

Estimating Ensemble Likelihoods for the Sentinel-1-Based Global Flood Monitoring Product of the Copernicus Emergency Management Service

Christian Krullikowski¹, Candace Chow², Marc Wieland³, Sandro Martinis⁴,
Bernhard Bauer-Marschallinger⁵, Florian Roth⁶, Patrick Matgen, Marco Chini⁷, *Senior Member, IEEE*,
Renaud Hostache⁸, Yu Li, and Peter Salamon

Abstract—The Global Flood Monitoring (GFM) system of the Copernicus Emergency Management Service addresses the challenges and impacts that are caused by flooding. The GFM system provides global, near-real-time flood extent masks for each newly acquired Sentinel-1 Interferometric wide swath synthetic aperture radar (SAR) image, as well as flood information from the whole Sentinel-1 archive from 2015 on. The GFM flood extent is an ensemble product based on a combination of three independently developed flood mapping algorithms that individually derive the flood information from Sentinel-1 data. Each flood algorithm also provides classification uncertainty information that is aggregated into the GFM ensemble likelihood product as the mean of the individual classification likelihoods. As the flood detection algorithms derive uncertainty information with different methods, the value range of the three input likelihoods must be harmonized to a range from low [0] to high [100] flood likelihood. The ensemble likelihood is evaluated on two test sites in Myanmar and Somalia, showcasing the performance during an actual flood event and an area with challenging conditions for SAR-based flood detection. The Myanmar use case demonstrates the robustness if flood detections in the ensemble step disagree and how that information is communicated to the end-user. The Somalia use case demonstrates a setting where

misclassifications are likely, how the ensemble process mitigates false detections and how the flood likelihoods can be interpreted to use such results with adequate caution.

Index Terms—Copernicus Emergency Management Service (CEMS), Earth observation, ensemble classification, flood monitoring, likelihoods, radar, uncertainties, sentinel-1.

I. INTRODUCTION

WITH an amount of 44 % [1] of all occurring disasters and produced economic losses of about 651 billion dollars, floods are among the most severe disasters worldwide. Although not the deadliest natural disaster, floods are affecting the largest number of people worldwide every year. With globally rising temperatures, Dottori et al. [2] predict an increase in human losses due to flooding by up to 70% to 83% and additional direct flood damages upward of 160% to 240%. Botzen et al. [3] identified population and economic growth in disaster-prone regions as key causes leading to this increase. Apart from human losses, floods may cause damages to (critical) infrastructure [4] and may lead to further cascading effects, such as the spread of infectious diseases [5].

Mitigating these effects requires coordinated action on multiple levels, including but not limited to the implementation of accurate early warning systems, constant monitoring of disaster-prone regions, and well-implemented risk management procedures [6]. Arguably, the monitoring requirement, especially at large scale, is currently best fulfilled through the utilization of Earth observation data. Grimaldi et al. [7] presented a review of different flood data sources and compared optical with synthetic aperture radar (SAR) sensors. Optical imagery mainly relies on cloud-free and illuminated data, whereas radar remote sensing satellites can operate day and night due to their ability to emit cloud penetrating microwaves.

Past studies already highlighted the potential of a synergetic use of optical and SAR data in flood mapping [8] and [9]. However, most studies focus on a single technology, most frequently microwave remote sensing [10], [11], [12], [13]. A comprehensive overview of advantages and limitations of different methods is found in [14]. The fusion of Sentinel-1 and Sentinel-2 data to

Manuscript received 30 March 2023; revised 23 May 2023; accepted 29 June 2023. Date of publication 18 July 2023; date of current version 2 August 2023. GFM is product and service of the Copernicus Emergency Management Service. The Joint Research Centre of the European Commission is the entrusted entity responsible for GFM in terms of management, technical implementation, and evolution. The provision of an Automated, Global, Satellite-based Flood Monitoring Product for the Copernicus Emergency Management Service (GFM) is supported under the Framework Contract number 939866-IPR-2020. (*Corresponding author: Christian Krullikowski.*)

Christian Krullikowski, Candace Chow, Marc Wieland, and Sandro Martinis are with the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), D-82234 Oberpfaffenhofen, Germany (e-mail: christian.krullikowski@dlr.de; candace.chow@dlr.de; marc.wieland@dlr.de; sandro.martinis@dlr.de).

Bernhard Bauer-Marschallinger and Florian Roth are with the Remote Sensing Research Group, Department of Geodesy and Geoinformation, TU Wien, 1040 Vienna, Austria (e-mail: bbm@geo.tuwien.ac.at; florian.roth@geo.tuwien.ac.at).

Patrick Matgen, Marco Chini, Renaud Hostache, and Yu Li are with the Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology, 4362 Esch-sur-Alzette, Luxembourg (e-mail: patrick.matgen@list.lu; marco.chini@list.lu; renaud.hostache@list.lu; yu.li@list.lu).

Peter Salamon is with the Joint Research Centre (JRC) of the European Commission, 21027 Ispra, Italy (e-mail: peter.salamon@ec.europa.eu).

Digital Object Identifier 10.1109/JSTARS.2023.3292350

derive reference water masks is described in [15]. The authors further elaborate on the different observation principles with the result of deviating water masks. This information seems useful for the utilization of validation data.

It can be noted that the aforementioned studies either focus on specific regions or were not implemented as operational services. Furthermore, a majority of the studies do not provide any information on the flood classification uncertainties. Clement et al. [10] highlighted several sources of uncertainty affecting SAR-based flood extent mapping, for example the ambiguities related to similar backscatter return over water look-alikes and dry soil, as well as areas with fuzzy backscatter response, e.g., dense vegetation, and areas with higher backscatter return over urban areas. In general, cases where SAR-based flood mapping may be hampered and the detection may become less confident, the classification necessitates and benefits from the inclusion of a dedicated uncertainty analysis [16]. This complementary output also supports the interpretation and use of SAR-based flood map products, where end-users can be alerted to flood features associated with lower confidence, which should be treated with more caution with respect to risk assessment.

The absence of a fully operational flood service that also returns confidence information culminated in the request of the Copernicus Emergency Management Service (CEMS) to integrate technically mature and scientifically validated flood detection algorithms into the Global Flood Awareness System (GloFAS)¹ and the European Flood Awareness System (EFAS).² Instead of utilizing an approach based on a single retrieval algorithm, the Joint Research Centre (JRC) as the contracting authority adopted an ensemble approach, which merges the results of three matured and independently developed flood algorithms.

A. Conceptual Basis of GFM Ensemble

The Global Flood Monitoring (GFM) product of the CEMS continuously processes and analyzes all incoming Sentinel-1 ground range detected (GRD) interferometric wide swath (IW) data, aiming to detect and monitor flood events in near real-time at global scale.

The GFM product builds on an ensemble approach that combines three mature and independently developed flood detection algorithms provided by the German Aerospace Center (DLR), Luxembourg Institute of Science and Technology (LIST) and Vienna University of Technology (TUW). The flood ensemble is computed pixelwise and based on a majority voting system, where at least two algorithms must classify a pixel as flooded or non-flooded. Further insight into the flood ensemble algorithm is described by [17] and the GFM product definition document [18]. Hostache et al. [19] presented a performance analysis of each of the individual flood detection algorithms and shows similar accuracies over three test cases that were investigated in the GFM project.

Besides a pixel-based flood classification, each flood detection algorithm generates classification uncertainty information

in the form of likelihoods. The GFM ensemble algorithm then combines the three individual layers of uncertainty information into a single layer termed ensemble likelihood. Although, an uncertainty analysis is performed, the term likelihood is used instead of uncertainty; most users have an a priori understanding about likelihoods, whereas uncertainties describe a negation which may not be understood as intuitively.

The GFM product is composed of different layers supporting the interpretation of flood situations using remote sensing data. Besides the actual flood extent layer, users can also download the likelihood data from GloFAS and EFAS. In addition, a downloadable exclusion layer informs about regions where no flood delineation was possible. These areas correspond to nodata values in both of the flood and likelihood products.

B. Objectives of Ensemble Flood Detection and Interpretation of Ensemble Likelihoods

Two sets of objectives drive the ensemble-based flood detection and the manner in which ensemble likelihoods are intended to be interpreted and applied by two user communities: 1) integrating results into further processes or studies and 2) utilizing results for decision-making processes.

The first objective addresses the first community, consisting of algorithm developers and scientists. Ensemble likelihoods may be used by them to identify subsets of pixels associated with low confidence values as a way to gain insights about opportunities for improving algorithms so that they return more accurate predictions. The individual and combined likelihoods may also serve as a basis for intercomparing the results obtained by different approaches, thereby potentially improving our understanding of their strengths and weaknesses.

The second community are data (end-)users. They may use ensemble likelihoods to minimize adverse consequences of making potentially costly decisions based on highly uncertain information. Results of this study may thus provide a basis on how the two aforementioned use cases can be used to support decision-making for flood and non-flood events.

In particular, to evaluate the impact of differential decisions made based on flood classifications with and without consideration of ensemble likelihood values, we examine two land cover types (i.e., agricultural lands, built environments) of particularly high economic importance and social consequences. In effect, mean likelihoods serve as a heuristic indication of overall confidence in the flood prediction, based on an average of available flood and respective uncertainty outputs from contributing individual flood algorithms. The values also function as an indicator of the current capacities/confidence of ensemble algorithms to detect water over certain types of land covers and uses.

This study focuses on gaining insights on scenarios that result in regular and overdetections with respect to dominant land covers. Based on the results, benefits, and limitations of the ensemble likelihood approach are highlighted and provide a starting point to guide further developments and applications in the two communities.

The application of ensemble likelihoods is evaluated with two use cases exemplifying flood (Myanmar use case) and

¹[Online]. Available: <https://www.globalfloods.eu/>

²[Online]. Available: <https://www.efas.eu/>

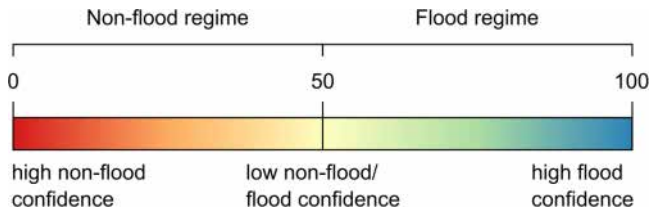


Fig. 1. Confidence distribution of likelihood values. Likelihood values toward 0 correspond to higher confidence in non-flood classifications, whereas values toward 100 correspond to higher confidence in flood classifications. Low confidence in both classifications is indicated by likelihood values toward 50.

non-flood (Somalia use case) events, respectively. The Myanmar use case was also investigated during the development phase of the project.

In particular, the objectives are to accurately delineate flood extent, while minimizing over and underdetection. Extension of case-based assessments provide useful insights on the generalizability of the flood monitoring algorithm on a global scale with respect to a more comprehensive range of land covers/uses.

Subsequent sections provide detailed descriptions about how each individual flood algorithm generates uncertainty information as in Section II, the generation and evaluation of ensemble likelihoods are presented in Section III. Data used to conduct the study is described in Section IV, followed by results in Section V and discussion in Section VI. Finally, Section VII concludes this article.

II. ALGORITHM LIKELIHOODS

The GFM flood ensemble likelihood product attributes to each valid pixel a likelihood of being flooded given its recorded Sentinel-1 backscatter value and ancillary data inputs. The term valid refers to pixels that are considered to be potentially flooded and included in the computation. Invalid pixels are excluded through an exclusion mask. This mask excluded areas blocked by radar shadow, regions of no Sentinel-1 SAR sensitivity toward flood dynamics, or areas that are considered non-floodable, as they are located too far away from the next drainage [20].

Ensemble likelihoods are defined in the interval $[0, 100]$. Likelihood values toward 0 represent lowest confidence in the ensemble flood classification, whereas values toward 100 represent highest confidence. Ensemble likelihoods are used to convert the set of ensemble classifications into a single binary flood classification, representing non-flood pixels as 0 and flood pixels as 1, respectively. In this binarization step, a likelihood value of 50 is defined as the threshold value that separates the two classes (i.e., non-flood pixels from the interval $[0, 49]$ and flood pixels from the interval $[50, 100]$). Confidence about the detection of each respective class increases with likelihood values toward the lower or higher class boundaries (see Fig. 1).

The ensemble likelihood value is computed pixelwise as the mean of the likelihood values attributed to each valid pixel by the three algorithms. The following sections describe the independent generation of each set of values.

In general, all likelihoods are in reference to a flood classification. If the likelihood is low over a certain pixel or feature, the classification confidence that the pixel or feature is flooded

is also low. The specific terms that are further evaluated in the next sections are defined as follows.

- 1) *Individual Likelihood Values*: refer to pixelwise likelihood information generated by each of the three flood algorithms (i.e., DLR, LIST, TUW).
- 2) *Initial Mean Likelihood Values*: are computed pixelwise based on the average of all available individual likelihood information, ideally generated by all three flood algorithms, prior to the application of the ensemble algorithm.
- 3) *Ensemble Likelihood Values*: are updated likelihood values based on the initial mean likelihood after the ensemble algorithm is applied.

The ensemble algorithm combines the flood detection and likelihood outputs of the individual flood algorithms. Although a majority voting system is implemented, split situations, i.e., cases of classification disagreement occur where a majority cannot be achieved, e.g., when one out of three algorithms yields a nodata pixel.

Post-processing steps involve the exclusion of subareas within a given Sentinel-1 scene that overlap with the reference water and exclusion masks. Clusters with a size less than a defined threshold of flood pixels are assumed to be unlikely flooded and relabeled as non-flood pixels. This action is termed a blob removal step and eliminates small fragmented patches. Results following these steps are then referred to as ensemble classifications, see [21] and [18]. Likelihood values corresponding to formerly flooded but excluded pixels are set to a likelihood value of 0. Likelihood values corresponding to formerly flooded but blob-removed pixels are set to a likelihood value of 49, i.e., expressing the lowest confidence in the non-flood regime.

The ensemble algorithm can be applied based on two approaches: 1) split and 2) consensus. The split approach considers the likelihood values associated with the respective classification of each flood algorithm and favors the classification with the highest confidence. The consensus approach is based on majority voting, which sets all split situations to non-flooded classifications, since only 2/3 flood algorithms generate valid but conflicting pixelwise classifications. In effect, a flood classification is only returned when there is an agreement. The following subsections describe the individual likelihood layers produced by the individual flood detection algorithms.

A. Computation of DLR Fuzzy Values

The flood detection algorithm by DLR is a single scene approach, i.e., the main data input for flood inundation is a single Sentinel-1 observation. The DLR algorithm applies fuzzy logic post-processing to measure and to reduce the uncertainty associated with the water classification, originally described by [22] and [23]. The algorithm was originally built as part of a flood service, hence, using a single scene approach. The development of the system is given with [24] and [25]. Martinis et al. [26] further outlined advantages and limitations of various flood detection approaches.

Three cases influence classification uncertainty. In particular, the likelihood of a pixel being classified as water is low:

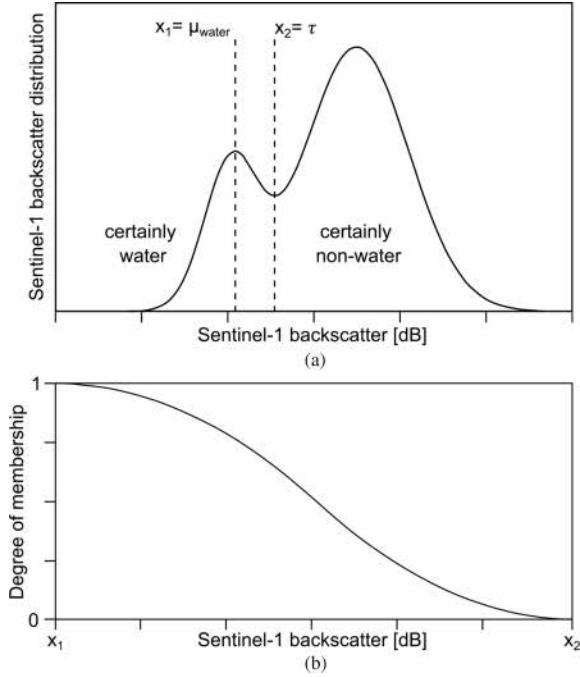


Fig. 2. Fuzzy logic approach for discriminating between water and non-water classes based on SAR backscatter values, denoted as σ^0 Sentinel-1 backscatter [dB].

- 1) if its radar backscatter is close to the automatically derived threshold τ , separating water and non-water;
- 2) if the slope at that location is high, since steeper surfaces are unlikely to retain water;
- 3) if that pixel is connected to other neighboring water pixels and the resulting area is relatively small. On the contrary, the uncertainty is low if the pixel is connected to other neighboring water pixels and the resulting area is relatively large.

The three cases or parameters, namely, backscatter of the normalized radar cross section (NRCS), slope, and minimum mapping unit are evaluated separately, resulting in the generation of three fuzzy layers. The concept of the fuzzy logic step is exemplified with the consideration of SAR backscatter values.

Fig. 2 illustrates the application of the fuzzy logic approach to address the first case where SAR backscatter is uncertain. In Fig. 2(a), the water/non-water separating threshold τ is defined as the upper fuzzy value x_2 . This value represents the boundary between both classes, where the likelihood of a correct classification is the lowest. The mean backscatter value of the class water μ_{water} is associated with a minimum fuzzy value x_1 . Pal and Rosenfeld [27] described the negative S-function that maps numeric to fuzzy values which is also depicted in Fig. 2(b).

As the majority of water pixels have backscatter values around the mean backscatter value, the uncertainty of a correct classification of these water pixels is low. The fuzzy logic approach maps high uncertainties to low degrees of membership to a particular class. For instance, high uncertainty of a correct pixelwise classification to the water class corresponds to a low degree of membership to that class. The pixel is, therefore, assigned a lower fuzzy value. The converse is also true, where a low

uncertainty corresponds to a high degree of membership of a given pixel to the water class; it is assigned a high fuzzy value.

The three individual fuzzy membership functions are in the range [0, 1]. For easier interpretation and lower storage requirements, float values were rescaled to the range [0, 100]. The resulting fuzzy layer is computed as the mean of all three individual fuzzy layers. A defuzzification value of 60 is defined as the threshold to mark the distinction between water and non-water classes.

Pixels classified as water with a fuzzy value of ≥ 60 are treated as water detections of high confidence.

B. Computation of LIST Probabilities

The flood detection algorithm by LIST applies a change-detection approach [28], i.e., the flood inundation is performed by detecting backscatter changes of two consecutive Sentinel-1 observations, the most recent SAR scene I_{t-1}^{S1} with the overlapping SAR scene acquired from the same orbit called reference SAR scene I_{t0}^{S1} . As it is a change detection algorithm, it aims at detecting and mapping all decreases of backscattering values with respect to a reference one. A change detection approach is adopted because it allows to differentiate floodwater from permanent water bodies and, at the same time, filter out classes having water-like backscattering values, such as shadows or smooth surfaces. The floodwater extent for the actual event is described with I_{t0}^{S1} and the image difference to the preevent situation is described with $I_D^{S1} = I_{t-1}^{S1} - I_{t0}^{S1}$. The likelihood of floodwater classification is characterized by flood probability.

Both I_{t0}^{S1} and I_D^{S1} are used for likelihood estimation, the pixels that have high posterior probability of both water class and change class are likely to be real flooded pixels. The probability of being flooded for a given pixel (1) is defined as the minimum value between the conditional probability of the water class $p(W|\sigma^0)$ and the conditional probability of the changed class $p(C|\Delta\sigma^0)$ with regards to the Sentinel-1 backscatter σ^0

$$p(F|\sigma^0, \Delta\sigma^0) = \min(p(W|\sigma^0), p(C|\Delta\sigma^0)) \quad (1)$$

where $p(\sigma^0)$ is the marginal distribution of backscatter values in I_{t0}^{S1} and $p(\Delta\sigma^0)$ is the marginal distribution of backscatter difference values in I_D^{S1} . In case a pixel is also flooded in the reference image, only I_{t0}^{S1} is considered for likelihood estimation of flood classification as follows:

$$p(F|\sigma^0) = p(W|\sigma^0). \quad (2)$$

As in this case, the likelihood is only calculated from the backscatter value in I_{t0}^{S1} , false high flood probability can be caused by permanent water and other water look-alike dark areas, these false alarms in binary map has been removed by comparing the resulting flood map with the previous flood map. To reduce these false high probabilities in current likelihood map, for non-flood pixels in the new flood map, their flood probability is the minimum value between $p(W|\sigma^0)$ and the value in the latest previous likelihood map.

C. Computation of TUV Uncertainties

The flood algorithm by TUV is based on a data cube approach, as introduced by [29], and builds upon a priori probability parameters for flood and non-flood conditions generated from Sentinel-1 time series. Incoming Sentinel-1 scenes that are subject to flood mapping are classified by means of Bayesian inference, which is not only computationally slim and NRT-suitable, but also intrinsically yields likelihood values in terms of posterior probabilities of the class allocation. For each pixel in a new Sentinel-1 backscatter measurement, the probability of belonging to either the flood or the non-flood class is inferred.

Based on the Bayes decision rule, higher (“winning”) posterior probabilities define then the class allocation. In addition, the conditional error $p(\text{error}|\sigma^0)$ can be defined by the lower posterior probability as follows:

$$p(\text{error}|\sigma^0) = \min [p(F|\sigma^0), p(NF|\sigma^0)] \quad (3)$$

where $p(F|\sigma^0)$ describes the probability of the flood class and $p(NF|\sigma^0)$ the probability of the non-flood class with respect to the Sentinel-1 backscatter σ^0 .

The conditional error as direct measure for *uncertainty* enables direct quantification of the lack of confidence with respect to a given decision. Since posterior probabilities sum up to 1, a higher posterior probability for one class results in a lower posterior probability for the other class in the binary classifications. Uncertainty is thus defined between 0.0 and 0.5. A value close to zero represents high confidence, since the probabilities for both classes (flood and non-flood) indicate a clear decision. High conditional errors (i.e., close to 0.5) indicate uncertain decisions, as the new observation is falling into the overlap of the local flood/no-flood distributions and hence no class is much more probable than the other one. In such a situation, the Bayes decision is very uncertain and the classification is not meaningful.

For all pixels of the incoming Sentinel-1 image, the conditional errors $p(\text{error}|\sigma^0)$ are forwarded to the ensemble algorithm, which represent the pixelwise uncertainties associated with the flood map of TUV’s algorithm. For easier interpretation and lower storage requirements, the uncertainties are scaled to values between 0 and 100.

The TUV flood mapping algorithm features some internal masking of conditions not well represented by the a priori probability parameters. This includes an internal uncertainty mask based on the statistical Sentinel-1 backscatter model of TUV is applied in this algorithm to exclude poorly based decisions (i.e., with low reliability), defined by an upper limit of 0.2 for the conditional error, reflecting a 4:1 probability that the assigned class is correct.

D. Fusion of Likelihoods

In the context of this study, likelihood values of different origins are fused to a single quantity. Probability and fuzziness can be considered equal in terms of the numerical expression of the likelihood that is represented in the unit interval [0, 1]. However, they have to be differentiated in the manner in which the two measures handle the semantic classes water and

TABLE I
LIKELIHOOD QUANTIFICATIONS OF THE THREE INDIVIDUAL FLOOD ALGORITHMS

Algorithm	Value ranges [low, threshold, high]	Type of likelihood
DLR	[0, 60, 100]	Fuzzy value
LIST	[0, 50, 100]	Probability
TUV	[100, 50, 0]	Uncertainty

non-water [30]. Given the same probability and fuzzy values of for example 0.8, the representation of likelihood is clarified as follows.

A probability of 0.8 represents an 80% chance of pixelwise water detection, where the value is determined based on pixel frequencies. The likelihood about the chance of a water or a non-water detection can be maximized as more observations become available and the pixelwise water detection is built on a broader data base.

A fuzzy value of 0.8 represents a pixel that is 80% water, describing the degree of membership belonging to that class, based on its properties. Such properties are defined through the uncertainty analysis, e.g., the DLR algorithm attributes a pixel with Sentinel-1 backscatter, slope, and size information that declare its membership to the class water. The fuzzy value expresses the degree to which it can be considered to be a (pure) water pixel. The uncertainty about class ambiguity persists even if more observations become available. Maximizing the likelihood is, however, possible by introducing additional auxiliary datasets that add to the pixel properties.

This study acknowledges the mathematical and ontological complexities that characterize the formulation of the two aforementioned types of likelihood generation. However, in order to return actionable and interpretable information, the GFM likelihood product simplifies the fusion of likelihood values by computing the average value of the three algorithm likelihood outputs. While more advanced approaches have been proposed to bring the two measures of likelihoods together, e.g., by [31], [30], and [32], this approach addresses the need for practicality in crisis information management. This objective is characterized by the need to make time-critical decisions with informative and also more easily interpretable products to support decision-making. Furthermore, the harmonized GFM likelihood product summarizes the likelihood inputs from the three water detection algorithms, thereby minimizing cognitive overload for end-users.

III. ENSEMBLE LIKELIHOODS

A. Generation of Ensemble Likelihoods

Combining the likelihood information generated by each flood algorithm requires value harmonization. Table I outlines each of the three outputs with respective value ranges. The value ranges indicate the lowest and highest classification confidences as well as the threshold distinguishing flood from non-flood. Since the TUV algorithm outputs uncertainties, a pixel value of 100 represents a maximum uncertainty value that is comparable to an LIST probability or a DLR fuzzy value of 0.

The DLR and LIST flood algorithms produce uncertainties that are numerically similar to likelihoods, with low values indicating low flood classification confidence and vice versa. The uncertainty analysis of TUV produces an inverse value range, where low values indicate high likelihood or high flood classification confidence and vice versa.

By definition, a threshold value of 50 separates the flood and non-flood pixels in the ensemble likelihood layer. Fuzzy values F generated from the DLR algorithm are adapted to this scheme, based on (4). With that, fuzzy values in the flood regime are stretched, whereas fuzzy values in the non-flood regime are compressed by an absolute value range of 10. The application of several exclusion layers to non-flood regions, however, removes certain locations from the non-flood regime which compensates for the compression of fuzzy values

$$\text{likelihood}_{\text{DLR}} = \begin{cases} 100 - 1.25 \cdot (100 - F), & F \geq 60 \\ \frac{F}{1.2}, & F < 60. \end{cases} \quad (4)$$

The TUV uncertainties U are inverted, following (5).

$$\text{likelihood}_{\text{TUV}} = \begin{cases} 100 - U, & \text{flood}_{\text{TUV}} = 1 \\ U, & \text{flood}_{\text{TUV}} = 0. \end{cases} \quad (5)$$

Once likelihood values from all three algorithms are represented in the same range, the ensemble likelihood is computed as the mean of the individual likelihood layers, irrespective of nodata values. The individual algorithm layers are not weighted, i.e., each flood and likelihood layer are treated equally both in the ensemble flood and likelihood computation.

The ensemble algorithm computes the result on pixel level and requires two sets of three input layers each for flood and likelihood computations, respectively. If two flood algorithms fail to output data, the required flood and likelihood layers are generated automatically with the same geometry as the input Sentinel-1 scene and filled with zero values stating no flood for the entire scene, accompanied with zero values stating low likelihoods.

A valid flood or a non-flood pixel is always connected to a valid likelihood pixel. If a flood pixel holds a nodata value, the corresponding likelihood pixel also stores a nodata value.

This behavior has implications for the statistical robustness of the ensemble results. For instance, a flood pixel that is based on three valid individual classifications is considered to be statistically more robust compared to a flood pixel that is based on only two valid individual classifications (and one nodata classification). The latter is a so-called split situation that is resolved through a consensus approach, i.e., the ensemble algorithm marks that pixel as not flooded.

Fig. 3 illustrates three different cases (C1, C2, C3) to demonstrate the ensemble classification scheme and further elaborates on the majority voting system. In the first case C1, three out of three of the individual algorithms return a flood classification which is a full consent. In case C2, two out of three of the individual algorithms return a classification that disagrees with the third algorithm. This is a major consent and the pixel is classified as flood or non-flooded, respectively, depending on the majority vote, e.g., [flood, flood, non-flood] or [non-flood, non-flood, flood]. In case C3, one algorithm returns a nodata

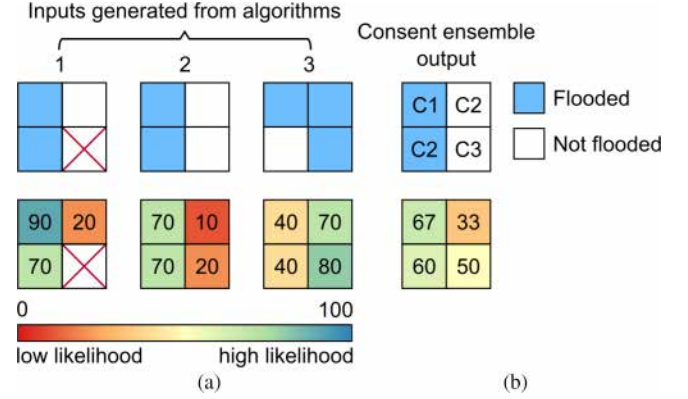


Fig. 3. Sample inputs generated by each of the three individual flood algorithms with (a) available flood and likelihood data per algorithm on pixel level and (b) possible changes in the final classification. The cases C1 and C2 handle full and majority agreement, respectively. The case C3 demonstrates a split situation that is resolved through a conservative agreement with a non-flood result.

value and the remaining algorithms disagree on the classification, e.g., [nodata, flood, non-flood]. The ensemble algorithm resolves the split situation through a conservative approach and marks the pixel as non-flooded.

The steps described in this section define initial flood mapping likelihoods that are to be corrected with auxiliary data masking out error-prone regions and excluding areas of no interest, e.g., reference water that is not flooded per definition. If these pixels are to be excluded and were classified as non-flooded, the initial likelihood value remains unchanged. If pixels to be excluded were classified as flooded, the respective likelihood value is changed to the value 49, i.e., the most unconfident likelihood value for the non-flood class.

B. Evaluation of Likelihoods

Evidence of the successful generation of algorithm and ensemble likelihoods is given with the creation of quantile-quantile plots, as described by [33]. The plot compares the computed individual algorithm and ensemble likelihood values with empirical probabilities that are computed from validation data. Quality control of Sentinel-1-based flood extent is performed with Sentinel-2 Level-1 C flood extent that was mapped manually. Hostache et al. [19] presented a general set of quantile-quantile plots during the development stage of the GFM project.

All incoming algorithm and ensemble likelihood values are transformed to prediction probabilities $p(P|L)$ (6) that represent the mean likelihood P for the respective and iterative likelihood range L . The total likelihood range is defined as the interval $[0, 1]$ with ten iteration steps

$$p(P|L) = \mu(P|L). \quad (6)$$

The binary Sentinel-2 water extent is transformed to empirical probabilities $e(W|L)$ (7) that represent the mean flood coverage F for the respective and iterative likelihood range L . The total likelihood range is defined as the interval $[0, 1]$ with ten iteration steps

$$e(F|L) = \mu(F|L). \quad (7)$$

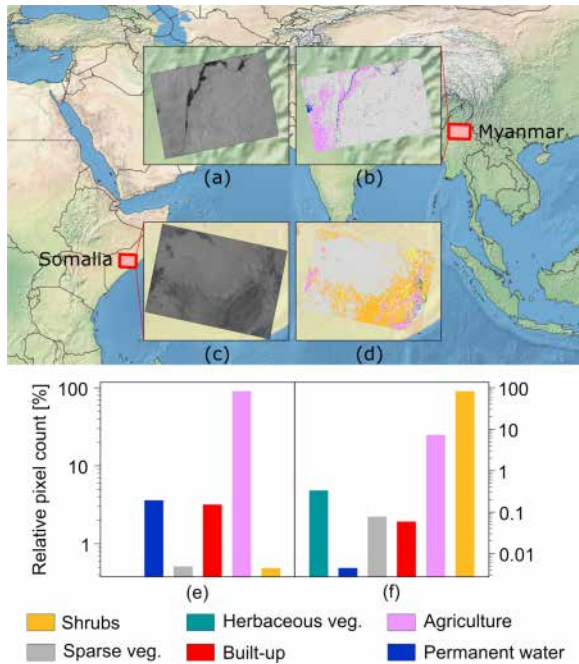


Fig. 4. Location of the study areas in Myanmar and Somalia. The Sentinel-1 scene over (a) Myanmar was acquired on 2019-07-16 11:39:44. The Sentinel-1 scene over Somalia (c) was acquired on 2019-03-16 02:46:06. Both study areas are accompanied with an overview of predominant land cover maps (b) and (d). An exclusion mask is applied and shows areas where flood computation is performed. This information is in line with the pixelwise distribution of land cover classes for (e) Somalia and (f) Myanmar, valid for non-excluded regions.

IV. DATASETS

This section gives an overview on the used datasets and how they were processed within this study. Two use cases are presented that showcase a flood event in Myanmar and a non-flood situation in the semi-arid climate zone of Somalia (see Fig. 4). The preprocessing of the Sentinel-1 IW GRDH datasets is described by the overview given in [34].

The individual flood algorithms exploit Sentinel-1 GRD IW data which are shown in Fig. 4(a) and (c). Sentinel-1 data over Myanmar was acquired on 2019-07-16 11:39:44. Sentinel-1 data over Somalia was acquired on 2019-03-16 02:46:06. The fuzzy logic step of the DLR flood algorithm uses slope information derived from Copernicus DEM data [35]. All input datasets were resampled to a common pixel spacing of 20×20 m in the Equi7Grid projection [36].

Land cover information for this study is based on the global Copernicus Land Cover product from 2019 [37] with an original pixel spacing of 100×100 m that has been resampled to 20×20 m. As for this study, it was decided to focus on selected predominant land cover types that are either of particular socio-economic interest or likely to be affected from flooding. Fig. 4(b) and (d) depict the spatial distribution of predominant land cover types for the study areas, followed by the class frequencies in Fig. 4(e) and (f). An exclusion mask is applied, leaving land cover types that are part of this analysis.

For the Myanmar use case, the land cover type *agriculture* dominates the study area with a pixel coverage of approx. 90% followed by the classes of *built-up* and *permanent water*, sharing

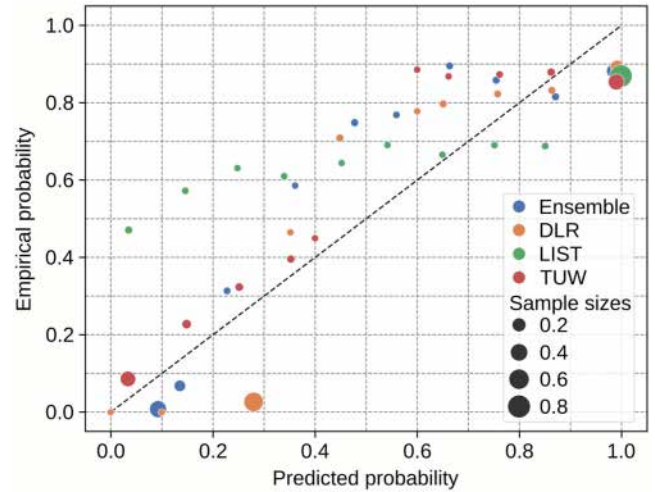


Fig. 5. Quantile-quantile plot comparing predicted probabilities (ensemble together with algorithm likelihood values) with empirical probabilities from validation data. Each marker plots into the respective bin range, e.g., $[0.0, 0.1]$. The marker size denotes the relative number of samples per algorithm. Low predicted probabilities denote low algorithm or ensemble likelihood values, which maps to higher flood classification confidence. Low empirical probabilities denote greater flood extent for the same pixel locations.

less than 5% coverage each. Land cover types depicting forests and similar classes are not considered, although shown in green colors in the map, as they represent a state of dense vegetation that is mostly excluded from the flood computation.

For the Somalia use case, the land cover type *shrubs* dominates the study area with a pixel coverage of approx. 90% followed by the class *agriculture* with a pixel coverage of approx. 10%.

A crucial part of the presented data relies on a consistent and robust reference water dataset that was computed prior to the release of the GFM products. The GFM reference water dataset exploits a two-years' time series of Sentinel-1 median backscatter images that were aggregated for each month. Thus, the reference water reflects permanent water which is stable over the reference period of two years and seasonal water bodies that are periodically flooded over the duration of the reference period.

An exclusion layer defines pixels that are not included into the final ensemble output, as examined by [38]. As already mentioned in Section II, the exclusion layer contains information about radar shadows, dense vegetation, and permanent low backscatter, i.e., regions where the flood inundation is hampered, as well as topographic regions that are not prone to flooding.

In order to evaluate the ensemble likelihood values validation data covering the Myanmar study area is introduced. The data consists of a binary flood extent map derived from Sentinel-2 Level-1 data, which was acquired on 2019-07-15 with a one-day delay to the acquisition of Sentinel-1 data over Myanmar.

V. RESULTS

In relation to the validation data, this study further examines a quantile-quantile plot supporting the evaluation of likelihood values (see Fig. 5). The quantile-quantile plot shows the

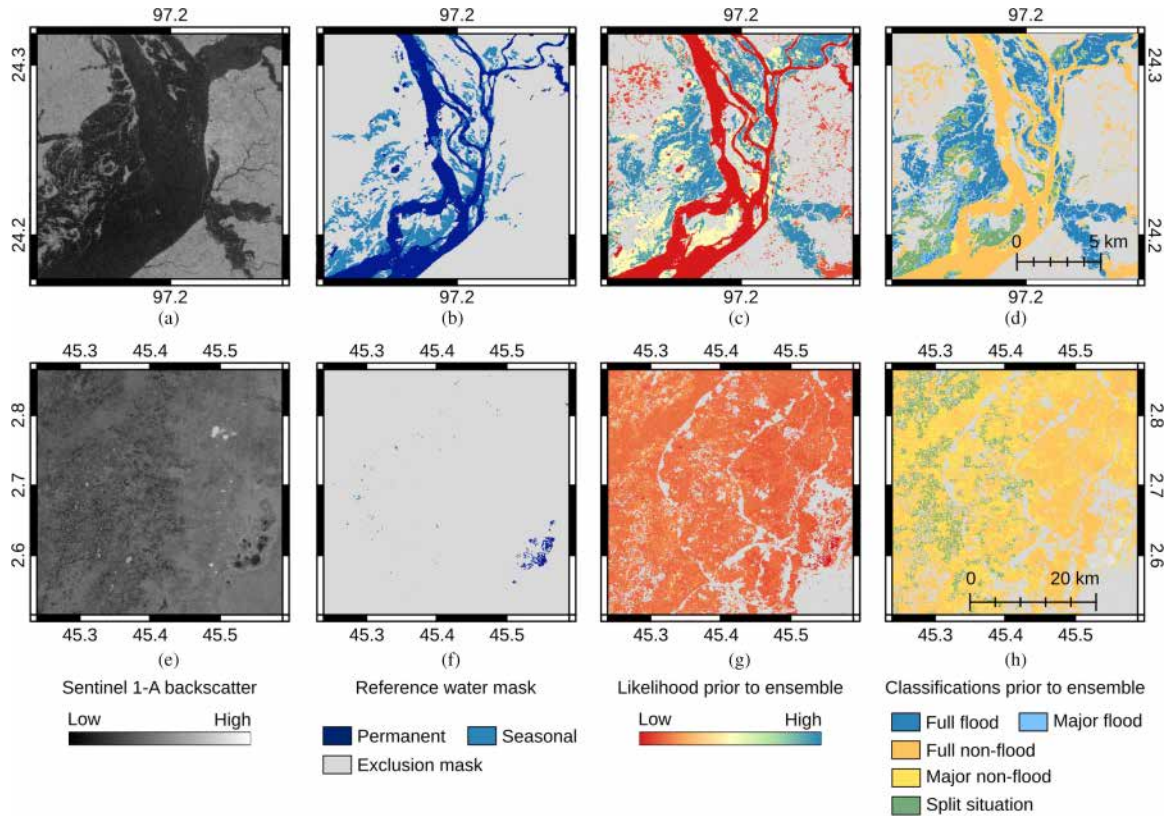


Fig. 6. Series of image chips sampled from the (a)–(d) Myanmar and (e)–(h) Somalia use case: Sentinel-1A image with backscatter values (a) and (e), reference water and exclusion mask (b) and (f), likelihoods prior to ensemble (c) and (g), and spatial distribution of flood classifications prior to ensemble (d) and (h).

agreement of the computed individual algorithm and ensemble likelihood values with empirical probabilities. The plot shows a clustering of low and high probabilities. Low predicted probabilities tend to map to low empirical probabilities and vice versa. The majority of samples show higher predicted probabilities compared to empirical samples. Samples where the empirical probability exceeds the predicted probability occur with a low pixel count within the likelihood interval $[0.3, 0.9]$. It should also be noted that the empirical probability reaches maximum values of only 0.9.

To address the objectives defined in Section I-B, scenarios are examined to identify specific likelihood regimes reflecting the number of algorithms that were used to compute the pixel-wise likelihood.

The Myanmar use case represents a known flood event on July 16, 2019, see Fig. 6(a)–(d).

The likelihoods in Fig. 6(c) represent the initial mean likelihood values that were computed as the average of all available pixelwise flood algorithm likelihood values, prior to the application of the ensemble algorithm.

The classification Fig. 6(d) illustrates the initial flood classification prior to the application of the ensemble algorithm. Full consent marks pixels where three out of three algorithms agree on a flood and non-flood classification, respectively. Major flood and major non-flood indicate pixels where two out of three algorithms agree on the classification of flood or non-flood, respectively, i.e., [flood, flood, non-flood], and vice versa with

[non-flood, non-flood, flood]. It should be noted that [flood, flood, nodata] also results in a major flood decision; the same applies to major non-flood decisions with [non-flood, non-flood, nodata]. Split situations mark pixels where one algorithm cannot classify a pixel, and thus, outputs nodata, while the remaining algorithms disagree on the classification, i.e., [flood, non-flood, nodata].

Very low initial mean likelihood values are observed over permanent water features, see Fig. 6(b). These values primarily correspond to areas that are excluded in the post-processing step and with initial full non-flood classifications, see Fig. 6(d). Much higher likelihood values are observed over image features that correspond to full or major flood classifications. Pixels with medium likelihood values around 50 correspond to split situations located along and within seasonal water bodies. The consensus approach resolves these split situations to non-flood decisions.

Fig. 7 illustrates split situations in green colors with likelihoods <50 that are remapped to non-flood decisions. Thus, 100% of the split pixels are reclassified to non-flood, thereby increasing the share of non-flood pixels for that particular likelihood value. Fig. 7(a) also shows major flood pixels in the likelihood range $[50, 80]$ and full flood pixels with likelihoods >80 . A small number of major flood pixels with likelihoods >50 are excluded from the final ensemble results and therefore marked as superior non-flood with reassigned likelihood values of 49. As can be seen in the top-right bar plot of Fig. 7(b), the

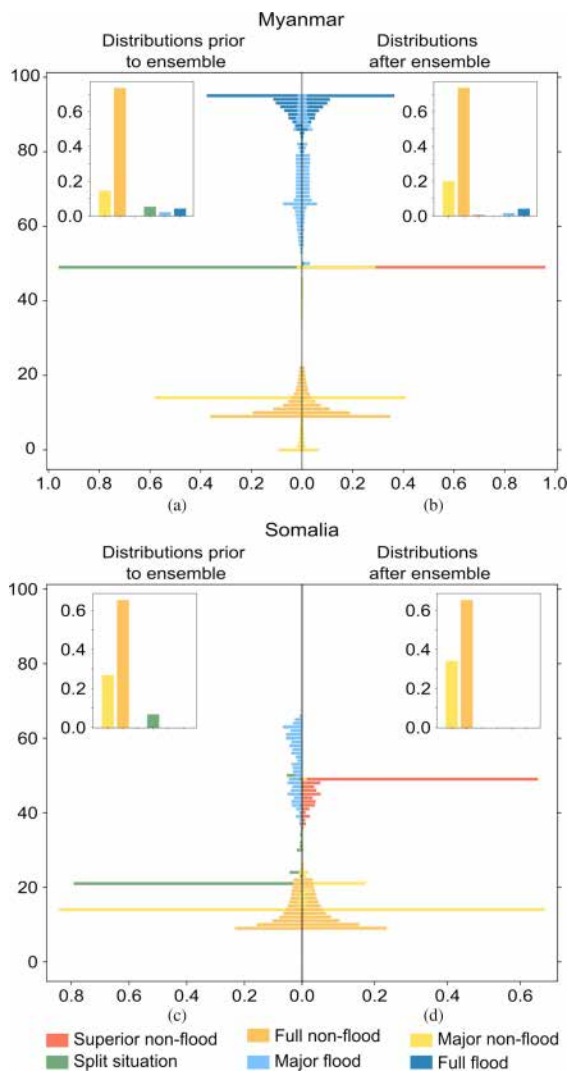


Fig. 7. Histograms of likelihood distributions for the (a) and (b) Myanmar and (c) and (d) Somalia use case. Each box contains a histogram pair (a) and (c) prior to the ensemble and (b) and (d) after the ensemble algorithm was applied. Pixels that were initially marked as flooded and have been overwritten by the exclusion layer to non-flooded are marked as superior non-flood. Depending on the distribution of the likelihood in the initial flood classification, ensemble flood likelihoods < 50 can occur, e.g., a classification with [flood, flood, non-flood] has the likelihoods [50, 50, 40] with a mean likelihood of 47. Each colored group sums up to 100 %, i.e., the bar widths are not comparable but give an indication about the likelihood distribution within that group. The share of each group on the total pixel count is given with a bar plot for each of the histograms.

count of initial flood pixels that are remapped to non-flood pixels is very low, as indicated through the small size of the green bar. Also, the full non-flood class clearly dominates the classification types, followed by major non-flood decisions. In the Myanmar study area, about 10% of all pixels are classified as flooded.

The Somalia use case represents a regular monitoring observation (i.e., non-flood event) on March 16, 2019, where overdetections are likely to be observed as the environmental setting mostly covers dry soil, see Fig. 6(e)–(h). In effect, the use case contains a very limited number of water pixels in general, which is also reflected in the reference water mask, see Fig. 6(f). Very low initial mean likelihoods, see Fig. 6(g), highlight the

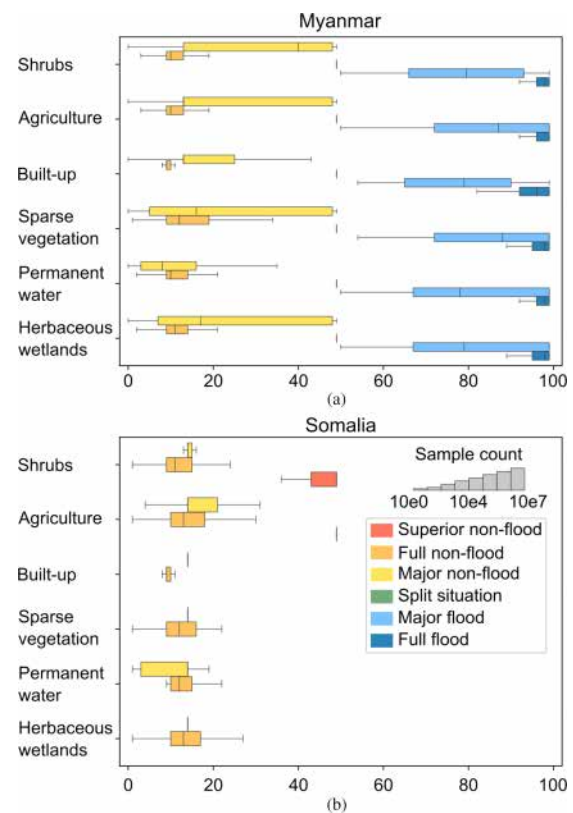


Fig. 8. Boxplots of ensemble likelihood distributions with respect to land cover/uses in the (a) Myanmar and (b) Somalia use case. The sample count for each box is indicated through the box height. Boxplots of extreme low variance are located along likelihoods of 50 and mark superior non-flood pixels. The legend in plot (b) applies for both plots.

fact that the majority of the pixels are initially classified as non-flooded, see Fig. 6(h). A substantial number of split situations are observed along meandering river channels. These split situations represent potential overdetections that appear as fragmented clusters and do not follow morphological shapes, e.g., depression boundaries, and generally correspond to dark Sentinel-1 backscatter features, i.e., potential water look-alikes.

Fig. 7(c) and (d) show medium-to-low likelihood values for split pixels that are remapped to major non-flood pixels and therefore increase the amount of major non-flood pixels, as depicted by the top-right bar plot of Fig. 7(d). It is also clear that very few flood pixels are removed at the end of the ensemble algorithm with reassigned likelihood values. Any initial flood pixel with a likelihood > 50 is reassigned with a likelihood value of 49 which is depicted in Fig. 7(d). However, the amount of these remapped pixels is still low and the majority of pixels belongs to non-flood classes.

The next set of results supports the identification of land cover classes that are associated with different likelihood values (see Fig. 8). This analysis is performed with both use cases and aims to focus on land covers with relatively high economic and social impacts to end-users, e.g., agriculture and built-up.

Fig. 7 already depicts low likelihood values for non-flood classifications for both use cases, which also dominate the

amount of considered pixels for these classes. In congruence to that, Fig. 8 shows low variance in likelihood values for full non-flood decisions across all land cover types and for both use cases. The situation differs for major non-flood decisions where the Myanmar use case shows greater likelihood variance across all land cover classes [Fig. 8(a)], in contrast to the Somalia use case [Fig. 8(b)].

As stated in Section IV, the land cover type agriculture dominates the valid pixels in the Myanmar study area and attributes to the majority of full non-flood classifications that are depicted with likelihoods of 20 and less [Fig. 7(b)]. This is also shown in Fig. 8(a), where full non-flood classifications over the land cover type agriculture show very low variance.

Major non-flood classifications show greater likelihood variance that originates from split situations, which are remapped to major non-flood in the ensemble algorithm.

For the Myanmar use case, the less dominant flood pixels show low variance for full flood and greater variance for major flood decisions across all land cover types. Superior non-flood decisions are depicted with extremely low variance across all land cover types.

As stated in Section IV, the land cover type shrubs dominates the Somalia study area and attributes to the majority of full non-flood classifications that are depicted with likelihoods of approx. 20, and less [Fig. 7(d)]. This is also shown in Fig. 8(b), where full non-flood classifications over the land cover type shrubs show very low variance.

Major non-flood classifications of greater variance are depicted for the land cover types agriculture, of which approx. 10 % of pixels build the study area, and permanent water, that is almost not present in the region. However, both major non-flood clusters are rather small.

As shown in Figs. 7(d) and 8(b), superior non-flood classifications occur and remap any potential flood classification prior to the application of the ensemble algorithm if they were masked out by the exclusion layer. Although only given with a very low number of pixels, their majority plots into the dominating land cover class shrubs and shows a likelihood cluster of low variances.

VI. DISCUSSION

The results provide a basis to obtain insights about the correlation of results prior to and after the application of the ensemble algorithm. In particular, the evaluations aim to link majority-based classifications and their statistical robustness to an explanatory variable, namely, dominant land cover types.

On the stability of the ensemble likelihood methodology, a discussion of such lacks a profound definition of the term *stable* and could express spatial, timely, or computational stability. Benefits and limitations of ensemble likelihood values for various land covers are presented in the previous and will be discussed in this section. As for timely stability, the ensemble results follow a seasonal meteorological pattern that becomes most visible for the reference water masks, which already includes a seasonal component.

A. Quantile–Quantile Plot

The quantile–quantile plot reflects the statistical reliability of the probabilistic prediction as the majority of predicted data is in close proximity to the one-to-one line.

The part of the data showing higher predicted probability compared to the empirical probability marks a flood over detection with reference to the validation data. The same applies to low predicted probabilities that mark a non-flood over detection with reference to the validation data.

The part of the data showing higher empirical probability with medium predicted probability marks flood under detections with reference to the validation data. These regions represent the minority of the data and mostly locate along the edges of detected flood patches and along river channels. It is likely that the exclusion layer does not cover these regions although they are likely to introduce misclassifications.

It should be noted that the validation data is not real ground truth data but relies on the flood extent manually derived from Sentinel-2 data. As it has been described in Section I, fusing Sentinel-1 with Sentinel-2 data not only provides a profound data basis but also contains deviating results due to the observation procedure. Furthermore, the validation data shows a one-day delay to the acquisition of Sentinel-1 data, reflecting a probable change of the flood situation. Therefore, flood patches missing in the validation data can be considered as possible sources of flood over detection in the prediction data. Clouds and cloud shadows are a known source of optical flood under detections which accounts for an over detection with SAR data when both data sources are compared.

Showcasing strengths and limitations of algorithm and ensemble likelihoods is hardly possible when assessing only a single quantile–quantile plot. However, presenting further data in higher quantities exceeds the objectives of this publication. Yet, examining empirical with predicted probabilities is a proof of concept that the underlying likelihood values are comparable with probability values from another data source.

B. Map Subfigures

Examining the series of subfigures for both use cases supports an initial assessment of the scenarios under which majority flood/non-flood classifications are identified, prior to the application of the ensemble algorithm. These results contain a significant number of split pixels where one out of three algorithms return a nodata value and two out of three disagree on the flood classification. This behavior does not indicate a failure of the system and is not to be mistaken as an inaccurate result. Nodata classifications occur if an algorithm is unable to classify a pixel with a robust likelihood, i.e., if the result shows a significantly low classification confidence. Such an output is observed over challenging SAR image features that could not be excluded from the ensemble result. The ensemble algorithm translates these unconfident results to non-flood with a likelihood value of 49, i.e., the most unconfident likelihood of the non-flood class.

In the Myanmar flood event use case, the relatively high number of split situations coincide with the location of seasonal reference water bodies and with the presence of dense

vegetation. Flood waters, including those classified based on majority decisions, also correspond with seasonal water bodies. Flood classifications over these areas represent a degree of disagreement among the contributing flood algorithms. The resulting mean likelihoods may be used to caution users to consider verifying these flood hotspots with additional data prior to making decisions, e.g., on resource allocation. Permanent water features, on the other hand, are classified as non-flooded despite one out of three algorithms classifying these features as flooded. It should be noted that the application of the reference water mask in the ensemble post-processing step reassigns the likelihood values of the permanent water features to 0 to indicate high confidence of being non-flooded. In these instances, the ensemble likelihood makes the non-flood classification explicit, regardless of the number of individually contributing flood algorithms over these pixels.

In the Somalia use case, the relatively high number of split situations, in addition to major non-flood classifications, are located along former meandering river channels that seemed to have dried out and share the same SAR signal responses as bare soils. Herbaceous vegetation, shrubs, and agriculture on dry soil are well-known challenges for flood detection based on Sentinel-1 backscatter, where the low radar backscatter tends to often result in overdetections. The consensus approach of the ensemble algorithm, although being a conservative measure, reduces these overdetections significantly and demonstrates the advantage of the ensemble algorithm over the application of a single measure alone.

C. Histograms

For the Myanmar use case, flood classifications are associated with ensemble likelihoods between 50 and 100; major floods detected with a certain degree of disagreement are associated with a wider range of lower ensemble likelihoods, while full floods are associated with a narrower range of higher confidence ensemble likelihoods between 80 and 100. The expression of agreement, as a form of confidence in flood detections, is useful information that can be consulted to support any decision-making by end users during the onset of reported flood events. It should be noted that a major flood decision originates from one out of three algorithms classifying nodata or non-flood. Regardless of the scenario, such a likelihood value is rather low and clearly indicates lower confidence for the ensemble flood classification compared to a full flood agreement, which builds on a broader data basis.

For the Somalia use case, no flood was classified, which is also a result of resolving split situations with the consensus approach. A different approach would have been to resolve split situations by favoring the flood or non-flood classification with the highest confidence in the respective class. Although less conservative, it would have been more likely to miss critical overdetections.

D. Boxplots

The pair of boxplots compare ensemble likelihood distributions with respect to land cover/uses in the Myanmar and Somalia use cases.

In the Myanmar use case, ensemble likelihoods of flood detection increase from the lower mid ranges corresponding to the initial mean likelihoods for all dominant land cover classes to notably higher ranges and demonstrate a medium spread of classification confidence.

In the Somalia use case, no floods could be observed but in comparison to the Myanmar use case, the likelihood variances are rather low and demonstrate higher confidence of the non-flood classifications.

At this stage, it should be noted that the exclusion layer also masks out regions that are likely to hamper the flood classification. It cannot be ruled out that for the Somalia use case, as a representative for challenging SAR-based flood detection conditions, a greater range of likelihood variance would be possible if another exclusion mask would have been applied to the data. However, this set demonstrates the advantage of including auxiliary data like an exclusion mask to focus on flood-prone regions.

In comparison with the utilized land cover data, the permanent water class reveals regions of disagreement with the GFM flood product. Although the GFM flood and likelihood products are not relevant for permanent water features, an intersection of these datasets shows different results due to the significantly enhanced spatial resolution of 20 m in the GFM dataset compared to the spatial resolution of 100 m in the land cover dataset.

Furthermore, the land cover class *built-up* is to be excluded with the exclusion layer. However, the same diversity in spatial resolution applies to this case as well as the definition of built-up areas that are not meant to be included in the computation. Apart from a small number of towns, both study areas also contain light settlements that are not covered by the exclusion mask. Having the likelihood values for these built-ups is useful information for the end-users as this land cover type is of high socioeconomic interest.

VII. CONCLUSION

Within this article, we describe a methodology to combine flood ensemble likelihoods of the Sentinel-1 GFM product. We further highlight the importance of interpretable and robust likelihood values to guide end-users and decision makers in their processes.

The computation of likelihoods informs on the robustness of flood classifications. While various methods have been proposed in literature to combine different types of uncertainty information origins, e.g., probabilities and fuzzy values, their computation and fusion is rather complex and arguably hampers their interpretability and a straight forward crisis response. In contrast, the method presented here is easy to interpret and its application is straight forward, as it is solely based on the computation of the arithmetic mean of the individual flood algorithm likelihoods. Considering the value range, the classification of a flood pixel with an associated likelihood close to 100 is considered to be more confident than a flood pixel with a likelihood of close to 50, and is based on a wider range of input data. Furthermore, flood classifications with low likelihood values, i.e., values in the range [50, 60], originate from an

ensemble configuration with one algorithm classifying a pixel as non-flooded or nodata. Consequently, the ensemble likelihood product alerts end-users to the presence of non-consent flood classifications, which should be treated with care in decision-making.

The first results show how ensemble likelihoods function as a heuristic to identify and provide a first indication of the performance, as well as the agreement among the three algorithms contributing to the final flood classification. These aggregated likelihood values capture cumulative uncertainties from data, model architecture, algorithmic level and interpretation. Further reduction of uncertainties requires more dedicated investigative methods. Once generated, this kind of uncertainty information can be used by both of the aforementioned communities. In particular, researchers or algorithm developers are offered guidance to investigate how to minimize uncertainties with respect to certain explanatory variables, e.g., land cover types. End users may consult likelihood information that supports cautioning against the direct use of flood classification product in areas with low likelihoods of flood classification, or where classifications are based on majority, rather than full detections. For resource allocation, it may be sufficient to identify areas of certain extents as potential hotspots, even if likelihoods associated with individual pixels in the vicinity of the areas of interest are lower.

Based on the preliminary results, the benefits and limitations of the ensemble likelihood approach are highlighted and provide a starting point for further developments and applications in the two research and end-user communities. Further assessments may be conducted to include additional variables, in addition to extending the number and variety of use cases. A broader analysis with higher quantities of input data, a wide coverage of land uses and a long-term assessment would answer the thesis that the individual flood algorithms bring specific benefits and limitations that are directly connected to the underlying flood detection mechanism. Such a long-term assessment was not possible until the GFM service became recently operational. At locations where a single algorithm is sensitive to misclassifications, applying an ensemble approach balances out such limitations.

REFERENCES

- [1] United Nations Office for Disarmament Affairs, *Human Cost of Disasters: An Overview of the Last 20 Years 2000-2019*, New York, NY, USA: United Nations, 2020. [Online]. Available: <https://www.un-ilibrary.org/content/books/9789210054478>
- [2] F. Dottori et al., "Increased human and economic losses from river flooding with anthropogenic warming," *Nature Climate Change*, vol. 8, no. 9, pp. 781–786, 2018.
- [3] W. J. W. Botzen, O. Deschenes, and M. Sanders, "The economic impacts of natural disasters: A review of models and empirical studies," *Rev. Environ. Econ. Policy*, vol. 13, no. 2, pp. 167–188, 2019.
- [4] F. Kadri, B. Birregah, and E. Châtelet, "The impact of natural disasters on critical infrastructures: A domino effect-based study," *J. Homeland Secur. Emerg. Manage.*, vol. 11, no. 2, pp. 217–241, 2014.
- [5] J. E. Suk, E. C. Vaughan, R. G. Cook, and J. C. Semenza, "Natural disasters and infectious disease in Europe: A literature review to identify cascading risk pathways," *Eur. J. Public Health*, vol. 30, no. 5, pp. 928–935, 2020.
- [6] "Sendai framework for disaster risk reduction 2015–2030," 2015. [Online]. Available: <http://digitallibrary.un.org/record/793460>
- [7] S. Grimaldi, Y. Li, V. R. N. Pauwels, and J. P. Walker, "Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges," *Surv. Geophys.*, vol. 37, no. 5, pp. 977–1034, 2016.
- [8] M. Chini, "The Sentinel-1 global flood monitoring system of the copernicus emergency management service: Introducing an ensemble approach based on three independent retrieval algorithms," presented at the ESA Living Planet Symp., 2022.
- [9] T. Spasova and R. Nedkov, "On the use of SAR and optical data in assessment of flooded areas," *Proc. SPIE*, vol. 11174, 2019, Art. no. 111740V.
- [10] M. Clement, C. Kilsby, and P. Moore, "Multi-temporal synthetic aperture radar flood mapping using change detection," *J. Flood Risk Manage.*, vol. 11, no. 2, pp. 152–168, 2018.
- [11] M.-J. Jo, B. Osmanoglu, B. Zhang, and S. Wdowinski, "Flood extent mapping using dual-polarimetric SENTINEL-1 synthetic aperture radar imagery," *ISPRS-Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 3, pp. 711–713, Apr. 2018.
- [12] L. Pulvirenti, N. Pierdicca, M. Chini, and L. Guerriero, "An algorithm for operational flood mapping from synthetic aperture radar (SAR) data using fuzzy logic," *Natural Hazards Earth Syst. Sci.*, vol. 11, no. 2, pp. 529–540, 2011.
- [13] C. Huang, Y. Chen, S. Zhang, and J. Wu, "Detecting, extracting, and monitoring surface water from space using optical sensors: A review," *Rev. Geophys.*, vol. 56, no. 2, pp. 333–360, 2018.
- [14] G. J.-P. Schumann and D. K. Moller, "Microwave remote sensing of flood inundation," *Phys. Chem. Earth, Parts A/B/C*, vol. 83–84, pp. 84–95, 2015.
- [15] S. Martinis, S. Groth, M. Wieland, L. Knopp, and M. Röttich, "Towards a global seasonal and permanent reference water product from Sentinel-1/2 data for improved flood mapping," *Remote Sens. Environ.*, vol. 278, 2022, Art. no. 113077.
- [16] V. Hertel, C. Chow, O. Wani, M. Wieland, and S. Martinis, "Probabilistic SAR-based water segmentation with adapted Bayesian convolutional neural network," *Remote Sens. Environ.*, vol. 285, 2023, Art. no. 113388.
- [17] P. Salamon et al., "The new, systematic global flood monitoring product of the copernicus emergency management service," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1053–1056.
- [18] P. Matgen et al., "Provision of an automated, global, satellite-based flood monitoring product for the Copernicus Emergency Management Service: Product definition document (PDD)," 2021. [Online]. Available: <https://extwiki.eodc.eu/GFM/PDD>
- [19] R. Hostache et al., "A first evaluation of the new CEMS systematic global flood monitoring product," presented at the Connecting Global to Local Hydrological Modelling and Forecasting: Scientific Advances and Challenges, 2021.
- [20] A. Nobre et al., "Height above the nearest drainage—A hydrologically relevant new terrain model," *J. Hydrol.*, vol. 404, no. 1, pp. 13–29, 2011.
- [21] M. Chini et al., "The sentinel-1 global flood monitoring system of the Copernicus Emergency Management Service: Introducing an ensemble approach based on three independent retrieval algorithms," in *Proc. ESA Living Planet Symp.*, 2022.
- [22] S. Martinis, J. Kersten, and A. Twele, "A fully automated TerraSAR-X based flood service," *ISPRS J. Photogrammetry Remote Sens.*, vol. 104, pp. 203–212, 2015.
- [23] A. Twele, W. Cao, S. Plank, and S. Martinis, "Sentinel-1-based flood mapping: A fully automated processing chain," *Int. J. Remote Sens.*, vol. 37, no. 13, pp. 2990–3004, 2016.
- [24] S. Martinis, A. Twele, and S. Voigt, "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data," *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 2, pp. 303–314, 2009.
- [25] S. Martinis, A. Twele, C. Strobl, J. Kersten, and E. Stein, "A multi-scale flood monitoring system based on fully automatic MODIS and TerraSAR-X processing chains," *Remote Sens.*, vol. 5, no. 11, pp. 5598–5619, 2013.
- [26] S. Martinis et al., "Comparing four operational SAR-based water and flood detection approaches," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3519–3543, 2015.
- [27] S. K. Pal and A. Rosenfeld, "Image enhancement and thresholding by optimization of fuzzy compactness," *Pattern Recognit. Lett.*, vol. 7, pp. 77–86, 1988.

- [28] M. Chini, R. Hostache, L. Giustarini, and P. Matgen, "A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6975–6988, 2017.
- [29] B. Bauer-Marschallinger et al., "Satellite-based flood mapping through Bayesian inference from a Sentinel-1 SAR datacube," *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3673.
- [30] B. Kosko, "Fuzziness vs. probability," *Int. J. Gen. Syst.*, vol. 17, no. 2/3, pp. 211–240, 1990.
- [31] Y. He, M. Mirzargar, and R. M. Kirby, "Mixed aleatory and epistemic uncertainty quantification using fuzzy set theory," *Int. J. Approx. Reason.*, vol. 66, pp. 1–15, 2015.
- [32] U. Hillenbrand, "On the relation between probabilistic inference and fuzzy sets in visual scene analysis," *Pattern Recognit. Lett.*, vol. 25, no. 15, pp. 1691–1699, 2004.
- [33] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proc. 22nd Int. Conf. Mach. Learn.*, Bonn, Germany, Aug. 2005, pp. 625–632.
- [34] W. Wagner et al., "A Sentinel-1 backscatter datacube for global land monitoring applications," *Remote Sens.*, vol. 13, no. 22, 2021, Art. no. 4622.
- [35] "Copernicus DEM-global and European digital elevation model (COP-DEM)," doi: [10.5270/ESA-c5d3d65](https://doi.org/10.5270/ESA-c5d3d65).
- [36] B. Bauer-Marschallinger, D. Sabel, and W. Wagner, "Optimisation of global grids for high-resolution remote sensing data," *Comput. Geosci.*, vol. 72, pp. 84–93, 2014.
- [37] M. Buchhorn et al., "Copernicus global land service: Land cover 100 m: Collection 3: Epoch 2019: Globe," 2020, doi: [10.5281/zenodo.3939050](https://doi.org/10.5281/zenodo.3939050).
- [38] J. Zhao, F. Roth, B. Bauer-Marschallinger, W. Wagner, M. Chini, and X. Zhu, "A preliminary comparison of two exclusion maps for large-scale flood mapping using Sentinel-1 data," in *Proc. ISPRS Geospatial Week*, submitted for publication, 2023.



Christian Krullikowski received the B.Sc. degree in geological sciences and the M.Sc. degree in geological sciences with a specialization in hydrogeology from Freie Universität Berlin, Germany, in 2013 and 2016, respectively.

In 2017, he joined the Department Geo-Risks and Civil Security, German Remote Sensing Data Center, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His research interests include thematic processing with rule-based and machine learning methods for efficient disaster response.



Candace Chow received the B.Sc. degree in physical geography from the University of Toronto, Toronto, ON, Canada, in 2009, the M.Sc. degree from the IHE Delft Institute for Water Education, The Netherlands, in 2014, with works on enhancing the thematic accuracy of SAR-based flood delineation algorithms, and the Ph.D. degree from the Universität Bern, Switzerland, in 2019, with works on the physical vulnerability to natural hazards.

She participated in an Erasmus Mundus program specializing in flood risk management at Universität

Bern. In 2020, she joined the Geo-Risks and Civil Security Department, German Remote Sensing Data Center, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. She worked at the interface of research and operations, focusing on uncertainty quantification and communicating time-sensitive information to end users, particularly within the context of the International Charter "Space and Major Disasters".



Marc Wieland received the Diploma degree in geography from Ruprecht Karls Universität, Heidelberg, Germany, in 2009, and the Ph.D. degree from the Technical University of Berlin, Berlin, Germany, in 2013, working on exposure estimation from multi-source imaging for rapid seismic vulnerability assessment.

In 2010, he joined the German Research Centre for Geoscience, Potsdam, Germany. In 2015, he joined Chiba University, Japan, to work on statistical pattern recognition in SAR time series. In 2016, he joined the

University of Oxford, U.K., as a Postdoctoral Researcher. He is currently with the German Remote Sensing Data Center, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His research interests include machine learning techniques for emergency response.



Sandro Martinis received the Diploma degree in geography, physics, and remote sensing and the Ph.D. degree from the University of Munich, Munich, Germany, in 2006 and 2010, respectively, working on automatic flood detection using high-resolution X-band SAR satellite data at the German Aerospace Center (DLR).

From 2006 to 2007, he was a Research Associate with the University of Munich working on the development of remote sensing-based methods for the monitoring of glacier motions and subglacial volcanic eruptions. Since 2013, he has been the Head of the research group "Natural Hazards" within the Department Geo-Risks and Civil Security, DLR. Since 2016, he has been leading the operational activities of Germany's contribution to the International Charter "Space and Major Disasters."



Bernhard Bauer-Marschallinger received the Ph.D. degree in geodesy and geoinformation from the Technische Universität Wien (TU Wien), Wien, Austria, in 2018.

As part of the TU Wien Remote Sensing group of the Department of Geodesy and Geoinformation, he has been involved i.a. in projects from ESA, EU, Copernicus Global Land Services (CGLS), Copernicus Emergency Management Services (CEMS), and Austrian Research Promotion Agency (FFG). His focus is on the exploitation of satellite data and

the development of operational processing software for synthetic aperture radar observations, aiming for algorithms that operate globally and in near real-time. With his team, he pursues the enhancement of the Sentinel-1 SAR processing and of scalable and efficient spatial reference systems. His research interests include the retrieval of geophysical variables like soil moisture, water bodies, and flood dynamics, employing signal and image analysis methods driven by mathematics and physics.



Florian Roth received the B.Sc. and M.Sc. degrees in geodesy and geoinformation from the Technische Universität Wien (TU Wien), Wien, Austria, in 2016 and 2020, respectively.

In 2020, he joined the Remote Sensing Group, Department of Geodesy and Geoinformation, TU Wien, in 2020. His research interests include retrieving and evaluating flood information from SAR-based satellite data.



Patrick Matgen received the M.Sc. degree in environmental engineering with the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2002, and the Ph.D. degree in water engineering from the Technical University, Delft, The Netherlands, in 2011.

Since January 2020, he has been leading the group "Remote Sensing and Natural Resources Modelling," Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg. The group's aim is to develop the synergistic use, processing, and interpretation of data from multiple complementary active and passive sensors

installed on both space and airborne platforms and to provide deeper insights into the relations between spectral features and properties of the Earth's natural resources. His research interests include the assimilation of remote sensing-derived observations into hydrodynamic models and the assessment of flood hazard and risk at large scale.



Marco Chini (Senior Member, IEEE) received the M.Sc. degree in electronic engineering from the Sapienza University of Rome, Rome, Italy, in 2003, and the Ph.D. degree in geophysics from the University of Bologna, Bologna, Italy, in 2008.

From 2003 to and 2012, he collaborated with the Sapienza University of Rome. From 2006 to 2008, he was a Visiting Researcher with the University of Colorado, USA. From 2008 to 2012, he was with INGV, Italy. Since 2013, he has been with the LIST, Luxembourg. His research interests include the anal-

ysis of multitemporal data, machine learning using SAR and optical data, soil moisture retrieval, and InSAR technique for geophysical applications.

Dr. Chini is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



Renaud Hostache received the M.Sc. degree in mechanics of geophysical media and environment from Joseph Fourier University, Grenoble, France, in 2003, the M.E. degree in hydraulics and environmental sciences from the Polytechnic Institute of Grenoble, Grenoble, France, in 2003, and the Ph.D. degree in water sciences from AgroParisTec, Montpellier, France, in 2006.

From 2007 to 2014, he was with the Centre de Recherche Public Gabriel Lippman, Belvaux, Luxembourg. From 2015 to 2021, he was a Researcher

with the Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg. Since 2022, he has been a Research Professor with the French Research Institute for Sustainable Development, Espace-Dev Research Unit, Montpellier, France. His research interests include exploitation of satellite Earth observation for flood and drought monitoring and integration of remote sensing observations in hydrodynamic models for reducing prediction uncertainty.



Yu Li received the B.Sc. degree in geoinformation science and technology and the M.Sc. degree in mathematical geology from the China University of Geosciences, Wuhan, China, in 2012 and 2015, respectively, and the Ph.D. degree in physical geography from the Ludwig Maximilian University of Munich, Germany, in 2020.

From 2016 to 2019, he was a Ph.D. Sstudent with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He is currently a Junior Researcher and Technology Associate with the “Remote Sensing and Natural Resources Modeling” Group, Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg. His research interests include machine/deep learning, multitemporal image analysis, and natural disaster monitoring with SAR data.



Peter Salamon received the M.Sc. degree in applied environmental geoscience from the Eberhard-Karls University, Tübingen, Germany, in 1999, and the Ph.D. degree in hydraulic and environmental engineering from the Polytechnic University of Valencia, Valencia, Spain, in 2006.

He is currently a Scientific Project Manager with the Joint Research Centre of the European Commission, Ispra, Italy, where he is coordinating the Copernicus Emergency Management Service and its components the of European and Global Flood Awareness

System. During his research activities and professional career, he has gained experience in policy support in the area of disaster risk management at European and global level. He has authored and coauthored more than 50 peer-reviewed articles. His research interests include hydrology, numerical modeling, hazard and risk mapping, climate change impacts, uncertainty assessment, and operational flood forecasting.