



Sea State Estimation from Uncalibrated, Monoscopic Video

Antonis Loizou¹ · Jacqueline Christmas¹

Received: 29 January 2021 / Accepted: 29 May 2021
© The Author(s) 2021

Abstract

Video of the ocean surface is used as a means for estimating the sea state. Time series of pixel intensity values are given as input to a method that uses the Kalman filter and the least squares approximate solution for estimating the 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. The results are tested against two sets of video data, and buoy measurements in both cases are solely used for indicating the true state. For significant wave height values between 0.5 and 3.6 m, the maximum observed value of root mean square error is 0.37 m and of mean absolute percentage error 16%.

Keywords Ocean video · Sea state · Filtering algorithms · Kalman filter · Amplitude scaling · Geophysical image processing

Introduction

Video of the ocean surface in real environments is investigated as a means of estimating the sea state. A methodology is proposed that estimates the 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.

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 [9, 14]. Structures are vulnerable to damage caused by the unpredictable nature of the seas, and some marine operations have to be performed in certain desirable (benign) sea states. For these operations, it is important to have on-board real time information of the state of the sea.

Any system design is an optimisation process [5]. In the case of designing and building harbours and other offshore structures, the effects of waves are a primary constraint [9].

This is why knowledge of the state of the ocean in an area over preferably 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 defined. These procedures include restrictions associated with the state of the sea, wind speed and direction [17]. 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 [4] and for improving the efficiency of wave energy converters [1, 29]. 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.

Information about the sea state is in most cases achieved with the use of in situ devices, such as wave buoys. Remote sensing methods have also been proposed for obtaining the sea state, for example, using satellite or radar images [15], or stereo images [12, 28]. In the case of stereo, detailed elevation information becomes available.

The problem that is investigated in this work is, with the use of a single uncalibrated camera in real environments,

✉ Antonis Loizou
al553@exeter.ac.uk

Jacqueline Christmas
j.t.christmas@exeter.ac.uk

¹ Department of Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Streatham campus, Rennes Drive, Exeter EX4 4RN, UK

whether it is possible to estimate useful sea state information. In early work based on photographs [26], the geometric configuration of the scene and theory of physics were used to answer this question. In another category [8], the sun's glitter pattern was used. In this work we focus on the use of information that is available from the video without the requirement of glitter pattern being present. The science of data processing and modelling has advanced greatly since these early works, and it would be logical to now have better means to answer the same questions.

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 [7], 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. Efforts are made to minimise the effects of noise with spatial and temporal smoothing techniques.

The work by [31] utilises videos of the ocean in real environments. The authors propose a methodology that uses the extended maximum likelihood method (EMLM), the Bayesian directional method (BDM) and a windowing process to estimate the directional energy spectrum of the ocean in shallow water from a video camera. Configurations of arrays of pixel intensity values are used in the input. In another published work, the same authors [32] 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 [7], 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 to only the customised ocean model SWAN.

Time series of pixel intensity values from coastal surveillance systems in the nearshore are utilised by [20] 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 main challenges of this work are the identification of the low cut-off frequencies of the filter and the isolation of the information of surface variation from environmental brightness fluctuations. The authors observe that the applicability of the technique is not valid for low and high sea states, but rather only medium sea states. In [27] the authors propose a methodology 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 to get the phase speed, which is then used with the dispersion relation to get the bathymetry of the scene.

The information of wave crests is used in [30] to estimate the dominant period from video in the surf zone. First, low-pass filtering and backward frame differencing are performed to remove noise caused by foam. Then, a thresholding process 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 particle image velocimetry (PIV) algorithm is used in [13] to get the phase speed information from videos of the nearshore ocean. The foam noise is removed with the same method as in [30], the backward frame differencing method. 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 [30]. 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.

In the case of using a single camera, and without taking into account physical phenomena associated with the scene geometry, such as refraction of light, the problem to solve is the correct distinction of the information from video that is caused by the movement of the ocean from all other irrelevant information. In Spencer et al. [25], the authors utilise the phase speed information from pixel time series and the Phillips ocean energy spectrum for acquiring information about the pixel to metre scale of the video and the sea state. In one of our earlier works [18], we extended Spencer et al.'s work using radar images and proposed a method for getting the sea state from video that focuses solely on the dominant waves present and the spatial information. In [23], the authors investigated and proposed improvements to Spencer et al.'s work with the use of airborne video data. The dominant frequency is determined as the centroid of a selection of frequencies near the peak frequency obtained with an iterative thresholding process.

This work uses the same type of input data, namely time series of pixel intensity values, and will aim to provide estimations of the sea state. First, an uncalibrated amplitude spectrum is formed with the use of a method based on the Kalman filter and least squares approximation. The configuration of the filtering algorithm aims to distinguish the ocean movement elements from all other irrelevant elements from video. Then,

a scaling process is proposed for transferring the uncalibrated amplitude spectrum to metres. From this, the significant wave height is estimated.

Ocean Theory

The dominant frequency is the sinusoidal element with the highest energy in the ocean energy spectrum. Two other components from ocean theory are introduced briefly here; these are used later in the methodology for scaling the uncalibrated amplitude spectrum from ocean video (see “Scaling to metres”). For full details see Kinsman [16].

Energy to Amplitude in Ocean Waves

In general, the energy of a wave is directly proportional to the square of its amplitude. In ocean theory, in [6] the authors used the first law of hydrodynamics and represented ocean wave energy as the sum of kinetic and potential energy. From their findings, a relationship can be formed that connects the total energy, E , of a wave to its amplitude, A :

$$E = \frac{1}{2} \rho g A^2, \tag{1}$$

where ρ is the water density and g is the gravitation constant.

Pierson–Moskowitz Spectrum

The Pierson–Moskowitz spectrum is an empirical energy spectrum of the ocean formed from data of shipborne wave recorders on ships in the North Atlantic ocean [22]. The energy spectrum of the ocean S according to this work is equal to:

$$S(\omega) = \frac{\gamma g^2}{\omega^5} \exp \left\{ -\beta \left(\frac{g/U}{\omega} \right)^4 \right\}, \tag{2}$$

where ω is the angular frequency, $\gamma = 8.1 \times 10^{-3}$, $\beta = 0.74$ and U is the wind speed at a height of 19.5 m above the ocean surface.

Methodology

The input of the presented methodology is video of the ocean surface from a single camera. Figures 1c and d show example stills from such video. First, any camera motion is stabilised. Then, if the camera is placed at an angle from the ocean surface, perspective transformation is performed to fix the scale issue (see “Preprocessing”). A single row of pixels from each frame is used for computational efficiency; perspective transformation allows all pixels to be used if desired.

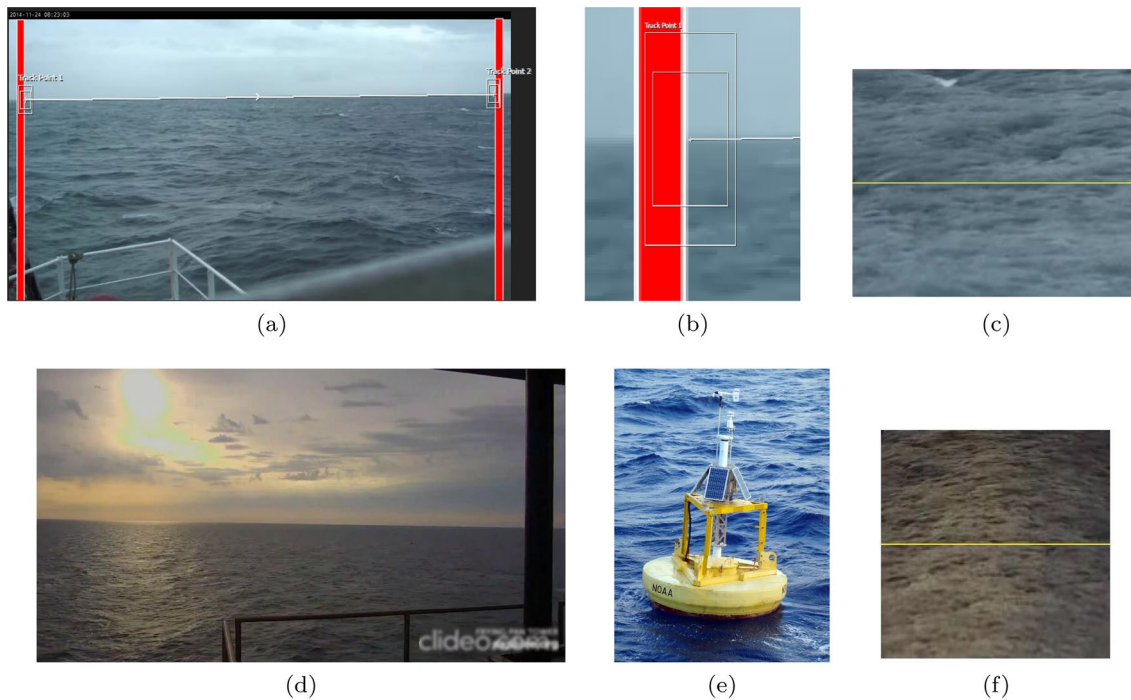


Fig. 1 Shipborne and tower video data. **a** Ship video with horizon stabilisation. **b** Ship video left tracking point. **c** Ship video after perspective transformation. **d** Tower video. **e** Buoy station 41013. **f**

Tower video after perspective transformation. Lines in **c** and **f** depict the pixel locations used in the experimental results

Table 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

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 “Kalman Filter” and the least squares approximate solution (“Least squares approximate solution”). Then, an averaged shape is formed using the results of multiple time series (“Wave height shape”), and scaling is performed to transfer this shape to metres (“Scaling to metres”). From the scaled to metres video amplitude spectrum, the significant wave height is estimated (“Significant wave height”).

Kalman Filter

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 done separately for each frequency.

The frequency domain of the wave spectrum 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 [24]. The Kalman filter as specified in the following text is performed for each frequency 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. This is not to imply that the algorithm correctly captures the true state. The environment 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:

$$x_t^* = A \sin(\omega t + \phi) + \epsilon_t, \tag{3}$$

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 (assumed to be zero-mean Gaussian distributed). The true signal is defined as:

$$x_t = A \sin(\omega t + \phi), \tag{4}$$

where x_t is the pixel intensity at time t , 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:

$$\dot{x}_t = A\omega \cos(\omega t + \phi) \tag{5}$$

The second derivative of the signal is equal to:

$$\ddot{x}_t = -A\omega^2 \sin(\omega t + \phi), \tag{6}$$

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

$$\ddot{x} = -\omega^2 x. \tag{7}$$

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 . And 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:

$$\begin{pmatrix} \dot{x} \\ \ddot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} \tag{8}$$

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

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \tag{9}$$

The process noise could be used 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. “Appendix 1” provides further information for solving the Kalman filter with the environment definition described here.

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:

$$x_i = A \sin(\omega t_i + \phi) \quad i = 0, 1, 2, \dots, \quad (10)$$

and the amplitude A is to be calculated. With three position estimations (here labelled as $i = 0, 1, 2$) the vector form of their differences $\Delta \mathbf{x}$ can be expressed as: $\Delta \mathbf{x} = \mathbf{J} \cdot \mathbf{y}$

$$\begin{pmatrix} x_1 - x_0 \\ x_2 - x_1 \\ A \cos \phi \\ A \sin \phi \end{pmatrix} = \begin{pmatrix} \sin \omega t_1 - \sin \omega t_0 & \cos \omega t_1 - \cos \omega t_0 \\ \sin \omega t_2 - \sin \omega t_1 & \cos \omega t_2 - \cos \omega t_1 \\ A \cos \phi \\ A \sin \phi \end{pmatrix} \quad (11)$$

where $y_0 = A \cos \phi$ and $y_1 = A \sin \phi$. The amplitude is then equal to $A = \sqrt{y_0^2 + y_1^2}$.

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

$$\begin{pmatrix} x_1 - x_0 \\ x_2 - x_1 \\ x_3 - x_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sin \omega t_1 - \sin \omega t_0 & \cos \omega t_1 - \cos \omega t_0 \\ \sin \omega t_2 - \sin \omega t_1 & \cos \omega t_2 - \cos \omega t_1 \\ \sin \omega t_3 - \sin \omega t_2 & \cos \omega t_3 - \cos \omega t_2 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} A \cos \phi \\ A \sin \phi \end{pmatrix} \quad (12)$$

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

$$\mathbf{y} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \Delta \mathbf{x}. \quad (13)$$

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.

Wave Height Shape

Once the Kalman filter has been applied to each frequency in the defined domain for one pixel intensity time series, and

the least squares approximate solution used for estimating the amplitude, the results from all frequencies are combined to give an amplitude spectrum. If multiple pixel intensity time series are selected from each video, multiple shapes are averaged to extract a final averaged video amplitude spectrum. Example of such shapes are given in Fig. 4b, e and h, with the difference that until this step, the shape is given in an uncalibrated metric. Now useful information about the sea state will be derived from this shape. In the next step, the shape will be used to scale the results into metres (“Scaling to metres”) and then to estimate the significant wave height (“Significant wave height”).

Scaling to Metres

Based on the uncalibrated, averaged amplitude spectrum obtained from the previous steps, the amplitude multiplier variable, α , is introduced here, 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:

$$\alpha = \frac{a_{pm}}{a_u} \quad (14)$$

where a_{pm} is an amplitude value in metres from the Pierson–Moskowitz spectrum (“Pierson–Moskowitz spectrum”) and a_u is a value from the uncalibrated spectrum.

The key in this process is the calculation of a_u ; thereafter, 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’s dominant frequency (“Ocean theory”). 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:

$$\xi = \frac{a_p - a_c}{a_p} \quad (15)$$

where a_p is the peak amplitude and a_c is the current amplitude. Then a threshold of 30% is applied: if ξ is below this threshold then the frequency associated with the current amplitude is selected; if it exceeds the threshold then the procedure ends. The amplitudes of the selected frequencies are then averaged to give the value of a_u in Eq. (13). 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.

To continue the process, the average of the selected frequencies is then given as an input to Eq. (2) 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 Eq. (1).

Having both a_{pm} and a_u of Eq. (13), the peak multiplier α is calculated, and is used for transferring the uncalibrated amplitude spectrum to metres. Examples of the final scaled video amplitude spectrum are given in Fig. 4b, e and h.

Significant Wave Height

The significant wave height is a key descriptor of the sea state, and is found from the video amplitude spectrum in metres (i.e. after the scaling procedure) as the average wave height (trough to peak) of the highest one third of the sea waves.

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

Experimental Results

The proposed methodology is demonstrated with real video and with the presence of in situ buoy devices. The first set of videos is taken from a moving, shipborne camera (“[Ship video data](#)”) and the second set from live video footage from a tower, i.e. a fixed location (“[Tower video data](#)”).

In terms of the validation process, each video estimation is validated against the corresponding buoy values of the significant wave height. No videos were excluded from the data collection process. The definition of the accuracy measure metrics used is provided in “[Appendix 2](#)”.

Sample Size

The sample size of the first data set (shipborne video) comprises of 71 1-min videos captured in the time span of 2–3 h of the same day. From these, 50 videos have corresponding values of significant wave height from 2 buoys, indicating the true sea state for validation. The frame rate of the shipborne videos is 15 frames per second. The significant wave height from the two buoys is between 3.1 and 3.4 m.

The sample size of the second data set comprises of 45 one-minute videos captured from a tower in the time frame of 10 months, each video capturing the ocean surface at a different date. Each tower video has a corresponding value of significant wave height for validation from a nearby buoy. The frame rate of the tower video is 30 frames per second. The buoy significant wave height values range between 0.5 and 3.6 m.

Ship Video 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 was isolated and split into one minute videos. During the experiments, two buoys recorded concurrently the wave elevation at a nearby location.

Preprocessing

A preprocessing step was performed for the ship data, to overcome the ship’s rotational movement and the perspective scale problem. To stabilize the ship video, the software Adobe After Effects [2] was used. The rotational tracker was used 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 Fig. 1a and the tracking points are selected as in Fig. 1b.

Next, for fixing the scale problem, trapezoidal shaped areas are selected to be perpendicular to the horizon and perspective transformation is performed. Figure 1c shows the result after this process. From a video, image sequences in the form of image Fig. 1c are given as output from the preprocessing step. A single row of pixel locations is used for computational efficiency. Perspective transformation enables us to use all pixels if this is desired.

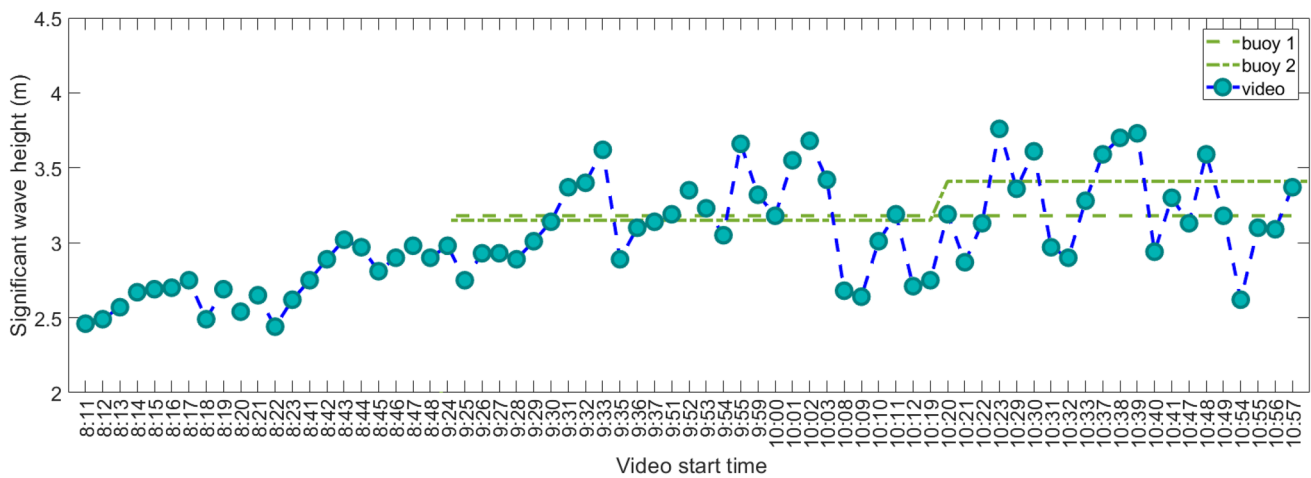


Fig. 2 Shipborne video showing stability of video estimations of a similar sea state and correlation of the significant wave height estimation between video and buoys. The buoys were measuring at 9:15

a.m. It is a well known assumption that the sea state will not change significantly in the period shown in this graph [16]

Main Methodology

The shipborne data results are presented in Fig. 2. 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 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 for a sea state of significant wave height 3.1 m, as well as MAPE of 9.86% indicate that the methodology is promising. The small positive value of MPE 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 video method will be examined for a variety of sea states.

Tower Video Data

To examine the performance of the methodology for a variety of sea states, video captured from the Frying Pan Ocean tower located 85ft above the Atlantic ocean was used. This is a lighthouse 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 [11] that captures the ocean surface. Sea state information is also

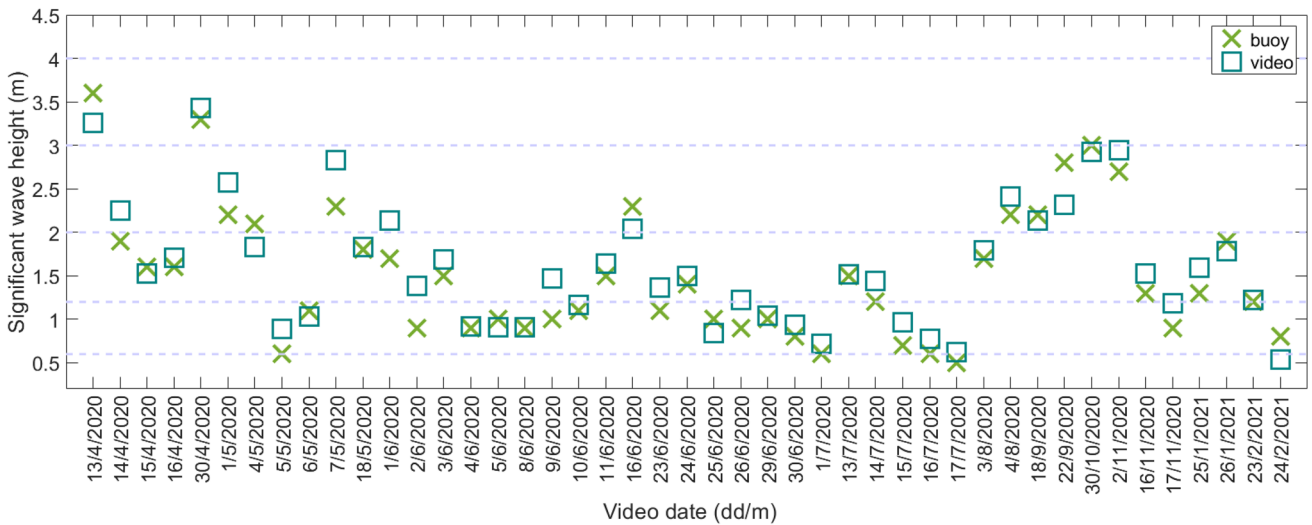
available in the tower’s website [10]. A buoy device ‘Station 41013’, owned and maintained by the National Data Buoy Center [19], provides detailed sea state information, and is located close to the tower. Figure 1d shows an example frame from the tower video.

Although the tower camera is panning from side to side most of the time, showing a panoramic view of the area, for some time intervals the camera remains still, and one minute videos were captured throughout many days. In cases of high local wind being present, the shaking of the camera was fixed with the stabilisation features of Adobe After Effects. In a similar manner as before, perspective transformation is performed and the result is shown in Fig. 1d, which is what is given in the input of the methodology.

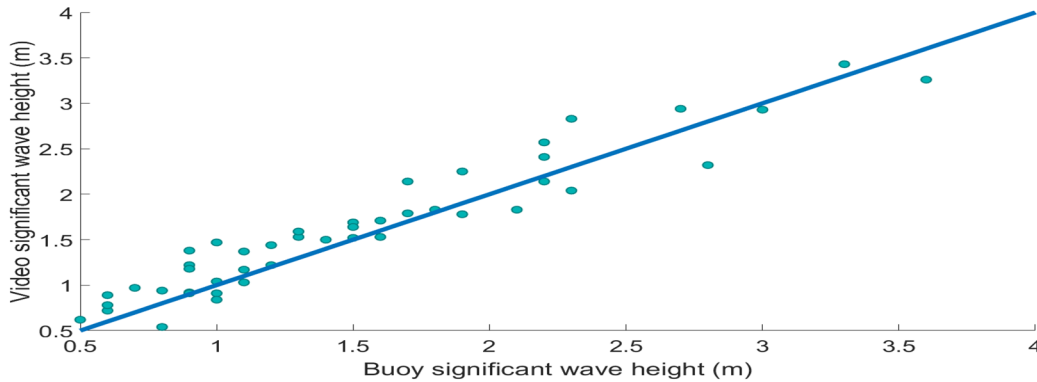
The tower video results are presented in Fig. 3. In each of these dates, one video is used, the significant wave height is estimated and is compared with the buoy’s significant wave height of the same time. As the videos are captured in different dates, the proposed method is tested in a variety of sea states.

From Fig. 3b 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 Fig. 3a it can be observed that the video methodology works across a significant time period in different sea states.

Figure 4 shows the buoy energy spectrum from three days with different sea states and the scaled video amplitude spectra. From these it can be observed that the shape of the buoy energy spectrum is similar to the video energy spectrum, converted from the video amplitude spectrum. Additionally,



(a)



(b)

Fig. 3 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

as the sea state decreases, the peak from both the buoy and video energy spectra is located in higher frequencies.

Table 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 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.

The value of the MAPE from the tower video (16%) is larger than the value of the error from the shipborne video (9%). Possible reason for this increase of error with the tower video data set is the nature of the MAPE

metric. Specifically, this metric gives higher error values as the estimated-true pairs decrease in value. Since with the tower video the methodology is tested for a range of sea states that also includes lower states (e.g. significant wave height of 0.5 m), it is expected that the MAPE value would be larger. Although a 16% MAPE still shows that the methodology performs well, the values from the MAE metric (0.2 m) and the RMSE metric (0.24 m) are more representative of the good performance of the methodology, as they are not affected by the decrease of values of the true-estimated pairs.

With values of significant wave height between 0.5 and 3.6 m, the results are very promising. With the ship data it was possible to check the variability of the estimations across multiple videos with approximately the same sea state, but this option is not available with the tower video due to the nature of the way the data are recorded.

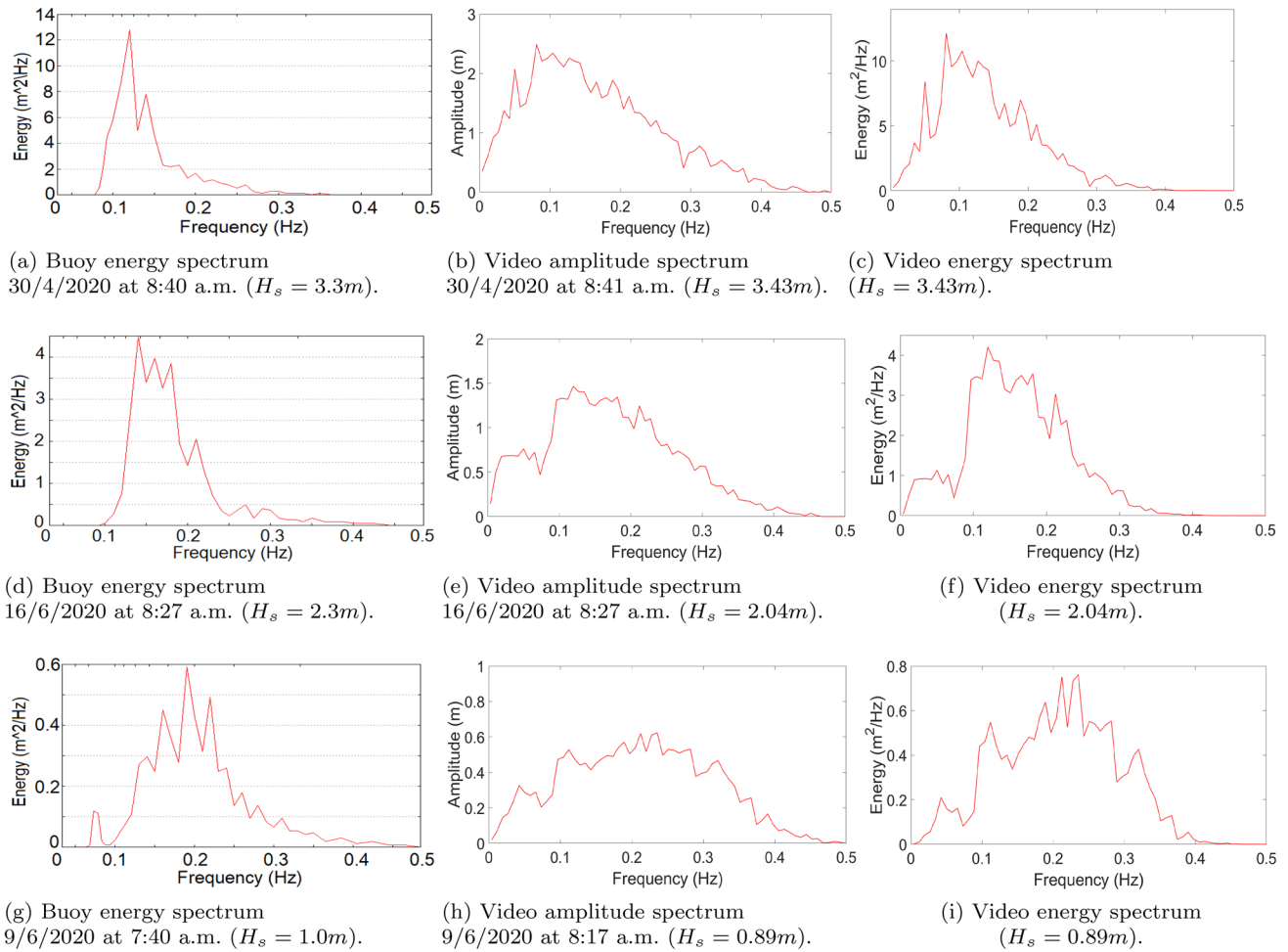


Fig. 4 Tower video energy and amplitude spectra comparison with buoy energy spectra from [19], showing a correspondence between the peaks of video and buoy across a variety of sea states

Complexity

In terms of time complexity, the runtime duration increases linearly with the increase of number of frames and the number of pixel locations utilised. That is, the present methodology runs in linear time. In terms of space complexity, the memory allocation increases in a similar manner.

Table 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

Experimental Details

In terms of the hyperparameter configuration, the results shown here are retrieved with a standard deviation of the measurement noise of the Kalman filter equal to 1. For sensor data, the value for this hyperparameter is usually selected arbitrarily by users to be equal to values near 0 [21]. The range of values considered in this work for this hyperparameter is between 0.5 and 20. Larger values were not considered as the noise would be too large in the context of this work. From these values, the results were found to be consistently the same. That is, the video estimations are not sensitive to the selection of the value of this hyperparameter in the specified range.

Very large values are set on the main diagonal of the initial covariance matrix of the Kalman filter to reflect the uncertainty of the initial state values (the variance of each variable). The process noise could be used 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.

Experiments were ran in Matlab R2017b on a standard office laptop with Windows 10 and Intel Core i5 processor. With these specifications, the average runtime is approximately 5 min.

Conclusion

The applicability of the method for higher sea states is not known at this point. It is likely that it will be adversely affected by, for example, white-capping, which is common in high sea states. However, being able to track sea state information continuously, and being able to identify when the high sea states are being observed, is very useful information for deciding whether key maritime operations can safely be completed.

The present work has the advantage of forming the shape of the ocean video amplitude spectrum that is localised to each video in the input and does not use the shape of a generalised empirical ocean spectrum for the sea state estimation. The empirical energy spectrum is used only for calibrating the specific ocean video amplitude spectrum to metres. Since the shape used is not a general one but specific formed from the video information, it is expected to provide more accurate results. It is expected that a localised shape of spectrum formed from the ocean surface information would describe more accurately the ocean than a generalised spectrum. Localised energy spectrum of the ocean can be measured reliably with wave buoys. However, these devices require funds for their acquisition, maintenance and deployment. Additionally, extreme weather conditions may cause damages to these devices. Cameras for capturing ocean video require less funds and their deployment is a more straightforward task.

This work also has the advantage of using ocean theory instead of in situ devices (in comparison to e.g. [7]) for calibration and it focuses on video in real environments, expanding the possibilities for practical utilisation of the proposed methodology. The use of in situ devices (such as wave gauges) for calibration introduces some challenges, such as the identification of the pixel locations where these devices are located in the video, since the video is not capturing the ocean at the same time as the in situ devices.

The dependence of the calibration process on the localised ocean video amplitude spectrum can be considered a drawback in cases where the estimated shape does not reflect the actual ocean amplitude spectrum. From the experimental results of this work however we observe that the video

estimations are close to the buoy measurements, which is interpreted as the calibration process working correctly and the shape of the ocean video amplitude spectrum being close enough to the shape of the true ocean.

As mentioned in the introduction, the estimation of the sea state is important for the construction of offshore structures, sea vessels and platforms, and the execution of critical maritime operations. It is also useful for improving the performance of wave energy converters. Thus, oceanographers and other scientists in these fields are interested in the acquisition of the sea state from the ocean surface in specific locations. Compared to the in situ measurement of the ocean (with e.g. wave buoys), the use of remote sensing has the advantage of measuring the ocean in a non-intrusive way. Additionally, the utilisation of a monoscopic camera for measuring the ocean surface requires less funds for acquisition and maintenance than buoys.

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 of Beaufort 10 (very high sea state) 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 of Beaufort 1 (very low sea state) 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 technique can be used as a useful guide.

Some of the practical applications mentioned (e.g. building of offshore structures) require a continuous measurement of the sea state over extended periods of time. In these cases, a camera measuring the ocean surface can be installed on a more permanent basis and methodologies such as the one presented in this work can be used for translating the streaming information from video to an approximation of the sea state.

This work proposes the modelling of the ocean surface in real environments from video with the ocean video amplitude spectrum and the use of this structure with ocean theory to get the significant wave height. Future work will investigate improvements to the development of the ocean video amplitude spectrum so that the similarity between video spectrum and true ocean spectrum is increased. Future work will also investigate possible improvements in the calibration process. Specifically, different processes than the one presented here (that uses the position of the peak and the distance between the amplitude of the peak from the amplitude of waves with lower amplitude) are to be investigated with the goal of acquiring smaller values of error metrics. Although already the values of the error metrics are in a satisfactory degree showing that the methodology is doing

something meaningful, there are always opportunities for improvement.

Future work will investigate the applicability of the methodology for higher sea states. Due to the fact that this work utilises videos of the ocean surface in real environments, this requires the capture of the ocean with video in these higher sea states, while also having concurrent knowledge of the sea state from in situ devices to validate the estimations.

Future work will also include the installation of a camera in a stable location (e.g. a pier) and the deployment of a buoy device in a nearby location to acquire additional video data. This will allow a more thorough testing of the methodology with an additional data set, where multiple consecutive short videos (e.g. 60 one-minute videos in a time frame of an hour) can be used for testing the variability of the results from multiple videos of approximately the same sea state. Although this is already achieved with the use of the shipborne data, it would be interesting to test that this behaviour of providing good approximations for consecutive videos of the same sea state is applicable for lower and higher sea states.

Appendix 1: Kalman filter solution

This section provides a further elaboration and solution to the Kalman filter algorithm with the environment definition described in “[Kalman Filter](#)”.

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:

$$\Phi_k = \begin{pmatrix} \cos \omega \Delta t & \frac{\sin \omega \Delta t}{\omega} \\ -\omega \sin \omega \Delta t & \cos \omega \Delta t \end{pmatrix}. \quad (16)$$

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

$$H = (1 \ 0). \quad (17)$$

The measurements are related to the states according to:

$$z_k = Hx_k + \zeta, \quad (18)$$

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

The noise matrices are equal to:

$$W = E[\epsilon \epsilon^T] \quad (19)$$

$$V = E[\zeta \zeta^T], \quad (20)$$

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

The process noise could be used 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. [3] provides further general information for solving the Kalman filter with the environment definition described here.

Appendix 2: Estimation Accuracy Measures

The following measures are used in the experimental results (“[Experimental results](#)”) for the evaluation of the accuracy of the proposed methodology against the sea state measurements of buoy devices.

The buoy data are considered to indicate the true sea state. 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.

Mean Absolute Error (MAE)

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_t - y_p|, \quad (21)$$

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.

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:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_t - y_p)^2}. \quad (22)$$

Mean Absolute Percentage Error (MAPE)

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

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_t - y_p}{y_t} \right|. \quad (23)$$

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.

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. This is useful for showing whether the approximations are overestimating (negative error) or underestimating (positive error). The equation is:

$$\text{MPE} = \frac{100}{n} \sum_{i=1}^n \frac{y_t - y_p}{y_t}. \quad (24)$$

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.

Funding The authors received no specific funding for this work.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aderinto T, Li H. Ocean wave energy converters: status and challenges. *Energies*. 2018;11(5):1250.
- Adobe: After Effects CC 2019. <https://adobe.com/products/after-effects> (Accessed 18 Feb 2021).
- Alonzo K. A 3D state space formulation of a navigation Kalman filter for autonomous vehicles, vol 2. Carnegie Mellon University, Technical Report; 1994. p. 105.
- Bitner-Gregersen EM, Dong S, Fu T, Ma N, Maisondieu C, Miyake R, Rychlik I. Sea state conditions for marine structures analysis and model tests. *Ocean Eng*. 2016;119:309–22.
- Breitkopf P, Coelho RF. Multidisciplinary design optimization in computational mechanics. New York: Wiley; 2013.
- Calisal S. A note on the derivation of potential energy for two-dimensional water waves. *Ocean Eng*. 1983;10(2):133–8.
- Chou CR, Yim JZ, Huang WP. Determining the hydrographic parameters of the surface of water from the image sequences of a CCD camera. *Exp Fluids*. 2004;36(4):515–27.
- Cox C, Munk W. Measurement of the roughness of the sea surface from photographs of the sun's glitter. *J Opt Soc Am*. 1954;44:838–50.
- Diab H, Younes R, Lafon P. Survey of research on the optimal design of sea harbours. *Int J Nav Archit Ocean Eng*. 2017;9(4):460–72.
- Explore Oceans: Frying pan camera. <https://explore.org/livecams/frying-pan/frying-pan-cam>. Accessed 18 Feb 2021.
- Explore Oceans: Frying pan ocean cam powered by explore.org. <https://youtube.com/watch?v=vGimsBVpuN8>. Accessed 18 Feb 2021.
- Fedele F, Benetazzo A, Gallego G, Shih PC, Yezzi A, Barbarioli F, Arduini F. Space-time measurements of oceanic sea states. *Ocean Model*. 2013;70:103–15.
- Fritz HM, Yoo J, Haas K, Work P, Barnes C. Particle image velocimetry applied to nearshore video wave images. *Proc. 5th COPRI International Conference on Ocean Wave Measurement and Analysis*, Madrid, Spain, 3-7 July, 2005.
- Holthuijsen L. *Waves in oceanic and coastal waters* by Leo H. Holthuijsen. Cambridge University Press, ISBN-10. ISBN-13; 2007. p. 404
- Huang NE. *Measuring ocean waves: proceedings of a symposium and workshop on wave measurement technology*. Washington, DC: The National Academies Press; 1982.
- Kinsman B. *Wind waves: their generation and propagation on the ocean surface*. Chelmsford: Courier Corporation; 1984.
- Langlois R. *Ship flight deck motion parameters for ensuring safety of helicopter operation*. Mississauga: Indal Technologies Inc.; 2003.
- Loizou A, Christmas J. 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*; 2018. p. 3513–3516.
- National Data Buoy Center: Station 41013 Frying Pan Shoals, NC. https://ndbc.noaa.gov/station_page.php?station=41013. Accessed 18 Feb 2021.
- Osorio AF, Montoya-Vargas S, Cartagena CA, Espinosa J, Orfila A, Winter C. Virtual buoy: a video-based approach for measuring near-shore wave peak period. *Comput Geosci*. 2019;133:104302.
- Park S, Gil MS, Im H, Moon YS. Measurement noise recommendation for efficient Kalman filtering over a large amount of sensor data. *Sensors*. 2019;19(5):1168.
- Pierson WJ Jr, Moskowitz L. A proposed spectral form for fully developed wind seas based on the similarity theory of SA Kitai-gorodskii. *J Geophys Res*. 1964;69(24):5181–90.

23. Rosser K, Chahl J. Wave scale, speed and direction from airborne video of maritime scene. In: 2019 digital image computing: techniques and applications (DICTA). IEEE; 2019. p. 1–7.
24. Shannon CE. Communication in the presence of noise. *Proc IRE*. 1949;37(1):10–21.
25. Spencer L, Shah M, Guha RK. Determining scale and sea state from water video. *IEEE Trans Image Process*. 2006;15(6):1525–35.
26. Stilwell DO Jr, Pilon R. Directional spectra of surface waves from photographs. *J Geophys Res*. 1974;79:1277–84.
27. Stockdon HF, Holman RA. Estimation of wave phase speed and nearshore bathymetry from video imagery. *J Geophys Res Oceans*. 2000;105(C9):22015–33.
28. de Vries S, Hill D, de Schipper M, Stive M. Remote sensing of surf zone waves using stereo imaging. *Coast Eng*. 2011;58(3):239–50.
29. Xie J, Zuo L. Dynamics and control of ocean wave energy converters. *Int J Dyn Control*. 2013;1(3):262–76.
30. Yoo J, Fritz H, Haas K, Work P, Barnes C. Wave property estimation using linear feature extraction from nearshore wave images. In: *The 5th international symposium on ocean wave measurement and analysis*; 2005. p. 3–7.
31. Zikra M, Hashimoto N, Yamashiro M, Yokota M, Suzuki K. Analysis of directional wave spectra in shallow water areas using video image data. *Coast Eng J*. 2012;54(3):1250020–1.
32. Zikra M, Hashimoto N, Yamashiro M, Yokota M, Suzuki K. Application of video images for monitoring coastal zone in Hasaki beach, Japan. *Coast Eng J*. 2012;1(33):43.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.