Proceedings of Australasian Conference on Robotics and Automation, 2-4 Dec 2013, University of New South Wales, Sydney Australia

# An Analysis of Regression Models for Predicting the Speed of a Wave Glider Autonomous Surface Vehicle

**Phillip Ngo[a], Wesam Al-Sabban[a], Jesse Thomas[b], Will Anderson[b], Jnashewar Das[c] and Ryan N. Smith[d†]**
[a]Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD, Australia.
[b]Liquid Robotics, Inc. Sunnyvale, CA, USA.
[c]Robotics Embedded Systems Laboratory, University of Southern California, Los Angeles, CA, USA.
[d]Physics and Engineering Department, Fort Lewis College, Durango, CO, USA.
[†]Corresponding author email: rnsmith@fortlewis.edu

## Abstract

An important aspect of robotic path planning for is ensuring that the vehicle is in the best location to collect the data necessary for the problem at hand. Given that features of interest are dynamic and move with oceanic currents, vehicle speed is an important factor in any planning exercises to ensure vehicles are at the *right place at the right time*. Here, we examine different Gaussian process models to find a suitable predictive kinematic model that enable the speed of an underactuated, autonomous surface vehicle to be accurately predicted given a set of input environmental parameters.

## 1 Introduction

By harvesting abundant natural energy from ocean waves, the Liquid Robotics Wave Glider (see Fig. 1) provides a persistent ocean presence for detailed ocean observation and studies on persistent robotic control. With a demonstrated endurance exceeding one year, the Wave Glider (WG) offers a unique platform for ongoing engineering development and a range of new applications for robotics research and ocean scientists alike [Hine *et al.*, 2009; Manley and Wilcox, 2010].

The WG is an interesting vehicle from the design, operation and control perspectives. The vehicle is composed of a two-part architecture; a surface component housing the electronics, and a subsurface wing system that generates locomotion. This vehicle provides an interesting path planning problem since it is underactuated, and although the on-board controller maintains an accurate heading, the speed of the vehicle is entirely environment-dependent. The primary open problem in controlling the WG is predicting its speed, given a desired heading and known or predicted environmental conditions, e.g., significant wave height, wave peak period, wind speed and direction, and the speed and direction of the ocean surface and subsurface currents.

Figure 1: The Wave Glider platform.

The focus of this paper is primarily concerned with the development of a predictive kinematic model with the purpose to infer the speed of the WG autonomous surface vehicle using environmental variables as inputs. This work builds upon the conclusions found in [Smith *et al.*, 2011a], following up on the key points the authors identified for further research. We use Gaussian Processes to derive a nonlinear, probabilistic, nonparametric model to predict vehicle velocity from environmental parameters recorded on-board.

## 2 Background

Path planning for autonomous underwater vehicles (AUVs) is required for a wide variety of applications, such as mine countermeasures, ecosystem monitoring, locating hydrothermal vents, and tracking dynamic features. One important aspect of the planning process is making sure that the vehicle is in the best location to col-

lect the data necessary for the problem at hand [Smith *et al.*, 2010]. Given that dynamic features of scientific points of interest may move with oceanic currents, vehicle speed is an important factor in any planning exercises to ensure AUVs are at the *right place at the right time* [Smith *et al.*, 2011b].

The propulsion mechanism for the WG, and in turn its speed, are completely dependent upon the surrounding environment, knowing the expected speed can assist in fault detection. The ocean environment raises challenges that are generally not encountered in conventional land based automated robots given its dynamic and complex nature.

**The Wave Glider** The Wave Glider consists of two primary subsystems, a submerged glider and a surface float connected together by a tether approximately 7 m long. The key innovation of the Wave Glider is its ability to mechanically convert energy from ocean waves to provide essentially limitless propulsion.

The surface float is lifted by a rising wave, which, in turn lifts the submerged glider. As it does so, specially designed wings on the sub glider are engaged to convert this upward motion of the sub glider to an up and forward motion, thus also pulling the surface float forward and subsequently off the wave. Similarly, as the surface float comes off the wave, the sub glider drops along with it and its wings will pivot to cause it to move in a down and forward motion, again pulling the surface float along. This process is repeated as long as there is wave motion on the surface, even the smallest amount. This process is illustrated in Figure 2 on the following page.

Several articles have been published detailing the architecture, mechanics, and capabilities of the Wave Glider including [Daniel *et al.*, 2011] and [Hine *et al.*, 2009]. Being a relatively new and promising technology, the Wave Glider has been the subject of a number of case studies, particularly in the area of data collection pertaining to oceanography. [Daniel *et al.*, 2011] illustrates the potential value this platform has to offer to the ocean science community by reporting on the positive outcomes of a number of scientific studies that utilised data collected by the Wave Glider. This has been a primary motivation for [Smith *et al.*, 2011a] to find a predictive model for the Wave Gliders speed from its surrounding environmental parameters.

**Vehicle and Environmental Parameters** Throughout this project, a number of different data streams, obtained from various sources, is considered. These include vehicle specific data collated from field experiments performed off the coast of Monterey Bay and off the coast of the island of Hawai'i, as well as environmental data collected by third party sources and/or measured by equipment on-board the Wave Glider itself.

A study conducted by [Smith *et al.*, 2011a] makes use of environmental data that is readily available from various sources for a trial period in Fall 2010. In this work, we re-examine the same data to determine whether a more suitable predictive model can be found. The data for this field trial includes wave data, water currents, and wind data collected within a one week period starting on 20th October, 2010 in Monterey Bay, California. Details on these data can be found in [Smith *et al.*, 2011a].

During a six month period between $15_{th}$ June and $15_{th}$ December of 2012, a Wave Glider, named Kaimana, was deployed off the Island of Hawaii. Throughout this period, the Wave Glider collected multiple data streams that were used in our analysis. The environmental data in this trial was entirely gathered by instruments on-board the vehicle, contrary to the Monterey field trial that utilized data from external sources, e.g., ocean models and moorings. This dataset will form the basis of the analysis presented here.

## 3 Technical Approach

In this section we will discuss the details of the dependent and independent variables along with a description of the regression analysis techniques used.

### 3.1 Environmental Parameters

**Wave Spectrum** Wave spectrum is measured by a Datawell sensor on-board the Wave Glider. This allows the sea state of the immediate area of the Wave Glider to be considered; in contrast to the data collected from the Monterey Bay experiment, which were sourced from various third parties and may have been subjected to secondary processes such as spatiotemporal interpolation. The spectrum data available spans a period of three continuous weeks between the months of June and July.

The wave spectrum includes power spectral density (PSD), ellipticity, skewness, directional spread (degrees), kurtosis, and direction (degrees) per frequency bin. The spectrum is generated every 30 minutes from 256, or 128 samples, depending on GPS quality, and spans a frequency range of 0 to 1 Hz, with two samples taken per frequency bin. Figure 2 below shows the power spectral densities that were measured during this period.

The spectrum also provides an enriched collection of wave parameters including peak period, peak wave direction, and the spectral analogues of mean period and significant wave height.

**Vehicle Parameters** In addition to the wave spectrum data, the vehicle coordinates and speeds were made available during the six month period, which encompasses the three weeks the wave spectrum data was available. The distribution of WG speeds is shown in Fig. 3, with a representative vehicle course displayed in Fig. 4.

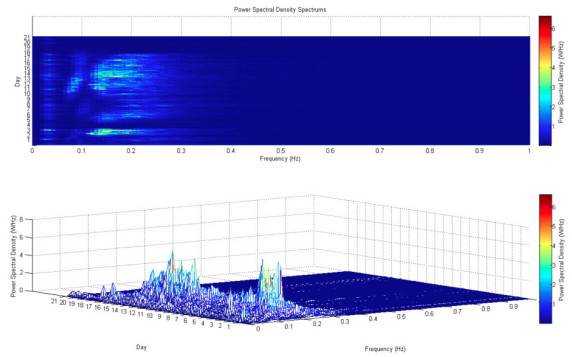


Figure 2: Power spectrum densities measured by the Wave Glider during the Hawai'i field trial.

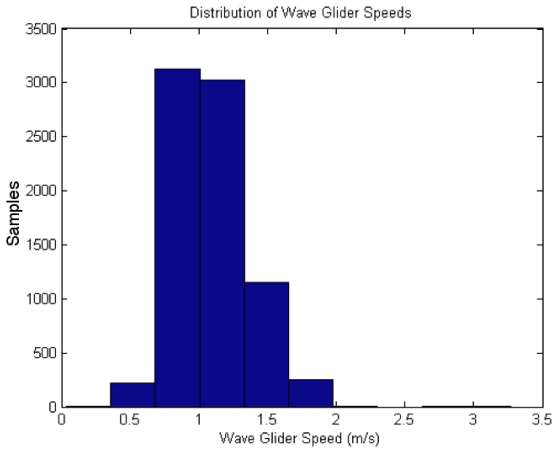


Figure 3: Distribution of Wave Glider speed during the Hawai'i field trial.

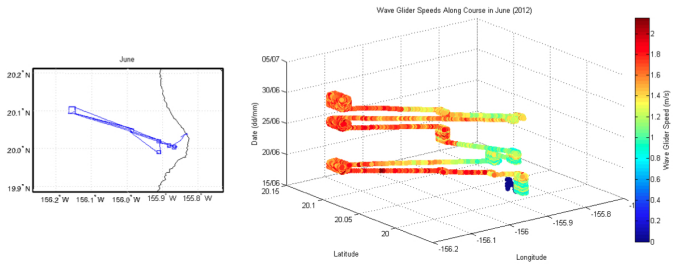


Figure 4: Top view (left) of the WG path and positioning with speed (right) during June 2012 of the Hawai'i field trial.

**Vehicle Speed Prediction** In [Manley and Wilcox, 2010], the forward speed of a WG is reported to be dependent upon the amplitude of the surface waves, the overall buoyancy force provided by the float, and the weight of the vehicle [Manley and Wilcox, 2010]. Despite this knowledge, it is generally known that models based on pure dynamics do not tend to perform well in real world applications; primarily because of the difficulty in identifying and incorporating every contributing factor into such models. With this, the use of regression analysis methods has been proven to be popular in this area, where relationships among variables are inferred without the need to completely understand these unknowns.

In supervised learning, a function is inferred from labelled training data, where a set of input data points and associated output values, generally obtained through experimentation, are available. In the case of this study, the input data points will be vectors with $D$ number of different environmental parameters, $x \in \Re^D$, with corresponding scalar output values $y \in \Re$ representing the speed of the WG. Hence, the objective is to learn the model parameters for the following simplified model:

$$y = f(x) \tag{1}$$

The problem of inferring falls into either the regression or classification cases depend on whether the output values are either continuous or discrete, respectively [Rasmussen and Williams, 2006]. For this study, the output vehicle speed will be treated as a continuous function of the environmental parameters, hence only the regression problem will be investigated.

A previous study has shown that a linear model will not be sufficient for WG speed prediction and a non-linear regression model needs to be explored [Smith *et al.*, 2011a]. As recommended, this project will attempt to make use of Gaussian Processes to define a non-parametric, non-linear, probabilistic regression model. In [Smith *et al.*, 2011a], the authors hypothesised that one could exploit the high-dimensional input space more effectively than with a simple linear regression. From their conclusions, there is evidence that suggests that other factors may not have been captured in their model and that a non-parametric method will be able to compensate for these unknowns.

## 3.2 Linear Regression

A basic linear regression model is initially developed using the procedure outlined in [Smith *et al.*, 2011a], in order to, not only confirm their reported findings, but to also provide an initial benchmark for comparison.

Specifically,the function,$f(x)$ ,is defined by a linear coefficient vector,$w \in \Re_D$, (consider $\Re^D$)such that:

$$y = f(x) = w^T x \tag{2}$$

For an appropriate training dataset, $X \in \Re^{D \times n}$ and $Y \in \Re^n$ where $X$ and $Y$ contain the input environmental

parameters and output Wave Glider speeds respectively for $n$ number of observations. Applying the method of least squares allows the model parameters, $w$,to be calculated using:

$$w = (X^T X)^{-1} X^T Y \tag{3}$$

Then for any unobserved inputs, $x^*$, corresponding predictions of $y^*$, are simply then calculated as:

$$y^* = f(x^*) = w^T x^* \tag{4}$$

### 3.3 Gaussian Process Regression

Rather than attempting to fit the data to a specific model such as described for the linear system above, non-parametric methods such as Gaussian process regression (GPR) allows the data to play a more prominent role in regression analysis.

Among the first to describe Gaussian process regression and classification in a machine learning context is [Rasmussen and Williams, 2006]. Their text, which has been heavily cited, aims to give a systematic and unified treatment of the area, as well as show connections to other existing and related models [Rasmussen and Williams, 2006]. Although Gaussian processes (GPs) of many types have been well-studied and used in statistics; being generally recognised as one of the most flexible and accurate regression methods available, its use in the machine learning context has only just been recently explored. The authors of [Rasmussen and Williams, 2006] hypothesize several reasons for this, including its associated costs of possibly $O(n^3)$ computing time, as it requires an expensive inversion of the training data covariance matrix, and a lack of understanding when comes to selecting suitable covariance functions [Rasmussen and Williams, 2006].

The following section will aim to provide a brief introduction to Gaussian process regression.

**Gaussian Processes** Gaussian processes are probabilistic models that have historically proven popular in the area of non-parametric regression with wide applications in areas such meteorology, and geostatistics where it is known as kriging. Gaussian processes can be thought of as a generalisation of the well-known Gaussian probability distribution to infinite dimensionality. Whereas a Gaussian distribution is concerned with only a single random variable, a Gaussian process is associated with a collection of random variables that produces a pool of functions relevant for prediction. Hence, a common view adopted is that Gaussian process can be thought of as a distribution over functions [Rasmussen and Williams, 2006].

[Rasmussen and Williams, 2006] defines a Gaussian process as a collection of random variables, any finite number of which has a joint multivariate Gaussian distribution. A simplified notation commonly used for a Gaussian process is given as:

$$f(x) \sim GP(m(x), k(x, x^{'})) \tag{5}$$

Where $m(x)$ and $k(x, x^{'})$ is its mean and covariance functions, respectively. Just as the mean and covariance can fully define a Gaussian distribution, the same can be said about the mean and covariance functions for Gaussian processes.

**Mean Function** The mean function,$m(x)$, is defined as the expected value of $f(x)$ :

$$m(x) = \mathbb{E}[f(x)] \tag{6}$$

It is common practice, as will be the case in this project, to assume the initial mean function to be zero for simplicity; that is, the average value of the functions at each $x$ in the prior Gaussian distribution is equal to zero. There are cases where explicitly specifying a non-zero mean function may be beneficial, but a zero mean function will not place a limit on the posterior process [Rasmussen and Williams, 2006]. Effectively, this then leads the system to be entirely determined by the covariance function.

**Covariance Function** The covariance function plays the most crucial role in GPR, as evidently suggested when an initial zero mean function is assumed. Although GPR are generally considered a non-parametric method of regression, there are still assumptions that need to be made about the model trying to be inferred. The covariance function encodes these assumptions about the predictor, by dictating the relationships between observations [Rasmussen and Williams, 2006]. The covariance function is defined as follows.

$$\begin{gathered} cov(f(x), f\left(x^{'}\right)) = k(x, x^{'}) = \\ \mathbb{E}\left[(f(x) - m(x))\left(f\left(x^{'}\right) - m(x^{'})\right)\right] \end{gathered} \tag{7}$$

This is the covariance between a pair of output values at inputs $x$ and $x^{'}$. Using this, a covariance matrix, $K$, for $n$ number of observations can be built where each entry is the calculated covariance between the output values of every $x_i$ and $x_j$:

$$K_{i,j} = k(x_i, x_j); \forall i, j = 1....n \tag{8}$$

For a covariance function to be valid within the GPR framework, it should have certain characteristics. More specifically, the resulting covariance matrix, $K$, should be positive semi-definite, as well as symmetric. Multiple covariance functions can also be combined to form other covariance functions to obtain desired characteristics [Rasmussen and Williams, 2006].

With this, there are many different possible covariance functions to suit the needs of the application. It has even been proven that many other different processes, such as Brownian motion and Wiener processes, have since been found to be Gaussian processes, just with different covariance functions [Rasmussen and Williams, 2006].

In training the model, the predictive mean function is obtained by conditioning on the training data. By doing so, the best set of functions is selected that passes through, or closely to, each of the training data points according to the assumptions encoded in the specified covariance function. For instance, the prior may favour smooth or stationary functions, or functions with certain length scales. All this is reflected in the choice of the covariance function, which essentially captures prior beliefs about the function to be inferred [Rasmussen and Williams, 2006].

After analysing the general characteristic shape of the Wave Gliders speed, the covariance functions primarily used in this project will be the squared exponential and the Matern class of covariance functions.

**Squared Exponential Covariance Function** The squared exponential covariance function is a widely used stationary covariance function that favours a smooth and continuous posterior by enforcing a maximum covariance of $\sigma_f^2$ between closely neighbouring observations [Rasmussen and Williams, 2006]. For this project, only the isotropic variety, in which the function can be defined as a function of $x - x^{'}$, will be considered. The squared exponential covariance function is given as:

$$k_{SE}(x, x^{'}) = \sigma_f^2 e^{-\frac{(x-x^{'})^2}{2l^2}} \tag{9}$$

This function has two parameters, also known as its hyperparameters, $\sigma_f^2$ and $l$, where these are the variance and characteristic length scale, respectively, and are configurable, depending on the desired characteristics.

**Matern Class of Covariance Functions** Where smoothness may not be a desirable characteristic, the use of Matern class covariance functions has been recommended [Rasmussen and Williams, 2006]. This class of covariance functions has the general form:

$$k_{matern}(x, x^{'}) = \frac{2^{1-\upsilon}}{\Gamma(\upsilon)} \left( \frac{\sqrt{2\upsilon}(x - x^{'})}{l} \right)^2 \times K_\upsilon \left( \frac{\sqrt{2\upsilon}(x - x^{'})}{l} \right)$$

With positive parameters $\upsilon$ and $l$, and the function, $K_\upsilon$, is a modified Bessel function [Rasmussen and Williams, 2006]. For this project, the three Matern class covariance functions examined are when $\upsilon = 1/2$, $\upsilon = 3/2$, $\upsilon = 5/2$ in which the above can be reduced to the following respectively:

$$k_{\upsilon=1/2}(x, x^{'}) = \sigma_f^2 e^{-\frac{(x-x^{'})}{l}} \tag{10}$$

$$k_{\upsilon=3/2}(x, x^{'}) = \sigma_f^2 \left( 1 + \frac{\sqrt{3}\left(x - x^{'}\right)}{l} \right) e^{-\frac{\sqrt{3}\left(x-x^{'}\right)}{l}} \tag{11}$$

$$k_{\upsilon=5/2}(x, x^{'}) = \sigma_f^2 \left( 1 + \frac{\sqrt{5}\left(x - x^{'}\right)}{l} + \frac{5\left(x - x^{'}\right)^2}{3l^2} \right) \times e^{-\frac{\sqrt{5}\left(x-x^{'}\right)}{l}}$$

**Noisy Data** For the system analysed here, it is assumed that the data has some element of noise,$\epsilon$, involved such that:

$$y = f(x) + \epsilon \tag{12}$$

Assuming that this noise has a Gaussian distribution with zero mean, and is additive independent identically distributed (i.e. $\epsilon \sim N(0, \sigma_n^2)$), it can be combined with the covariance function such that:

$$cov(f(x), f\left(x^{'}\right)) = k\left(x, x^{'}\right) + \sigma_n^2 \zeta\left(x, x^{'}\right) \tag{13}$$

Where $\sigma_n^2$ is the noise variance and $\zeta\left(x, x^{'}\right)$ is the Kronecker delta function (i.e. where $\zeta\left(x, x^{'}\right) = 1$ at $x = x^{'}$ and is 0 everywhere else). Alternatively, this could be written with the covariance matrix, $K$ , and identity matrix, $I$:

$$cov(y) = K + \sigma_n^2 I \tag{14}$$

This essentially means that the value to be predicted, for some new given input, $x^*$, is now $y^*$ instead of just $f(x^*)$. It could be shown that the expected values of $y^*$ and $f(x^*)$ are identical, but their variances differ owing to the added observational noise process [Rasmussen and Williams, 2006].

**Regression Analysis** Since the data is assumed to be jointly multivariate Gaussian, for an unobserved input,$x^*$, and its predicted output ,$y^*$, it could be written:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K^{*T} \\ K^* & K^{**} \end{bmatrix}\right) \tag{15}$$

Where $K^{**} = k(x^*, x^*)$ and $K^* = k(x^*, x_i)$; $\forall i = 1....n$

In regression, the actual prediction of $y^*$ for unknown input, $x^*$, is given by the conditional probability distribution of observing the predicted value, $y^*$,given the training data and observations, $X$ and $Y$ respectively [Rasmussen and Williams, 2006]. Therefore, from joint multivariate Gaussian theory, this conditional probability distribution could then be written as:

$$p\left(y^*|x^*, y, X\right) \sim N\left(\overline{y^*}, cov\left(y^*\right)\right) \tag{16}$$

Where the posterior mean distribution is found to be:

$$\overline{y^*} = K^*\left(K + \sigma_n^2 I\right)^{-1} y \tag{17}$$

And the variance is:

$$cov\left(y^*\right) = K^{**} - K^*\left(K + \sigma_n^2 I\right)^{-1} K^{*T} \tag{18}$$

Note that since the posterior distribution is Gaussian, the mean of this distribution is also its maximum a posteriori (MAP) estimate [Rasmussen and Williams, 2006].

**Selection of Hyperparameters** In keeping with a non-parametric approach, throughout this study the hyperparameters, commonly defined as a vector,$\theta$, of the covariance functions used are allowed to be naively conditioned, or learnt, based on the training data. A Bayesian model selection method, known as the marginal likelihood maximisation method, is commonly used to infer these parameters.

The maximum a posteriori estimate of the hyperparameters,$\theta$, occurs when the conditional probability of having a set of hyperparameters given the training data (i.e $p\left(\theta|X, y\right)$) is at its highest [Rasmussen and Williams, 2006]. Using Bayes theorem, and assuming little prior knowledge about the hyperparameters, this would correspond to maximising the log marginal likelihood found to be given as:

$$\log p\left(y|X, \theta\right) = -\frac{1}{2} y^T \left(K + \sigma_n^2 I\right)^{-1} y - \frac{1}{2}\log\left|K + \sigma_n^2 I\right| - \frac{n}{2}\log 2\pi \tag{19}$$

Therefore, the resulting MAP estimate of the hyperparameters is those that correspond to the maximum of this log marginal likelihood [Rasmussen and Williams, 2006].

**Feature Space Reduction** One of the reasons that Gaussian processes have only been recently considered in Machine Learning, despite its prominence in statistics, is the given relatively large computational costs associated with them. This is largely because of the need to calculate the inverse of the covariance matrix dictated by the equations above. Therefore, given the dimensionality and number of data points available in the datasets considered here, it was deemed relevant to ensure that the feature space contained only the parameters that will have an actual influence over to the predicted values. Initial analysis also shows that there is significant variability in the data and so it would be prudent to ensure that an appropriate sample is taken for training that would adequately represent the entire system. As such, it was deemed necessary to employ various sampling configurations.

**Bootstrap Aggregating** In [Chen and Ren, 2009], the application of Bootstrap Aggregating, or Bagging, to Gaussian process models have been found to result in more robust and accurate predictions [Chen and Ren, 2009]. As such, this method was briefly examined in this project to determine whether model accuracy could be improved.

The method of Bootstrap Aggregating in Gaussian process regression is very straightforward, where it simply involves training multiple Gaussian processes using randomly sampled data points in the available data set. Where, in turn, the predicted values from these models are then combined using a suitable averaging rule. The simple averaging rule for multi-model combination is the original method, in which the predicted output is the average of the predicted values [Chen and Ren, 2009].

Another averaging rule suggested by [Chen and Ren, 2009] is a weighted averaging rule where the average of the predictions, weighed by the inverse of their respective calculated predicted variance, is taken as the predicted output. [Chen and Ren, 2009] hypothesised that this method of weighted averaging will allow the prediction uncertainty of the models to be automatically accounted for [Chen and Ren, 2009].

## 4 Results and Analysis

### 4.1 Monterey Bay, California

Initial analysis of the data pertaining to the Monterey Bay experiments show that there is evidence of correlation between the Wave Glider speed with both the significant wave height and wave peak period, as also reported by [Smith *et al.*, 2011a]. This can be easily seen from Figs. 5, 6, and 7, where the Wave Glider speed is shown against each of the environmental data after the data has been demeaned and normalised.

A summary of the calculated correlation coefficients are collated in Table 1 below, with significant wave height and wave peak period both exhibiting relatively high correlation coefficients, R=0.5640 and R=0.1193, respectively, with small p-values ($p < 0.05$). The p-values presented here are the probability of observing the calculated correlation when the true correlation is zero. The relatively high observed correlation calculated

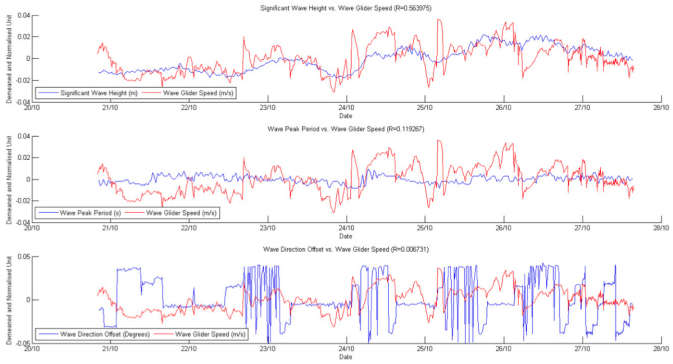


Figure 5: Correlation between Wave Glider Speed and wave input parameters during the Monterey Bay field trial.

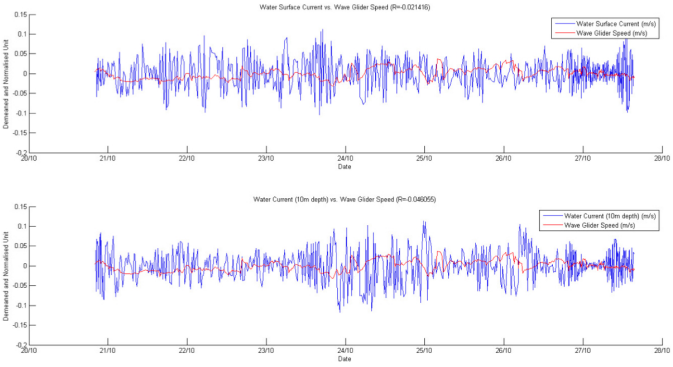


Figure 6: Correlation between Wave Glider Speed and water current speed input parameters during the Monterey Bay field trial.

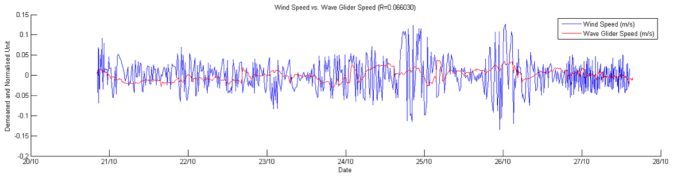


Figure 7: Correlation between Wave Glider Speed and wind speed during the Monterey Bay field trial.

| *Input* | *Correlation* | *p-value* |
|---|---|---|
| Significant Wave Height (m) | 0.5640 | 0.0000 |
| Wave Peak Period (s) | 0.1193 | 0.0028 |
| Wave Direction | 0.0067 | 0.8663 |
| Wind Speed Projection (m/s) | 0.0660 | 0.0983 |
| Water Surface Current (m/s) | -0.0461 | 0.2491 |
| Subsurface Current (10m depth) (m/s) | -0.0214 | 0.5922 |

Table 1: Calculated correlation coefficients and p-values from observed data between proposed inputs, environmental parameters, and desired output, Wave Glider speed.

from projected wind speed has a much higher p-value which therefore leads to the belief that this value has relatively lower significance.

**Linear Regression** In accordance with the procedure outlined in [Smith *et al.*, 2011a], a training dataset, generated by taking every fifth data point in the available data, was used to create a linear model with the remaining used for validation. The result was found to be the following linear regression equation:

$$y_{sog} = 0.1623x_{wht} + 0.0086x_{wpp} + 0.0001x_{wdir} + 0.0124x_{wnd} - 0.0557x_{adcp} - 0.2660x_{hfr} \quad (20)$$

Where $y_{sog}$ is the predicted Wave Glider speed,$x_{wht}$ is the significant wave height,$x_{wpp}$ is the wave peak period, wwdir is the wave direction offset,$x_{wdir}$ is the projected wind speed,$x_{adcp}$ is the projected water surface current, and $x_{hfr}$ is the projected subsurface water current at the depth of 10 m.

The computed results differ from the results presented in [Smith *et al.*, 2011a] with root square mean errors of 0.122 and 0.137 for the test and training datasets respectively, which are noticeably better than what was previously reported (i.e. RMSE = 1.0012 and RMSE = 1.0004 for the testing and training datasets, respectively). These differences may be the result of different interpolation methods used when treating the raw data, different methods of handling anomalies or outliers in the data, and/or discrepancies in the actual raw data itself. It should also be noted that during this investigation that there was inconsistencies identified in the raw data, such as incorrect timestamps that was subsequently resolved by further processing and/or sourcing fresh data.

**Gaussian Process Regression** A number of Gaussian processes were produced using the various covariance functions, with a zero mean function, mentioned above. The hyperparameters were allowed to be naively

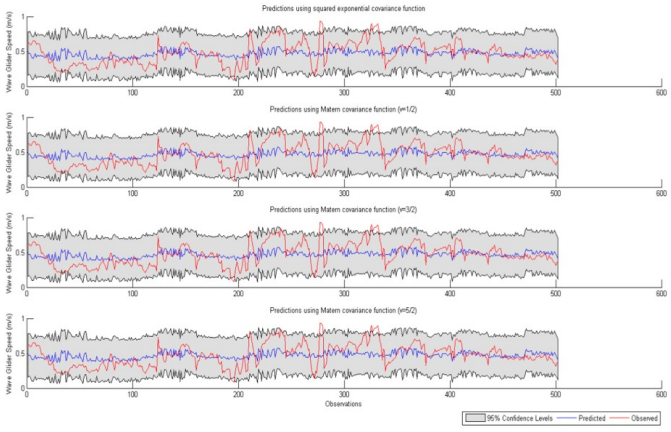


Figure 8: Predicted versus actual Wave Glider speeds from Gaussian process models using different covariance functions during the Monterey Bay field trial.

| *Method* | *RSME* | *R* |
|---|---|---|
| Linear Regression | 0.1369222000 | 0.5410785571 |
| GPR | | |
| Squared Exponential | 0.1476193620 | 0.4144884772 |
| Matern (v=1/2) | 0.1485630500 | 0.4087973842 |
| Matern (v=3/2) | 0.1482156653 | 0.4086340682 |
| Matern (v=5/2) | 0.1480338464 | 0.4099077562 |

Table 2: Calculated root squared mean errors and correlation coefficients between predicted and actual Wave Glider speeds by regression method.

conditioned using the same training dataset employed to produce the linear model in the previous section. Figure 8 illustrates the predictions against the observed Wave Glider speeds for each of the squared exponential and Matern class covariance functions.

In Fig. 8, the calculated variance is used to find the 95% confidence interval of the prediction, of which is approximately two times the standard deviation of the posterior distribution, and is represented by the grey areas. These graphs show relatively high standard deviations, which is reflective of the high variability in the data.

The calculated root square mean errors and correlation coefficients for the predicted and actual Wave Glider speeds are presented in Table 2 on the proceeding page for comparison.

Based on these metrics, it could be said that there is little difference in predictor performance across each of the Gaussian processes considered. To determine whether an improvement can be found, feature space reduction was explored.

A greedy forward feature selection approach is utilised where, because of their relatively high observed correlations, the significant wave height and wave peak period was initially considered. From this, it was discovered that only a slight improvement is observed in all Gaussian processes, the best given from Matern covariance function (v=5/2), with a calculated root mean square error of 0.1480 and correlation coefficient of 0.4099 between the predicted and observed data; shown in Fig. 9.

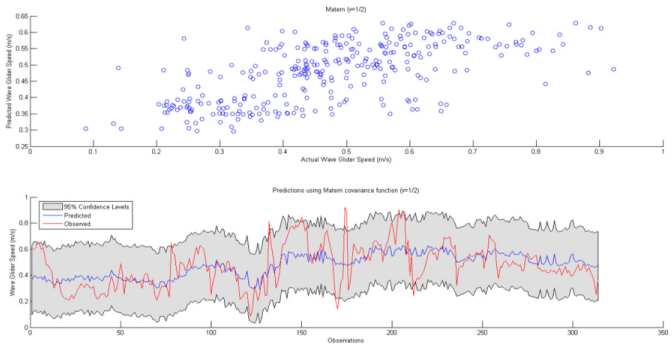


Figure 9: Predicted versus actual Wave Glider speeds from Gaussian process models using different covariance functions during the Monterey Bay field trial.

## 4.2 Hawaii Field Trial Results

The regression analysis using data collated from the Monterey Bay field experiments have shown to produce relatively unreliable predictive models for Wave Glider speed. Justification in using sparse environmental data from third party sources in the area was given by assuming that sea state features are relatively constant across low spatial and temporal resolutions [Smith *et al.*, 2011a].

In addition to assessing the validity of this assumption, this project has considered environmental data collected by the Wave Glider itself, via on-board instruments, during Hawaii field experiments in the three weeks from June to July of 2012.

**Linear Regression** In a similar approach with the Monterey Bay data analysis above, a linear model was initially developed using the data available from the Hawaii field experiments to provide an initial baseline for comparison. Each of the different spectrum types were considered individually, as well as in various combinations, from which a subset of the results are presented in Table 3.

The apparent poor results above motivated an investigation into other sampling and segmentation configurations in order to try and capture as much variability as possible to provide a better comparison. With this, more reasonable model accuracy can be achieved with a 70/30 training and testing dataset configuration, where the first 7 of every 10 data points is used for training as

| *Spectrum Type* | *RSME* | *R* |
|---|---|---|
| Power Spectral Density | 2.9956 | 0.1342 |
| Ellipticity | 0.8052 | 0.2100 |
| Skewness | 1.2052 | 0.1932 |
| Kurtosis | 4.7828 | 0.0343 |
| Direction | 1.3280 | 0.0425 |
| Spread | 0.5639 | 0.0061 |
| Power Spectral Density and Ellipticity | 5.2448 | 0.0061 |

Table 3: Calculated root squared mean errors and correlation coefficients between predicted and actual Wave Glider speeds by input spectrum type.

shown in Table 4.

| *Spectrum Type* | *RSME* | *R* |
|---|---|---|
| Power Spectral Density | 0.6776 | 0.5504 |
| Ellipticity | 0.2386 | 0.6500 |
| Skewness | 0.4619 | 0.5778 |
| Kurtosis | 1.7025 | 0.0592 |
| Direction | 0.3955 | 0.6218 |
| Spread | 0.1791 | 0.6916 |
| Power Spectral Density and Ellipticity | 0.3224 | 0.5641 |

Table 4: Calculated root squared mean errors and correlation coefficients between predicted and actual Wave Glider speeds by input spectrum type using 70/30 sample segmentation configuration.

**Gaussian Process Regression** While finding a Gaussian process regression model, a greedy forward feature selection approach was again adopted in an attempt to reduce the feature space to only include the more significant environmental parameters. Training datasets were generated using the same sampling and segmenting configuration from above (i.e. 70/30). Using a Matern covariance function ($v = 3/2$), the results are listed in Table 5. From which, the best results were found to be returned with the power spectral density and ellipticity spectrums; see Fig. 10.

Using the power spectral density and ellipticity spectrum, GP models were then created using each of the covariance functions to be considered. In addition to this, as mentioned above, Bootstrap Aggregating was also briefly explored to determine whether the practice improved the accuracy of the proposed model. Here, five GP models, using 20 data points randomly selected from the full data set, was generated for this purpose with the results presented in Table 6.

For Bootstrap Aggregating, both the simple and weighted averaging rules were found to be equivalent,

| *Spectrum Type* | *RSME* | *R* |
|---|---|---|
| Power Spectral Density | 0.0879 | 0.9310 |
| Ellipticity | 0.1322 | 0.8339 |
| Direction | 0.2386 | NaN |
| Spread | 0.2386 | 0.1566 |
| Kurtosis | 0.2318 | 0.2468 |
| Skewness | 0.1647 | 0.7254 |
| Power Spectral Density and Ellipticity | 0.1647 | 0.9422 |

Table 5: Calculated root squared mean errors and correlation coefficients between predicted and actual Wave Glider speeds using Gaussian process model with Matern covariance function (v=3/2).

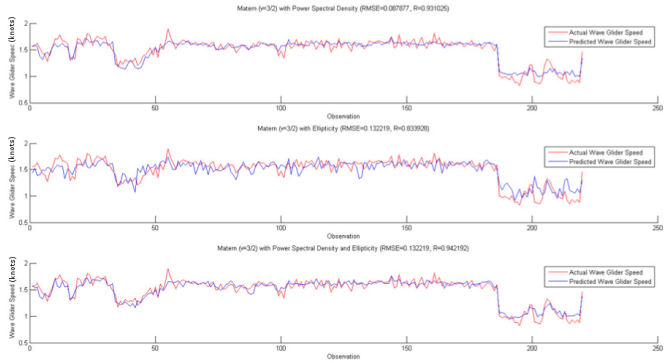


Figure 10: Predicted and actual Wave Glider speeds using Gaussian process model with Matern covariance function (v=3/2).

| *Spectrum Type* | *RSME* | *R* |
|---|---|---|
| Squared Exponential | 0.0831 | 0.9436 |
| Matern (v=1/2) | 0.0801 | 0.9422 |
| Matern (v=1/2) | 0.0807 | 0.9421 |
| Matern (v=1/2) | 0.0809 | 0.9390 |
| Matern (v=1/2) with Bootstrap Aggregating | 0.0544 | 0.9728 |

Table 6: Calculated root squared mean errors and correlation coefficients between predicted and actual Wave Glider speeds using Gaussian process models by covariance function (including Bootstrap Aggregating).

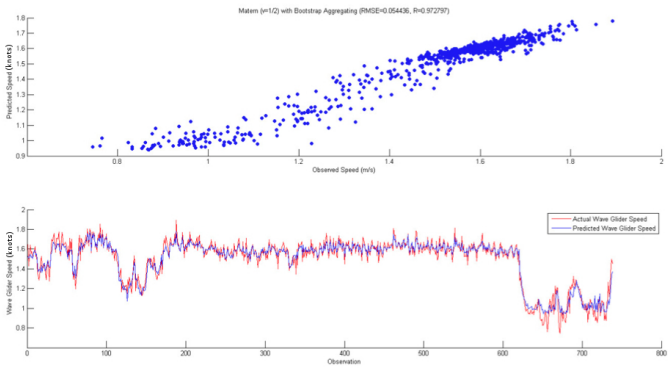


Figure 11: Predicted and actual Wave Glider speeds using Gaussian process model with Matern covariance function (v=1/2) using Bootstrap Aggregating.

since the calculated uncertainty for the predictions from each model was the same. However, it should still be noted that Bootstrap Aggregating has provided significantly more accurate predictions using training datasets that were much smaller than those used in building the other models, see Fig. 11. This was likely due to the smaller models being able to better capture more of the variability in the data. As such, an emphasis in looking at different sampling and segmentation techniques in building training datasets would be important when dealing with high variability data.

### 4.3 Future Work

Methods of reinforcement learning could also be investigated to look at course planning optimisation, such that course proposals may be found that would allow the vehicle to reach a certain point in space at the minimal, or at a specified, time. This would effectively allow the Wave Glider to make intelligent decisions on its own in regards to the best course of action to fulfil its functional objectives, of which, many would agree, is to be at the *right place at the right time*.

In relation to forecasting, this project has only studied the case where only a single time series is concerned. Methods exist for the modelling and interpretation of vector-valued multivariate time series, which could also be explored. Additionally, alternative wave models that provide prediction data at higher temporal and spatial resolutions should also be considered.

## Acknowledgements

This work was supported in part by both an International Fellowship through the Queensland Smart State Program and a grant from the Early Career Academic Recruitment and Development (ECARD) Program of the Queensland University of Technology. This work was also supported in part by the NOAA MERHAB program under grant NA05NOS4781228 and by NSF as part of the Center for Embedded Network Sensing (CENS) under grant CCR-0120778.

## References

[Chen and Ren, 2009] Tao Chen and Jianghong Ren. Bagging for gaussian process regression. *Neurocomputing*, 72(7-9):1605 – 1610, 2009.

[Daniel *et al.*, 2011] Tom Daniel, Justin Manley, and Neil Trenaman. The wave glider: enabling a new approach to persistent ocean observation and research. *Ocean Dynamics*, 61(10):1509–1520, 2011.

[Hine *et al.*, 2009] R. Hine, S. Willcox, G. Hine, and T. Richardson. The wave glider: A wave-powered autonomous marine vehicle. In *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, pages 1 –6, oct. 2009.

[Manley and Wilcox, 2010] Justin Manley and Scott Wilcox. The wave glider, a persistent platform for ocean science. In *Proceedings of the MTS/IEEE OCEANS*, Sydney, Australia, May 2010.

[Rasmussen and Williams, 2006] C. E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, 2006.

[Smith *et al.*, 2010] Ryan N. Smith, Yi Chao, Peggy P. Li, David A. Caron, Burton H. Jones, and Gaurav S. Sukhatme. Planning and implementing trajectories for autonomous underwater vehicles to track evolving ocean processes based on predictions from a regional ocean model. *International Journal of Robotics Research*, 29(12):1475–1497, October 2010.

[Smith *et al.*, 2011a] Ryan N. Smith, Jnaneshwar Das, Graham Hine, Will Anderson, and Gaurav S. Sukhatme. Predicting wave glider speed from environmental measurements. In *MTS/IEEE OCEANS*, Kona, Hawaii, Sep 2011.

[Smith *et al.*, 2011b] Ryan N. Smith, Mac Schwager, Stephen L. Smith, Burton H. Jones, Daniela Rus, and Gaurav S. Sukhatme. Persistent ocean monitoring with underwater gliders: Adapting sampling resolution. *Journal of Field Robotics*, 28(5):714 – 741, September/October 2011.