Fish4Knowledge Deliverable D2.3

Appendix I - How UI features address biologists requirements

Principal Author: J. van Ossenbruggen, L. Hardman, E.
                     Beauxis-Aussalet, J. He
Contributors: CWI, UCAT, UEDIN
Dissemination: PU

Abstract: This document describes how the Fish4Knowledge user interface addresses the user requirements stated in the Deliverables 2.1 and 2.3.

1 Overview of the requirements

The user requirements are mainly expressed in the Deliverable 2.1. Some of them were excluded from the scope of the UI because of the comparatively high implementation costs. These requirements are complemented by the Deliverable 2.3, because initially users were not able to envision the uncertainties related to the errors introduced by our video analysis components, nor the certainty scores we implement to support them in understanding these uncertainties. We summarize the user needs of the Deliverables 2.1 and 2.3 into 5 general requirements. The following section details these requirements, the UI features that address them, and the status of their implementation.

2 Addressing the requirements

2.1 Support the analysis of population dynamics

2.1.1 Definition of the requirement

D2.1-A: We must support the analysis of population dynamics by providing the following metrics: abundance, species composition, species richness. These metrics can be calculated on specific datasets that target fish with specific characteristics (e.g., species, behavior, location...) *These metrics refer to the functionalities specified in Section 3 of Appendix VI of D2.1.*

2.1.2 UI features and implementation status

This requirement is addressed in the UI by the features that allow users to 1) select the metric of interest in the *Zone C* indicated in Fig.1; and 2) calculate the metric for the dataset of interest, selected through the parameters available in the *Zone D*. For more details about these features, see Figures 3, 15-20 and 23 in D2.3. These features are being implemented, and the *abundance* in counts of fish is the first metric currently available.

2.2 Support browsing of videos of interest

2.2.1 Definition of the requirement

D2.1-B: We must support browsing of videos of interest, especially the videos that correspond to the datasets that are visualized. *This refers to the video search functionality specified in the Appendix VI of D2.1.*

2.2.2 UI features and implementation status

This requirement is addressed in the UI by the features that will be contained in the tab called "Videos", as displayed in the *Zone A* in Fig.1. This tab allow users to browse the videos that correspond to the data that are currently visualized. This part of the UI is not yet designed.
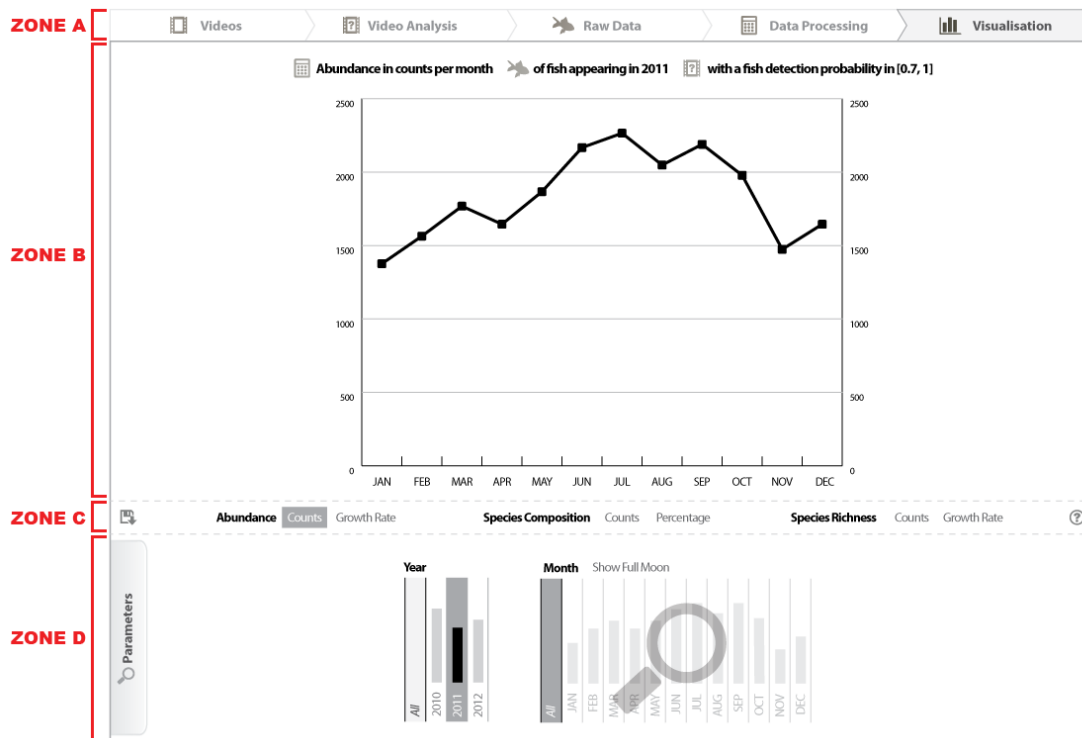
Figure 1: **Counts of fish over the month of the year** (i.e., the *abundance in counts*). This illustrates the basic organization of the UI, and how it addresses the requirement **D2.1-A**.

2.3   Support the identification and the correlation of trends

2.3.1   Definition of the requirement

D2.1-C: We must support the identification of trends and correlations of trends in specific datasets and metrics of interest. *This refers to the most important queries envisioned by biologists, e.g., as detailed in Section 5 of Appendix VI of D2.1. Most of these queries involve the identification and the correlation of trends.*

2.3.2   UI features and implementation status

This requirement is addressed in the UI by the features that allow users to 1) overlay metrics that are calculated for 2 different datasets of interest, as illustrated in Fig.2; and 2) choose the *x-axis variable* as illustrated in the Fig.3. The first feature, the overlay of 2 metric's results, is not yet implemented, and the second feature, the *x-axis variable* is partially implemented.

2.4   Provide an overview of the uncertainty inherent to video analysis components

2.4.1   Definition of the requirement

D2.3-A: For each video analysis component, we must provide an overview of the potential errors that are inherent to their specific video analysis function. This overview must report on i) the ground truth dataset, its inherent errors due to involvement of human judgment; ii) the related ROC evaluation of the components; and iii) the certainty score profile of the components. *This refers to the requirements expressed in sections 3.1 and 3.2 of D2.3.*

2.4.2   UI features and implementation status

This requirement is addressed in the UI by the features that will be contained in the tab called "Video Analysis", and displayed in the *Zone A*. This tab allow users to browse the ROC evaluation and the certainty profiles of the components that produced the data currently visualized. This part of the UI is not yet designed. This is partially because the design of the computation of the certainty scores is not yet finished.

2.5   Provide an estimation of the potential errors contained in visualized datasets

2.5.1   Definition of the requirement

D2.3-B: We must provide an estimation of the potential errors that are likely to be contained in the specific datasets that are currently visualized. This estimation of errors is expressed using 2 types of inter-related metrics, i) the certainty scores, and ii) the estimated numbers of True Positives, False Positives and False Negatives. From a high-level point of view, this is meant for users to evaluate the level of confidence in the trends and correlations of trends they identify. *This refers to the requirements expressed in section 3.2 of D2.3.*

2.5.2   UI features and implementation status

This requirement **D2.3-B** is addressed in the UI by the features that allow users to 1) display an estimation of True Positives, False Positives and False Negatives along with the fish counts currently visualized; and 2) display an estimation of the errors involved by the certainty score profile of the fish counts currently visualized. Figures 4 and 5 illustrate these features.

The estimation of the level of confidence in the identified trends is complemented by basic statistical analyses as illustrated in Figures 28-30 of D2.3. The tabs named "Raw Data" and "Data Processing" will also contain information that help users in understanding the uncertainties contained in the data, and how it propagates to the metrics they are visualizing.

This part of the UI is not yet designed. This is partially because the design of computation of certainty scores and the related estimation of fish counts are not yet finished.
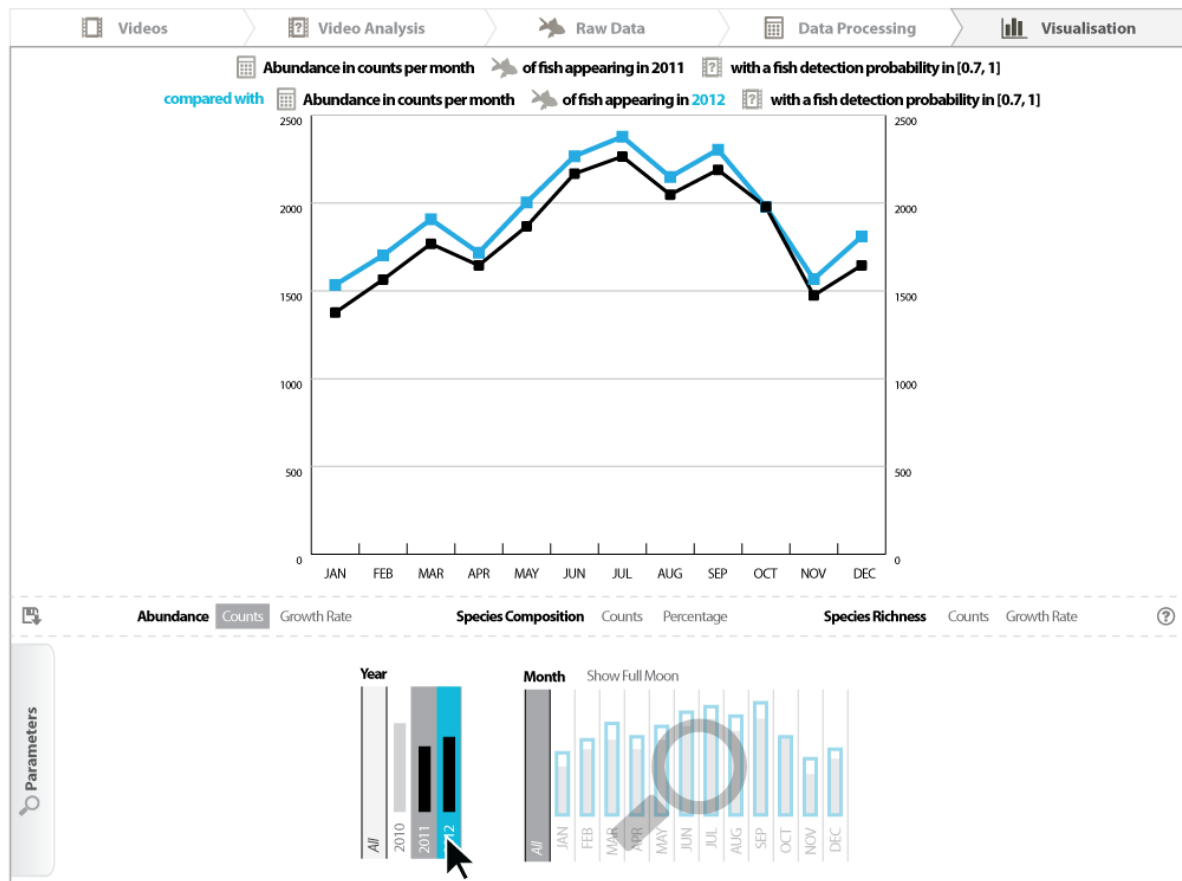
Figure 2: **Two comparable sets of population metrics**, i.e. the *abundance in counts* for each month of 2011 and 2012. The alternative set of population metrics is obtained by a rollover on the alternative variable value. In this example, the user rolls over the year 2012 in the *"Year"* widget on the bottom left. The alternative variable value is highlighted in blue. A new set of population metrics is calculated using all the other variables used for the previous set of metrics. The new set of population metrics is displayed in blue in the main graph. The title of the graph describes the 2 sets of population metrics that are compared. This addresses the requirement **D2.1-C**.
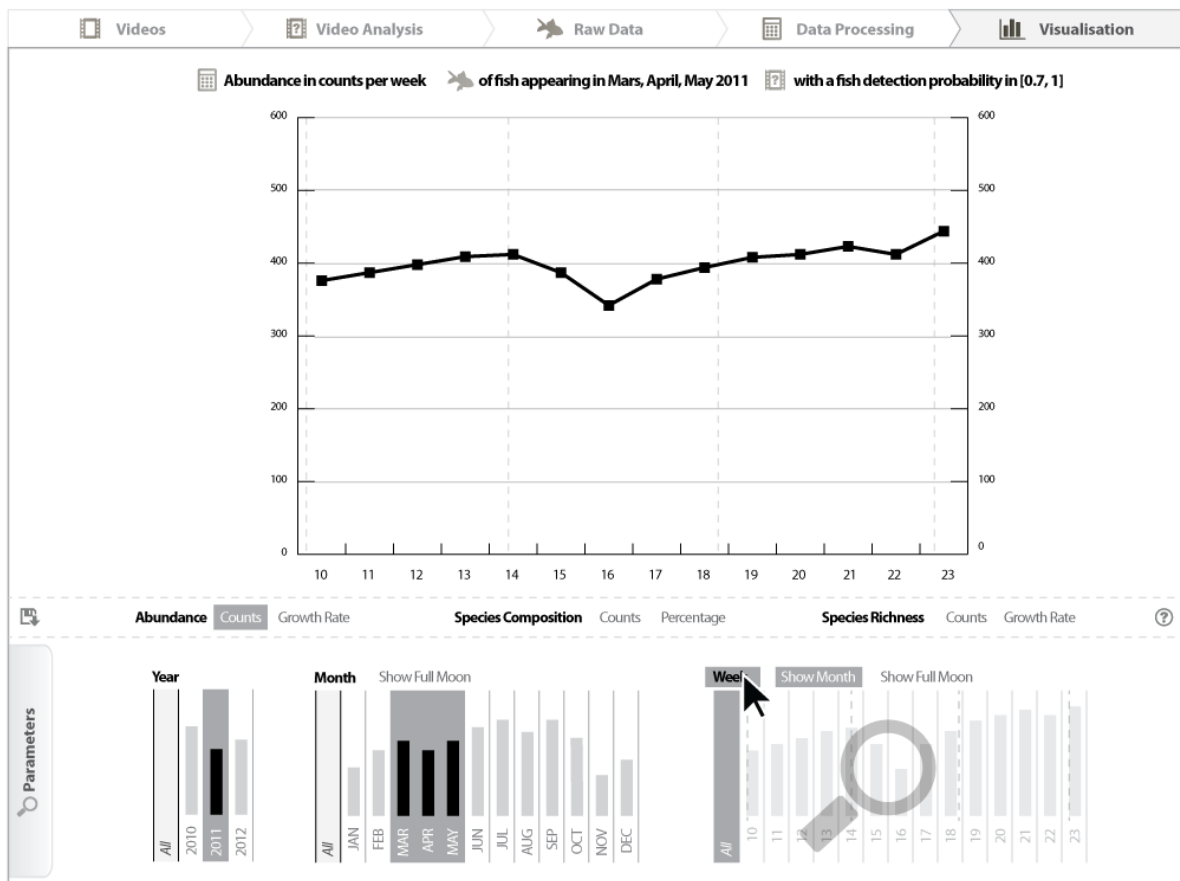
Figure 3: **Selection of the *x-axis variable*.** In this example, the user sets the x-axis to weekly counts of fish. This is done by clicking on the title of the *"Week"* widget (on the bottom right of the UI). The x-axis displays the week numbers for each week of the timeframe of interest (e.g., March, April and May 2011). This addresses the requirement **D2.1-C**. For instance, by selecting "cameras" as the x-axis variable, users can identify and compare trends over different locations.
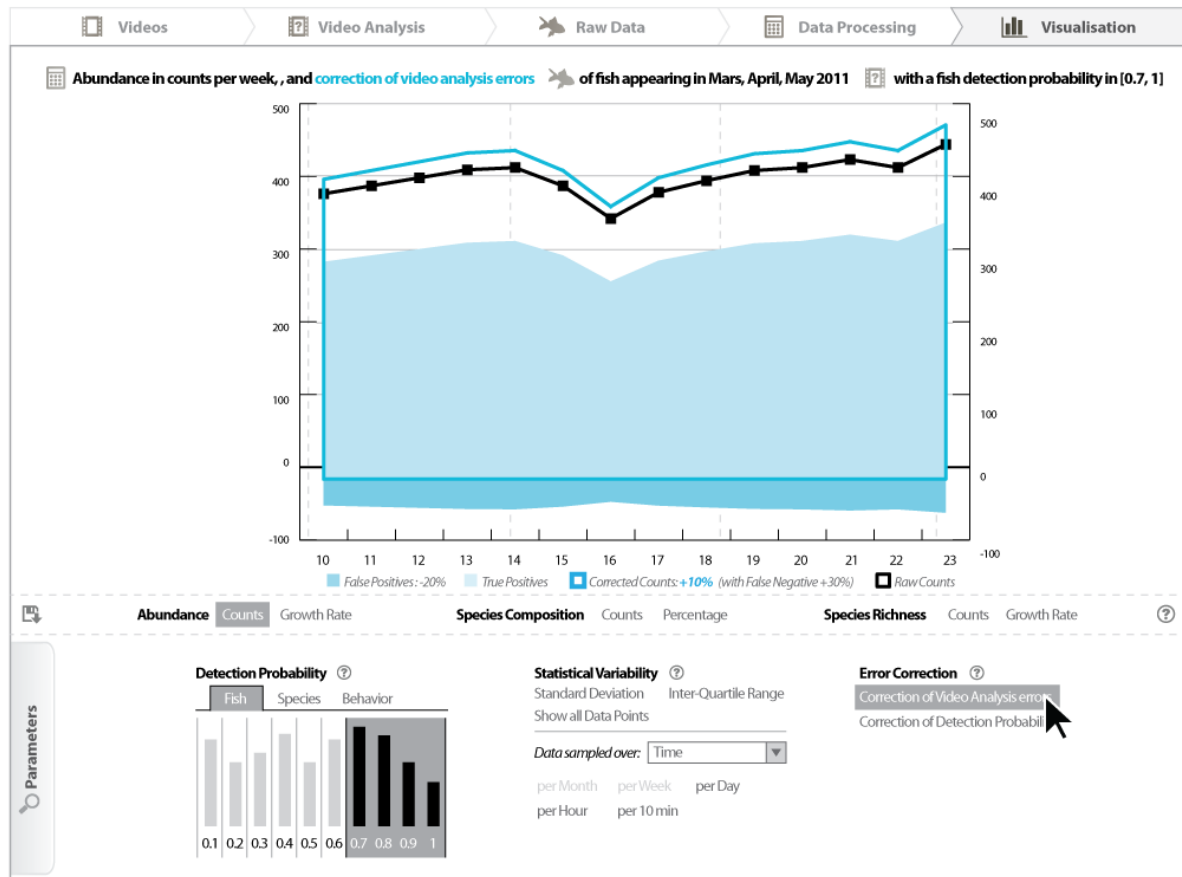
Figure 4: **Estimation of *video analysis errors*.** As described in section 4.1.2 of D2.3, the estimated number of True Positives, False Positives and False Negatives are reported on the population metric results. In the example, the user rolls over the *"Correction of Video Analysis errors"* button (in the bottom right of the UI). This triggers the display of the estimated errors and the corrected count of fish. If the user clicks on the error correction button, this causes the calculation of population metrics using only corrected counts of fish, without the estimated video analysis errors. This addresses the requirement **D2.3-B**.
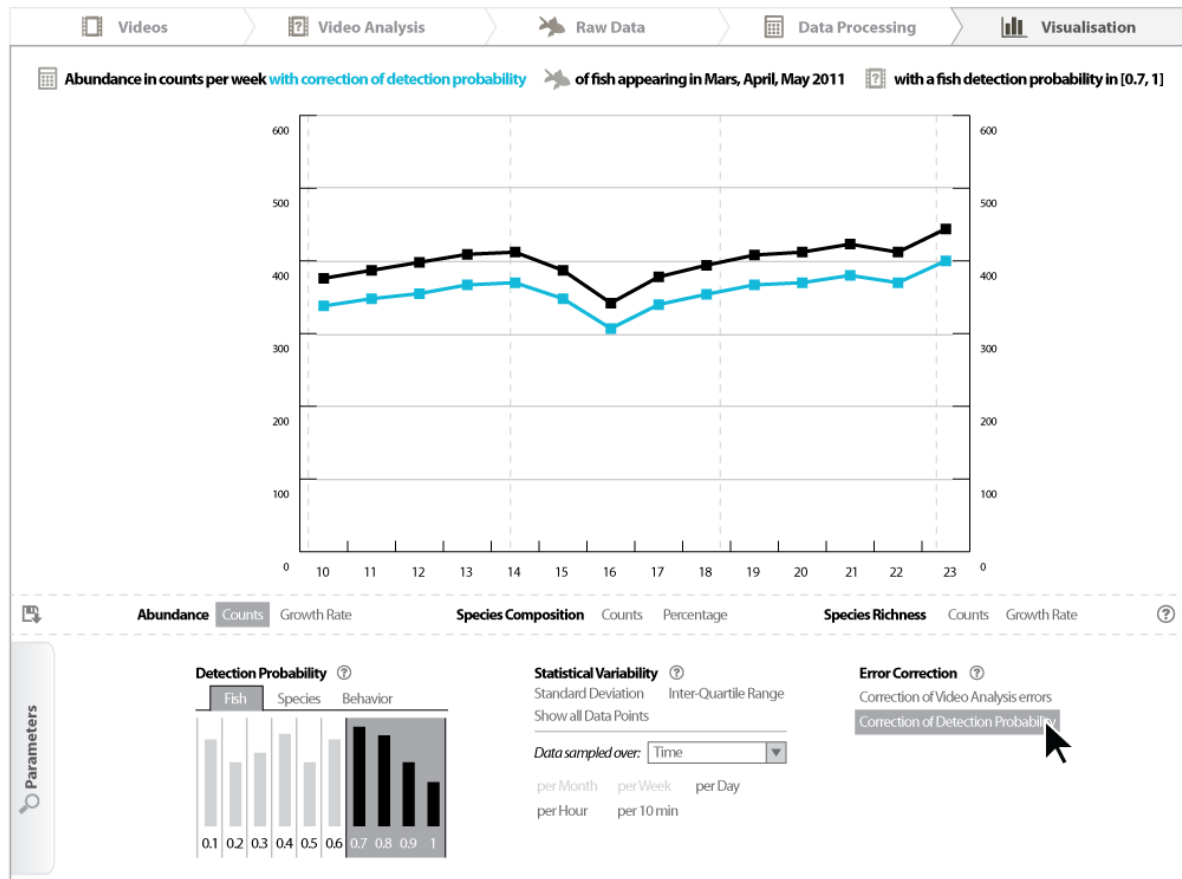
Figure 5: **Estimation of *detection probability errors*.** As described in section 4.1.2 of D2.3, the estimated errors due to imperfect detection probability are reported on the population metric results. In the example, the user rolls over the *"Correction of Detection Probability"* button (in the bottom right of the UI). This triggers the display of corrected count of fish. The counts of fish are corrected using the method defined in section 4.1.2 of D2.3. If the user clicks on the error correction button, this causes the calculation of population metrics using only corrected counts of fish, without the estimated detection probability errors. This addresses the requirement **D2.3-B**.