# Supporting Non-Experts' Awareness of Uncertainty: Negative Effects of Simple Visualizations in Multiple Views

Emma Beauxis-Aussalet
emalb@cwi.nl

Elvira Arslanova

Lynda Hardman
lynda.hardman@cwi.nl

CWI
Science Park 123
1098XG Amsterdam, The Netherlands

## ABSTRACT

Video analysis tools can provide valuable datasets for a wide range of applications, such as monitoring animal populations for ecology research, while reducing human efforts for collecting information. Transferring such technology to novel application domains implies exposing non-expert users to unfamiliar datasets and technical concepts. Existing data analysis practices must adapt to the new data characteristics and technical constraints. With such changes, uncertainty is of major concern as it can yield misinterpretation of data, or distrust and rejection of valid results. We present a study of an interactive visualization of computer vision results and uncertainty. We evaluate the correctness of users' interpretation of data, and their confidence in their interpretation. We compare the impact of either data features (i.e., the true level of uncertainty) or visualization features on user perception of uncertainty. Visualization features had a similar impact on user responses than the data uncertainty itself, thus biasing user awareness of uncertainty. We conclude with the opportunities (intuitive navigation in complex unfamiliar data) and limitations (poor extrapolation and memory loss) of our visualization design which integrates simple graphs in coordinated multiple views. Our design and insights contribute to other cases where non-experts need to familiarize with novel datasets and explore their uncertainty.

## Categories and Subject Descriptors

I.3.6 [**Computer Graphics**]: Methodology and Techniques—*Interaction techniques*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Video*

## 1. INTRODUCTION

The Fish4Knowledge project collected data from 9 underwater cameras which continuously recorded coral reef fish every day during 3 years. The collection, too large for manual analysis, was processed by computer vision tools. They extracted fish occurrences from each species, which is a key asset for studying population dynamics, food chains, reproduction and migration [3]. The video processing consists of 3 steps: storing video streams as 10-minute clips, detecting fish against background objects and recognizing their species. Each step introduces potential errors that may impact data usage for scientific research. Previous user studies highlighted the needs for understandable evaluation of video analysis errors [3], and for handling the fragmentary processing of video streams [4] (e.g., the fewer the videos, the fewer the fish, and the more sensitivity to outliers). We developed a visualization interface for exploring the video data and its uncertainties. We harness simplicity in our design so as not to overwhelm users who are dealing with unfamiliar, complex data. We investigate how our design supports information seeking tasks related to fish abundance (e.g., What is the size of fish populations?), sampling size (e.g., What is the number of processed videos?), and reliability of species recognition (e.g., Does the system over- or under-estimate the population size for *species X*?). Particular attention is drawn to differentiating interaction design and layout design (e.g., confusing interactions may be implemented on a clear layout), and to supporting data complexity (e.g., a design may be suitable for common trivial information, but not for unfamiliar technical information). We set up an evaluation framework inspired from the situation awareness domain, for its treatment of uncertainty and information processing issues. We recruited ecologists studying the ecosystem monitored by our cameras. We analyzed the usability issues they encountered, and the factors impacting the perception of uncertainty. We conclude by discussing the strength and weaknesses of our design, and the directions for future work.


*ECCE 2015, July 01 - 03, 2015, Warsaw , Poland*

DOI: http://dx.doi.org/10.1145/2788412.2788432

## 2. RELATED WORK

**Visualizing Uncertain Data** - Visualization encodes information into the visual dimensions offered by human perception: color, size, value (e.g., transparency), texture, orientation and shape (e.g., glyph) [6]. Multiple views can be used for complex multidimensional visualization problems. [22] (discussing opportunities and risks of multiple views) and [2, 18] (describing the types of analytic tasks) highlight that locating pieces of information and characterizing their relationships is a core cognitive task. Techniques for uncertainty visualization represent uncertainty as extra dimensions of canonical graphical representations [12, 17].

| *Uncertainty Factor* | *Description* | *Stage* |
|---|---|---|
| **Uncertainty due to computer vision algorithms** | | |
| Ground-Truth Quality | Fish examples may be scarce, erroneous, or show odd fish appearances. | 1 |
| Fish Detection Error | Some fish may be undetected, and non-fish objects may be detected as fish. | 1 |
| Species Recognition Error | Some species may not be recognized, or confused with another. | 1 |
| Image Quality | Degraded images (e.g., lighting, fuzziness) may impact computer vision errors. | 1 |
| **Uncertainty due to in-situ system deployment** | | |
| Field of View | Observed ecosystems may be heterogeneous, and over- or under-represent species. | 3 |
| Fragmentary Processing | Some videos may be unprocessed, missing, or unusable (e.g., encoding errors). | 1 |
| Duplicated Individuals | Fish swimming back and forth are repeatedly recorded. Rates of duplication vary among species behaviour (e.g., sheltering in coral head). | 3 |
| Sampling Coverage | The numbers of videos may not provide statistically representative datasets. | 1 |
| **Uncertainty due to both computer vision algorithms and deployment conditions** | | |
| Biases Emerging from Noise | Errors may be random (noise) or systematic (bias). Biases may emerge from a combination of factors (*Image Quality, Field of View, Duplicated Individuals, Computer Vision Errors*). | 3 |
| Uncertainty in Specific Output | Uncertainty in specific outputs may be extrapolated from errors measured in test conditions, compared to the output characteristics (e.g., fewer low quality images yield fewer biases). | 2 |

**Table 1: Uncertainty factors arising at each information processing steps, identified along the design stages.**

Uncertainty is itself multidimensional: it is of various types and can be introduced at any information processing step [7, 20]. A major challenge is to deal with the complexity of uncertainty (e.g., propagation through information processing steps) and visualization (e.g., cluttered display), especially for non-experts. [11] highlights three main issues for non-expert users: *translating questions into data attributes* (e.g., selecting datasets of interest), *constructing visualization* (e.g., mapping data into visual templates) and *interpreting the visualizations*. The core strain is *converting concepts* of different natures: those of users' mental model, of data attributes, or of visual features. [8] reports that non-experts requested a reminder of data filters to be always displayed, and propagated to all views of dashboards, echoing novices' need for *contextual information* explaining the visualized data [13]. Constraining dashboards to display views of the same dataset (propagating the filters) limits the expressibility of multiple views, but increases usability as it helps manipulating *data attributes*. Limiting expressibility (e.g., each multiple view must display aspects of the same species population) in favor of usability is reasonable for an audience of non-expert users [19].

**Situation Awareness** - Endsley distinguishes 3 levels in situation awareness related to information processing [9], which echo visual analytical tasks (locate and associate information [2, 18, 22]): **Perception** of cues, occurring when information is simply read, without further interpretation or correlation; **Comprehension**, i.e., integrating multiple pieces of information ; and **Projection**, i.e., forecasting elements of an unknown situation (e.g., future events, interpretations of uncertain data). Misinterpretations are frequently caused by data overload, complex systems that are poorly understood, and users' limited working memory [23]. Memory-loss (forgetting already perceived information) is yielded by delays between receiving information and using it, e.g., due to intermediate interaction with the system, or confusing layout. Working memory is crucial in novel systems such as ours, as users are not familiar with the computer vision techniques and must combine, interpret and apply unusual information within the limits of their working memory. Methods for evaluating Situation Awareness rely on the usage of *probes* (i.e., predefined states of the system in which users are emerged), and consider the types and the levels of uncertainty in users' knowledge of a situation [14, 15]. Probes were used to expose users to specific interface layouts, prior to letting users interact with the system. It allowed to distinctively evaluate our layout and interaction designs.

## 3. INTERFACE DESIGN

**Design Rationale** - Prior studies identified two core user information needs: i) counting fish from each species, and for specific time periods and locations; and ii) estimating the uncertainty of fish counts [3, 4]. Information related to uncertainty concerns multiple factors (Table 1). This study focuses on two factors: fragmentary processing (i.e., missing videos) impacting the sample size, and species recognition errors. Our design decisions targeted three key challenges. **C1:** Users deal with unfamiliar data, they need to understand the concepts of computer vision techniques, and their implications for their data analysis tasks. It demands important, sometimes overwhelming cognitive efforts [3]. **C2:** Users deal with uncertainty occurring at different information processing steps (Table 1), yielding multidimensional data. The resulting complexity is a major challenge. **C3:** Users may have a variety of research goals which can not all be addressed with a one-fits-all visualization. Specific data analysis methods may be necessary to complete their tasks. Our information layout and interactions are designed for reducing complexity *[C1, C2]*. We use simple graphs and handle multidimensionality with multiple views *[C2]*. We target basic data analysis tasks (*Retrieve, Filter, Determine Range, Correlate Information* [2]), and exclude advanced data mining and statistical representations *[C1,C3]*. We expose uncertainty factors at each information processing step *[C2]*. [22] rule of *Diversity* (separate kinds of information) inspired the organization of information into 5 tabs. The tabs reflect the sequence of information processing steps: data collection (*Video* tab), data processing (*Video Analysis, Raw Data* tabs), and data interpretation (*Visualization, Report* tabs). They provide contextual information about the video data, as recommended for non-expert users [13]. This study focuses only on the *Visualization* tab, which is described in Fig. 1. A description of the remaining tabs can be found in [5]. Our interaction design with on-demand widget display follows the rule of *Decomposition* (create manageable chunks). The rule of *Complementarity* (expose relationships) led to designing i) linkage between views (e.g., main graph and widgets sharing the same Y axis metric[1] and data filters), and ii) interactions for accessing visualization variants (e.g., the same visualization for varying datasets, by changing the filters; or varying visualizations of the same dataset, by changing the main graph in Zone A or opening widgets). Users navigate through visualization variants by swapping graph axes and types of graph. This *"swapping"* interaction mode genuinely synthesizes features from ManyEyes [21] (swap axes) and Tableau [1] (swap graph templates). It offers a large scope of possible data associations and comparisons, while limiting cluttered display and information overload. This is desired in a context where ecologists pursue a variety of research goals, while being unfamiliar with video data.

**Usage Scenario** - This section describes the interactions we seek to evaluate, i.e., that supporting information seeking tasks related to fish abundance, sampling size (i.e., the number of video samples), and reliability of species recognition. When visualizing the fish counts in Fig. 1-a, users can wonder if abundance drops in weeks 35 and 45 are due to missing videos. Using the *Y Axis* menu in *Zone B*, they can display the numbers of videos from which fish counts were extracted (Fig. 1-b). As no videos were processed for Week 45, no insight can be drawn on fish abundance at this period. Considering the high variability of video numbers, visualizing average numbers of fish per video is preferable to visualizing absolute fish counts. The *Y Axis* menu provides this visualization (Fig. 1-c). Considering that the abundance pattern is not due to varying numbers of videos, users should question the reliability of species recognition. The widget *Certainty Score* shows the quality of fish appearances (Fig. 3). The more fish with high certainty scores, the more reliable the species recognition. Users can use the certainty scores to estimate potential biases due to species recognition errors. E.g., Fig. 5-7 compares recognition reliability for species 1 and 2. Higher certainty scores are observed for species 2, thus its recognition is likely to be more reliable than for species 1. Similarly, users can compare the certainty scores for week 35 with other weeks. For comparing relative fish abundance (i.e., species composition), users can use the *Species* widget or select *Stacked chart* in the *Chart Type* menu. For studying specific species, the *Species* widget supports the exclusion of irrelevant species (Fig. 5-7).

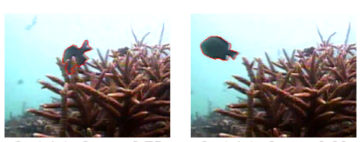

Certainty Score: 0.75 Certainty Score: 0.89

**Figure 3: Scores reflecting fish appearance quality.**

# 4. EVALUATION

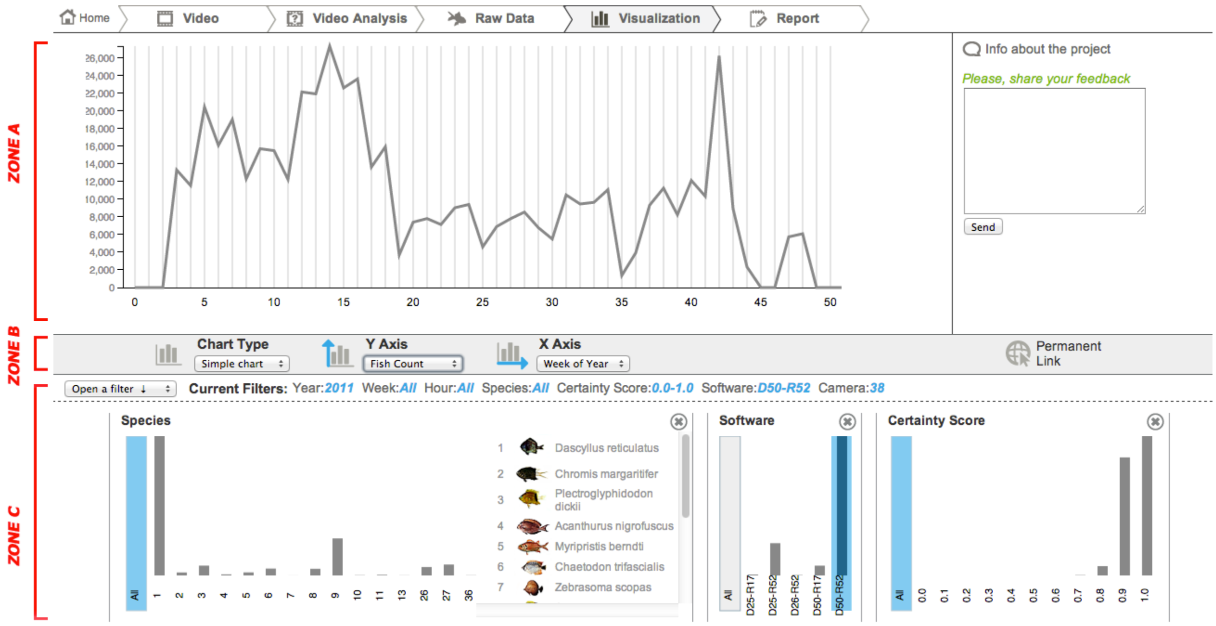
**Experimental Setup** - We recruited 10 users from the marine ecology community in Taiwan. A 20 minute tutorial introduced the interface and the concept of certainty score. Users learned the interactions needed to perform the usage scenario described in Section 3: i) display visualization variants with fish counts, video counts, or average number of fish per video; ii) display variants using simple chart or stacked chart; iii) use filter widgets to select datasets of interest; and iv) use filter widgets to compare fish distributions. They also learned how to watch videos in the *Video* tab, since users recurrently request to check the footage. We then conducted an experiment inspired by situation awareness methods. We exposed users to 3 *probes* (predefined states of the interface with preselected filters and graph options), showing real data extracted by our computer vision system, and asked a total of 20 questions. Users indicated their confidence in their answers using a 5-grade scale (*Very Low, Low, Moderate, High* or *Very High* confidence). The questions dealt with various complexity of *Fact* assessment (levels F1 to F3), *Uncertainty* evaluation (levels U1 to U3), and *Interaction* with the interface (levels I1 to I2). The levels are specified in Table 2 and their related questions in Table 3. Levels F1-3 refer to levels of situation awareness postulated by Endsley [10]. Levels U1-3 and I1-2 were created for our use case. Dealing with uncertainty (U2-3) implies dealing with complex facts (F3), as assessing uncertainty requires extrapolation. Thus all questions from levels U2 or U3 are also from level F3, and task complexity is synthesised in 4 levels F1, F2, U2, or U3. The questions were designed to draw attention on the uncertainty factors: *species recognition errors* and *fragmentary processing*. The latter was particularly emphasized: questions Q5 and Q11-13 ask for video counts before questions asking about fish abundance. Prior to Q5 users deal with absolute fish count and later with average fish count per 10-minute video samples, hence showing the effect fragmentary processing. Q13 explicitly examines the suitability of sampling size for scientific research. Guiding users' attention may artificially enhance their reactivity to uncertainty. This was desired both a priori, as fragmentary processing is an unfamiliar concept, and a posteriori, considering the poor user reactivity to this uncertainty factor. Usability issues and wrong answers were reported. Under uncertainty (levels U2-3), answers such as *"I don't know"* were considered as one of the possible valid answers. Our setup allowed to observe: i) how users interact with the visualization when seeking information (e.g., using widget overviews or main graph); ii) which usability issues arise with layout or interactions (e.g., before or after modifying a predefined state of the interface); iii) how user confidence varies among the levels of information complexity (levels F1-2 and U2-3); and iv) how interaction complexity impacts user answers (levels I1-2). Our small user group and the number of observations collected for each condition (levels F1-F3, U1-U3, I1-I2) may not represent the general population of marine ecologists. However, they are suitable for identifying major usability issues, and the means to support user awareness of uncertainty.

| | |
|---|---|
| **F1** Perception | Read one single information |
| **F2** Comprehension | Compare several information |
| **F3** Projection | Extrapolate from given information |
| **U1** Acknowledment | Only one answer is entirely true. |
| **U2** Ambivalence | Several answers are valid. Enough information can inform user choice. |
| **U3** Assumption | Several answers are valid. Limited information can inform user choice. |
| **I1** No Interaction | No UI manipulation is needed |
| **I2** Exploration | UI manipulations are needed |

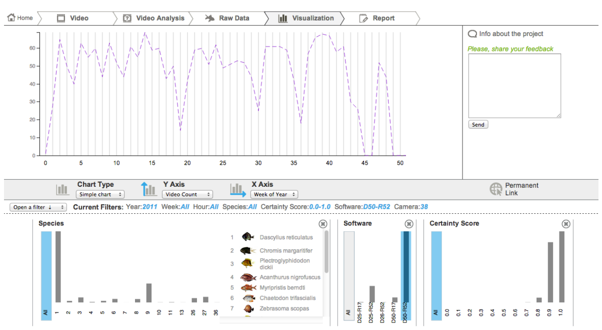
**Table 2: Levels of complexity exposed to users.**

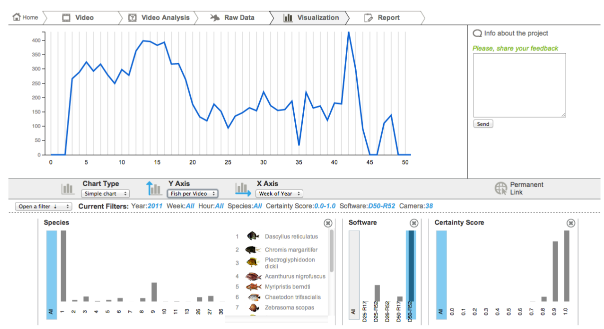
**Experiment Results** - The results are reported in Table 4 and summarized in Fig. 8. Question Q13 was discarded since answer correctness is ambiguous: the most precise answer is *"It depends on research goals"* as replied by one single

[1]We elicited this feature after conducting our experiment




a) The Visualization tab, as shown for the first probe of the experiment, and needed for answering questions 1-4.





b) Visualization for answering questions 5 and 7.

c) Visualization for answering questions 6 and 7.

The *Visualization* tab lays out information in 3 zones (Fig. 1-a). **Zone A** contains the main graph, which can be adapted to fit user interests using options in **Zone B**. Users can specify what is represented by the axes of the main graph. E.g., while the Y axis represents fish counts, the X axis can be swapped for weeks of the year, hours of the day or locations, to show various fish distributions. Similarly, the Y axis can be swapped to video counts or average number of fish per video while keeping the same X axis (Fig. 1-a and b). Swapping axes is the originality of this design. Users can gradually navigate through the dimensions of the data. Users can also select other types of graph (e.g., stacked chart or boxplot) while keeping the settings of the X and Y axes. E.g., fish counts can be stacked by species for studying species composition. Dedicated menus are displayed for adapting further the visualizations, e.g., to stack fish counts by species or locations. Swapping both axes and types of graph helps manipulating *data attributes* and *visual representations* which are major strains for non-experts [11]. **Zone C** contains *filter widgets* for i) selecting datasets of interest, and ii) overviewing datasets over several dimensions. There are widgets for each dimension of the data, namely: Year, Week of Year, Hour of Day of fish occurrence, Camera (i.e., location), Species, Certainty Score, Software Version. Filter widgets are displayed on-demand. The selected filter values are highlighted in blue and summarized just below *Zone B*. The filter summary is especially important for non-experts [8]. The widget histograms display fish counts, hence showing fish distribution over several dimensions. The filters selected for the main graph are applied to the dataset shown in widget histograms. E.g., in Fig. 1-a, the *Zone A* graph and the *Zone C* histograms all show a dataset of fish detected by software *D50-R52*, occurring during 2011 at Camera 38, and belonging to all species, certainty scores, weeks of year and hours of day. In this example, the *Software* widget shows a dataset from all software, keeping all other applicable filters. After conducting our experiment, we concluded that widget histograms must display the same metric as the main graph (i.e., same Y axis), to ensure high consistency between views as recommended by [8, 22].

**Figure 1: Visualizations for exploring the impact of fragmentary processing.**

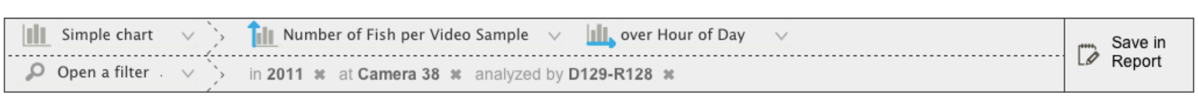



**Figure 2: Improved version of *Zone B*.**

| | *Question* | *Complexity* |
|---|---|---|
| **Probe 1** (Fig. 1) | | |
| 1 | What is the number of fish for the week 12? | F1 U1 I1 |
| 2 | For which cameras are we counting the fish? | F1 U1 I1 |
| 3 | Which week of the year has the most fish? | F2 U1 I1 |
| 4 | At which period of the year can we observe the highest fish count? | F3 U2 I2 |
| 5 | How many videos were analyzed for the week 12? | F1 U1 I2 |
| 6 | What is the number of fish per video for the week 12? | F1 U1 I2 |
| 7 | What is the fish abundance for the week 45? | F3 U2 I2 |
| 8 | Which week of the year has the highest number of fish per video? | F2 U1 I1 |
| 9 | What is the period of the year for which the fish population is the most abundant? | F3 U2 I1 |
| 10 | Is it the same period of time for the camera 37? | F3 U3 I2 |
| **Probe 2** (Fig. 4) | | |
| 11 | Is the number of video samples constant over hours of the day? | F2 U1 I1 |
| 12 | Is the number of video samples constant over weeks of the year? | F2 U1 I2 |
| 13 | Is the amount of video samples suitable for scientific research? | F3 U3 I1 |
| 14 | Which is the most abundant species in HoBiHu? | F2 U1 I1 |
| 15 | Which camera has the most abundant fish population from the species 2 (Chromis Margaritifer)? | F3 U3 I2 |
| 16 | Do fish from species Chromis Margaritifer generally have high certainty scores? | F2 U1 I2 |
| 17 | Is the abundance of species 2 (Chromis Margaritifer) lower than species 1 (Dascillus Reticulatus)? | F2 U1 I1 |
| 18 | Is it because the video analysis may not correctly detect the species 2 (Chromis Margaritifer) | F3 U3 I2 |
| **Probe 3** (Fig. 6) | | |
| 19 | Is there a correlation in the occurrence of fish from species 9, 26 and 27 over weeks of the year? (considering the entire dataset, for all time periods and all cameras) | F3 U3 I2 |
| 20 | Is there a correlation in the occurrence of fish from species 9, 26 and 27 over hours of the day? | F3 U3 I2 |

**Table 3: The 20 questions asked to participants, and their levels of complexity.**

| | | User 1 | | | User 2 | | | User 3 | | | User 4 | | | User 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Level | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. |
| Q1 | F1 I1 | | 5 | | | 3 | | | 5 | | | 5 | | | 5 | |
| Q2 | F1 I1 | | 5 | | | 5 | | | 5 | | | 5 | | | 5 | |
| Q3 | F2 I1 | | 5 | | | 5 | X | | 5 | | | 5 | | | 5 | |
| Q4 | U2 I1 | | 3 | | | 4 | | | 5 | | | 5 | | | 4 | |
| Q5 | F1 I2 | | 4 | | | 4 | | | 5 | | | 5 | | W | 5 | |
| Q6 | F1 I2 | | 4 | | | 4 | X | | 4 | | | 4 | | | 5 | |
| Q7 | U2 I2 | W | 5 | | W | 5 | | W | 4 | | W | 5 | | W | 5 | |
| Q8 | F2 I1 | | 5 | | | 5 | | | 5 | | | 5 | | | 5 | |
| Q9 | U2 I1 | | 3 | | | 4 | | | 5 | | | 5 | | | 5 | |
| Q10 | U3 I1 | | 4 | | W | 4 | X | W | 5 | | W | 4 | X | W | 5 | X |
| Q11 | F2 I1 | | 5 | | W | 5 | | W | 3 | | | 5 | | | 5 | |
| Q12 | F2 I2 | | 4 | | | 5 | | W | 4 | | | 5 | | | 5 | |
| Q14 | F2 I1 | | 5 | | | 5 | | | 5 | | | 5 | | | 5 | |
| Q15 | U3 I2 | | 4 | | | 4 | | | 5 | | | 4 | | | 3 | |
| Q16 | F2 I2 | | 4 | | | 3 | | | 4 | | | 5 | | | 5 | |
| Q17 | F2 I1 | | 4 | | | 4 | | | 5 | | | 5 | | W | 5 | |
| Q18 | U3 I2 | W | 1 | | W | 3 | | W | 4 | | | 5 | | W | 2 | |
| Q19 | U3 I2 | | 4 | | | 4 | | | 4 | | | 5 | | | 4 | |
| Q20 | U3 I2 | | 4 | | | 3 | | | 4 | | | 5 | | | 5 | |

| | | User 6 | | | User 7 | | | User 8 | | | User 9 | | | User 10 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Level | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. | Err. | Conf. | Usa. |
| Q1 | F1 I1 | | 5 | | | 5 | | | 5 | | | 4 | | | 5 | |
| Q2 | F1 I1 | | 5 | | | 5 | | | 4 | | | 4 | | | 5 | |
| Q3 | F2 I1 | | 5 | | | 5 | | | 5 | | | 4 | | | 5 | |
| Q4 | U2 I1 | | 3 | | | 4 | | W | 5 | | | 3 | | | 4 | |
| Q5 | F1 I2 | W | 4 | | W | 3 | | W | 2 | | | 4 | | | 3 | |
| Q6 | F1 I2 | | 4 | | | 4 | | | 1 | X | | 4 | | | 5 | |
| Q7 | U2 I2 | W | 3 | X | W | 5 | | W | 3 | X | W | 2 | | W | 5 | |
| Q8 | F2 I1 | | 5 | | | 5 | | W | 5 | X | | 4 | | | 5 | X |
| Q9 | U2 I1 | | 4 | | | 5 | | | 4 | X | | 4 | | | 4 | |
| Q10 | U3 I1 | | 5 | | | 5 | | | 4 | | W | 4 | X | W | 5 | X |
| Q11 | F2 I1 | W | 4 | | | 5 | | | 4 | | | 4 | | | 4 | |
| Q12 | F2 I2 | | 5 | | | 5 | | W | 3 | X | | 4 | | | 4 | |
| Q14 | F2 I1 | | 5 | | | 5 | | | 5 | | | 4 | | | 5 | |
| Q15 | U3 I2 | | 4 | | | 2 | | W | 4 | X | | 4 | | | 3 | |
| Q16 | F2 I2 | | 5 | | | 5 | | | 5 | | | 4 | | | 5 | |
| Q17 | F2 I1 | W | 5 | | | 5 | | W | 3 | X | | 3 | | | 5 | |
| Q18 | U3 I2 | | 5 | | | 4 | | | 4 | | | 3 | | W | 2 | |
| Q19 | U3 I2 | | 4 | | | 5 | | | 3 | | | 3 | | | 5 | |
| Q20 | U3 I2 | | 5 | | | 5 | | | 4 | | W | 2 | X | W | 2 | X |

**Table 4: User responses, incl. invalid answers (Err.), user confidence (Conf.) and usability issues (Usa.).**

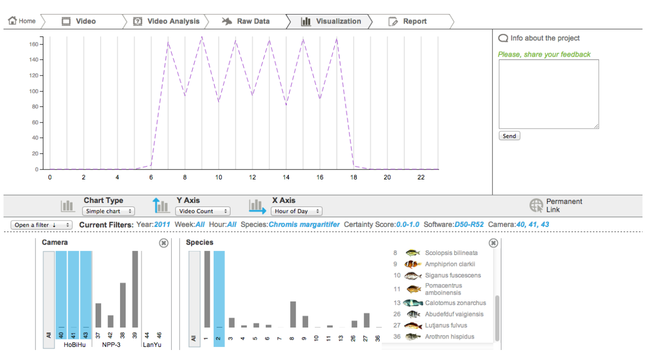


**Figure 4: The second probe of the experiment.**

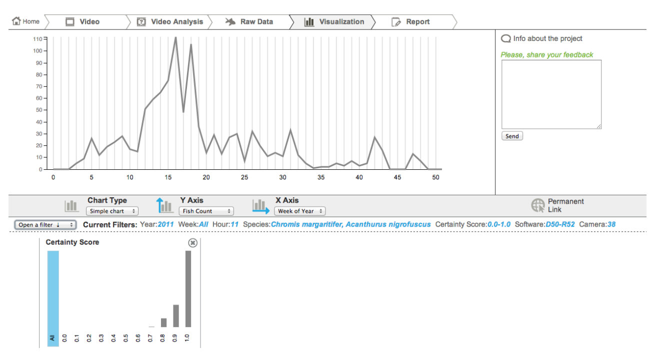


**Figure 6: The third probe of the experiment.**

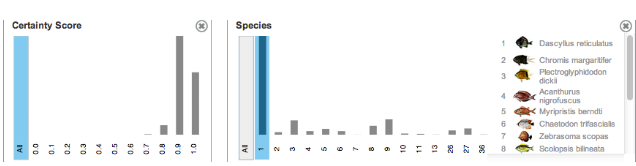



**Figure 5: Visualizations for answering question 18.**

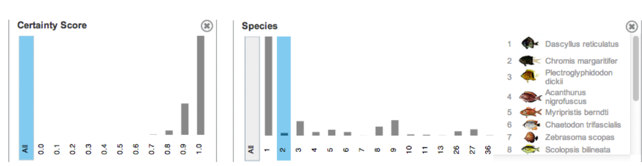



**Figure 7: Visualizations for answering question 18.**

user. We partitioned questions into groups representing task complexity (F1, F2, U2, or U3), interaction complexity (I1 or I2), answers' validity (Right or Wrong), and usability (Issue or No issue). These groups contain distinct questions and separate conditions involving uncertainty (U2-3) or not (F1-2). Fig. 9 shows the confidence values per user and group of questions (mean and 95%CI for each group).

Users were generally highly confident, even when answers were wrong or uncertainty was high. Level 5 is often the default answer, but some users consider level 4 as the default, making comparisons difficult, e.g., for *User 4*, level 4 is weak confidence while being the optimal confidence for *User 9*. To compare user confidence over question groups, we focused on *confidence drifts* (i.e., relative changes in confidence) rather than absolute confidence levels. Confidence drifts were computed for each user, e.g., 1) average a user's confidence for groups F1 and F2 distinctively; 2) subtract the averages to get the user's confidence drift; 3) repeat 1-2 for all users; 4) analyze confidence drifts over all users (e.g., median and inter-quartile range in Fig. 10). We analyzed confidence drifts between question groups *F1-F2, F2-U2, U2-U3, I1-I2, R-W* (Right or Wrong answers), *NoIssue-Us.Issue* (occurrence of usability issues). As we focus on the effect of either uncertainty itself or the user interface, we considered three additional question groups: *Certain-Uncertain* (F1∪F2 against U2∪U3), *CertainI1-CertainI2* (Certain∩I1 against Certain∩I2) and *UncertainI1-UncertainI2* (Uncertain∩I1 against Uncertain∩ I2). Fig. 10 shows the corresponding confidence drifts.

Except for the groups *F1-F2*, increasing question complexity yields a decrease in user confidence. It happens whether complexity arises from the information (*F2-U2, Certain-Uncertain, Right-Wrong*) or from the interface (*I1-I2, Us. Issue-NoIssue, CertainI1-CertainI2, UncertainI1-Uncertain I2*). However, the statistical significance of confidence drifts is not verified. Using Welch t-test (compensating for the unequal variance shown in Fig. 9), we tested the confidence drifts of each user. E.g., for each user, aggregate confidence measured for group F1 and F2 distinctively, and apply Welch test. P-values are generally $>.05$ (Fig. 11). User confidence is generally high with mostly the same value (e.g., level 5), thus in general confidence levels are not significantly different over question groups. However, two observations give credence in concluding that uncertainty and interactivity had similar effects on user confidence. **O1**: except for groups *F1-F2*, confidence was lowered by questions' complexity. If the effect was random, confidence drifts would show as many increases than decreases. **O2**: confidence drifts are the most significant for the groups *I1-I2*, with the lowest variance and p-values, and a median similar to that of the groups *Certain-Uncertain*. We noticed that wrong answers and usability issues had an important effect on user confidence (Fig. 9), but are outlying conditions (low numbers of observations). Further, wrong answers and usability issues often occurred together (Fig. 8). We repeated the analysis on right answers with no usability issue, and obtained similar observations (O1-2). We conclude that either interacting with the visualization, or analysing uncertain data had similar effects on user perception of uncertainty. This biases user awareness of uncertainty. Low user confidence may not assess the strength of their data interpretation, but may reflect difficulties in locating information via the interface.

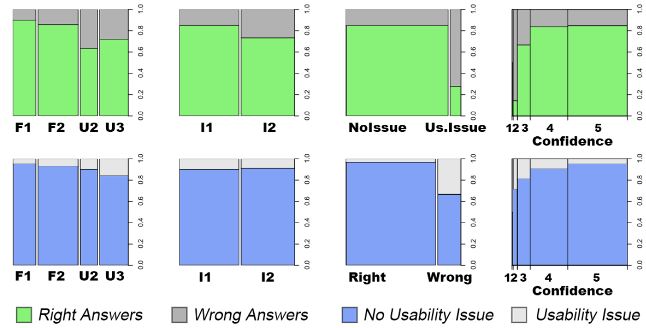



**Figure 8: Proportion of answers in each condition.**

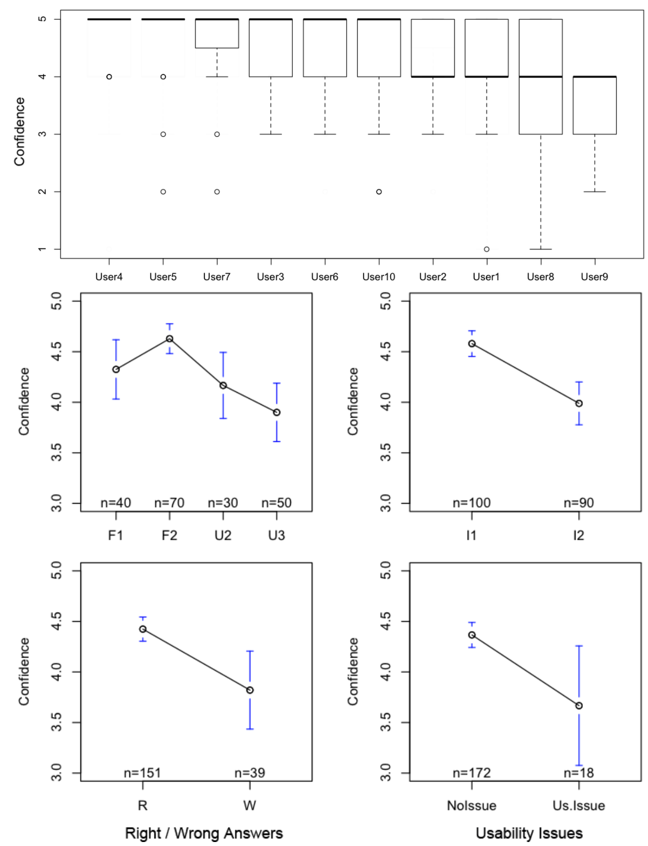



**Figure 9: Confidence levels per users and per groups of questions.**

**Interpretation and Recommendations** - User confidence was generally high, even for wrong answers and uncertain information. Over-confidence may be due to cultural factors. We expected skepticism from scientific users, but paradoxically they may feel obliged to express only sure answers. It may also be due to the presence of observers during the task, inducing a will to perform well (Hawthorne effect). Taiwanese culture may reinforce the incentive for providing solid, valuable answers to not let interrogators helpless (Taiwanese people were always extremely helpful with us as visitors). We recommend that studies of user awareness of uncertainty give strong incentives for users to express their low confidence. The 5-grade likert scale may be reduce to a 2-grade (confident or not) or 4-grade scale.

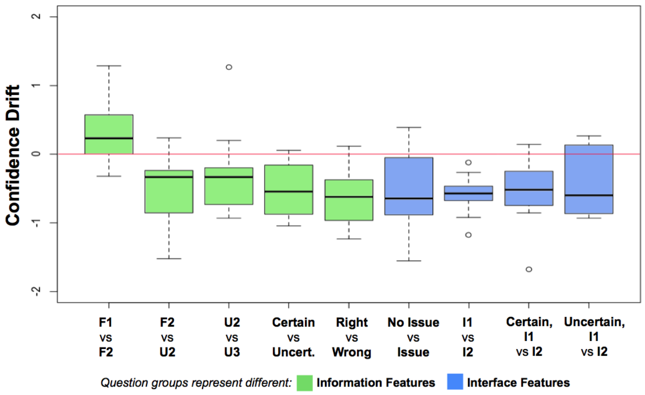



**Figure 10: Confidence drifts per question groups.**

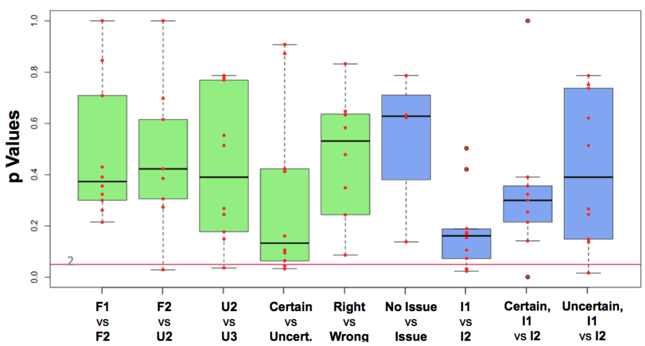



**Figure 11: p-values from Welch t-tests (each point represents a user).**

Users overlooked uncertainty due to fragmentary processing. No spontaneous *Projection* (F3) of possible video numbers occurred: questions Q7, 10, 19, 20 did not show numbers of videos, and users did not extrapolate the potential imbalance in numbers of videos. Yet *Perception* (F1) and *Comprehension* (F2) of numbers of videos were correct (Q5, 11, 12). Hence no right answers were given to Q7. Otherwise answers were fortuitously correct regardless of fragmentary processing. Question formulation may be ambiguous, e.g., *"What is the fish abundance?"* may be interpreted as a need for raw fish count. Widgets' histograms displayed only raw fish counts, thus it may seem the main metric for fish abundance and eclipse the average fish counts per video. Fragmentary processing is similar to sampling size issues (e.g., not enough observations), a well-known concern in marine ecology. Yet it is specific to computer vision, and not assimilated by ecologists, e.g., they may expect video stream processing to be continuous, rarely missing videos. Hence we recommend that this uncertainty factor is always uncovered. Raw fish counts can be harmful for studying fish populations, and by default, should not be displayed. Average fish counts per video should be displayed together with an indication of the sampling size (e.g., encoded as an extra visual dimension like transparency). Terms closer to ecology vocabulary may be preferable (e.g., *sampling size* rather than *numbers of videos*).

No users spontaneously considered certainty scores, which is an unfamiliar, complex concept. However, some users anticipated uncertain factors we overlooked: issues with fields of view, duplicated individuals and differentiating biases from noise (Table 1). All uncertainty factors were not foreseen at the first stage of the requirements analysis, neither by domain experts nor by system experts. They were uncovered at each stage of the design process (Table 1, last column). User-Centered Design had limitations for introducing this technology to a novel application domain. Requirements from non-experts only would produce *incoherent design* as [16] notifies. System experts were particularly needed for specifying *Uncertainty* and *Error Modeling* [7]. However, they did not identify nor quantify uncertainties beyond general-purpose computer vision evaluation (e.g., heterogeneous *Fields of View, Duplicated Individuals*). We recommend that user needs regarding uncertainty management be iteratively investigated by a multidisciplinary team involving domain and system experts, designers, and novices. For instance, it is a coral ecologist, novice to fish ecology, who uncovered the crucial issue with *duplicated individuals*.

The increase in confidence observed between F1 and F2 questions, although not statistically signifiant (p=.07, Welch t-test), suggests an effect of the learning curve. Experiencing and overcoming slightly higher complexity may induce a sentiment of higher level of expertise. User confidence may be reinforced, while the information they had dealt with do not justify it. Similarly, interacting with the interface and using unknown functionalities, may reduce user confidence on a short-term basis. As they need to learn the interaction features, users are not confident that their interactions, and thus the obtained information, are correct. Thus we highly recommend providing a tutorial and a brief memo summarizing the uncertainty exploration steps needed for valid data analysis. These should be easily accessible from the user interface, for quick checks while interacting with the data. It can help users building their confidence beyond dealing with the interface and its interactivity.

Predefined filters of the 3rd *probe* were often overlooked, most probably because users did not set them up themselves. However, it suggests potential attention tunneling issues with the layout design. User attention may be directed to more salient features of the interface, e.g., the main graph, rather than the selected filters. In the next version of the interface, filters were reinforced and included in *Zone B*. It now describes both main graph and filters in natural language (Fig. 2), and can serve as a title for the main graph. The dimensions not used for filtering (e.g., all years, all certainty scores) can saturate users' working memory, and are no longer displayed. Users tried to click on the filter summary, thus we added interactions for reseting the filters (cross buttons). The interaction design for manipulating the widgets and main graph was welcome and easily understood (*"It is very nice, I can display anything I want."*). Participants used either histograms and main graph when appropriate. It suggests that our interaction design is reasonable, while our layout design raised most of the usability issues.

We recommend that uncertainty is always salient in the interface. It may complicate the layout design, yet it may be the best tradeoff regarding the high risk of misinterpretation. Our design of simple graphs in multiple views is intuitive and quickly understood. But it may over-simplify data exploration at the cost of concealing the uncertainties. Over-simplification may enhance attention tunneling, memory loss and over-confidence.

## 5. CONCLUSION

We presented and evaluated an interactive visualization for exploring multidimensional data containing multiple uncertainty factors. It aims at limiting potential data overload and interface cluttering, while facilitating the exploration of data dimensions with flexible visualizations. It supports preliminary data analyses for a wide range of potential usage of the dataset, which may be achieved with specialized data analysis techniques. Our design for preliminary data exploration can help users to familiarize with novel datasets, and identify issues and uncertainties that may impact further data analyses. Our interaction design was found intuitive and easy to understand, although the dataset was unfamiliar to users. Our design can contribute to similar use cases, possibly within domains other than marine ecology. Our evaluation method, inspired from the Situation Awareness domain, can contribute to other evaluations of uncertainty awareness, or to distinguish issues with either layout or interaction design. Our main finding is that user confidence is generally high, and subjectively influenced by the interactions with the visualization: interaction complexity had similar effects than uncertainty itself. Using simple graphs with multiple views achieves high intuitiveness but may have negative effects on user awareness of uncertainty. The simplicity of the graphs and interactions may have contributed to overconfidence through a sentiment of mastering the interface and its information, which led to overlooking uncertainty issues. Furthermore, uncertainty assessment requires the visualization of several graphs within multiple views, which may yield attention tunneling and memory loss, and induce misinterpretations and unawareness of crucial information on uncertainty.

## 6. REFERENCES


[1] Tableau software, http://www.tableausoftware.com.

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Symposium on Information Visualization (Infovis)*, pages 111–117. IEEE, 2005.

[3] E. Beauxis-Aussalet, E. Arslanova, L. Hardman, and J. van Ossenbruggen. A case study of trust issues in scientific video collections. In *2nd International Workshop on Multimedia Analysis for Ecological Data (MAED)*. ACM, 2013.

[4] E. Beauxis-Aussalet, E. Arslanova, L. Hardman, and J. van Ossenbruggen. A video processing and data retrieval framework for fish population monitoring. In *2nd International Workshop on Multimedia Analysis for Ecological Data (MAED)*. ACM, 2013.

[5] E. Beauxis-Aussalet and L. Hardman. Interactive visualization of video data for fish population monitoring. In *CHI Sparks Conference*, pages 152–155, 2014.

[6] J. Bertin. Semiology of graphics: diagrams, networks, maps. 1983.

[7] C. D. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *Symposium on Visual Analytics Science and Technology (VAST)*, pages 51–58. IEEE, 2009.

[8] M. Elias and A. Bezerianos. Exploration views: understanding dashboard creation and customization for visualization novices. In *Human-Computer Interaction–INTERACT*, pages 274–291. Springer, 2011.

[9] M. R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 32, pages 97–101. SAGE Publications, 1988.

[10] M. R. Endsley. Situation awareness global assessment technique (SAGAT). In *Proceedings of the National Aerospace and Electronics Conference (NAECON)*, pages 789–795. IEEE, 1988.

[11] L. Grammel, M. Tory, and M. Storey. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):943–952, 2010.

[12] H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *SimVis*, pages 143–156, 2006.

[13] J. Heer, F. van Ham, S. Carpendale, C. Weaver, and P. Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In *Information Visualization*, pages 92–133. Springer, 2008.

[14] A.-L. Jousselme, P. Maupin, and É. Bossé. Uncertainty in a situation analysis perspective. In *Proceedings of the Sixth International Conference of Information Fusion*. IEEE, 2003.

[15] B. McGuinness. Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings. DTIC Document, 2004.

[16] D. A. Norman. Human-centered design considered harmful. *Interactions*, 12(4):14–19, 2005.

[17] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997.

[18] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Symposium on Visual Languages*, pages 336–343. IEEE, 1996.

[19] D. Tang, C. Stolte, and R. Bosch. Design choices when architecting visualizations. *Information Visualization*, 3(2):65–79, 2004.

[20] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Visualization and Data Analysis (VDA)*, pages 146–157, 2005.

[21] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[22] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced Visual Interfaces (AVI)*, pages 110–119. ACM, 2000.

[23] C. D. Wickens and C. M. Carswell. Information processing. *Handbook of human factors and ergonomics*, pages 89–122, 1997.