Applied Computing and Geosciences 12 (2021) 100072

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

# An interactive sequential-decision benchmark from geosteering☆

Sergey Alyaev[a,*], Sofija Ivanova[c], Andrew Holsaeter[a], Reidar Brumer Bratvold[b], Morten Bendiksen[c]

[a] NORCE Norwegian Research Centre, Postboks 22 Nygårdstangen, 5838, Bergen, Norway
[b] University of Stavanger, Postboks 8600 Forus, 4036, Stavanger, Norway
[c] Bendiksen Invest og Konsult, St. Olavs vei 156, 5063 Bergen, Norway

ARTICLE INFO

*Keywords:*
Interactive benchmark
Sequential geosteering decisions
Uncertainty quantification
Expert decisions
Experimental study
Decision support system

ABSTRACT

During drilling, to maximize future expected production of hydrocarbon resources, the experts commonly adjust the trajectory (geosteer) in response to new insights obtained through real-time measurements. Geosteering workflows are increasingly based on the quantification of subsurface uncertainties during real-time operations. As a consequence, operational decision-making is becoming both better informed and more complex. This paper presents an experimental web-based decision support system, which can be used to both teach expert decisions under uncertainty or further develop decision optimization algorithms in a controlled environment. A user of the system (either human or AI) controls the decisions to steer the well or stop drilling. Whenever a user drills ahead, the system produces simulated measurements along the selected well trajectory which are used to update the uncertainty represented by model realizations using the ensemble Kalman filter. To enable informed decisions the system is equipped with functionality to evaluate the value of the selected trajectory under uncertainty with respect to the objectives of the current experiment.

To illustrate the utility of the system as a benchmark, we present the initial experiment, in which we compare the decision skills of geoscientists with those of a recently published automatic decision support algorithm. The experiment and the survey after it showed that most participants were able to use the interface and complete the three test rounds. At the same time, the automated algorithm outperformed 28 out of 29 human participants.

Such an experiment is not sufficient to draw conclusions about practical geosteering but is nevertheless useful for geoscience. First, this communication-by-doing made 76% of respondents more curious about and/or confident in the presented technologies. Second, the system can be further used as a benchmark for sequential decisions under uncertainty. This can accelerate development of algorithms and improve the training for decision making.

## 1. Introduction

Geosteering is the intentional control of a well trajectory based on the results of down-hole real-time geophysical measurements (Shen et al., 2018). Traditionally, research in geosteering has been focused on the interpretation of log measurements. As a result, during the last decade, there has been a steady growth of automated methods for measurement inversion and interpretation which yield steadily growing amounts of data that need to be handled by the decision-makers. This data opens the possibility to target the oil-bearing zones which were not economically viable previously (Larsen et al., 2016). At the same time, this also makes the decision-making more complex by adding more relevant information to consider and evaluate in real-time (Hermanrud et al., 2019). Complex uncertainties that can impact decisions on one hand and contradicting objectives on the other (see e.g. Halset et al. (2020)), pose difficulty for decision-makers and requires new workflows and/or training.

The literature review in Kullawan et al. (2016) showed that there was hardly any prior publication that considered a consistent framework for geosteering decision-making with several objectives. The authors prepared an alternative decision-focused approach addressing geosteering as a sequential decision problem. Specific to optimization of sequential decisions under uncertainty is that the acquired data, and hence the uncertainty for subsequent decisions, will depend on the previous decisions.

☆ Link to the code: https://github.com/NORCE-Energy/geosteering-game-gui, see Section 8 for description.

* Corresponding author.
*E-mail address:* saly@norceresearch.no (S. Alyaev).

https://doi.org/10.1016/j.acags.2021.100072
Received 6 January 2021; Received in revised form 16 August 2021; Accepted 31 October 2021
Available online 3 November 2021

During the last seven years, more publications were addressing the optimization of geosteering decisions. Chen et al. (2015); Luo et al. (2015) considered an ensemble-based method for optimization of reactive steering under uncertainty. Kullawan et al. (2018) demonstrated the application of dynamic programming for finding optimal long-term decision strategies for a certain set of geosteering problems. In Alyaev et al. (2019), a simplified dynamic programming algorithm was used in a context of a more general geosteering problem with several targets. Veettil and Clark (2020) developed a Bayesian estimator of stratigraphy that can be further extended with forward well planing. Kristoffersen et al. (2020) proposed a reinforcement-learning-based approach to steering based on the initial field planning.

To test the developed methodologies, the papers reviewed above relied either on a static benchmark sacrificing uncertainty updates (Kristoffersen et al. (2020) used Olympus (Fonseca et al., 2018)) or needed to develop a testing platform alongside the decision methodology (Chen et al., 2015; Luo et al., 2015; Kullawan et al., 2018; Alyaev et al., 2019). On one hand, the descriptive optimization benchmarks either contain static pre-defined environment which is not influenced by decisions (e.g. Olympus, see Fonseca et al. (2018)) or are ambiguous due to the open choice of both model-updating and decision strategies (e.g. Brugge, see Peters et al. (2013)). On the other, the benchmark tool-kits for developing and comparing sequential decision agents, such as OpenAI Gym (Brockman et al., 2016), are primarily targeting reinforcement learning for games and simple physical simulators.

While the individual benchmarks have been derived for the newly developed algorithms, the publications addressing the expert decisions are still limited to summaries of current best practices, see e.g. Przybylo (2019). The experience from other branches of geosciences shows that decision-making can be significantly improved and de-biased when supported by information technologies and artificial intelligence (Wilson et al., 2019). Thus, there is a need for benchmark systems providing an uncertain geological environment where the uncertainty changes consistently in response to taken decisions.

For this study, we have developed a web-based platform that can update a multi-realization 2D geological model in response to decisions and share the current state of the system via an Application Programming Interface (API) and a Graphical User Interface (GUI). The GUI uses a newly-developed visualization that shows how the geometric uncertainty relates to the expected value of the planned well. Thus, the platform serves as a benchmark for decision-agent development and as a tool that can compare algorithms' to experts' decisions.

We also present the first experimental use of the web-based platform, in which formation evaluation and geosteering experts competed alongside the fully automated system from Alyaev et al. (2019) to get the highest well-value in synthetic drilling operations with several possible target layers. Unlike a more typical task of following a layer, the multi-target geosteering cannot be solved optimally by experts through a visual analysis of the average image. Neither can it be solved by a convex optimization algorithm. The purpose of the experiment was to compare the decisions of the experts with the fully automated algorithm in controlled conditions. The web platform handled the complex updates of the uncertainty and only left the decision-making process to the experts/algorithms.

This paper is organized as follows: First, we describe the experimental setup which is currently used for the platform. After that, Section 3 explains the discretization and the automatic updates performed by the system during the experiment. Section 4 presents the GUI and the API available to the users of the platform (the corresponding code is linked in Section 8). In Section 5 we present the first experiment run on the platform, including the results of the experts and the feedback that we have received. Finally, the findings of the paper and further perspectives for the platform usage are summarized in Section 6.

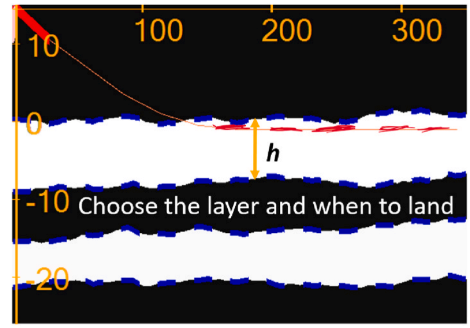


**Fig. 1.** An example of synthetic truth and a possible steering trajectory in the X-Z coordinates. The dashed part of the trajectory is within a sand layer (white with thick dashed boundary) and gives a positive score. $h$ is the thickness of the sand layer.

## 2. Description of the experiment

In this section, we describe the setup of the geosteering decision-making experiment that runs in our sandbox environment.

An experiment is split into rounds. The objective for the decision-making task in each round is to make landing and steering decisions in a multi-layer geological setting. Each round consists of several (at most 12) decisions for changing direction or stopping which can be made in sequence. After every decision, the system drills a virtual drill stand and updates the uncertainty according to the simulated data. Thus, every next decision would be influenced and informed by the choices from the previous decisions. The decision locations are evenly spaced along the X-axis.

### 2.1. Objective

The pre-drill model (as well as the synthetic truth) contains five alternating layers: shale-sand-shale-sand-shale. The goal is to maximize an approximate Net Present Value (NPV) of the well. This is done by landing and staying near the roof of a sand layer with considerations of layer thickness and drilling costs. More specifically, the objective score is calculated by the following rules: the participant gets:

- $h$ points for every meter in sand layer (along X-axis), where $h$ is the layer thickness
- $2*h$ points when they drill in the sweet-spot near the roof (0.5 m–1.5 m from the top boundary of sand)
- negative $c$ points is the cost of drilling every meter, where $c = 0.086$.

This objective can be written in the form of an equation:

$$NPV = \sum_{i=1}^{2}\left[\int_{z_i^{top}(x)>y(x)>z_i^{bottom}(x)} h_i(x)dx\right] \tag{1}$$

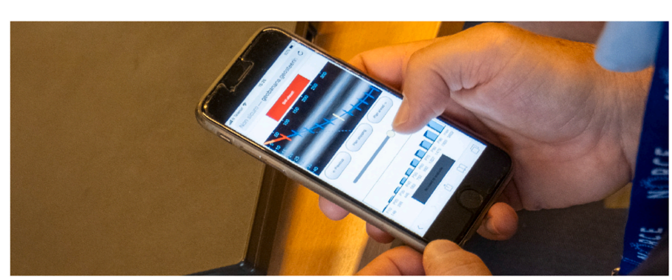

**Fig. 2.** The GUI running on a mobile phone during the experiment.

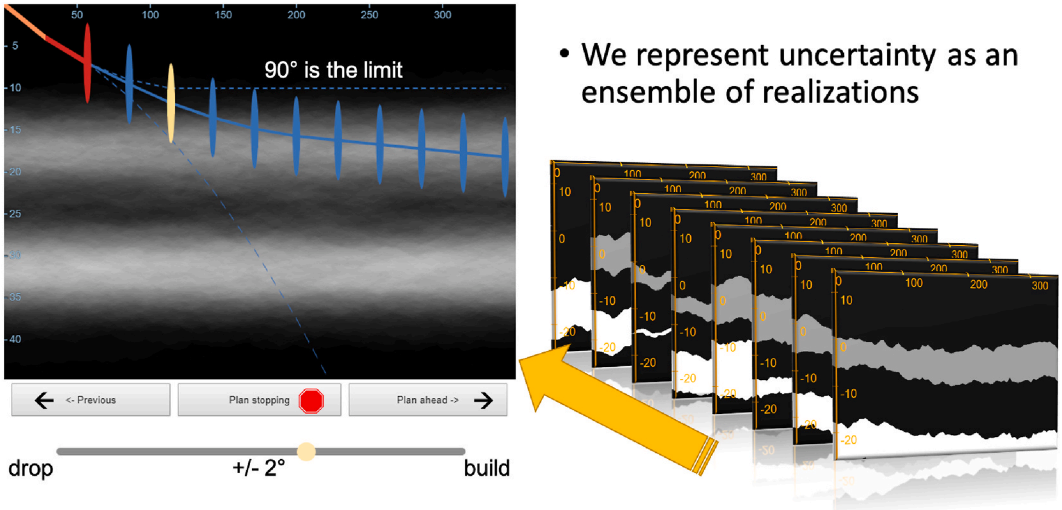


**Fig. 3.** A part of the GUI showing geological uncertainty, controls for steering, and steering limits. The uncertainty is represented by an ensemble of realizations and visualized as an overprint of all ensemble members. The ellipses represent decision points and their size indicates the look around of the EM tool. The dashed blue line represents the steering potential of the well after the current decision; the steering is limited by dog-leg severity and the 90-degree inclination. The orange part of the trajectory is already drilled; the red part is the next decision to commit to; the blue part is the plan ahead. The yellow ellipse shows the selected point at which the steerer can adjust the dip. The selection can be moved by the buttons. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$+\int_{z_i^{top}(x)+0.5>y(x)>\min(z_i^{top}(x)+1.5,z_i^{bottom}(x))} h_i(x)dx \Bigg] \tag{2}$$

$$+0.086\int_{x_0}^{x_n}\sqrt{1+(y'(x))^2}dx, \tag{3}$$

where $y(x)$ is the well trajectory, $y'(x)$ is its derivative, $z_i^{top/bottom}$ are the positions of the boundaries for sand layers indexed by $i$, and $x^0$ and $x^n$ are the start and the end of the trajectory. We use $y(x)$ for the depth value of the trajectory to distinguish it from the layer geometry. An example of synthetic truth and a possible steering trajectory is shown in Fig. 1.

## 3. Uncertainty and automatic updates

Like any decision, geosteering decisions are made under uncertainty. The main uncertainty during geosteering is the lack of complete knowledge of geology through which the well will be drilled. We represent uncertainty using an ensemble of 120 realizations of the layered geology in 2D.

The original ensemble is based on a prior distribution which is also used to generate the synthetic truth in the experiment. The layer boundaries are discretized along the X-direction. Y-positions for every boundary are generated using a variogram model with kriging as described in Alyaev et al. (2019).

The ensemble of realizations (each parametrized as Y-positions of the boundaries) is updated automatically following each decision using the Ensemble Kalman Filter (EnKF) algorithm described in Luo et al. (2015).[1] For the update, we use measurements produced from a synthetic EM tool which is located at the drill-bit and has a look-around capability of 4.8 m up, down, and sideways. The model for the tool is described in Chen et al. (2015). The resistivity is assumed constant within each layer for all models. It equals 10 Ωm in shales and 150 Ωm in sands. The system performs one update between decision points which uses measurements in three equally distributed locations.

## 4. Description of the web-based platform

One of the main purposes of the developed web-based platform is to enable comparison between experts and algorithms when it comes to decision-making for geosteering. The developed web-based platform includes both a GUI for experts and an API for AI bots. In this section, we first describe the capabilities of the platform through its GUI and then explain how corresponding API can be used by a bot. We finish the section with a summary of the implementation of the Decision Support System optimization algorithm (DSS-1) from Alyaev et al. (2019).

### 4.1. The GUI

To evaluate the decision-making strategies of the experts, we developed a simple online decision-support GUI. The online mobile application gives the contestants the same information that a geosteering decision algorithm would get, wrapped into a user-friendly GUI, see Fig. 2.

#### 4.1.1. Uncertainty visualization

One of the primary sources of information for the system users is the graphical display of the uncertain earth model. The users can view an overprint of the ensemble which provides a display of the (white) sand layers' location uncertainties (Fig. 3).

While the overprint gives an intuitive representation of the mean positions of layers and the uncertainty magnitude, this information is insufficient to take quantitative decisions in scenarios with several targets. This is somewhat similar to interpretation difficulties observed when interpreting a seismic image without context (Bond et al., 2007).

#### 4.1.2. Planning and committing to a decision

The discrete locations of the geosteering decisions to be made in this round are shown in Fig. 3.

Until the well is finalized, the steerer can plan the entire well (ahead of the current point) by changing the dip of the well in the decision points (ellipses in Fig. 3). Alternatively, the contestant can decide to stop

[1] The EnKF is a Monte-Carlo (discrete) approximation of the Kalman Filter. It gives an approximation of a Bayesian update with Gaussian priors and likelihoods.

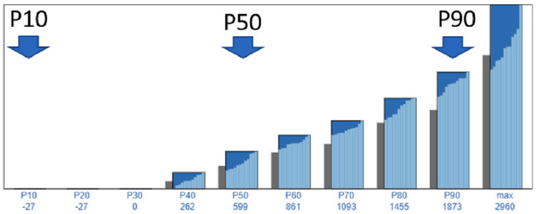


**Fig. 4.** A score distribution diagram shown to a user based on the current set of realizations representing the uncertainty. Each thin column shows a value based on one realization, which is sorted from small to large. The columns are grouped for convenience to approximate P10, P20, etc from the ensemble. Since percentiles are not exact, but ensemble estimates the right column is denoted as 'max' and shows the maximum value observed in the ensemble.

drilling at any of the points. The latter might be optimal if, for example, the well entered the under-burden.

At each decision point the participant must commit to a decision: choose whether to adjust the dip for the next drill-stand or stop drilling. After every decision to drill ahead, the web server updates the uncertain realizations using the data along the chosen segment of the trajectory (see Section 3). Stopping decisions implies that the well is finalized, and no further drilling steps can be taken.

#### 4.1.3. *Tools to make informed decisions*

To aid in their decision-making, the contestants are presented with a visual decision support tool in the GUI. The GUI dynamically updates uncertainty as the well is drilled and helps to estimate the well-value.

Once a well-trajectory is planned, it can be evaluated using the scoring function with respect to the current understanding of uncertainty (the ensemble). The results of this evaluation are summarized in a bar diagram as shown in Fig. 4. The diagram shows a cumulative density diagram based on the 120 ensemble members (light blue). The results are grouped into percentiles of value (P10 – P90) shown in dark blue. Note that the highest value based on the ensemble approximation is not equivalent to P100; that is why it is denoted by 'max'. The GUI also shows the percentile values for the previous evaluation as gray bars in the background. Fig. 5 shows the evaluation before and after the trajectory update. The gray bars appearing on the right figure are scaled comparatively to the new values in the diagram.

The percentile - cumulative density diagram is interactive. The user can select a percentile to see the subset of realizations that give the selected value range, e.g. between P70 and P80 (Fig. 6).

### 4.2. *The API*

In this section, we describe the API, which is used by the GUI of the system and can also be used for developing decision agents/AI bots.

#### 4.2.1. *Uncertainty communication*

The uncertainty is communicated to the system using the native format of the ensemble-based earth model. The system transmits an array of 120 realizations, each containing the grid of X-positions in an ascending order together with four boundaries represented as a

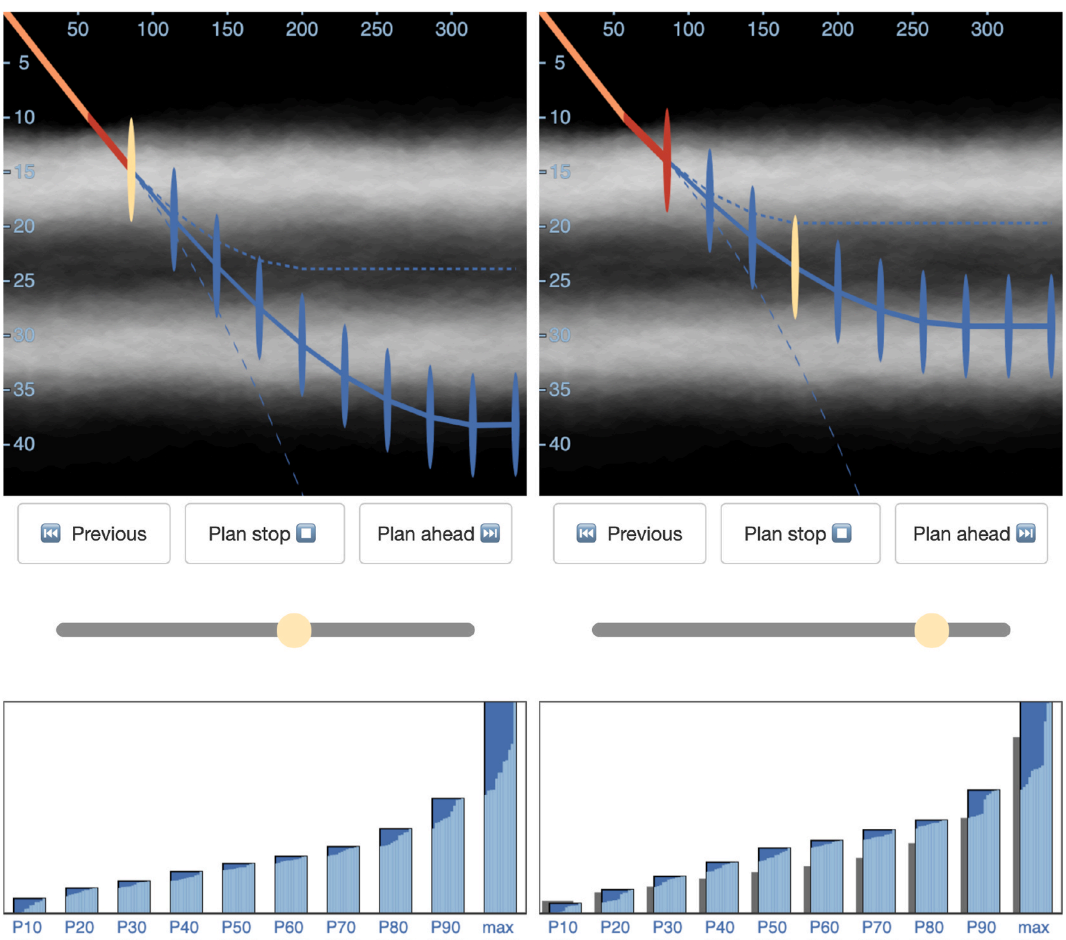


**Fig. 5.** An example of change in the score distribution diagram after changing the trajectory and recalculating the score. The left panel shows a trajectory and a distribution before an update. The right pane shows the updated score distribution overlayed on top of the gray bars which correspond to the score distribution before the update, scaled to the new distribution.

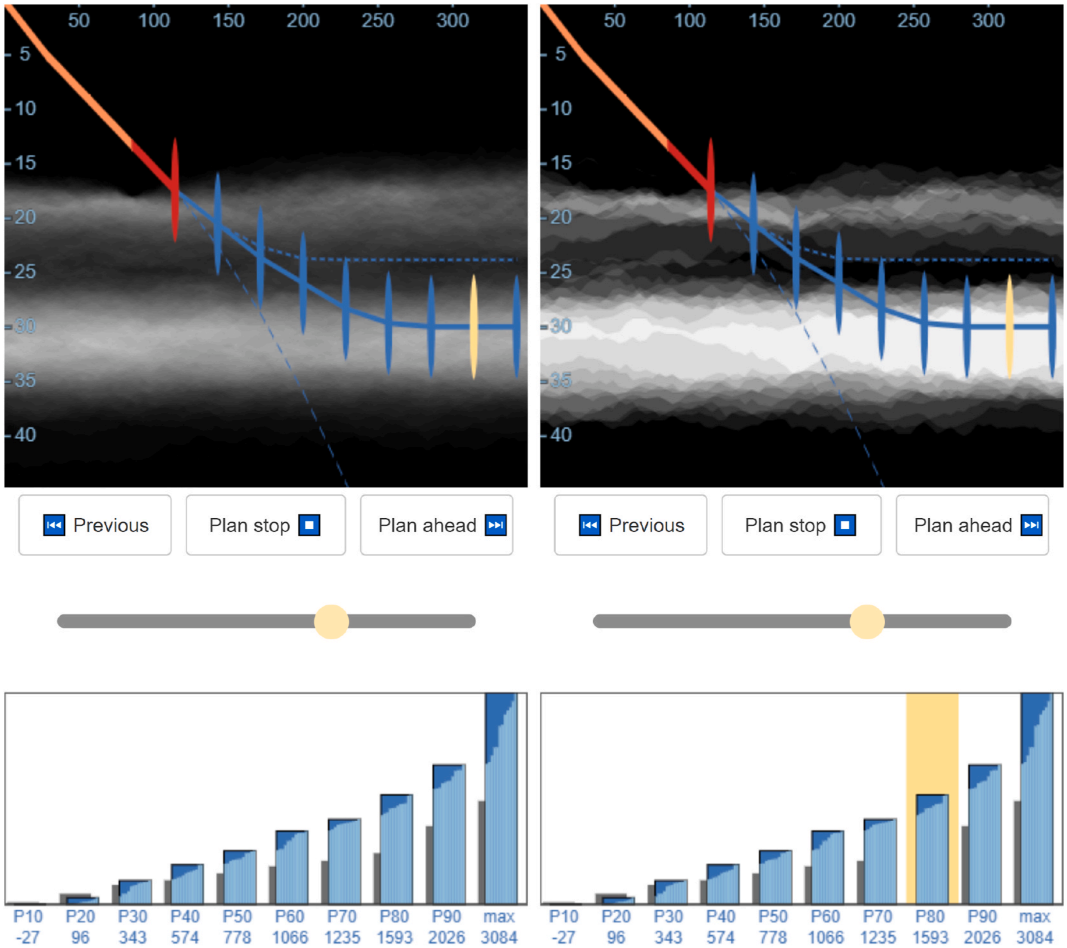


**Fig. 6.** An example of selecting a subset of realizations using the interactive score distribution diagram explained in Fig. 4. The left panel shows the full uncertainty distribution in the geology overprint from 120 realizations. The right panel shows the overprint of 12 realizations, which give a score between P70 and P80 (selected in the distribution diagram by yellow). In the selected optimistic scenario the trajectory enters a thick bottom sand layer. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

sequence of corresponding Y-positions. The first two boundaries are the roof and the floor of the top sand layer, and the second two correspond to the bottom layer.

#### 4.2.2. Committing a decision

For each sequential decision point, the API gives the trajectory that has already been drilled. A bot needs to commit to a decision of either continuing or stopping.

The continuing decision should be represented by the coordinates of the committed point $(x, y)$, where $x$ is fixed for a decision step and $y$ should be within the limit on the dog-leg-severity angle ( $\pm 2°$). After the decision to continue, the API will return the updated realizations based on the acquired data along the drilled segment.

After stopping, the system receives the score obtained within the synthetic truth as well as the rating relative to other participants of the round. The actual score can be used for the training of data-based algorithms.

#### 4.2.3. Making informed decisions

Similar to human participants, a bot can request the evaluation of a score for a given trajectory. Given a sequence of $(x, y)$ pairs, the API returns an evaluation of the given trajectory represented by a sequence of points for each member in the ensemble of realizations. The response is an array of pairs, each consisting of the score and the index of realization for which that score was observed.

### 4.3. Implementation of the DSS-1 bot

The web-based platform was developed from the functionalities that have been used by DSS-1 described in Alyaev et al. (2019). The update part of the DSS-1 workflow is used by the web platform directly and was described in Section 3. Here we summarize the optimization algorithm of DSS-1.

At every decision point, the optimization is divided into two steps: global deterministic optimization for each realization and robust optimization for the immediate decision.

On the first step DSS-1 considers all discrete trajectories until the end of the operation. The discretization is made so that for the decision step $x_0$ the system uses a finite number of depth possibilities $(y_i)$ distributed on a selected regular grid. For every one of these trajectories, the system stores the score of the well for each of the realizations. In this paper, we use a version with a discount factor for future decisions. That is, every next segment's value is multiplied by 0.9 and thus has an influence on the score of the well, which compensates for the uncertainty of the future learning. The details about the efficient implementation of DSS-1 are presented in Alyaev et al. (2019).

On the second step DSS-1 considers all discrete alternatives for the next (immediate) decision, i.e. all $(x_1, y_i)$ within the geometric constraints and stopping and chooses $y_i^{\text{opt\_next}}$, which gives best decision on average:

$$y_i^{\mathrm{opt\_next}} = \arg\max_{y_i} \frac{1}{n}\sum_{j=1}^{n}\left\{O([(x_0,y_0),(x_1,y_i)]|M_j) + \gamma O([(x_1,y_i),\ldots]|M_j)\right\}, \quad (4)$$

where $n = 120$ is the number of realizations, $O(\psi|M_j)$ is the objective function computed for trajectory $\psi$ against realization $M_j$, $[(x_0, y_0), (x_1, y_i)]$ is a segment from current point to the next decision point ending in $y_i$, $O([(x_1, y_i), \ldots]|M_j)$ is the optimal value for trajectory starting in $(x_1, y_i)$ in the realization $M_j$ computed in the first step, and $\gamma$ is the discount factor. Here we use $\gamma = 0.9$.

Based on the description above, we can summarize the following properties of the DSS-1 decisions:

- The DSS-1 optimization algorithm described above is a deterministic function of the discretization, realizations, and the objective function.
- The DSS-1 only uses the evaluation of the objective function and not the representation of the earth model. This makes it flexible with respect to earth model implementation and the objective function. At the same time, it does not use all the information available in the experiment.
- The global optimization used by DSS-1 requires up to 100 000 evaluations in 40 s (Intel XEON 8168 2.7 GHz single core[2]) of the objective function for every decision which is impossible for a human. We note that the required number of evaluations is controlled by discretization and can be reduced if the objective functions are more costly. For practical applications, fast proxy models of NPV would be required for real-time performance.

## 5. Results and discussion

The web-based platform has been developed to compare the DSS-1 to human experts, and through this comparison, to communicate the concepts related to real-time decision making under uncertainty. In this section, we first present the result of the first experiment which was held as a plenary session of the Formation Evaluation and Geosteering Workshop 2019 by NFES and NORCE held in Stavanger Norway (NORCE, 2019). We evaluate the decision quality of experts and DSS-1 based on the collected data. After that, we show the results of the survey among the participants of the experiment which addresses the usefulness of such an experiment in the communication of the research. Finally, we summarize feedback about the web-based platform in the context of ease of use and application to training for geosteering.

### 5.1. *The first experiment*

The rounds of the experiment were organized as follows. After the presentation of the rules (Section 2) and the GUI (Section 4.1), the contestants had a practice period of 15 min to familiarize themselves with the competition set-up. During this period an expert of the system was showing the usage of the features of the GUI and a possible strategy for geosteering on a big screen.

Following the demonstration, there were three scoring rounds of approximately 6 min each. All rounds had an identical ensemble of starting realizations but a different synthetic truth unknown to the participants. To evaluate the consistency of decisions, the truths were chosen as:

- Round 1) The bottom sand layer was optimal
- Round 2) The top sand layer was optimal

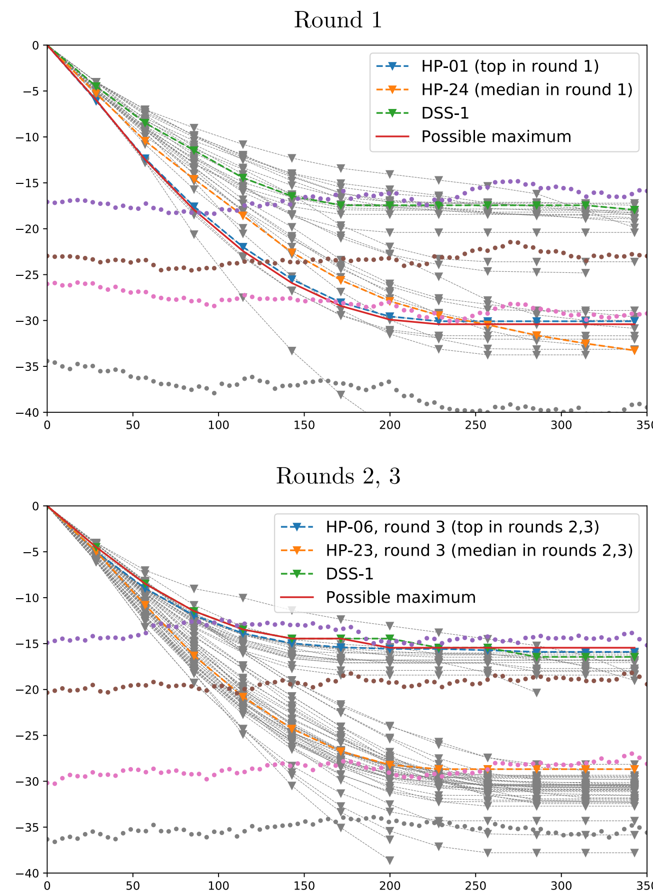


**Fig. 7.** The well-trajectories drilled by participants in each of the three rounds and corresponding synthetic-truth boundaries as dotted lines. The highlighted trajectories show the top participant result; the median participant result; DSS-1 result; and the solution obtained by optimization assuming perfect information (Possible maximum). Note, that since rounds 2 and 3 had identical setups, the trajectories are summarized from both rounds.

- Round 3) Identical to Round 2, to allow comparison of the consistency of contestants' decisions under the same conditions.

Note that the sub-optimal layer was also giving a positive score as described by the objective function in Section 2. The synthetic truths and the optimal solution computed by deterministic optimization on the synthetic truth, are shown in Fig. 7.

Out of the 75 workshop participants, 55 participated in all three rounds. The wells drilled by all the participants are compared with the optimal trajectory in Fig. 7. A fraction of the 55 participants were affected by software issues (22 participants) or did not reach any of the sand layers in one or more rounds (3 participants). For fairness, we disregard them from the results and consider the remaining 30 participants, whom we call **qualified participants**. Among the qualified participants was DSS-1 described in Section 4.3. Since the majority of participants not affected by the software issues managed to drill meaningful wells for in all three rounds, we consider that the basics of the user interface were understood from the original demonstration.

Clearly, in a real-world drilling operation with full complexity, the geosteering experts would outperform an automated system if they possess geological knowledge and operational experience beyond what is built into an automated system. In this controlled experiment, we put the decision-makers and the system in equal conditions in terms of information availability.

[2] The optimization can run all realization in parallel, but single-core set-up is chosen to prioritise server responsiveness to human participants.

#### 5.1.1. Discussion of decision strategies

Decision analysis defines a good decision as the one that is logically consistent with the alternatives (steering choices), information (representation of geological uncertainty), and values (objectives) brought to the decision (Bratvold and Begg, 2010; Abbas and Howard, 2015). In the decision analysis process, the outcome of a single decision does not imply the quality of that decision. That is, given the uncertainty, a good decision may lead to a bad outcome and vice versa. The decision analysis framework allows the identification of good decisions before knowing their outcome by recording the principal inputs of the decision-making process and the corresponding decision strategies.

The decision strategy of DSS-1 (see Section 4.3) derived and analyzed in detail in Alyaev et al. (2019) is designed on the principles of robust optimization, which are known to lead to better decisions under uncertainty. The decision strategies of human participants, however, are not so easy to deduce and analyze. One possible approach is to survey participants about their strategies (Welsh et al., 2005; Alyaev et al., 2021a). Conversely, the ambition of the experiment presented here is to assess the quality of decision strategies of the participants based on the decision-outcome data recorded in a sand-box environment.

Nevertheless, before describing the ranking let us hypothesize some possible strategy elements that can be employed by the participants. Simplistically the decision-making during the experiment can be divided into three sub-tasks selecting 'the best' layer and landing it and following the roof of the layer. The tasks of landing a well and following the top boundary are typical for geosteering training (Alyaev et al., 2021a). Selecting a layer based on the prior uncertainty was a task that should have been new to most of the participants.

For selecting the layer, there are three main possible strategies.

1. Randomly pick one of the layers for every round. Then steer aggressively dismissing the other sand layer and using the data to land the well in the optimal spot in the selected layer.
2. Select the layer by analyzing the prior and try to optimally land into this same layer in all three rounds.
3. Repeatedly evaluate uncertainty in the objective, and allow for landing in either layer using the information about steering potential. Once the tool reaches the top sand layer, use the reduced uncertainty to decide on the landing target. The decision about the target can be possibly made until the end of the operation, but due to steerability limitations, the target layer would have been selected by the inclination around 150 m along X-axis.

Prior information indicated that either layer gives a positive score, and in practice, all three strategies yielded positive gain. However, only strategies 2 and 3 will result in consistent decisions. Strategy 1 is partially inconsistent due to the random selection of the layer, but possibly a consistent landing strategy in the selected layer. In the presented experiment the optimization used by DSS-1 used Strategy 3. According to the interview (Alyaev et al., 2021a), the top-scoring participant used the inconsistent strategy 1.

The outcome of every decision is a result of both skill and chance. Therefore, achieving good results over a short experiment with a partially inconsistent strategy is quite possible. We admit, however, that the results might not objectively reflect the decision strategies for real geosteering because the controlled environment was not familiar to the experts and since the competition format might influence the participants' behavior.

#### 5.1.2. Ranking

There is no unique method to assess the results of the competition over several distinct rounds as it requires scaling the results by a chosen metric. As a primary simple metric (selected prior to the experiment), we used the percentage of maximal possible results for each round which was averaged to give the final ranking. The results are scaled by the 100% result obtained by discrete optimization on the synthetic truth for

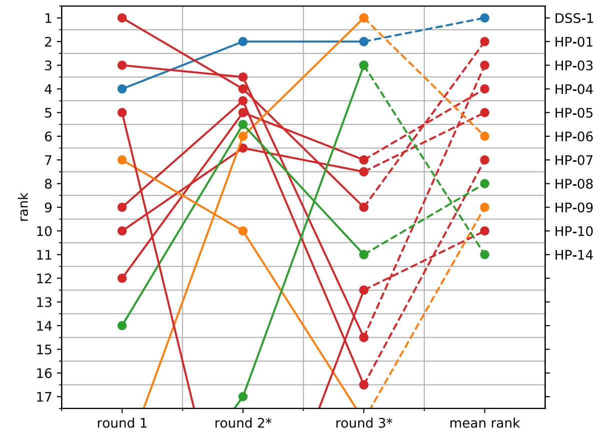


**Fig. 8.** The comparative ranking of the top participants over three rounds of the experiment. Rounds 2 and 3 followed the identical setup and rank for these rounds includes participants' results from both rounds (60 results). The rank* is derived by scaling this rank to 30 participants, resulting in fractional values. The mean rank is the rank (1–30) based on the mean of the three rounds. For convenience, this figure uses the same color-coding as Fig. 9. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

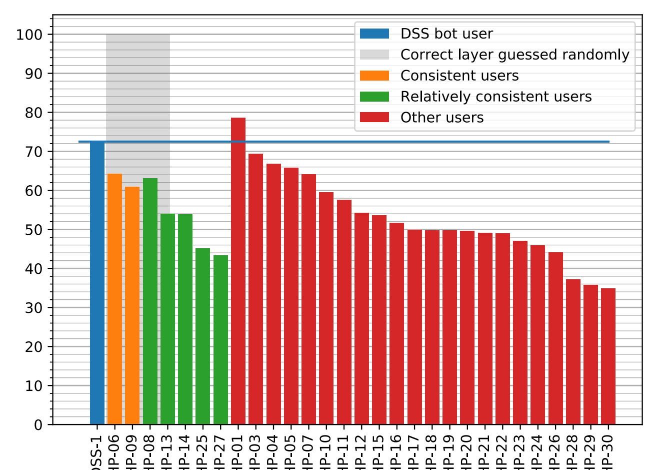


**Fig. 9.** The average performance of the participants over three rounds compare to theoretically maximum value, %. The colors show levels of consistency of decisions when geosteering in the same environment for all the participants: DSS and Human Participants (HPs). Consistent decisions result in the trajectories that are the same (DSS), almost the same (Consistent), or the same layer (Relatively consistent), see Fig. 10. The gray area indicates the number of participants who would have guessed the correct layer if they were guessing at random. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

each round, thus representing a close approximation to the theoretically possible maximum. The data from the experiment, including scoring, is available in the linked repository, see Section 8. The human participants are identified as HP-*n*, where *n* is the rank (1–30) according to this metric. The fully automated decision system performed better than 93% of the participants placing 2nd among the 30 qualified participants.

Another possibility to compare the results of different rounds is to consider the ranking within the population. The rank in the population is the position of the participant in the round among the other participants relative to the total size of the population. In the case of our experiment, we take advantage of two identical rounds and arrive at rank*, common

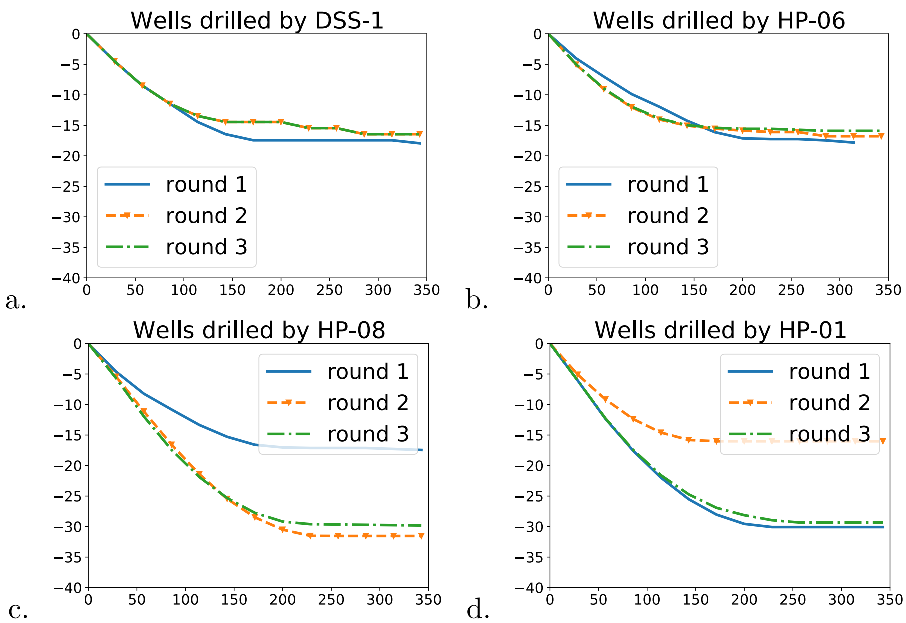


**Fig. 10.** Examples of trajectories drilled by participants over the three rounds. Rounds 2 and 3 have the same test set up, where the best solution was to drill into the top sand layer. Round 1 has the other test, where the best solution was to drill into the bottom sand layer. Consistent (predictable) decision-making results in identical or similar results for rounds 2 and 3. Each plot showing the best participant in its consistency group (see Fig. 9): a. absolutely consistent; b. consistent; c. relatively consistent; d. other.

for rounds 2 and 3. This type of ranking of the top 11 participants is shown in Fig. 8.

While DSS-1 was second in the simple ranking discussed previously, it gets the top position in this alternative ranking. The simple ranking is highly influenced by results in a single round. HP-01 got a near-perfect score (92%) in round 1, making him the top simple-ranked participant. At the same time, neither HP-01, nor any other participant has beaten DSS-1 in more than one round, bringing DSS-1 to the top of the comparative ranking.

As we see, comparative ranking is more objective as it reduces the influence of chance from a single round. Therefore, we are planning to adopt comparative ranking as primary for future experiments.

### 5.1.3. *Consistency of decision strategies*

The automated DSS is built to ensure consistency. That is, the system is guaranteed to make the same decisions given the same input parameters. This means repeatability, and hence, predictability of automated decisions. The same perfect consistency would be impossible for human participants due to various factors, including the GUI limitations. Nevertheless, a good geosteerer should make his decisions based on a process consistent with their objectives, alternatives, and information. As uncertainty is an element of every decision, being consistent in the decision making itself does not guarantee a good outcome for every decision (sequence). A consistent decision making process will ensure good outcomes in the long run (on average). The experiment allowed us to test the extent to which humans were consistent in their decision-making by comparing their decisions in identical rounds 2 and 3. Fig. 10 shows the trajectories drilled by several participants with different levels of consistency:

- DSS-1, which produces identical trajectories for identical set-ups.
- Consistent users, for whom the distance between trajectories for the same set-up was less than 0.5 m on average.
- Relatively consistent users, for whom the consistency was worse than for consistent users, but the trajectories were in the same layer for the same round. More formally, the distance between the trajectories in the same test was at least 2std lower than between different cases. Note that *std* here is the standard deviation of the average distance between the pairs of trajectories based on all combinations of tests 1–3.
- Other users for whom the consistency for the same set-ups was not observed.

All the results arranged by the level of consistency are shown in Fig. 9. For comparison, the figure shows a gray area of selecting the optimal layer purely by chance. [3] Thus, if all the participants did not use any relevant knowledge and tried to land and drill in a layer chosen randomly, about four of them should have selected the correct layer in all three rounds. From Fig. 9 the number of consistent users is lower than the probability of random guessing. The consistent and relatively consistent users together are still within possible error given the relatively small number of total participants.

Another important observation from Fig. 9 is that the consistent users ended up with a relatively low total score. Conversely, the top-scoring participants (expected to be better geosteerers) should consistently maintain their good decision strategies, while the low-scoring participants should learn and improve (in this particular setup the consistent strategies translate to similar outcomes). We observe that for HPs, the strategy, which scored highest, involved chance (early betting on which layer to land). This is confirmed by the interview of the top-scoring participants (Alyaev et al., 2021a). The interviews also reveal that the competition setting of the experiment made some of the participants alter their strategies towards more risk-taking compared to realistic geosteering (e.g. randomly selecting the sand layer and trying to guess the layer's top to get to the sweet spot for double score). Therefore, the observed results do not necessarily reflect what experts would have done in a real operation.

## 5.2. *An interactive experiment as means of communication*

After the experiment, we asked the workshop participants to respond to an anonymous survey. The results in this section are based on responses from 21 participants. The respondents rated the experiment as part of the workshop above 4 out of 5 on average. The other questions were designed to evaluate the web-based experiment platform itself as well as its applicability for training and communication of research.

[3] By guessing, one has a 50% chance to guess and aim for the optimal layer in each round. Given three rounds, a participant has a 1/8th chance to aim for the optimal layer all three times.

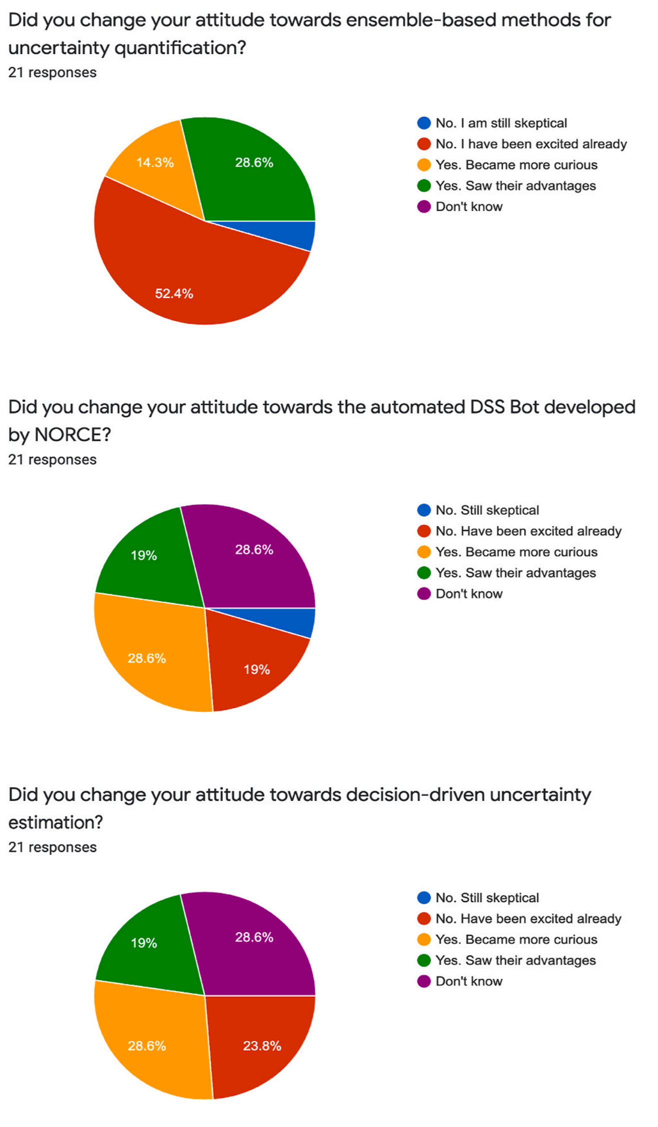


**Fig. 11.** The value of the experiment in terms of communication of the research based on the survey.

In this subsection, we present the feedback about the usefulness of such an experiment for communication of the research. Connected to the experiment, we presented the ensemble-based geosteering workflow which was used behind the scenes in the web-based platform as well as the optimization behind the DSS-1.

In the survey, we tried to see if the interactive experiment allowed us to make people more interested in the related research. Fig. 11 shows whether and how the respondents changed their attitude towards the presented research concepts. The groups who were 'engaged' include those who became more curious and those who saw the advantages of the presented research. We see that 47.6% of respondents were engaged by automated decision-making and decision-driven uncertainty evaluation. The lower 42.9% engagement for the ensemble-based uncertainty quantification can be related to the much higher percentage (52.4%) of the respondents who were already excited about these technologies. While we lack comparative data for standard workshop presentations, the communication-by-participation seems very powerful based on the results of the survey.

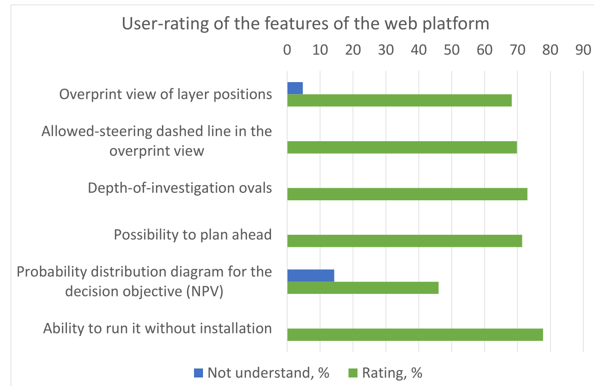


**Fig. 12.** The relative importance of the features of the web-based platform according to the conducted survey. 100% corresponds to all respondents considering a feature "Very useful". The diagram also shows the percentage of respondents who did not understand a feature.

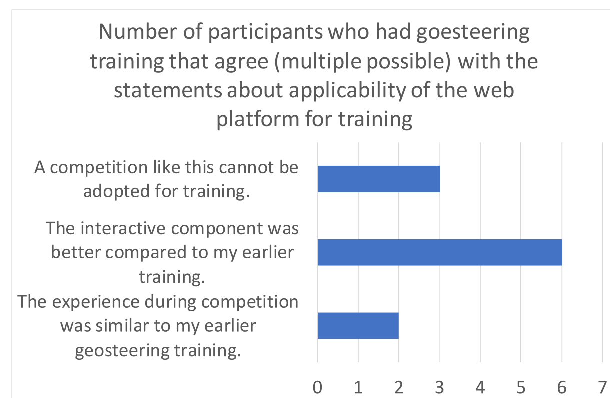


**Fig. 13.** Value of the web-based platform for geosteering training. This was an optional question to those who had prior training in goesteering.

### *5.3. User feedback to the web-based platform*

It is known that uncertainty visualization influences decision-making (Viard et al., 2011). Therefore, in the second part of the questionnaire, we asked the participants to evaluate the usefulness of different elements of the platform to allow its further improvement. This included the elements of the GUI, but also the ability to use the platform from any browser-equipped device.

The respondents evaluated every feature of the platform on the following scale:

● I did not understand it
0 Not useful
1 Somewhat useful
2 Quite useful
3 Very useful

Fig. 12 shows the relative importance of each feature as the sum of the scores by the rules above in percent. I.e. 100% correspond to all participants answering "Very useful". It also shows the fraction of the participants who did not understand a feature.

The ability to run the GUI without installation has the highest relative importance. This supports our design decision to develop the web-based GUI for the experimental platform as a paradigm for user-based experiments.

Looking at the rest of the distribution in Fig. 12, we observe that most features of the GUI have a rating of around 70%. This indicates that overall communication of the GUI during the experiment was successful despite its relatively short format. The lowest scores correspond to probability distribution and the overprint display. These two features of the GUI were designed to communicate the uncertainty, which is not straightforward. Further experimental studies are required to improve uncertainty communication.

We also asked how such an experiment would stand against the current training practices in geosteering (for the participants with prior training). Fig. 13 shows that only three respondents think that the experiment cannot be adopted for training in geosteering. Most of the respondents, however, think that the interactive component is similar (two participants) or better (six participants) compared to standard training practices. We anticipate that the data from the presented survey can be used for further improvement of the platform towards a useful educational tool.

## 6. Conclusions

In this paper, we have presented a web-based platform that provides users with the opportunity to perform assisted decision-making under uncertainty in a benchmark environment. While there have been previous benchmark environments made for development and training AI agents, to our knowledge, the presented platform is novel in the geoscience/geosteering context. Moreover, it provides the possibility to compare the results of decision-making by human experts with automated algorithms providing similar information to both.

In this paper, we present the first experiment, which put 29 geoscientists against the DSS algorithm from Alyaev et al. (2019). The results show that DSS-1 outperformed all but one qualified participant considering relative wells' value; and every one based on comparative rating. The experiment shows that even decisions of the highest-scoring participants are influenced by random chance, confirming more general work (Kahneman, 2011) in the context of geosteering under uncertainty. However, according to an interview (Alyaev et al., 2021a), the strategies of participants could have been affected by the artificial setting of the experiment and thus cannot directly yield conclusions about operational decision making.

The feedback collected from the participants indicates that the experiment provided both entertaining and educational value. As a result, 76% of respondents became more curious and/or saw the advantages of the technologies related to the experiment. These include ensemble-based representation and updating of uncertainty, decision-driven uncertainty estimation, and technologies behind DSS-1. Thus, we can conclude that the platform is valuable as a medium for research communication. Future series of experiments can provide more detailed experts' reactions to the technology of updating and visualization of uncertainty in geosteering operations.

Based on the feedback, future work should include improvement of visualization and/or communication of uncertainty. This includes also more complex scenarios, such as 3D uncertainty and complex sedimentary environments. The updating and static visualization for these scenarios were demonstrated in Fossum et al. (2021) and Alyaev et al. (2021b) respectively.

## Authorship statement

Sergey Alayev: Algorithm and web-platform development, set-up of the experiment and the survey, analysis of the experimental results and the survey, and writing the original draft. Sofija Ivanova: UX design of the web platform, presentation of the GUI of the web-based platform, set-up of the survey, and revising the draft. Andrew Holsaeter: set-up of the experiment and the survey, analysis of the experimental results, and revising the draft. Reidar Bratvold: set-up of the survey, analysis of the experimental results, writing decision analysis description, and revising the draft. Morten Bendiksen: web-platform development, analysis of the experimental results.

## Computer code availability

The web-based API and GUI described in this paper is developed by the authors of the paper in 2019 and, since October 2020, is available as a stand-alone repository "API and GUI for GEOSTEERING benchmark the NORCE way" at https://github.com/NORCE-Energy/geosteering-game-gui under the MIT license. The code contains a reference implementation of a decision agent in Python 3 (API), and the Python server that runs a copy of the GUI of the web-based in JavaScript/HTML/CSS locally. The code has been tested with Python 3.7 and Google Chrome 86.0.4240.111 on macOS v.10.15.7 and Windows 10 v.1709.

The repository also contains the files with the anonymized results of the first experiment described in Section 5.1 and a script for their playback.

For any questions related to the repository, please, contact the corresponding author: saly@norceresearch.no, +47 518 75 610.

## 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

We thank Robert Ewald for help with the deployment of the platform for the experiment.

This work was supported by the research project 'Geosteering for IOR' (NFR-Petromaks2 project no. 268122) which is funded by the Research Council of Norway, Aker BP, Equinor, Vår Energi, and Baker Hughes Norge.

## References

Abbas, A.E., Howard, R.A., 2015. Foundations of Decision Analysis. Pearson Higher.

Alyaev, Sergey, Holsaeter, Andrew, Bratvold, Reidar Brumer, Ivanova, Sofija, Morten, Bendiksen, March 2021. Systematic Decisions Under Uncertainty: An Experiment Towards Better Geosteering Operations. In: Paper presented at the SPE/IADC International Drilling Conference and Exhibition, Virtual.

Alyaev, S., Suter, E., Bratvold, R.B., Hong, A., Luo, X., Fossum, K., 2019. A decision support system for multi-target geosteering. J. Petrol. Sci. Eng. 183 (December), 106381.

Alyaev, S., Tveranger, J., Fossum, K., Elsheikh, A.H., 2021b. Probabilistic forecasting for geosteering in fluvial successions using a generative adversarial network. First Break 39 (7), 45–50.

Bond, C.E., Gibbs, A.D., Shipton, Z.K., Jones, S., 2007. What do you think this is "conceptual uncertainty" in geoscience interpretation. GSA Today (Geol. Soc. Am.) 17 (11), 4.

Bratvold, R., Begg, S., 2010. Making Good Decisions. Society of petroleum engineers.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W., 2016. Openai Gym. CoRR abs/1606.01540.

Chen, Y., Lorentzen, R.J., Vefring, E.H., apr 2015. Optimization of well trajectory under uncertainty for proactive geosteering. SPE J. 20 (2), 368–383.

Fonseca, R., Della Rossa, E., Emerick, A., Hanea, R., Jansen, J., 2018. Overview of the olympus field development optimization challenge. In: ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery, vol. 2018. European Association of Geoscientists & Engineers, pp. 1–10.

Fossum, K., Alyaev, S., Suter, E., Tosi, G., Mele, M., 2021. Reducing 3d uncertainty by an ensemble-based geosteering workflow: an example from the goliat field. In: Accepted to Third EAGE/SPE Geosteering Workshop.

Halset, G., Perazzi, F., Horstmann, M., Mele, M., 2020. Advanced geosteering technologies to maximize well design while minimizing costs and risks; marulk field (Norwegian sea). In: 82nd EAGE Annual Conference & Exhibition, vol. 2020. European Association of Geoscientists & Engineers, pp. 1–5.

Hermanrud, K., Antonsen, F., Teixeira De Oliveira, M.E., Petersen, S.A., Constable, M., 2019. Future Geosteering and Well Placement Solutions from an Operator Perspective. Society of Petroleum Engineers - SPE Annual Caspian Technical Conference 2019. CTC 2019.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York. URL. https://www.amazon.de/Thinking-Fast-Slow-Daniel-Kahneman/dp/0374275637/ref=wl_it_dp_o_pdT1_nS_nC?ie=UTF8&colid=151193SNGKJT9&coliid=I3OCESLZCVDFL7.

Kristoffersen, B.S., Bellout, M.C., Berg, C.F., 2020. Automatic Well Planner and Well Placement Optimization. Accepted in Computational Geosciences.
Kullawan, K., Bratvold, R., Bickel, J., aug, 2016. Value creation with multi-criteria decision making in geosteering operations. International Journal of Petroleum Technology 3 (1), 15–31.
Kullawan, K., Bratvold, R.B., Bickel, J.E., 2018. Sequential geosteering decisions for optimization of real-time well placement. J. Petrol. Sci. Eng. 165 (January), 90–104.
Larsen, D., Antonov, Y., Luxry, P., Skillings, J., Skaug, M., Wagner, V., 2016. Navigating the horizontal section in a heterogeneous formation while using Extra Deep Azimuthal Resistivity for optimizing the wellbore placement within a narrow TVD window. In: SPWLA 57th Annual Logging Symposium, June 25-29, 2016 Navigating, pp. 1–12. URL. https://www.onepetro.org/conference-paper/SPWLA-2016-IIII.
Luo, X., Eliasson, P., Alyaev, S., Romdhane, A., Suter, E., Querendez, E., Vefring, E., 2015. An ensemble-based framework for proactive geosteering. In: SPWLA 56th Annual Logging Symposium, July 18-22, 2015, pp. 1–14. URL. https://www.onepetro.org/conference-paper/SPWLA-2015-KKKK.
NORCE, 2019. formation evaluation and geosteering workshop 2019. NORCE. URL. https://geosteering.no/workshop2019/.
Peters, E., Chen, Y., Leeuwenburgh, O., Oliver, D.S., 2013. Extended Brugge benchmark case for history matching and water flooding optimization. Comput. Geosci. 50, 16–24.
Przybylo, P., 2019. Horizontal well geosteering guidelines. Blurb, incorporated. URL. https://books.google.no/books?id=DYHEwgEACAAJ.
Shen, Q., Wu, X., Chen, J., Han, Z., Huang, Y., 2018. Solving geosteering inverse problems by stochastic Hybrid Monte Carlo method. J. Petrol. Sci. Eng. 161 (August 2017), 9–16.
Veettil, D.R.A., Clark, K., 2020. Bayesian geosteering using sequential Monte Carlo methods. Petrophysics 61 (1), 99–111. URL. https://www.onepetro.org/journal-paper/SPWLA-2020-v61n1a4.
Viard, T., Caumon, G., Lévy, B., 2011. Adjacent versus coincident representations of geospatial uncertainty: which promote better decisions? Comput. Geosci. 37 (4), 511–520.
Welsh, M.B., Bratvold, R.B., Begg, S.H., 2005. Cognitive biases in the petroleum industry: impact and remediation. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, p. 10.
Wilson, C.G., Bond, C.E., Shipley, T.F., 2019. How can geologic decision-making under uncertainty be improved? Solid earth 10 (5), 1469–1488.