Science of Remote Sensing 11 (2025) 100210

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Full length article

Evaluating the robustness of Bayesian flood mapping with Sentinel-1 data: A multi-event validation study

Florian Roth [a,*], Mark Edwin Tupas [a,b], Claudio Navacchi [a], Jie Zhao [a], Wolfgang Wagner [a], Bernhard Bauer-Marschallinger [a]

[a] Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstr 8, Vienna, A-1040, Austria
[b] Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City, 1101, Philippines

ARTICLE INFO

Keywords:
Flood mapping
Remote sensing
Validation
SAR
Sentinel-1

ABSTRACT

The impact of recent extreme flood events has once again highlighted the importance of accurate near-real-time flood information. Consequently, a number of operational services have been established that primarily use Synthetic Aperture Radar (SAR) data to map flood extent. Among them is the Global Flood Monitoring (GFM) service that is part of the Copernicus Emergency Management Service (CEMS). Using the systematic monitoring capabilities of Sentinel-1, it is the first service to deliver flood maps fully automatic on a global scale. To automatically and reliably monitor flood extent worldwide, the strengths and weaknesses of flood mapping methods need to be known under various and sometimes challenging conditions. To examine the performance of the TU Wien Bayesian flood mapping algorithm, which is one of the scientific flood algorithms used operationally in the CEMS GFM service, we designed this validation study in which we compare our results with all compatible Sentinel-1-based flood events of the CEMS on-demand mapping (ODM) service between January 2021 and January 2022. In total, the study investigates 18 events from five continents. In addition to computing common accuracy metrics, eight representative events were analysed in detail to understand the reasons for the differences found, identify potential improvements for the method, and gain generic insights for radar-based flood mapping. Most differences are caused by the use of the VH polarization in some of the ODM reference maps, while the GFM service so far relies exclusively on VV polarization due to computational costs. The impact of using two polarizations can be seen in particular over vegetation or in case of windy conditions. Furthermore, while the post-processing strategy applied in the TU Wien algorithm helps to prevent speckle impact, it also smooths out important details in small-scale flood events. Nonetheless, the automatic TU Wien algorithm achieved a Critical Success Index (CSI) of over 70% against the semi-automatic reference in 10 of 18 flood events. It exceeds this mark for all large-scale events and in cases without vegetation close to the flooded surfaces. Overall, the median User's Accuracy (UA) is 84.0 %, the Producer's Accuracy (PA) is 72.9% and the Overall Accuracy (OA) is 85.3%. The results demonstrate that the GFM service would benefit for using both VV and VH polarization and relaxing filters applied in the SAR processing workflow.

1. Introduction

Satellite-based flood mapping has gained considerable momentum with the growing availability of new satellites, operational services, and advanced methodologies. The information provided by operational flood forecasting and monitoring services, such as those offered by the Copernicus Emergency Management Service (CEMS), is crucial as more people are impacted by floods (Tellman et al., 2021), and the frequency of severe events is expected to rise due to climate change (on Climate Change (IPCC), I.P., 2023). Flood mapping results, obtained from Synthetic Aperture Radar (SAR) observations, serve a critical function in emergency management and can contribute to flood risk modelling or prediction efforts. Most commonly, the flooded area is retrieved based on the SAR backscatter, since it generally features a high contrast between land and water. This characteristic results from the radar's high sensitivity to surface roughness, whereby calm open water surfaces act as smooth mirror-like reflectors, showing low backscatter in contrast to non-inundated land. This allows flooded areas to be detected well based on the drop of backscatter. Nevertheless, there are still some challenging conditions or land cover types that need to be taken into account in advanced flood mapping methods (Chini et al., 2021; Amitrano et al., 2024). Vegetation is associated with more complex scattering mechanisms, which can lead to both missed flood

* Corresponding author.
E-mail address: florian.roth@geo.tuwien.ac.at (F. Roth).

https://doi.org/10.1016/j.srs.2025.100210
Received 5 December 2024; Received in revised form 24 January 2025; Accepted 13 February 2025
Available online 22 February 2025

detections and false alarms (Pulvirenti et al., 2016; Moya et al., 2019). Similar effects can be observed with other dynamic phenomena like wind or snow cover. Finally, the lack of sensitivity in densely vegetated or urban areas limits the applicability of many methods.

Due to its ability to ensure weather-independent observations of flooded areas, SAR data is particularly suitable for operational services, despite the mentioned challenges. Westerhoff et al. (2013) and Salamon et al. (2021) distinguish between two main strategies for operational flood mapping services: on-demand mapping and systematic services. On-demand mapping is commonly performed for the limited area and at the time of a specific hazard event, while systematic services routinely provide automatically produced maps without the need for a manual activation. The most significant methodological difference between the two service types is that in the case of on-demand services, manual interventions by experts are possible (semi-automatic), whereas the algorithms in systematic services provide results automatically and consistently everywhere. To meet the requirements of systematic services, automatic methods must function under many different and also challenging conditions. For this reason, only certain methods are suitable in this context, and their performance must be validated accordingly.

Numerous flood mapping methods are known in the literature, with some providing solutions for specific problems, others optimized for particular regions, and still others representing more universal methods. However, there are significant differences between the methods in terms of technological readiness. The proof-of-concept of new methods is often given for a limited number of large-scale events, sometimes in a regional context (e.g. Clement et al., 2018; Huang and Jin, 2020; Katiyar et al., 2021; Wu et al., 2023), and many of these methods often do not progress beyond this status. Other methods are meant to solve a specific challenge (e.g. Li et al., 2019; Zhao et al., 2022; Garg et al., 2024) or are optimized for specific regions (e.g. Ohki et al., 2020; Landuyt et al., 2021; Colacicco et al., 2024). While those methods are typically evaluated in the context of their specific application, their broader performance and generalizability remain to be demonstrated, where relevant. Finally, more universal or mature methods are evaluated for many different flood events and conditions (e.g. Martinis et al., 2014; Bereczky et al., 2022; Ghosh et al., 2024), which requires a significant resource investment that is not always available. These methods show the highest technology readiness and are best suited for global operational services. In addition to the more advanced validation, some of the studies involved also question the design of the validation, such as the choice of accuracy metrics used or the use of sampling.

These differences in the validation status of flood mapping methods have already been noted in the literature. Schumann (2019) and Amitrano et al. (2024) argue that the validation and uncertainty assessment is often neglected or not thoroughly conducted. The Committee on Earth Observation Satellites (CEOS) subgroup for Land Product Validation (LPV) defined five stages to characterize the advancement of a satellite-based product's validation (Nightingale et al., 2010). Flood mapping is currently not included in the listed products, although its inclusion would certainly be desirable. However, most flood mapping methods can obviously be categorized in Stage 1, which is defined by a few validation locations and time periods. In contrast to that, other products, such as surface soil moisture (e.g. Jackson et al., 2010; Colliander et al., 2017; Gruber et al., 2020) or land cover mapping (e.g. Stehman, 2009; Zhao et al., 2014; Tsendbazar et al., 2021), can be classified at Stage 2–4 nowadays. At these stages, many different locations and time periods are used, with Stage 4 already requiring systematic validation for each new version of a method. As a consequence of the lack of comprehensive validation efforts, we often do not know how well methods perform under challenging conditions or different land cover types, when validation is only performed for individually selected events or regions. This makes it more difficult to assess the scientific value of a method. However, knowing the performance of a method does not mean that all weaknesses of the method have been resolved.

One major obstacle, challenging a higher number of comprehensive validation efforts, is the scarcity of truly independent reference data for flood situations. The challenge arises from the impracticality of directly and dynamically measuring flood extents from the ground. Consequently, independent ground-truth data must be obtained from other measurement sources (e.g. water level, drone imagery), which are often only available locally due to the resource-intensive and costly nature of data collection. Further, the data is often captured at different times than the satellite data, which introduces complication given the dynamic nature of flood events. This limits the possibilities of wide-ranging validation efforts for other satellite-based products.

1.1. Our global performance evaluation study

Our Bayesian flood mapping algorithm is described in Bauer-Marschallinger et al. (2022). It is currently utilized as one of the three scientific algorithms collectively detecting flooding in the Copernicus Emergency Management Service (CEMS) Global Flood Monitoring (GFM) (Salamon et al., 2021). The quality of the GFM products is evaluated and published on a regular basis, but the performance of the individual algorithms is only be determined indirectly. To bridge this validation gap, this study aims for the following research objectives:

1. To evaluate and present the strengths and weaknesses of our flood mapping algorithm under a wide range of environmental conditions.
2. To analyse the impact of the operational setup in GFM, specifically the provided input data, on the performance of the algorithm.

Section 2 describes the selected study areas, the reference data and the flood mapping algorithm. Details of the performance evaluation are given in Section 3. The results and their interpretation are provided in Sections 4 and 5.

2. Study areas, data and evaluated algorithm

An overview of the considered study areas is given in Section 2.1, while the reference data of the performed evaluation is presented in Section 2.2. Relevant input data for the algorithm is described in Section 2.3 and a summary of the algorithm is given in Section 2.4.

2.1. Study areas

Study areas were selected by retrieving flood events in the time period between January 2021 and January 2022 from the CEMS On-demand mapping (ODM) archive.[1] For avoiding sensor-specific or temporal differences, the selection was limited to events where Sentinel-1 has been used to produce the ODM flood map and a corresponding observation in the VV polarization is available (details given in Section 2.3). For two events (EMSR498 and EMSR504) an obvious shift between the ODM flood map and the Sentinel-1 observation was detected, which is why they are removed for the evaluation. As a result, 18 events covering five continents were identified for the performance assessment (see Fig. 1 or Table 1). Some events within this selection contain multiple Sentinel-1-based flood maps, which is why the one showing the largest extent was selected based on visual interpretation to test the performance based on the maximum available impact of the event.

The performed selection method aims for an unbiased selection of test sites, while providing many flood events with direct access to corresponding reference data. The events were caused by different triggers (e.g., La-Nina, cyclone), covering different land cover types (e.g., herbaceous vegetation, crop land, forests) and have different scales (from 89.7 ha to 11750.8 ha of flooded area).

[1] https://emergency.copernicus.eu/mapping/

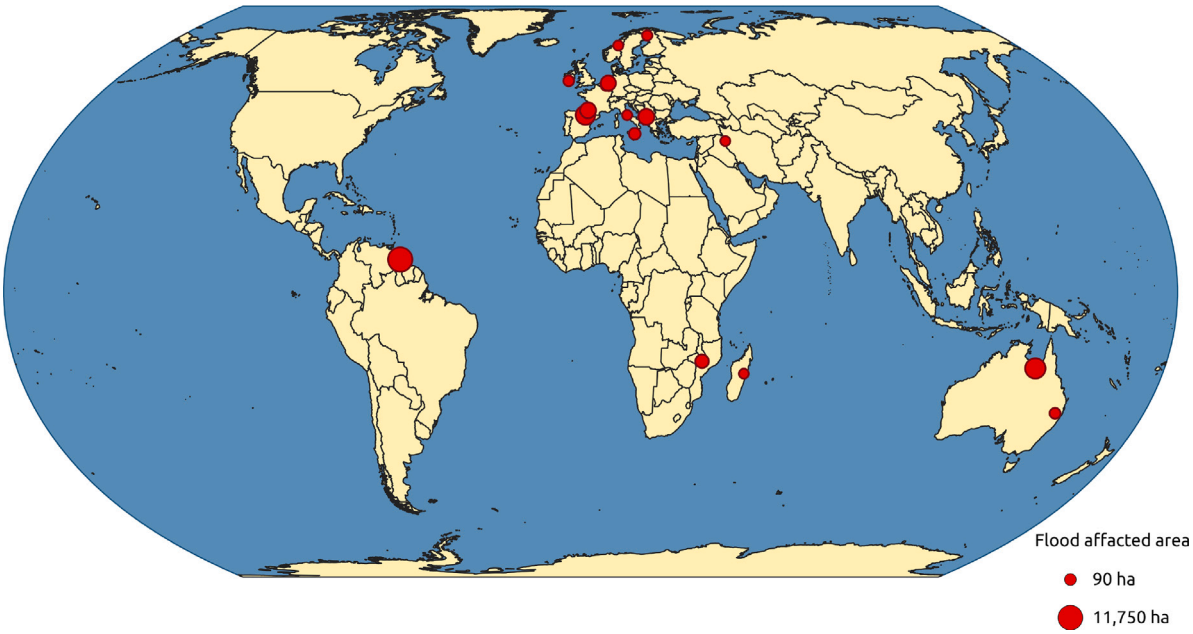

Fig. 1. Global overview of selected flood activations/events.

Table 1
Details and evaluation results of the study's flood events (coloured events, are qualitative evaluated and further discussed in Sections 4.1–4.4).

ID	Country	S-1 acquisition	Flood area [ha]	UA [%]	PA [%]	Bias	CSI [%]	OA [%]
EMSR492	France	2021/01/02	4899.0	93.1	82.6	0.89	81.1	90.6
EMSR496	Italy	2021/01/27	307.0	79.2	98.1	1.24	97.2	98.0
EMSR497	Germany	2021/02/03	4371.7	82.9	93.9	1.13	94.2	96.6
EMSR501	Albania	2021/02/12	5038.7	85.0	96.0	1.13	93.8	96.2
EMSR502	Ireland	2021/02/24	751.7	93.0	84.3	0.91	84.1	91.9
EMSR511	Finland	2021/05/22	490.3	91.8	41.9	0.46	41.9	71.3
EMSR514	Guyana	2021/06/06	11750.8	82.2	43.2	0.53	43.1	71.4
EMSR517	Germany	2021/07/15	981.3	57.1	68.9	1.21	66.5	84.0
EMSR518	Belgium	2021/07/15	752.6	88.7	72.7	0.82	70.8	86.4
EMSR520	Netherlands	2021/07/16	1197.3	90.6	62.6	0.69	63.4	81.3
EMSR548	Italy	2021/10/31	1565.8	95.5	42.5	0.45	41.3	71.1
EMSR551	Iraq	2021/11/21	89.7	36.6	17.4	0.48	17.1	57.9
EMSR554	Australia	2021/11/30	951.5	78.9	73.0	0.93	73.2	85.2
EMSR555	Spain	2021/10/12	7806.6	91.4	50.6	0.55	51.1	74.9
EMSR557	Norway	2022/01/13	259.9	73.8	2.9	0.04	2.8	51.0
EMSR559	Madagascar	2022/01/27	113.8	55.8	92.1	1.65	91.7	95.4
EMSR561	Malawi	2022/01/26	2974.5	93.7	93.0	0.99	92.1	94.6
EMSR562	Australia	2022/02/03	8508.4	74.1	73.3	0.99	71.5	85.4

2.2. On-demand mapping reference data

The Copernicus Emergency Management Service (CEMS) provides an on-demand mapping (ODM) service as well as a systematic service with the Global Flood Monitoring (GFM). As described by Wania et al. (2021), the ODM provides crisis information in maximum 24 h after activation. The service can make use of satellite imagery gathered from 30 different Copernicus Contributing Missions (CCM) but flood mapping is mostly performed based on SAR observations, due to their all-weather and -light capabilities. It provides globally distributed activations collected in a database that goes back to 2012. Hence, the service allows a systematic evaluation that would not be possible with other reference data options (e.g. airborne images or citizens science in its current state).

According to Ajmar et al. (2017), the flood extent provided by ODM relies on several semi-automatic mapping methods hosted by different service providers. Thresholds and other parameters are defined by means of visual interpretation of qualified experts. SAR data input can utilize different microwave wavelengths and polarizations. In the case of Sentinel-1, one can rely on VV and VH polarization, and sometimes on HH, when a corresponding observation is requested for emergency management.

As described in the ODM product manual (Inès et al., 2020), the service provides separate datasets for each activation (identified by a unified ID "EMSRxxx") over user-defined areas of interests (AOI). For each AOI, up to four product types are provided: reference, first estimate, delineation and grading. Since the delineation product (example given in Fig. 2) represents the event's flood extent, it is most suitable as reference data within this study. For the purpose of the flood validation, we utilized the flood extent ("observedEventA"), the AOI ("areaOfInterestA") and the permanent water bodies ("hydrographyA"). Details (acquisition time, resolution etc.) about the used satellite source are provided in a separate database-file. Some delineation products consist of multiple flood extents featuring different satellite sources and acquisition times.

The design of the ODM considering expert knowledge allowing the correction of misclassifications, which potentially occur due to challenging local characteristics within the affected area. Automatic algorithms used in systematic mapping services have to rely on appropriate models instead, as no manual intervention is possible. The goal of ODM is to support emergency management, which is why they are designed for speed rather than accuracy. However, quality

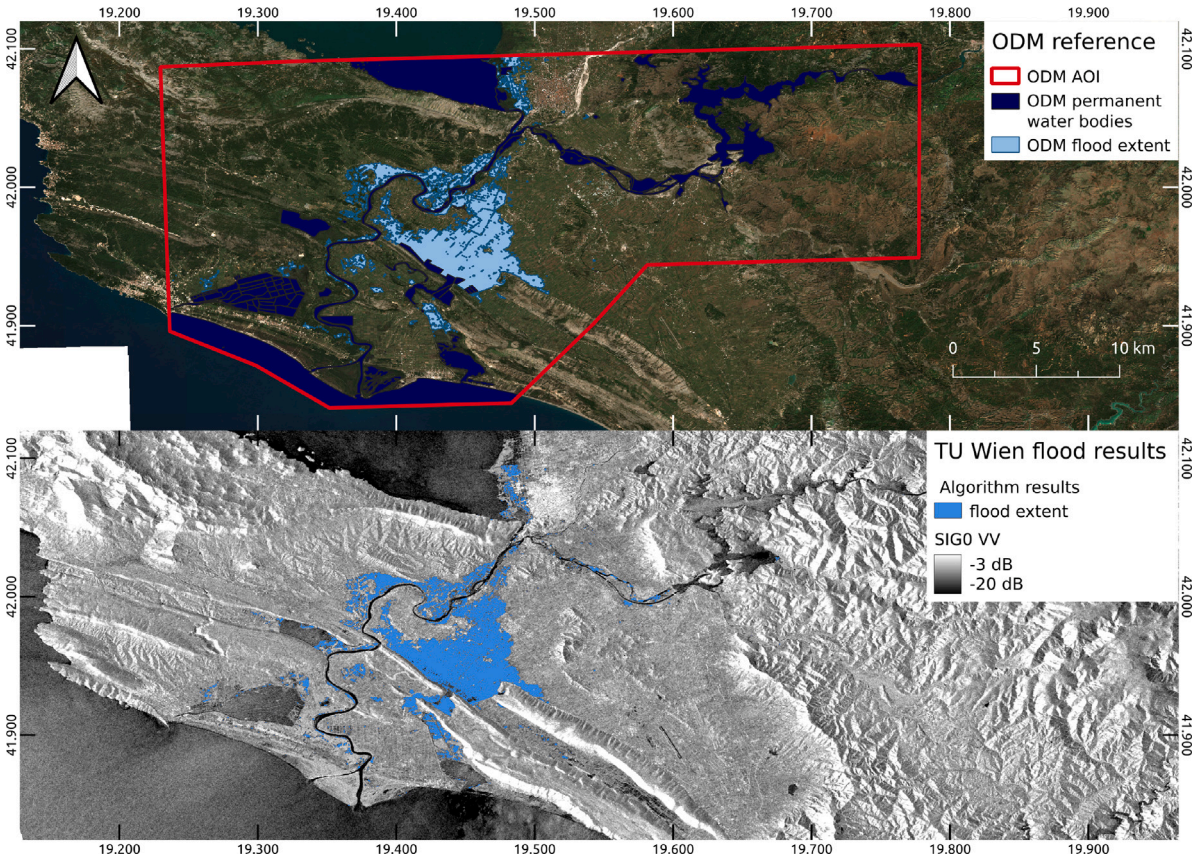

Fig. 2. Exemplary reference data for the Albania/EMSR501 event compared to corresponding result of the TU Wien flood mapping algorithm. ODM data: Copernicus Emergency Management Service, [EMSR501] Albania: Delineation Map, Monitoring 1. Upper background: ESA WorldCover Sentinel-2 True Colour Composite (TCC) from year 2021 (Zanaga et al., 2022).

control is conducted by both the service providers and the CEMS, and defective products will be replaced if necessary (Inès et al., 2020). Limiting the reference to products produced by Sentinel-1 allows for the exclusion of sensor-specific and temporal differences and focuses on the validation of the automatic approach of our algorithm. On the other hand, this restricts the independence of the reference data. Due to the mentioned advantages, as well as the lack of completely independent reference data, we consider the expert-generated reference to be a very suitable method for testing the performance of our automatic algorithm (presented in Section 2.4).

2.3. Input data for Sentinel-1 flood mapping algorithm

The Sentinel-1 observations passed to the flood mapping algorithm (described in Section 2.4) are provided from the Sentinel-1 datacube (Wagner et al., 2021) that underpins the GFM service. The datacube includes preprocessed sigma nought (σ^0 or SIG0) at a pixel-size of 20 m and derived from the GRDH product in Interferometric Wide-swath (IW). The GFM service currently relies only on the VV polarization (vertically transmitted, vertically received), primarily due to high computational costs of including multiple polarizations. Since VV shows better flood mapping accuracy compared to VH (vertically transmitted, horizontally received) when a single polarization of Sentinel-1 is used (Twele et al., 2016; Clement et al., 2018), it is the polarization of choice. Additionally, the algorithm requires the mean projected local incidence angles (PLIA) and pre-derived parameters of the utilized backscatter models (see Section 2.4). The datacube structure based on the Equi7Grid (Bauer-Marschallinger et al., 2014) allows for efficient pixel-wise comparison of all input parameters and straight-forward access to multi-temporal backscatter data.

2.4. Sentinel-1 based flood mapping algorithm

The Bayesian algorithm (see Fig. 3) performs a pixel-wise decision about flood or non-flood conditions. The availability of historic Sentinel-1 measurements within our global Sentinel-1 datacube allows the statistical description of the backscatter signature of the two states. While the backscatter signature of the flood state (being in this respect simply a water surface) is described by its linear relationship with the incidence angle, non-flood conditions are estimated by the harmonic model for non-flooded land (described below). As the backscatter behaviour from water is the same everywhere, the linear water backscatter model is defined globally through the use of observations of permanent water bodies in the years 2015–2016 from several regions in Europe. The backscatter behaviour from non-flooded land is dominated by the local surface, and consequently the non-flood signature is defined locally for each pixel individually. Utilizing the mean PLIA and the acquisition time of the incoming Sentinel-1 scene, the backscatter signatures are parametrized specifically for the flood mapping situation. Using the Bayes inference, the incoming Sentinel-1 scene is pixel-wise compared to the modelled backscatter signature of flood and non-flood conditions, and a relative probability of the pixel to belong to each class is assigned. Finally, the more probable condition is chosen. Since the relative probabilities of the classes sum to one, the probability of the less probable state can be used as a measure of the closeness of the decision, which is further referred to as uncertainty. The harmonic model (shown in Eq. (1)) aims at providing an estimate of backscatter under non-flooded conditions by considering its historic seasonality. In practice, a harmonic model of third order has proven to be effective, as it captures processes that repeat at a $\approx$ 4 months cycle, thus approximating the seasonal backscatter pattern (Schlaffer et al., 2015). The model is defined for any day-of-year t_{doy} and locally

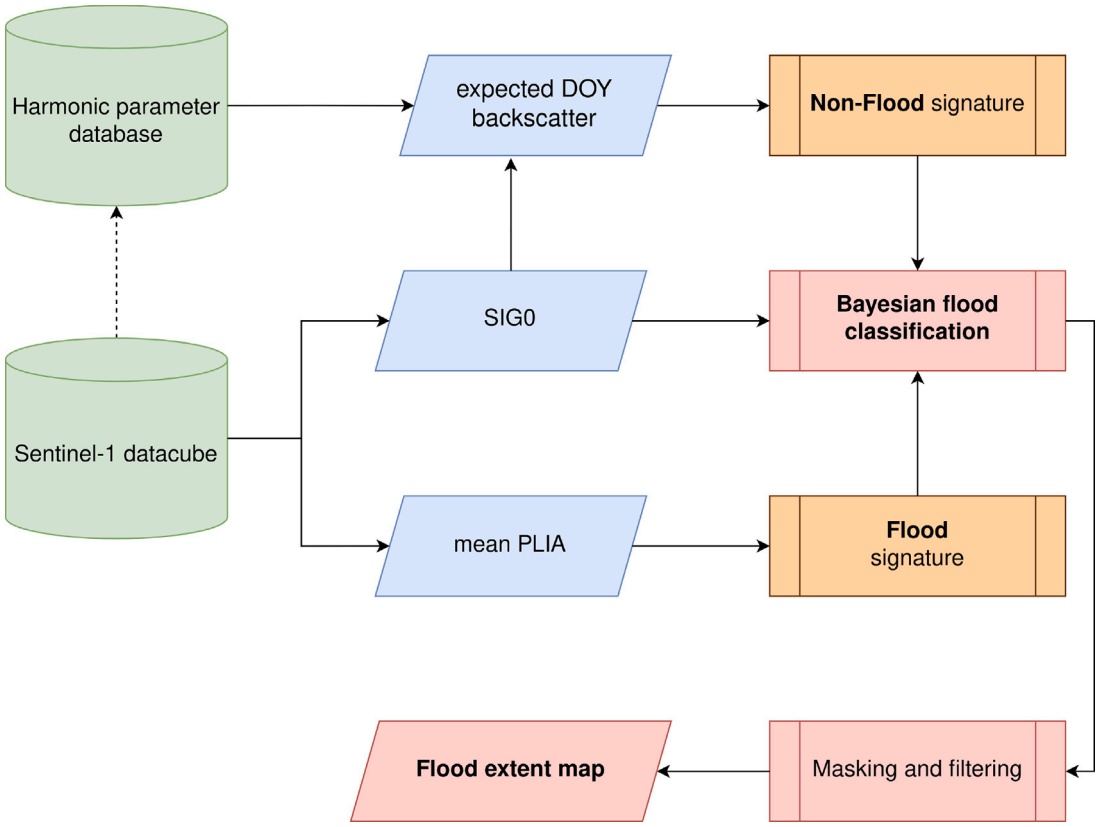

Fig. 3. Flowchart of the flood mapping algorithm which is developed by TU Wien (green = data sources, blue = individual input layers, orange = retrieved model input, red = flood results).

computed harmonic parameters C_j and S_j as well as the average SAR backscatter $\overline{\sigma}^0$ (Bauer-Marschallinger et al., 2022):

$$\overline{\sigma}^0(t_{doy}) = \sum_{j=1}^{3} \left\{ C_j \cos(j\,\nu) + S_j \sin(j\,\nu) \right\} + \overline{\sigma}^0 \quad \text{with} \quad \nu = \frac{2\pi}{365} \tag{1}$$

The required parameters of the harmonic model are precomputed and stored in the global harmonic parameter database (Roth et al., 2023) utilized within the GFM service. To ensure similar observation geometries, the parameters are defined separately for each relative orbit of Sentinel-1. For data examined here, the database's parameters were retrieved from Sentinel-1 observations covering the years 2019–2020.

As the harmonic model cannot reproduce the backscatter dynamic for all circumstances and the linear water backscatter model is limited to flat surfaces, some conditions are not supported by the models used. Furthermore, decisions showing close to equal probabilities for both classes (high uncertainty), and outlier measurements which are not well represented by the two classes, are excluded. The side-looking measuring geometry of Sentinel-1 observes the Earth's surface in viewing angles of 29° to 46°. As PLIA is defined as the angle between the surface normal and the looking direction, flat areas feature only values in this range. Consequently, non-flat areas (including a buffer, defined by: $PLIA < 27°$ or $PLIA > 48°$) are masked as exceeding PLIA. Although the harmonic model is used in the context of the non-flood class, we have calculated it for all pixels worldwide, excluding the oceans. Some pixels permanently show a low backscatter (e.g. permanent water bodies, tarmac surfaces), similar to water, while other pixels (e.g. seasonal water, wetlands) show a periodic low backscatter. In these cases, the non-flood signature retrieved from the harmonic model strongly overlaps the water signature. Since we cannot detect a temporary inundation from an already water-like low backscatter, these areas are masked (conflicting distributions).

Due to the coherent nature of the SAR signal, speckle can lead to arbitrary low and high backscatter pixels. To avoid corresponding false-positive alarms, a spatial majority filter is applied on the binary flood extent. Finally, the result is limited to flood prone areas, which is achieved by applying a topography mask. The mask excludes pixels with a vertical distance larger than 15 m above the nearest drainage using the HAND index (Rennó et al., 2008).

The flood extent map shows flood surfaces of a Sentinel-1 specific observation and is stored in the same tile structure and projection as the input data (see Section 2.3).

3. Evaluation method

For the evaluation of the Sentinel-1 based flood mapping algorithm, the flood extent maps generated by ODM (further referred to as reference data) and by the TU Wien algorithm (further referred to as classification results) are compared for the selected 18 events (listed in Section 2.1). To enable comparison, the classification results are brought to the same projection and the same extent as the reference data. Further, the permanent water layer of the ODM dataset is applied as a mask, to focus on the flood areas only. The performance is assessed by computing common accuracy metrics. While these metrics allow a straightforward comparison to other methods and an estimate of the overall performance, they lack information about the spatial distribution and reasons behind the found differences. Consequently, we perform a visual inspection of the underlying contingency maps to bridge this gap. The inclusion of many well-distributed events, with a

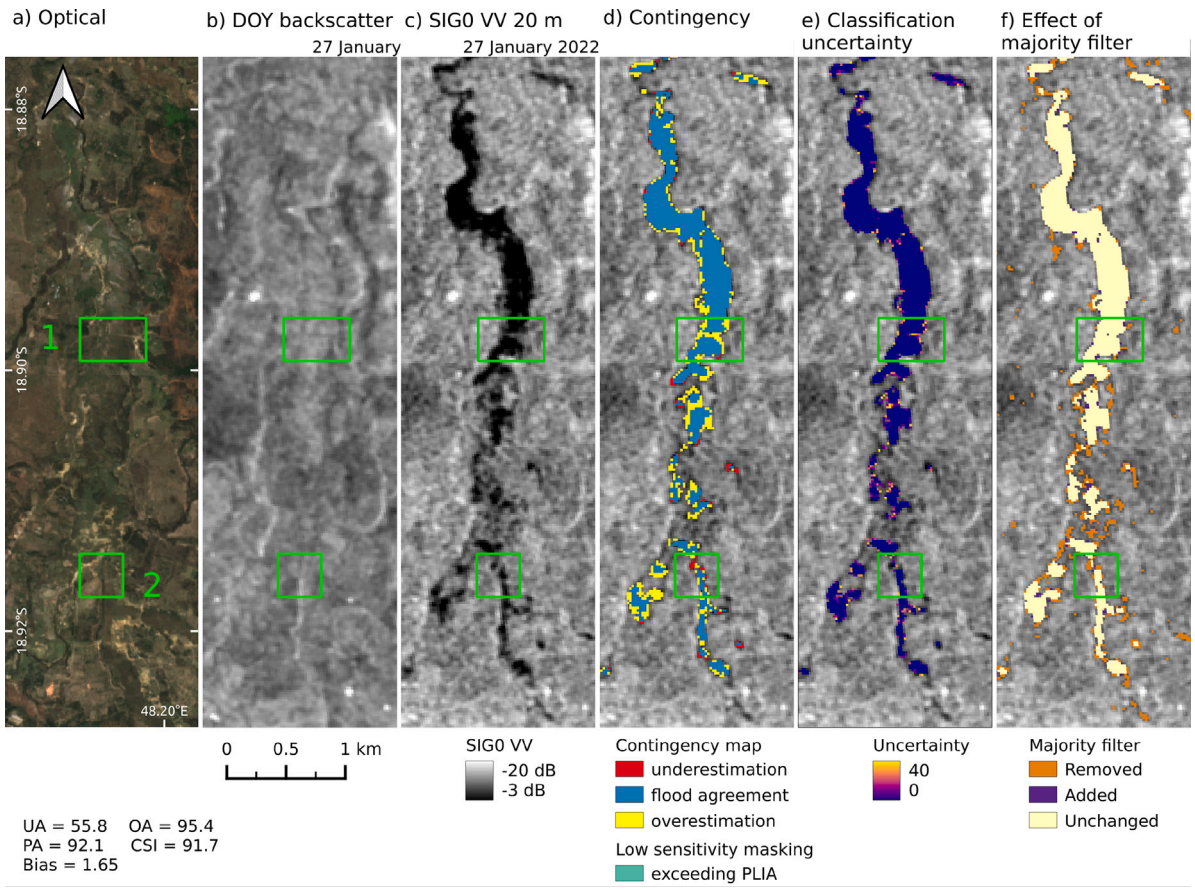

Fig. 4. Details on the flood event in Madagascar/EMSR559 from the 2022/01/27 02:19 UTC. **(a)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(b)** Expected DOY backscatter from the harmonic model **(c)** SIG0 VV 20 m flood scene **(d)** Contingency map **(e)** Classification uncertainty **(f)** Effect of majority filter.

substantial proportion of them being qualitatively analysed, contributes to the distinctiveness of our rigorous validation approach.

Two commonly used class-specific metrics are the User's and Producer's Accuracy (UA and PA). As our focus is to optimize the classification of flood pixels, both metrics are calculated by focusing on the flood class. While a high number of the UA indicates a low overestimation, a high number of the PA indicates a low underestimation of the flood class. To easily distinguish if an event tends towards over- or underestimation, the bias can be consulted. A value larger than 1 indicates an overall overestimation, while a value lower than 1 corresponds to an overall underestimation. It needs to be noted that all presented metrics strongly depend on the chosen extent of the AOI.

Besides class-specific accuracy metrics, many studies rely on global metrics like the Overall Accuracy (OA), the Critical Success Index (CSI) or the F1-score. Since the non-flood class usually consists of much more pixels compared to the flood class, some metrics like the OA or the F1 are biased towards the dominant class (Stephens et al., 2014). Besides UA and PA, the CSI is known for being independent of the proportion of the two classes. However, the CSI is known to show a bias towards larger floods and overdetection (Landuyt et al., 2019). In detail, considering the same amount of pixels being false positive (overestimation) or false negative (underestimation), the CSI values will be higher for overestimation compared to underestimation. Additionally, the CSI value for the same number of falsely classified pixels in a large scale event will be higher compared to the same number in case of a small-scale event.

Following the suggestion from Landwehr et al. (2024), we computed the accuracy metrics using an appropriate sampling design. To compensate for the unequal proportions of the two classes, we applied a random stratified sampling with an equal number of samples for each class (2500 pixels per class). To test the chosen sampling design, we repeated the random selection of samples 1000 times (i.e., bootstrapping) and calculated the standard deviation of the different metrics. Since the standard deviation does not exceed 0.01 for any metric, the design is considered appropriate. The UA, PA and bias are chosen as class-specific metrics, and we kept the CSI to allow comparisons to other studies. Given the sampling methodology and bootstrapping results, the OA is a more robust global metric and is chosen to balance the CSI. The selected accuracy metrics are computed for all 18 events, while the interpretation of the contingency maps is presented for 8 exemplary events.

The selected metrics commonly rely on the confusion matrix, consisting of the number of true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) pixels, and they are defined by Eq. (2) and (3).

$$UA = \frac{TP}{TP+FP} \qquad PA = \frac{TP}{TP+FN} \qquad bias = \frac{TP+FP}{TP+FN} \tag{2}$$

$$OA = \frac{TP+TN}{TP+TN+FP+FN} \qquad CSI = \frac{TP}{TP+FP+FN} \tag{3}$$

4. Results and discussion

The performance of our automatic flood mapping algorithm is evaluated by comparing its results to the semi-automatic results of the ODM for 18 flood events. The performance metrics of all these events as well as further details of the individual events are summarized in Table 1. In order to improve readability of the provided values, the Critical Success Index (CSI), the Overall Accuracy (OA), the User's (UA) and the Producer's Accuracy (PA) are scaled between 0 and 100%.

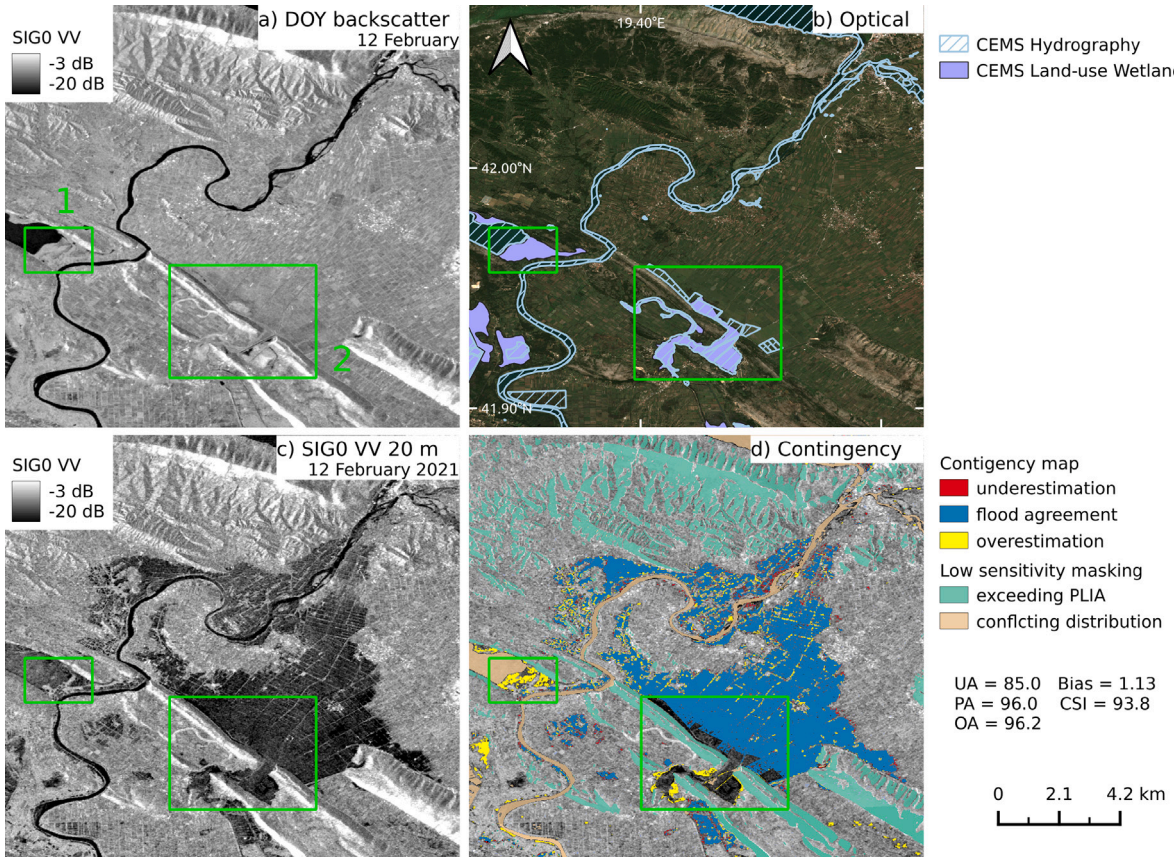

Fig. 5. Details on the flood event in Albania/EMSR501 at 2021/02/12 16:32 UTC. **(a)** Expected DOY backscatter from the harmonic model **(b)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) overlayed by ODM land-use layers **(c)** SIG0 VV 20 m flood scene **(d)** Contingency map.

Overall, the CSI exceeds a value of 70% for 10 out of 18 events. We see this as a positive result, considering that the algorithm's output was compared with a semi-automatic data approach. The OA values behave similarly to the CSI, generally being slightly higher, and in the case of smaller events with poorer performance, not dropping off as much as the CSI (also mentioned by Tupas et al., 2023). This can be particularly observed when there is predominantly an underestimation for an event (e.g. Finland/EMSR511). Without a suitable sampling approach, the OA is usually less effective for flood mapping validation, as the dominant no-flood class can mask misclassifications of the flood class. Therefore, the OA values of different events often turn out quite similar. However, it can be seen here that clear differences exist between the individual events and the OA is shown to be a valid accuracy metric for flood mapping applications.

The amount of over- or underestimation can be estimated by the UA and PA, while the overall trend can be shown by the use of the bias metric. Besides a number of events matching the reference closely (e.g. Germany/EMSR497, Malawi/EMSR561), a general trend towards underestimation can be observed (e.g. Iraq/EMSR551, Norway/EMSR557, Spain/EMSR555). Additionally, a smaller number of events show an overall overestimation (e.g. Albania/EMSR501, Germany/EMSR517, Madagascar/EMSR559). The limitations of the bias metric can be seen with the events Australia/EMSR554 and Australia/EMSR562, where the bias is close to 1 because of the roughly identical size of overestimation and underestimation. For a better understanding of the differences, we selected eight events from the total of 18 events and analysed them spatially and in greater detail (see Sections 4.1–4.4). For some events, the shown figures do not represent the whole area of the event, but a cutout to emphasize the differences of the two datasets.

4.1. Events showing overestimation

4.1.1. Madagascar/EMSR559

Heavy rainfall hit Madagascar and caused severe flooding at the beginning of the year 2022. The situation close to the city of Moramanga, in the Alaotra-Mangoro region, in eastern Madagascar is covered by the Sentinel-1 observation from the 27th of January (shown in Fig. 4). As the detected inundated area results in 113.8 ha, the event is considered small-scale relative to the other investigated events.

Overall, the tested flood mapping algorithm shows an overestimation compared to the reference data (details are given in Fig. 4d). One can see that areas with overestimation generally show a slightly higher backscatter (see Fig. 4c, especially in the green box 1) compared to the areas where classification and reference data agree on flooding. These pixels represent a close decision where our algorithm still assigns the pixels with a slightly higher backscatter to the flood class, while they are already classified as non-flood in the reference data. This conclusion is supported by the observation of an increased classification uncertainty of our result (Fig. 4e) in the affected areas, indicating a more similar probability of the pixel being part of the flood or non-flood class. The ambiguity may be explained by mixed water-land pixels along the outline of the flood body, with a mixed SAR backscatter signal. In a small-scale event like this, such outline pixels have more impact compared to an event of larger scale.

Fig. 4f shows the extension and reduction of the detected flood area due to the algorithm's majority filter applied in the post-processing. Although the filter successfully compensates for SAR speckle and local effects, large water bodies are reduced as well. Within the green box 2, an underestimation caused by the majority filter is shown. A similar small-scale situation and trend towards underestimation is present in the flood event Italy/EMSR496 (inundated area of 307.0 ha), which is not shown visually and in detail in this study.

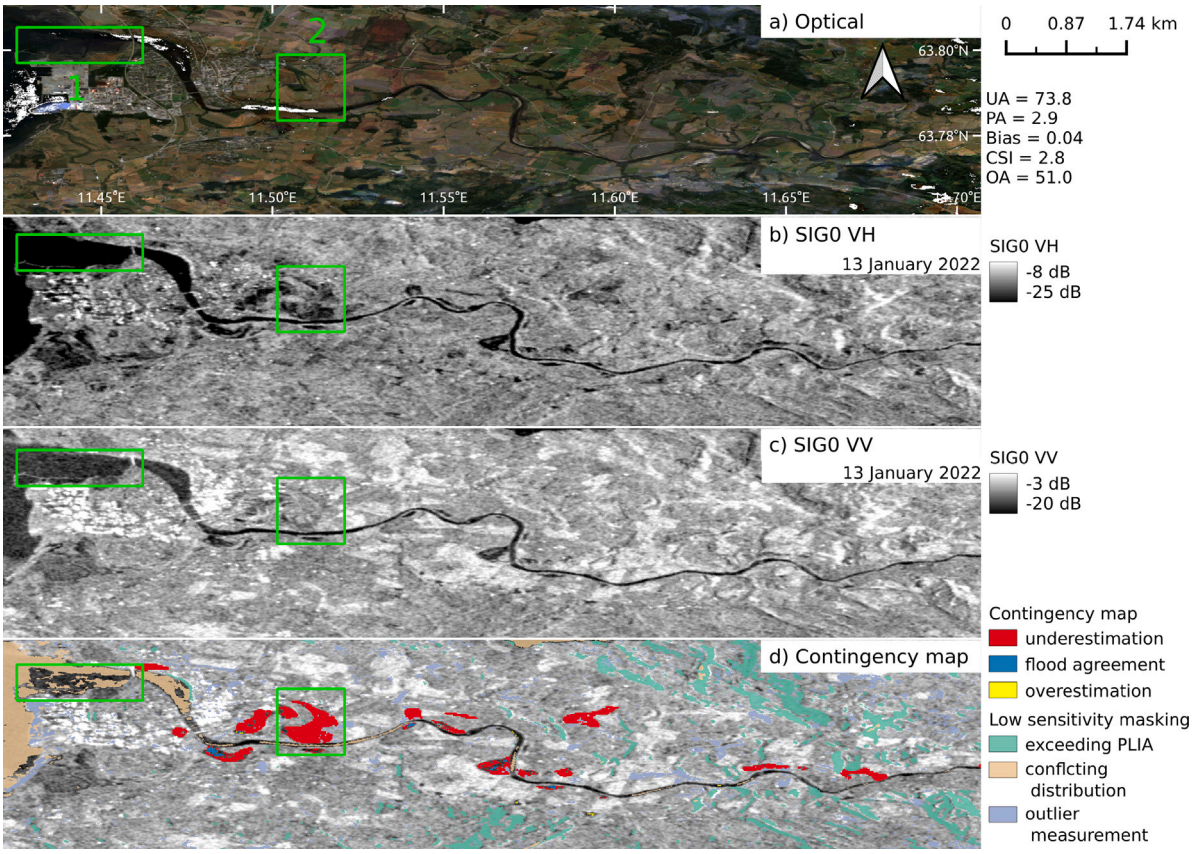

Fig. 6. Details on the flood event in Norway/EMSR557 from the 2022/01/13 16:47 UTC. **(a)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(b)** SIG0 VH 20 m flood scene **(c)** SIG0 VV 20 m flood scene **(d)** Contingency map.

4.1.2. Albania/EMSR501

The ODM was activated for the flood event in Albania on the 12th of January 2021 and Sentinel-1 covered the affected area on the same day. As this flood from heavy rains covered a comparatively large area (5038.7 ha), we categorize the event as large scale. Overall, our algorithm correctly classified the vast majority of pixels, resulting in a CSI of 93.8 . In Fig. 5a and 5b permanent water bodies on the top (Lake Skadar) and in the centre of the plot (Lake Šas) are visible, as well as River Buna crossing the area. The impact of wind-induced waves can be seen in Fig. 5c due to an increased backscatter over the permanent and inundated water bodies. However, there is no negative impact visible for this event.

Fig. 5d shows the contingency map overlaid by the low sensitivity mask. It is evident that sloped areas are accurately excluded as exceeding PLIA and most permanent water bodies are correctly masked due to conflicting distributions (details are given in Section 2.4). Nevertheless, the area emphasized by the green box 2 is not excluded by the low sensitivity mask, but by the hydrography mask (Fig. 5b) provided by the ODM. To mask this wetland area, water-like backscatter needs to be present in both the DOY backscatter of the harmonic model (Fig. 5a) and the actual backscatter observation (Fig. 5c), which is not the case. The water body's shape suggests an artificial construction that was potentially being built after the training period of the harmonic model (years 2019–2020). Furthermore, an error in the reference data cannot be excluded.

Some local overestimation is present within the wetland areas shown in the green boxes 1 and 2 of Fig. 5d. In Fig. 5b wetland areas indicated by the ODM are emphasized and contain the areas of overestimation well. Since the post-flood observation (Fig. 5c) shows low backscatter within the green boxes and the corresponding expected backscatter (Fig. 5a) is high in comparison, the algorithm's decision is plausible. One possible explanation for why these areas are missing in the ODM reference flood map, is the semi-automatic approach used by ODM. The areas might have been manually removed based on other data sources, as land cover or optical data.

4.2. Events showing underestimation

4.2.1. Norway/EMSR557

The combination of heavy rain and snow melt caused flooding in multiple locations of Norway on 11th of January 2022, which is why the ODM was activated. Fig. 6 shows the Sentinel-1 observation from the 13th of January 2022 and the performance of the flood mapping algorithm (Fig. 6d). The flooded area detected by ODM is with 259.9 ha relatively small.

Similar to the event Albania/EMSR501 (presented in Section 4.1.2), the existing permanent water bodies are expected to be masked due to conflicting distributions. The mouth of River Verdalselva (see the green box 1 in Fig. 6c) is more complex, since some areas are only inundated seasonally. As those areas were not inundated during the years 2019 and 2020, which is the time period the harmonic model is based on, the expected backscatter of the harmonic model is larger than the expected water backscatter. Consequently, those areas are not masked due to low sensitivity. Furthermore, it can be seen that the river is only partly visible in the Sentinel-1 observation at a spatial resolution of 20 m, and therefore it is not masked by the low sensitivity mask either.

The contingency map (Fig. 6d) shows that the majority of flooded areas are missed by the TU Wien flood mapping algorithm. A comparison of the Sentinel-1 observation in VV (Fig. 6c) and VH (Fig. 6b) polarization indicates a strong difference in backscatter over the flooded areas. For example, water-like low backscatter is found for VH, while this is not the case for VV. In general, the reference data seems to match the VH observation much more closely compared to VV. Since

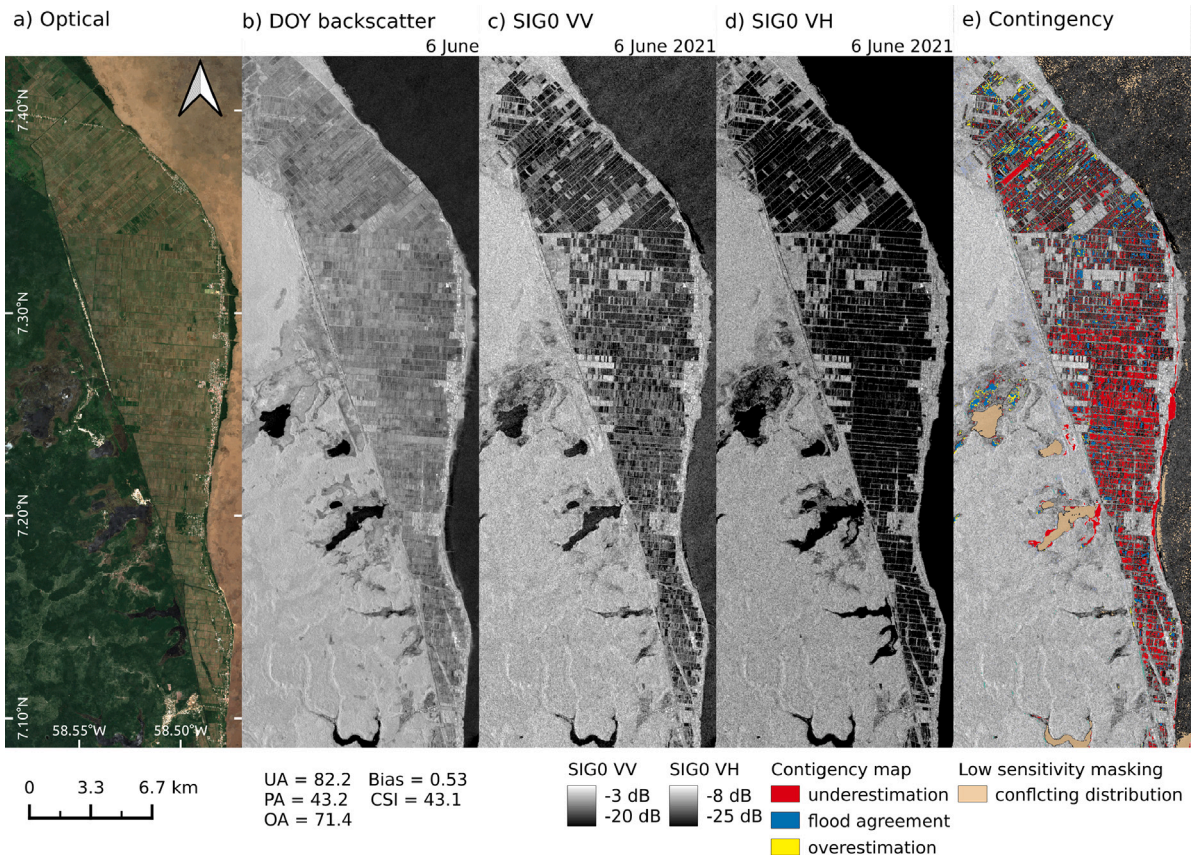

Fig. 7. Details on the flood event in Guyanna/EMSR514 from the 2021/06/06 22:08 UTC. **(a)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(b)** Expected DOY backscatter from the harmonic model **(c)** SIG0 VV 20 m flood scene **(d)** SIG0 VH 20 m flood scene **(e)** Contingency map.

the algorithm uses solely VV as an input, the underestimation can be attributed to the difference in the input data. The most likely reason for this difference between the polarizations is the presence of low vegetation. As stated by Twele et al. (2016) and Vreugdenhil et al. (2020), the backscatter from VH and VV polarization shows strong differences over vegetation. VV shows a higher penetration depth into the vegetation, which results in a higher contribution of the ground-vegetation interaction (e.g. double-bounce effect). If the vegetation is low and dominated by the signal coming from the ground, the double-bounce effect between water and vegetation could lead to an increased VV backscatter, while the VH backscatter is less affected and provides a less disturbed signal from the water surface. This assumption corresponds with the findings of Landuyt et al. (2020). When comparing the observations in VV and VH, it can be seen that VV is more affected by the impact of wind, visible as increased backscatter over permanent water bodies.

Other events, for which we assume that observations based on the VH polarization were used in the reference data (only a fraction of them is presented visually and detailed in this study), are: Belgium/EMSR518, Finland/EMSR511, Guyanna/EMSR514 (discussed in Section 4.2.2), Italy/EMSR548, Malawi/EMSR561 (discussed in Section 4.4.2) and Spain/EMSR555.

4.2.2. Guyanna/EMSR514

Fig. 7 presents the Sentinel-1 observation from the 6th of June 2021 covering one part of the corresponding flood event over Guyanna. According to the ODM, the flood water covered a total area of 11750.8 ha, which represents the largest event by area investigated by this study.

Similar to the observation of the event in Norway (described in Section 4.2.1), the Sentinel-1 observation of the event in Guyanna shows a similar difference between the VV (Fig. 7c) and VH (Fig. 7d) polarization. The difference in both events is obviously caused by vegetation, but in case of Guyanna the vegetation is located in an agricultural context. The resulting underestimation of our VV-based result can be seen in Fig. 7e. Agriculture presents a major challenge for C-band SAR flood mapping, since water-like low backscatter can occur due to multiple reasons during the plant cycle (more details are given in Section 4.3.1). In this case, the fields were flooded and the vegetation penetrates the water surface in some regions, preventing the specular reflection generally causing the low backscatter of water. A similar backscatter behaviour can be related to wind, which was present at the time, as an increased backscatter is observed over the ocean. Based on the radar data alone, one is not able to tell which of the two effects caused the underestimation. However, the observation in the VH polarization shows less impact of these effects.

The automatic low sensitivity masking again excluded permanent water bodies (visible in Fig. 7e). The chosen approach shows obvious limitations over the ocean, due to the turbulent water surface causing strong variability of the backscatter over the ocean. Consequently, the harmonic model (Fig. 7b) does not depict the ocean to be a homogeneous low backscatter area. Because of this, many pixels show no clear overlap between flood and non-flooded backscatter signature, preventing them of being masked.

4.3. Events showing both over- and underestimation

4.3.1. Germany/EMSR517

In July 2021 extreme rainfall caused exceptional flooding in Europe, and the ODM was activated for Belgium/EMSR518, Germany/EMSR517 and the Netherlands/EMSR520. On 15th of July 2021 Sentinel-1 covered the affected area in Western Germany and the eastern part of the area can be observed in Fig. 8. The comparison of classification result

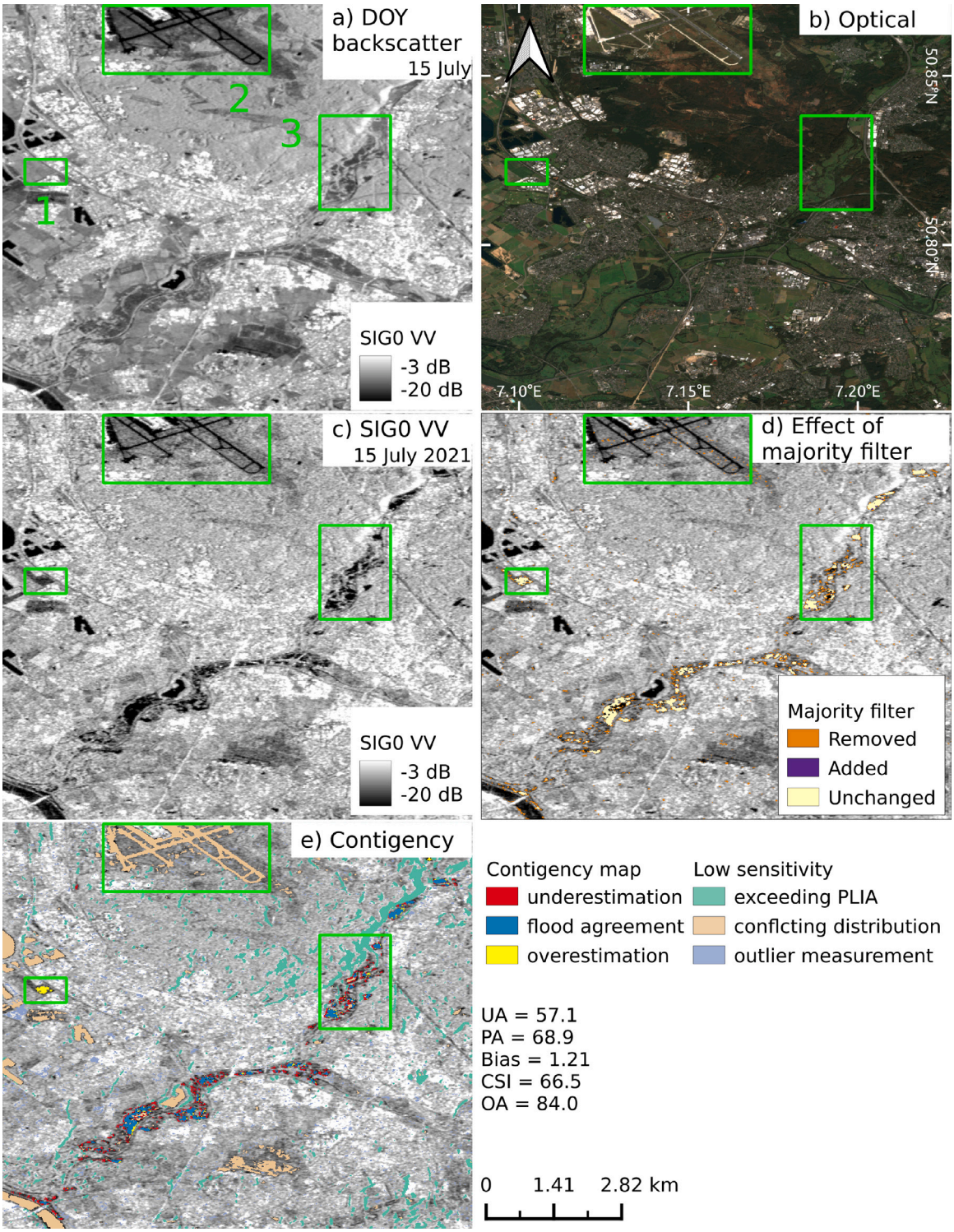

Fig. 8. Details on the flood event in Germany/EMSR517 from the 2021/07/15 05:50 UTC. **(a)** Expected DOY backscatter from the harmonic model **(b)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(c)** SIG0 VV 20 m flood scene **(d)** Effect of the applied majority filter **(e)** Contingency map.

and reference data revealed some over- and underestimation (Fig. 8e), mostly over vegetation.

Due to the complex interaction of the radar signal with the vegetation volume and the ground, situations of both high or low backscatter can be caused by change in vegetation (Vreugdenhil et al., 2018; Harfenmeister et al., 2019). If the backscatter decreases temporarily over vegetation, it is difficult to distinguish whether this is caused by changes in vegetation state or an actual flood is happening. This backscatter behaviour is generally caused by signal attenuation or smooth soil (e.g. after ploughing of agricultural fields). As the harmonic model represents the seasonality of the backscatter based on the years 2019 and 2020 (expected backscatter visible in Fig. 8a), low backscatter situations following the same seasonality would be expected by the model and a correct classification can be made. On the other hand, there are two conditions where misclassifications occur (confusion map in Fig. 8e): First, a low backscatter situation not occurring during the

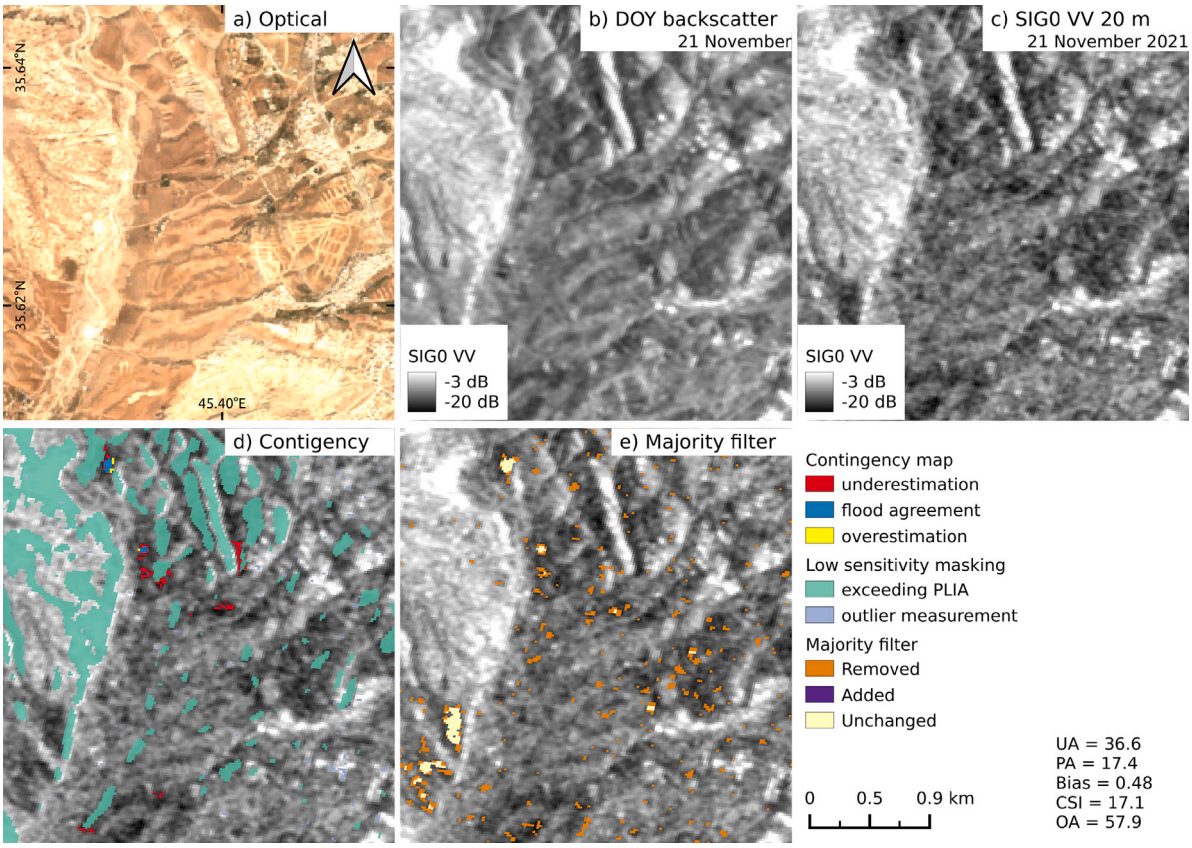

Fig. 9. Details on the flood event in Iraq/EMSR551 from the 2021/11/21 03:01 UTC. **(a)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(b)** Expected DOY backscatter from the harmonic model **(c)** SIG0 VV 20 m flood scene **(d)** Contingency map **(e)** Effect of majority filter.

expected time of the year or with a larger amplitude than expected, would lead to an overestimation (seen in the green box 1 in Fig. 8). Second, the harmonic model expects low backscatter, but an actual flood is happening, which leads to an underestimation because of the masking of conflicting distributions (see green box 3 in Fig. 8).

The flooded area of this event consists of a number of small, distinct polygons, instead of a continuous surface. To reduce the influence of the speckle effect, our algorithm always applies a majority filter onto the binary flood maps during the post-processing step (see Section 2.4 for more details). The impact of the majority filter can be seen in Fig. 8d. Unfortunately, it is currently not feasible for the algorithm to distinguish between false flood detections, small areas due to speckle, or those which are actually flooded. Consequently, the application of the majority filter increases the underestimation of our result.

The green box 2 in Fig. 8 covers the airport Cologne/Bonn. Since the flat tarmac surfaces of the runways show a low backscatter in both the expected backscatter and the post-flood observation of Sentinel-1, the area is correctly masked due to conflicting distributions. Similar observations can be made for the River Rhine in the bottom of the plot.

4.3.2. Iraq/EMSR551

The flood-prone district Sulaymaniyyah in Iraq was hit by a flood on 21st of November 2021. The detected flooded area make up 89.7 ha, which is the smallest event based on the affected area within the selection of this study. The optical satellite data (Fig. 9d) as well as the expected backscatter (Fig. 9a) show a mountainous region. Consequently, large parts of the area of interest are masked out, as the PLIA exceeds the valid range of the used water backscatter model. As the remaining flooded area consists of small, distinct polygons, they are removed by the majority filter during post-processing (see Fig. 9e) similarly to the event Germany/EMSR517 (see Section 4.3.1). Nevertheless, attention must also be paid to the advantages of the filter, as the influence of speckle could be significantly reduced here.

4.4. Events being close to reference

4.4.1. Germany/EMSR497

In February 2021 the River Rhine overflowed its banks within the district of Düsseldorf and caused moderate flooding. Within the selected area of the ODM, a flooded area of 4371.7 ha was affected, which suggests considering it as a large scale event. Overall, the TU Wien flood mapping algorithm's result matches the reference data closely (Fig. 10d). Next to the river, many permanent water bodies are present and being masked correctly (compare Fig. 10b and 10d). While the riparian area is mostly covered by built-up areas and pastures, the majority of the plot consists of agricultural areas. In contrast to the Germany/EMSR517 event (see Section 4.3.1), the presence of vegetation has not caused any significant over- or underestimation. As the flood happened during Central Europe's winter, the plant height is assumed to be low and no signal attenuation or ploughing caused unexpected low backscatter measurements.

4.4.2. Malawi/EMSR561

The tropical storm Ana caused a flood in Malawi, which was covered by Sentinel-1 on 26th January 2022. The event is considered to be large scale, as the detected flood area affected 2974.5 ha and the algorithm's result matches the reference data closely. Nevertheless, some regions of over- and underestimation are visible in the contingency map (see Fig. 11d). Firstly, a noise-like area of overestimation can be seen within the green box 1. As the area of interest is covering steppe and agricultural land is present as well, vegetation could be the reason for the overestimation. Alternatively, a false negative result of the CEM-ODM can be present in this part of the event. A similar issue is detected in the event Germany/EMSR517 (see Section 4.3.1). Secondly, areas of underestimation are found along the rivers Lukhubula, Mwamphanzi and Mapalera. Since the area next to these rivers is dominated by

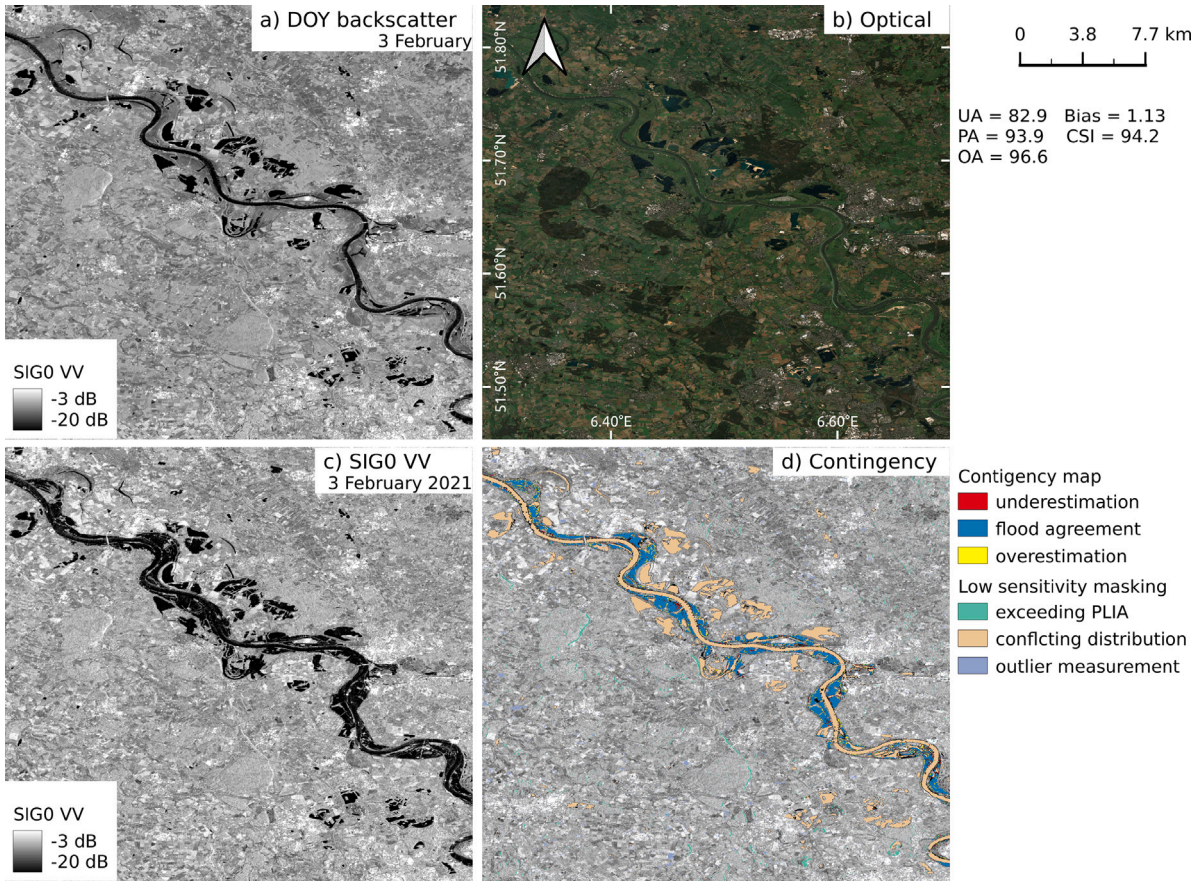

Fig. 10. Details on the flood event in Germany/EMSR497 from the 2021/02/03 05:49 UTC. **(a)** Expected DOY backscatter from the harmonic model **(b)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(c)** SIG0 VV 20 m flood scene **(d)** Contingency map.

shrubland and shows a higher backscatter in VV (Fig. 11d), while VH (Fig. 11c) measures a low backscatter, the underestimation most likely related to flooded vegetation. Similar issues appear and being described for the event Norway/EMSR557 and other events listed in Section 4.2.1. As for the event Albania/EMRS501 (see Section 4.1.2), we can observe a high CSI of 92.1 , despite some discussed deviations from the reference data. In both cases, the deviations do not have much impact due to the large scale of the overall event.

5. Conclusions

In this study, we evaluated our automatic Sentinel-1-based flood mapping algorithm by comparing its results to the semi-automatic results of the Copernicus Emergency Management Service (CEMS) on-demand mapping (ODM). Overall, we included 18 globally-distributed flood events between January 2021 and January 2022. For each event common performance metrics were retrieved and for eight events we performed a qualitative evaluation based on the spatial differences between our and the semi-automatic results.

The TU Wien flood mapping algorithm shows a robust performance, considering that it is an automatic algorithm and not using the VH polarization. An overview of the performance metrics of all events is given in Table 1. The best performing event is Malawi/EMSR561 with a UA of 93.7% and a PA of 93.0%. Calculating the median over all events, the UA results in 84.0%, while the median PA results in 72.9%. Other studies featuring an automatic approach presented values in an equal order of magnitude (UA /PA): 88%/65% (Chini et al., 2017), 94.3%/71.5 % (Clement et al., 2018), or 82.4-98.5% /83.6-92.6 % (Martinis et al., 2014). Generally, the causes of over- or underestimations can be traced back to the complex interaction of the radar signal with certain land-cover types (e.g. vegetation) as well as the impact of the algorithm's post-processing facing challenges under certain conditions (e.g. small non-connected flood surfaces). As no systematic overestimation or -masking is noticed, we conclude that the harmonic model estimates the expected backscatter for non-flooded conditions well. Furthermore, the usage of backscatter with a pixel-size of 20 m as an input to the algorithm shows no apparent disadvantages in the flood detections compared to 10 m data, while clearly reducing the required data volume and the impact of the speckle effect. Considering regional instead of global applications, the systematical utilization of 10 m pixel-size could nevertheless be of interest for automatic algorithms as well.

Some potential improvements for the TU Wien flood mapping algorithm are identified based on the detailed investigation of detected differences. The main cause of differences between classification and reference data is related to the occasional usage of VH polarized measurements in case of the reference data. Overall, we identified two conditions where VH influences the flood mapping result: wind impact (see Section 4.1.2, 4.2.1 or 4.2.2) and flooded vegetation (see Section 4.2.1 or 4.2.2). The VV backscatter is generally increased by the double-bounce effect present for flooded vegetation, as well as by the roughening of the surface due to wind. The interpretation of the tests sites showed that our algorithm performs well even when wind is present. However, to enhance performance in the presence of vegetation, one should consider incorporating VH in the future or capturing an increased VV. Since this step might lead to an increased false positive rate, thorough research must be conducted initially.

In general, large scale events show better performance metrics compared to small-scale events, although this bias is reduced by the applied sampling. This is because small deviations from the reference

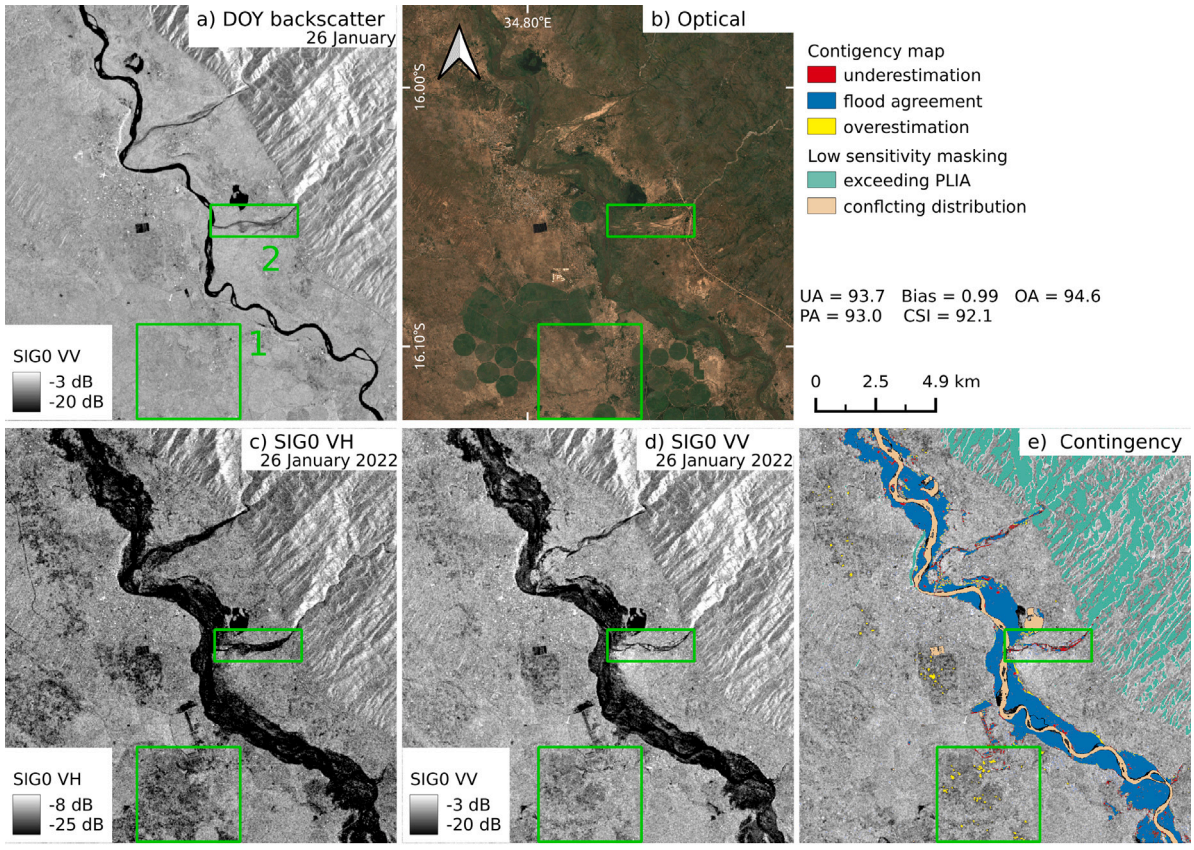

Fig. 11. Details on the flood event in Malawi/EMSR561 from the 2022/01/26 03:16 UTC. **(a)** Expected DOY backscatter from the harmonic model **(b)** ESA WorldCover Sentinel-2 TCC year 2021 (Zanaga et al., 2022) **(c)** SIG0 VH 20 m flood scene **(d)** SIG0 VV 20 m flood scene **(e)** Contingency map.

do not influence large scale events by the same relative impact as small-scale events. Furthermore, this evaluation identified the majority filter applied during the post-processing routine as a weakness of the current TU Wien algorithm implementation, which especially affects small-scale events (see Sections 4.1.1, 4.3.1 and 4.3.2). While the impact of the speckle effect is clearly reduced, the filter limits the size of unconnected water bodies that are classified as flood. Furthermore, the reduction of large water bodies and extension of the flooded area without being based on the backscatter data is considered as deterioration of the performance. Consequently, alternative post-processing routines such as applying a minimum mapping unit or a region growing approach are planned to be tested in future studies.

There are multiple reasons for low backscatter observations over vegetation, in particular over agricultural areas (e.g. signal attenuation, ploughing, frost). These reasons show a strong seasonal behaviour, but the harmonic model is not fully able to cover them (see Section 4.3.1), since the low backscatter situations emerge quickly and are hard to predict based on historic measurements. Consequently, some overestimation might appear if the harmonic model's expectation does not match the corresponding year, or wrong masking could be performed due to the same cause. To overcome this issue, the consideration of temporal and spatial patters and their differences compared to actual flooding should be studied. Therefore, a non-flooded reference with the ability to be updated in near-real-time is currently being investigated (Tupas et al., 2024).

The low sensitivity masking aims for excluding situations where the underlying assumptions of the applied statistical models are not met. Overall, the masking was effective across all areas that could be compared to external data, such as permanent water bodies and mountainous regions. Only the automatic masking of the ocean area performed poorly (see Sections 4.2.1 and 4.2.2), due to the rough surface and temporal variability of the water. Fortunately, many external datasets can be utilized to mask the ocean, which is why the masking can be supported easily and a dedicated optimization over the ocean is not deemed necessary.

While there are studies that have conducted a comparable quantitative evaluation, to our knowledge there is no study that has conducted a comparable qualitative evaluation of a specific flood mapping algorithm. Since test site selection was independent of the strengths and weaknesses of the algorithm, the evaluation is considered to be unbiased. Furthermore, the test sites covered multiple continents and climatic conditions, although the majority of sites are located in Europe. This is caused by the reliance on the CEMS on-demand mapping, which is a European service and the majority of activations happen on the very same continent. Further, the majority of flood events are of the fluvial type, whereas pluvial and coastal floods are relatively underrepresented. However, the included number and spatial distribution of test sites allowed an overview of the performance of the algorithm and provided suggestions for future improvements. Although some of the findings are specific to the algorithm tested (e.g. issues with the majority filter or harmonic model), others remain valid for other radar-based methods (e.g. impact of the VH polarization or challenges raised by vegetation). Given the valuable insights about the performance of our algorithm, we consider the study's design to be well suited for the evaluation of our automatic flood mapping algorithm evaluation.

CRediT authorship contribution statement

Florian Roth: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Mark Edwin Tupas:** Writing – review & editing, Validation, Software. **Claudio Navacchi:** Writing – review & editing, Software, Data curation. **Jie Zhao:** Writing – review & editing. **Wolfgang Wagner:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Bernhard Bauer-Marschallinger:** Writing – review & editing, Visualization,

Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Florian Roth reports financial support was provided by European Commission Joint Research Centre Ispra. Florian Roth reports financial support was provided by Austrian Research Promotion Agency. Florian Roth reports financial support was provided by TU Wien Bibliothek. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

TU Wien acknowledges co-funding from the project “Provision of an Automated, Global, Satellite-based Flood Monitoring Product for the Copernicus Emergency Management Service” (GFM), Contract No. 939866-IPR-2020 for the European Commission’s Joint Research Centre (EC-JRC) and the project “Preparing for Scaling Up Flood Monitoring Capabilities with ROSE-L and Sentinel-1 NG” (ScaleFloodS), Project No. 47817167, founded by the Austrian Research Promotion Agency (FFG). Further, the authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program. We like to thank the whole GFM consortium for their work on building a global flood mapping service.

Data availability

Data will be made available on request.

References

Ajmar, A., Boccardo, P., Broglia, M., Kucera, J., Giulio-Tonolo, F., Wania, A., 2017. Response to flood events: the role of satellite-based emergency mapping and the experience of the copernicus emergency management service, In: Molinari, D., Menoni, S., Ballio, F. (Eds.), First Edition In: Geophysical Monograph Series, Wiley, pp. 211–228. http://dx.doi.org/10.1002/9781119217930.ch14, https://agupubs.onlinelibrary.wiley.com/doi/10.1002/9781119217930.ch14.

Amitrano, D., Di Martino, G., Di Simone, A., Imperatore, P., 2024. Flood detection with SAR: A review of techniques and datasets. Remote. Sens. 16, 656. http://dx.doi.org/10.3390/rs16040656, https://www.mdpi.com/2072-4292/16/4/656.

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. Sens. 14, http://dx.doi.org/10.3390/rs14153673.

Bauer-Marschallinger, B., Sabel, D., Wagner, W., 2014. Optimisation of global grids for high-resolution remote sensing data. Comput. Geosci. 72, 84–93. http://dx.doi.org/10.1016/j.cageo.2014.07.005.

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 J. Sel. Top. Appl. Earth Obs. Remote. Sens. 15, 2023–2036. http://dx.doi.org/10.1109/JSTARS.2022.3152127, https://ieeexplore.ieee.org/document/9714780/.

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 Trans. Geosci. Remote Sens. 55, 6975–6988. http://dx.doi.org/10.1109/TGRS.2017.2737664.

Chini, M., Pelich, R., Li, Y., Hostache, R., Zhao, J., Di Mauro, C., Matgen, P., 2021. Sar-based flood mapping, where we are and future challenges. In: 2021 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, IEEE, pp. 884–886. http://dx.doi.org/10.1109/IGARSS47720.2021.9554975.

Clement, M.A., Kilsby, C., Moore, P., 2018. Multi-temporal synthetic aperture radar flood mapping using change detection. J. Flood Risk Manag. 11, 152–168. http://dx.doi.org/10.1111/jfr3.12303.

Colacicco, R., Refice, A., Nutricato, R., Bovenga, F., Caporusso, G., D’Addabbo, A., La Salandra, M., Lovergine, F.P., Nitti, D.O., Capolongo, D., 2024. High-resolution flood monitoring based on advanced statistical modeling of sentinel-1 multi-temporal stacks. Remote. Sens. 16, 294. http://dx.doi.org/10.3390/rs16020294, https://www.mdpi.com/2072-4292/16/2/294.

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S., Cosh, M., Dunbar, R., Dang, L., Pashaian, L., et al., 2017. Validation of smap surface soil moisture products with core validation sites. Remote Sens. Environ. 191, 215–231. http://dx.doi.org/10.1016/j.rse.2017.01.021.

Garg, S., Dasgupta, A., Motagh, M., Martinis, S., Selvakumaran, S., 2024. Unlocking the full potential of sentinel-1 for flood detection in arid regions. Remote Sens. Environ. 315, 114417. http://dx.doi.org/10.1016/j.rse.2024.114417, https://linkinghub.elsevier.com/retrieve/pii/S0034425724004437.

Ghosh, B., Garg, S., Motagh, M., Martinis, S., 2024. Automatic flood detection from sentinel-1 data using a nested unet model and a NASA benchmark dataset. PFG – J. Photogramm. Remote. Sens. Geoinf. Sci. http://dx.doi.org/10.1007/s41064-024-00275-1, https://link.springer.com/10.1007/s41064-024-00275-1.

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., et al., 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors? Remote Sens. Environ. 244, 111806. http://dx.doi.org/10.1016/j.rse.2020.111806.

Harfenmeister, K., Spengler, D., Weltzien, C., 2019. Analyzing temporal and spatial characteristics of crop parameters using sentinel-1 backscatter data. Remote. Sens. 11 (1569), http://dx.doi.org/10.3390/rs11131569.

Huang, M., Jin, S., 2020. Rapid flood mapping and evaluation with a supervised classifier and change detection in shouguang using sentinel-1 sar and sentinel-2 optical data. Remote. Sens. 12 (2073), http://dx.doi.org/10.3390/rs12132073.

Inès, J.B., Annett, W., Simone, D., 2020. Manual for cems-rapid mapping products. http://dx.doi.org/10.2760/29876, JRC121741.

Jackson, T.J., Cosh, M.H., Bindlish, R., Starks, P.J., Bosch, D.D., Seyfried, M., Goodrich, D.C., Moran, M.S., Du, J., 2010. Validation of advanced microwave scanning radiometer soil moisture products. IEEE Trans. Geosci. Remote Sens. 48, 4256–4272. http://dx.doi.org/10.1109/TGRS.2010.2051035.

Katiyar, V., Tamkuan, N., Nagai, M., 2021. Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. Remote. Sens. 13 (2334), http://dx.doi.org/10.3390/rs13122334.

Landuyt, L., Van Coillie, F.M., Vogels, B., Dewelde, J., Verhoest, N.E., 2021. Towards operational flood monitoring in flanders using sentinel-1. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 11004–11018. http://dx.doi.org/10.1109/JSTARS.2021.3121992.

Landuyt, L., Van Wesemael, A., Schumann, G.J.P., Hostache, R., Verhoest, N.E.C., Van Coillie, F.M.B., 2019. Flood mapping based on synthetic aperture radar: An assessment of established approaches. IEEE Trans. Geosci. Remote Sens. 57, 722–739. http://dx.doi.org/10.1109/TGRS.2018.2860054, https://ieeexplore.ieee.org/document/8432448/.

Landuyt, L., Verhoest, N.E., Van Coillie, F.M., 2020. Flood mapping in vegetated areas using an unsupervised clustering approach on sentinel-1 and-2 imagery. Remote. Sens. 12 (3611), http://dx.doi.org/10.3390/rs12213611.

Landwehr, T., Dasgupta, A., Waske, B., 2024. Towards robust validation strategies for EO flood maps. Remote Sens. Environ. 315, 114439. http://dx.doi.org/10.1016/j.rse.2024.114439, https://linkinghub.elsevier.com/retrieve/pii/S0034425724004656.

Li, Y., Martinis, S., Wieland, M., Schlaffer, S., Natsuaki, R., 2019. Urban flood mapping using SAR intensity and interferometric coherence via Bayesian network fusion. Remote. Sens. 11, 2231. http://dx.doi.org/10.3390/rs11192231, https://www.mdpi.com/2072-4292/11/19/2231.

Martinis, S., Twele, A., Voigt, S., Strunz, G., 2014. Towards a global sar-based flood mapping service. In: 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, pp. 2355–2358. http://dx.doi.org/10.1109/IGARSS.2014.6946944.

Moya, L., Endo, Y., Okada, G., Koshimura, S., Mas, E., 2019. Drawback in the change detection approach: False detection during the 2018 western Japan floods. Remote. Sens. 11, 2320. http://dx.doi.org/10.3390/rs11192320, https://www.mdpi.com/2072-4292/11/19/2320.

Nightingale, J., Schaepman-Strub, G., Nickeson, J., Baret, F., 2010. Assessing Satellite-Derived Land Product Quality for Earth System Science Applications: Results from the CEOS LPV Sub-Group. Rome, Italy, https://library.wur.nl/WebQuery/wurpubs/402578.

Ohki, M., Yamamoto, K., Tadono, T., Yoshimura, K., 2020. Automated processing for flood area detection using ALOS-2 and hydrodynamic simulation data. Remote. Sens. 12, 2709. http://dx.doi.org/10.3390/rs12172709, https://www.mdpi.com/2072-4292/12/17/2709.

on Climate Change (IPCC), I.P., 2023. Weather and Climate Extreme Events in a Changing Climate. Cambridge University Press, pp. 1513–1766.

Pulvirenti, L., Chini, M., Pierdicca, N., Boni, G., 2016. Use of SAR data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence. IEEE Trans. Geosci. Remote Sens. 54, 1532–1544. http://dx.doi.org/10.1109/TGRS.2015.2482001, http://ieeexplore.ieee.org/document/7299665/.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., 2008. Hand, a new terrain descriptor using srtm-dem: Mapping terra-firme rainforest environments in amazonia. Remote Sens. Environ. 112, 3469–3481. http://dx.doi.org/10.1016/j.rse.2008.03.018.

Roth, F., Tupas, M.E., Reuß, F., Bauer-Marschallinger, B., Navacchi, C., Wagner, W., 2023. Sentinel-1 global harmonic parameters: A seasonal model for flood mapping and more. http://dx.doi.org/10.48436/x8p2j-1tj74.

Salamon, P., Mctlormick, N., Reimer, C., Clarke, T., Bauer-Marschallinger, B., Wagner, W., Martinis, S., Chow, C., Böhnke, C., Matgen, P., et al., 2021. The new, systematic global flood monitoring product of the copernicus emergency management service. In: 2021 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, IEEE, pp. 1053–1056. http://dx.doi.org/10.1109/IGARSS47720.2021.9554214.

Schlaffer, S., Matgen, P., Hollaus, M., Wagner, W., 2015. Flood detection from multi-temporal sar data using harmonic analysis and change detection. Int. J. Appl. Earth Obs. Geoinf. 38, 15–24. http://dx.doi.org/10.1016/j.jag.2014.12.001.

Schumann, G.J., 2019. The need for scientific rigour and accountability in flood mapping to better support disaster response. Hydrol. Process. 33, 3138–3142. http://dx.doi.org/10.1002/hyp.13547, https://onlinelibrary.wiley.com/doi/10.1002/hyp.13547.

Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30, 5243–5272. http://dx.doi.org/10.1080/01431160903131000.

Stephens, E., Schumann, G., Bates, P., 2014. Problems with binary pattern measures for flood model evaluation. Hydrol. Process. 28, 4928–4937. http://dx.doi.org/10.1002/hyp.9979.

Tellman, B., Sullivan, J., Kuhn, C., Kettner, A., Doyle, C., Brakenridge, G., Erickson, T., Slayback, D., 2021. Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80–86. http://dx.doi.org/10.1038/s41586-021-03695-w.

Tsendbazar, N., Herold, M., Li, L., Tarko, A., De Bruin, S., Masiliunas, D., Lesiv, M., Fritz, S., Buchhorn, M., Smets, B., Van De Kerchove, R., Duerauer, M., 2021. Towards operational validation of annual global land cover maps. Remote Sens. Environ. 266, 112686. http://dx.doi.org/10.1016/j.rse.2021.112686, https://linkinghub.elsevier.com/retrieve/pii/S0034425721004065.

Tupas, M.E., Roth, F., Bauer-Marschallinger, B., Wagner, W., 2023. An intercomparison of sentinel-1 based change detection algorithms for flood mapping. Remote. Sens. 15, 1200. http://dx.doi.org/10.3390/rs15051200, https://www.mdpi.com/2072-4292/15/5/1200.

Tupas, M.E., Roth, F., Bauer-Marschallinger, B., Wagner, W., 2024. Assessment of time-series-derived no-flood references for sar-based bayesian flood mapping. GIScience Remote. Sens. http://dx.doi.org/10.1080/15481603.2024.2427304.

Twele, A., Cao, W., Plank, S., Martinis, S., 2016. Sentinel-1-based flood mapping: a fully automated processing chain. Int. J. Remote Sens. 37, 2990–3004. http://dx.doi.org/10.1080/01431161.2016.1192304.

Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Pfeil, I., Dorigo, W., Wagner, W., 2020. Sentinel-1 cross ratio and vegetation optical depth: A comparison over europe. Remote. Sens. 12 (3404), http://dx.doi.org/10.3390/rs12203404.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., Strauss, P., 2018. Sensitivity of sentinel-1 backscatter to vegetation dynamics: An austrian case study. Remote. Sens. 10 (1396), http://dx.doi.org/10.3390/rs10091396.

Wagner, W., Bauer-Marschallinger, B., Navacchi, C., Reuß, F., Cao, S., Reimer, C., Schramm, M., Briese, C., 2021. A sentinel-1 backscatter datacube for global land monitoring applications. Remote. Sens. 13 (4622), http://dx.doi.org/10.3390/rs13224622.

Wania, A., Joubert-Boitat, I., Dottori, F., Kalas, M., Salamon, P., 2021. Increasing timeliness of satellite-based flood mapping using early warning systems in the copernicus emergency management service. Remote. Sens. 13 (2114), http://dx.doi.org/10.3390/rs13112114.

Westerhoff, R., Kleuskens, M., Winsemius, H., Huizinga, H., Brakenridge, G., Bishop, C., 2013. Automated global water mapping based on wide-swath orbital synthetic-aperture radar. Hydrol. Earth Syst. Sci. 17, 651–663. http://dx.doi.org/10.5194/hess-17-651-2013.

Wu, X., Zhang, Z., Xiong, S., Zhang, W., Tang, J., Li, Z., An, B., Li, R., 2023. A near-real-time flood detection method based on deep learning and SAR images. Remote. Sens. 15, 2046. http://dx.doi.org/10.3390/rs15082046, https://www.mdpi.com/2072-4292/15/8/2046.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O., 2022. Esa worldcover 10 m 2021 v200. http://dx.doi.org/10.5281/zenodo.7254221.

Zhao, Y., Gong, P., Yu, L., Hu, L., Li, X., Li, C., Zhang, H., Zheng, Y., Wang, J., Zhao, Y., Cheng, Q., Liu, C., Liu, S., Wang, X., 2014. Towards a common validation sample set for global land-cover mapping. Int. J. Remote Sens. 35, 4795–4814. http://dx.doi.org/10.1080/01431161.2014.930202, https://www.tandfonline.com/doi/full/10.1080/01431161.2014.930202.

Zhao, J., Li, Y., Matgen, P., Pelich, R., Hostache, R., Wagner, W., Chini, M., 2022. Urban-aware U-net for large-scale urban flood mapping using multitemporal sentinel-1 intensity and interferometric coherence. IEEE Trans. Geosci. Remote Sens. 60, 1–21. http://dx.doi.org/10.1109/TGRS.2022.3199036, https://ieeexplore.ieee.org/document/9857936/.