Inland Excess Water (IEW) Monitoring Using Sentinel-1/2: A SplitClass Segmentation and Temporal Gap-Filling Approach

Supplementary Material

1. Supplementary Material Overview

This supplementary material provides additional details and visualizations to support the main paper. We begin by presenting the methodology used to generate the Inland Excess Water (IEW) map, supported by illustrative examples. Next, we provide further clarification on the separation of training and testing datasets, including a summary table of acquisition dates and a figure showcasing representative samples from each subset. We then analyze the effect of the threshold selection in the proposed SplitClass segmentation approach, along with a comparison between the student and teacher models. Finally, we include high-resolution versions of the segmentation results presented in the main paper to enable more detailed visual inspection.

2. Creation of the Inland Excess Water map (IEW)

The classification workflow of producing IEW Ground Truth Maps required high-level expert control. This methodology was developed by the Lechner Knowledge Center (Lechner Tudásközpont, LTK). Accordingly, the following procedural steps were implemented:

2.1. Threshold-Based Classification of Sentinel-2 Imagery

Sentinel-2 L2A images (Bottom of the atmosphere (BOA) reflectance) are processed in the Google Earth environment to produce the 1st-level of IEW maps. Supported by visual interpretation, pixel color and threshold combinations of

- S2 Band 11 (Short Wave Infrared band, SWIR 1).
- and several spectral indices are considered in the process:
  - Normalized Difference Vegetation Index (NDVI)
  - A custom index was applied operationally within the Lechner Knowledge Center defined as

$$\frac{\text{SWIR1} - \text{RED}}{\text{SWIR1} + \text{RED}}$$

  - Normalized Difference Moisture Index (NDMI) was used occasionally in certain spring periods, when additional information was needed to increase the accuracy of the classification.

Thresholds are suited to the actual characteristics of the area in focus. At this level, no masks are applied. Clouds and cloud shadows are not mapped either. Altogether six thematic classes are distinguished. Figure S1 presents the pseudo-color representation of the input image (using Sentinel-2 bands 8, 11, and 4 in order) alongside the thresholding result based on color information.

2.2. Masking built-up areas and forests:

A mask of built-up and forested areas was applied to the first-level maps to filter out all areas irrelevant to the project (forests, built-up and other sealed areas, roads, and railways). Masked regions are merged with the category of "not affected areas" to generate the 2nd-level IEW map. The mask is derived from the Harmonized National High-resolution Land Cover Layer of Hungary. Figure S2 shows the masked regions on the left and the filtered IEW map with these areas merged into the "not affected areas" category on the right.

2.3. Cloud and cloud shadow mapping

In the 3rd level of the IEW map generation. The clouds and cloud shadows were delineated using the relevant classes of the Sentinel-2 Scene Classification Maps. In most cases sentinel-2 scene classification data had to be corrected and completed manually by LTK experts based on visual interpretation of the original S2 images. The delineation of the clouds and cloud shadows is a key step in the processing because of the modifying effect on the incoming and reflected radiation. However, in the case of cirrus-type (often semi-transparent) clouds, the delineation is associated with high uncertainty. Figure S3 illustrates the cloud and cloud shadow delineation process. The left panel shows the original Sentinel-2 Scene Classification Layer (SCL) with initial cloud and shadow classes, while the right panel displays the manually corrected and completed mask after expert refinement by LTK specialists.

2.4. Integration of cloud and cloud shadow categories into the 2nd level IEW maps

As the last step, the manually corrected cloud and cloud shadow layer was integrated into the 2nd level IEW maps to produce the final level of IEW maps. The number of thematic categories in the final map is extended with clouds and shadows. Figure S4 presents a comparison between the input images and the final IEW maps. On the left, the RGB image and the pseudo-color input (using Sentinel-2 bands 8, 11, and 4 in order) used for processing are shown. On the right, the final level of IEW map is displayed, incorporating the manually corrected cloud and cloud shadow layer. The figure also includes a color legend describing the thematic categories, now extended to include clouds and shadows.

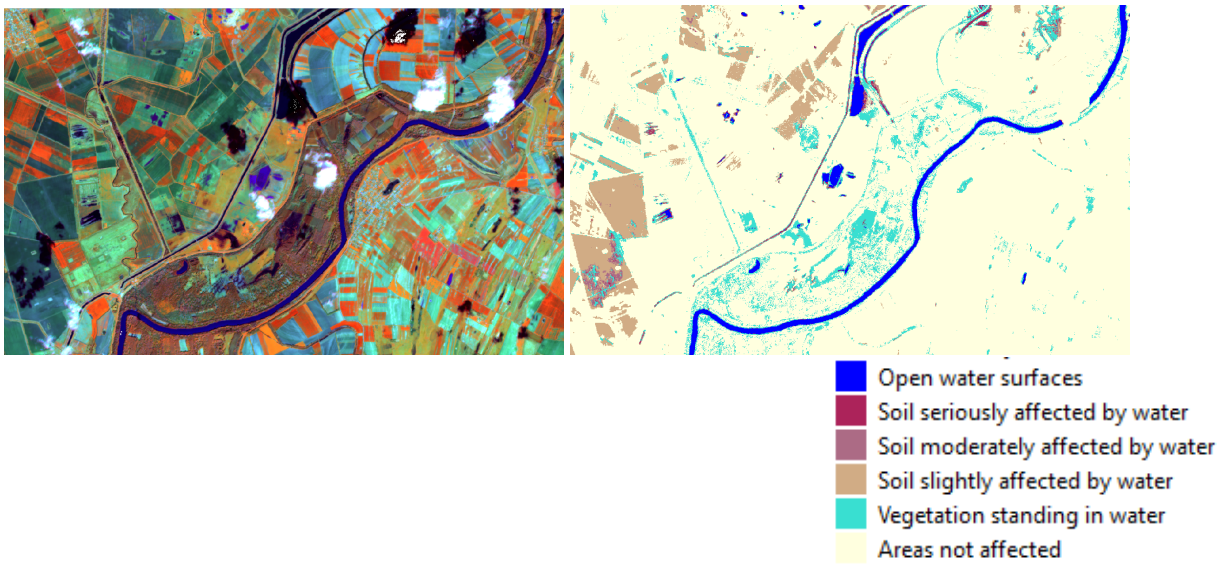


Figure S1. Pseudo-color representation of the input image and the corresponding thresholded output based on color information.

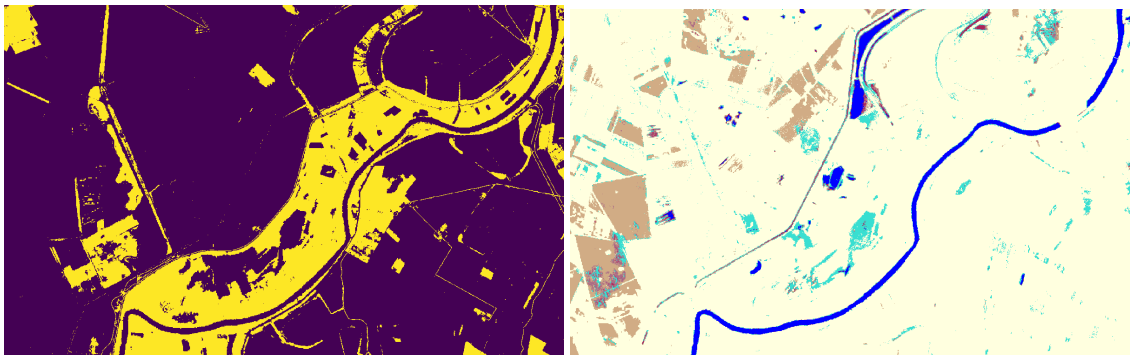


Figure S2. Masked Regions (left) and Filtered IEW Map Output (right) after Applying Built-up and Forest Area Mask.

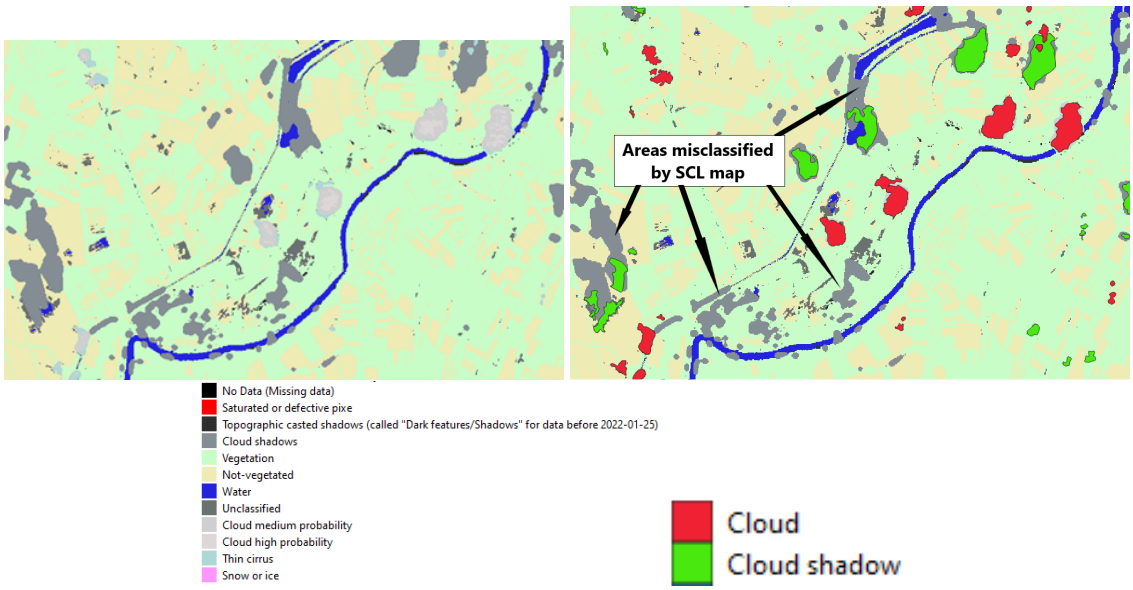


Figure S3. Sentinel-2 Scene Classification Map (left) and Expert-Corrected Cloud and Shadow Mask (right).

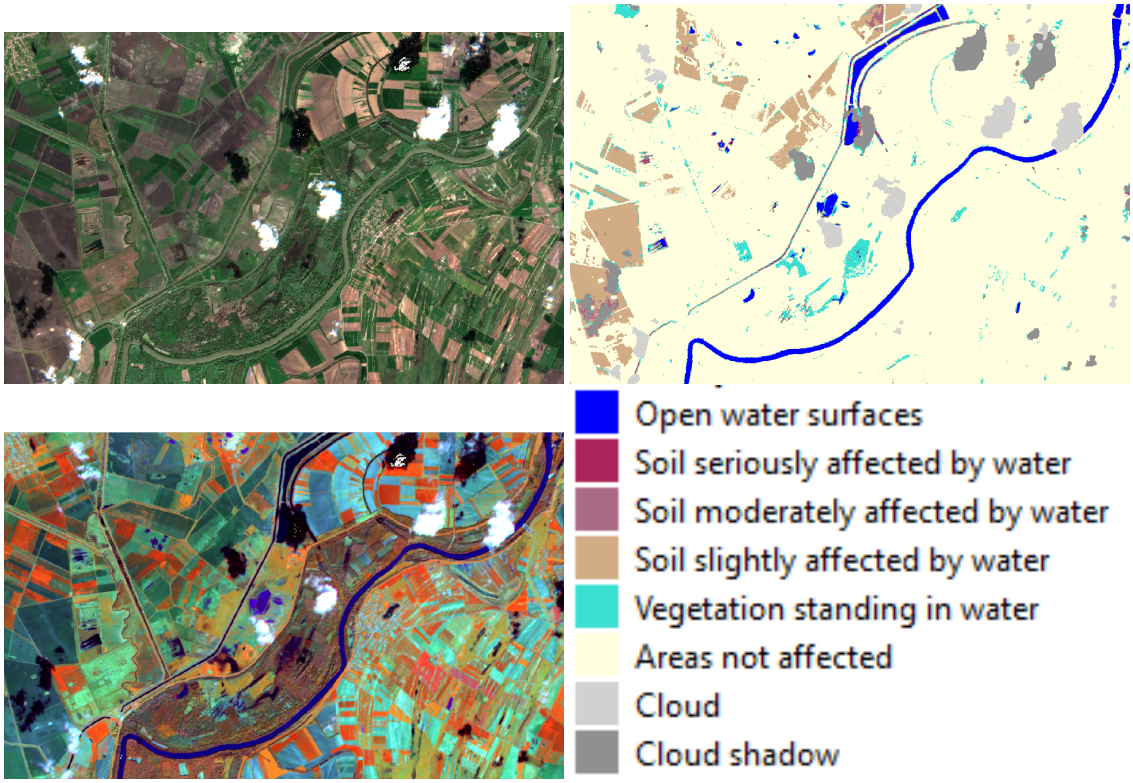


Figure S4. Input RGB and Pseudo-Color Images (left) and Final IEW Map with Clouds and Shadows (right) Along with Color Legend

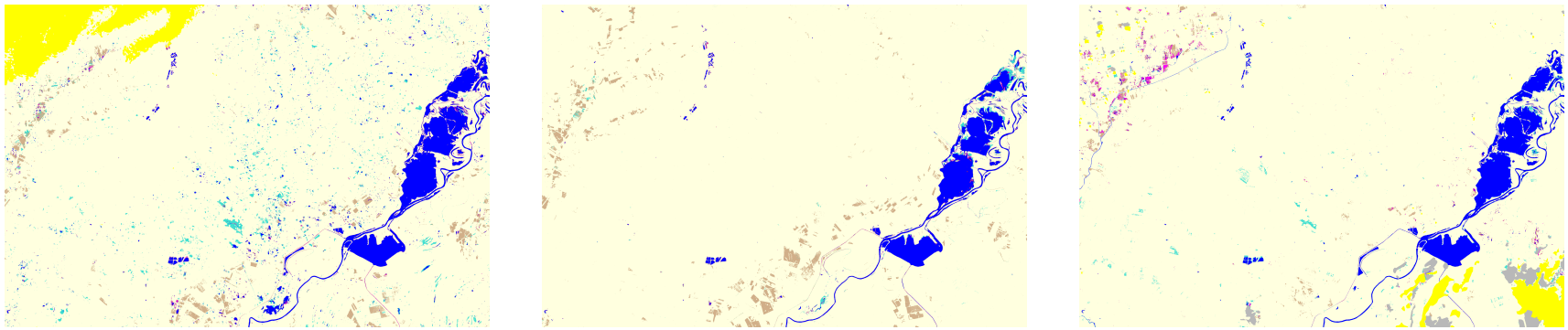
3. Dataset Composition and Temporal Coverage

To ensure a fair and temporally independent evaluation of model performance, we divided the dataset into separate training and testing sets for each study region. The regions include the Heves Steppes in the Bükk National Park (BNPI) and the sodic lakes of Upper-Kiskunság in the Kiskunság National Park (KNPI). This division was designed to cover diverse seasonal and interannual conditions while maintaining consistent spatial coverage. The details of this split, including the years and seasons of image acquisitions, are summarized in Table S1.

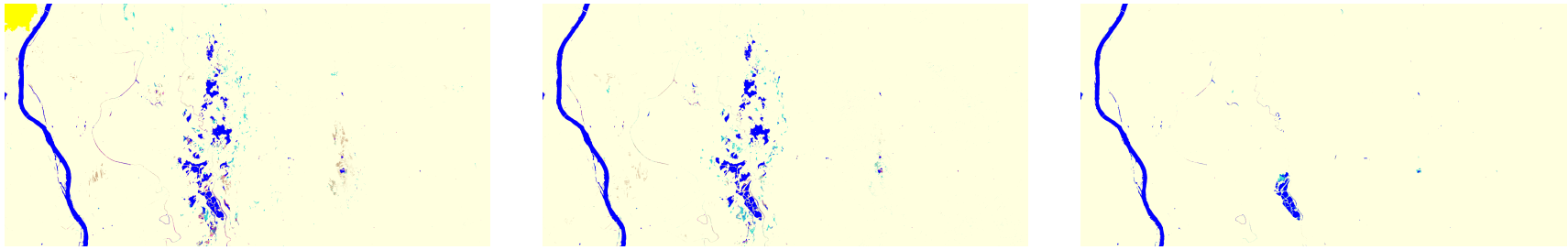
A visual overview is provided in Figure S5, which shows representative samples from both regions. The first row presents examples from the Heves Steppes (BNPI), while the second row displays maps from the sodic lakes area in Upper-Kiskunság (KNPI). These examples illustrate the geographical consistency and seasonal diversity of the dataset used for training and testing.

		Train	Test
BNPI	No. samples	7	4
	Dates	2021-02-23, 2021-03-10, 2023-02-08, 2023-02-13, 2024-01-09 , 2023-05-19 2024-02-03	2022-02-08, 2022-03-25, 2023-11-25, 2024-01-29,
KNPI	No. samples	7	5
	Dates	2021-02-18, 2021-02-23, 2022-01-19, 2022-02-13, 2022-03-10, 2022-03-25, 2022-07-23	2023-02-13, 2023-06-03, 2024-01-04, 2024-01-29, 2024-02-03

Table S1. Training and testing dataset distribution for BNPI and KNPI. Dates are formatted as YYYY-MM-DD



BNPI 2021-03-10 BNPI 2022-02-08 BNPI 2023-05-19



KNPI 2021-02-23 KNPI 2022-01-19 KNPI 2022-07-23

Figure S5. Temporal Examples of IEW Maps for BNPI (Top) and KNPI (Bottom) Across Training and Testing Dates. Dates are formatted as YYYY-MM-DD.

4. Quantitative Evaluation of Soft Segmentation Performance

In this section, we present a quantitative evaluation of the proposed SplitClass segmentation approach. Table S2 shows the performance of the teacher and student models based on Top-2 class predictions, while Table S3 presents the effect of varying the confidence threshold τ on the student model's segmentation performance.

Table S2 compares the performance of the teacher and student models based on Top-2 predictions, where a prediction is considered correct if the ground truth class is among the top two predicted classes. Precision (Pr), Recall (Rc), and F1-score are reported per class and overall. The student model performs competitively, slightly outperforming the teacher in several categories—especially in moderately and slightly waterlogged soil classes—highlighting the effectiveness of knowledge distillation and the model's generalization ability.

Notably, the teacher model also shows significantly improved scores under the Top-2 evaluation compared to its Top-1 (argmax) predictions (See Table 2 in the main paper). This indicates that the second-ranked predictions of the pretrained teacher model often include the correct class, particularly in ambiguous or transitional areas. Therefore, applying the SplitClass segmentation—even without retraining or modifying the teacher—can extract additional value from the model's predictive uncertainty. These results strongly support the practical usefulness of SplitClass as a lightweight enhancement strategy, capable of improving segmentation reliability in complex scenarios such as waterlogged soil classification.

Table S3 presents the student model's performance under three different confidence thresholds: $\tau = 0.5$, $\tau = 0.8$, and $\tau = 0.98$. These thresholds define when a second class is added to the final prediction based on the model's confidence. As the threshold decreases, the model becomes more selective, leading to reduced recall and F1-score in some affected classes (e.g., Classes 3–5). These results confirm that the SplitClass approach effectively manages uncertainty, particularly for the underrepresented classes or those rarely appearing due to data imbalance, such as transitional or mixed-category pixels in waterlogged regions.

5. High-Resolution Versions of Selected Figures

The following figures present high-resolution versions of key results from the main paper to support detailed visual inspection. Figure S6 corresponds to Figure. 4 in the main text and shows the combined temporal predictions for waterlogging detection, integrating both radar-based results and SplitClass outputs. Figure S7, aligned with the top-right panel of Figure. 5, provides a detailed visualization of the model outputs, overlaid with expert-annotated ground truth classes and field photo identifiers. Lastly, Figure S8 offers a high-resolution version of Figure. 6, illustrating model performance in an unseen rice field near Doñana, Spain, and comparing the WIW method, our Top-1 prediction, and the proposed SplitClass segmentation. These figures provide greater visual detail to support the interpretation of the results.

ID	Teacher Top2			Student Top2		
	Pr ↑	Rc ↑	F1 ↑	Pr ↑	Rc ↑	F1 ↑
2	0.99	1.00	0.99	0.99	1.00	**1.00**
3	0.90	0.89	0.89	0.91	0.90	**0.90**
4	0.87	0.84	0.85	0.90	0.84	**0.87**
5	0.87	0.92	0.89	0.94	0.97	**0.95**
6	0.92	0.99	**0.95**	0.93	0.96	0.94
7	1.00	0.99	**1.00**	1.00	1.00	**1.00**
All	0.99	0.99	**0.99**	1.00	1.00	**1.00.**

Table S2. Performance Top2 classes for the student and the teacher. Bold values indicate the best performance for each class.

ID	$\tau = 0.5$			$\tau = 0.8$			$\tau = 0.9$		
	Pr ↑	Rc ↑	F1 ↑	Pr ↑	Rc ↑	F1 ↑	Pr ↑	Rc ↑	F1 ↑
2.0	0.99	0.99	0.99	0.99	1.00	0.99	0.99	1.00	0.99
3.0	0.78	0.71	0.74	0.87	0.84	0.86	0.90	0.87	0.88
4.0	0.67	0.63	0.65	0.83	0.77	0.80	0.88	0.81	0.84
5.0	0.59	0.69	0.63	0.73	0.79	0.76	0.79	0.83	0.81
6.0	0.80	0.78	0.79	0.89	0.86	0.88	0.91	0.90	0.91
7.0	0.99	0.98	0.99	1.00	0.99	0.99	1.00	0.99	0.99
All	0.98	0.97	0.98	0.99	0.98	0.98	0.99	0.99	0.99

Table S3. Performance Comparison student model with different thresholding for Waterlogging Detection, by class. blue for the best, orange for the second best.

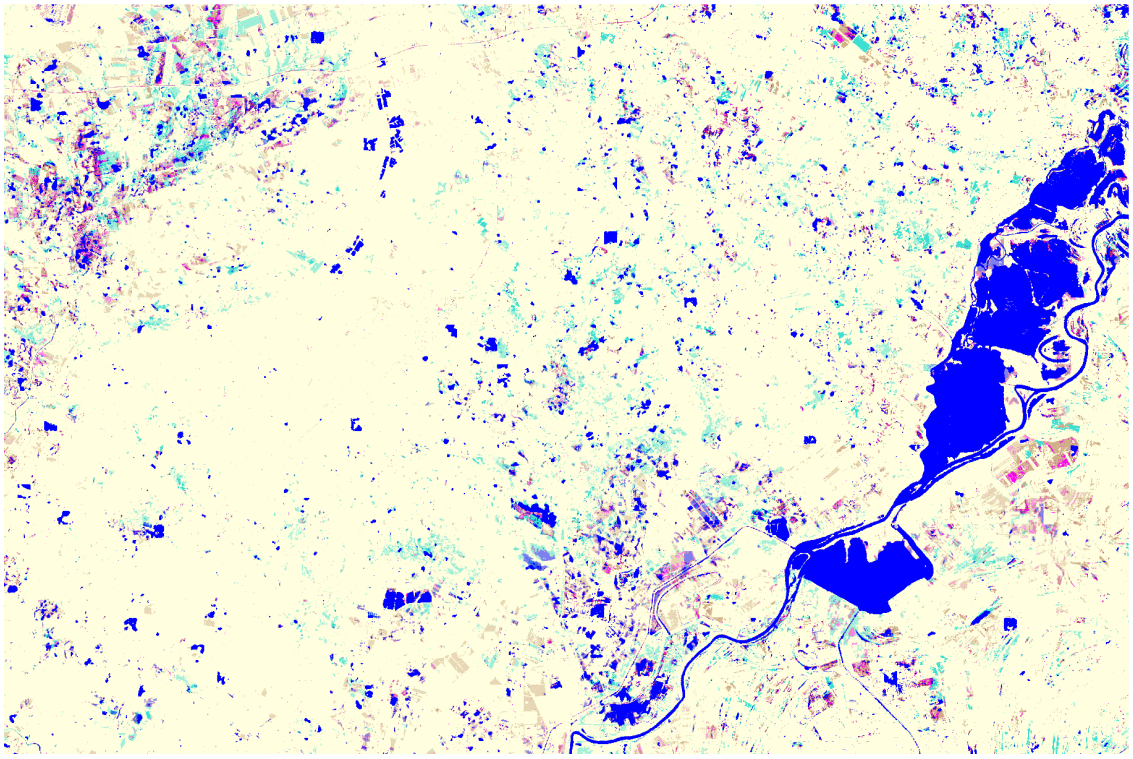


Figure S6. High-resolution version corresponding to Figure. 4 (Final result) in the main paper. Temporal Analysis for Waterlogging Detection: Combined Temporal 2 Class Predictions and Radar Result

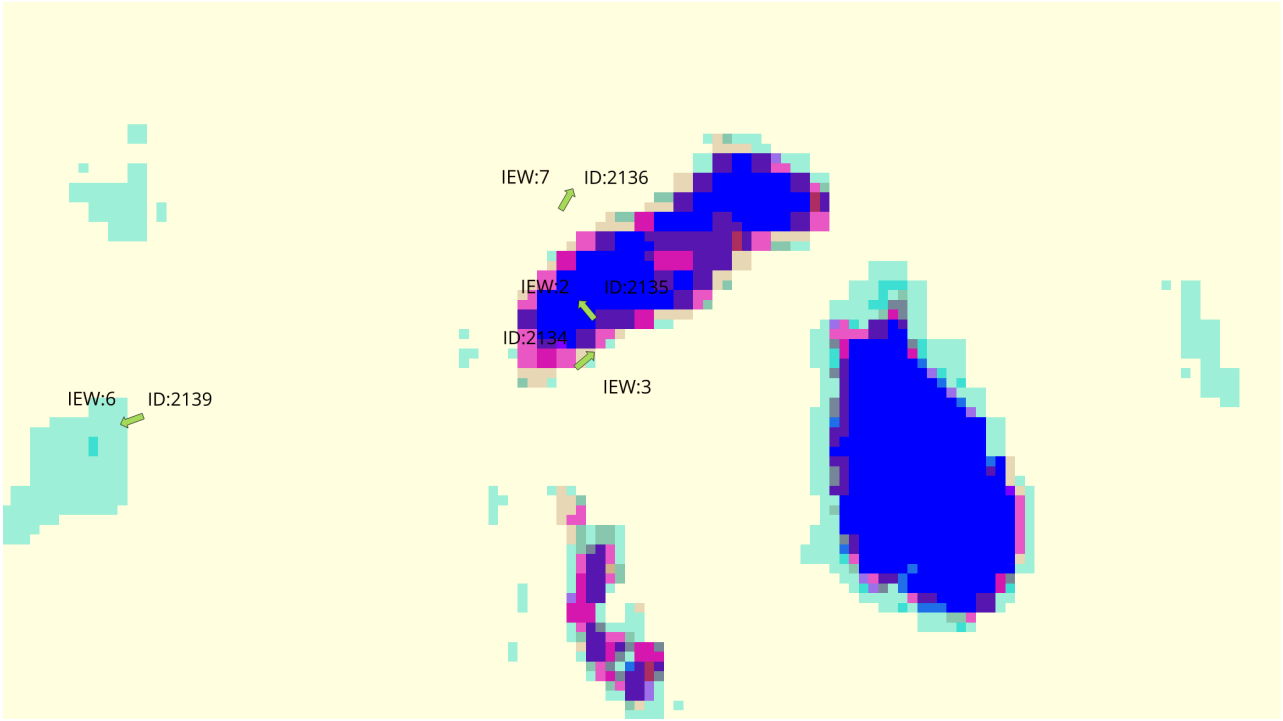


Figure S7. High-resolution version corresponding to Figure. 5 (Top Right) in the main paper. Background: Model Output; Arrows: Expert GT Classes & Field Photo IDs

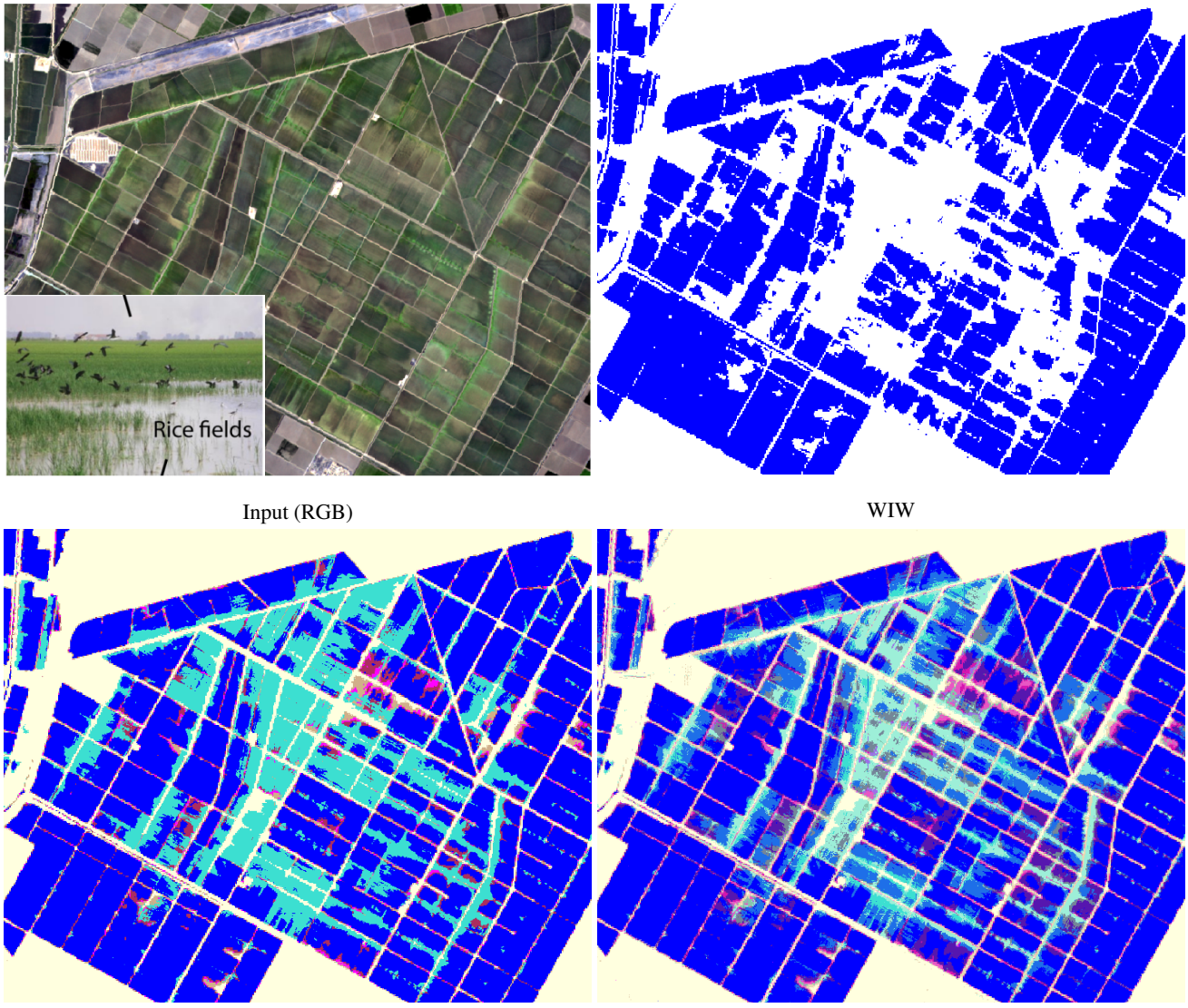


Figure S8. High-resolution version corresponding to Figure 6 in the main paper, Waterlogging Detection in Unseen Rice Field near Doñana, Spain (June 1, 2019): Comparison of WIW Method, Ours Top-1 Prediction, and Ours SplitClass Segmentation.