Solid Earth Discuss.,
https://doi.org/10.5194/se-2019-70-AC2, 2019

Solid Earth Discussions

Interactive comment on "How can geologic decision making under uncertainty be improved?" *by* Cristina G. Wilson et al.

Cristina G. Wilson et al.

cristina.wilson@temple.edu

Received and published: 1 August 2019

The second part of the manuscript is then focused on two case studies on the combination of "AI" with human interpretations in order to improve decision making. In this part, I have some problems in following the argumentation of the authors. I understand that any help with reducing cognitive overloading ("busy editor") can potentially help. But especially in the first case, I do not quite see how an automated sampling strategy can help here. For sure, an optimised sampling is interesting in itself - but how does this address the three forms of bias presented above, as opposed to a pure random sampling or the commonly used regular flight paths (option "C" in Fig. 4)? The only added benefit I see (maybe because this is a simplified example) is the reduced time of sampling. Even

more: couldn't one also argue that any form of "AI" is prone to introducing additional bias, as it is based on an underlying algorithm that may also be biased? Also, you argue in line 13 (pg. 19) that the expert (user) should retain the ability to interact and adjust the flight path - but wouldn't this then again be prone to the biases described before?

We have improved the connection between the case study section and the preceding bias review section by explicitly detailing how automated flights can address susceptibility to anchoring bias in field decisions about where to fly. The title of case study 1 was changed to: *"Optimizing field data collection with UAVs to minimize anchoring bias".* We have also removed many extra details about UAVs not directly pertinent to the case study. Most of the changes were to the first three paragraphs:

"In this case study, we describe how automated UAV navigation could be used to nudge geoscientists to be more efficient when making decisions regarding reconnaissance and mapping and mitigate against anchoring bias. The advent of better mobile robot platforms has allowed for the deployment of robots by ground, sea, and air to collect field data at a high spatial and temporal resolution. Here, we focus on the use of aerial robots (semi-autonomous or autonomous UAVs) for data collection, but the conclusions we draw are likely applicable to other mobile robot platforms (i.e., underwater autonomous vehicles, ground robots).

Currently, the majority of geoscience research with UAVs is non-autonomous, i.e., user-controlled. Efforts have been made to automate interpretation of geological data from UAV imagery or 3D reconstruction with some success (Thiele et al., 2017; Vasuki, Holden, Kovesi, Micklethwaite, 2014; Vasuki, Holden, Kovesi, Micklethwaite, 2017), and the application of image analysis and machine learning techniques continue to be developed (Zhang, Wang, Li, Han, 2018). In reconnaissance and geologic mapping, the decision of where to go and how to fly there is made by the expert – either the expert fly's the UAV and makes navigation decisions in-situ or they pre-set a flight path for the UAV to follow semi-autonomously (cf. Koparan et al., 2018; Ore, Elbaum, Burgin,

Detweiler, 2015). However, a UAV that is capable of attending to measurements in real time and reacting to local features of measurement data could navigate autonomously to collect observations where they are most needed. Such autonomous workflows should increase the efficiency of data collection, and could be designed to mitigate against potential biases. Here, we consider how an automated UAV navigation nudge could reduce the tendency to anchor field exploration based on existing models and hypotheses.

In our hypothetical example, a UAV surveys a large bedding surface with the aim of identifying fracture orientations. The bedding surface exposure is large, but split into difficult to access exposure, e.g., due to cliff-sections or vegetation (see Column A, Figure 4). A birds-eye view afforded by the UAV improves the ability to observe fractures, which would otherwise require time-costly on-foot reconnaissance to different outcrops of the bedding surface. Note that in our hypothetical example we assume that fracture information is obtained only when the flight path crosses fractures (e.g., Column B, blue flight path), thereby representing a high level reconnaissance rather than a flight path in which overlapping imagery is collected. When the UAV flight path is user-controlled, the decision of where and how to fly is unlikely to be optimal: users could be distracted by irrelevant information in UAV view, and are likely biased towards exploring certain features and ignoring others (see Andrews et al., 2019). For example fractures may only be sampled where fracture data is dense, or in an orientation that maximizes sample size but not the range in orientation (see Watkins, Bond, Healy, Butler, 2015), or when it fits with a hypothesis (e.g. tensional fractures parallel to the axial trace of a fold). These strategies are all informed by expectations, leaving the geoscientist vulnerable to anchoring her sampling behavior to align with initial interpretations and hypotheses. This anchoring bias is visualized in Column B (blue flight path), where the user detects two unique fracture orientations (a, b) on the first exposure visited, but then spends needless time (T1 to T2) at exposure that offers no new information, before finally visiting exposure that features the previously identified orientations (a, b) and a novel N-S fracture orientation. This novel orientation is not detected in the

user's flight path – the accompanying certainty plot in Column B shows that time spent at uninformative exposure (T1 to T2) results in increased certainty that all orientations have been sampled, when in fact they have not (i.e., the threshold of confidence is reached before sampling the N-S orientation). This is reflected in the rose diagrams in Column B, which show the orientation of fractures and the relative number of fractures sampled in each orientation; even at time T3 the three fracture sets (as shown in the rose diagram in Column A) are not represented."

Regarding AI introducing additional bias – Yes, this is possible, dangers and necessary precautions (i.e., explainability) are discussed in the conclusion.

Regarding the user retaining autonomy to make biased decisions – this is a classic principle of the choice architecture approach, i.e., freedom of choice must never be encroached upon. As we state in section 4.2, *"It is the role of the choice architect…to influence people's decision making such that their well-being (and the well-being of others) is maximized, without restricting the freedom to choose. Importantly, there is no such thing as neutral choice architecture; the way the environment is setup will guide decision making, regardless of whether the setup was intentional on the part of the architect, e.g., descriptions of risk will be framed in terms of gains or losses, a wise architect chooses the framing that will maximize well-being."*

The aspect of fault interpretations in seismic data, explained in case study 2, is more obvious to me - although here the question could also be how much bias is in the initial choice of a fault displacement model (which can be based on physical principles, but the potential interactions can also quickly become very complex when considering fault networks, relay structures, etc.). But here, the point of flagging potential areas of problems is an interesting aspect of "digital nudging" (if I understand it correctly), and similar to the example from Polson and Curtis (2010) and the "bias warning" point in the expert decision-making process.

We have improved the connection between the case study section and the preceding bias review section by explicitly detailing how seismic interpretation aids (built into software) can address susceptibility to availability bias during interpretation. The title of case study 2 was changed to: *"Fault interpretations in 3D seismic image data to minimize availability bias".* We also removed details about the technique of automated horizon tracking (including the accompanying Figure 5) which are not directly pertinent to the case study. Most of the changes were to the first three paragraphs:

"In this case study, we consider how software interpretations of seismic image data, and the information derived from them, could be used to nudge geoscientists to consider alternative models and minimize availability bias. Understanding of the geometries of sub-surface geology is dominated by interpretations of seismic image data, and these interpretations serve a critical role in important tasks like resource exploration and geohazard assessment. 3D seismic image volumes are analyzed as sequences of 2D slices. Manual interpretation involves visually analyzing a 2D image, identifying important patterns (e.g., faulted horizons, salt domes, gas chimneys) and labeling those patterns with distinct marks or colors; then, keeping this information in mind while generating expectations about the contents of the next 2D image. Given the magnitude and complexity of this task, there has been a strong and continued interest in developing semi-autonomous and autonomous digital tools to make seismic interpretation more efficient and accurate (e.g., Araya-Polo et al., 2017; Di, 2018; Farrokhnia, Kahoo, Soleimani, 2018).

Here, we consider how 3D information could be used with digital nudge technology to inform fault interpretations in a 3D seismic image volume. Simple normal fault patterns show a bull's-eye pattern of greatest displacement in the center of an isolated fault, decreasing towards the fault-tip (see Image A, Figure 6). Consider interpreting 2D seismic image lines across the fault starting at in-line A (Image A) and working towards in-line F: with each subsequent line the displacement of horizons across the fault should increase and then decrease, although this pattern will not be known until

the interpretation is completed. Holding this information on displacements for individual faults between in-line interpretations in complicated seismic image data (e.g. with multiple faults per seismic section, Image B, Figure 6) is incredibly challenging even for the well-practiced expert. We imagine a digital nudge that alerts users to discrepancies in fault displacement patterns, and prompts consideration of alternative fault patterns, thereby relieving some of the cognitive burden of 3D interpretation from the expert and guarding them against availability bias by encouraging consideration of models beyond what is most readily accessible to the mind.

In our hypothetical example, a geoscientist analyzes a 3D seismic volume, interpreting in a series of 2D in-line images faults and horizon off-sets. As subsequent in-lines (A-F) are interpreted, fault displacement patterns are co-visualized, so inconsistencies from normal fault displacement can be clearly seen. Fault 1 (Image B) conforms to a simple fault-displacement pattern (see Fault 1 displacement-distance plot). Fault 2 appears to conform to a similar pattern until in-line D when the interpreted displacement decreases; on interpretation of in-line E, the displacement on Fault 2 increases again, further highlighting the displacement anomaly on in-line D. Reduced displacement in itself does not highlight an issue, but consideration of the displacement-distance plot for Fault 1 suggests that if the interpreted displacement for Fault 2 is correct then the two faults are behaving differently. In our imagined digital tool, this discrepancy in displacement between nearby faults would be flagged for further consideration by the user, and potential alternative models could be highlighted. You can see the hypothetical conclusion certainty plots for the interpreter for the two faults (Fault 1 = green line, Fault 2 = pale blue line) during the interpretation process. Note the decrease in certainty of the interpreter for Fault 2, as they interpret in-lines D and E, in comparison to the increasing certainty for Fault 1 as consecutive interpreted in-lines conform to a simple normal fault displacement pattern. At in-line E the co-visualized displacement-distance plot nudges the interpreter to consider a new interpretation for Fault 2 at in-line D. Certainty in this new interpretation (displayed as dark blue dashed line on certainty plot), now increases as subsequent in-line interpretations conform to expected displacements."

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-70, 2019.