# PICANTEO: A Modular Change Detection Framework for Remote Sensing Applications

Dimitri Lallement * [1], Christian Hümmer * [1,2], David Youssefi[1], Paul Mauduit[3], Agnès Verzeni[1]

[1] CNES (French Space Agency), 18 Avenue Eduard Belin, 31400, Toulouse, France
[2] DLR (German Aerospace Center), D-82234 Oberpfaffenhofen, Germany
[3] Thales Services SAS, 31670 Labège, France

(dimitri.lallement, david.youssefi, agnes.verzeni)@cnes.fr
christian.huemmer@dlr.de
paul.mauduit@thalesgroup.com

**Keywords:** Multi-Modal Change Detection, Natural Disaster, Uncertainty, 3D, Deep Learning.

**Abstract**

This paper presents PICANTEO, a modular and multi-modal change detection framework designed for remote sensing applications in natural disaster response. The framework aims to support damage assessment during both the rapid mapping phase, which occurs in the immediate aftermath of a disaster, and the longer recovery phase. PICANTEO provides automated, reliable disaster-related change detection maps and associated impacted areas to support a wide range of disaster monitoring activities. The integration of uncertainty and ambiguity concepts ensures reliable and qualified results.
PICANTEO handles multi-modal remote sensing data, including very high-resolution optical imagery, Digital Surface Models, and Synthetic Aperture Radar (SAR) data. Its modular architecture enables users to apply ready-to-use pipelines or implement their own workflows. The provided scalable components can be combined or extended by custom methods to define new applied pipelines. Several real-world case studies demonstrate PICANTEO’s ability to address various disaster scenarios across diverse geographic contexts. Source code is available at: `https://github.com/CNES/picanteo`.

## 1. Introduction

The first application developed within PICANTEO was designed to support damage assessment activities during the rapid mapping phase, spanning from the initial response phase to the first days following the disaster. During this critical period, operational services such as the International Charter “Space and Major Disasters” (Bessis et al., 2004), the Copernicus Emergency Management Service (CEMS) (European Commission, 2015), or Sentinel Asia (Kaku et al., 2013) are activated to produce initial change maps and damage grading as soon as possible. As one of its core concepts, PICANTEO aims to assist human annotation experts in their manual mapping activities by providing qualified damage assessment.

Beyond rapid mapping, the framework also targets the recovery phase, which can last several months. During this period, data acquisition is less time-constrained and benefits from improved imaging conditions. In particular, stereoscopic acquisitions become more frequent, enabling a more accurate damage assessment through the integration of 3D information. The arrival of new missions and constellations, such as CO3D (Lebègue et al., 2020), will also increase the number of available 3D acquisitions, which could potentially even be acquired in selected rapid mapping cases.

PICANTEO is a modular and multi-modal change detection framework that emphasizes flexibility and re-usability. It allows users to build custom processing pipelines from predefined, configurable steps for efficient and scalable analysis. By integrating multi-modal data fusion such as 2D/3D optical imagery or Synthetic Aperture Radar (SAR) data, the framework supports diverse disaster response scenarios. PICANTEO covers the entire processing chain, from data pre-processing and deep learning analysis to 3D reconstruction, uncertainty estimation, and visualization. Its modular architecture allows users to extend the framework by incorporating new applications, ensuring adaptability to evolving operational response needs.

Section 2 provides an overview of the PICANTEO architecture and the key features that enable modularity and flexibility. Section 3 lists the main methods included in the framework to process multi-modal data using multiple approaches. Rather than aiming to outperform existing algorithms, this work demonstrates the framework’s capacity to handle diverse application scenarios and to gather multi-source, multi-event pipelines within a unified platform. The section also highlights the interest of integrating uncertainty and ambiguity estimation into the disaster response framework, where informed decision making based on reliable qualified predictions is crucial and severe domain shifts occur regularly. Finally, Section 4 presents several real use cases in different regions and types of disasters.

## 2. Framework overview

PICANTEO is a framework designed to address specific change detection applications, such as post-disaster building damage assessment or flood detection. For each predefined application, PICANTEO provides a dedicated processing *pipeline*. If users require specific applications that are not covered by the provided pipelines, they can implement their own using the provided structure.

Each pipeline is composed of configurable components, called *steps*, which represent methodological processing units. These steps are organized by functional categories and grouped in

* These authors contributed equally to this work.


https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1291-2026  

a toolbox, as described in Section 3. Their interactions and data flow are defined at the pipeline level. Users can extend the existing toolbox by implementing new steps, following the input/output PICANTEO interface, to integrate missing data processing routines. This flexibility allows PICANTEO to adapt to a wide range of disaster types while ensuring an efficient workflow.

PICANTEO's steps are designed to be scalable and include parallel computation as well as tiling strategies when necessary, to handle large satellite images. In particular, for the 2D change detection workflows described in Section 3.2, the inference steps integrate overlapping tiling schemes, padding strategies to mitigate border effects, and weighting mechanisms to ensure proper aggregation of predictions across tiles.

Figure 1 illustrates an example of a bi-temporal 2D/3D uncertainty-aware very high-resolution optical change detection pipeline. The blue boxes depict different categories of steps, including data extraction, data filtering, and data fusion. This example demonstrates how a specific application can be implemented within a structured PICANTEO pipeline by decomposing the method into modular components, following the step-based architecture of the framework.

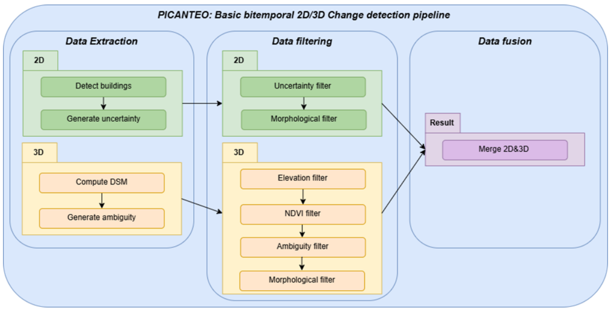


Figure 1. PICANTEO's bi-temporal 2D/3D uncertainty-aware very high-resolution optical change detection pipeline.

## 2.1 Configuration driven approach

PICANTEO relies on a hierarchical configuration file structure. When a developer introduces a new processing step, they provide a corresponding configuration template, which is then populated with the required parameters within the pipeline.

The top-level configuration file used by the end user defines the overall pipeline, including input data and output directories. It also specifies the set of processing steps to be executed. Thanks to its nested configuration system, input parameters defined at the pipeline level are automatically propagated to the underlying step configuration templates. In addition, the internal configuration engine enables dynamic linking of inputs and outputs between steps directly within the main configuration file.

As a result, users can define a complete PICANTEO pipeline through a single configuration file, specifying the required parameters: input data, output directory, processing steps, and mapping of inputs and outputs across steps.

## 2.2 Data visualization

To facilitate user interaction, PICANTEO includes a modular visualization step with synchronized user-defined panels. This dashboard simplifies navigation across application-specific layers, such as 2D changes, building footprints, uncertainty maps, and 3D changes. In the case of the previously described 2D/3D change detection pipeline, the dashboard is composed of three main panels: pre-disaster 2D layers, post-disaster 2D layers, and 3D layers, as illustrated in Figure 2.

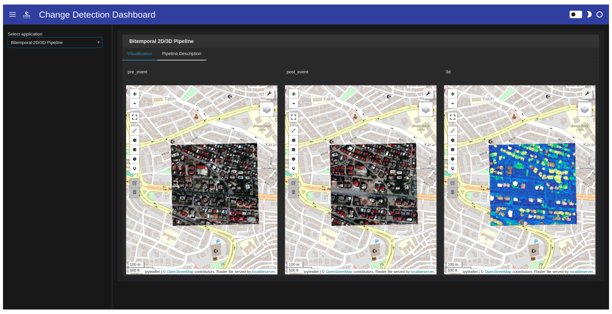


Figure 2. PICANTEO's visualization dashboard. Left: pre-event image with the associated layers; Middle: post-event image with the associated layers; Right: 3D change detection.

# 3. Methodology

## 3.1 Data pre-processing

PICANTEO integrates a set of standard pre-processing operations to facilitate the use of its processing pipelines. These include re-projection of input layers when data are not defined in a common Coordinate Reference System (CRS), cropping to ensure consistent spatial extents between pre- and post-event images, resampling to harmonize spatial resolution, and co-registration to correct potential misalignment between input images.

In addition to these geometric pre-processing steps, PICANTEO also incorporates radiometric pre-processing steps. These include data normalization, which enables proper visualization within the interface described in Section 2.2, and ensures that input data fits the value range required by some downstream steps such as deep learning-based detection models.

## 3.2 2D Data processing

Many approaches in the literature present advanced ensemble model architectures (Durnov, 2020) and optimized baseline training strategies (Hafner et al., 2025) to obtain building segmentation, building damage assessment (BDA) or generalized change detection models, leveraging Convolutional Neural Networks, Vision Transformers or hybrid approaches (Zhu et al., 2026). Usually change detection methods are based on early- or mid-level feature fusion, often leveraging Siamese architectures (Caye Daudt et al., 2018) with shared encoder weights. Our framework does not intend to propose new state-of-the-art architectures, but provides easy-to-use steps and ready-to-use application-specific pipelines with representative baseline models. Thanks to the modular design of the framework, users can easily integrate their own models, as long as they follow the expected I/O interface of the pipeline.

In PICANTEO, a first building detection approach is formulated as a binary semantic segmentation task, where each pixel is classified as either building or non-building. This formulation is well suited to very high-resolution (VHR) remote sensing imagery, with more available training data compared to a direct change detection task, and provides building masks that can subsequently be compared across different dates to obtain change indicators.

A second approach formulates the change detection task directly in a bi-temporal input manner using the aforementioned Siamese architectures. The framework provides a baseline UNet model (Ronneberger et al., 2015) with MA-Net backbone (Fan et al., 2020), trained on publicly available BDA datasets and optimized to provide robust VHR building segmentation across multiple sensors, using strong data augmentations in the training phase. MA-Net introduced self-attention mechanisms that model feature inter-dependencies in spatial dimensions between pixels in a global view via Position-wise Attention Blocks and additional multi-scale Fusion Attention Blocks that capture channel dependencies between multi-scale feature maps. A second internal sensor-specific model trained on Pléiades data and fine-tuned on a manually annotated building footprints dataset was used for evaluation purposes, but is not made available within the pipeline due to licensing restrictions.

In addition to traditional model architectures, recent foundation models, i.e. large models pre-trained on broad and diverse image corpora in a self-supervised way, offer strong transferable visual representations for this task and can be adapted either by full fine-tuning of all model parameters or by parameter-efficient strategies such as adapters or LoRA (Hu et al., 2021). Generalist vision foundation models such as SAM2 (Ravi et al., 2024) and DINOv3 (Siméoni et al., 2025) provide powerful encoders that transfer well to remote sensing tasks. We tested this approach with the integration of a SAM2-UNet model (Xiong et al., 2026), which pairs SAM2’s hierarchical image encoder with a UNet-style decoder. Complementing these generalist approaches, geospatial foundation models pre-trained specifically on large corpora of remote sensing imagery such as SatMAE(Cong et al., 2022) or Prithvi-EO-2.0 (Szwarcman et al., 2026), learn spectral-spatial representations that may be better suited to satellite data characteristics, including multi-spectral bands and varying spatial resolutions. These models have shown strong transfer performance on downstream segmentation and change detection tasks.

**3.2.1 Uncertainty Estimation** Deep learning models often suffer from under-/overconfident predictions and lack a reliable representation of uncertainty (Gawlikowski et al., 2023), although the notion of uncertainty alongside the prediction plays a crucial role in critical applications for reliable decision making, especially in the context of disaster management.The following section gives a brief overview of well established and easily-implementable baselines for uncertainty-aware deep neural networks, which can be integrated into the framework.

As one of the main emerging concepts, Bayesian neural networks allow for a natural probabilistic representation by incorporating priors and placing distributions over model parameters, as well as marginalizing these parameters to derive a predictive distribution. As the posterior over those distributions often becomes intractable for modern deep learning architectures, Bayesian approaches are usually based on approximate methods. In Variational inference, the posterior is approximated by a simpler approximating parametric distribution, e.g. a Gaussian variational distribution (Graves, 2011), which can be derived during usual network optimization, using approaches like Bayes by Backprop (Blundell et al., 2015) and the reparameterization trick (Kingma et al., 2015). After training, multiple stochastic forward passes for one input image yield an approximate predictive posterior, in contrast to the deterministic counterparts, which only obtain one prediction per input sample. As a very basic approximation, Monte Carlo Dropout (MCDO) (Gal and Ghahramani, 2015) stands out as one of the most widely applied approaches for approximate BNNs due to its simplicity. MCDO can be interpreted as a simple variational inference approach, where dropout layers are formulated as Bernoulli distributed random variables (Gawlikowski et al., 2023), with a Bernoulli distribution representing the variational distribution over the network weights. Keeping the dropout layers active during inference time can be seen as approximate Bayesian inference in probabilistic deep Gaussian processes, where each forward pass is interpreted as a single Monte Carlo sample from the posterior distribution, which could also be interpreted as an ensemble approach.

As a scalable alternative approach following a Bayesian perspective, Stochastic Weight Averaging Gaussian (SWAG) (Maddox et al., 2019) uses low-rank Gaussian distributions as approximations to the posterior over the model parameters. SWAG fits a Gaussian distribution to the first and second moments of recorded model weights obtained by the common regularization technique of Stochastic Weight Averaging (Izmailov et al., 2018) with a modified learning rate schedule, using a diagonal or low-rank covariance approximation in order to form a Gaussian approximation to the posterior distribution over the network weights. Sampling from this Gaussian distribution results in a well-calibrated approximate Bayesian inference scheme, while only adding little additional computational cost to standard network training.

Instead of combining the different outputs of multiple forward passes of a probabilistic network, deep ensembles (Lakshminarayanan et al., 2017) combine predictions of independently trained deterministic networks during inference time and aggregate multiple outputs over a collection of models. Although adding an important computational overhead in the training phase, they offer a simple, robust and scalable alternative to BNN’s for estimating predictive uncertainty while improving predictive performance and showing improved robustness to dataset shift.

Similarly to Ensemble methods, test-time augmentations do not consider multiple outputs from different forward passes through a probabilistic model but aim to use the spread of aggregated predictions from augmented versions of the input image passed through a single deterministic network as a proxy for uncertainty. However, its uncertainty estimation capabilities might be limited by the variety and validity of applied augmentations, especially in the field of remote sensing, where many imaging conditions are restricted by the sensor type and associated characteristics, and thus not all augmentations commonly used in Computer Vision might be valid and applicable.

Uncertainty is generally categorized into Epistemic (model) uncertainty, which can be reduced with more training data, and Aleatoric uncertainty, which is caused by noisy measurements. Sampling from posterior distributions with multiple forward passes during inference time yields a predictive distribution, where the average reflects the final network prediction and additional uncertainty measures describing the predictive uncertainty can be derived (Gal, 2016), such as the entropy of the predictive distribution, which is a measure of the predictive uncertainty (epistemic and aleatoric), or the mutual information between the predictive distribution and the posterior, which rather captures epistemic uncertainty .

**3.2.2 Model Comparison** Specific performance metrics, which assess how reliably a model’s confidence or uncertainty estimates reflect its predictive performance, enable a quantitative comparison of different uncertainty estimation methods in several disaster response pipelines.

One of the most common calibration measures, the Expected Calibration Error (ECE) (Guo et al., 2017), can be derived from visually intuitive reliability diagrams, which plot accuracy as a function of confidence. The model's pixel-wise predictions are partitioned into m=1..M equally-sized bins based on their confidence value, for which an accuracy score per bin is calculated. Deviations from the identity function indicate over- or under-confidence. The ECE summarizes the reliability diagram into a single score, taking the weighted average of differences between the accuracy and confidence per bin:

$$\mathrm{ECE} \;=\; \sum_{m=1}^{M} \frac{|B_m|}{N} \Big|\mathrm{acc}(B_m) \;-\; \mathrm{conf}(B_m)\Big|. \tag{1}$$

The Brier-Score (BS) (Brier, 1950), another established measure that assesses model calibration, is defined as the mean squared error (MSE) between the predicted probability vector and the one-hot-encoded ground-truth labels over all pixels for each class. An adapted version, the Stratified Brier-Score, accounts for class-imbalance by computing the BS per class and averaging over the class-wise scores, such that each class contributes equally to the final score.

Mukhoti et al. (Mukhoti and Gal, 2018) proposed the Patch Accuracy versus Patch Uncertainty (PAvPU) metric to directly evaluate the uncertainty maps of Bayesian Deep Learning models, considering an inverse relation between accuracy and uncertainty. Their proposed metrics and sub-metrics are based on the principle that confident model predictions should be accurate, whereas inaccurate predictions should come along with a high uncertainty.

Taking into account spatial context, these properties are expressed by two patch-level conditional probabilities. First,

$$\mathrm{p(uncertain} \mid \mathrm{inaccurate)} = \frac{n_{iu}}{(n_{ic} + n_{iu})}, \tag{2}$$

with the number of patches being inaccurate and uncertain $n_{iu}$ and the number of patches being inaccurate and certain $n_{ic}$, and second,

$$\mathrm{p(accurate} \mid \mathrm{certain)} = \frac{n_{ac}}{(n_{ac} + n_{ic})}, \tag{3}$$

with the number of patches being accurate and certain $n_{ac}$ and the number of patches being inaccurate and certain $n_{ic}$. While in our context, especially the conditional probability of $p(uncertain \mid inaccurate)$ is highly relevant, both conditional probabilities can be further combined into the global PAvPU metric:

$$\mathrm{PAvPU} = \frac{(n_{ac} + n_{iu})}{(n_{ac} + n_{au} + n_{ic} + n_{iu})}, \tag{4}$$

which additionally includes the number of patches that are accurate and uncertain $n_{au}$.

In preparation of our framework, we evaluated different baseline uncertainty estimation methods (BNN-VI - Bayes By Backprop, MCDO, deep ensembles, TTA) in a previous work (Hümmer, 2025) ona dataset of manually annotated optical VHR Pléiades imagery under severe domain shift. As expected, a first "in-domain" setting trained on globally distributed cities and evaluated on unseen patches from the same regions resulted in comparable scores for segmentation performance (IoU), model calibration (ECE, Brier), and uncertainty quantification (PAvPU) between the deterministic baseline model and uncertainty-aware approaches. Both Bayesian approximations BNN-VI and MCDO performed comparably, with no significant improvements over the deterministic baseline model, whereas TTA and the ensemble model yielded slightly improved performance over the deterministic baseline, unsurprisingly. In contrast, a second "out-of-domain" setting with severe geographical domain shift, evaluating patches from unseen regions, revealed clear improvements of all uncertainty-aware approaches over the deterministic baseline, especially when using a limited training subset, with ensemble models outperforming the other models in all settings. The change from the "in-domain" to "out-of-domain" setting was clearly reflected in the metrics, with calibration scores increasing from 2.8% - 8.2% ECE $\downarrow$ in the first setting to 21.9% - 30.0% ECE in the second setting or the conditional probabilities $p(\mathrm{U} \mid \mathrm{I}) \uparrow$ decreasing from $0.82 - 0.84$ to $0.64 - 0.75$ accordingly. In the "out-of-domain" setting, BNN-VI (24.8% ECE) yielded better calibrated models than MCDO (26.9% ECE), both improving the deterministic baseline (30.0%), but being outperformed by the model ensemble (21.9%). A similar ranking was observed for uncertainty quantification with $p(\mathrm{U} \mid \mathrm{I})$ scores improving from 0.64 (baseline), to 0.68 (MCDO) / 0.69 (BNN-VI) to 0.75 (ensemble).

### 3.3 3D Data processing

For 3D processing, PICANTEO currently integrates CARS (Youssefi et al., 2020), a photogrammetric pipeline designed to generate Digital Surface Models (DSM) from at least one stereo pair, as illustrated in Figure 3. Additional acquisitions can be incorporated to reduce occlusions in the scene (tri-stereo configuration). These images may be acquired by either a single agile satellite performing along-track maneuvers to observe the same area from different viewing angles (e.g., Pléiades, WorldView-3), or by satellite constellations capturing the scene simultaneously from multiple perspectives (e.g., CO3D). As highlighted in Section 2, PICANTEO is based on a modular architecture, allowing future integration of alternative 3D reconstruction approaches such as NeRF-based methods (Mildenhall et al., 2020) or inSAR.

CARS follows a classical photogrammetric workflow while being specifically designed to process large-scale satellite imagery. The 3D reconstruction relies on several steps: first, the stereo images are resampled into a common epipolar geometry, in which disparities are constrained to horizontal shifts. This transformation simplifies the matching step, where each pixel in the reference image is associated with its corresponding pixel in the secondary image. Due to the epipolar constraint, the search is reduced to a one-dimensional problem, significantly decreasing the computational cost. This matching process is performed using the Pandora stereo framework (Cournet et al., 2020) inspired by the modular taxonomy(Scharstein et al., 2001). The first step is to compute the cost volume that is the similarity or cost function $cv(x, y, d)$ of each pixel $(x, y)$ in one stereo image to each pixel within a subset $[d_{\min}, d_{\max}]$ in the other stereo image. PANDORA then uses this cost-volume in the optimization step with the well-known SGM algorithm (Hirschmuller, 2008). This cost volume also contains information called the ambiguity measure to quantify uncertainty during stereo matching (Sarrazin et al., 2021). From one pixel, the ambiguity curve is computed using the following formula:

$$\begin{aligned}\mathrm{Amb}(x, y, \eta) = Card(\{d \in [d_{\min}, d_{\max}] \mid \\ cv(x, y, d) < \min_{d} cv(x, y, d) + \eta\Big\})\end{aligned} \tag{5}$$

Ambiguity is an increasing function that grows more rapidly as other potential matches become as similar as the best match. In other words, the area under this curve increases as the match becomes more ambiguous. That is why to obtain a scalar value for the ambiguity, one simply needs to calculate the area under the curve:

$$\mathrm{Amb}_{\mathrm{Int}}(x, y) = \int \mathrm{Amb}(x, y, \eta)\, d\eta \tag{6}$$

If we consider the extreme cases: all points have the same cost, so the integral of ambiguity is 1. If a single point has a cost of zero and all the others have the maximum cost, this integral is 0.

This ambiguity information can be propagated through the CARS pipeline and exploited in downstream processing, particularly for filtering purposes, as described in Section 3.5. Once the disparity is estimated in epipolar geometry, it is converted into geographic coordinates (longitude, latitude, altitude) using the sensor geometric models. This triangulation step produces a geo-referenced point cloud. To enable efficient analysis within PICANTEO, and in particular pre- and post-event comparisons, the final step of the CARS pipeline rasterizes this point cloud into a raster DSM.

In PICANTEO, 3D change detection is not necessarily performed through a direct difference between pre- and post-event DSMs. Instead, normalized DSMs (nDSM) are computed for each acquisition and compared. This approach improves robustness to topographic changes in the scene, preventing misinterpretation of phenomena such as landslides as building damage, and reducing artifacts commonly observed in raw DSM differencing, such as uncorrected satellite vibrations.

To derive nDSMs from DSMs, we use the Bulldozer tool (Lallement et al., 2023), a morphological method designed to estimate Digital Terrain Models (DTM) without relying on external data such as water masks or land cover maps (Figure 3). Similarly to CARS, Bulldozer is scalable and well-suited for large satellite images. The method simulates a cloth draped over an inverted DSM that has been pre-processed to remove outliers. The cloth fits the large-scale topography while ignoring elevated structures such as buildings and vegetation. After this simulation, post-processing steps are applied to smooth residual artifacts. The resulting DTM captures the underlying terrain, including low-frequency vibrations that may also reflect acquisition-related distortions.

Once DTMs are obtained for both pre- and post-event datasets, the corresponding nDSMs are computed and subtracted to produce a raw 3D change detection map.

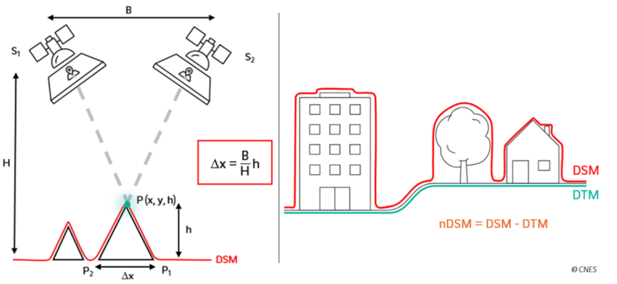


Figure 3. Left: Elevation measurement of a point using photogrammetry. Right: Comparison between Digital Surface Model (DSM) and Digital Terrain Model (DTM).

### 3.4 Flood segmentation

In addition to a comprehensive multi-modal building damage assessment at object level, which emphasizes model generalization across diverse natural disaster types, our framework further integrates a full thematic event-specific flood segmentation module. The flood segmentation module combines flood masks derived from multiple sources, leveraging existing open-source frameworks and operational services, with optional user-specified custom segmentation models for high precision flood delineation and reliable base masks for informed decision-making in flood-related disaster management. Specifically, a user-specific segmentation model can utilize the framework's 2D-pipeline to generate a flood and permanent water segmentation mask. A data fusion step merges the derived map with aggregated thematic layers from three main sources: A multiclass flood segmentation layer predicted from the CNES FloodML model, the EU's Copernicus Emergency Management Service Global Flood Monitoring (GFM) (Salamon et al., 2021) product within the Global Flood Awareness System (GloFAS)*, as well as the DLR's Surface Water Inventory and Monitoring (SWIM) (Groth et al., 2025) water masks hosted on the DLR EOC Geoservice*. All algorithms and auxiliary products are based on the integration of synthetic aperture radar data, allowing for reliable flood mapping regardless of weather, atmospheric conditions, and cloud cover, which is especially important for the specific application in disaster monitoring for flood events.

Launched by the Space for Climate Observatory (SCO) FloodDAM and FloodDAM-DT projects, FloodML * is designed for flood detection and monitoring. Using a Machine Learning (ML) approach with Earth Observation (EO) satellite data, FloodML produces estimated flood extent maps as well as permanent water maps.

The algorithm can utilize various sensors including Sentinel-1 (VV/VH), Sentinel-2, TerraSAR-X, Landsat 8, Landsat 9 and incorporates Digital Elevation Models and auxiliary land cover information. Additional EO products can be integrated to improve spatial and temporal coverage and to enhance the detection of floodwater beneath vegetation.

The ML algorithm is based on a Random Forest and trained on flood events covered by the Copernicus Emergency Management Service (CEMS) before 2020, selected in diverse regions, topography, and climatic conditions. Global Surface Water Occurrence (GSOW) (Pekel et al., 2016), a surface water world map

* https://global-flood.emergency.copernicus.eu/
* https://geoservice.dlr.de/web/datasets/swim
* https://github.com/CNES/floodml

based on 30 years of Landsat data, was used to train permanent water detection. In addition, Multi-Error-Removed Improved-Terrain (MERIT) DEM model (Yamazaki et al., 2017), and Height Above Nearest Drainage (HAND) (Nobre et al., 2011) are used for training. The Sentinel-2 model can be directly applied to Landsat 8 and 9 images, as they share similar spectral bands. The Sentinel-1 model has been adapted to enable inference on TerraSAR-X data, despite differences in frequency bands and calibration.

FloodML represents a static ad-hoc on-demand solution for specific regions of interest, whereas the additional thematic products of GFM and SWIM extend the mapping capabilities by near real-time continuous monitoring of floods worldwide in fully automated operational pipelines. GFM aims at providing near real-time robust flood and water extent maps by following an Ensemble approach to maximize the classification quality, fusing the outputs of three algorithms, which process VV-polarized Sentinel-1 SAR data in Interferometric Wide-swath mode as detected, multi-looked, and projected GRD products. The final ensemble consensus map fuses the outputs of the three submodules from (Chini et al., 2017), (Martinis et al., 2015) and (Bauer-Marschallinger et al., 2022), accompanied by a likelihood map indicator, qualifying its reliability.

The SWIM Water Extent (Groth et al., 2025) service offers an additional global surface water product based on convolutional neural networks (CNN) (Bereczky et al., 2022), regularly publishing large-scale continuous flood monitoring assets through OGC-Webservices on the DLR EOC Geoservice platform. Its dynamically updated water extent masks highlight open surface water bodies for specific dates and locations, whereas reference water masks with permanent and seasonal water bodies derived from previous observation periods enable the analysis of temporal water dynamics for water monitoring and rapid flood mapping in disaster response scenarios. All assets can be accessed via spatio-temporal asset catalogs (STAC), which allows for an easy and seamless integration into different pipelines and the visualization dashboard of our framework, using a simple and re-usable STAC-query step given a region of interest.

### 3.5 Data filtering

PICANTEO integrates a set of basic steps for filtering the outputs of data extraction steps. In the previous sections, we described how raw change maps can be generated. However, these maps may contain artifacts. For example, in 2D change detection, slight misalignments between pre- and post-event building detection masks can result in building edges being falsely detected as changes.

To mitigate these effects and refine the detection maps, several filtering approaches are implemented in PICANTEO. These include simple threshold-based filters, where pixel values are compared to a fixed value. This is the case in 3D change detection, where only elevations below a given threshold (in meters) are retained. Alternatively, pixel values can be compared with those from another layer, for example, to remove uncertain detections in 2D building extraction using an uncertainty map. In addition, morphological filters are available to remove small residual artifacts.

Figure 4 illustrates an example of filtering a raw 3D change detection map to extract destroyed buildings. First, a threshold-based filter on elevation values is applied to retain only significant changes, excluding minor variations that may result from elevation uncertainties in the 3D models. Next, a filter based on an auxiliary vegetation layer (NDVI) is used to remove pixels corresponding to vegetation. Similarly, the ambiguity measure introduced in Section 3.3 is used to retain only areas where the pre- and post-event Digital Surface Models are considered reliable. Finally, a morphological filter is applied to remove all changes with an area below a predefined threshold.

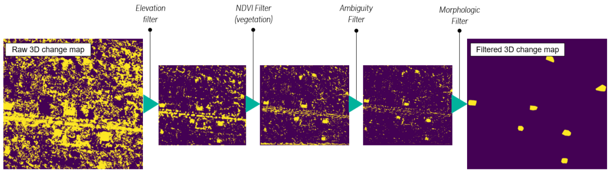


Figure 4. Example of 3D change map refinement using elevation, NDVI, ambiguity and morphologic filters.

## 4. Applications

### 4.1 Optical 2D Change Detection - Mayotte

Tropical Cyclone Chido struck the island of Mayotte on 14 December 2024, causing massive damage to buildings and infrastructure. In the following days, the first very high-resolution satellite images were acquired to locate and estimate the damages.

Since only 2D imagery was available, damage detection was performed using the Pléiades model described in Section 3.2 with the framework’s 2D pipeline. This model was previously applied in (Hümmer et al., 2024) to a different type of natural disaster in another region: a wildfire case in Chile. Figure 5 illustrates PICANTEO’s ability to process large areas efficiently. The framework’s output, generated in a few minutes, is compared with the manual annotations produced by the Copernicus Emergency Management Service (CEMS) during the rapid mapping phase. We can observe in Figure 6 that in several locations, PICANTEO detected destroyed buildings that were not labeled by CEMS, likely due to the operational constraints in delivering maps under time pressure or outdated pre-event imagery.

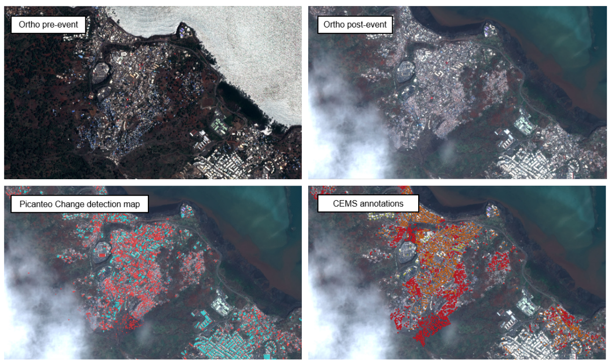


Figure 5. Building destruction on Mayotte island in France. Top left: Pre-event orthoimage; Top right: Post-event orthoimage; Bottom left: PICANTEO bi-temporal 2D detection pipeline output; Bottom right: Ground truth from Copernicus Emergency Management Services. Pléiades ©CNES 2024/2025, Distribution CNES PWH.

This example highlights PICANTEO’s capability to operate in rapid mapping scenarios, where input data are degraded, for

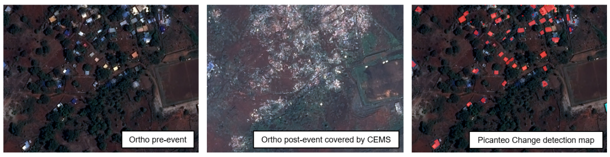


Figure 6. Missing destroyed building by Copernicus Emergency Management Services. Left: Pre-event orthoimage; Middle: Post-event orthoimage; Right: PICANTEO bi-temporal 2D detection pipeline output. Pléiades ©CNES 2024/2025, Distribution CNES PWH.

instance due to cloud cover. It also demonstrates the advantage of automated methods for quick and exhaustive damage assessment over large areas.

### 4.2 Optical 2D/3D Change Detection - Kahramanmaraş

On 9 February 2023, a series of earthquakes with a magnitude of 7.8 struck Turkey and Syria, causing widespread destruction and significant loss of life in numerous cities. To evaluate the performance of PICANTEO for building destruction detection during the recovery phase, a stereoscopic acquisition was carried out on 15 March 2023 over the city of Kahramanmaraş in Turkey.

In our previous work (Hümmer et al., 2024), we demonstrated the benefit of combining 2D and 3D change detection, resulting in an approximately 7% increase in the number of destroyed buildings detected compared to using a purely 2D or purely 3D approach, leading to an overall detection rate of 95.2% (494/519). For this evaluation, we asked the Icube-SERTIT annotation experts to produce building annotations, which were then used as ground truth and compared to our predictions, as illustrated in Figure 7. For the 2D part, the integration of uncertainty-based filtering helped to improve noise reduction in the intermediate change maps while only slightly lowering the recall. The final 2D change detection map (binary classification of destroyed buildings vs background) improved from $IoU - 0.55$ to $IoU - 0.63$ and $F1 - 0.71$ to $F1 - 0.78$. In the final framework, the uncertainty filter should be adjustable by the end-user, as suitable filter thresholds may vary depending on the scene and the degree of domain shift.

We used a bi-temporal 2D/3D uncertainty-aware very high-resolution optical change detection pipeline, using the Pléiades model for 2D change detection described in Section 3.2, combined with the 3D change detection pipeline presented in Section 3.3. This case study highlights PICANTEO's ability to handle multi-modal data and to fuse change detection outputs in order to provide end users with the most accurate and complete damage assessment maps possible.

### 4.3 Multi-modal Building Damage Assessment - Acapulco

In October 2023, category 5 hurricane Otis hit the Pacific Coast of Mexico and made landfall near the city of Acapulco, causing severe damage to buildings, infrastructure, and surrounding forests. (Chen et al., 2025) and (Wang et al., 2025) included the Acapulco hurricane event in their multi-modal building damage assessment datasets BRIGHT and DisasterM3, providing fused multi-source building damage annotations with optical pre-event imagery and VHR SAR post-event acquisitions. This application highlights the easy and seamless integration of multi-modal data sources, combining very high-resolution optical and synthetic aperture radar imagery for all-weather disaster response.

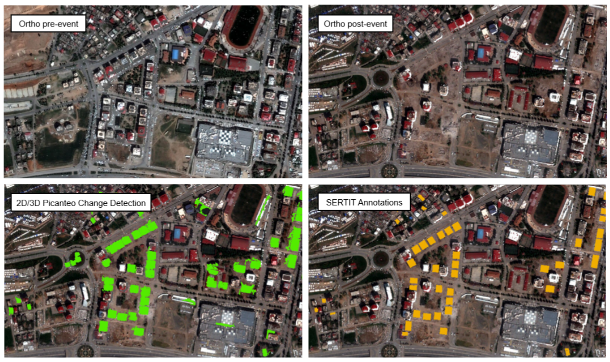


Figure 7. Building destruction on Kahramanmaraş city in Turkey. Top left: Pre-event orthoimage; Top right: Post-event orthoimage; Bottom left: PICANTEO bi-temporal 2D/3D change detection pipeline output; Bottom right: Ground truth from Icube-SERTIT. Pléiades ©CNES 2016/2023, Distribution CNES PWH.

For the following example, we trained uncertainty-aware optical-SAR baseline UNet models on the BRIGHT dataset with a two-fold training schedule. First, a pre-training phase follows the BRIGHT zero-shot cross-event transfer setup not including the Acapulco area. A second fine-tuning phase adapts the pre-trained model weights, only using a few annotated tiles of the new target scene as training samples (few-shot) instead of re-training the model on the whole dataset. This approach emulates a more realistic rapid-mapping workflow in which timeliness is critical, annotation experts can quickly integrate a modest number of labeled tiles through active learning, enabling the pre-trained model to be fine-tuned on a limited subset rather than requiring full re-training on the entire dataset. Uncertainty-aware models can help to minimize the impact of severe domain shifts, guide active learning processes, or filter change detection maps, while providing qualified predictions.

The fine-tuning step not only results in performance gains in terms of building damage assessment ($mIoU$ 0.27→0.43), but also shows improved model calibration and uncertainty estimation, with the Expected Calibration Error improving from $ECE - 16.8\%$ to $ECE - 2.7\%$ and the predictive entropy reflecting more reliable uncertainty estimates with less overconfidence. For the multi-modal damage assessment use-case in the transfer learning Acapulco setting, a baseline SWAG model slightly outperformed MCDO. An example prediction of the fine-tuned SWAG model, based on the lightning UQ-toolbox wrapper (Lehmann et al., 2025), is illustrated in figure 8, where misclassified predictions occur mostly in areas with high uncertainty. Even after fine-tuning, the Acapulco event remains a very challenging application case due to its complex scene characteristics, acquisition conditions, and damage signature. In addition, a comparison of the ground truth annotations with the optical Pléiades post-event images revealed strong inherent label noise/ambiguities, which further hinders successful model transferability.

As an additional experiment, we evaluated an optical-only building damage assessment network on the Acapulco study area. The model is based on a shared DINOv2 (Oquab et al., 2024) encoder and a lightweight feature pyramid to extract hierarchical features from bi-temporal imagery. A localization branch predicts building footprints from the pre-event image, while a damage assessment branch combines multi-scale features from the pre-event and post-event images to infer pixel-wise damage classes. The two tasks share the encoder but rely on separ-

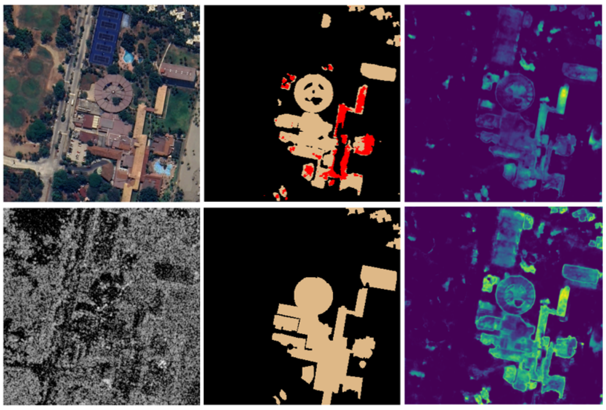

Figure 8. Acapulco example prediction with a fine-tuned SWAG model. First column: Pre- (upper) and post-event (lower) image. Second column: Prediction (upper) and ground truth (lower). Third column: Mutual information (upper) and predictive entropy (lower). Pre-/post event images and annotations from DisasterM3.

ate DPT-based decoders (Ranftl et al., 2021) and task-specific heads, enabling efficient joint learning of localization and damage assessment. Figure 9 shows example predictions before data filtering.

Figure 9. Building damage assessment on Acapulco. Top: Pre-event and post-event images; Bottom: Predictions (red destroyed, orange damaged). Pléiades ©CNES 2023, Distribution CNES PWH.

### 4.4 Flood Detection - Marmande

This example of storm Nils, which hit the Bordeaux region of France in early 2026, causing severe floodings, integrates the previously described FloodML pipeline alongside auxiliary continuous monitoring layers from GFM and SWIM. The Copernicus Emergency Management Service (CEMS) has been activated and their rapid flood mapping products* are shown alongside the pipeline results and auxiliary layers.

Figure 10 shows that the three layers are generally well aligned with the semiautomatically generated CEMS product, although each shows a few misclassifications. The integration of multiple flood mapping sources allows for a better qualification of water masks and the ad-hoc Flood-ML method could potentially integrate water masks derived from additional input sources (e.g. VHR SAR TerraSAR-X) if available, thus extending the coverage of available continuous monitoring layers (GFM/SWIM).

* https://mapping.emergency.copernicus.eu/activations/EMSR866/

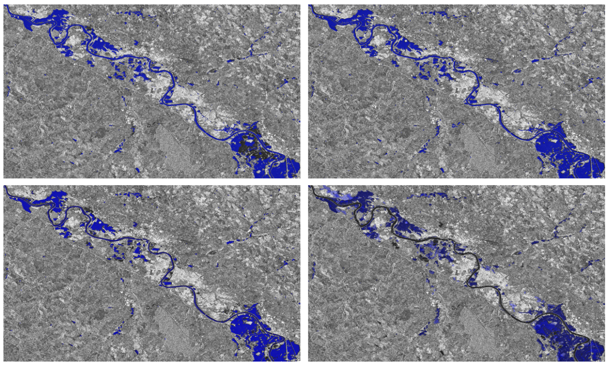

Figure 10. Visual comparison of different flood masks for the Flooding event in Marmande, France, in February 2026. Top left: SWIM water extent; Top right: FloodML; Bottom left: GFM; Bottom right: Semi-automatic CEMS mask based on Sentinel-1 acquisition of the same day. Contains modified Copernicus Sentinel data 2026. Background picture: Sentinel-1 IW-GRD.

## 5. Conclusion and perspectives

This work presents the PICANTEO framework, a unified open-source platform for change detection, with a particular focus on natural disaster response. It features a collection of pre-configured pipelines tailored to various disaster scenarios, with the possibility to build custom pipelines leveraging reusable and extensible processing steps, such as multi-modal data fusion, uncertainty-aware change indicators, and an interactive visualization dashboard.

Section 4 highlights PICANTEO's ability to address a wide range of disaster types across diverse geographic regions and temporal scales, from rapid mapping to the recovery phase. It presents several pipelines including very high-resolution optical 2D change detection following a cyclone, earthquake-related optical 2D/3D change detection, optical-SAR-based hurricane damage assessment, as well as SAR-based flood mapping. It provides templates for future user applications and illustrates PICANTEO's ability to process multi-modal data through dedicated pipelines available within the framework. We will integrate more pipelines to support a broader range of multi-resolution and multi-modal data from heterogeneous sensors.

PICANTEO is still in development, and we plan to train more robust deep learning models to handle challenging acquisition conditions, including cloud-covered imagery and high off-nadir views. We also aim to strengthen generalization across diverse geographic regions. Future works will further focus on the implementation of more advanced uncertainty estimation techniques to enhance the reliability of result qualification and the detection of out-of-distribution samples.

The framework enables the integration of new applications, actively encouraging contributions from end users to develop new pipelines or extend existing functionalities and processing steps, as described in Section 2.1. In parallel, the authors collaborate with domain experts in disaster management to optimize existing pipelines for operational deployment and integrate new use cases.

**References**

Bauer-Marschallinger, B., Cao, S., Tupas, M. E., Roth, F., Navacchi, C., Melzer, T., Freeman, V., Wagner, W., 2022. Satellite-Based Flood Mapping through Bayesian Inference from a Sentinel-1 SAR Datacube. *Remote Sensing*, 14(15).

Bereczky, M., Wieland, M., Krullikowski, C., Martinis, S., Plank, S., 2022. Sentinel-1-Based Water and Flood Mapping: Benchmarking Convolutional Neural Networks Against an Operational Rule-Based Processing Chain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2023-2036.

Bessis, J.-L., Béquignon, J., Mahmood, A., 2004. The International Charter “Space and Major Disasters” initiative. *Acta Astronautica*, 54(3), 183-190.

Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D., 2015. Weight uncertainty in neural network. F. Bach, D. Blei (eds), *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research, 37, PMLR, Lille, France, 1613–1622.

Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1-3. https://api.semanticscholar.org/CorpusID:122906757.

Caye Daudt, R., Le Saux, B., Boulch, A., 2018. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, 4063–4067.

Chen, H., Song, J., Dietrich, O., Broni-Bediako, C., Xuan, W., Wang, J., Shao, X., Wei, Y., Xia, J., Lan, C., Schindler, K., Yokoya, N., 2025. BRIGHT: a globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response. *Earth System Science Data*, 17(11), 6217–6253.

Chini, M., Hostache, R., Giustarini, L., Matgen, P., 2017. A Hierarchical Split-Based Approach for Parametric Thresholding of SAR Images: Flood Inundation as a Test Case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 6975-6988.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., Ermon, S., 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35, 197–211.

Cournet, M., Sarrazin, E., Dumas, L., Michel, J., Guinet, J., Youssefi, D., Defonte, V., Fardet, Q., 2020. Ground truth generation and disparity estimation for optical satellite imagery. *ISPRS Archives*, XLIII-B2-2020, 127–134.

Durnov, V., 2020. xView2: 1st place solution. *GitHub*. https://github.com/vdurnov/xview2_1st_place_solution.

European Commission, 2015. Copernicus emergency management service mapping user guide. Available online: https://emergency.copernicus.eu.

Fan, T., Wang, G., Li, Y., Wang, H., 2020. MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. *IEEE Access*, 8, 179656-179665.

Gal, Y., 2016. Uncertainty in Deep Learning. PhD thesis, University of Cambridge.

Gal, Y., Ghahramani, Z., 2015. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*.

Gawlikowski, J., Tassi, C., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., Zhu, X., 2023. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56, 1-77.

Graves, A., 2011. Practical variational inference for neural networks. *Neural Information Processing Systems*.

Groth, S., Wieland, M., Martinis, S., 2025. Surface water inventory and monitoring (swim): Hands-on examples for improved flood mapping and water resource monitoring. *ESA Living Planet Symposium 2025*.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K. Q., 2017. On calibration of modern neural networks. D. Precup, Y. W. Teh (eds), *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, 70, PMLR, 1321–1330.

Hafner, S., Gerard, S., Sullivan, J., Ban, Y., 2025. DisasterAdaptiveNet: A robust network for multi-hazard building damage detection from very-high-resolution satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 143, 104756.

Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 2021. LoRA: Low-Rank Adaptation of Large Language Models. https://arxiv.org/abs/2106.09685.

Hümmer, C., 2025. Qualification and evaluation of uncertainty estimation concepts for automatic change detection. Proceedings of the International Conference on Advanced Remote Sensing (ICARS 2025), MDPI: Basel, Switzerland,, Barcelona.

Hümmer, C., Lallement, D., Youssefi, D., 2024. Uncertainty-aware 2d/3d change detection for natural disaster response. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2024, Athens, Greece, July 7-12, 2024*, IEEE.

Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D. P., Wilson, A. G., 2018. Averaging weights leads to wider optima and better generalization. *Conference on Uncertainty in Artificial Intelligence*.

Kaku, K. et al., 2013. Sentinel Asia: A space-based disaster management support system in the Asia-Pacific region. *International Journal of Disaster Risk Reduction*, 6, 1–10.

Kingma, D. P., Salimans, T., Welling, M., 2015. Variational dropout and the local reparameterization trick. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 28, Curran Associates, Inc.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 30, Curran Associates, Inc.

Lallement, D., Lassalle, P., Ott, Y., 2023. Bulldozer, a free open source scalable software for DTM extraction. *ISPRS Archives*, XLVIII-4/W7-2023, 89–94.

Lebègue, L., Cazala-Hourcade, E., Languille, F., Artigues, S., Melet, O., 2020. CO3D, a worldwide one-meter accuracy DEM for 2025. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* http://dx.doi.org/10.5194/isprs-archives-XLIII-B1-2020-299-2020.

Lehmann, N., Gottschling, N. M., Gawlikowski, J., Stewart, A. J., Depeweg, S., Nalisnick, E., 2025. Lightning UQ Box: Uncertainty Quantification for Neural Networks. *Journal of Machine Learning Research*, 26(54), 1–7. http://jmlr.org/papers/v26/24-2110.html.

Maddox, W. J., Garipov, T., Izmailov, P., Vetrov, D. P., Wilson, A. G., 2019. A simple baseline for bayesian uncertainty in deep learning. *Neural Information Processing Systems*.

Martinis, S., Kersten, J., Twele, A., 2015. A fully automated TerraSAR-X based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 203-212.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. *European Conference on Computer Vision (ECCV)*, Springer, 405–421.

Mukhoti, J., Gal, Y., 2018. Evaluating Bayesian Deep Learning Methods for Semantic Segmentation. *ArXiv*, abs/1811.12709. https://api.semanticscholar.org/CorpusID:54208798.

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Medeiros, G., Silveira, A., Waterloo, M., Saleska, S., 2011. Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *Journal of Hydrology*, 404, 13-29.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2024. DINOv2: Learning Robust Visual Features without Supervision. https://arxiv.org/abs/2304.07193.

Pekel, J., Cottam, A., Gorelick, N., Belward, 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540, 418–422.

Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision transformers for dense prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12159–12168.

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., Feichtenhofer, C., 2024. SAM 2: Segment Anything in Images and Videos. https://arxiv.org/abs/2408.00714.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Salamon, P., McCormick, N., Reimer, C., Clarke, T., Bauer-Marschallinger, B., Wagner, W., Martinis, S., Chow, C., Böhnke, C., Matgen, P., Chini, M., Hostache, R., Molini, L., Fiori, E., Walli, A., 2021. The new, systematic global flood monitoring product of the copernicus emergency management service. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1053–1056.

Sarrazin, E., Cournet, M., Dumas, L., Defonte, V., Fardet, Q., Steux, Y., Jimenez Diaz, N., Dubois, E., Youssefi, D., Buffe, F., 2021. AMBIGUITY CONCEPT IN STEREO MATCHING PIPELINE. *ISPRS Archives*, XLIII-B2-2021, 383–390.

Scharstein, D., Szeliski, R., Zabih, R., 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, 131–140.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P., 2025. DINOv3. https://arxiv.org/abs/2508.10104.

Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, E., Blumenstiel, B., Ghosal, R., de Oliveira, P. H., de Sousa Almeida, J. L., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C., Kumar, A., Gaur, V., Truong, M., Godwin, D., Khallaghi, S., Lee, H., Hsu, C.-Y., Asanjan, A. A., Mujeci, B., Shidham, D., Balogun, R. O., Kolluru, V., Keenan, T., Arevalo, P., Li, W., Alemohammad, H., Olofsson, P., Mayer, T., Hain, C., Kennedy, R., Zadrozny, B., Bell, D., Cavallaro, G., Watson, C., Maskey, M., Ramachandran, R., Moreno, J. B., 2026. Prithvi-EO-2.0: A Versatile Multitemporal Foundation Model for Earth Observation Applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64, 1-20.

Wang, J., Xuan, W., Qi, H., Liu, Z., Liu, K., Wu, Y., Chen, H., Song, J., Xia, J., Zheng, Z., Yokoya, N., 2025. DisasterM3: A Remote Sensing Vision-Language Dataset for Disaster Damage Assessment and Response. https://arxiv.org/abs/2505.21089.

Xiong, X., Wu, Z., Tan, S., Li, W., Tang, F., Chen, Y., Li, S., Ma, J., Li, G., 2026. SAM2-UNet: segment anything 2 makes strong encoder for natural and medical image segmentation. *Visual Intelligence*, 4(1). http://dx.doi.org/10.1007/s44267-025-00106-w.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., Bates, P. D., 2017. MERIT DEM: A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, 44(11), 5844-5853.

Youssefi, D., Michel, J., Sarrazin, E., Buffe, F., Cournet, M., Delvit, J.-M., L'Helguen, C., Melet, O., Emilien, A., Bosman, J., 2020. Cars: A photogrammetry pipeline using dask graphs to construct a global 3d model. *IGARSS 2020*, 453–456.

Zhu, D., Huang, X., Huang, H., Cheng, Q., Huang, Z., Shao, Z., 2026. ChangeViT: Unleashing plain vision transformers for change detection in remote sensing images. *Pattern Recognition*, 172, 112539.