W. Rosenthal. Wind- and wave-field measurements using marine X-band radar-image sequences. *IEEE Journal of Oceanic Engineering*, 30(3):534–542, July 2005.

[38] H. Dankert and W. Rosenthal. Ocean surface determination from X-band radar-image sequences. *Journal of Geophysical Research: Oceans*, 109(C4), 2004.

[39] S. de Vries, D.F. Hill, M.A. de Schipper, and M.J.F. Stive. Remote sensing of surf zone waves using stereo imaging. *Coastal Engineering*, 58(3):239 – 250, 2011.

[40] S De Vries, DF Hill, MA De Schipper, and MJF Stive. Remote sensing of surf zone waves using stereo imaging. *Coastal Engineering*, 58(3):239–250, 2011.

[41] Robert G Dean and Robert A Dalrymple. *Water Wave Mechanics for Engineers and Scientists*, chapter Engineering wave properties. World scientific, 1991.

[42] Hassan Diab, Rafic Younes, and Pascal Lafon. Survey of research on the optimal design of sea harbours. *International Journal of Naval Architecture and Ocean Engineering*, 9(4):460 – 472, 2017.

[43] Michele Drago, Andrea Del Guzzo, Luigino Vitali, and Roberto Bruschi. Weather stand-by assessment in offshore operations using motion limit criteria. In *The 27th International Ocean and Polar Engineering Conference*. OnePetro, 2017.

[44] Michele Drago, Luigino Vitali, Andrea Del Guzzo, and Federico Gaggiotti. Decisional criteria for off-shore operations interruption due to adverse weather. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 58851, page V07BT06A007. American Society of Mechanical Engineers, 2019.

[45] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer, 1992.

[46] Bruce R. Ellingwood and Kursat Kinali. Quantifying and communicating uncertainty in seismic risk assessment. *Structural Safety*, 31(2):179–187, 2009. Risk Acceptance and Risk Communication.

[47] Explore Oceans. Frying pan ocean cam powered by explore.org. Available at `https://youtube.com/watch?v=vGimsBVpuN8`.

[48] Eric Firing and R Lee Gordon. Deep ocean acoustic doppler current profiling. In *Proceedings of the IEEE Fourth Working Conference on Current Measurement*, pages 192–201. IEEE, 1990.

[49] Matt Folley. *The Wave Energy Resource*, pages 43–79. Springer International Publishing, 2017.

[50] Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 10(4):439–451, 1988.

[51] Hermann Fritz, Jeseon Yoo, Kevin Haas, Paul Work, and C. Barnes. Particle image velocimetry applied to nearshore video wave images. In *Waves 2005*, 01 2005.

[52] Harold Q Fuller, Richard M Fuller, and Robert G Fuller. *Physics, including human applications*, chapter Traveling waves. HarperCollins Publishers, 1978.

[53] CG Gelpi, BC Schuraytz, and ME Husman. Ocean wave height spectra computed from high-altitude, optical, infrared images. *Journal of Geophysical Research: Oceans*, 106(C12):31403–31413, 2001.

[54] Nina Golyandina, Anton Korobeynikov, and Anatoly Zhigljavsky. *Singular spectrum analysis with R*, chapter 1 Introduction: Overview. Springer, 2018.

[55] Nina Golyandina, Anton Korobeynikov, and Anatoly Zhigljavsky. *Singular spectrum analysis with R*. Springer, 2018.

[56] Nina Golyandina, Vladimir Nekrutkin, and Anatoly A Zhigljavsky. *Analysis of time series structure: SSA and related techniques*, chapter Basic SSA: Basic Capabilities. CRC press, 2001.

[57] Nina Golyandina, Vladimir Nekrutkin, and Anatoly A Zhigljavsky. *Analysis of time series structure: SSA and related techniques*, chapter SSA Forecasting. CRC press, 2001.

[58] Nina Golyandina, Vladimir Nekrutkin, and Anatoly A Zhigljavsky. *Analysis of time series structure: SSA and related techniques*, chapter Homogeneity and Heterogeneity. CRC press, 2001.

[59] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, volume 120, chapter Trends of different resolution, page 19. Springer, 2013.

[60] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, volume 120, chapter Smoothing, page 20. Springer, 2013.

[61] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, volume 120, chapter Extraction of seasonality components, page 21. Springer, 2013.

[62] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, volume 120, chapter Extraction of periodic components, page 21. Springer, 2013.

[63] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*. Springer Science & Business Media, 2013.

[64] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, chapter 2.3.3 Separability of Components of Time Series. Springer Science & Business Media, 2013.

[65] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*, chapter 2.4.1.4 Elementary Reconstructed Components and w-Correlation Matrix. Springer Science & Business Media, 2013.

[66] BL Gotwols and GB Irani. Optical determination of the phase velocity of short gravity waves. *Journal of Geophysical Research: Oceans*, 85(C7):3964–3970, 1980.

[67] N Haritos. Modelling ocean waves and their effects on offshore structures. In *Australian Earthquake Engineering Society 2010 Conference*, pages 1–7, 2010.

[68] Noriaki Hashimoto. Estimation of directional spectrum using the bayesian approach, and its application to field data analysis. *Rept. PHRI*, 26(5):57–100, 1987.

[69] Hossein Hassani. Singular spectrum analysis: Methodology and comparison. *University Library of Munich, Germany, MPRA Paper*, 5, 01 2007.

[70] Klaus et al Hasselmann. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut. Hydrogr. Z.*, 8:1–95, 01 1973.

[71] Jon C Helton, Jay D Johnson, William L Oberkampf, and Cedric J Sallaberry. Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, 39(6):605–646, 2010.

[72] David Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998.

[73] Leo H Holthuijsen. *Waves in oceanic and coastal waters*, chapter Observation techniques, page 1023. Cambridge university press, 2010.

[74] BA Hooper, B Baxter, C Piotrowski, JZ Williams, and J Dugan. An airborne imaging multispectral polarimeter (AROSS-MSP). In *OCEANS 2009*, pages 1–10. IEEE, 2009.

[75] Brett A Hooper, Becky Van Pelt, JZ Williams, JP Dugan, M Yi, CC Piotrowski, and C Miskey. Airborne spectral polarimeter for ocean wave research. *Journal of Atmospheric and Oceanic Technology*, 32(4):805–815, 2015.

[76] John H. Hubbard and Beverly H. West. *Linear Dynamical Systems*, pages 121–129. Springer New York, New York, NY, 1993.

[77] Steven A Hughes. The TMA shallow-water spectrum description and applications. Technical report, Coastal Engineering Research Center Vicksburg MS, 1984.

[78] Jennifer L Irish, Jennifer M Wozencraft, A Grant Cunningham, and Claudine Giroud. Nonintrusive measurement of ocean waves: Lidar wave gauge. *Journal of Atmospheric and Oceanic Technology*, 23(11):1559–1572, 2006.

[79] Chester Ismay and Albert Y Kim. *Statistical inference via data science: a ModernDive into R and the tidyverse*, chapter 8 Bootstrapping and Confidence Intervals. CRC Press, 2019.

[80] M Isobe, K Kondo, and K Horikawa. Extension of mlm for estimating directional wave spectrum. In *Proc. Symp. on Description and modeling of directional seas*, pages 1–15, 1984.

[81] Bernd Jähne and Klaus S Riemer. Two-dimensional wave number spectra of small-scale water surface waves. *Journal of Geophysical Research: Oceans*, 95(C7):11531–11546, 1990.

[82] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193. International Society for Optics and Photonics, 1997.

[83] Dan Kalman. A singularly valuable decomposition: the svd of a matrix. *The college mathematics journal*, 27(1):2–23, 1996.

[84] Rudolph Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35–45, 1960.

[85] Durga Rao Karanki, S Rahman, Vinh N Dang, and O Zerkak. Epistemic and aleatory uncertainties in integrated deterministic and probabilistic safety assessment: Tradeoff between accuracy and accident simulations. *Reliability Engineering & System Safety*, 162:91–102, 2017.

[86] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54. Cambridge university press, 1997.

[87] D Kiefhaber, S Reith, R Rocholz, and Bernd Jähne. High-speed imaging of short wind waves by shape from refraction. *Journal of the European Optical Society-Rapid publications*, 9, 2014.

[88] Youngjoo Kim and Hyochoong Bang. Introduction to kalman filter and its applications. *Introduction and Implementations of the Kalman Filter*, 1:1–16, 2018.

[89] Ruth King and Rachel McCrea. Chapter 2 - Capture–Recapture Methods and Models: Estimating Population Size. In *Integrated Population Biology and Modeling, Part B*, volume 40 of *Handbook of Statistics*, pages 33–83. Elsevier, 2019.

[90] Blair Kinsman. *Wind waves: their generation and propagation on the ocean surface*, chapter Introduction. Courier Corporation, 1984.

[91] SA Kitaigordskii, VP Krasitskii, and MM Zaslavskii. On phillips' theory of equilibrium range in the spectra of wind-generated gravity waves. *Journal of Physical Oceanography*, 5(3):410–420, 1975.

[92] Oleg Konstantinov. Measurements of sea level, wave height and slopes by analysis of sea surface image sequence from polarization camera. *Pacific Oceanological Institute, Vladivostok Russia*, 2003.

[93] Mariya V Kosnik and Vladimir A Dulov. Extraction of short wind wave spectra from stereo images of the sea surface. *Measurement science and technology*, 22(1):015504, 2010.

[94] H. Lamb. *Hydrodynamics 6th edition*. Cambridge University Press, 1994.

[95] David Lane. *Online statistics education: A multimedia course of study*, chapter 10 Estimation E Confidence Intervals 2 Confidence Interval for the Mean. Association for the Advancement of Computing in Education (AACE), 2003.

[96] RG Langlois. Ship flight deck motion parameters for ensuring safety of helicopter operation. *Indal Technologies Inc.*, 2003.

[97] Nathan JM Laxague, Brian K Haus, Darek Bogucki, and Tamay Özgökmen. Spectral characterization of fine-scale wind waves using shipboard optical polarimetry. *Journal of Geophysical Research: Oceans*, 120(4):3140–3156, 2015.

[98] Tine Lefebvre*, Herman Bruyninckx, and Joris De Schutter. Kalman filters for non-linear systems: a comparison of performance. *International journal of Control*, 77(7):639–653, 2004.

[99] Chuan Li, Martin Shaw, David Pickup, Darren Cosker, Phil Willis, and Peter Hall. Realtime video based water surface approximation. In *2011 Conference for Visual Media Production*, pages 109–117. IEEE, 2011.

[100] Qiulin Liu, Tony Lewis, Yongliang Zhang, and Wanan Sheng. Performance assessment of wave measurements of wave buoys. *International Journal of Marine Energy*, 12:63–76, 2015. Special Issue on Marine Renewables Infrastructure Network.

[101] Yichao Liu, Sunwei Li, Qian Yi, and Daoyi Chen. Wind Profiles and Wave Spectra for Potential Wind Farms in South China Sea. Part II: Wave Spectrum Model. *Energies*, 10(1), 2017.

[102] Antonis Loizou and Jacqueline Christmas. Estimating pixel to metre scale and sea state from remote observations of the ocean surface. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 3513–3516. IEEE, 2018.

[103] Regan M Long, Don Barrick, John L Largier, and Newell Garfield. Wave observations from central California: SeaSonde systems and in situ wave buoys. *Journal of Sensors*, 2011, 2011.

[104] Douglas A Lyon. The discrete fourier transform, part 4: spectral leakage. *Journal of object technology*, 8(7), 2009.

[105] David R Lyzenga, Okey G Nwogu, Robert F Beck, Andrew O'Brien, Joel Johnson, Tony de Paolo, and Eric Terrill. Real-time estimation of ocean wave fields from marine radar data. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3622–3625. IEEE, 2015.

[106] O M. Phillips. On the generation of waves by turbulent winds. *Journal of Fluid Mechanics*, 2:417 – 445, 07 1957.

[107] Gary A. Mastin, Peter A. Watterberg, and John F. Mareda. Fourier synthesis of ocean scenes. *IEEE Comput. Graph. Appl.*, 7(3):16–23, March 1987.

[108] Hermann G Matthies. Quantifying uncertainty: modern computational representation of probability and applications. In *Extreme man-made and natural hazards in dynamics of structures*, pages 105–135. Springer, 2007.

[109] Bruce A McElhoe. An assessment of the navigation and course corrections for a manned flyby of mars or venus. *IEEE Transactions on Aerospace and Electronic Systems*, (4):613–623, 1966.

[110] Chiang C. Mei. Lecture notes on wave propagation, Fall 2004.

[111] Chiara Mellucci, Prathyush P Menon, Christopher Edwards, and Peter Challenor. Oceanic feature boundary mapping with an autonomous underwater glider. In *2018 Annual American Control Conference (ACC)*, pages 5338–5343. IEEE, 2018.

[112] Edward C. Monahan. Oceanic whitecaps. *Journal of Physical Oceanography*, 1(2):139–144, 1971.

[113] FM Monaldo and RS Kasevich. Daylight imagery of ocean surface waves for wave spectra. *Journal of Physical Oceanography*, 11(2):272–283, 1981.

[114] Lionel Moskowitz. Estimates of the power spectrums for fully developed seas for wind speeds of 20 to 40 knots. *Journal of geophysical research*, 69(24):5161–5179, 1964.

[115] Joshua Mullins, You Ling, Sankaran Mahadevan, Lin Sun, and Alejandro Strachan. Separation of aleatory and epistemic uncertainty in probabilistic model validation. *Reliability Engineering & System Safety*, 147:49–59, 2016.

[116] National Data Buoy Center. Station 41013 Frying Pan Shoals, NC. Available at `https://ndbc.noaa.gov/station_page.php?station=41013`.

[117] J.C. Nieto-Borge, K. Hessner, P. Jarabo-Amores, and D. de la Mata-Moya. Signal-to-noise ratio analysis to estimate ocean wave heights from X-band marine radar image time series. *IET Radar, Sonar and Navigation*, 2:35–41(6), February 2008.

[118] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[119] Michel K Ochi. *Ocean waves: the stochastic approach*, chapter Spectral analysis. Number 6. Cambridge University Press, 2005.

[120] MK Ochi. A series of JONSWAP wave spectra for offshore structure design. In *Proceedings of the Second International Conference on the Behaviour of Off-Shore Structures, held at Imperial College, London, England*, 1979.

[121] Andrés F Osorio, Sebastian Montoya-Vargas, Cesar A Cartagena, Jairo Espinosa, Alejandro Orfila, and Christian Winter. Virtual buoy: A video-based approach for measuring near-shore wave peak period. *Computers & Geosciences*, 133:104302, 2019.

[122] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[123] B Pérez, E Álvarez Fanjul, S Pérez, M de Alfonso, and J Vela. Use of tide gauge data in operational oceanography and sea level hazard warning systems. *Journal of Operational Oceanography*, 6(2):1–18, 2013.

[124] Oliver M Phillips. The equilibrium range in the spectrum of wind-generated waves. *Journal of Fluid Mechanics*, 4(4):426–434, 1958.

[125] Owen M Phillips. *The dynamics of the upper ocean*. CUP Archive, 1966.

[126] Owen M Phillips. *The dynamics of the upper ocean (2nd ed.)*, chapter 3.2. Cambridge University Press, 1977.

[127] Willard J. Pierson Jr. and Lionel Moskowitz. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *Journal of Geophysical Research*, 69(24):5181–5190, 1964.

[128] Tsai Ping-Sing and Mubarak Shah. Shape from shading using linear approximation. *Image and Vision computing*, 12(8):487–498, 1994.

[129] A.J. Plater and J.R. Kirby. 3.03 - sea-level change and coastal geomorphic response. In Eric Wolanski and Donald McLusky, editors, *Treatise on Estuarine and Coastal Science*, pages 39–72. Academic Press, Waltham, 2011.

[130] Lawrence Rabiner and Biinghwang Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[131] K Reichert, K Hessner, JC Nieto Borge, J Dittmer, et al. Wamos ii: A radar based wave and current monitoring system. In *The Ninth International Offshore and Polar Engineering Conference*. International Society of Offshore and Polar Engineers, 1999.

[132] Maria Isabel Ribeiro. Kalman and extended kalman filters: Concept, derivation and properties. *Institute for Systems and Robotics*, 43, 2004.

[133] Kent Rosser and Javaan Chahl. Wave scale, speed and direction from airborne video of maritime scene. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7. IEEE, 2019.

[134] Sam Roweis and Zoubin Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, 1999.

[135] Sujit K Sahu and Peter Challenor. A space-time model for joint modeling of ocean temperature and salinity levels as measured by Argo floats. *Environmetrics: The official journal of the International Environmetrics Society*, 19(5):509–528, 2008.

[136] John Robert Schott. *Fundamentals of polarimetric remote sensing*, volume 81. Spie Press, 2009.

[137] Christian M Senet, Jorg Seemann, and Friedwart Ziemer. Hydrographic parameter maps deduced from ccd image sequences of the water surface supplemented by in-situ wave gauges. In *IEEE 2000 International Geoscience and Remote Sensing Symposium*, volume 2, pages 843–846. IEEE, 2000.

[138] CM Senet, J Seemann, and F Ziemer. Dispersive surface classification: Local analysis of optical image sequences of the water surface to determine hydrographic parameter maps. In *OCEANS 2000*, volume 3, pages 1769–1774. IEEE, 2000.

[139] R Serway and C Vuille. College physics singapore: Cengage learning. 2012.

[140] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[141] Omar H Shemdin, H Minh Tran, and SC Wu. Directional measurement of short ocean waves with stereophotography. *Journal of Geophysical Research: Oceans*, 93(C11):13891–13901, 1988.

[142] Martin Skaldebø, Albert Sans Muntadas, and Ingrid Schjølberg. Transfer learning in underwater operations. In *OCEANS 2019-Marseille*, pages 1–8. IEEE, 2019.

[143] Gerald L Smith, Stanley F Schmidt, and Leonard A McGee. *Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle*. National Aeronautics and Space Administration, 1962.

[144] Arnold Sommerfeld. *Mechanics of deformable bodies: Lectures on Theoretical Physics*, volume 2. Elsevier, 2016.

[145] Concetto Spampinato, Yun-Heh Chen-Burger, Gayathri Nadarajan, and Robert B Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP (2)*, 2008(514-519):1, 2008.

[146] L. Spencer, M. Shah, and R. K. Guha. Determining scale and sea state from water video. *IEEE Transactions on Image Processing*, 15(6):1525–1535, June 2006.

[147] Lisa Spencer and Mubarak Shah. Water video analysis. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 4, pages 2705–2708. IEEE, 2004.

[148] K. Stephan. Kinetic and potential energy in the first law of thermodynamics. *Wärme - und Stoffübertragung*, 8(1):1–10, Apr 1975.

[149] Gilbert W Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566, 1993.

[150] Robert Stewart. *Introduction To Physical Oceanography*. 01 2008.

[151] Denzil Stilwell Jr. Directional energy spectra of the sea from photographs. *Journal of Geophysical Research*, 74(8):1974–1986, 1969.

[152] Denzil Stilwell Jr and R O. Pilon. Directional spectra of surface waves from photographs. *Journal of Geophysical Research*, 79:1277–1284, 03 1974.

[153] Hilary F Stockdon and Rob A Holman. Estimation of wave phase speed and nearshore bathymetry from video imagery. *Journal of Geophysical Research: Oceans*, 105(C9):22015–22033, 2000.

[154] Yasuhiro Sugimori. A study of the application of the holographic method to the determination of the directional spectrum of ocean waves. *Deep Sea Research and Oceanographic Abstracts*, 22(5):339 – 350, 1975.

[155] Edward J Tarbuck, Frederick K Lutgens, and Dennis Tasa. *Earth science*, chapter The dynamic ocean. Prentice Hall New Jersey, 1997.

[156] Ming Yang Teng, Ruby Mehrubeoglu, Scott A King, Kirk Cammarata, and James Simons. Investigation of epifauna coverage on seagrass blades using spatial and spectral analysis of hyperspectral images. In *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE, 2013.

[157] Gabriel A Terejanu et al. Extended kalman filter tutorial. *University at Buffalo*, 2008.

[158] Richard E Thomson and William J Emery. *Data analysis methods in physical oceanography*, chapter Data acquisition and recording. Newnes, 2014.

[159] Harold V Thurman. *Essentials of oceanography*, chapter Waves and water dynamics, pages 236–264. Pearson, 2019.

[160] Ryota Tsubaki and Ichiro Fujita. Stereoscopic measurement of a fluctuating free surface with discontinuities. *Measurement Science and Technology*, 16(10):1894, 2005.

[161] Malcolm John Tucker and Edward G Pitt. *Waves in ocean engineering*, pages 35–36. Number Volume 5. 2001.

[162] A_P Varga and RK Moore. Hidden markov model decomposition of speech and noise. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 845–848. IEEE, 1990.

[163] Robert Vautard, Pascal Yiou, and Michael Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1-4):95–126, 1992.

[164] K Viriyakijja and C Chinnarasri. Wave flume measurement using image analysis. *Aquatic Procedia*, 4:522–531, 2015.

[165] Jeffrey Walker. *Elements of Statistical Modeling for Experimental Biology*, chapter 5 Variability and Uncertainty. 2018.

[166] Jenny Walker, Takaki Yamada, Adam Prugel-Bennett, and Blair Thornton. The effect of physics-based corrections and data augmentation on transfer learning for segmentation of benthic imagery. In *2019 IEEE Underwater Technology (UT)*, pages 1–8. IEEE, 2019.

[167] Justin Wanek and Chin Wu. Automated trinocular stereo imaging system for three-dimensional surface wave measurements. *Ocean Engineering*, 33:723–747, 04 2006.

[168] Chau-Chang Wang, Po-Chi Chen, and Chen-Yi Liao. Application of ccd cameras as a versatile measurement tool for flume tank. *Ocean engineering*, 42:71–82, 2012.

[169] Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on acoustics, speech, and signal processing*, 33(2):387–392, 1985.

[170] Graham H White. Basics of estimating measurement uncertainty. *The Clinical Biochemist Reviews*, 29(Suppl 1):S53, 2008.

[171] Christian E Willert and Morteza Gharib. Digital particle image velocimetry. *Experiments in fluids*, 10(4):181–193, 1991.

[172] TY Wu. Dispersion of water waves. *Journal of Geophysical Research*, 68(4):1195–1197, 1963.

[173] Jingjin Xie and Lei Zuo. Dynamics and control of ocean wave energy converters. *International Journal of Dynamics and Control*, 1(3):262–276, 2013.

[174] Takaki Yamada, Adam Prügel-Bennett, and Blair Thornton. Learning features from georeferenced seafloor imagery with location guided autoencoders. *Journal of Field Robotics*, 38(1):52–67, 2021.

[175] Jeseon Yoo, Hermann Fritz, Kevin Haas, Paul Work, and Christopher Barnes. Wave property estimation using linear feature extraction from nearshore wave images. In *The 5th International Symposium on Ocean Wave Measurement and Analysis*, pages 3–7, 2005.

[176] Christopher J Zappa, Michael L Banner, Howard Schultz, Andres Corrada-Emmanuel, Lawrence B Wolff, and Jacob Yalcin. Retrieval of short ocean wave slope using polarimetric imaging. *Measurement Science and Technology*, 19(5):055503, 2008.

[177] X Zhao, Roger Xu, and Chiman Kwan. Ship-motion prediction: algorithms and simulation results. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages V–125. IEEE, 2004.

[178] Muhammad Zikra, Noriaki Hashimoto, Masaru Yamashiro, Masaki Yokota, and Kojiro Suzuki. Analysis of directional wave spectra in shallow water areas using video image data. *Coastal Engineering Journal*, 54(3):1250020–1, 2012.

[179] Muhammad Zikra, Noriaki Hashimoto, Masaru Yamashiro, Masaki Yokota, and Kojiro Suzuki. Application of video images for monitoring coastal zone in Hasaki beach, Japan. *COASTAL ENGINEERING*, page 2, 2012.

[180] Martin Zurowietz, Daniel Langenkämper, Brett Hosking, Henry A Ruhl, and Tim W Nattkemper. Maia—a machine learning assisted image annotation method for environmental monitoring and exploration. *PloS one*, 13(11):e0207498, 2018.