



Dictionary-based experiment design for estimation of marine models

Fredrik Ljungberg^{a,*}, Jonas Linder^b, Martin Enqvist^a, Kalevi Tervo^c

^a Linköping University, Linköping, Sweden

^b ABB Corporate Research, Västerås, Sweden

^c ABB OY, Industrial Automation, Marine and Ports, Helsinki, Finland

ARTICLE INFO

Keywords:

System identification
Experiment design
Nonlinear models
Greybox modeling
Marine systems

ABSTRACT

In this work, experiment design for marine vessels is explored. A dictionary-based approach is used, *i.e.*, a systematic way of choosing the most informative combination of independent experiments out of a predefined set of candidates. This idea is quite general but is here tailored to an instrumental variable (IV) estimator with zero-mean instruments. This type of estimator is well-suited to deal with parameter estimation for second-order modulus models, which is a class of models often used to describe motion of marine vessels. The method is evaluated using both simulated and real data, the latter from a small model ship as well as from a full-scale vessel. Further, a standard motion-planning problem is modified to account for the prior-made choice of information-optimal sub-experiments, which makes it possible to obtain a plan for the complete experiment in the form of a feasible trajectory.

1. Introduction

Ship motion and control have engaged researchers for at least a century, see for example [Minorsky \(1922\)](#) for an early reference. Even if the controller concept has hardly changed since then, there have been great advances made. One major difference in modern-time automatic steering of ships is that many control methods are model based. Therefore, having accurate simulation models available is becoming increasingly important as ships are becoming more automated. There are essentially two classes of experiments that can be employed to generate data that such models can be based on. Firstly, there are experiments where motion-yielding forces are produced by some external equipment, such as towing-tank tests. Very accurate parameter estimates can be obtained in this way, but the experiments are usually expensive and time-consuming. Moreover, since a down-scaled replica is often used in this type of experiment, modeling errors due to mismatches between the scale model and the full-size vessel are common ([Yoon & Rhee, 2003](#)). An often cheaper way of obtaining models is through system identification, *i.e.*, by modeling based on data collected from tests where the vessel's own propulsion system is used to generate the motion.

It is well known that the choice of input signal during the data acquisition is a significant factor for the result of the parameter estimation in system identification, see for example [Ljung \(1999\)](#). For a data-collection experiment, there are both practical and statistical considerations to be made. The practical considerations relate to physical limitations, for example, traffic regulations, actuator saturations

and avoiding of obstacles such as shallow water, islands and other vessels. The statistical aspect of the design relates to maximizing the information gained from the experiment and can often be studied in a more general and abstract fashion. Input design for general linear time-invariant systems has been studied for some time and is by now a fairly well-understood topic, see for example [Goodwin and Payne \(1977\)](#) for an early reference and [Gevers, Bombois, Hildebrand, and Solari \(2011\)](#) for a more recent overview. However, accurate ship modeling often requires that nonlinear hydrodynamic forces and moments are taken into account if the model should cover the ship's full speed envelope. Experiment design for nonlinear systems is not yet as well understood, and most previous works have dealt with experiment design for certain subclasses of nonlinear systems. In an early work regarding experiment design for nonlinear systems, [Hjalmarsson and Mårtensson \(2007\)](#) discussed a two-step approach to solve the input-design problem, one step in which the optimal probability density function of the input is computed and one in which the signal is realized based on the found probability density. This idea is inspired by how experiment design is commonly performed for linear systems and was extended in [Larsson, Hjalmarsson, and Rojas \(2010\)](#). There, input design for general nonlinear finite-input-response (NFIR) models was considered, a model class which was also dealt with in [De Cock, Gevers, and Schoukens \(2016\)](#). Another often studied subclass of nonlinear systems is interconnections of linear systems and static nonlinearities. Experiment design for such systems was explored in [Mahata, Schoukens, and De Cock \(2016\)](#) and [Vincent, Novara, Hsu, and Poolla \(2010\)](#). Further, for static nonlinear

* Corresponding author.

E-mail addresses: fredrik.ljungberg@liu.se (F. Ljungberg), jonas.x.linder@se.abb.com (J. Linder), martin.enqvist@liu.se (M. Enqvist), kalevi.tervo@fi.abb.com (K. Tervo).

<https://doi.org/10.1016/j.conengprac.2023.105528>

Received 3 October 2022; Received in revised form 6 February 2023; Accepted 6 April 2023

Available online 13 April 2023

0967-0661/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

systems, which are linear in the parameters, the input design can be formulated as a convex optimization problem. This property was used in [Forgione, Bombois, Van den Hof, and Hjalmarsson \(2014\)](#) to perform input design for dynamical systems by considering a fixed number of amplitude levels of the input signal. A similar approach was taken in [Valenzuela, Rojas, and Hjalmarsson \(2015\)](#), where input design for nonlinear output-error (NOE) models was considered using graph theory. In [Gopaluni, Schön, and Wills \(2011\)](#), a particle-filter approach was used and in [Umenberger and Schön \(2019\)](#), the input-design problem was transformed to an optimal control problem using tools from statistical inference. Both these approaches consider fairly general classes of nonlinear models but rely on stochastic approximation techniques and require initial guesses for the input signals to deal with the resulting non-convex formulations.

For most model structures, the optimal experiment design will depend on the parameters of the system to be identified, *i.e.*, the quantities that it is desired to estimate. This can seem a bit restrictive, but for many applications there is some nominal model available that the experiment design can be based on. In other cases a multi-stage approach can be employed, where an estimation step and a design step are alternated iteratively, as in [Gerencsér, Hjalmarsson, and Mårtensson \(2009\)](#). A third option is a robust design, where the experiment-design criterion is optimized over a distribution of possible parameter values. This last approach was explored for nonlinear models in [Valenzuela, Dahlin, Rojas, and Schön \(2017\)](#).

Parameter estimation for nonlinear models of industrial manipulators was discussed in [Wernholt and Moberg \(2011\)](#). A two-step solution was used, where frequency-response estimates for local linear models were first estimated for a chosen number of operating points. Suitable values for parameters of a nonlinear model were then found in a second step by minimizing the discrepancy between these locally valid frequency-response estimates and the parametric frequency response of the nonlinear model. Experiment design for such an approach was explored in [Wernholt and Löfberg \(2007\)](#). In this case, the design reduces to a selection between a predefined set of operating points, where tools from experiment design for linear systems can be applied for finding a suitable input signal for each operating point. This is appealing because the Fisher information matrix of a sequence of sub-experiments can be written as a convex combination of the Fisher information matrices of the individual sub-experiments. As a result, the optimal experiment-design problem is convex and can be solved efficiently.

The importance of collecting informative data has also been stressed in works regarding marine modeling. In [Blanke and Knudsen \(1999, 2006\)](#), surface vessels were considered and the informativity gains from zig-zag, circle and spiral maneuvers as well as telegraph input signals, were compared. Further, an optimal zig-zag maneuver was found in [Yoon and Rhee \(2003\)](#) by empirically trying out different candidates of varying amplitude and frequency. More recently, an approach for finding optimized input signals for nonlinear models of underwater vehicles was presented in [Nouri, Valadi, and Asgharian \(2018\)](#). Their proposed idea was to parameterize a step-wise changing signal in terms of a finite number of amplitude levels and dwell times, *i.e.*, an approach similar to [Forgione et al. \(2014\)](#) and [Valenzuela et al. \(2015\)](#). A comparable approach was applied for surface vessels in [Wang, Soares, and Zou \(2020\)](#). Both [Nouri et al. \(2018\)](#) and [Wang et al. \(2020\)](#) obtain optimization problems that can be solved in reasonable time by imposing restrictions on the input signal, such as having the propeller thrust fixed and only varying the rudder angle. In both cases, the resulting optimization problems are solved using numerical methods and the optimized input signals are in simulation experiments shown to give more accurate parameter estimates than zig-zag maneuvers. Lastly, the benefits of having the vessel's own propulsion system generate motion during data collection were shown in a comparative study between model estimation following towing tests and self-actuated tests for an underwater vehicle in [Lack, Rentzow, and Jeansch \(2019\)](#).

As seen in the references above, data for estimation of marine models are often collected using standard maneuvers, such as turning circles, spiral motions and zig-zag tests. These maneuvers were not initially developed for generating informative data but for evaluating a ship's maneuvering capabilities ([Wang et al., 2020](#)). A standard maneuver alone does not necessarily excite all relevant system modes, but when multiple maneuvers are performed in sequence, informative data can often be obtained. Based on this observation and inspired by [Wernholt and Löfberg \(2007\)](#), this work evaluates a dictionary-based approach for experiment design of marine vessels, *i.e.*, a systematic way of choosing the most informative combination of independent sub-experiments out of a predefined set of candidates. This idea is tailored to an instrumental variable (IV) estimator, which due to high accuracy under general disturbance conditions has been shown to work well for models of marine vessels in [Ljungberg and Enqvist \(2022\)](#). There, it was noted that choosing instruments with zero mean can be of importance for an IV estimator when applied to such models. Consequently, a technique to account for this in the experiment design is suggested. The proposed method is evaluated in an experimental study using both simulated and real data from a scale ship as well as a full-size marine vessel. Lastly, modifications are made to a standard motion-planning problem to account for the prior-made choice of information-optimal sub-experiments. This makes it possible to obtain a plan for the complete experiment in the form of a feasible trajectory and the potential of the idea is shown in a simulation example, where the dictionary-based method for choosing a set of sub-experiments is supplemented with a lattice-based motion planner.

One contribution of this work is the adaptation of the dictionary-based approach in [Wernholt and Löfberg \(2007\)](#) to time-domain estimation using an IV estimator with zero-mean instruments. Another contribution is the observation that there are multiple conceivable ways of subtracting the instrument mean when data is collected from different sub-experiments and that this choice is important for the experiment design. Additionally, the modifications to the motion-planning problem for (indirectly) taking informativity into account are novel.

In the following, some brief theoretical preliminaries are given in Section 2. After that, the considered dictionary-based approach for experiment design is outlined in Section 3. This approach is then adapted to work well for marine models in Section 4. Following this, the results from experimental work are shown in Section 5. Further, the method for generating a plan for the complete experiment in the form of a feasible trajectory is presented in Section 6. Lastly, the paper is concluded in Section 7.

2. Problem formulation and preliminaries

Accurate ship modeling requires that nonlinear hydrodynamic forces and moments are taken into account if the model should cover the ship's full speed envelope. To this end, consider a nonlinear discrete-time state-space system

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k), \boldsymbol{\theta}_0) + \mathbf{w}(k), \quad (1a)$$

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{e}(k), \quad (1b)$$

where $\mathbf{u}(k) \in \mathbb{R}^{n_u}$ is a known input signal, $\mathbf{x}(k) \in \mathbb{R}^{n_x}$ is a vector consisting of the latent system states, all of which are measured directly (with noise) and collected in the output vector $\mathbf{y}(k) \in \mathbb{R}^{n_x}$. Moreover, $\mathbf{e}(k) \in \mathbb{R}^{n_x}$ and $\mathbf{w}(k) \in \mathbb{R}^{n_x}$ are disturbance signals and $\boldsymbol{\theta}_0 \in \mathbb{R}^{n_\theta}$ is a vector of parameters, which is assumed to not vary over time.

2.1. Instrumental variable estimator

Consider a predictor model on regression form

$$\hat{\mathbf{y}}(k, \boldsymbol{\theta}) = \boldsymbol{\Phi}^T(k) \boldsymbol{\theta}. \quad (2)$$

Here $\Phi(k) \in \mathbb{R}^{n_\theta \times n_x}$ is a function of system inputs and outputs. The IV estimate of the model parameters is defined as the solution to the optimization problem

$$\hat{\theta}_N^{IV} = \underset{\theta}{\operatorname{argmin}} V_N^{IV}(\theta), \quad (3)$$

where the cost function is

$$V_N^{IV}(\theta) = \left\| \frac{1}{N} \sum_{k=1}^N Z(k) (y(k) - \Phi^T(k)\theta) \right\|_2^2. \quad (4)$$

The instrument matrix, $Z(k) \in \mathbb{R}^{n_\theta \times n_x}$, plays an important role for the accuracy of the resulting models and should ideally be correlated with the system states but uncorrelated with the disturbances. The consistency of an IV estimator was shown in Ljungberg and Enqvist (2022) for second-order modulus models, *i.e.*, models with signed-square regressors, such as $y(k)|y(k)|$. This model class is often used for describing motion of marine vessels. Consequently, the IV estimator will be employed in this work as well.

3. Basic method

If the system can be fully described by the model structure, it is possible to evaluate the usefulness of an estimator based on the bias and covariance of the parameter estimates. In this paper, the focus is on the covariance since the consistency of the estimator was explored in Ljungberg and Enqvist (2022). A common starting point when formulating an experiment-design problem for minimizing the variance is to assume that the estimator is asymptotically efficient, which means that the covariance matrix of the estimated parameters will converge to the inverse of the Fisher information matrix as the amount of estimation data increases, see for example Ljung (1999). In this case, making the covariance matrix as small as possible is equivalent to making the information matrix as large as possible, at least asymptotically. This is appealing because it is usually easier to obtain an expression for the information matrix, which to a large extent depends on the Hessian of the cost function, in comparison to directly finding an analytical expression for the covariance matrix. Despite the fact that the IV method gives increased possibilities of obtaining consistency, IV estimators are generally not asymptotically efficient. Therefore, the conventional arguments in terms of Fisher information do not apply to the estimators studied in this work. However, maximizing the Hessian of the cost function with respect to the parameters is useful from a practical point of view since it makes the output of the estimator sensitive to changes in the parameters. Thereby, performing experiments in such a way should generally give more distinct optima to the resulting parameter estimation problems. To this end, let

$$G(N) \triangleq \frac{1}{2} \frac{\partial^2}{\partial \theta^2} V_N^{IV}(\theta), \quad (5)$$

denote the Hessian of the cost function and note that

$$G(N) = \left[\frac{1}{N} \sum_{k=1}^N \Phi(k) Z^T(k) \right] \left[\frac{1}{N} \sum_{k=1}^N \Phi(k) Z^T(k) \right]^T.$$

Even if this is technically not the Fisher information matrix it will subsequently be referred to as the information matrix. The fact that maximizing $G(N)$ reduces the uncertainty of the parameter estimates is shown with a simulation example in Section 5.1.

Ideally, the optimization problem

$$\begin{aligned} \mathbf{u}^* &= \underset{u(1), \dots, u(N)}{\operatorname{argmax}} \quad \ell(G(N)) \\ \text{s.t.} \quad & u(1), \dots, u(N) \in \text{Feasible input,} \\ & y(1), \dots, y(N) \in \text{System dynamics,} \end{aligned}$$

would be solved for some scalar function $\ell(\cdot)$. However, this is generally a non-convex problem and solving it would require a good initial guess for the input signal. Such an initial guess is not that easy to obtain for

marine vessels because they are often equipped with multiple thrusters, which means that there are many degrees of freedom (DOF) in the input design. A simpler alternative is to use a dictionary-based approach where a sub-optimal input sequence is constructed by selecting the most informative combination of input signals out of a predefined set. An important step of such an approach is the selection of signals in the dictionary because if the dictionary does not include candidate signals that can be combined to obtain informative data, the optimized design will not be informative either.

Here, it is assumed that candidate signals with this property exist and that they are represented by the set

$$\mathcal{U} = \{u_1, \dots, u_Q\}. \quad (6)$$

Subsequently, the motion trajectories obtained when applying these will be referred to as experiment primitives. Moreover, introduce the variable

$$\tilde{\Phi}(k) = \Phi(k) Z^T(k), \quad (7)$$

for simplified notation and let

$$G_q(N_q) = \left[\frac{1}{N_q} \sum_{k=1}^{N_q} \tilde{\Phi}_q(k) \right] \left[\frac{1}{N_q} \sum_{k=1}^{N_q} \tilde{\Phi}_q^T(k) \right], \quad (8)$$

be the Hessian of the cost function when using the candidate signal $u_q(k)$ and instrument matrix $Z_q(k)$ for N_q time steps. If all the candidate input signals are applied in sequence, the total information gain is

$$G(N) = \Gamma(N) \Gamma^T(N), \quad (9)$$

where

$$\Gamma(N) = \frac{1}{N} \left(\sum_{k=1}^{N_1} \tilde{\Phi}_1(k) + \dots + \sum_{k=1}^{N_Q} \tilde{\Phi}_Q(k) \right), \quad (10)$$

and where $N = N_1 + \dots + N_Q$, *i.e.*, the total experiment time. Since the Hessian is a matrix it is necessary to choose some scalar measure of it to optimize. There are multiple different such criteria that can be used, some of which are discussed in Mehra (1974). One choice is to maximize the determinant. This is called a D-optimal design and will be used in this work. Geometrically, a D-optimal design is comparable to minimizing the uncertainty volume of the estimated parameters.

Further, the matrix $\Gamma(N)$ defined above has the same number of rows and columns. Consequently, it holds that

$$\det(G(N)) = \det(\Gamma(N) \Gamma^T(N)) = \det(\Gamma(N))^2. \quad (11)$$

and thereby that maximizing the determinant of $G(N)$ is equivalent to maximizing the determinant of $\Gamma(N)$ if $\Gamma(N) \geq 0$. Since $\Gamma(N) \geq 0$ cannot be guaranteed in a general case, the maximum of $|\det(\Gamma(N))|$ will be sought instead. Now, let

$$\Gamma_q(N_q) = \frac{1}{N_q} \sum_{k=1}^{N_q} \tilde{\Phi}_q(k), \quad (12)$$

be the information gain from using $u_q(k)$ and $Z_q(k)$. If stationary signals are considered, it will asymptotically be the case that

$$\lim_{N_q \rightarrow \infty} \frac{1}{N_q} \sum_{k=1}^{N_q} \tilde{\Phi}_q(k) \triangleq \bar{\Gamma}_q, \quad (13)$$

and consequently it holds that

$$N_q \frac{1}{N_q} \sum_{k=1}^{N_q} \tilde{\Phi}_q(k) \approx N_q \bar{\Gamma}_q, \quad (14)$$

for significantly large N_q . Under assumption that each candidate signal is applied for a sufficiently long time, it is therefore reasonable to

consider the optimization problem

$$\begin{aligned} (N_1^*, \dots, N_Q^*) = \underset{N_1, \dots, N_Q}{\operatorname{argmax}} \quad & \log |\det(\mathbf{T})| \\ \text{s.t.} \quad & \mathbf{T} = N_1 \bar{\mathbf{F}}_1 + \dots + N_Q \bar{\mathbf{F}}_Q, \\ & N_1 + \dots + N_Q = N, \\ & N_q \geq 0, \quad q = 1, \dots, Q, \end{aligned} \quad (15)$$

where $\bar{\mathbf{F}}_1, \dots, \bar{\mathbf{F}}_Q$ are estimated beforehand, either by initial experiments with the real platform or in a simulation environment with a nominal model. It can be noted that

$$\mathbf{T} \approx N \mathbf{\Gamma}(N), \quad (16)$$

which, since N is not a free variable during the optimization, means that the solution to the problem above will maximize the determinant of $\mathbf{G}(N)$. The last constraint in (15) is a relaxation of forcing each N_q to be a non-negative integer. With this formulation, a combinatorial problem is avoided, which makes the optimization easier to carry out. The found values of N_1, \dots, N_Q can readily be truncated to adjacent integer values in a second step, and the relaxation typically has a negligible effect on the outcome for reasonably long experiments.

Further, it can be noted that the approximation (14) is justified by asymptotic arguments and that the information gained from transient effects following switches between different experiment primitives is neglected in (15). Applying the experiment primitives for short times naturally makes these transient effects more notable in comparison to the experiment primitives themselves, and a possible remedy to this issue is to include constraints in the optimization problem that force each N_1, \dots, N_Q to either be identically zero or above a chosen threshold. This threshold should then be selected such that the time of the transient effects becomes small with respect to the length of each sub-experiment. Forcing the optimization variables to be identically zero or above a chosen threshold would, however, give a non-convex feasible set and make the optimization problem harder to solve with high precision. Another option is to skip the threshold constraints in the optimization and instead have small values of N_1, \dots, N_Q approximated with zero afterwards. This strategy can be problematic if the neglected sub-experiments carry unique information that otherwise is not attained but turned out to work well for the examples studied in this work.

There are a couple of small modifications that can be made for (15), which are worth mentioning. Firstly, the variable N corresponding to the experiment length does not affect the outcome of the optimization beyond scaling, and for $N = 1$, the optimization variables can be interpreted as the time fractions of the total experiment that should be spent in various modes. Secondly, under the assumption that N_{prior} data points have already been collected, it is possible to compute $\bar{\mathbf{T}}_{\text{prior}}$ based on these. By then augmenting the objective function with $\bar{\mathbf{T}}_{\text{prior}}$ as an extra term, the optimization instead yields the best mix of experiment primitives to complement the prior-collected data. Thirdly, it is possible to formulate an optimization problem similar to (15) that depends on prior-simulated covariance matrices instead of the expected information gains that are considered here. A notable drawback with this approach is that the sub-experiments all need to be sufficiently informative by themselves for the corresponding sub-covariance matrices to be well defined. The interested reader will find some more details regarding these modifications in Ljungberg (2022).

4. Method adapted for marine vessels

Most nonlinear model structures used for describing motion of marine vessels at speed stem from one of two basic ideas. The first was suggested by Abkowitz (1964) and is based on a Taylor expansion. In this case, the even-order terms are often neglected to enforce that the model behaves in the same way for positive and negative relative velocities, something that is necessary due to ship symmetry. The other type of model structure was first proposed by Fedyaevsky and Sobolev

(1964) and is based on a combination of physical effects such as circulation and cross-flow drag principles. These properties are often well-described by quadratic functions and the constraint of having a symmetric model is therefore instead resolved by use of the modulus function. Models of this type typically do not include any terms of higher order than two and are therefore referred to as second-order modulus models. Moreover, experimental ship data is often collected under presence of environmental disturbances, such as wind and water currents. Correctly dealing with these disturbances is challenging already in the linear case and becomes even harder when models are nonlinear. Doing so is, however, important to not obtain a biased model.

4.1. Two ways of subtracting the instrument mean

The possibility of obtaining consistency by using an IV estimator for dealing with these challenges was explored in Ljungberg and Enqvist (2022). In general, the accuracy of an IV estimator is highly dependent on the choice of instruments and a common way of obtaining instruments in practice is by simulation of a nominal model with crude parameter values. In this case, the instrument matrix can be formed as a noise-free version of the regression matrix, see for example Thil, Gilson, and Garnier (2008). However, in Ljungberg and Enqvist (2022) it was noted that this was not sufficient and the additional step of forcing the instrument matrix to have zero mean was proposed as a remedy. This turned out to be an efficient way of asymptotically eliminating the influence of certain disturbances on the parameter estimates. When data is collected from different sub-experiments, there are multiple ways in which the instrument mean can be subtracted. One way is to subtract it from one batch of data at a time

$$\tilde{\mathbf{Z}}_q(k) = \mathbf{Z}_q(k) - \frac{1}{N_q} \sum_{k=1}^{N_q} \mathbf{Z}_q(k). \quad (17)$$

Another alternative is to subtract the mean of the complete dataset at once

$$\tilde{\mathbf{Z}}_q(k) = \mathbf{Z}_q(k) - \underbrace{\frac{1}{N} \sum_{q=1}^Q \sum_{k=1}^{N_q} \mathbf{Z}_q(k)}_{\triangleq \bar{\mathbf{Z}}}. \quad (18)$$

It turns out that the latter option often gives a reduced variance for the estimated parameters and that better results can be obtained if this is taken into consideration already during the experiment design. The motivation for this is illustrated by the following simple example.

Example 1. Consider the data-generating system

$$y(k) = \theta_0 u(k) + e(k), \quad (19)$$

and the predictor model

$$\hat{y}(k) = u(k)\theta \stackrel{\Delta}{=} \varphi^T(k)\theta. \quad (20)$$

For a general instrument vector, $\zeta(k)$, the IV estimate is

$$\begin{aligned} \hat{\theta}_N^{IV} &= \left(\frac{1}{N} \sum_{k=1}^N \zeta(k) \varphi^T(k) \right)^{-1} \left(\frac{1}{N} \sum_{k=1}^N \zeta(k) y(k) \right) \\ &= \theta_0 + \frac{\frac{1}{N} \sum_{k=1}^N \zeta(k) e(k)}{\frac{1}{N} \sum_{k=1}^N \zeta(k) u(k)}. \end{aligned} \quad (21)$$

Further, assume that the input has a time-varying component $\tilde{u}(k)$ with zero mean and an excitation offset that alters sign halfway through the experiment

$$u(k) = \begin{cases} \tilde{u}(k) + \bar{u}, & \text{if } k = 1, \dots, N/2, \\ \tilde{u}(k) - \bar{u}, & \text{if } k = N/2 + 1, \dots, N. \end{cases} \quad (22)$$

Furthermore, assume that $\bar{u}(k)$ is repeated in both experiment halves, such that $\bar{u}(k + N/2) = \bar{u}(k)$ for $k = 1, \dots, N/2$. Now, divide the analysis into two cases. Firstly, consider the case where $\zeta(k) = u(k)$ and note that this corresponds to a complete subtraction of the instrument mean (the mean of $u(k)$ is zero by design). Then, it holds that

$$\hat{\theta}_N^{IV_1} = \theta_0 + \frac{\frac{1}{N} \left(\sum_{k=1}^N \bar{u}(k)e(k) \right) + \bar{u}'_N}{\frac{1}{N} \left(\sum_{k=1}^N \bar{u}(k)^2 \right) + \bar{u}^2} \quad (23)$$

where $\bar{u}'_N = \frac{1}{N} \left(\sum_{k=1}^{N/2} e(k) - \sum_{k=N/2+1}^N e(k) \right)$. Secondly, assume that $\zeta(k) = \bar{u}(k)$, which corresponds to a batchwise subtraction of the instrument mean. In this case, it holds that

$$\hat{\theta}_N^{IV_2} = \theta_0 + \frac{\frac{1}{N} \sum_{k=1}^N \bar{u}(k)e(k)}{\frac{1}{N} \left(\sum_{k=1}^N \bar{u}(k)^2 \right)} \quad (24)$$

It can be noted that the variance expression of $\hat{\theta}_N^{IV_1}$ has an additional term, \bar{u}^2 , in the denominator, which means that it will decrease to zero when \bar{u} increases. Therefore, $\hat{\theta}_N^{IV_1}$ is a better choice in comparison to $\hat{\theta}_N^{IV_2}$ for significant values of \bar{u} .

4.2. Compensating for zero-mean instruments

Compensating for a batchwise subtraction of the mean in the experiment design would be straightforward since the relevant signal levels associated with each experiment primitive can be found during the initial experiments where $\bar{r}_1, \dots, \bar{r}_Q$ are estimated. However, if the mean is subtracted for the complete experiment at once, the usefulness of each sub-experiment can only be assessed in combination with the other sub-experiments. Compensating for a complete subtraction of the mean in the experiment design is therefore a bit more challenging. If the mean is subtracted as in (18), it is the case that

$$\bar{r}_q = \bar{X}_q - \bar{Y}_q \bar{Z}_q^T, \quad (25)$$

where

$$\bar{X}_q = \lim_{N_q \rightarrow \infty} \frac{1}{N_q} \sum_{k=1}^{N_q} \Phi_q(k) Z_q^T(k), \quad (26)$$

$$\bar{Y}_q = \lim_{N_q \rightarrow \infty} \frac{1}{N_q} \sum_{k=1}^{N_q} \Phi_q(k). \quad (27)$$

Similar to how \bar{r}_q was estimated for each experiment primitive earlier, \bar{X}_q, \bar{Y}_q and

$$\bar{Z}_q = \lim_{N_q \rightarrow \infty} \frac{1}{N_q} \sum_{k=1}^{N_q} Z_q(k), \quad (28)$$

can be estimated for each experiment primitive based on initial experiments or a nominal model. Consequently, the optimization problem

$$\begin{aligned} (N_1^*, \dots, N_Q^*) = & \operatorname{argmax}_{N_1, \dots, N_Q} \log \left| \det \left(T - S \bar{Z}^T \right) \right| \\ \text{s.t. } & T = N_1 \bar{X}_1 + \dots + N_Q \bar{X}_Q, \\ & S = N_1 \bar{Y}_1 + \dots + N_Q \bar{Y}_Q, \\ & \bar{Z} = \frac{1}{N} (N_1 \bar{Z}_1 + \dots + N_Q \bar{Z}_Q), \\ & N_1 + \dots + N_Q = N, \\ & N_q \geq 0, \quad q = 1, \dots, Q, \end{aligned} \quad (29)$$

can be considered as an alternative to (15), where the fact that the first-order moment of the instrument matrix will be removed during the parameter estimation is acknowledged in the experiment design. This optimization problem is non-convex. However, no initial guess for the optimal input signal is required unlike many other non-convex problem formulations used for experiment design. Instead, only a guess

for how long the candidate signals should be applied with respect to each other is needed. This is a significantly milder requirement and based on experimental work, it seems to be sufficient to assume that all candidate signals are used equally much.

5. Experimental work

In order to illustrate the potential of the proposed method, experiments were performed using both simulated and real data.

5.1. Results from simulation experiments

The simulation experiments were carried out using a maneuvering model of a marine surface vessel given by

$$\begin{aligned} u(k+1) &= u(k) + \mathcal{X}_u u_r(k) + \mathcal{X}_{|u|} u_r(k) |u_r(k)| \\ &\quad + \mathcal{X}_{vr} v_r(k) r(k) + \mathcal{X}_\tau \tau_1(k), \\ v(k+1) &= v(k) + \mathcal{Y}_v v_r(k) + \mathcal{Y}_{ur} u_r(k) r(k) + \mathcal{Y}_\tau \tau_2(k), \\ r(k+1) &= r(k) + \mathcal{N}_r r(k) + \mathcal{N}_{uv} u_r(k) v_r(k) + \mathcal{N}_\tau \tau_3(k), \\ y(k) &= [u(k) \quad v(k) \quad r(k)]^T + e(k). \end{aligned}$$

Here, $u(k)$, $v(k)$ and $r(k)$ constitute surge, sway and yaw rate, respectively, and $\tau(k) = [\tau_1(k), \tau_2(k), \tau_3(k)]^T$ are input forces and moments caused by the ship's actuators. The subscripted r signifies relative velocity such that $u_r(k) = u(k) - u_c(k)$ and $v_r(k) = v(k) - v_c(k)$, where $u_c(k)$ and $v_c(k)$ are velocity components of an ocean current. The velocity of the ocean current as well as the uncertainty associated with the measurement of the system states, $e(k)$, were treated as disturbance signals. Note that quadratic damping was only considered in surge, where the vessel was assumed to reach the highest speeds. Except for this, the model is in agreement with the well-adopted framework found in Fossen (2011) and the definition of a second-order modulus model in Ljungberg and Enqvist (2022). The parameter values, $\mathcal{X}_u = -0.06$, $\mathcal{X}_{|u|} = -0.01$, $\mathcal{X}_{vr} = 0.08$, $\mathcal{X}_\tau = 1.4 \cdot 10^{-5}$, $\mathcal{Y}_v = -0.1$, $\mathcal{Y}_{ur} = -0.006$, $\mathcal{Y}_\tau = 1.4 \cdot 10^{-5}$, $\mathcal{N}_r = -0.35$, $\mathcal{N}_{uv} = -0.03$, $\mathcal{N}_\tau = 3 \cdot 10^{-4}$, were chosen based on earlier work with the small-scale ship presented in Section 5.2, with a sampling rate of $f_s = 8$ Hz.

Data was generated based on 11 different excitation signals and the system states resulting from applying them in undisturbed simulations, i.e., the expected experiment primitives, are given below.

- (τ_1) Decelerating motion:
 $u(k) : 1 \rightarrow 0, v(k) = 0, r(k) = 0.$
- (τ_2) Accelerating motion:
 $u(k) : 0 \rightarrow 1, v(k) = 0, r(k) = 0.$
- (τ_3) Slow/flat zig-zag motion
 $u(k) \approx 0.35, v(k) \in [-0.1, 0.1], r(k) \in [-0.2, 0.2].$
- (τ_4) Moderate/flat zig-zag motion:
 $u(k) \approx 0.8, v(k) \in [-0.1, 0.1], r(k) \in [-0.2, 0.2].$
- (τ_5) Fast/flat zig-zag motion:
 $u(k) \approx 1.15, v(k) \in [-0.1, 0.1], r(k) \in [-0.2, 0.2].$
- (τ_6) Slow/steep zig-zag motion:
 $u(k) \approx 0.35, v(k) \in [-0.15, 0.15], r(k) \in [-0.35, 0.35].$
- (τ_7) Moderate/steep zig-zag motion:
 $u(k) \approx 0.7, v(k) \in [-0.15, 0.15], r(k) \in [-0.35, 0.35].$
- (τ_8) Fast/steep zig-zag motion:
 $u(k) \approx 1, v(k) \in [-0.15, 0.15], r(k) \in [-0.35, 0.35].$
- (τ_9) Slow (inward-bound) spiral motion:
 $u(k) \approx 0.4, v(k) \approx 0.05, r(k) : 0 \rightarrow -0.6.$
- (τ_{10}) Moderate (inward-bound) spiral motion:
 $u(k) \approx 0.75, v(k) \approx 0.125, r(k) : 0 \rightarrow -0.6.$
- (τ_{11}) Fast (inward-bound) spiral motion:
 $u(k) \approx 1, v(k) \approx 0.2, r(k) : 0 \rightarrow -0.6.$

Here, all values for $u(k)$ and $v(k)$ are given in m/s and the values for $r(k)$ in rad/s. Each of the input signals were applied for about 40 s (300 samples) and the matrices in (26)–(28) were estimated. The problem

(29) was solved in MATLAB using the function *fmincon* and the interior-point algorithm, which resulted in an optimized design where $\tau_1(k)$ is used for 16% of the total experiment time, $\tau_2(k)$ for 4%, $\tau_5(k)$ for 9%, $\tau_6(k)$ for 42%, $\tau_8(k)$ for 4% and $\tau_{11}(k)$ for 25%. It can be noted that the most used experiment primitive in the optimized design is the slow and steep zig-zag motion associated with $\tau_6(k)$. This is a bit surprising, considering that the signal magnitudes are strictly higher when input signals $\tau_7(k)$ or $\tau_8(k)$ are applied.

The predictor model

$$\hat{y}(k) = \begin{bmatrix} \varphi_u^T(k) & 0 & 0 \\ 0 & \varphi_v^T(k) & 0 \\ 0 & 0 & \varphi_r^T(k) \end{bmatrix} \theta, \quad (30)$$

with

$$\varphi_u(k) = [y_1(k) \quad y_1(k)|y_1(k)| \quad y_2(k)y_3(k) \quad \tau_1(k)]^T,$$

$$\varphi_v(k) = [y_2(k) \quad y_1(k)y_3(k) \quad \tau_2(k)]^T,$$

$$\varphi_r(k) = [y_3(k) \quad y_1(k)y_2(k) \quad \tau_3(k)]^T,$$

was used, which is able to describe the data-generating system in undisturbed conditions. As discussed earlier, the instruments were generated by simulating the output of this model with crude parameter values θ' ($\theta'_1 = \theta'_5 = -0.08$, $\theta'_4 = \theta'_7 = 10^{-5}$, $\theta'_8 = -0.4$, $\theta'_{10} = 3.5 \cdot 10^{-4}$ and $\theta'_2 = \theta'_3 = \theta'_6 = \theta'_9 = 0$). As discussed before, the accuracy of an IV estimator is highly dependent on the instruments and a too imprecise nominal model can make the estimator diverge. Despite neglecting all nonlinear effects, the nominal model above was sufficiently accurate to yield a reliable estimator. Further, an optimized input sequence of 125 s (1000 samples) was formed, which respected the found ratios and new data was collected where $u_c(k) \sim \mathcal{N}(0, 0.025)$, $v_c(k) \sim \mathcal{N}(0, 0.025)$ and $e_i(k) \sim \mathcal{N}(0, 0.025)$ for $i = 1, 2, 3$. The models obtained from 500 Monte Carlo simulations using this optimized design were compared with models obtained using a random design. The random design was formed by arbitrarily choosing 1 out of the 11 input signals for each of 5 equally long segments (200 samples each). Which input signals that were used and in what order, was varied over the Monte Carlo simulations. For each Monte Carlo iteration, parameter estimation errors were computed and ideally, histograms showing the distribution for these individually would be provided. Instead, due to limited space, a normalized root-mean-squared error (RMSE) for each estimated parameter was computed and the norm of all these is shown in Fig. 1. By visual inspection it can be seen that the random design often gives higher estimation errors and that the improved design is a more robust choice in this regard. The improved design gives an error that is smaller than 5 in more than 99% of the cases, whereas the random design only does so in 49% of the cases.

The accuracies of the estimated models were also evaluated by cross-validation (CV) between simulated model output and the output of the true system. In the validation dataset, a more conventional input signal was used, which corresponded to smoother motion. This input signal was used in order to evaluate the models in a more realistic scenario. Most importantly, the input signal used in the validation was different from all the experiment primitives considered for the experiment design. The RMSE metric was used in this case as well and in Fig. 2, the errors for each DOF is shown and in Fig. 3, the norm of these is shown. It can be seen that the simulated model output is more often in agreement with the system output when the optimized experiment design is used. Similar to when parameter errors were considered, it can in Fig. 3 be noted that the norm of the simulation errors is smaller than 0.15 in 99% of the cases when the optimized design is used but only in 69% of the cases when the random design is used. Similar observations can be made for each individual DOF in Fig. 2.

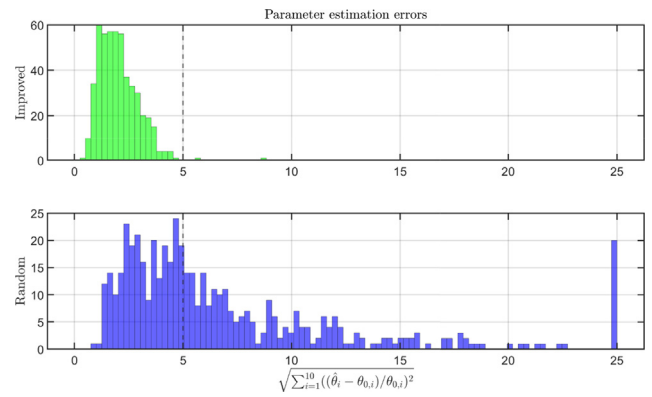


Fig. 1. Norm of RMSE for normalized estimation errors between estimated parameters, $\hat{\theta}$, and true parameters, $\theta_0 = [\mathcal{N}_u \dots \mathcal{N}_r]^T$, over 500 Monte Carlo simulations. The top plot shows an optimized design and the bottom plot a random design. Errors above 25 are truncated to 25.

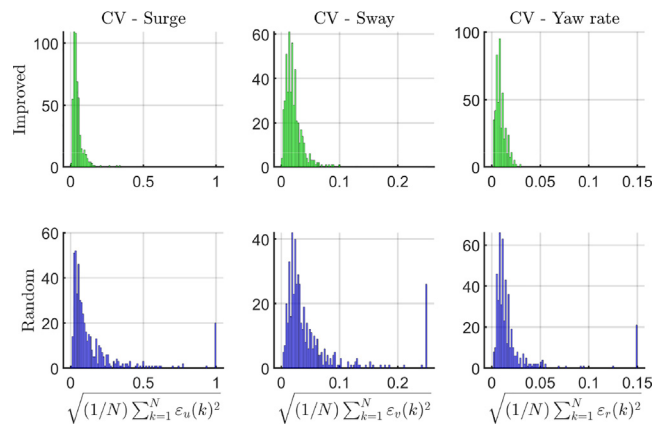


Fig. 2. RMSE between simulated model output and output of the data-generating system for each DOF, $\varepsilon_u(k) = \hat{u}(k) - u(k)$, $\varepsilon_v(k) = \hat{v}(k) - v(k)$, $\varepsilon_r(k) = \hat{r}(k) - r(k)$. Errors above 1, 0.25 and 0.15 are truncated, respectively.

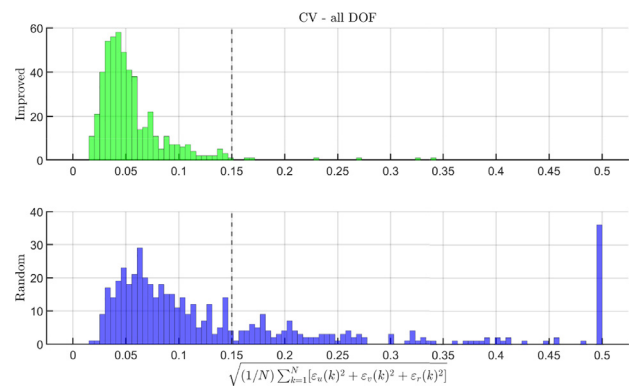


Fig. 3. Norm of RMSE between simulated model output and output of the data-generating system, $\varepsilon_u(k) = \hat{u}(k) - u(k)$, $\varepsilon_v(k) = \hat{v}(k) - v(k)$, $\varepsilon_r(k) = \hat{r}(k) - r(k)$. Errors above 0.5 are truncated to 0.5.

5.2. Results from model ship

Experimental data was also collected using a small-scale surface ship, which is about 1 m long, 0.3 m wide and weighs roughly 14 kg. This ship is actuated by two azimuth thrusters mounted in the aft and a tunnel thruster located in the bow. Further, the ship is equipped with an inertial measurement unit (IMU) for orientation, a GNSS receiver for positioning, as well as rotational-speed sensors for the thrusters. The

surge and sway speeds were estimated using Euler's explicit method with the position and orientation measurements. In order to make the analysis simple, the tunnel thruster was not used during the experiments and the input forces and moments, $\tau_1(k)$, $\tau_2(k)$ and $\tau_3(k)$, were generated solely based on the two azimuth thrusters. A simple static model for describing forces and moments obtained from azimuth thrusters is

$$\tau_1(k) = \sum_{i=1}^2 n_i(k) \cos(\alpha_i(k)),$$

$$\tau_2(k) = \sum_{i=1}^2 n_i(k) \sin(\alpha_i(k)),$$

$$\tau_3(k) = \sum_{i=1}^2 l_{y,i} n_i(k) \cos(\alpha_i(k)) + l_{x,i} n_i(k) \sin(\alpha_i(k)),$$

where $l_{x,i}$ and $l_{y,i}$ define the (known) mounting position of azimuth thruster i in x and y -direction, respectively. The control signals, $n_1(k)$, $n_2(k)$, $\alpha_1(k)$ and $\alpha_2(k)$, corresponding to propeller speeds and azimuth angles, were based on this model chosen to yield experiment primitives in agreement with the simulation experiments shown above. For example, the thrusters were rotated about 15° in each direction for the flat zig-zag motion and 25° for the steep zig-zag motion. Moreover, the model structure was the same as in the simulation experiments and the instruments were again generated by use of a nominal model. This time, the crude parameter values used for the nominal model were obtained by the least-squares estimate, *i.e.*, by letting $Z(k) = \Phi(k)$.

The collected data was divided into 55 parts (75 samples each), 5 for each of the 11 experiment primitives. Using these shorter datasets for estimating (26)–(28) and solving (29), gave an optimized design where $\tau_1(k)$ is used for 20% of the total experiment time, $\tau_6(k)$ for 49%, $\tau_9(k)$ for 13% and $\tau_{11}(k)$ for 18%. This optimized design was then approximated with a design where $\tau_6(k)$ is used for half of the total time and $\tau_1(k)$, $\tau_9(k)$ and $\tau_{11}(k)$ are used for one sixth of the time each. In this way, multiple different datasets could be formed from the collected data where the optimized ratios were fulfilled. The fact that multiple different realizations of the optimal experiment could be formed made it easier to discuss the results and to make statistical conclusions, something that is otherwise only possible in simulation experiments where more data is available.

The optimized datasets were formed by picking 3 sub-experiments where $\tau_6(k)$ is used and 1 sub-experiment where $\tau_1(k)$, $\tau_9(k)$ and $\tau_{11}(k)$ are used each, *i.e.*, 6 shorter sub-experiments were used in total. This way of forming the experiment was compared to a random design where 6 sub-experiments were selected randomly. For both the optimized design and the random design, the set of sub-experiments was varied 500 times in a Monte Carlo manner. Similarly to the simulation experiments, for each such Monte Carlo iteration, the accuracies of the estimated models were evaluated by CV between simulated model output and estimates obtained from the measured signals. The results are given in Figs. 4 and 5. It can be seen that the simulated model output is more often in agreement with the measured system output when the optimized experiment design is used. Here, it is most notable because the error following the random design takes on degenerate values in about 20% of the cases.

The validation data was made up by all the 55 sub-experiments, which means that the validation data was not completely independent of the estimation data ($6/55 \approx 11\%$ overlap). This small overlap was deemed justifiable because the alternative would be a folded type of CV, where the sub-experiments currently used for estimation were kept out and not included in the validation set. Since there were only 5 sub-experiments of each type available, this would mean that some maneuvers occasionally were not present at all in the validation data and would therefore possibly result in a misleading evaluation.

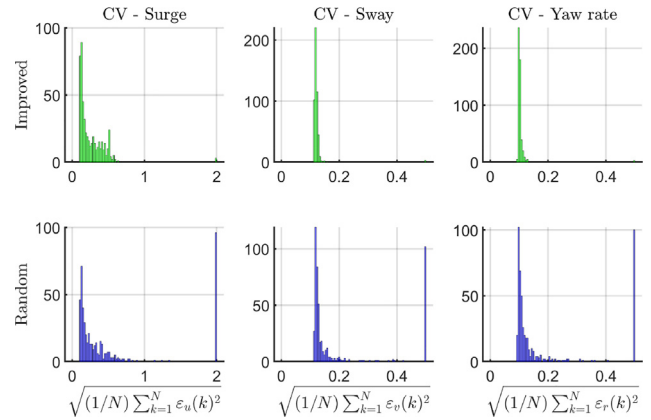


Fig. 4. RMSE between simulated model output and estimates obtained from measured signals in each DOF from the small-scale model ship. Errors above 2, 0.5 and 0.5 are truncated, respectively.

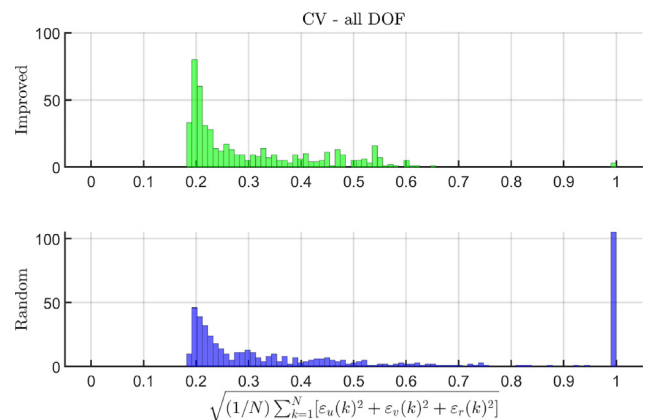


Fig. 5. Norm of RMSE between simulated model output and estimates obtained from measured signals from the small-scale model ship. Errors above 1 are truncated to 1.

5.3. Results from full-scale ship

There was also data available from a full-scale ship. This ship is roughly 30 m long and has an actuator setup with two azimuth thrusters mounted along the centerline, one at the front and one in the rear. This configurations makes it possible to excite the ship in sway without having a forward speed. The full-scale ship is equipped with a GNSS receiver with two antennas, which provides estimates of both its position and yaw angle. Unlike the small model ship discussed before, all signals were sampled at the lower rate of $f_s = 1$ Hz. Notably, the data from the full-scale ship was not collected for the sake of this work in particular and consequently the data does not include the previously discussed experiment primitives. Instead, the following motion types were identified.

- (τ_{12}) Surge accelerations/decelerations:
 $u(k) \in [0, 3.5]$, $v(k) = 0$, $r(k) = 0$.
- (τ_{13}) Sway accelerations/decelerations:
 $u(k) = 0$, $v(k) \in [-1, 1]$, $r(k) = 0$.
- (τ_{14}) Slow sway motion:
 $u(k) = 0$, $|v(k)| \in [0.5, 0.8]$, $r(k) = 0$.
- (τ_{15}) Moderate sway motion:
 $u(k) = 0$, $|v(k)| \approx 0.9$, $r(k) = 0$.
- (τ_{16}) Fast sway motion:
 $u(k) = 0$, $|v(k)| \in [1.1, 1.2]$, $r(k) = 0$.
- (τ_{17}) Slow/high frequent ($T \approx 30$) zig-zag motion:
 $u(k) \approx 3$, $v(k) \in [-0.5, 0.5]$, $r(k) \in [-0.03, 0.03]$.

- (τ_{18}) Fast/high frequent ($T \approx 30$) zig-zag motion:
 $u(k) \approx 4, v(k) \in [-0.5, 0.5], r(k) \in [-0.03, 0.03]$.
- (τ_{19}) Slow/low frequent ($T \approx 100$) zig-zag motion:
 $u(k) \approx 3, v(k) \in [-0.5, 0.5], r(k) \in [-0.03, 0.03]$.
- (τ_{20}) Fast/low frequent ($T \approx 100$) zig-zag motion:
 $u(k) \approx 4, v(k) \in [-0.5, 0.5], r(k) \in [-0.03, 0.03]$.

As before, all values for $u(k)$ and $v(k)$ are given in m/s and the values for $r(k)$ in rad/s. Moreover, T is the period time of the zig-zag motion in seconds. The simple model structure (30) turned out to be sufficient for describing the collected data and was therefore used here as well and the instrument matrix was generated in the same way as for the small-scale ship. Further, the data was divided into different parts for the different maneuvers. This time, 4 sub-experiments of each type were formed, each corresponding to 100 s of maneuvering (*i.e.*, 100 samples). Estimating (26)–(28) and solving (29) gave an optimized design where $\tau_{13}(k)$ is used for 4% of the total experiment time, $\tau_{14}(k)$ for 2%, $\tau_{16}(k)$ for 30%, $\tau_{18}(k)$ for 22%, $\tau_{19}(k)$ for 11% and $\tau_{20}(k)$ for 31%.

Analogously to before, the optimal design was approximated with a design where $\tau_{16}(k)$ and $\tau_{20}(k)$ are used for one third each while $\tau_{18}(k)$ and $\tau_{19}(k)$ are used for one sixth each. In this case, realizations of an optimal experiment can be formed by picking 2 of the sub-experiments where $\tau_{16}(k)$ is used, 1 where $\tau_{18}(k)$ is used, 1 where $\tau_{19}(k)$ is used and 2 where $\tau_{20}(k)$ is used. Furthermore, the optimized experiment design was compared with a random design, where 6 sub-experiments were selected randomly and the accuracies of the estimated models were evaluated by CV between simulated model output and estimates obtained from the measured signals. For both the optimized design and the random design, the set of sub-experiments was varied 500 times and the results are given in Figs. 6 and 7. Here it is actually the case that the most accurate models are obtained with the random design. The worst-case errors following the random design are, however, also higher and sometimes take on degenerate values. In this sense, the optimized design is a more solid choice.

It can be remarked that the data from the full-scale ship was collected on two separate occasions with quite different weather conditions. Parts of the data was collected during a day with wind speeds of less than 3 m/s, *i.e.* light breeze by the definition in Fossen (2011), whereas the other parts were collected during a day where the wind speed was about 10 m/s, which corresponds to fresh breeze by the same set of definitions. Further, $\tau_{16}(k) - \tau_{18}(k)$ were only applied during the windy day and consequently, the data associated with those experiment primitives is more noisy in comparison to the other data. Due to the way that the realizations of the optimal design are formed, by concatenation of sub-experiments where both $\tau_{16}(k)$ and $\tau_{18}(k)$ are used, they are bound to always include at least 50% of data from the windy day. The realizations of the random design are, however, occasionally made up exclusively of data from the calmer day, which partly explains why the most accurate models are obtained with the random design.

6. Generating an experiment trajectory

Most previous work about experiment design have focused on obtaining an optimal input signal. There are, however, many practical considerations to be made for a data-collection experiment. In addition to being time-efficient, it is, for example, important that the experiment is planned with respect to geographical constraints and operational safety. The feasibility of the resulting experiment trajectory is therefore equally important as the informativity content of the collected data. Motion planning is a vast field of research and the goal of this work was not to develop new theory in that regard. Instead, the focus is on showing how the previously discussed dictionary-based method for finding an excitation signal can be combined with a motion-planning framework to obtain a trajectory that is both informative and feasible.

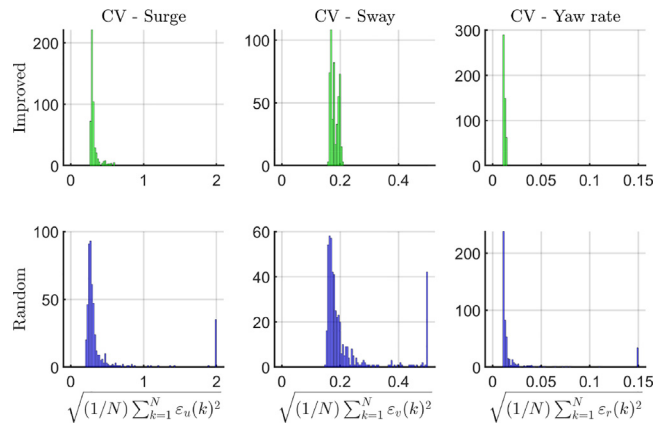


Fig. 6. RMSE between simulated model output and estimates obtained from measured signals in each DOF from the full-scale ship. Errors above 2, 0.5 and 0.15 are truncated, respectively.

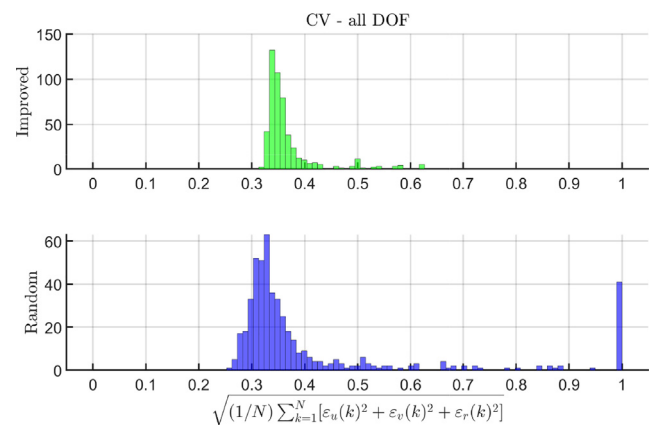


Fig. 7. Norm of RMSE between simulated model output and estimates obtained from measured signals from the full-scale ship. Errors above 1 are truncated to 1.

In motion planning, a state-lattice-based approach uses a library of motion primitives that are combined to obtain a path from a starting position to a goal position, see for example Pivtoraiko, Knepper, and Kelly (2009). Similarly to how the candidate excitation signals, $u_1(k), \dots, u_Q(k)$, previously have been associated with information gains, each candidate signal can by use of a nominal model, or by initial experiments with the real platform, be associated with an expected change in position and attitude. These motion segments can then be used in lattice-based motion planning to find a suitable realization under constraints, such as spatial limitations or safety. The standard formulation of a lattice-based motion-planning problem is

$$\text{minimize}_{\{m_k\}_{k=0}^{M-1}} \sum_{k=0}^{M-1} J(m_k) \quad (31a)$$

$$\text{s.t.} \quad \mathbf{x}_0 = \mathbf{x}_s, \quad \mathbf{x}_M = \mathbf{x}_f, \quad (31b)$$

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, m_k), \quad (31c)$$

$$m_k \in \mathcal{P}, \quad (31d)$$

$$c(m_k, \mathbf{x}_k) \in \mathcal{X}_{\text{free}}. \quad (31e)$$

adopted from Bergman (2021). Here, $J(m_k)$ is a cost function and the decision variables are the motion primitive sequence $\{m_k\}_{k=1}^M$ and its length M . Further, the constraints in (31b) define the initial and final states, the state-transition in (31c) gives the successor state \mathbf{x}_{k+1} after m_k is applied in \mathbf{x}_k , the set \mathcal{P} in (31d) includes all applicable motion primitives and lastly the constraint (31e) ensures that there is no collision with obstacles when m_k is applied in \mathbf{x}_k .

6.1. Motion planning for experiment design

Normally, the state vector contains information about position and attitude, which for motion planning in the horizontal plane includes positional coordinates, (x, y) , and heading angle, ψ . Therefore, introduce the notation $\eta_k = [x_k, y_k, \psi_k]^T$. The key difference between motion planning for dictionary-based experiment design and regular motion planning lies in the requirement of respecting the ratios between N_1, \dots, N_Q found when solving (15) or (29). There are multiple ways to include such a constraint in the problem (31). One solution is to augment the state vector with additional elements that count how many times each experiment primitive has been used. This approach is conceptually appealing because of its interpretability. Thereby, let g_k^q be the number of times experiment primitive q has been used at time k . Including this, the state vector becomes

$$\mathbf{x}_k \stackrel{\Delta}{=} \begin{bmatrix} \eta_k^T & g_k^1 & \dots & g_k^Q \end{bmatrix}^T. \quad (32)$$

The ratios between N_1, \dots, N_Q can then be respected by enforcing that each experiment primitive is used a chosen number of times

$$\mathbf{x}_s = \begin{bmatrix} \eta_s \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{x}_f = \begin{bmatrix} \eta_f \\ n_1 \\ \vdots \\ n_Q \end{bmatrix}. \quad (33)$$

The number of times experiment primitive q is applied in total can be selected such that $n_q N'_q \approx N_q$. In this case, N'_1, \dots, N'_Q become design variables, which declare how long each motion primitive should be. Due to the augmented state vector, the state lattice will be of higher dimension in comparison to a conventional lattice-based motion-planning problem and if too many experiment primitives (large Q) are considered and each is divided into a large number of parts (small N'_1, \dots, N'_Q), the state lattice will grow quickly. Consequently, the trade-off between feasibility and computational complexity that follows the choice of N'_1, \dots, N'_Q is important. Moreover, there are some general difficulties associated with a dictionary-based experiment design. This is because different concatenation orders will lead to different signal transitions, which affect the informativity of the design. Also, applying different signal types after each other might lead to abrupt changes and consequently high-frequency transients. These problems become more apparent when the experiment primitives are divided into smaller parts. To summarize, large values of N'_1, \dots, N'_Q make the motion-planning problem easier to solve whereas small values of N'_1, \dots, N'_Q give better possibilities of finding feasible solutions but increase the computational complexity and make the expected information gain from the experiment harder to predict. From a more practical point of view, some sub-experiments, such as zig-zag maneuvers, have natural breaking points at the end of each period. Choosing N'_1, \dots, N'_Q in agreement with these is therefore a reasonable choice.

The objective function

$$J = \sum_{k=0}^{M-1} J(\mathbf{m}_k), \quad (34)$$

is made up of the accumulated cost of applying a sequence of M motion primitives. In a conventional motion-planning problem, the cost of applying a particular motion primitive can, for example, be related to travel time or fuel cost. Performing motion planning for experiment design is a bit different. Ideally, a data-collection experiment should only be made up by information-yielding maneuvers, but in practice this is not always possible. This is because it is sometimes necessary to reposition the ship prior to executing the next maneuver due to space constraints. One way of dealing with this issue is to assume that there are additional basic motion primitives that can be used to connect the informative sequences. In this case, the complete set of motion primitives is

$$P = \left\{ \underbrace{m^1, \dots, m^Q}_{\text{informative}}, \underbrace{m^{Q+1}, \dots, m^{Q+B}}_{\text{basic}} \right\}, \quad (35)$$

where the basic motion primitives can be of types that are often used in conventional motion planning, such as straight driving as well as plain left and right turns. The goal with the experiment design can then be formalized as using each information-yielding primitive a given number of times, while using the basic primitives as sparsely as possible. A simple objective function that caters to this goal is

$$J(\mathbf{m}^\ell) = \begin{cases} 0 & \text{if } \ell \leq Q, \\ \bar{L} & \text{if } \ell > Q. \end{cases} \quad (36)$$

In other words, have the cost of applying an information-yielding primitive be 0 and the cost of applying a basic primitive be $\bar{L} > 0$.

For high-dimensional and large state lattices, an exhaustive search for a solution is time consuming. In this case, the planning can be carried out using the A^* -search algorithm, as formulated in Bergman (2021). The online computational efficiency of such a search method relies on a well-informed heuristic function to guide the search towards the solution. The heuristic function takes as input two arbitrary states in the state lattice and returns an estimated cost of going from one to the other. Typically, the second input argument is the goal node, such that the heuristic function returns the estimated cost of reaching the goal. In general, it is only possible to guarantee optimality if the heuristic function is both admissible and consistent. It is, however, not easy to come up with a non-trivial heuristic function that fulfills these criteria for the planning problem considered in this work. This is because the cost of reaching the goal node from another node can be 0, even if this other node happens to be far away from the goal in terms of geographical distance. In fact, considering (36), the cost will be 0 if the goal position can be reached by using any combination of the information-yielding experiment primitives that are still left to be used. This means that it is hard to say for sure that a solution is optimal unless most other options have been explored.

6.2. Simulation example

To show that the proposed way of generating an experiment trajectory is reasonable, a simulation experiment was performed based on a sea-depth map over the port of Helsinki. Three different experiment primitives were considered, two types of zig-zag motion in surge direction with low and high amplitude for the turns, respectively, as well as one zig-zag motion in sway direction. Each of these were applied 3 times. In addition to the information-yielding maneuvers, the ship was allowed to move straight forward and to rotate on the spot (both clockwise and anticlockwise), *i.e.*, 3 basic motion primitives were considered as well. The obstacles were represented by bounding boxes and additional bounding boxes associated with the changes in position from the motion primitives were computed. The constraint (31e) could then be checked by examining whether the bounding box associated with a particular motion primitive overlapped with any bounding box associated with an obstacle. The state lattice was discretized with 35×44 positional coordinates and 4 different attitude levels. In this simulation experiment, a heuristic function with three terms was implemented. The first two terms were based on the Euclidian distance and deviation in attitude with respect to the goal, which constitutes standard costs in motion planning. The third term was an artificial cost associated with the remaining required use of the information-yielding experiment primitives. By giving significant weight to the third term, a greedy behavior where the planner prioritized using information-yielding maneuvers early was obtained. This improved the speed of the search for a solution. The considered heuristic function was neither admissible nor consistent but gave good results in practice.

A planned trajectory is shown in Fig. 8, where the ship was assumed to start in the upper-left corner, which is a harbor and to finish up in a position in the lower parts of the map. The goal position was chosen arbitrarily and an arguably more realistic scenario would be to both start and end in the harbor. Planning with the same node as start and goal is, however, by no means more difficult and choosing the goal

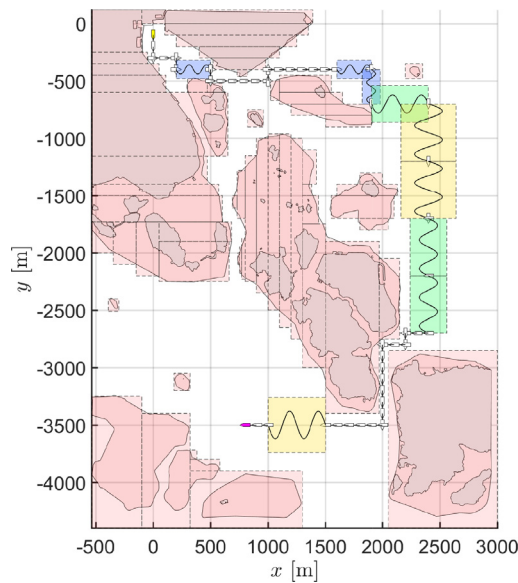


Fig. 8. Planned trajectory using 3 different experiment primitives; steep zig-zag motion in surge direction (blue), wide zig-zag motion in surge direction (orange) and wide zig-zag motion in sway (green). The start position is marked in yellow and the goal position is marked in purple. Islands and shallow water are treated as obstacles and are marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

node in this way had the benefit of the trajectory not overlapping itself, which makes the result much easier to illustrate in a plot. It can be seen that the planner chooses to use the most space-consuming maneuvers in the wide area at the top right. Further, the importance of including the basic motion primitives is clear, because without them it would, for example, not be possible to pass by the narrow region around (2000, -3500). Finding this trajectory took about 40 seconds on a modern laptop when using a completely plain implementation of the A^* -search algorithm in MATLAB. A more state-of-the-art implementation would probably speed up the search significantly.

7. Conclusions

The usefulness of a dictionary-based approach for experiment design has been analyzed and promising results have been shown on both simulated and real data from marine vessels. Further, a technique for combining the approach with a motion-planning framework to obtain an experiment trajectory that is both informative and feasible has been demonstrated. Selecting a set of informative maneuvers in the proposed way can probably in itself give improvements in terms of model accuracy in comparison to what is currently done in practice and if a more generally optimal input signal is sought, the suggested method can be viewed as a way of obtaining an initial guess for warm-starting a more flexible optimizer. Evaluating the potential benefits of such an approach is relevant future work. Further, a key step in a dictionary-based approach for experiment design is the selection of candidate input signals. For ships, natural choices are excitation signals corresponding to standard maneuvers, such as zig-zag, circle and spiral motions of varying speeds, frequencies and turn rates. In this work, it was assumed that the set of candidate signals to consider was predetermined, and important future work is to explore ways of automating the choice. The first step in this direction could be to adapt the set of candidate excitations based on the data that has thus far been collected during the experiment. Furthermore, in this work, the trajectory generation was performed after choosing the combination of experiment primitives. In future work, the possibility of formulating a joint problem where the informativity and spatial feasibility are dealt with simultaneously should be explored.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the competence center LINK-SIC, Sweden.

References

- Abkowitz, Martin A. (1964). *Lectures on ship hydrodynamics – steering and manoeuvrability: Technical report HY-5*, Lyngby, Denmark: Hydro- and Aerodynamics Laboratory.
- Bergman, Kristoffer (2021). *Exploiting direct optimal control for motion planning in unstructured environments* (Ph.D. thesis), Linköping, Sweden: Linköpings universitet.
- Blanke, Mogens, & Knudsen, Morten (1999). Optimized experiment design for marine systems identification. In *Proceedings of the 14th IFAC world congress* (pp. 8249–8254). Beijing, China:
- Blanke, Mogens, & Knudsen, Morten (2006). Efficient parameterization for grey-box model identification of complex physical systems. In *14th IFAC symposium on system identification* (pp. 338–343). Newcastle, Australia.
- De Cock, Alexander, Gevers, Michel, & Schoukens, Johan (2016). D-optimal input design for nonlinear FIR-type systems: A dispersion-based approach. *Automatica*, 73, 88–100.
- Fedyayevsky, K. K., & Sobolev, G. V. (1964). *Control and stability in ship design*. St. Petersburg, Russia: State Union Ship-building House.
- Forgione, Marco, Bombois, Xavier, Van den Hof, Paul M. J., & Hjalmarsson, Håkan (2014). Experiment design for parameter estimation in nonlinear systems based on multilevel excitation. In *European control conference* (pp. 25–30).
- Fossen, Thor I. (2011). *Handbook of marine craft hydrodynamics and motion control*. Chichester, U.K.: John Wiley & Sons.
- Gerencsér, László, Hjalmarsson, Håkan, & Mårtensson, Jonas (2009). Identification of ARX systems with non-stationary inputs—asymptotic analysis with application to adaptive input design. *Automatica*, 45(3), 623–633.
- Gevers, Michel, Bombois, Xavier, Hildebrand, Roland, & Solari, Gabriel (2011). Optimal experiment design for open and closed-loop system identification. *Communications in Information and Systems*, 11(3), 197–224.
- Goodwin, Graham C., & Payne, Robert L. (1977). Dynamic system identification: experiment design and data analysis. *Mathematics in Science and Engineering*, 136.
- Gopaluni, R. Bhushan, Schön, Thomas B., & Wills, Adrian G. (2011). Input design for nonlinear stochastic dynamic systems—a particle filter approach. In *Proceedings of the 18th IFAC world congress* (pp. 13191–13196). Milano, Italy.
- Hjalmarsson, Håkan, & Mårtensson, Jonas (2007). Optimal input design for identification of non-linear systems: Learning from the linear case. In *American control conference (ACC)* (pp. 1572–1576).
- Lack, Sven, Rentzow, Erik, & Jeinsch, Torsten (2019). Experimental parameter identification for an open-frame ROV: Comparison of towing tank tests and open water self-propelled tests. *IFAC-PapersOnLine*, 52(21), 271–276.
- Larsson, Christian A., Hjalmarsson, Håkan, & Rojas, Cristian R. (2010). On optimal input design for nonlinear FIR-type systems. In *9th IEEE conference on decision and control* (pp. 7220–7225). Atlanta, GA, USA.
- Ljung, Lennart (1999). *System identification: theory for the user* (2nd ed.). Upper Saddle River, N. J.: Prentice Hall.
- Ljungberg, Fredrik (2022). *Identification of nonlinear marine systems* (Ph.D. thesis), Linköping, Sweden: Linköpings Universitet.
- Ljungberg, Fredrik, & Enqvist, Martin (2022). Consistent estimators for nonlinear vessel models. arXiv:2204.04973.
- Mahata, Kaushik, Schoukens, Johan, & De Cock, Alexander (2016). Information matrix and D-optimal design with Gaussian inputs for Wiener model identification. *Automatica*, 69, 65–77.
- Mehra, Raman (1974). Optimal input signals for parameter estimation in dynamic systems—survey and new results. *IEEE Transactions on Automatic Control*, 19(6), 753–768.
- Minorsky, Nicolas (1922). Directional stability of automatically steered bodies. *Journal of the American Society for Naval Engineers*, 34(2), 280–309.
- Nouri, Nowrouz Mohammad, Valadi, Mehrdad, & Asgharian, Jafar (2018). Optimal input design for hydrodynamic derivatives estimation of nonlinear dynamic model of AUV. *Nonlinear Dynamics*, 92(2), 139–151.
- Pivtoraiko, Mihail, Knepper, Ross A., & Kelly, Alonzo (2009). Differentially constrained mobile robot motion planning in state lattices. *Journal of Field Robotics*, 26(3), 308–333. <http://dx.doi.org/10.1002/rob.20285>.
- Thil, Stéphane, Gilson, Marion, & Garnier, Hugues (2008). On instrumental variable-based methods for errors-in-variables model identification. In *Proceedings of the 17th IFAC world congress* (pp. 426–431). Seoul, Korea.

- Umenberger, Jack, & Schön, Thomas B. (2019). Nonlinear input design as optimal control of a Hamiltonian system. *IEEE Control Systems Letters*, 4(1), 85–90.
- Valenzuela, Patricio E., Dahlin, Johan, Rojas, Cristian R., & Schön, Thomas B. (2017). On robust input design for nonlinear dynamical models. *Automatica*, 77, 268–278.
- Valenzuela, Patricio E., Rojas, Cristian R., & Hjalmarsson, Håkan (2015). A graph theoretical approach to input design for identification of nonlinear dynamical models. *Automatica*, 51, 233–242.
- Vincent, Tyrone L., Novara, Carlo, Hsu, Kenneth, & Poolla, Kameshwar (2010). Input design for structured nonlinear system identification. *Automatica*, 46(6), 990–998.
- Wang, Zihao, Soares, C. Guedes, & Zou, Zaojian (2020). Optimal design of excitation signal for identification of nonlinear ship manoeuvring model. *Ocean Engineering*, 196, Article 106778.
- Wernholt, Erik, & Löfberg, Johan (2007). Experiment design for identification of nonlinear gray-box models with application to industrial robots. In *46th IEEE conference on decision and control* (pp. 5110–5116). New Orleans, LA, USA.
- Wernholt, Erik, & Moberg, Stig (2011). Nonlinear gray-box identification using local models applied to industrial robots. *Automatica*, 47(4), 650–660.
- Yoon, Hyeon Kyu, & Rhee, Key Pyo (2003). Identification of hydrodynamic coefficients in ship maneuvering equations of motion by estimation-before-modeling technique. *Ocean Engineering*, 30(18), 2379–2404.