View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Physica A 415 (2014) 503–513

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Quantitative projections of a quality measure: Performance of a complex task

K. Christensen [a,b], Gisle Kleppe [c], Martin Vold [d], Vidar Frette [c,*]

[a] *Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, United Kingdom*
[b] *Complexity & Networks Group, Electrical & Electronic Engineering Building, Imperial College London, London SW7 2AZ, United Kingdom*
[c] *Department of Engineering, Stord/Haugesund College, Bjørnsonsgt. 45, N-5528 Haugesund, Norway*
[d] *Simsea as, Karmsundgata 72, N-5529 Haugesund, Norway*

HIGHLIGHTS

- Quantitative measures as approximations to qualitative concept: quality in task performance.
- Ship-bridge simulator as laboratory for humans interacting with advanced technology.
- The same simulator exercise recorded for several crews under identical conditions.
- Quantitative measures of task performance have been constructed.
- The crews differ significantly under these measures.

ARTICLE INFO

Article history:
Received 11 February 2013
Received in revised form 5 April 2014
Available online 15 July 2014

Keywords:
Human–technology interactions
Ship-bridge simulator
Maneuvering histories
Professional crews
Performance measures
(Qualitative) concept explored through quantitative measures

ABSTRACT

Complex data series that arise during interaction between humans (operators) and advanced technology in a controlled and realistic setting have been explored. The purpose is to obtain quantitative measures that reflect quality in task performance: on a ship simulator, nine crews have solved the same exercise, and detailed maneuvering histories have been logged. There are many degrees of freedom, some of them connected to the fact that the vessels may be freely moved in any direction. To compare maneuvering histories, several measures were used: the time needed to reach the position of operation, the integrated angle between the hull direction and the direction of motion, and the extent of movement when the vessel is to be manually kept in a fixed position. These measures are expected to reflect quality in performance. We have also obtained expert quality evaluations of the crews. The quantitative measures and the expert evaluations, taken together, allow a ranking of crew performance. However, except for time and integrated angle, there is no correlation between the individual measures. This may indicate that complex situations with social and man–machine interactions need complex measures of quality in task performance. In general terms, we have established a context-dependent and flexible framework with quantitative measures in contact with a social-science concept that is hard to define. This approach may be useful for other (qualitative) concepts in social science that contain important information on the society.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

* Corresponding author. Tel.: +47 52702673; fax: +47 52702601.
E-mail addresses: k.christensen@imperial.ac.uk (K. Christensen), gisle.kleppe@hsh.no (G. Kleppe), mv@simsea.no (M. Vold), vidar.frette@hsh.no (V. Frette).

http://dx.doi.org/10.1016/j.physa.2014.06.068
0378-4371/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Spatial and temporal structures that intuitively are referred to as complex arise in many fields of science [1,2]. Growth processes leading to fractal patterns [3,4], solutions to NP-complete problems [5], and the many geometrical conformations of macromolecules like DNA [6] are but a few examples. In some cases, simple quantitative measures can be found that allow characterization and comparison of structures. An example is the fractal dimension, a single numerical value calculated from complex geometries that allows sorting into universality classes [7].

Complex patterns are also generated from man–machine interactions, during operation of advanced technology. Such patterns reflect both strategic choices, in their large-scale features, and standard modes of operation, on a more detailed (small-scale) level. Analysis of these patterns is useful from a safety perspective, for training, and for technology development. In addition, it may serve as a laboratory for exploring new concepts and measures for complex structures. The data is qualitatively different from what one obtains from purely natural systems and from many-agent systems like the stock market. The differences apply to both large-scale features and the small-scale (noise) level. In the present contribution, we make some modest steps in characterizing geometrical structures generated from human interaction with advanced technology. The aim is to investigate whether these simple quantitative measures reflect quality in task performance.

On the methodological side, there may be two long-term benefits from the strategy we establish in this paper. Firstly, as suggested above, new tools for quantitative analysis may be developed as new types of quantitative data are analyzed. Secondly, and most important, implicit in our approach is an attempt to build bridges between 'classical' social science and statistical physics. Classical social science is very different from the studies published as sociophysics in statistical-physics journals. Much of classical social science is qualitative, and proceeds by using deep and complex concepts like intention, culture, group interest, collaboration, alienation, and solidarity. Concepts like these are not easy to define, but convey deep insight into social structure. Our approach is to explore such concepts by their reflections (or projections) on quantitative measures. In this paper the concept is quality in task performance. Note that in our approach, quantitative measures are used as tentative approximations to the concept under study, and are not of fundamental interest as such. Indeed, the main point of our approach is that we focus not so much on the representations (projections) themselves, but on the process of establishing, changing and adjusting them.

A ship bridge simulator is used as laboratory. Simulators allow man–man and man–machine interactions to be studied repeatedly under near identical conditions. The development of simulators as such (modeling of forces, display technology) and their use (didactic, integration into regulations from the authorities) have been extensively researched [8]. However, we are not aware of previous studies where the data logged by the simulator software has been analyzed quantitatively from a statistical-physics point of view.

We analyze complex maneuvering histories of ships obtained from training sessions on bridge simulators. The participants are professional seamen, and not students (that too often have been used in social-science studies). We suggest and develop measures that characterize variations in the way a task is solved by different crews, and, possibly, quality in performance. An improved understanding of man–machine interactions may be obtained through such quantitative measures. Recently, quantitative methods have been successfully used on social interactions of humans—temporal patterns in communication is one example [9].

The quantitative measures of task performance during maneuvering are compared to expert evaluations from experienced instructors. Expert evaluations are intuitive, integrating assessments of complex situations, based on extensive experience. They cannot always, at least not easily, be broken down to a series of parameter values. Thus, expert evaluations represent an alternative to purely quantitative measures of maneuvering histories seen as geometrical objects.

In many contexts, quality in task performance is a concept that is hard to define and hard to quantify. Similarly, there are many other concepts used in social science that are hard to quantify such as culture (specifically: the safety culture of a company) or alienation (classically split into the chain: alienated product, alienated process, alienated self, alienated relations). Still, such concepts convey important, integrated insights into social systems. In our approach, the quantitative measures are approximations to the basic quality concept. Thus, the main interest is not in the quantitative results as such, but in the tentative interpretations of the concept they offer. The choice of quantitative measures may be varied as the quality concept is explored. Assessment of validity is then not primarily based on quantitative consistency but on the depiction of the concept brought forward by the quantitative measures in combination. We develop this approach below using relatively simple quantitative measures. However, it should be possible to apply our context-dependent framework for other social-science concepts using more advanced quantitative measures from the statistical-physics toolbox.

Advanced ships are used in fields like offshore oil exploration: dive support vessels, supply vessels, anchor handling vessels, tugs, cable layers, and multi-purpose vessels. Due to high demands from the operations carried out, these ships need to have very high maneuverability. This is achieved through a propulsion system with several thrusters, water jets, and rudders in addition to standard propellers. For some operations, like maintenance of subsea installations, it is crucial that the ship accurately keeps a fixed position. Therefore, bridge systems usually incorporate equipment for Dynamic Positioning (DP).

DP is a method to automatically keep ships and semi submersible rigs in a fixed position using the propulsion systems instead of anchors. It may also be used for sailing a vessel from one position to another along a predefined route and, like an autopilot on an airplane, DP may operate without human involvement. The method relies on accurate determination of position from external reference systems like GPS, as well as a continuously adjusted mathematical model of the ship and external forces from wind, waves and currents. There are six degrees of freedom associated with the motion of a vessel:

translation along and rotation about each of three perpendicular axes. There is active regulation for three of these (translation along two perpendicular horizontal axes, rotation around the vertical axis), while for the other three degrees of freedom one needs accurate measurements to have a correct interpretation of external position information.

This paper is organized as follows. The data obtained from simulator sessions is described in Section 2 and some quantitative measures extracted from the data in Section 3. A discussion and summary is given in Section 4.

2. Maneuvering data

The simulator software allows logging of a significant number of observables during a training session. A typical data set consists of observables related to vessel position, to the status of the propulsion machinery, and to external conditions like wind and current. Typical behavior of the selection of observables we have concentrated on is given in Figs. 1 and 2.

2.1. Maneuvering task

Fig. 1 shows maneuvering during a simulator exercise, as solved by three different offshore crews. In this exercise, a ship is to be taken up to an installation consisting of three nearby platforms connected by bridges, where a subsea inspection is to be carried out. The installation used in the simulator exercise is an accurate mapping from the Frigg gas field, located at the border between the British and Norwegian sectors in the North Sea. However, this part of the offshore field is no longer in operation, and the installation has been dismantled.

The task consists of three phases: I, taking the vessel up to a position close to the inner corner of the platform installation, and, II, keeping the vessel in this position until a subsea operation (using an unmanned submarine, a so-called Remotely Operated Vehicle (ROV)) has been completed. When the position of operation has been reached and the submarine launched, all DP systems that yield external position references are disabled by the instructor. Thus phase III of the task is to maintain the vessel in position manually, without (full) support of the automation system (DP). The way the crew handles this (unexpected) loss of external reference is one component in an accompanying study based on interviews and observations. Here we concentrate on a quantitative description of the maneuvering histories, taken as geometrical objects.

2.2. Data processing

The data consists of observables recorded at different times t_i given in whole seconds. Measurements are triggered by a minimum change in one or more of the observables. Therefore, the time difference between two consecutive measurements $\Delta t = t_{i+1} - t_i$ is not constant; it has a distribution with a strong peak at 2 s and a tail extending at least up to 20 s. In addition, a small Δt introduces quantization effects on the small scale.

Both effects are undesirable for our investigation. To minimize their effect, we apply the technique of coarse graining, well known from statistical physics. The original data (o_i, t_i), where o_i is an observable recorded at time t_i, is coarse grained such that Δt is as close as possible to 20 s.

Hence, starting with $i = 1$, find the index j such that $t_j - t_i$ is as close as possible to $\Delta t = 20$ s. Then

$$o_{\text{new}} = \frac{1}{t_j - t_i} \sum_{k=i}^{j-1} o_k (t_{k+1} - t_k), \tag{1}$$

$$t_{\text{new}} = \frac{1}{2} \left(t_i + t_j \right), \tag{2}$$

where the observable o_k has been weighted by the time to the next measurement. This represents an average value, because the observable o_k has not changed by more than a (set) threshold value between two consecutive measurements. Now let $i = j$ and repeat the coarse graining.

2.3. Maneuvering characteristics

Due to the many options, including partly or full use of DP, as well as variations in offshore experience and collaboration in the crews, the chosen routes vary significantly. We take the chosen routes as given, basic data. Thus, we will make no attempt to correlate the details of the routes to parameters characterizing the crews.

In Fig. 1, full maneuvering histories are shown in the left column. For crew C1, the vessel is reversed into position, after a period of sideways motion. When the operation has been concluded, the vessel leaves through forward motion. On the other hand, for crews C2 and C3, the vessel is taken into position through forward motion. For crew C2 it is slow, while crew C3 had a relatively fast motion. In both these cases, the vessels leave through backwards motion.

In the right column of Fig. 1, the later phases II and III of the same maneuvering histories are shown in more detail. There is first a period (phase II) where the vessel is kept in position using the automation system (DP) (indicated by black color). After the (imposed) failure of the external position references, the vessel has to be kept in position manually (phase III, indicated by magenta color). These two phases can be characterized by circles indicating the radius of gyration for each phase. For all crews, the position is much better kept when automation (DP) is used than for manual maneuvering, which is no surprise.

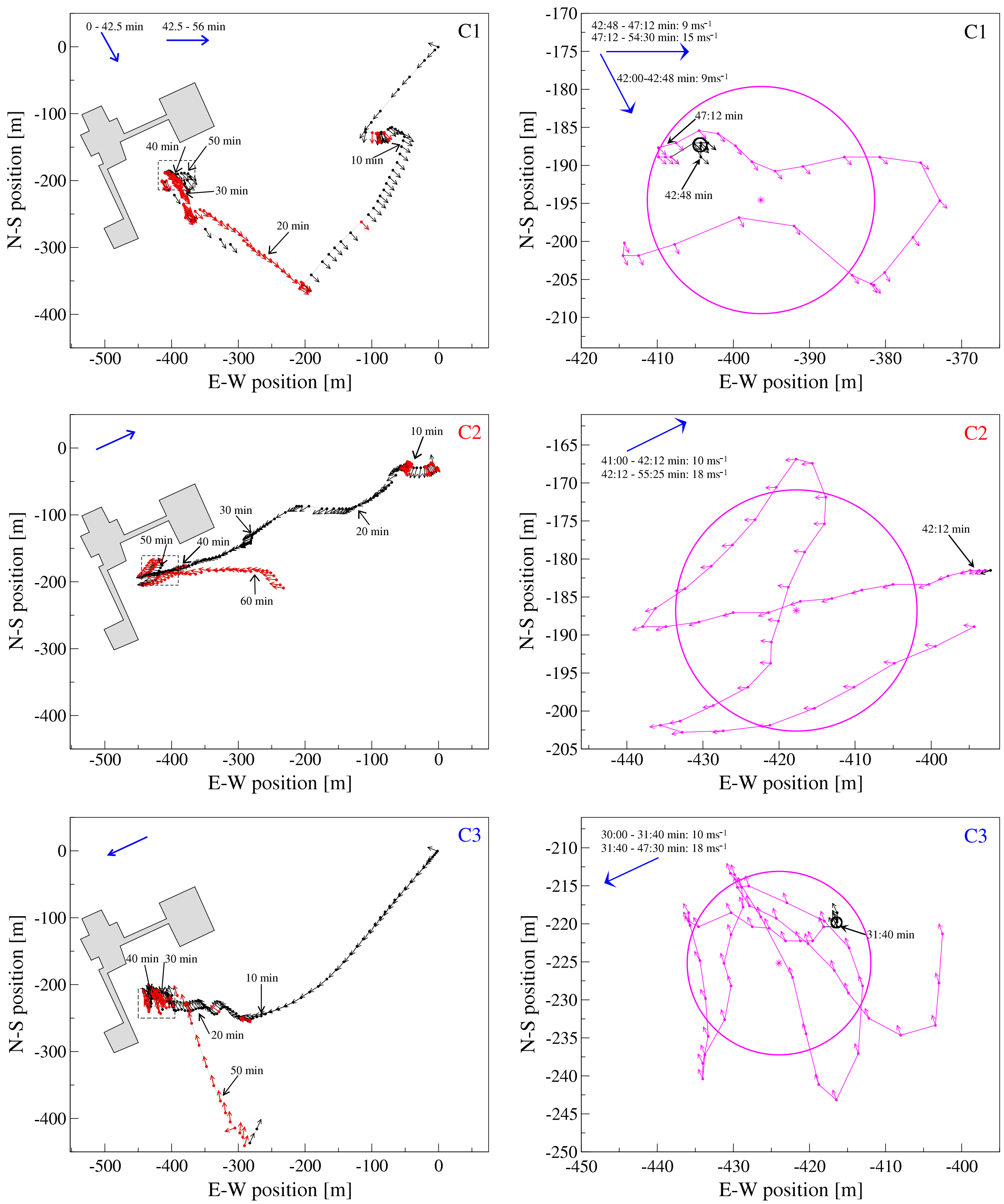

Fig. 1. Three maneuvering histories, for crews C1, C2, and C3, are shown (top to bottom). In the left column, the trace shows the motion of the vessel from an initial position in the upper right corner to a position near three platforms connected by bridges (in gray). The center of gravity is shown at a series of times, with arrows that indicate the direction of the hull. The points (and arrows) shown are separated by 20 s. Note that the direction of the hull often differs significantly from the direction of motion. The vessel was 100 m long, thus, maneuvering close to the platforms was demanding. Red indicates motion in the reverse direction, which corresponds to a negative value for the dot product between hull direction and vessel velocity. In the right column, details on the maneuvering close to the platforms are given. The regions shown have equal area and they are blown-up versions of the dashed regions in the left-column figures. The smaller (black) circle indicates the radius of gyration of the trace for phase II where maneuvering is still controlled by the automation system (DP), the large circle (magenta) manual maneuvering after loss of external position information (phase III). The large (blue) arrow indicates the wind direction, and wind velocities are specified. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

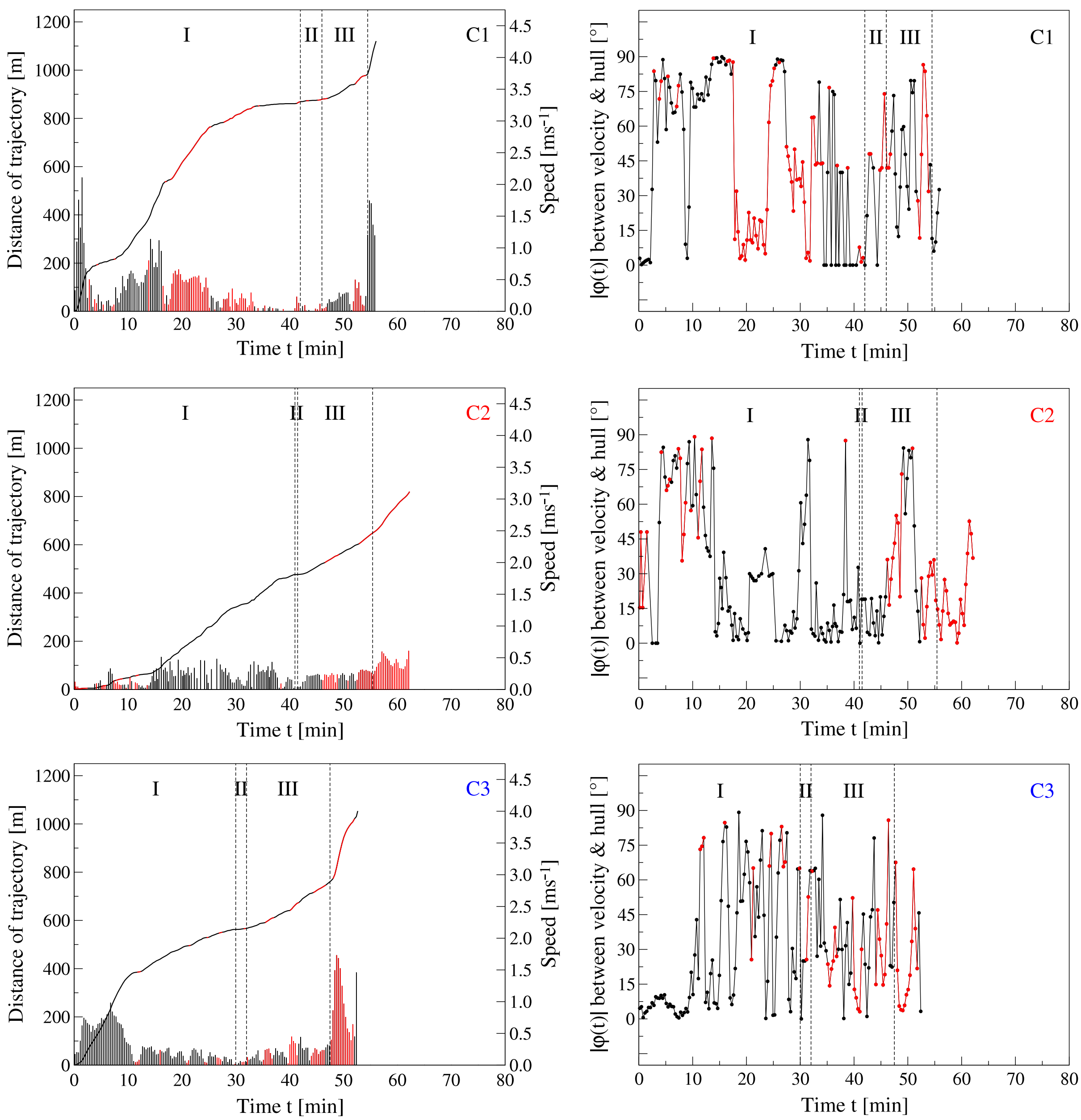

Fig. 2. (Color online) Data extracted from the maneuvering histories shown in Fig. 1. In the left column is shown the total distance the vessel has covered (curve) as well as its speed (spikes) as a function of time. In the right column, the absolute value of the smallest angle between the line given by the hull of the vessel and its instantaneous velocity is shown versus time. The color coding is the same as in Fig. 1. The three vertical lines indicate the end of phases I, II, and III, respectively.

We will have more to say about the radius of gyration in Sections 3.3–3.4. Note that in all cases, the direction of the hull (indicated by arrows) is almost constant while the position of the center of gravity of the vessel varies in an erratic way.

Fig. 2 shows two quantities extracted from the data in Fig. 1. In the left column of Fig. 2, the distance the vessel has covered along its trajectory is displayed, as a function of time (curve). The instantaneous speed is also shown (spikes). Obviously, the distance curve is obtained from the speed data by integration. Note the differences in the distance curves for the three cases, resulting from the speed profiles. There is a lot of fine structure in the speed data, including alternation between motion primarily forward and backwards. There are high velocities when leaving the position near the platforms at the end of the operation, this is in particular the case for crews C1 and C3. Note also the difference in duration of these three maneuvering histories.

In the right column of Fig. 2, the absolute angle between the direction of the hull of the vessel and its velocity (direction of motion) is shown as a function of time for the three crews. During maneuvering of a ship with a simple propeller–rudder

Table 1
Expert evaluation of total task performance, for all nine crews C1–C9. Three independent expert evaluations were averaged to obtain a final ranking of the crews. For the two-way tie 8 and 8, we assign the ranking 8.5 to both. For the three-way tie 6, 6, and 6, we assign the ranking 7 to all. This applies in a similar way for rankings below.

Crew	C1	C2	C3	C4	C5	C6	C7	C8	C9
Expert 1	8.5	5	2	3	1	8.5	5	5	7
Expert 2	1	3	9	7	2	6	4	5	8
Expert 3	5	9	4	3	1	7	8	6	2
Final ranking	3	7	4	2	1	9	7	5	7

design, this angle will be close to 0° most of the time. The vessel used in the simulator exercise we consider, on the other hand, has high maneuverability and is easily moved in any direction relative to the hull direction. The plots in the right column of Fig. 2 show that the crews use these options to a large extent. The pattern in angle variations differs between the three crews, with a more persistent signal for crew C2.

2.4. Data sets

We have recorded data for maneuvering carried out by nine crews while solving the exercise described in Section 2.1. There are significant variations between these nine cases for all representations used in Figs. 1 and 2. In addition to the maneuvering data, we have for each case a detailed record of observations made during the training sessions as well as post-session interviews. However, such qualitative data is not used in the analysis that follows below.

2.5. Crews

The exercise was carried out during DP courses for professional seamen. Many of the participants had significant experience at sea, but not with DP. The simulator courses we have analyzed are parts of the requirements to be certificated as a DP operator. Two seamen worked together during the exercise described above. The teamwork in these two-person crews varied much, from close collaboration, via one participant carrying out all tasks, to direct conflict.

2.6. Expert evaluations

In Section 3, we will consider quantitative measures of the maneuvering carried out by different crews. Expert evaluations, on the other hand, are qualitative assessments of performance. It is challenging to collect such evaluations in a systematic way. In the literature, there is not even a consensus on what an expert is.

We have used experienced instructors at the simulator as experts. They know well both the navigation trade and the training situation at the simulator. The ranking of the crews is based on input from three instructors, see Table 1. We decided to base these rankings solely on plots shown in Fig. 1, and the experts were given these plots a long time after the training sessions were carried out. This ensured that the rankings were established under identical conditions. We have additional, richer material relevant for quality in task performance through comments given during the training session, post-session interviews with instructors, and observations on communication, special events and similar that the plots do not reveal. To obtain transparent rankings, this additional material has not been used in the present context.

The assessments differ between our experts. It seems clear that they do not use the same references. These variations, as well as the arguments behind, are as such an interesting topic that will not be further pursued here.

3. Measures of performance

In the previous section, we emphasized the variations between the maneuvering histories for several features. The variability reflects differences in the strategic choices of the crews. We have characterized some of the features quantitatively, that is, we have constructed measures that extract parts of the information contained in structures (maneuvering histories) like those shown in Figs. 1 and 2.

These quantitative measures have been constructed in such a way that they, possibly, reflect quality in the performance of each crew. Three of these measures will be described below. We compare these measures to expert evaluations from experienced instructors, see Section 2.6.

Qualitatively, there are striking differences between the maneuvering histories in Figs. 1 and 2. Under the measures below, the crews also clearly differ. We take this as a first indication that the measures may be useful.

3.1. Phase I: Time spent

A very simple criterion is the time spent, during the first phase of the operation, to reach the position for the subsea inspection. The hypothesis is that a short time reflects quality in task performance. The times varied significantly: The shortest time spent was 30 min, the longest 73 min, see Table 2.

The assumption that the time spent on phase I of the task (reaching the position of operation) is an indication of quality is questionable. Obviously, a 'the-faster-the-better' approach may compromise safety. The high day rates paid for these vessels inevitably pushes towards fast operations. On the other hand, a more detailed and strict regulatory regime has been established in recent years, and fast operations have less emphasis than earlier.

3.2. Phase I: Integrated absolute angle

We consider now the angle between the hull direction and the direction of motion. As mentioned above, the vessel has high maneuverability and is easily moved in any direction relative to the hull direction. However, the options for sideways motion should not be used needlessly: First, fuel consumption is significantly higher during sideways motion. Second, extensive use of sideways motion necessitates more frequent maintenance. Therefore, the hypothesis is that low use of sideways motion reflects quality in task performance. In this quantitative measure, the absolute angle between the hull direction and the direction of motion is integrated along the trajectory.

Results for the nine crews are shown in Fig. 3(a). Here, the integrated angle is shown as a function of time. Note that the integrated angle is meaningful as a measure of quality in task performance only until the position of operation is reached, and this is how far each curve has been drawn. For each crew the value of the integral at the end point is taken as the measure. Values are given in Table 2, and they reflect the accumulated costs of sidewards motion. The crews differ under this measure: there is a factor 3.6 between the lowest (best) and highest (worst) value.

One may argue that this measure gives too much weight on the time spent to reach the position of operation: Fig. 3(a) shows that for all crews the integrated angle increases steadily. Thus, a crew that spends a long time, will also have a large value for the integrated angle, and one would expect time and integrated angle performance indicators to be highly correlated.

To mitigate for such an effect, one may divide the integral by the integration interval. This quantity is shown in Fig. 3(b), as a function of time. This quantity reflects the slopes in Fig. 3(a), which differ. However, it is difficult to obtain a unique ranking of the crews from the curves in Fig. 3(b).

3.3. Phase II: Radius of gyration with DP

The second and third phases of the task consist of keeping the vessel in a fixed position near the inner corner formed by the three platforms, see Section 2.1. During the second phase, this is achieved with the support of the automation system (DP), during the third phase without.

We have used the radius of gyration to quantify the linear extent of the irregular motion of the vessel during the second and the third phase of the operation. The radius of gyration R_g is the root mean square distance of the trajectory $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ measured from its center of mass $\mathbf{r}_{cm}$ [10], that is,

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_{cm})^2}, \tag{3}$$

$$\mathbf{r}_{cm} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i. \tag{4}$$

The number of points N in the trajectory increases with time, so effectively R_g and $\mathbf{r}_{cm}$ are functions of time. The radius of gyration while the vessel is kept in position using DP (referred to as the second phase) is shown in Fig. 4, as a function of time. Note that for most of the crews, R_g saturates at a value of (1 ± 0.5) m. This is the characteristic size of the random motion of the center of mass. Relative to the size of the vessel (which is 100 m long), DP keeps position extremely well.

The behavior of R_g during the second phase (under automation) is hardly a measure of quality in task performance. It will be used instead as a reference for the third phase of the operation.

3.4. Phase III: Radius of gyration without DP

The same measure, the radius of gyration, for the third phase of the operation (keeping the vessel in position without full support of the automation system) is shown in Fig. 5(a). One notes that R_g has gone up by a factor of ten compared with the value observed with DP (Fig. 4). This is not surprising, obviously, the automation system keeps the position better than can be done manually. It is also clear from Fig. 5(a) that there are significant differences between the crews.

Consider now three hypothetical but extreme cases where (1) the ship drifts off along a given direction (e.g. due to wind or current), (2) the ship initially drifts off but the crew manage to limit the trajectory to random motion within a given area, or (3) the ship initially drifts off but the crew manage to maneuver the ship back to the working position and stabilize it there. This would result in (1) $R_g(t)$ increasing linearly with time, (2) $R_g(t)$ initially increases with time but then crosses over to a constant value, or (3) $R_g(t)$ initially increases with time but then decreases again and goes towards zero. Hence, in order to evaluate the crew's performance, it makes sense to take notice of the history of the radius of gyration as a function of

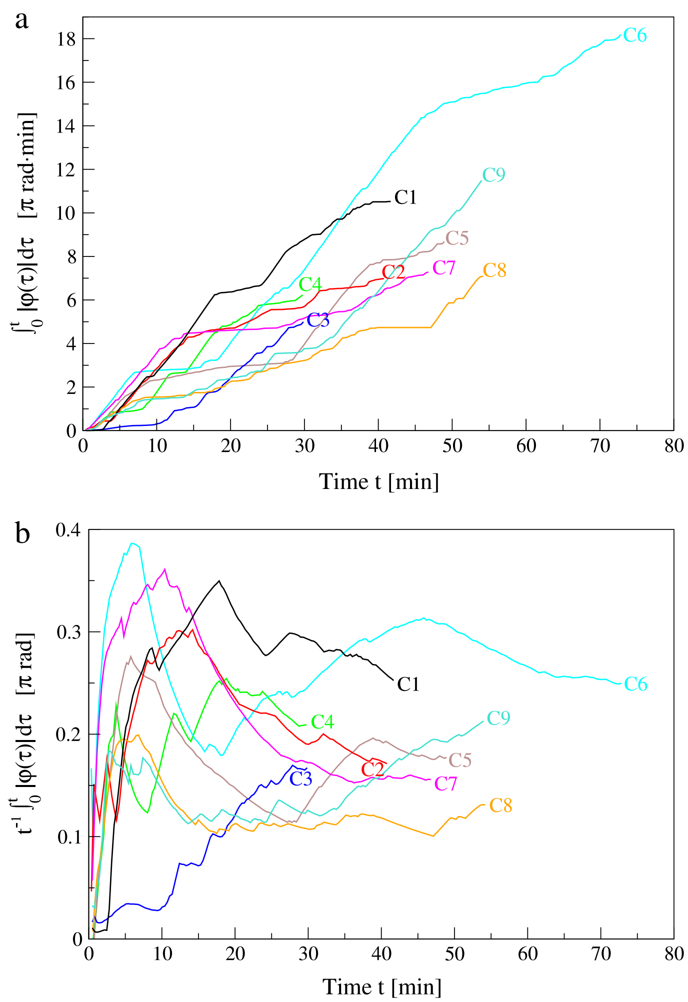

Fig. 3. (Color online) Angle between hull and velocity. (a) The integrated absolute angle between the direction of the hull and the direction of motion, versus time. Results for each of the 9 crews are marked and shown using different colors (the same color coding will be used in later figures). For each crew, the curve is drawn up to the time where the position of operation (close to the platforms) is reached. (b) Mean angle, obtained as the integrated absolute angle between the direction of the hull and the direction of motion (shown in the upper plot), divided by the integration time, as a function of time.

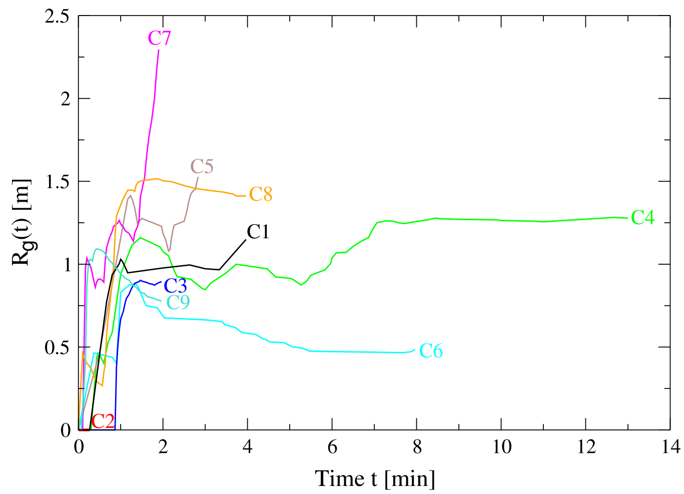

Fig. 4. (Color online) Radius of gyration, phase II. Radius of gyration versus time for the vessel trajectory for the second phase, where a fixed position is to be kept using DP. The radius of gyration remains bounded at a relatively small value of (1 ± 0.5) m for all crews.

Table 2
Values for each crew, for the three quantitative measures of performance: *Time* is the time used to reach the position of operation (phase I), *angle* is the integrated absolute angle between the direction of the hull and the direction of motion (phase I), and *radius* is the integrated radius of gyration after loss of position reference systems (phase III). The values for radius were taken at time 7.5 min in Fig. 5(b), that is, shortly before the C6 curve ended. For crew C8, the third phase was too short for a radius value to be obtained.

Crew	C1	C2	C3	C4	C5	C6	C7	C8	C9
Time (min)	42	41	30	30	49	73	47	54	54
Angle (π rad · min)	10.5	7.0	5.0	6.2	8.7	18.2	7.3	7.1	11.5
Radius (m · min)	67	75	47	8.5	23	16	55		19

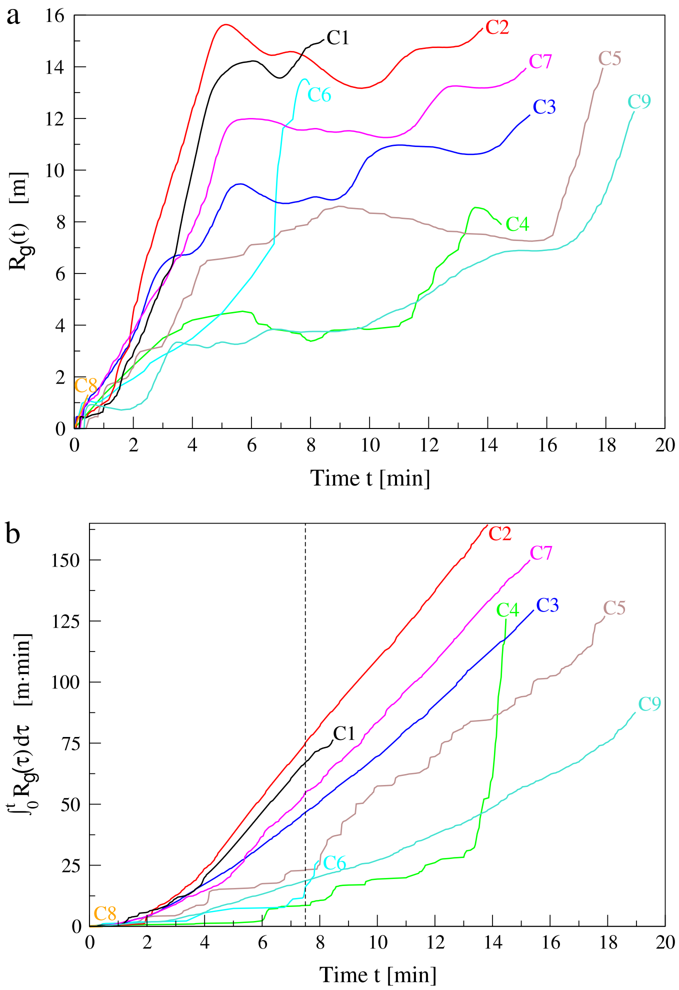

Fig. 5. (Color online) Radius of gyration, phase III. (a) Radius of gyration versus time for the vessel trajectory for the third phase of the operation, where the vessel is to be kept in a fixed position manually, without the use of DP. (b) The integrated radius of gyration $\int_0^t R_g(\tau)\mathrm{d}\tau$ versus time. The intersections of the vertical dashed line placed at 7.5 min and the curves yield the values given in Table 2 for the radius. Note that the rank is almost independent of when it is measured.

time. The simplest way to implement such a measure would be to integrate the radius of gyration $\int_0^t R_g(\tau)\mathrm{d}\tau$ as a function of time t.

A reasonable hypothesis is that a low value of $\int_0^t R_g(\tau)\mathrm{d}\tau$ is an indicator for high quality in task performance. This measure is shown in Fig. 5(b) and demonstrates significant differences in the quality of task performance. The values given in Table 2 were read off from Fig. 5(b) at 7.5 min, indicated by the dashed vertical line.

4. Discussion and summary

The data shown in Figs. 1 and 2 may be analyzed without much reference to how it was generated, for instance using concepts like interacting spins on clusters, persistence, and probability distributions. We have not used such an approach,

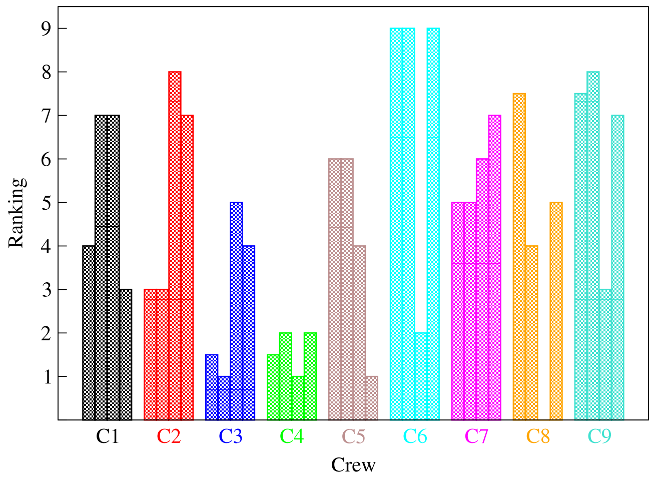

Fig. 6. (Color online) Rankings for the nine crews, in the order: time, angle, radius, expert. Note that time and angle both refer to phase I, radius to phase III while experts' ranking is based on phases I–III. The time, angle, and radius rankings were based on Table 2 while the expert ranking is the final ranking from Table 1. In each case, the best crew is given ranking 1, the worst ranking 9. Note that crew 8 has not been ranked under radius, since the third phase was very short in this case, see Fig. 5.

which would focus merely on abstract geometrical entities. In our opinion, an analysis should reflect the fact that the data is generated by humans interacting with advanced technology. Moreover, the data was generated as certain tasks were carried out, and we have incorporated quality in task performance in the analysis. While such a concept fits the data well, it is demanding to find an objective standard to measure against. Purely geometrical measures are easy to interpret, but it is difficult to find a basis for measuring quality in task performance. Expert evaluations do not provide such a basis, as discussed below.

Two perspectives on the performance of the nine crews have been discussed above. Expert evaluations were discussed in Section 2.6, with results in Table 1—and in Section 3, three quantitative measures were considered, with the values obtained for each crew given in Table 2. The results are summarized in Fig. 6, where the quantitative measures have been transformed to rankings. In all cases, 1 is the highest quality in the set, 9 the lowest.

Fig. 6 displays a reasonable coherence in the ranking of the crews along the dimensions *time*, *angle*, *radius*, and *expert*. If one simply averages ranking over these four dimensions for each crew one obtains an overall ranking: C4, C3, C5, C1 and C2, C8, C7, C9, C6.

However, on a more detailed level, there is a low degree of correlation between rankings under the four dimensions. Assuming linear relationships, we find support for a correlation only between time and angle ($R^2 = 0.73$) [11]. This is reasonable, since time is one component in the angle measure, see discussion in Section 3.2. All other combinations are dominated by outliers, and there are no correlations between the data sets (over the nine crews).

Thus, our data does not support definite functional relationships between the four dimensions or groups thereof (with one exception, as mentioned above). We do not believe this is merely a matter of obtaining larger data sets, but connected to fundamental differences between measures.

The expert evaluations were in principle based on all three phases of the exercise, taking into consideration any aspect of the task performance the expert found to be of importance. The experts had access to both types of plots shown in Fig. 1, but in effect based their evaluations only on plots like the ones in the left column. These plots are easily interpreted based on previous experience with maps, while plots like the ones in the right column in Fig. 1 are more unfamiliar. The orientation of the vessel seemed to be an important ingredient in the expert evaluations: one view was that it should be towards the wind, another that it should be in the direction of easy escape (in case of emergency), which means that the vessel should be reversed into position. We do not believe that a more coherent set of expert rankings would have emerged with more experts.

Time rank, angle rank, and radius rank all relate to only one of the phases and the measures were based on one observable expected to reflect quality. Time rank and angle rank are related to phase I and the integrated radius of gyration to phase III. One may speculate that quality in performance during phase I of the operation should be correlated with quality during phase III, but our data do not support such a conclusion.

In summary, we have analyzed detailed maneuvering histories obtained from simulator training sessions. This is a promising type of socio-physical data that allows detailed quantitative analysis and reflect strategic choices. Simulators are excellent environments for exploring complex situations that combine social interactions and man–machine interactions. Participants in the training sessions we followed were professional seamen, thus there was a high level of realism on the crew side. Experiments with different crews under close to identical conditions are easily carried out.

We have studied a simulator exercise that is relatively open, and thus allows the crews to pursue different strategies. Such an open design is interesting from a scientific point of view. On the other hand, educationally, challenges at several levels are required, from the training of elementary operations, where trainee actions are strongly prescribed, to more open situations. Should quantitative comparison between the performances of various crews be desirable in an educational setting, we suggest that the instructors apply a protocol for the training sessions that would ensure approximately equal duration of phase II for all crews and equal duration of phase III for all crews.

Using both quantitative measures and expert evaluations, we have ranked the performance of nine crews. There is a reasonable coherence in the rankings along the four dimensions, when taken together. However, there are no detailed correlations, the rankings along each dimension are not simple functions of each other, except for the case of time and integrated angle. Had there been simple functional relationships between time, angle, radius, and expert rankings, one of them would have been sufficient to characterize performance.

This may indicate that complex situations like the one we have studied necessitate complex measures of quality in task performance. Phrased differently, this concept could to some extent be approximated by the quantitative measures discussed above. However, there seems to be more to quality in task performance than these measures reveal. Thus, additional measures should be sought for the present or similar cases. It would be very interesting to study different, but open and complex, situations during simulator training.

Acknowledgments

KC greatly appreciates the hospitality of Stord/Haugesund College, Norway during research visits. This research has been supported by the MAROFF programme under the Research Council of Norway, as part of the project *Managing Complexity in Petro-Maritime Operations.* The Research Council of Norway have not influenced the way the research reported in this paper was designed or carried out.

References

[1] L.P. Kadanoff, From Order to Chaos. Essays: Critical, Chaotic and Otherwise, World Scientific, Singapore, 1993.
[2] K. Christensen, N.R. Moloney, Complexity and Criticality, Imperial College Press, London, 2005.
[3] J. Feder, Fractals, Plenum Press, New York, 1988.
[4] T. Vicsek, Fractal Growth Phenomena, World Scientific, Singapore, 1992.
[5] M. Mezard, G. Parisi, M.A. Virasoro, Spin Glass Theory and Beyond, World Scientific, Singapore, 1987.
[6] C.A. Laughton, S.A. Harris, WIREs Comput. Mol. Sci. 1 (2011) 590–600.
[7] B.B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman and Company, New York, 1977, 1982.
[8] The Conference Marsim 2012, see http://www.marsim2012.com/ (accessed 6th January 2013).
[9] H.H. Jo, M. Karsai, J. Kertész, K. Kaski, New J. Phys. 14 (2012) 013055.
[10] D. Stauffer, A. Aharony, Introduction to Percolation Theory, Taylor & Francis, London, 1994.
[11] The $R^2 = 0.73$ was obtained from the measured quantities of the observables time and angle. If one evaluates Spearman's rank correlation coefficient (done on the rankings) we find 0.82 that also indicates that the ranks to a good approximation are monotonically correlated.