Gaussian Process Regression for Vessel Performance Monitoring

Benjamin Pjedsted Pedersen, FORCE Technology, Copenhagen/Denmark, bpp@force.dk
Jan Larsen, DTU Compute, Technical University of Denmark, Copenhagen/Denmark, janla@dtu.dk

Abstract

It is showed how Gaussian Process Regression (GPR) can be used equally good or better than Artificial Neural Networks (ANN) for short and long term predictions of the energy consumption on a ship. Using different data sets, from five sister container ships, the method is tested with very different data quality, quantity and prediction horizon and shows that GPR is equally accurate and computational more efficient than ANN, but also offer the possible of finding the predictive variance. Furthermore, the used of so-called characteristic length-scales can be used for evaluating the importance of the input variables for the propulsion performance. A crude introduction to GPR is given together with how it has been applied for this purpose.

1. Introduction

The recent years focus on fuel efficient ship operations have increased significantly. Partly because of the increasing oil prices, but also in general to reduce cost in the economically strained situations many shipowners have been in recently. Furthermore, the need for shipowners to show awareness for the environment and climate changes is increasing. And finally the political pressure to reduce emissions from shipping have been concretized through the International Maritime Organisation (IMO) by introducing of the Energy Efficiency Design Index (EEDI), which apply to all new buildings from 2013.

Energy efficient operation of ships has many facets. It both comprises voyage planning, efficient loading and discharging, if possible optimal trim of the loading conditions, and long term factors as wear on the engines and other mechanical part together with fouling of the hull and propeller. Since the condition of the hull and propeller is difficult to assess it is also difficult to estimate the effect of fouling of the propulsion power.

Traditionally the effect of fouling has been evaluated by comparing e.g. the actual fuel consumption with a theoretical estimate based on empirical methods and other available data as e.g. model tests.

Two well-known and robust methods are *Holtrop (1984)* and *Harvald (1983)* that are based on statistical data from model tests of a large number of ships, which makes them more suitable for ship models and loading conditions resembling the ones tested. This is often not the case since model tests usually only are performed at the design and/or one ballast condition.
Furthermore corrections for the environmental conditions, as the wind speed and direction (*Isherwood (1972)*), the wave height, period and directions (from e.g. *Bhattacharyya (1978)*) where both methods also are based on limited data.

These factors often lead to large discrepancies from the operational measured energy consumption.

In *Pedersen and Larsen (2009)* different regression method was tested, but the Artificial Neural Network (*ANN*) was found superior in all tests. Later tests performed using the Gaussian Process Regression method showed similar or better performance than the *ANN*, which led to a more elaborate study which is given herein.

2. Regression methods

Gaussian Processes, GP, is non-parametric model that provides a flexible framework for regression. The regression function does not take a predetermined form but is constructed from information derived from the data. The definition by *Rasmussen and Williams (2006)* is: "A Gaussian process is a collection of random variables, any Gaussian process finite number of which have a joint Gaussian distribution." GP is the most flexible class of non-parametric function estimators which can be interpreted as an infinite ANN, *Neil (1994).*

A GP can also be described as a multivariate Gaussian distribution over functions (instead of a scalar or vectors) and the general form can be written as in:

$$f(\boldsymbol{x}) \approx GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \text{ or } N(\mu, \Sigma) \tag{1}$$

where $m(\boldsymbol{x})$ or $\boldsymbol{\mu}$ is the mean function $E[f(\boldsymbol{x})]$ and $k(\boldsymbol{x},\boldsymbol{x}')$ or $\sum$ is the covariance function $E[(f(\boldsymbol{x})-m(\boldsymbol{x}))(f(\boldsymbol{x}')-m(\boldsymbol{x}'))]$. The mean function can be set to zero, and the covariance function used for this problem is the Squared Exponential (K_{SE}).

$$cov(y) = K_{SE}(\boldsymbol{x}, \boldsymbol{x}') + \sigma_n^2, \tag{2}$$

$$K_{SE}(\boldsymbol{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{x_d - x'_d}{l_d}\right)^2\right] \tag{3}$$

where σ_f^2 is the predictive variance, σ_n^2 is the noise variance, and l is the characteristic length-scale, D is the dimension of the input variables x, σ_f^2, σ_n^2 and l are also referred to as the hyperparameters.

The length-scale can be thought of as the distance you have to move in input space for the function value to change. In the present problem, the input is multi-dimensional and thus there is one length-scale for each of the input variables and the hyperparameters thus defined as: $\theta = \{l^1, l^2, \ldots, l^d, \sigma_f^2, \sigma_n^2\}$. The model is trained by optimizing the covariance function with respect to the hyperparameters by maximum likelihood also referred to as Automated Relevance Determination routine, ARD. This leaves one optimum length-scale for each of the input variables $l^1, l^2, \ldots, l^d$.

Predictions by GPR are found by making a joint distribution of the training target values $\boldsymbol{y}=[y_1,\ldots,y_N]$ and the test values $y_t(x_t)$, and from this the predictive function value and variance can be derived to (4) and the predictive variance σ_t^2 can be found as in the sum of the predicted covariance matrix $cov(f(x_t))$ and the variance σ_n^2 (5).

$$f(\boldsymbol{x}_t) = K(\boldsymbol{x}_t, \boldsymbol{x})\boldsymbol{\alpha}\boldsymbol{y}, \tag{4}$$

$$\sigma_t^2 = K(\boldsymbol{x}_t, \boldsymbol{x}_t) - K(\boldsymbol{x}_t, \boldsymbol{x})\boldsymbol{\alpha}K(\boldsymbol{x}, \boldsymbol{x}_t), \tag{5}$$

where $\boldsymbol{\alpha} = (K(\boldsymbol{x}, \boldsymbol{x}) + \sigma_n^2\boldsymbol{I})^{-1}.$

GPR is a fast method, but the complexity increases with the amount of input $O(N^3)$, so it is not appropriate for analysis of large data sets. The training data is explicitly used for prediction, so these can be computationally expensive.

One of the strengths of GPR is that from ARD, the length-scale is found, which determines how relevant each input variable is for the regression. Input variables with a small length-scale have a higher influence than the variables with high length-scales due to the short distance the input have to move in order to change the function. In the variable analysis all the length-scales are presented as the logarithm *log(l)* due to the large variation in the length-scales.

Furthermore the prediction variance is calculated for every prediction.

A detailed description of GPR and the specific software implementation used in this project is found in *Rasmussen and Williams (2006)* and the author's website (www.gaussianprocess.org).

2.1. Training

In order to use the data set most efficiently, it was divided into several training/test sets. The most efficient use is by training with all the data except one (*N-1*) and test with the single remaining input/output variable that is not used for training. This can be done alternately N times so all data points are used for test at a time. This is referred to as "Leave-One-Out" (LOO).

Prediction errors are calculated for each training/test set, and the mean value of the test error from all the test sets is referred to as the "Cross-validation error".
The input x_i^d was normalized by the mean and standard deviation: $\tilde{x}_i^d = \frac{x_i^d - \bar{x}^d}{std(x^d)}$. This was done in order to avoid too large variations in the input variables leading to large variations in the length-scale and hence the predictions. Since GPR with zero-mean is assumed, the output variable is centred around *0* with following normalization: $\tilde{y} = y_i - \bar{y}$

The initial guess of the hyperparameters occasionally resulted in local minima depending on the data length which was discovered by the prediction variance being very small or zero. To overcome this, the training/test set was also split up into *MS* parts accumulating to the full dataset (7). Each training was restarted with the final hyperparameters from the previous run in total MS-times as in (7).
After a few tests, it was concluded that it was sufficient to restart the training twice depending on the data size, i.e. three trainings of the GPR, subsequently two restarts was used for all the trainings.

$$\theta\left(\boldsymbol{x}_{ms_{i+1}}\right) = \theta\left(\boldsymbol{x}_{ms_i}\right) \tag{6}$$
$$\boldsymbol{x}_{ms_i} \in \boldsymbol{x}\left(0: i\,\frac{|\boldsymbol{x}|}{MS}\right) \tag{7}$$

Where $\boldsymbol{x}$ is the total input data set and $\boldsymbol{x}_{msi}$ is the multi start subset.

3. Data

Based on the traditional performance evaluation using empirical methods for predicting the propulsion power the critical input variables have been identified. As expected speed through the water, draught (fore and aft), the wind sped and direction, sea state and temperatures are amongst the most important input variables for estimating the propulsion power. The specific variables are listed below in order of the expected importance.

1. Draught midship [m]
2. Trim Ta-Tf (Draught aft- draught fore) [m]
3. Ship speed [knots]
4. Relative wind speed [m/s]
5. Relative wind direction [degree]
6. Wave height [m]
7. Relative wave direction [degree]
8. Water temperature [degree C]
9. Air temperature [degree C]

The output variable is either the propulsion power measured by a torsiometer or the specific fuel consumption.

3.1. Container ship data

Noon reports from five sister container ships were systematically collected for a period of up to 10 years. The logging periods are long and contain both dry-dockings and hull cleanings, which give an interesting insight into how these operations influence the performance. The data is well organized, purged of irrelevant data and seems to be very consistent, especially the manual observations of the wave height/direction and wind speed/direction.

Unfortunately the dimensions and ship names had to be kept confidential, and only the performance data was available. This made it impossible compare with traditional performance methods where ship specific data are necessary.

An overview of the data sets are given in Table 1 where the logging period until the first dry-docking is presented together with the total number of data points, number of dry dockings and hull cleanings within the period.

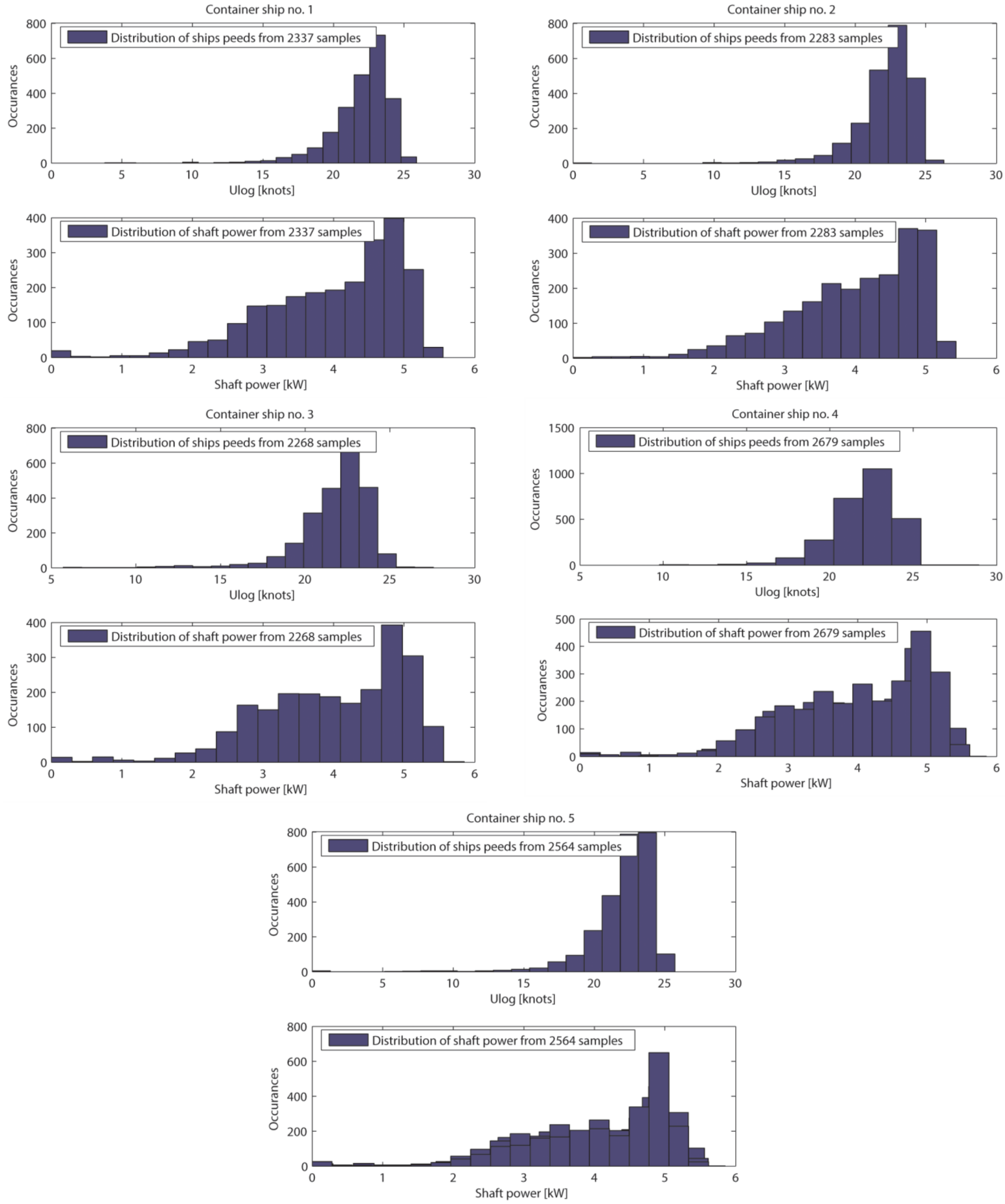

Fig. 1: Speed and power distribution for the five container ships

Table 1: Noon reports from the container ships

Ship ID	Period year to first docking	Total no. of NR	No. Docking	No. Hull cleaning
1	0-4.5	2337	2	7
2	0-4.5	2283	2	2
3	0-4.9	2268	1	3
4	0-5.0	2679	2	3
5	0-5.0	2564	2	4

Fig. 1 shows the distribution of the logged speed and shaft power, it indicates a narrow speed profile with a mean around 23 knots and with almost no occurrences of speeds out of the range of 20-25 knots. The power distribution was much broader which indicates that the power is adjusted to meet the speed for different draught, environmental conditions, etc.

In order to supplement the noon report data hindcast data. Hindcasts are weather information at a certain time and position in the past. The data is received by a tool developed for Seatrend (Performance Monitoring tool developed at FORCE Technology, www.force.dk) at FORCE Technology based on weather information from the NOAA (National Oceanic Atmospheric Administration, US Dept. of Commerce) data base. This adds several new variables as to the noon reports, but it also limits the number of data, since not all areas are covered by the NOAA database. Many of the smaller sea i.e. the Mediterranean, North sea, Baltic sea, are not included. This reduces number of data of approximately the half.

Table 2: Container ship input data

x	Data variable	Source	ID	Unit
1	Speed through water	Noon report	NRxls.Ulog	knots
2	Speed over ground	Noon report	NRxls.Uobs	knots
3	Sea water temperature	Noon report	NRxls.Tsw	deg
4	Mean draught (Ta+Tf)/2	Noon report	NRxls.Tm	m
5	Trim, Ta-Tf	Noon report	NRxls.Trim	m
6	True wind speed	Noon report	NRxls.WindSpeed	m/s
7	Relative wind direction	Noon report	NRxls.WindDir	deg
8	Average relative winds speed during report period	Hindcast	HC. Vrel	m/s
9	Average relative winds direction during report period	Hindcast	HC.gammarel	deg
10	Average significant wave height during report period	Hindcast	HC.mean.Hs	m
11	Average wave period during report period	Hindcast	HC.mean.Tp	s
12	Variance of the significant wave height during report period	Hindcast	HC.var.Hs	m^2
13	Variance of the wave period during report period	Hindcast	HC.var.Tp	s^2
14	Variance of the wave direction during report period	Hindcast	HC.var.Td	deg^2
15	Variance of the winds speed during report period	Hindcast	HC.var.Ws	m/s
16	Variance of the winds direction during report period	Hindcast	HC.var.gamma	deg
17	Report date and tim (Matlab numeric value)	Noon report	NRxls.UTC	numeric
18	Average winds speed during report period	Hindcast	HC.Ws	m/s
19	Average winds direction during report period	Hindcast	HC.gamma	deg
27	Sea state	Noon report	NRxls.SeaState	m
28	Relative sea direction	Noon report	NRxls.TrueRelativeSeaDirection	deg
36	Average shaft power	Noon report	NRxls.PropPower	kW

Since the weather observations are instantaneous values, usually taken around the reporting time, and other important variables are average values from the previous noon report time, it is not correct to use them together. It would have been ideal to have "noon reports" for every hour to increase the accuracy of the hindcast weather information. As this was not possible, hindcasts were made between every noon report, but with one-hour intervals and equivalent position, and then the average of the values between the present and the previous noon report were found. Similarly, the variance of the hindcasts for every noon report period was determined and made available for input to give more detailed weather information. The goal was to find weather that is equivalent to the effect spent in the same period, but since this is not possible, it is believed to be better than using only one observation (noon report time) to represent the past 24 hours.

Table 2 shows the list of the data used for the analysis.

4. Results

The evaluation was done by comparing the relative prediction errors of energy consumption. As described in above, cross-validation has been performed by Leave-One-Out. The cross-validation error $\overline{\omega_K}$ is the mean error of the mean ω_k from the k'th subsets of all the relative tests errors ω_{n_k}. For *LOO*, *K* is equal to the length of the total data set *N*.

Similarly the cross-validation value of the relative predictive standard deviations $\overline{\sigma_k}$ has been determined.

$$\omega_{n_k} = \frac{\widehat{EC}_{n_k} - EC_{n_k}}{EC_{n_k}} \tag{8}$$

$$\omega_k = \frac{1}{N_k} \sum_{n_k=1}^{N_k} \left|\omega_{n_k}\right| \tag{9}$$

$$\overline{\omega_K} = \frac{1}{K} \sum_{k=1}^{K} |\omega_k| \tag{10}$$

$$\sigma_k = \sqrt{\frac{1}{N_k} \sum_{n_k=1}^{N_k} \left(\omega_{n_k} - \omega_k\right)^2} \tag{11}$$

$$\overline{\sigma_k} = \frac{1}{K} \sum_{k=1}^{K} \sigma_k \tag{12}$$

Where:
$\widehat{EC}_{n_k}$ is the predicted energy consumption of the n'th input data for data subset *k*
EC is the measured energy consumption of the n'th input data for data subset *k*
N_k is the number of data points within each subset
K is the number of subset, *K*=*N* for for LOO, *N* being the number data in the full data set
ω_{n_k} is the relative prediction error of the n'th input data for data subset *k*
ω_k is the average of the relative prediction errors for the *k*'th subset
$\overline{\omega_K}$ is average of the relative prediction errors

4.1. Results from Container ship #1

The data from Container ship #1 was used to make initial tests in order to determine the best combinations of the input variables. Table 3 shows the input variable combinations tested, with the cross validation errors and standard deviations listed in Table 4. From that the best prediction of the energy consumption (the lowest prediction error) is found using the input combination *13*, *14* and *21*, all having relative prediction error around *4%* they all have in common that they are based on noon reports and hindcasts, and that the time is included as a variable. Omitting the time increase the prediction errors of up to *2.4%* (input combination *14* and *15*).

Using input data from the noon reports only the best prediction error is found to *4.9%* with input variable combination *18*, where all available noon report input, except the speed over ground, is used including the time variable. Without the *time* variable the input combination *18* is identical to *17* which gives significant higher errors of *6.7%*, for the noon reports data without *time* the rather simple input combination *9*, including only the logged speed, sea water temperature and the mean draught, yields the best prediction of an error of *5.5%*.

Table 3: Input variable setup combination for Noon Report data of the containership data

	Input combination ID	3	9	11	12	13	14	15	17	18	20	21
1	NR.Ulog	x	x	x	x	x	x	X	x	x	x	x
2	NR.Uobs				x	x	x	X			x	x
3	NR.Tsw	x	x	x	x	x	x	x	x	x	x	x
4	NR.Tm	x	x	x	x	x	x	x	x	x	x	x
5	NR.Trim	x		x	x	x	x	x	x	x	x	x
6	NR.True_wind_speed_m_s	x		x	x	x	x	x	x	x	x	x
7	NR.True_relative_wind_direction_deg	x		x	x	x	x	x	x	x	x	x
8	HC.Vrel				x	x					x	x
9	HC.gammarel				x	x					x	x
10	HC.mean.Hs				x	x	x	x			x	x
11	HC.mean.Tp				x	x	x	x			x	x
12	HC.var.Hs				x	x	x	x			x	x
13	HC.var.Tp				x	x	x	x			x	x
14	HC.var.Td				x	x	x	x			x	x
15	HC.var.Ws				x	x	x	x			x	x
16	HC.var.gamma				x	x	x	x			x	x
17	NR.UTC	**x**				**x**	**x**			**x**		**x**
18	HC.Ws						x	x				
19	HC.gamma						x	x				
27	NR.Sea_state_m								x	x	x	x
28	NR.True_relative_sea_direction_deg								x	x	x	x

Table 4: Relative cross validations errors predictive standard deviations.

Input variable setup	$\overline{\omega_K}$ %	$\overline{\sigma_k}$ %
3	5.81	6.11
9	5.45	6.76
11	8.01	9.59
12	5.40	5.22
13	3.96	3.99
14	4.00	3.96
15	6.43	6.21
17	7.64	7.81
18	4.92	5.45
20	4.52	5.05
21	3.98	4.03

Fig. 2 shows the relative prediction error with error bars based on the predicted standard deviation. It is seen how the error bars generally increase with high prediction errors or areas with sparse data density.

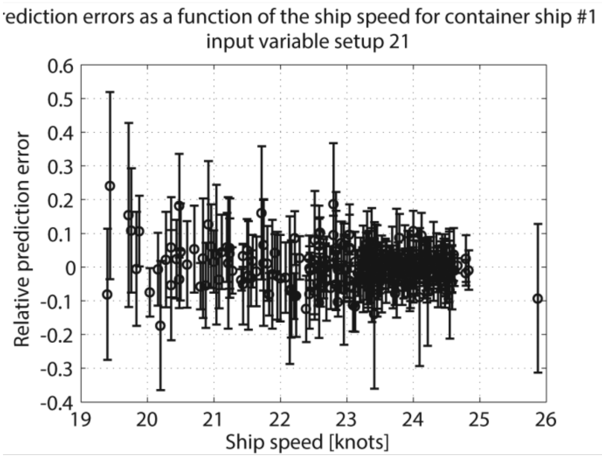

Fig. 2: Relative prediction errors with error from the predicted standard deviation (input 21, 0-4.5 years)

In order to validate the prediction errors an Artificial Neural Network (ANN) has been trained and tested with the same setups as the GPR. The ANN is a feed forward network as described in *Pedersen and Larsen (2009)*. Comparing the prediction methods in Fig. 3 show that the cross validation errors for GPR in most cases are lower or in the same order of magnitude as ANN, was satisfactory for the further study of GPR with performance data.

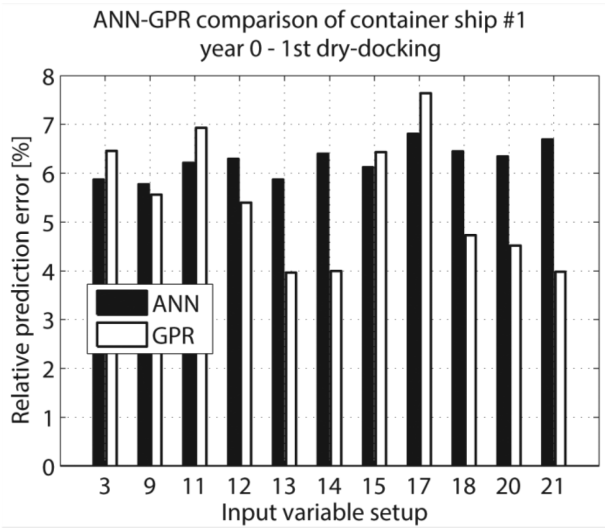

Fig. 3: Comparison of prediction methods GPR and ANN

4.2. Results from Container ship #1-#5

Based on the findings in the previous sections the remaining 4 container ships was trained and evaluated similarly as for container ship *#1*, i.e. using input variable combination *9, 18, 20* and *21*, in order to represent data with and without hindcast data, and with and without time as a variable. Furthermore each of the data set was split into five subset: one for the first year, another for the first two years of data, a third one for the first three year and so forth until the first dry-docking. Fig. 4 show that for the data with hindcast input (20 and 21) is only affected very little by the increasing number of data and the end up with prediction errors between *4* and 6*%*. For the data set using noon reports without the hindcast data there seems to be a significant benefit using more data, since the prediction errors decrease with time for most of the data set.

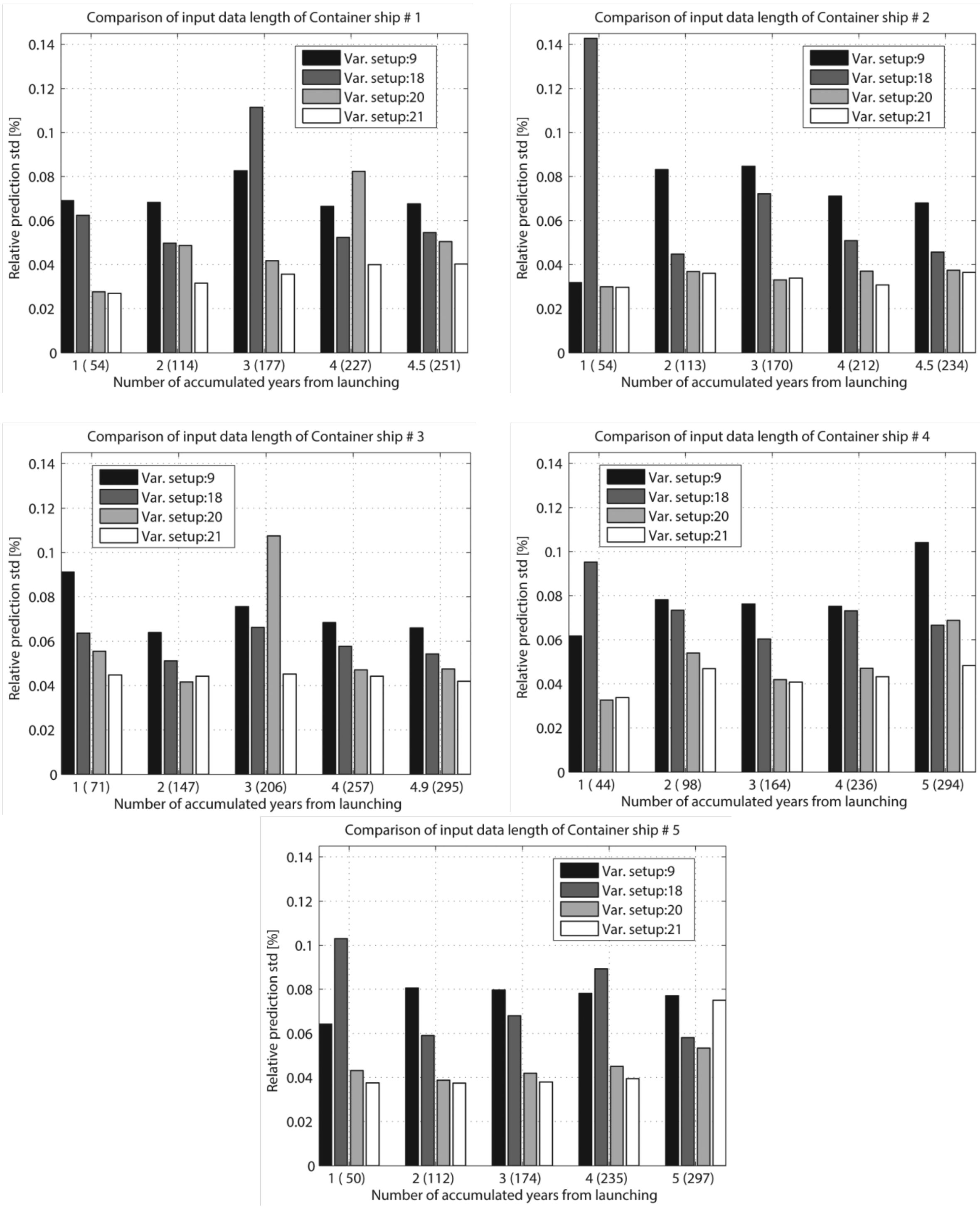

Fig. 4: Cross-validation errors of container ship #1-#5

5. Analysis of the Input-Output Variables

In order to evaluate the influence of each of the input variables, different trainings have been performed with Container ship #1. The length-scales for each of the input variables have then been evaluated for the different training setups. Initially the first year of container ship #1 data was used with all the relevant input variables including the hindcast data, see Table 1. In Fig. 2 the length-scales are illustrated in a bar diagram for the input variable setup where time is included in variable setup *21* and the one without time *20* after training the data from the launching until the first dry-docking. Here all the available input variables are used, i.e. many data points are dismissed because of lacking

hindcast availability. For both input *20* and *21* the figure show a clear trend of the logged speed has the highest influence (input 1). The speed over the ground (input 2) is slightly higher together with hindcast wind speed and direction, and significant wave height have an equally influence. The hindcast wave period is generally less important. When the time (input 17) is introduced in variable setup *21*, it has a significant influence itself, but is also influence the influence of other variables. The most dramatic drop between variable setup *20* and *21* is the seawater temperature (input 3). The time was introduced as an estimator for the hull fouling since the propulsion performance is expected to drop over time of this. The significant influence of the time variable confirms the strong relation between the expected to change in the performance over time.
The draught and trim (input *4* and *5*) have a smaller impact than expected but this might be due to only small variation for these variables in the current data set.

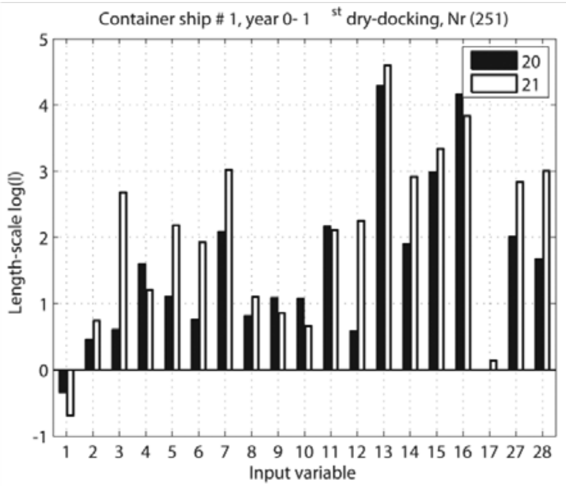

Fig. 5: Logarithmic length-scales of input variable setups 20 and 21 (with time as a variable)

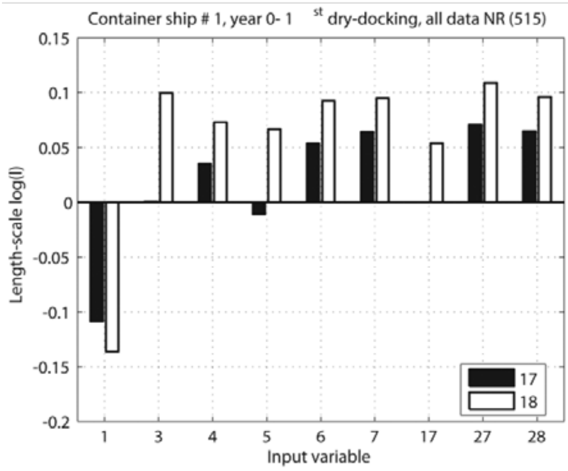

Fig. 6: Logarithmic length-scales of input variable setups 17 and 18(with time as a variable), using only noon report data.

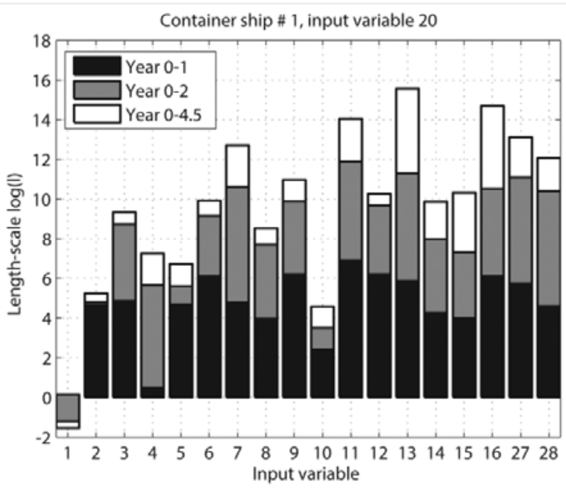

Fig. 7: The accumulated length-scales of input variable setup 20 (without *time*), trained for the three periods of time 0-1, 0-2, 0-4.5 years (4.5 year is the time of the first dry docking).

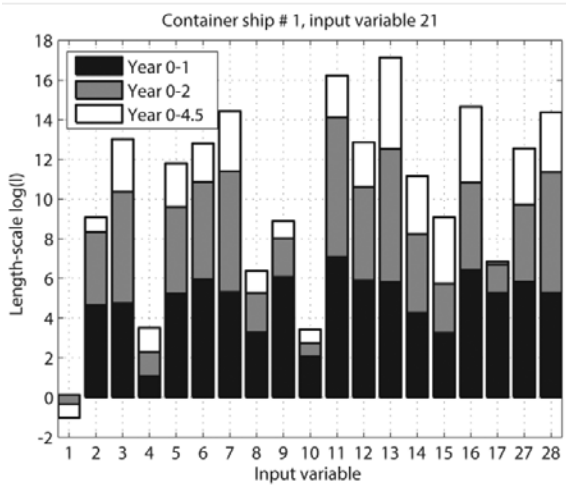

Fig. 8: The accumulated length-scales of input variable setup 21 (including time – input variable 17), trained for the three periods of time 0-1, 0-2, 0-4.5 years (4.5 year is the time of the first dry docking).

Fig. 6 is presenting the length-scales based on the training of data with all noon report data available and no hindcast data, input variable setup *17* and *18*, without and with the time as input variable. Again it is confirmed that the ship speed is the variable with the most significant influence with very small length-scales for both the input combinations. For the variable setup 18, with time, the remaining input variables are all with the same order of magnitude. For input variable setup 17, the seawater temperature (input *3*) and the trim (input *5*) indicates to have more influence than the remaining variables.

In order asses the trend for shorter logging periods the data was trained also trained for 0-1 year, 0-2 years and 0-until the first dry-docking about 4.5 years. This has been performed for the variable setups 20 and 21 and the accumulated bars are shown in Fig. 7 and Fig. 8. The figures show that except for a few incidents the length-scales becomes shorter for longer data series.

6. Detection of performance trend

The overall propulsion performance of a ship is expected to decrease over time mainly due to fouling of the hull and propeller. The change in performance can be determined by comparing the actual measured energy consumption *EC* propulsion power or fuel consumption with a calculated or predicted energy consumption $\widehat{EC}$ of how the vessel should be able to perform.
With *GPR*, $\widehat{EC}$ predictions can be based on training of a previous period of time. The difference between the predicted and actual values, i.e. the prediction error, is thus a measure of the vessel propulsion performance.

In the analysis, the relative prediction error $\omega = \left(\widehat{EC} - EC\right)/EC$ was used to evaluate the behaviour of the performance. The actual energy consumption *EC* is expected to increase over time due to the fouling, while ω is expected to decrease since the predicted values are based on the training data which are assumed not to be affected by fouling. It is thus desirable to train on the shortest possible period of time in order to limit the effect of a trend in the training data. Yet the training set should include a reasonable variation in the input variables.

The training was performed on the entire training set and tested on the remaining data set resulting in a predicted energy consumption $\widehat{EC}$. The change in propulsion performance due to fouling is assumed to be linear, and the relative prediction error ω can be estimated as a linear function of the time. In order to account for the predicted variance, a "weighted least square" regression is performed on ω as a function of the time where the weights, $\boldsymbol{w}$, are the inverse relative predicted variance $(1/\sigma_R^2)$. This gives the prediction errors with large variance less influence on the linear regression model. As discussed previously the majority of the larger prediction errors also have a large standard deviation σ_R^2, so letting the inverse variance being the weights makes these prediction errors less relevant.

Weighted least square regression is similar to normal least square, but with the residual $\left(\omega_i - f(x,\alpha)\right)$ multiplied by the weights, $\boldsymbol{w}_{ii}$ in the sum of square errors (13) which is minimized with respect to α. The *x*-axis is represented by the time t_i.

$$E(\alpha) = \sum_{i=1}^{n} \boldsymbol{w}_{ii}\left(\omega_i - f(t_i,\alpha)\right)^2 \tag{13}$$

The function values of the performance trend can be define as a Vessel Performance Index, *VPI:*

$$VPI(t) = \alpha t + \beta \tag{14}$$

The performance trend is only continuous in periods without external disturbances that change the propulsion performance. External disturbances can be known or unknown. The known disturbances are e.g. dry-docking and hull and propeller cleaning, and unknown disturbances could e.g. be sudden unknown damage of the propeller or rudder.

All known disturbances or events are available for the container ships and the trend was thus found between the known events, including dry-docking DD, hull cleaning HCL and propeller cleaning PCL for the container ship. All these events are expected to increase the relative prediction error ω, because the energy consumption is expected to drop. But for the hull and propeller cleaning, this effect can sometimes be difficult to detect due to its limited impact on the relatively large data scatter. Therefore the trend detection has been performed both between all the known events and the dry-dockings only.

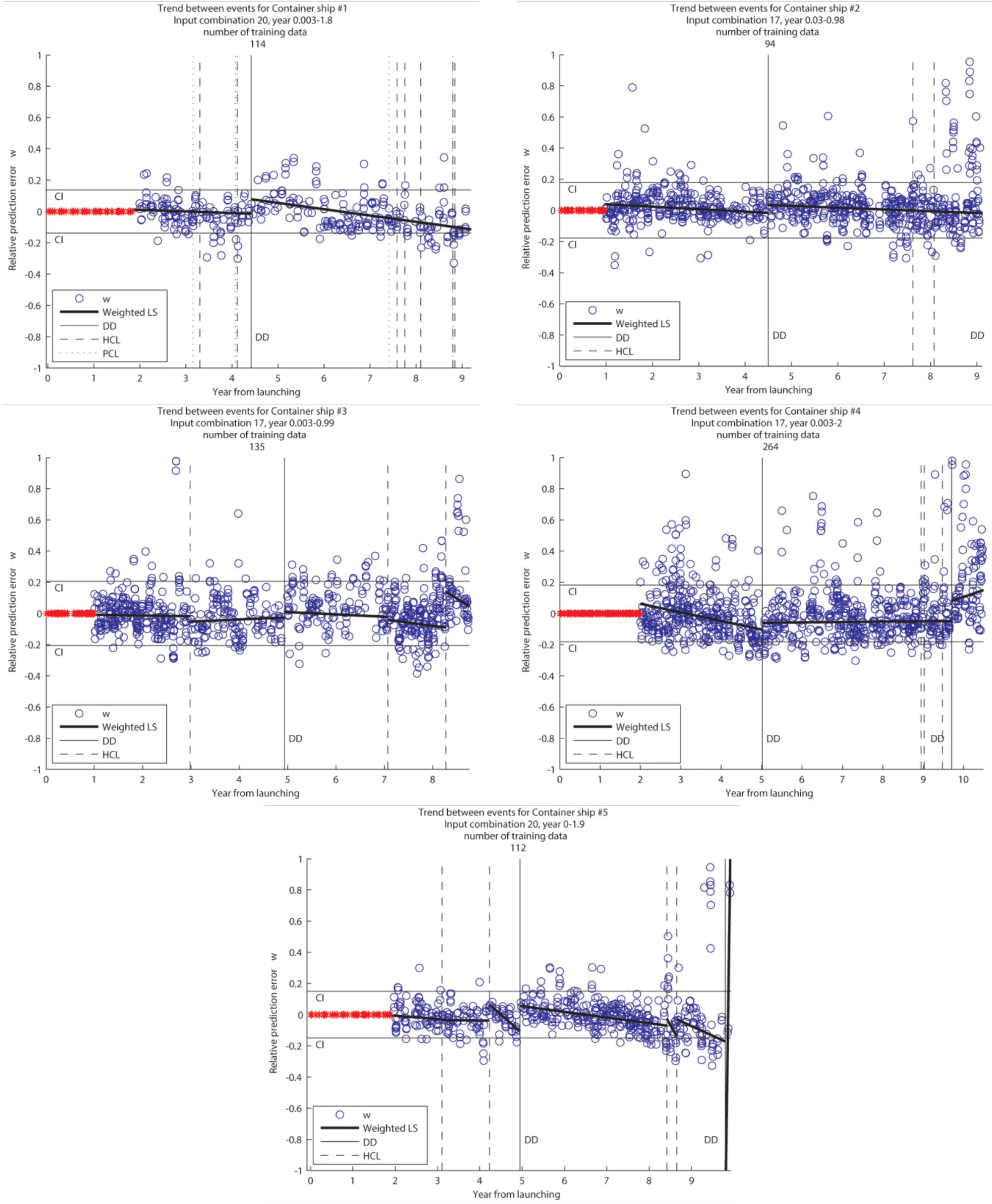

Fig. 9: Performance trends for container ship #1-#5

The time variable in input variable setups *18* and *21* increases linearly with time. This means that if time has relevance for the regression model, it will be dominant for prediction far into the future. Input combinations including the time were thus inapplicable for the trend detection. The best input variable setups without time was previously found to be *17* and *20*, with *17* based exclusively on noon report data, and *20* including the hindcast data.

Given the considerations described above, $\widehat{EC}$ for the five container ships were predicted based on two training periods, one and two years from launching, two input variable setups, *17* and *20*. The trends were detected only between all the known events and the dry-dockings.

The performance trends of container ship #1-#5 are illustrated in Fig. 9. For container ship #1 the intermediate events with short intervals gave too few data make reasonable trends from and subsequently only the dry-dockings was used, which gave a good picture of how the performance increase after a dry-docking, but afterwards drops with a higher rate than before. Container ship #2 has fewer events and thus more continuous development where only a small increase in the performance is detected.
Container ship #3 have more events, but with reasonable time between and it thus become easier develop trend for after the first hull cleaning there is a drop in the performance, indicating that the hull cleaning have actually had a negative effect on the propulsion performance. The following dry-docking has a small positive effect.
Container ship #5 show a very clear trend of the second hull-cleaning having a temporary significant positive effect, but the slope of the performance trend drops dramatically, indicating that the anti fouling paint might have been damaged. The following dry-docking is bringing the vessel back to state which is actually better than at new.

7. Discussion

With the methods described above, it is possible to detect the general trends of the change in the performance over time without any predefined definitions of the vessel. Using a combination of noon report data and hindcast data increases the prediction performance significant, even though it reduces the number of inputs.

The dry-docking, hull and propeller cleanings do not always have the intended effect. The hull cleanings may have a positive immediate effect, but the long-term trend can be negative. It might also be that the event has no immediate effect, but the long-term effect can be beneficial. In order to evaluate this, detailed information about the event is needed such as what part of the ship was cleaned and what equipment was used.

In general, the dry-dockings had a more consistent effect with a positive change in VPI after the docking, but the slope afterwards varied from being steeper than before, as for Container ship #1, to being flatter than the previous trend, as for Container ship #4.

References

BHATTACHARYYA, R., (1978), *Dynamics of Marine Vehicles*, Ocean Engineering, Wiley-Interscience, 1st Edition

HARVALD, S.A. (1983), *Resistance and Propulsion of Ships*, John Wiley & Sons

HOLTROP, J. (1984), A *statistical re-analysis of resistance and propulsion data*. International Shipbuilding Progress, p. 272.

ISHERWOOD, R.M. (1972), *Wind Resistance of Merchant Ships*, Proceedings of Royal Institution of Naval Architects.

NEIL, R. (1994), *Priors for Infinite Networks*, Technical Report CRG-TR-94-1, Department of Computer Science, University of Toronto

PEDERSEN, B.P.; LARSEN, J. (2009), *Prediction of Full-scale Propulsion Power Using Artificial Neural Networks*, 8th Int. Conf. Computer and IT Appl. Maritime Ind, Budapest, pp. 537-550

RASMUSSEN, C.; WILLIAMS, K. I. (2006), *Gaussian Processes for Machine Learning*, The MIT Presse, London, England, 1. Edition