ISSN 1831-9424

European Commission

Copernicus Emergency Management Service

Global Flood Monitoring (GFM)

Pre-operational Product and Service Quality Assessment Report

Prepared by the EXPERT FLOOD MONITORING ALLIANCE

Joint Research Centre

EUR 31425 EN

Copernicus Emergency Management Service

This publication is a Technical Report on the Global Flood Monitoring (GFM) product of the Copernicus Emergency Management Service, which is operated by an international consortium led by the Earth Observation Data Centre for Water Resources Monitoring GmbH (EODC), under a Framework Contract with the Joint Research Centre (JRC), the European Commission's science and knowledge service. The aim of the report is to provide a product and service quality assessment for the pre-operational phase of the GFM Product. In order to assess the thematic accuracy of the output layers of flood information provided by the GFM Product, a systematic check has been performed of the product quality and consistency with the Technical Specifications. This report describes the related quality assurance activities and achieved results, based on an initial analysis of the performance of the GFM Product for a representative set of Use Cases of worldwide flood events.

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC131351

EUR 31425 EN

PDF ISBN 978-92-76-99709-2 ISSN 1831-9424 doi:10.2760/362585 KJ-NA-31-425-EN-N

Luxembourg: Publications Office of the European Union, 2023

How to cite this report: Seewald, M., Riffler, M., Ralser, S., Innerbichler, F., Dulleck, B., Leitner, A., Gruber, C., Schleicher, C., Ziselsberger, M., McCormick, N., Salamon, P., *Global Flood Monitoring (GFM) – Pre-operational Product and Service Quality Assessment Report*. Publications Office of the European Union, Luxembourg, 2023. doi:10.2760/362585.

Contents

Abstract

This report presents the first results of the independent quality assessment of the new Global Flood Monitoring (GFM) Product of the Copernicus Emergency Management Service, based on a three-month period of pre-operational testing. The initial (i.e. Phase I) quality assessment focused on the GFM Product's main output layer - namely Observed Flood Extent - which provides a continuous monitoring of floods worldwide. The second (i.e. Phase II) quality assessment also included the validation of the GFM Product output layer Reference Water Mask, which delineates permanent and seasonal water. In the Phase II quality assessment, the general plausibility of the GFM Product output layer Exclusion Mask, which delineates areas where SAR-based water mapping is not feasible (due to a variety of static ground surface characteristics), was also considered.

1 Introduction

The Global Flood Monitoring (GFM) Product of the Copernicus Emergency Management Service[1] provides a continuous monitoring of worldwide flood events, by immediately processing and analysing in near real-time (NRT) all incoming Sentinel-1 Synthetic Aperture Radar (SAR) satellite imagery. Key elements of the GFM Product are the use of a historical time-series (or "data cube") of SAR backscatter intensity, enabling high product timeliness, and implementation of an ensemble method comprising three independent, state-of-the-art SAR-based water and flood mapping algorithms, in order to improve the robustness and accuracy of the flood and water extent maps and to build a high degree of redundancy into the service (Salamon et al., 2021; Matgen et al., 2020; Wagner et al., 2020). A brief technical overview of the GFM Product is provided in Section 2 below.

In accordance with the Technical Specifications (TS) for implementing and operating the GFM Product (European Commission, 2020), the technical and scientific quality of the NRT product generation and product access and dissemination is ensured through a well-defined product and service quality assurance (QA) procedure. Central to the QA procedure is a set of **Key Performance Indicators** (KPIs), used for the quarterly monitoring and reporting of various aspects of the service and product delivery performance, and providing the basis for the **annual product and service quality assessment reports** of the GFM Product. The annual quality assessment reports will provide a detailed quality assessment of the NRT product generation, and of the processed Sentinel-1 archive of flood events (see Section 2 below).

In addition to the annual product and service quality assessment report that must be prepared for each year of the implementation phase of the GFM Product, a **pre-operational product and service quality assessment report** must also be prepared for the set-up phase of the GFM Product, focused specifically on the **validation and assessment of the flood mapping algorithm** used in the NRT product generation. The present document constitutes the pre-operational product and service quality assessment report of the GFM Product.

As required by the TS, the KPI for **Product Quality** - which specifically refers to the thematic accuracy of the output layer Observed Flood Extent - is measured through the **Critical Success Index** (CSI), which is commonly used to evaluate the performance of dichotomous (Yes / No) classifications of low-frequency events (e.g. weather forecasts), and combines "hits", "false alarms" and "misses" into one score. As specified in the TS, the GFM Product is expected to have a Product Quality of **at least 70-80%**, as measured by the CSI.

According to the TS, the GFM Product must operationally deliver, in NRT, Sentinel-1 SAR-based output layers of flood information that are **of a high technical quality comparable with the best that can be obtained**, given the current state of research and production possibilities. As stated in the TS, the thematic accuracy of the GFM Product must be validated and assessed through regular offline interpretations **of the same Sentinel-1 scene for selected flood events**, based on comparison of the GFM's Observed Flood Extent with an appropriate sample of reference sites derived from automated and offline interpretation of selected Sentinel-1

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

images by experts, and using different data sources if available. Furthermore, the GFM Product validation must be performed for various Sentinel-1 datasets acquired in **different environments and various geographic locations** throughout the world.

The quality assessment of the pre-operational version of the GFM Product was carried out for two reporting periods: an initial (Phase I) accuracy assessment (completed in October 2021), and a later (Phase II) accuracy assessment (completed in March 2022). For the Phase I assessment, only the quality of the output layer **Observed Flood Extent** was considered, while for the Phase II assessment, the quality of the **Reference Water Mask** and **Exclusion Mask** was also analysed. It should be noted that during the set-up phase of the GFM Product, technical improvements were continuously made to the SAR-based flood and water detection methodology. However, it was not within the scope of the pre-operational product and service quality assessment report to consider the effects on thematic accuracy of specific improvements in the pre-operational GFM Product methodology.

In accordance with the TS, the quality assessment that is described in this report included both systematic automated and planned offline quality checks, performed continuously and considering all aspects of the GFM Product output layers, from data ingestion to data processing, data delivery, in addition to the thematic accuracy of the GFM Product.

The thematic accuracy assessment includes the two main output layers of the GFM Product:

a. The GFM Product output layer **Observed Flood Extent**, which provides continuous monitoring of floods worldwide, based on the processing and analysis in near real-time of all incoming Sentinel-1 SAR satellite images, acquired in Interferometric Wide Swath (IW) mode and provided as Ground Range Detected (GRD) products.

b. The GFM Product output layer **Reference Water Mask**, which shows pixels that are classified as water, both permanent and seasonal, based on Sentinel-1 SAR mean backscatter intensity over a two-year time-period.

As mentioned earlier, the quality assessment includes appropriate procedures for systematically assessing the product quality (i.e. thematic accuracy) based on a representative set of Use Cases of worldwide flood events. The derived analysis protocol for the GFM Product output layer Observed Flood Extent is designed to estimate objectively the accuracy, based on independent sample data, and to report on the service output quality by means of the proposed KPI for **Product Quality**, measured by the Critical Success Index (CSI).

It should be noted that the employed QA procedures follow the main principles for any validation exercise, which are supported by the INSPIRE directive that describes standard Implementing Rules (IR) to be adopted for Metadata, Data Specifications, Network Services, Data, and Service Sharing and Monitoring and Reporting[2], the GEO QA4EO guidelines describing the general principles for the validation and verification of Earth Observation products, and the principles of the CEOS Land Product Validation (LPV) group which has also defined the principles for validation activities in agreement with INSPIRE and QA4EO.

[2] http://inspire.jrc.ec.europa.eu/

The remainder of this pre-operational product and service quality assessment report is structured as follows:

- Section 2 provides a brief technical description of the GFM Product, including the main output layers of global flood information, the underlying state-of-the-art SAR-based flood mapping algorithms, and highlighting specific aspects designed to enhance product quality, i.e. the combination of the three flood mapping algorithms in an ensemble approach, and application of an Exclusion Mask.

- Section 3 describes the procedure for assessing the thematic quality of the output layer Observed Flood Extent, including how the reference dataset is generated, the analysed KPIs, and the validation strategy. As mentioned above, the quality assessment has been performed within two reporting periods (Phases I and II), and based on a sample of Use Cases of worldwide flood events, covering a diverse set of environmental conditions.

- Section 4 describes the initial (Phase I) accuracy assessment, including an overview of the six Use Cases and the detailed accuracy results.

- Section 5 describes the Phase II accuracy assessment that was performed for an additional three Use Cases, and the obtained results.

- Finally, the main results of the quality assessment of the pre-operational version of the GFM Product are discussed in Section 6.

The GFM Product has been developed and implemented under a Framework Contract with the European Commission's Joint Research Centre (European Commission, 2020), by an international consortium (called the "**Expert Flood Monitoring Alliance**") consisting of six partners:

- EODC (Earth Observation Data Centre for Water Resources Monitoring GmbH);
- GeoVille (GeoVille Information Systems and Data Processing GmbH);
- TU Wien (Technische Universität Wien);
- DLR (the German Aerospace Centre / Deutsches Zentrum für Luft- und Raumfahrt e.V.);
- LIST (Luxembourg Institute for Science and Technology);
- CIMA (Centro Internazionale in Monitoraggio Ambientale Research Foundation).

2 Technical overview of the GFM Product

The Global Flood Monitoring (GFM) Product of the Copernicus Emergency Management Service is an automated, global, flood monitoring system that provides a continuous (i.e. all-weather, day-and-night), systematic monitoring of all major global flood events, in near real-time (NRT), based on the latest Sentinel-1 Synthetic Aperture Radar (SAR) satellite images.

For each newly acquired Sentinel-1 SAR satellite image, the GFM Product provides 10 output layers of worldwide flood-related information, which are shown in Table 1. Central to the GFM Product are three state-of-the-art algorithms for the SAR-based detection and delineation of flooded areas, which were developed by members of the GFM consortioum (i.e. LIST, DLR, TUW), and which are briefly described in Table 2.

In order to ensure optimal accuracy of the derived flood (and water) extent maps, and to build a high degree of redundancy into the production service, the GFM Product deploys the three state-of-the-art flood mapping algorithms in an "**ensemble**" approach, whereby an area (i.e. image pixel) is considered to be flooded if (a) it is classified as flooded by **at least two of the three algorithms**, in the normal case when all three algorithms produce a result; (b) it is classified as flooded by **two algorithms**, in the exceptional case when only two of the three algorithms produce a result. (Note that in the pre-operational version of the GFM Product, which is the subject of this report, a slightly different ensemble approach was used: in those exceptional cases when only two of the three algorithms produce a result, the classification result of the algorithm with the **lower classification uncertainty** was selected).

In order to optimize further the quality of the results of the GFM Product, as can be seen in Table 1 an **Exclusion Mask** is used to exclude those areas where SAR-based water (and flood) detection is not technically feasible. The GFM Exclusion Mask is created by combining global information layers delineating the following ground surface characteristics:

1. **No sensitivity areas** (e.g. urban areas, dense vegetation), where Sentinel-1 SAR is not sensitive to flooding (or any other type of change) of the ground surface.
2. Water look-alikes (e.g. flat impervious areas, sand surfaces), which are indistinguishable from flooded areas due to a **low backscatter** signature.
3. Areas with strong topography (and low probability of flood occurrence), where the Sentinel-1 signals are affected by **topographic distortions**.
4. **Radar shadows** cast by mountains, high vegetation canopies or man-made structures.
5. Areas with **low coverage** (revisit frequency) of Sentinel-1 observations, where there is an inadequate historical time-series of SAR data available.

Finally, as specified in the Technical Specifications, in addition to the NRT generation of the 10 output layers of flood-related information, the GFM Product must also be used to generate a processed archive of worldwide observed flood events and water extent, from 1 January 2015 (until the start of the implementation of the NRT flood monitoring).

Full technical details on the GFM Product are provided on-line in the Product Definition Document (https://extwiki.eodc.eu/GFM/PDD).

Table 1: The GFM Product output layers of global flood-related information, generated in near real-time based on Sentinel-1 SAR satellite imagery.

#	GFM PRODUCT OUTPUT LAYER	DESCRIPTION AND NOTES
1	Observed flood Extent	Flooded areas mapped by applying the GFM ensemble flood mapping algorithm to the latest Sentinel-1 images of SAR backscatter intensity.
2	Observed water extent	Open and calm water mapped as the union of the observed flood extent and the reference water mask.
3	Reference water mask	Normal (i.e. permanent and seasonal) water mapped by applying the GFM ensemble water mapping algorithm to a historical time-series of Sentinel-1 images of SAR backscatter intensity.
4	Exclusion mask	Areas where SAR-based water mapping is not technically feasible, due to no sensitivity (e.g. urban areas, dense vegetation), low backscatter (e.g. flat impervious areas, sandy surfaces), topographic distortions, radar shadows, or low coverage of Sentinel-1.
5	Likelihood values	Estimated likelihood of flood classification, for all areas outside the exclusion mask.
6	Advisory flags	Flags indicating potential reduced quality of flood mapping, due to prevailing environmental conditions (e.g. wind, ice, snow, dry soil), or degraded input data quality due to signal interference from other SAR missions;
7	Footprint and metadata	Image boundaries of the Sentinel-1 data used, and in addition information on the “metadata”, i.e. information on the acquisition parameters of the Sentinel-1 data used.
8	Schedule	Next scheduled Sentinel-1 data acquisition.
9	Affected population	Number of people in flooded areas, mapped by a spatial overlay of observed flood extent and gridded population, from the Copernicus GHSL project.
10	Affected Landcover	Land cover / use (e.g. artificial surfaces, agricultural areas) in flooded areas, mapped by a spatial overlay of observed flood extent and the Copernicus GLS land cover.

Table 2: Overview of the GFM Product’s three underlying state-of-the-art algorithms for SAR-based flood mapping.

GFM ALGORITHM	MAIN TECHNICAL FEATURES	SCIENTIFIC REFERENCE
Algorithm 1 (LIST)	– Hierarchical split-based approach enabling re-calibration of parameters in NRT based on the most recent pair of S-1 images. – Uses a highly innovative sequence of hierarchical image splitting, statistical modelling and region-growing to delineate and classify areas that changed their flooding-related backscatter response between two image acquisitions from the same orbits.	Chini et al. (2017)
Algorithm 2 (DLR)	– Fuzzy logic-based approach enabling a post-classification and region-growing, taking advantage of topography-derived indices in addition to SAR backscatter.	Martinis et al. (2015)
Algorithm 3 (TUW)	– A fully automatic, pixel-based flood extent mapping workflow which exploits per-pixel full Sentinel-1 signal history in a data cube (time-series) of backscatter measurements; – Enables a very fast and scalable production of flood and water extent maps through pre-computed global parameters at high quality.	Bauer-Marschallinger et al. (2022)

3 Quality assessment methodology for the pre-operational GFM Product

In accordance with the Technical Specifications (TS) for implementing and operating the GFM Product (European Commission, 2020), the Key Performance Indicator (KPI) for **Product Quality** (i.e. thematic accuracy), which is measured through the Critical Success Index (CSI), is assessed through regular offline interpretations of the GFM production data (i.e. Sentinel-1) for selected global flood events (Use Cases), and an offline interpretation of selected events using other data sources (optical, radar, and in-situ measurements where available). The quality assessment must be based on an appropriate global sample of reference sites that are representative, in space and time, of the different scientific challenges to be addressed by the GFM Product's SAR-based flood mapping algorithms.

The GFM Product output layers to be validated are the Observed Flood Extent and Reference Water Mask, which are the two main outputs generated by the GFM Product. Observed Flood Extent indicates flooded areas mapped in near real-time from Sentinel-1 satellite imagery, and is derived from an ensemble approach integrating three algorithms developed independently by three leading research teams. The three algorithms run in parallel and have access to the same pre-processed Sentinel-1 input data. The three generated flood and water extent maps are systematically combined into a single product (i.e. "consensus map"). Briefly, each pixel in the combined product is accepted as flooded when a majority rule classifies it as flooded, and then subtracting the permanent or seasonal water grid-cells.

The GFM Product output layer Reference Water Mask, also validated for a defined area covering the selected flood events, delineates the permanent and seasonal water bodies. The Reference Water Mask is produced once every year and considers Sentinel-1 data from the two previous years. The permanent water extent mapping is based on the mean backscatter of all Sentinel-1 data from a reference period of two years. The seasonal reference water mapping uses as input the median backscatter of all Sentinel-1 data from a given month over the same reference period.

3.1 Validation procedure for the pre-operational GFM Product

The validation (i.e. thematic accuracy assessment) procedure takes the following workflows and related aspects into consideration: **(1) Sampling design** specifying the protocol for selecting locations at which the response (or reference) data are to be acquired; **(2) Response design** specifying the protocol used to determine the reference or ground condition label(s) and the definition of the agreement for comparing the map label(s) to the reference label(s); **(3) Analysis design** describing the set of analysis procedures and formulae for estimating the accuracy metrics of interest and their associated standard errors.

3.1.1 Sampling design based on Use Cases of worldwide flood events

The contents (i.e. all pixels and / or objects) included in the GFM output layers Observed Flood Extent and Reference Water Mask are too large for a complete survey, and sampling is therefore required to validate these output layers. The sampling needs to be carefully designed regarding the distribution and types of samples to be taken. The selection of a proper and efficient sampling design must adhere to procedures that ensure statistical rigor, yet still accommodate practical realities in terms of cost and time constraints.

The Use Case-based evaluation is performed for a selected number of worldwide flood events. For this set of particular events, a manually derived flood extent reference dataset within a locally confined area enables a pixel-based comparative analysis. The Global Environmental Stratification, as proposed by Metzger et al. (2018,) is used to ensure that the selected Use Cases are well distributed. This approach distinguishes 125 relatively homogeneous strata in bioclimatic conditions and can be aggregated into 18 environmental zones (Figure 1). Such a dataset allows us to group the samples into meaningful categories (i.e. strata), enabling us to identify issues within regions of similar environmental conditions. Furthermore, such a stratification makes the use-case evaluation more systematic than an approach that selects Use Cases randomly and ensures that an in-depth analysis of flood cases will be spread throughout the various environmental zones.

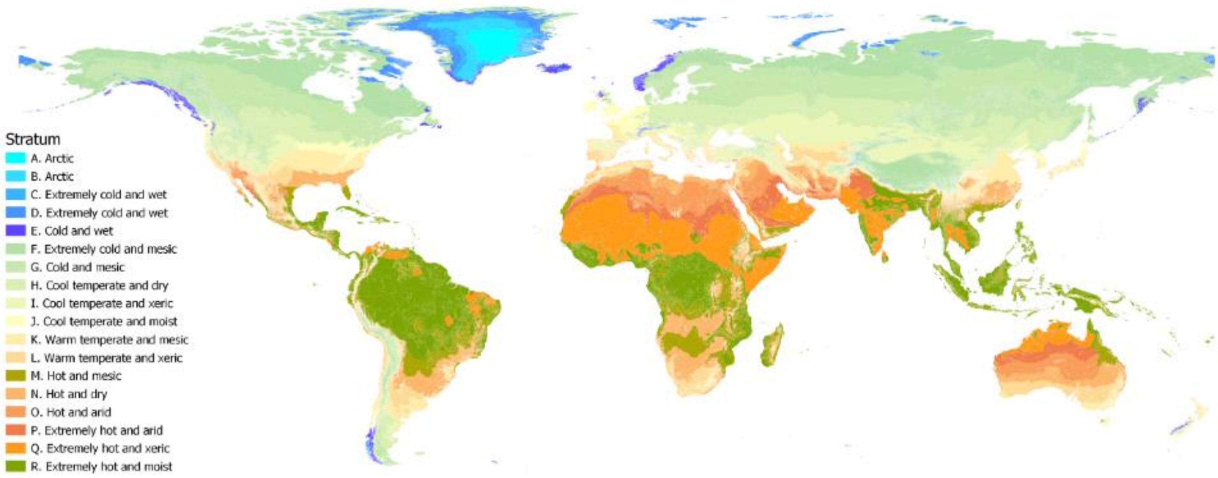


Figure 1: Global Environmental Stratification (Metzger, 2018).

Table 3: The nine Use Cases of worldwide flood events selected for the Phase I and Phase II accuracy assessment of the pre-operational GFM Product.

PHASE	*USE CASE*	*DATE OF FLOOD EVENT*	*LOCATION OF FLOOD EVENT*	*CONTINENT*
I	*1*	12.10.2021	Shanxi and Hebei, North-Eastern China	Asia
	2	15.09.2020	Bangi, Region of Tahoua, Niger	Africa
	3	07.11.2020	Villahermosa, States of Tabasco and Chiapas, southern Mexico	North America
	4	14.07.2021	Liege, Wallonia, Belgium	Europe
	5	25.10.2021	State of Kerala, southern India	Asia
	6	27.10.2021	Northern Luzon island, the Phillipines	Asia
II	*1*	16.11.2021	Bellingham and Mount Vernon, Washington State, USA	North America
	2	22.11.2021	Bentiu, Unity State, South Sudan	Africa
	3	02.03.2022	Grafton, New South Wales, Australia	Oceania

As stated in the TS, the accuracy assessment of the GFM Product must be performed through regular offline interpretations of the same Sentinel-1 scene for selected global flood events (Use Cases) that are representative of different environments and various geographic locations throughout the world. The nine Use Cases that were selected for the Phase I and Phase II accuracy assessment of the pre-operational version of the GFM Product are listed in Table 3, while the distribution of the selected Use Cases within the 18 global environment zones of Metzger et al. (2018) is shown in Table 4.

Table 4: Distribution of the nine Use Cases selected for the Phases I and II accuracy assessment within the 18 global environmental zones of Metzger et al. (2018).

#	*ENVIRONMENTAL ZONES*	*PHASE I USE CASES*						*PHASE II USE CASES*		
		1	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*
A	Arctic									
B	Arctic									
C	Extremely Cold & Wet									
D	Extremely Cold & Wet									
E	Cold & Wet									
F	Extremely Cold & Mesic									
G	Cold & Mesic									
H	Cool Temperate & Dry									
I	Cool Temperate & Xeric									
J	**Cool Temperate & Moist**							***X***		
K	**Warm Temperate & Mesic**	***X***								
L	Warm Temperate & Xeric									
M	Hot & Mesic									
N	**Hot & Dry**									***X***
O	Hot & Arid									
P	Extremely Hot & Arid									
Q	**Extremely Hot & Xeric**		***X***	***X***					***X***	
R	**Extremely Hot & Moist**			***X***		***X***	***X***			

3.1.2 Response Design: Reference Dataset for the GFM Product output layers

The analysis of the near real-time GFM Product's output layers Observed Flood Extent and Reference Water Mask compares the results of the ensemble approach with a locally trained and manually enhanced flood / water mask. The masks were created independently and without any knowledge of the methods used for the predicted map. The creation of the reference datasets is described below.

For the prediction area and the reference dataset, the Use Case area and date to be validated were the same. The comparison was conducted on a pixel level (binary flood / no flood for the Observed Flood Extent, respectively water / no water for the Reference Water Mask) and sampled on a dense regular grid of 100m. Further steps included:

- All no-data pixels have been excluded from the validation.
- All pixels marked by the exclusion mask have been excluded.
- Uncertainty values have not been considered.

(a) Reference dataset for the GFM Product output layer Observed Flood Extent:

To create the reference datasets for the observed flood extent, we used an independent semi-automated procedure combined with additional visual enhancements to get a high-quality reference flood mask. Therefore, dynamic local thresholding methods, mainly following the process described in Ludwig et al. (2019) and Twele et al. (2016), have been applied to Sentinel-1 imagery. For the generation of the reference dataset, pre-processed Sentinel-1 data was used (see the GFM Product Definition Document[3], under Satellite data pre-processing and ancillary data preparation).

The threshold derivation to distinguish between water and non-water pixels was conducted by tiling the pre-processed S1 images into 100x100 pixel patches which were further tiled into four sub-patches. Tiles that contain permanent water bodies (compared with an occurrence > 75% in the JRC's Global Surface Water Layer dataset) were removed beforehand from the threshold computation to ensure that only flooded pixels were considered. Tiles which potentially contain water are selected by analysing statistical relations between tiles and sub-tiles (Twele et al., 2016). Additionally, the Height Above Nearest Drainage (HAND) value for each patch was derived. The HAND index is used to exclude patches from the tile selection which cannot be flooded based on physical considerations. Therefore, only patches are considered with at least 20% of pixels with a HAND value lower than 15. The water / non-water threshold was computed by applying the Otsu algorithm (Otsu, 1979) at each of the selected 100x100 pixel tiles. Finally, Hartigan's Dip test values were calculated for each tile to measure the bi- / unimodality per tile (Hartigan and Hartigan, 1985).

The thresholds were then filtered by comparing the tile statistics with the statistics of the whole image and the Dip test values with a threshold that indicates high bimodality. The ten most viable tile thresholds were then averaged to get the final global threshold ultimately applied to the input backscatter image.

To facilitate the comparability of the predicted and reference flood masks, the created reference water masks were masked with the same layers (exclusion mask, permanent / seasonal water, topographic shadows) as the ensemble product.

Furthermore, manual enhancement was performed using Sentinel-2 imagery to remove false positives from the reference flood masks.

[3] https://extwiki.eodc.eu/en/GFM/PDD

(b) Reference dataset for the GFM Product output layer Reference Water Mask:

The generation of the reference dataset to evaluate the Reference Water Mask uses dynamic thresholding methods on both optical and SAR imagery separately, as described in Ludwig et al. (2019) and Martinis et al. (2009), making use of pre-processed Sentinel-1 and Sentinel-2 imagery. Sentinel-2 L1C data are atmospherically corrected using the Sen2Cor Processor (version 2.8), clouds and cloud shadows are masked applying the Sen2Cor Scene Classification (SCL). An additional cloud-shadow detection is applied on the whole time series to overcome omission errors due to the similar spectral behaviour of cloud-shadow and water areas (Ludwig et al., 2019). Further, seeded region growing is performed to fill gaps in incompletely detected shadows. Additionally, commission errors are removed by applying the Cloud Displacement Index (Frantz et al., 2018). Monthly images are then combined to image composites calculated by the geometric median (Roberts et al., 2017). Depending on the environmental conditions, the quality of the Sen2Cor classification can vary leading to artefacts in the resulting composites due to undetected clouds. In such a case, the optical water detection is performed on single imagery instead of image composites (e.g. Greenland).

Multispectral indices used for water detection (namely the Normalized Difference Water Index, Modified Normalized Difference Water Index, and Multi-Band Water Index) are derived from monthly image composites or single scenes. The choice of which multispectral indices are used depends on the land cover characteristics at the area of interest. The optical water detection is applied on equally sized tiles (e.g. 100x100 pixels) of the aggregated multispectral indices, whereas only those tiles with meaningful HAND values and variances higher than the 95th percentile of all tiles are considered to determine the global threshold using the median. The global threshold is then adapted for each tile by weighting it with the mean of the neighbouring local thresholds.

The SAR water detection uses monthly VV-polarized backscatter statistics as input. Sentinel-1 images are pre-processed using SNAP (Version 8). The processing steps are:

- Orbit corrections.
- Thermal / Border noise removal.
- Custom border noise removal (if needed).
- Radiometric calibration to Sigma Nought.
- Terrain Correction.
- Speckle noise reduction.

The water detection algorithm combines global and local image thresholding, seeded region growing and fuzzy logic postprocessing. Thresholding is performed as described above, except that instead of local Otsu thresholding, an adaptive thresholding method is used (Bradley and Roth 2007). Omitted water pixels are added to the water masks by a seeded region growing algorithm which is applied to each water body separately. Finally, a post-processing procedure, described in Martinis et al. (2009), is conducted to remove commission errors such as terrain shadows. The corresponding S1 and S2 water masks are fused by combining all water pixels.

3.1.3 Analysis design: Measures used to quantify the quality of the GFM Product

As outlined in the Technical Specifications (TS), the basis for the accuracy assessment of the pre-operational GFM Product is the 2x2 **error matrix** (also called confusion matrix or contingency table), as illustrated in Table 5. As can be seen, the error matrix is a simple cross-tabulation of the class labels FLOOD and NO FLOOD which are allocated by the GFM flood mapping algorithm, against those in the reference dataset, for each Use Case. The error matrix organises the acquired sample data to summarise critical results and aids in quantifying accuracy. The main diagonal of the error matrix highlights correct classifications, while off-diagonal elements show “omission” and “commission” errors, as explained below. The main accuracy metrics which can be derived from the 2-by-2 error matrix shown in Table 5, are summarised in Table 6. Although the metrics User’s Accuracy and Producer’s Accuracy are not used in the quality assessment of the pre-operational GFM product, they are included in Table 6 for the sake of completeness.

Table 5: Illustration of the 2-by-2 error matrix used in the quality assessment to compare the sample points (total = A+B+C+D) in the reference (observed) and classified (detected) datasets, for each Use Case.

		Observed	
		FLOOD	NO FLOOD
Detected	FLOOD	A	B
	NO FLOOD	C	D

Table 6: Summary of the main accuracy metrics which are derived from the 2-by-2 error matrix.

ACCURACY METRIC	DESCRIPTION	COMPUTATION (see Table 5)
Overall Accuracy	Proportion of the total number of sample points (FLOOD and NO FLOOD) that are correctly classified.	[A + D] / [A + B + C + D]
User’s Accuracy	Proportion of classified FLOOD pixels that are FLOOD in the observed pixels.	[A] / [A + B]
Commission Error (or false positive, over-detection)	Proportion of classified FLOOD pixels that are NO FLOOD in the observed pixels. (Complement of User’s Accuracy).	[B] / [A + B], or [100% - User’s Accuracy]
Producer’s Accuracy	Proportion of observed FLOOD pixels that are FLOOD in the classified pixels.	[A] / [A + C]
Omission Error (or false negative, under-detection)	Proportion of observed FLOOD pixels that are NO FLOOD in the classified pixels. (Complement of Producer’s Accuracy).	[C] / [A + C], or [100% - Producer’s Accuracy]
Critical Success Index (CSI)	Proportion of the total number of observed and classified FLOOD pixels that are correctly classified.	[A] / [A + B + C + D]

Bias or bias ratio (b)	Indication if errors are due to under-detection (b < 1), over-detection (b > 1), or are neutral (b = 1).	[A + B] / [A + C]

Additionally, accuracy estimations can be improved by using stratified estimators (for stratified random sampling) or post-stratified estimators if simple random sampling is used (Card, 1982; Olofsson, 2013). This means that the overall and per-class accuracy estimations used to evaluate the output products should include the known areas of each map category to improve the estimation of the proportion of correctly mapped samples. However, in the use-case-based evaluation, as performed, no stratification and weighting were necessary because the validation performed on pixel level is equivalent to a very dense random sampling scheme.

The error measures explained above to satisfy the principles of equivalence of events, i.e. flood and no-flood cases, are equally important. However, the latter class is usually dominant outside the flood extent, and the reported measure would likely indicate a biased result (towards no-flood accuracy). Therefore, the **Critical Success Index (CSI)**, also known as the threat score (Wilks, 2011), is a measure that is particularly useful when classified events occur substantially less frequently than the non-occurrence of the event.

CSI is the number of correct observations divided by the number of occasions on which a particular event was either detected or observed. It can be viewed as a proportion correct for the quantity to be detected after removing correct non-detected events from consideration. The CSI ranges between 0 (worst) and 1 (best).

In addition to the classical accuracy metrics (i.e. overall, producer's, and user's accuracies) and the CSI, another useful measure is the **bias** or **bias ratio** (b). A bias ratio of b = 1 means that the measured errors are "neutral", with errors of commission (false positives) and omission (false negatives) of the same magnitude. Cases with b < 1 or b > 1 indicate, respectively, an under-detection or over-detection of events. Therefore, the bias can be used to report the relation between both errors in a single metric and thus also helps to find an optimal (desired) solution between both cases.

The accuracy metrics described above will report the product performance in terms of thematic accuracy, and are used to compute KPI-3a, KPI-3b, KPI-3c, and KPI-3d (see Table 7 below) for both the GFM Product output layers Observed Water Extent and the Reference Water Mask (i.e. permanent and seasonal water).

Table 7: KPIs to assess the thematic accuracy of the GFM Product output layer Observed Flood Extent.

KPI #	Title	Description	Target KPI values
KPI-3a	Critical Success Index	Commonly used indicator to evaluate the performance of dichotomous (Yes / No) classifications of low-frequency events	70-80 %

KPI-3b	Bias	Ratio informing about under- or over-detection of flood extent areas	1
KPI-3c	Overall accuracy	Proportion correctly detected considering permanent and seasonal water bodies as well as flooded areas	>95 %
KPI-3d	Omission and commission errors	Miss rate (i.e. omission errors) and false alarm rate (commission errors), caused respectively by under- and over-detection of flooding.	<5 %

3.1.4 Automated file quality checks of output datasets of the GFM Product

The automated file quality checks are performed to ensure the consistent quality of all output datasets of the GFM Product. Therefore, each file is compared to the product technical specifications, which cover: geometric (spatial) resolution; coordinate reference system (CRS); coverage (extent of raster file); data type; raster coding; metadata; data format; and file-naming. The technical specifications for the GFM Product output layers “Observed Flood Extent”, “Reference Water Mask”, and “Exclusion Mask” are shown in Table 8, Table 9, and Table 10 below.

Table 8: Technical specifications – GFM Product output layer Observed Flood Extent.

PARAMETER	DEFINITION
Product acronym:	ENSEMBLE_FLOOD
Geometric resolution:	Pixel resolution 20m x 20m
Coordinate Reference System (CRS):	CRS of corresponding Sentinel-1 scene
Coverage:	Extent of corresponding Sentinel-1 scene
Data type:	8bit unsigned raster with LZW compression
Raster coding (thematic pixel values):	0: no flood. 1: flood. 255: nodata
Metadata:	JSON File (.json)
Data format:	GeoTIFF (.tif)
Filename:	[PRODUCT ACRONYM]_[SENTINEL-1 SCENE ID].tif

Table 9: Technical specifications – GFM Product output layer Reference Water Mask.

PARAMETER	DEFINITION
Product acronym:	REFERENCE_WATER
Geometric resolution:	Pixel resolution 20m x 20m
Coordinate Reference System (CRS):	CRS of corresponding Sentinel-1 scene
Coverage:	Extent of corresponding Sentinel-1 scene
Data type:	8bit unsigned raster with LZW compression
Raster coding (thematic pixel values):	0: no water. 1: permanent water. 2: seasonal water. 255: nodata.
Data format:	GeoTIFF (.tif)

Filename:	[PRODUCT ACRONYM]_OUT_S1_IW_GRDH_1SSV_ [START DATE]_[END DATE]_MONTH*.tif

Table 10: Technical specifications – GFM Product output layer Exclusion Mask.

PARAMETER	DEFINITION
Product acronym:	EXCLUSION_LAYER
Geometric resolution:	Pixel resolution 20m x 20m
Coordinate Reference System (CRS):	CRS of corresponding Sentinel-1 scene
Coverage:	Extent of corresponding Sentinel-1 scene
Data type:	8bit unsigned raster with LZW compression
Raster coding (thematic pixel values):	0: not excluded area. 1: excluded area. 255: nodata.
Data format:	GeoTIFF (.tif)
Filename:	[PRODUCT ACRONYM]_[SENTINEL-1 SCENE ID].tif

4 Phase I accuracy assessment of the pre-operational GFM Product

This Section describes in detail the Phase I accuracy assessment of the pre-operational version of the GFM Product, which was carried out in October 2021, and which considered the six selected Use Cases of worldwide flood events which are listed in Table 11. The six Use Cases are outlined and further discussed in Section 4.1 below.

Table 11: List of Use Cases for the Phase I accuracy assessment of the pre-operational GFM Product.

#	LOCATION OF USE CASE	DATE OF FLOOD EVENT
1	Shanxi and Hebei, North-Eastern China	12.10.2021
2	Bangi, Region of Tahoua, Niger	15.09.2020
3	Villahermosa, States of Tabasco and Chiapas, southern Mexico	07.11.2020
4	Liege, Wallonia, Belgium	14.07.2021
5	State of Kerala, southern India	25.10.2021
6	Northern Luzon island, the Phillipines	27.10.2021

4.1 Use Cases for Phase I accuracy assessment

This Section provides a short overview of the six investigated Use Cases of worldwide flood events, in terms of time, location, and a brief description of causing effects of the events.

4.1.1 Use Case 1: Flood event in China (12.10.2021)

Flood event	12.10.2021
Location	Shanxi and Hebei, North-Eastern China
Coordinates	36° 58' 28" N, 115° 46' 55" E
Global Env. Stratification	K – Warm temperate and mesic
Description	Throughout 2021 many flood events occurred in China. In October 2021, heavy rainfalls caused floods and landslides in the provinces of Shanxi and Hebei, forcing 120,000 people to be displaced and 60 coal mines to suspend operations.

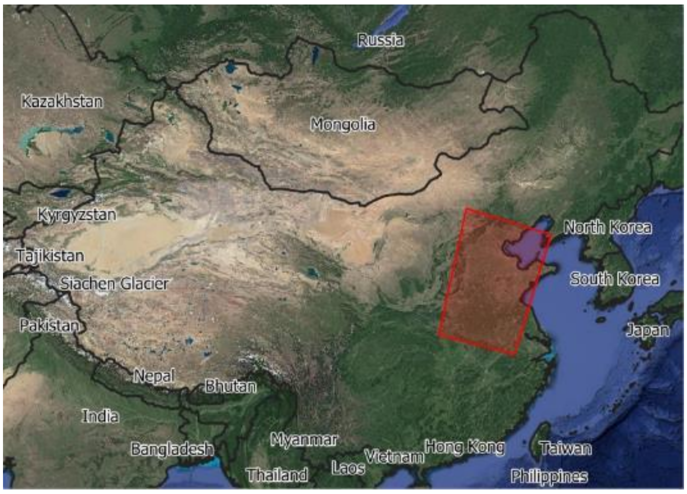


Figure 2: Area of interest for the flood event in China (12-10-2021).

4.1.2 Use Case 2: Flood event in Niger (15.09.2020)

Flood event	15.09.2020
Location	Bangi, Region of Tahoua, Niger.
Coordinates	6°7'44"E, 13°37'20"N
Global Env. Stratification	Q – Extremely hot and xeric
Description	After heavy rainfalls, the Niger River flooded surrounding areas. The floods caused the bursting of dams and dikes along the river and vast volumes of water to engulf people's homes, farms, and other buildings without warning. The floods destroyed 30,000 houses.

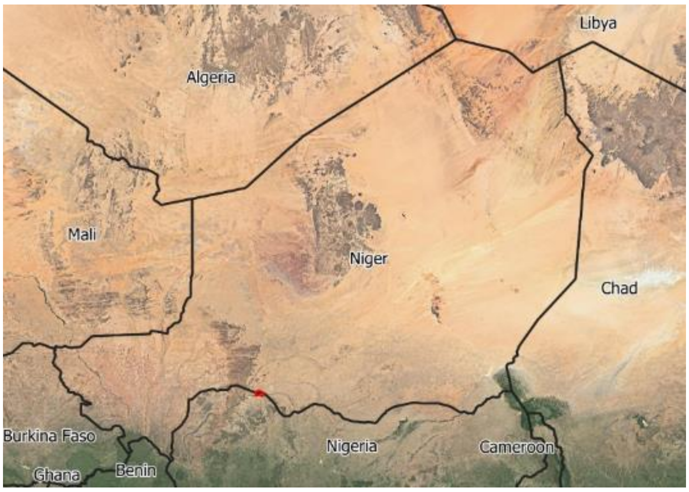


Figure 3: Area of interest for the flood event in Niger (15.09.2020).

4.1.3 Use Case 3: Flood event in Mexico (07.11.2020)

Flood event	07.11.2020
Location	Villahermosa, States of Tabasco and Chiapas, southern Mexico
Coordinates	92°54'30"W, 17°59'19"N
Global Env.	Q – Extremely hot and xeric, R – Extremely hot and moist

Stratification	
Description	Heavy rainfalls (>200mm rain in 24 hrs) triggered landslides, blocked roads, and the overflow of several rivers. Significant floods in Villahermosa were avoided by intentionally conducting water from dams to low areas.

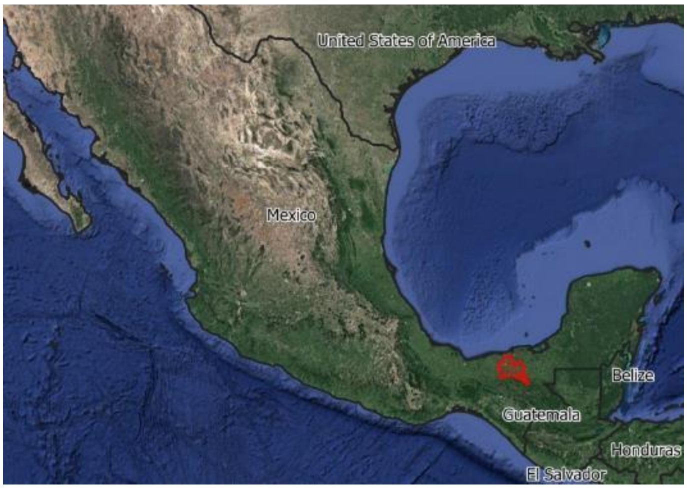


Figure 4: Area of interest for the flood event in Mexico (07-11-2020).

4.1.4 Use Case 4: Flood event in Belgium (14.07.2021)

Flood event	14.07.2021
Location	Liege, Wallonia, Belgium.
Coordinates	5°46'31"E, 51°342" N
Global Env. Stratification	J – Cool temperate and moist
Description	Due to persistently high moisture supply and slowly moving weather patterns, heavy rainfalls caused floods along rivers in the Wallonia region. The worst floods in decades caused 41 fatalities.

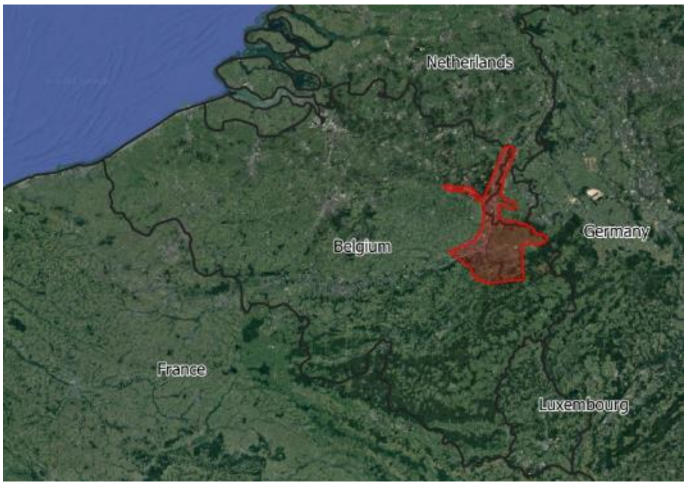


Figure 5: Area of interest for the flood event in Belgium (14.07.2021).

4.1.5 Use Case 5: Flood event in India (25.10.2021)

Flood event	25.10.2021
Location	State of Kerala, southern India
Coordinates	76°26'40"E, 10°38'17"N
Global Env. Stratification	R – Extremely hot and moist
Description	Days of heavy rainfall led to landslides and rivers overflowing. Dams were opened to reduce the risk of dangerous overflows. Wetlands and lakes that once acted as natural safeguards against floods have disappeared because of increasing urbanisation and construction.

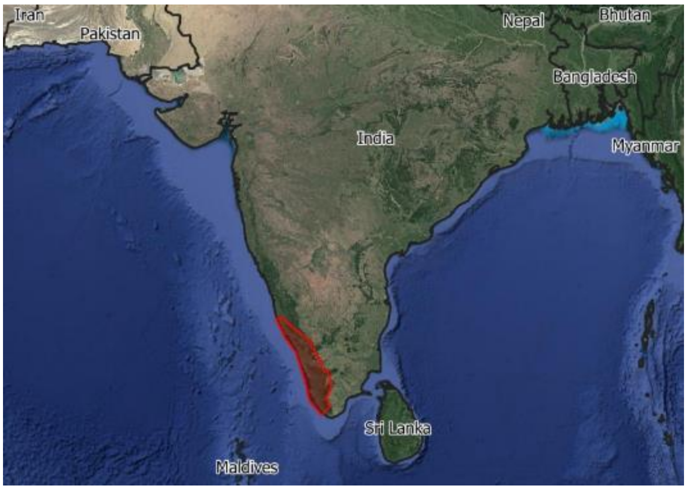


Figure 6: Area of interest for the flood event in India (25-10-2021).

4.1.6 Use Case 6: Flood event in the Philippines (27.10.2021)

Flood event	27.10.2021
Location	Northern Luzon island, the Phillipines
Coordinates	121°3'37"E, 16°38'47"N
Global Env. Stratification	R – Extremely hot and moist
Description	Tropical storm Maring hit Luzon Island, flooding large parts of agricultural land and causing high losses for farmers.

Figure 7: Area of interest for the flood event in the Philippines (27-10-2021).

4.2 Results of Phase I accuracy assessment for the six Use Cases

The Key Performance Indicators (KPIs) which were achieved for the six Use Cases of worldwide flood events included in the initial Phase I accuracy assessment, are summarised Table 12. As can be seen, the targeted value for KPI-3a (Critical Success Index), could not be reached for any selected cases based on the provided data. (This is in marked contrast with the good results achieved in the Phase II accuracy assessment, as reported in Section 5 below).

Further visual analysis shows that, especially for the selected flood events in China and Belgium, results are generally good and close to the targeted KPIs. The Bias (KPI-3b) indicates a general under-detection of flooded areas for the selected Use Cases. Considering the KPIs overall accuracy, it is evident that this metric is not very meaningful in the context of relatively rare events such as flooding, as non-flooded areas dominate the result.

The omission and commission errors are more useful, as they specify the number of false alarms and missed events in more detail. Again, for the investigated areas and based on the omission and commission errors, the results indicate that the GFM Ensemble generally under-estimates the flood extent.

Some further points regarding the presented results of the Phase I accuracy assessment are considered and discussed below in Section 4.3, and in Section 6.

Table 12: Achieved KPIs for the six Use Cases in the Phase I accuracy assessment.

KPI #	*Title*	*China*	*Niger*	*Mexico*
KPI-3a	Critical Success Index	65.8 %	50.1 %	28.5 %
KPI-3b	Bias	0.7131	0.5279	0.3032
KPI-3c	Overall accuracy	99.9 %	98.5 %	90.7 %

KPI-3d	Omission errors (no flood / flood)	0.02 % / 28,69 %	0,15 % / 47,21 %	0,89 % / 69,68 %
KPI-3d	Commission errors (no flood / flood)	0,07 % / 10,49 %	1,33 % / 9,36 %	8,97 % / 17,28 %
KPI #	***Title***	***Belgium***	***India***	***Philippines***
KPI-3a	Critical Success Index	69.3 %	54.9 %	25.2 %
KPI-3b	Bias	0.8262	0.6003	0.8569
KPI-3c	Overall accuracy	99.8 %	99.9 %	99.7 %
KPI-3d	Omission errors (no flood / flood)	0,09 % / 17,38 %	0.02 % / 39,97 %	0,33 % / 14,31 %
KPI-3d	Commission errors (no flood / flood)	0,08 % / 18,9 %	0.08 % / 13,5 %	0.02 % / 73,64 %

4.3 Discussion of results of Phase I accuracy assessment for the six Use Cases

The validation results for the Phase I accuracy assessment show a wide range of thematic accuracy values, especially when looking at KPI-3a (Critical Success Index) which has a target value of 70-80%. As is shown in Table 12, this target value was not reached for the selected Use Cases. However some general points should be noted, and the Use Cases themselves should be closely considered, in order to interpret correctly the results of the Phase I accuracy assesment.
As outlined earlier, the analysis approach is based on Use Cases. According to good practice guidelines (e.g. Olofsson et al., 2013 and 2014), product validation can be based either on an independent, higher quality Reference Dataset, if available, or on an independent higher quality production methodology. Because an independent, higher quality Reference Dataset was not available for any Use Case, the quality assessment conducted here applies a higher quality methodology to the production data. For this purpose, a semi-automated approach was used (see Section 3.1.2), tuned to the context of the Use Cases and visually controlled (and adjusted where required) in order to create a Reference Dataset of the best possible quality. However, due to the large areas of the six Use Cases, we cannot claim that the Reference Dataset is error-free. Therefore, while the Reference Dataet is used for validation, potential errors stemming from the semi-automated approach will contribute to the analysed errors, and lower the KPI values of the KPIs. In other words, it may occur that a flood is correctly detected by the GFM Product but not in the Reference Dataset.

A fully manual mapping of the presented Use Cases is obviously not feasible, due to the large areas covered and the uncertainty of exact delineation of contiguous flooded regions based on the Sentinel-1 backscatter data. Therefore, a complete picture of the achieved accuracy will only be available based on the planned validation approach based on sample points. Those investigated areas will be carefully selected and confined for future use-case-based evaluations, allowing for the best-possible visual control. Furthermore, a plausibility approach based on visually interpreted sample points at a larger scale and spatial context is less affected by a single-pixel noise, and higher accuracies can therefore be expected.

Despite the shortcomings of the approach used for the Phase I accuracy assessment, this preliminary evaluation indicates that the target 70% CSI may not be achieved for all regions using the current configuration of the GFM flood detection algorithms. Extending the evaluation to the entire water extent (i.e. permanent and temporary water) would enable a more in-depth analysis, by considering all areas of water occurrence for the validation.

A final general observation, which applies to all six Use Cases in the Phase I accuracy assessment, is that the flood detection is characterised by both omission and commission errors. While the sensitivity of the GFM algorithms spreads in both directions, there is a tendency to underestimate the flood extent (i.e. detected flood areas are statistically more minor than in the Reference Dataset). A more sensitive detection might reduce these underestimations, but will likely increase the detection of false-positive events.

All of the above general points must be borne in mind when interpreting the validation results (i.e. the KPIs and especially the Critical Success Index, with a target value of 70-80%).

Figure 8, Figure 9, Figure 10, and Figure 11 provide a visual comparison of the SAR-based water detection results obtained for the GFM ensemble algorithm and the Reference Dataset, and the corresponding Sentinel-1 backscatter images and resulting omission and commission errors (if any), for four of the Use Cases in the Phase I assessment: i.e. the flood events in China on 12.10.2021, Mexico on 07.11.2020, Belgium on 14.07.2021, and the Phillipines on 27.10.2021. (Note that in Figure 9, which presents the results for the Mexico Use Case, the Exclusion Mask has NOT been applied to the displayed Reference Dataset).
As can be seen in Table 12, for the Phase I assessment the two Use Cases in Mexico and the Philippines show the lowest Critical Success Index values (28.5% and 25.2%, respectively). Looking at the Mexico flood event, it is evident in Figure 9 that the Exclusion Mask used for the GFM Ensemble covers large areas. Therefore, many areas which are detected as water in the Reference Dataset are excluded from the GFM Ensemble. Figure 9 also clearly shows that the GFM ensemble approach shows some omission errors. The validation results for the Philippines flood event (Figure 11) how mainly commission errors, as the GFM Ensemble covers more flooded areas along the river. Unfortunately, due to the given short timeframe for the Phase I accuracy assessment, further refinement and analysis of the Reference Dataset were not possible. Furthermore, the cloud cover for the given flood event prevented a timely visual and manual interpretation. In this case a more in-depth analysis would be required in order to provide concrete and reliable conclusions from the validation.

As can be seen in Table 12, the best results for the Phase I accuracy assessment were achieved for the flood events in Belgium (CSI 69.3%) and China (CSI 65.8%). In the case of Belgium, the GFM Ensemble and Reference Dataset produce similar results (see Figure 10). As can be seen, fields along the river are inundated, and these flood areas are successfully covered by both approaches. The validation results for the China flood event tell a similar story (Figure 8). In this case, several different flood events occurred over a large area, and both the GFM Ensemble and the Reference Dataset captured a similar extent of floods.

Many retention basins along rivers were inundated, and here the GFM Ensemble and Reference Dataset show identical results (see Figure 8).

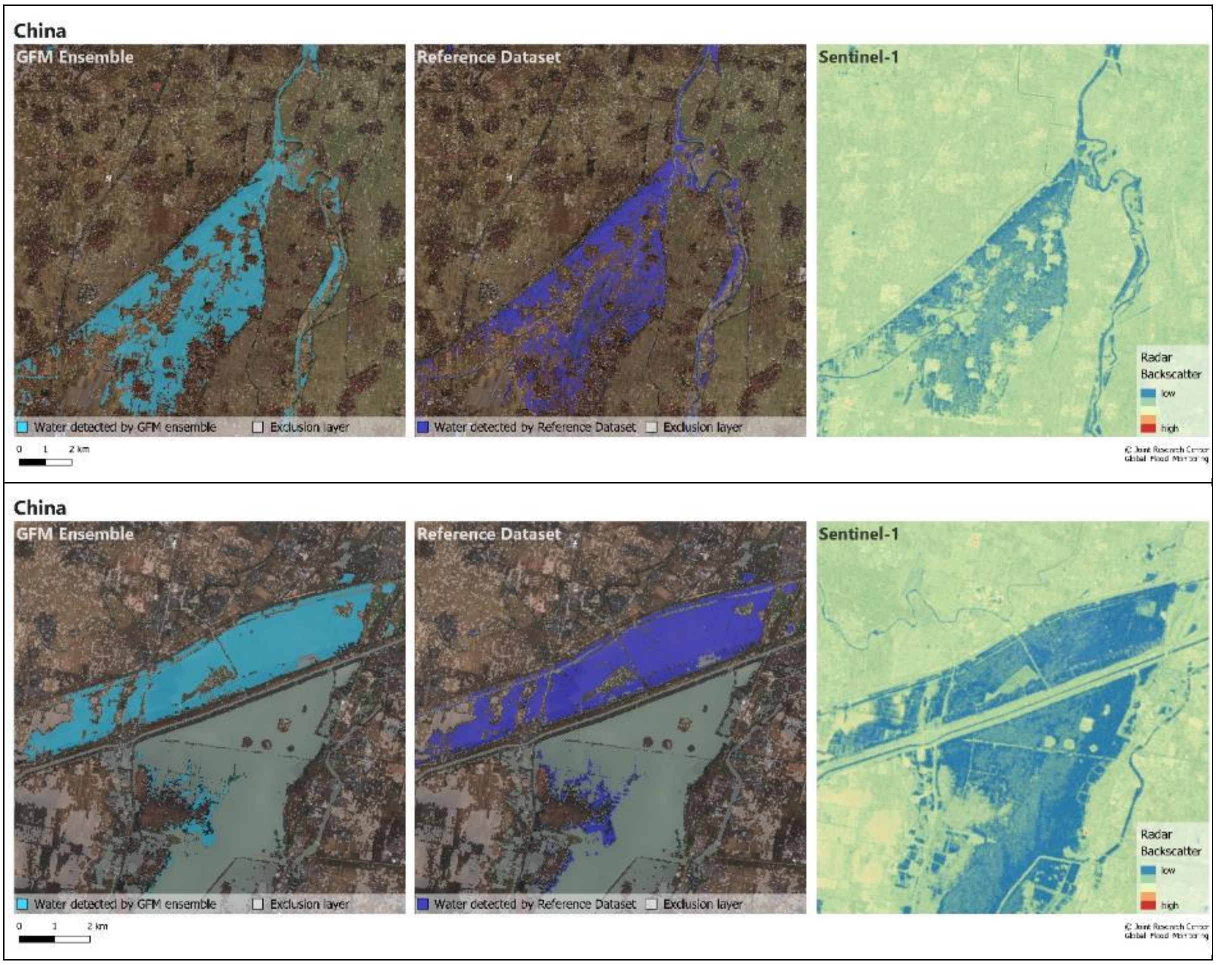


Figure 8: Flood event in China (12.10.2021) - SAR-based water detection results from the GFM Ensemble (left) and Reference Dataset (middle), and Sentinel-1 backscatter image (right).

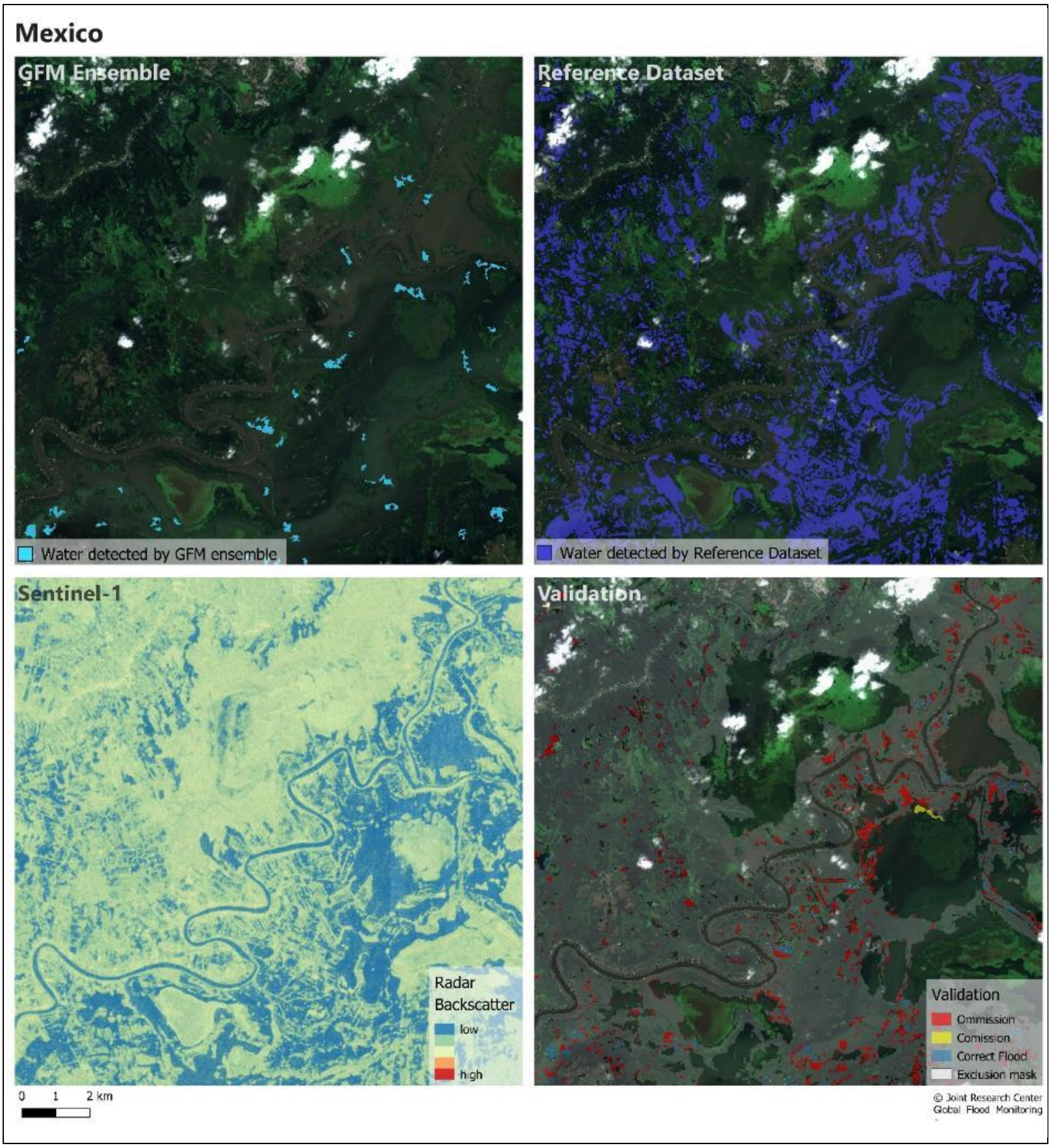


Figure 9: Flood event in Mexico (07.11.2020) - SAR-based water detection results from the GFM Ensemble (top-left) and Reference Dataset (top-right), Sentinel-1 backscatter image (bottom-left), and omission / commission errors (bottom-right).

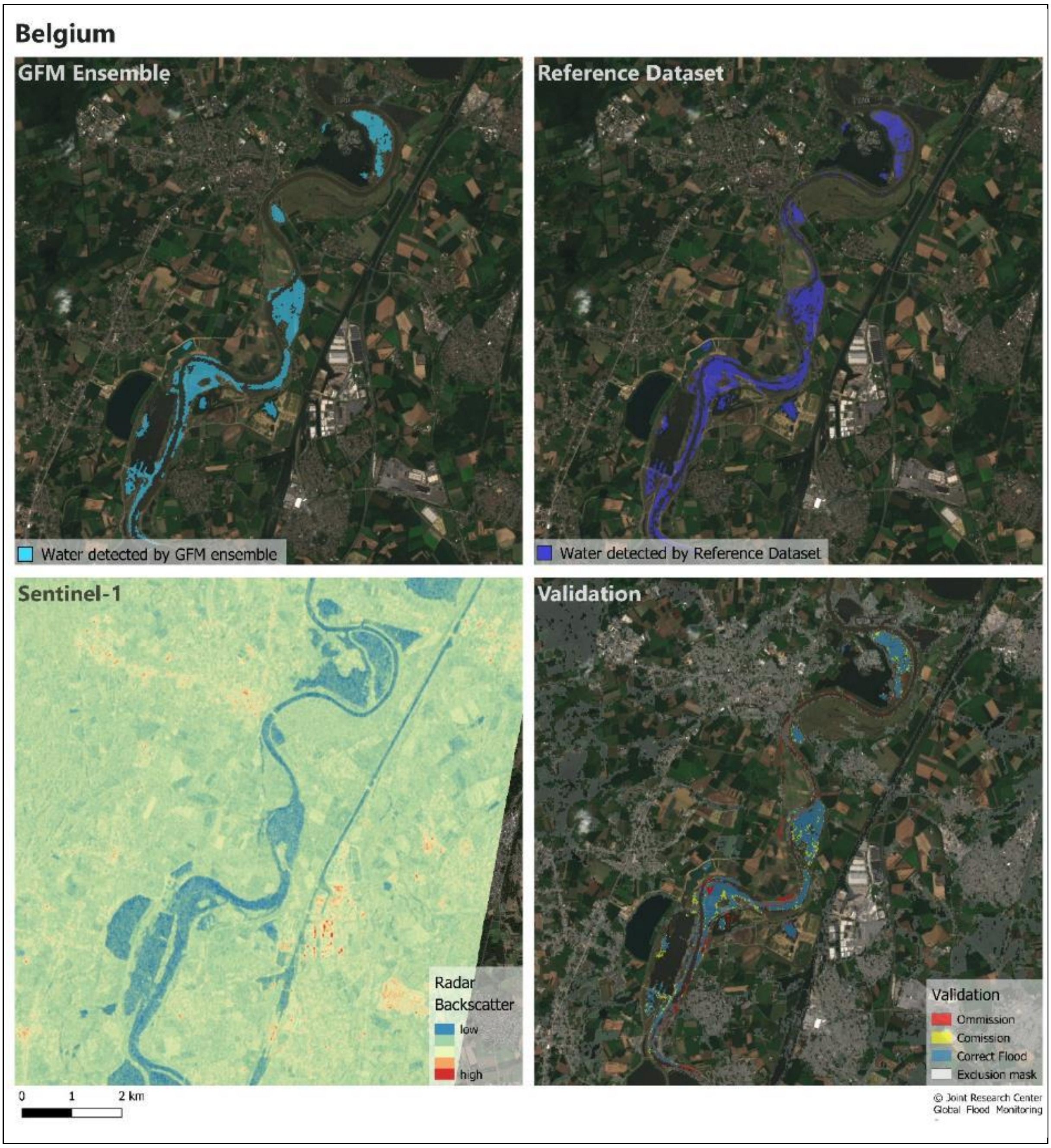


Figure 10: Flood event in Belgium (14.07.2021) - SAR-based water detection results from the GFM Ensemble (top-left) and Reference Dataset (top-right), Sentinel-1 backscatter image (bottom-left), and omission / commission errors (bottom-right).

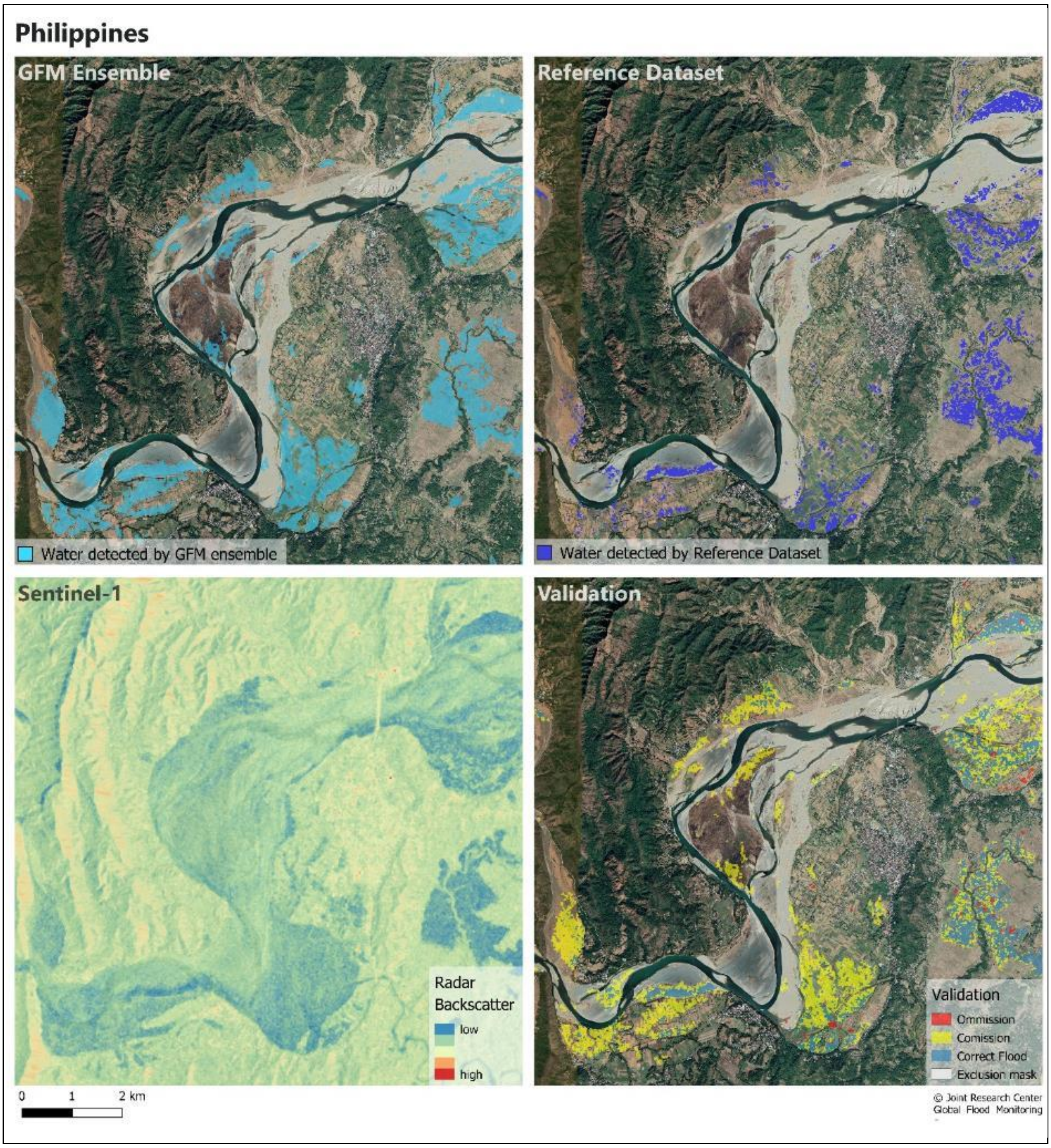


Figure 11: Flood event in the Philippines (27.10.2021) - SAR-based water detection results from the GFM Ensemble (top-left) and Reference Dataset (top-right), Sentinel-1 backscatter image (bottom-left) and omission / commission errors (bottom-right).

5 Phase II accuracy assessment of the pre-operational GFM Product

This Section describes the Phase II accuracy assessment of the pre-operational version of the GFM Product, for the three selected Use Cases of worldwide flood events that are listed in Table 13 below. For each Use Case, the time, location and causing effects of the flood event are first outlined. Then, for each Use Case the validation results for the GFM Product output layers "Observed Flood Extent' and "Reference Water Mask", and including the results of the general product checks for all GFM Product output layers, are summarized. The results of the Phase II accuracy assessment are then analysed and discussed in further detail. Finally, a short analysis of the general plausibility of the GFM Product output layer Exclusion Mask, for the three selected Use Cases, is presented. It should be noted that the general remarks on the validation procedure, which were presented earlier in Section 4.3, are equally relevant for the Phase II accuracy assessment.

Table 13: List of Use Cases for the Phase II accuracy assessment of the pre-operational GFM Product.

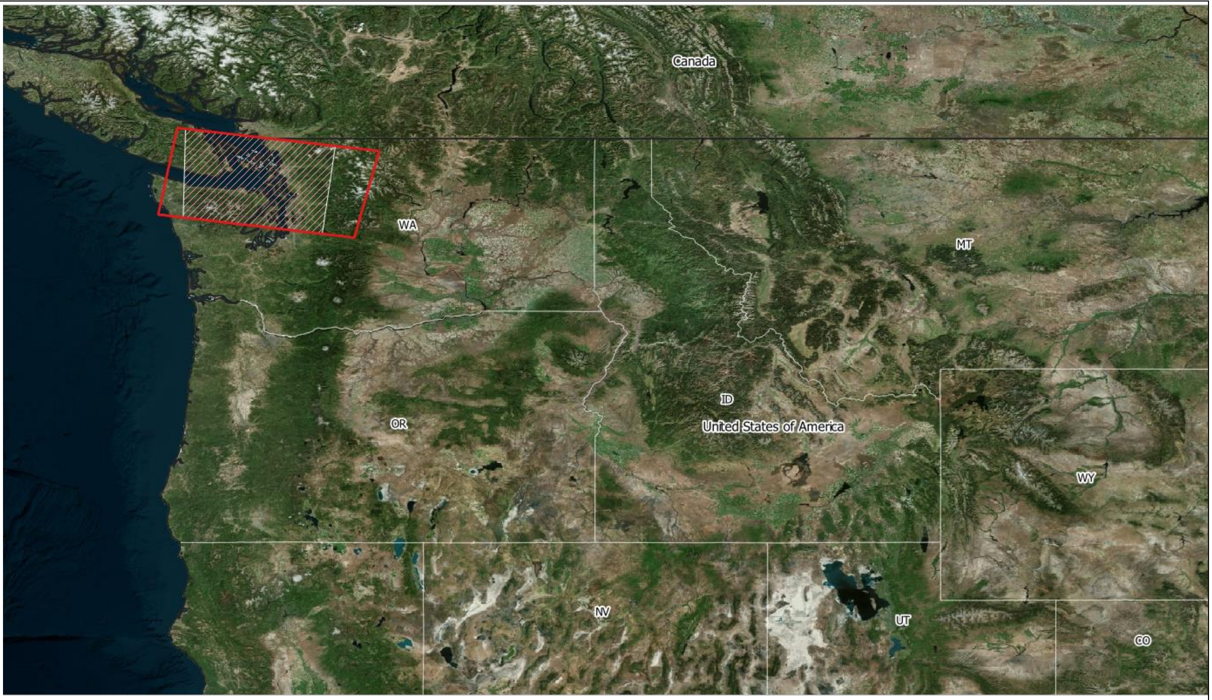
#	LOCATION OF USE CASE	DATE OF FLOOD EVENT
1	Bellingham and Mount Vernon, Washington State, USA	16.11.2021
2	Bentiu, Unity State, South Sudan	22.11.2021
3	Grafton, New South Wales, Australia	02.03.2022

5.1 Use Cases for Phase II accuracy assessment

5.1.1 Use Case 1: Flood event in USA (16.11.2021)

Flood event	16.11.2021
Location	Bellingham and Mount Vernon, Washington State, USA
Coordinates	48° 27' 23" N, 122° 20' 11" W
S-1 scene	S1B_IW_GRDH_1SDV_20211116T142056_20211116T142121_029614_0388BC_B07E
Global Env. Stratification	J – Cool temperate and moist
Description	British Columbia (Canada) and Washington State (USA) were affected by serious floodings in November 2021 following days of severe rain caused by an atmospheric river.[4] The selected Sentinel-1 scene from November 16, 2021, covers floodings around Bellingham and Mount Vernon in Washington State.

[4] ESA (2021), Washington state flooding [https://www.esa.int/ESA_Multimedia/Images/2021/11/Washington_state_flooding], accessed 18.03.2022; NASA Earth Observatory (2021), Severe Flooding in the Pacific Northwest, [https://earthobservatory.nasa.gov/images/149100/severe-flooding-in-the-pacific-northwest], accessed 18.03.2022.



Sentinel-1 scene Area of Interest

Figure 12: Area of interest for the flood event in USA (16.11.2021). (Basemap: Bing satellite imagery).

5.1.2 Use Case 2: Flood event in South Sudan (22.11.2021)

Flood event	22.11.2021
Location	Bentiu, Unity, South Sudan
Coordinates	9° 11' 41" N, 29° 38' 56" E
S-1 scene	S1B_IW_GRDH_1SDV_20211122T034954_20211122T035019_029695_038B51_9E17
Global Env. Stratification	Q – Extremely hot and xeric
Description	South Sudan experienced severe floodings throughout the country in 2021, starting in May with the beginning of the rainy season. Bentiu is amongst the worst impacted states.[5] The identified area of interest covers a flood event recorded on November 22, 2021 close to Bentiu.

[5] Floodlist (2021), South Sudan – Over 800,000 Affected by Worst Flooding in 60 Years, [https://floodlist.com/africa/south-sudan-floods-update-december-2021], accessed 18.03.2022; CNN (20221), The world's newest nation is both drying up and drowning, [https://edition.cnn.com/2021/12/06/africa/south-sudan-floods-climate-cmd-intl/index.html], accessed 18.03.2022.

Sentinel-1 scene *Area of Interest*

Figure 13: Area of interest for the flood event in South Sudan (22.11.2021). (Basemap: Bing satellite imagery).

5.1.3 Use Case 3: Flood event in Australia (02.03.2022)

Flood event	02.03.2022
Location	Grafton, New South Wales, Australia
Coordinates	29° 36' 50" S, 153° 0' 36" E
S-1 scene	S1A_IW_GRDH_1SDV_20220302T190635_20220302T190700_042146_050598_23F0
Global Env. Stratification	N – Hot and Dry
Description	In March 2022, Australia's east coast (Southeast Queensland, New South Wales) was hit by severe flooding affecting thousands of people. The rainfall event was associated with the La Nina event.[6] This Use Case covers flooding around Grafton, New South Wales, on March 2, 2022.[7]

[6] Copernicus Emergency Management Service – Mapping (2021), EMSR567: Floods in Queensland Australia, [https://emergency.copernicus.eu/mapping/list-of-components/EMSR567], accessed 18.03.2022.

[7] Copernicus Emergency Management Service – Mapping (2021), [EMRS567], Grafton: Delineation Product, Monitoring 1, version 1, release 1, RTP Map #01, [https://emergency.copernicus.eu/mapping/ems-product-component/EMSR567_AOI08_DEL_MONIT01_r1_RTP01/1], accessed 18.03.2022.

Sentinel-1 scene Area of Interest

Figure 14: Area of interest for the flood event in Australia (02.03.2022). (Basemap: Bing satellite imagery).

5.2 Results of Phase II accuracy assessment for the three Use Cases

5.2.1 Results for flood event in USA (16.11.2021)

For the flood event in the USA / Washington State, the Key Performance Indicators (KPIs) quantifying the thematic accuracy of the GFM Product output layers Observed Flood Extent and Reference Water Mask (which comprises both permanent and seasonal water), are presented in Table 14, Table 15, and Table 16.

As can be seen, for both the Observed Flood Extent and the permanent water class of the Reference Water Mask, KPI-3a (Critical Success Index or CSI) shows very good results that exceed the target value (70-80%). As can be seen in Table 14, the CSI for the Observed Flood Extent reaches 86.8 %, while KPI-3b (Bias) indicates a slight under-detection compared with the independent reference dataset.

The Reference Water Mask that was evaluated and considered for the flood event was computed considering the years 2019 and 2020. The permanent water class was computed across all Sentinel-1 images available in 2019 and 2020, while the seasonal water mask is based on all Sentinel-1 images for each respective month in 2019 and 2020.

Concerning the Permanent Water class, as can be seen in Table 15 there is very good agreement (CSI = 98%, Bias = 1.01) between the GFM Product and the independent reference dataset. In contrast, as can be seen in Table 16, the twelve Seasonal Water classes show a low CSI (< 10%). The Bias is constantly below 1.0, suggesting that the seasonal water classes underestimate the water extent compared with the monthly independent reference datasets. The CSI shows even a value of zero for the June seasonal water mask, and the Bias is not applicable, as the reference dataset does not include any seasonal water pixels within the defined, validated area.

However, in the case of the seasonal reference masks, the overall low results for KPI-3a and -3b should be considered cautiously. Our general observation is that the Sentinel-1-based computation of the reference water extent underestimates the true water extent, particularly along river courses. The validation product considers both optical and radar imagery, and from the investigated Use Cases, the optical data supplies additional observations to compute the seasonal reference masks and seems to track rivers more accurately. On the other hand, static water bodies (i.e. lakes) appear slightly overestimated by the radar-based methodology compared to optical imagery. In addition to that, it is evident that some rivers show substantial intra-year variability.

Therefore, the question of how to best map seasonal water occurrence - which explains some of the significant differences observed (e.g. see Figure 15) - needs to be raised. The resulting "salt-and-pepper" differences in the monthly water occurrences along the water / non-water pixel border explain the generally low CSI and off Bias values when only considering the month-specific water extent (i.e. observed water extent for a particular month minus the permanent water). If we also considered those parts in the monthly reference maps that are permanent water bodies year-round, the resulting CSI and Bias would be close to those computed for the permanent water mask.

Intra-annual variations of river-flow are highlighted in Figure 15, which shows the river flow differences between November 2019 and November 2020 for the Use Case in USA, and also compares the seasonal Reference Water Masks computed by the GFM algorithm (based on SAR data only) and the Reference Dataset (which combines optical and SAR data).

For the Use Case in USA, automated file quality checks for the GFM output layers Observed Flood Extent, Reference Water Mask, and Exclusion Mask show that the files follow the technical specifications regarding resolution, data format, and further quality parameters.

Table 14: Flood event in USA (16.11.2021) - achieved KPIs for Observed Flood Extent.

KPI #	*Title*	*Observed Flood Extent*
KPI-3a	Critical Success Index	86.8 %
KPI-3b	Bias	0.987
KPI-3c	Overall accuracy	99.1 %
KPI-3d	Omission errors (no flood / flood)	0.4 % / 7.7 %
KPI-3d	Commission errors (no flood / flood)	0.5 % / 6.5 %

Table 15: Flood event in USA (16.11.2021) – achieved KPIs for permanent water of the Reference Water Mask.

KPI #	*Title*	*Permanent Water*
KPI-3a	Critical Success Index	98 %
KPI-3b	Bias	1.013
KPI-3c	Overall accuracy	99.4 %
KPI-3d	Omission errors (no water / water)	0.7 % / 0.3 %
KPI-3d	Commission errors (no water / water)	0.1 % / 1.6 %

Table 16: Flood event in USA (16.11.2021) - achieved KPIs for monthly masks of seasonal water of the Reference Water Mask.

KPI #	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*
KPI-3a [%]	0.5	0.7	0.2	0.5	5.1	0
KPI-3b	0.093	0.131	0.087	0.156	0.171	NA
KPI-3c [%]	99.8	99.8	99.9	99.9	99.8	100
KPI-3d om. [%]	0 / 99.4	0 / 99.2	0 / 99.8	0 / 99.4	0 / 94.3	0 / NA
KPI-3d com. [%]	0.2 / 93.7	0.2 / 94.2	0.1 / 97.4	0.1 / 96	0.1 / 66.9	0 / 100
KPI #	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
KPI-3a [%]	6.6	7.4	5.9	2.8	0.3	0.2
KPI-3b	0.186	0.197	0.175	0.201	0.119	0.073
KPI-3c [%]	99.9	99.9	99.9	99.9	99.9	99.9
KPI-3d om. [%]	0 / 92.7	0 / 91.8	0 / 93.5	0 / 96.7	0 / 99.7	0 / 99.8
KPI-3d com. [%]	0.1 / 60.6	0.1 / 58.4	0.1 / 62.8	0.1 / 83.8	0.1 / 97.1	0.1 / 96.8

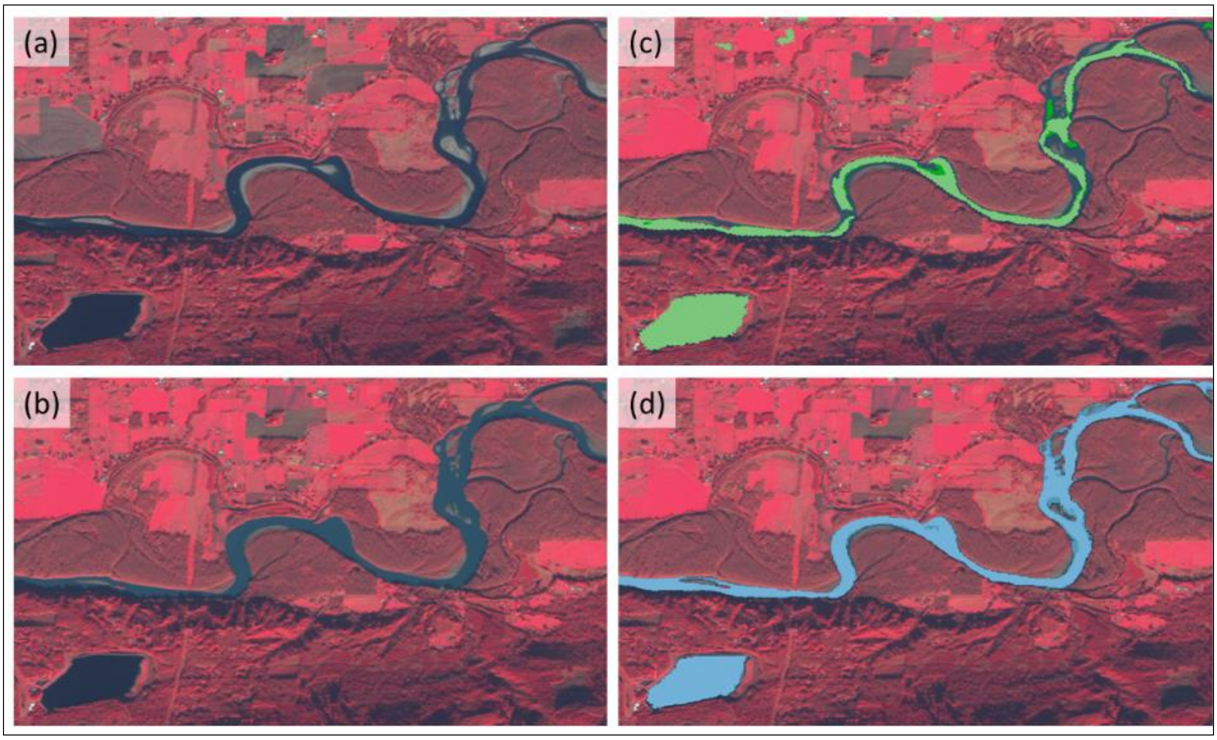


Figure 15: Flood event in USA (16.11.2021) - Permanent and monthly reference water for (a) November 2019 and (b) November 2020, and seasonal water computed by (c) the SAR-based GFM water detection algorithm, and (d) the Reference Dataset, which uses both optical and SAR data.

5.2.2 Results for flood event in South Sudan (22.11.2021)

For the flood event in South Sudan, the KPIs quantifying the thematic accuracy of the GFM Product output layers Observed Flood Extent and Reference Water Mask (which comprises both permanent and seasonal water), are presented in Table 17, Table 18, and Table 19.

For both the Observed Flood Extent and the permanent water class in the Reference Water Mask product, KPI-3a (Critical Success Index or CSI) shows very good results that exceed the target value (70-80%). As can be seen in Table 17, the CSI for the Observed Flood Extent reaches 76.9 %, while the Bias (KPI-3b) of 0.88 indicates an underestimation compared with the independent reference dataset.

The Reference Water Mask that was evaluated and considered for the flood event was computed considering the years 2019 and 2020. The permanent water mask was computed across all Sentinel-1 images available in 2019 and 2020, while the seasonal water mask is based on all Sentinel-1 images for each respective month in 2019 and 2020. Concerning the permanent water class, as can be seen in Table 18 there is a good agreement (CSI = 75.5%) between the GFM Product and the independent reference dataset, while the bias of 1.15 suggests that the permanent water class is slightly overestimated compared with the independent reference dataset.

As can be seen in Table 19 - and similar to the previously described Use Case in the USA - the twelve seasonal water masks show a rather low CSI value, ranging between 10-20% for most months. KPI-3b (Bias) is mostly below the target value of 1, which indicates that the water extent has been underestimated compared with the independent reference datasets. A more detailed analysis and explanation of this observation was provided earlier for the Use Case in USA (see Section 5.2.1).

For the Use Case in South Sudan, automated file quality checks for the GFM output layers Observed Flood Extent, Reference Water Mask, and Exclusion Mask show that the files follow the technical specifications regarding resolution, data format, and further quality parameters.

Table 17: Flood event in South Sudan (22.11.2021) – achieved KPIs for Observed Flood Extent.

KPI #	*Title*	*Observed Flood Extent*
KPI-3a	Critical Success Index	76.9 %
KPI-3b	Bias	0.879
KPI-3c	Overall accuracy	99.6 %
KPI-3d	Omission errors (no flood / flood)	0.1 % / 18.3 %
KPI-3d	Commission errors (no flood / flood)	0.3 % / 7 %

Table 18: Flood event in South Sudan (22.11.2021) – achieved KPIs for permanent water of the Reference Water Mask.

KPI #	*Title*	*Permanent Water*
KPI-3a	Critical Success Index	75.5
KPI-3b	Bias	1.151
KPI-3c	Overall accuracy	99.8
KPI-3d	Omission errors (no water / water)	0.1 % / 7.4 %
KPI-3d	Commission errors (no water / water)	0 % / 19.6 %

Table 19: Flood event in South Sudan (22.11.2021) – achieved KPIs for monthly masks of seasonal water of the Reference Water Mask.

KPI #	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*
KPI-3a [%]	9.5	10.2	7.5	10.6	12.5	14
KPI-3b	0.695	1.307	1.262	1.076	0.69	0.564
KPI-3c [%]	99.8	99.8	99.8	99.8	99.9	99.9
KPI-3d om. [%]	0.1 / 85.3	0.1 / 78.6	0.1 / 84.3	0.1 / 80.2	0.1 / 81.2	0 / 80.7
KPI-3d com. [%]	0.1 / 78.8	0.1 / 83.6	0.1 / 87.5	0.1 / 81.6	0.1 / 72.8	0.1 / 65.8
KPI #	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
KPI-3a [%]	19.7	15.6	16.3	13.7	12.4	17.2
KPI-3b	0.655	0.476	0.495	0.4	0.369	0.651
KPI-3c [%]	99.9	99.9	99.9	99.9	99.8	99.8
KPI-3d om. [%]	0 / 72.8	0 / 80.1	0 / 79	0 / 83.1	0 / 84.9	0.1 / 75.8
KPI-3d com. [%]	0.1 / 58.5	0-1 / 58.2	0.1 / 57.6	0.1 / 57.7	0.1 / 59	0.1 / 62.8

5.2.3 Results for flood event in Australia (02.03.2022)

For the flood event in Australia, the KPIs quantifying the thematic accuracy of the GFM Product output layers Observed Flood Extent and Reference Water Mask (which comprises both permanent and seasonal water), are presented in Table 20, Table 21, and Table 22.

As can be seen in Table 20, the CSI for the Observed Flood Extent is 87.6 %, thereby exceeding the target value of 70-80%, while the bias slightly below 1 indicates an under-detection compared with the independent reference dataset.

The Reference Water Mask that was evaluated and considered for this flood event was computed based on years 2019 and 2020. The permanent water mask has been computed across all Sentinel-1 images available in 2019 and 2020, while the seasonal water mask is based on all Sentinel-1 images for each respective month in 2019 and 2020. Concerning the Permanent Water class, as can be seen in Table 21, similar to the GFM Product output layer Observed Flood Extent, the CSI value of 95.8% indicates large agreement between the GFM Product and the independent reference dataset. A Bias (KPI-3b) of 1.03 implies that the extent of the permanent water class is similar to that of the independent reference dataset.

As can be seen in Table 22, similar to the findings for the Use Cases in USA and South Sudan, the GFM's twelve Seasonal Water Masks show a low CSI value, and a Bias that is constantly below the target value of 1, indicating that seasonal water is under-estimated compared with the monthly independent reference datasets. A more detailed analysis and explanation of this observation has been provided earlier for the Use Case in USA (Section 5.2.1).

For the Use Case in Australia, automated file quality checks for the GFM output layers Observed Flood Extent, Reference Water Mask, and Exclusion Mask show that the files follow the technical specifications regarding resolution, data format, and further quality parameters.

Table 20: Flood event in Australia (02.03.2022) – achieved KPIs for Observed Flood Extent.

KPI #	*Title*	*Observed Flood Extent*
KPI-3a	Critical Success Index	87.6 %
KPI-3b	Bias	0.895
KPI-3c	Overall accuracy	98.3 %
KPI-3d	Omission errors (no flood / flood)	0.2 % / 11.5 %
KPI-3d	Commission errors (no flood / flood)	1.8 % / 1.1 %

Table 21: Flood event in Australia (02.03.2022) – achieved KPIs for permanent water of the Reference Water Mask.

KPI #	*Title*	*Permanent Water*
KPI-3a	Critical Success Index	95.8
KPI-3b	Bias	1.03
KPI-3c	Overall accuracy	99.6
KPI-3d	Omission errors (no water / water)	0.3 % / 0.7 %
KPI-3d	Commission errors (no water / water)	0.1 % / 3.5 %

Table 22: Flood event in Australia (02.03.2022) – achieved KPIs for monthly masks of seasonal water of the Reference Water Mask.

KPI #	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*
KPI-3a [%]	1.3	2.9	3	3.3	2	2.7
KPI-3b	0.176	0.8	0.39	0.508	0.722	0.604
KPI-3c [%]	99.9	99.9	99.9	99.9	99.9	99.9
KPI-3d om. [%]	0 / 98.4	0 / 95	0 / 96	0 / 95.2	0 / 96.7	0 / 95.8
KPI-3d com. [%]	0.1 / 91.1	0.1 / 93.7	0.1 / 89.7	0.1 / 90.6	0.1 / 95.4	0.1 / 93.1
KPI #	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
KPI-3a [%]	1.8	1.8	0.3	1.3	0.7	1.9
KPI-3b	0.433	0.451	0.085	0.309	0.147	0.335
KPI-3c [%]	99.9	99.9	99.9	99.9	99.9	99.9
KPI-3d om. [%]	0 / 97.5	0 / 97.5	0 / 99.7	0 / 98.3	0 / 99.2	0 / 97.5
KPI-3d com. [%]	0.1 / 94.2	0.1 / 94.4	0.1 / 96.5	0.1 / 94.7	0.1 / 94.6	0.1 / 92.6

5.3 Discussion of results of Phase II accuracy assessment for the three Use Cases

5.3.1 Discussion of results for flood event in USA (16.11.2021)

As can be seen in Figure 16 below, the GFM Product output layer Observed Flood Extent and the independent Reference Dataset show a high level of agreement for the three main flooded areas that have been detected within the Use Case. As is shown in Figure 17, the GFM Ensemble differs from the independent Reference Dataset only for small patches and border pixels, but the overall flooding pattern is well captured.

The validation results for the permanent water class of the GFM Product output layer Reference Water Mask, also indicate considerable agreement between both datasets. As can be seen in Figure 18, slight differences exist – for example, sometimes what is classified as permanent water by the GFM Product is classified as seasonal water in the Reference Dataset, and vice versa. This applies in particular to border pixels of water bodies, transition zones in

river deltas and coastlines, as is highlighted in Figure 18. This is presumably due to the uncertainty in selecting the threshold between water and non-water areas. Based on a visual inspection of the data, and without access to very high resolution reference information for the flood events, it is not possible to specify exactly the correct threshold.
In Figure 18, which shows the reference (i.e. permanent and seasonal) water for the month of November, the indicated omission errors are classified in the GFM dataset as seasonal water for most months throughout the observation period.

As was mentioned earlier (Section 5.2.1), differences in the reference (i.e. permanent and seasonal) water between the GFM Ensemble and Reference Dataset are particularly evident along river sections. Rivers are generally well detected in both datasets. However, in some cases (see Figure 19) the GFM Product tends to underestimate river sections in the permanent water class, instead classifying them in some of the seasonal (monthly) masks. Including Sentinel-2 data with 10 metres resolution in the processing chain of the independent Reference Dataset (compared with the GFM Product's resolution of 20 metres), might allow better detection of such areas. Conversely, the GFM Product classifies some water bodies as permanent water which are not covered by the Reference Dataset.

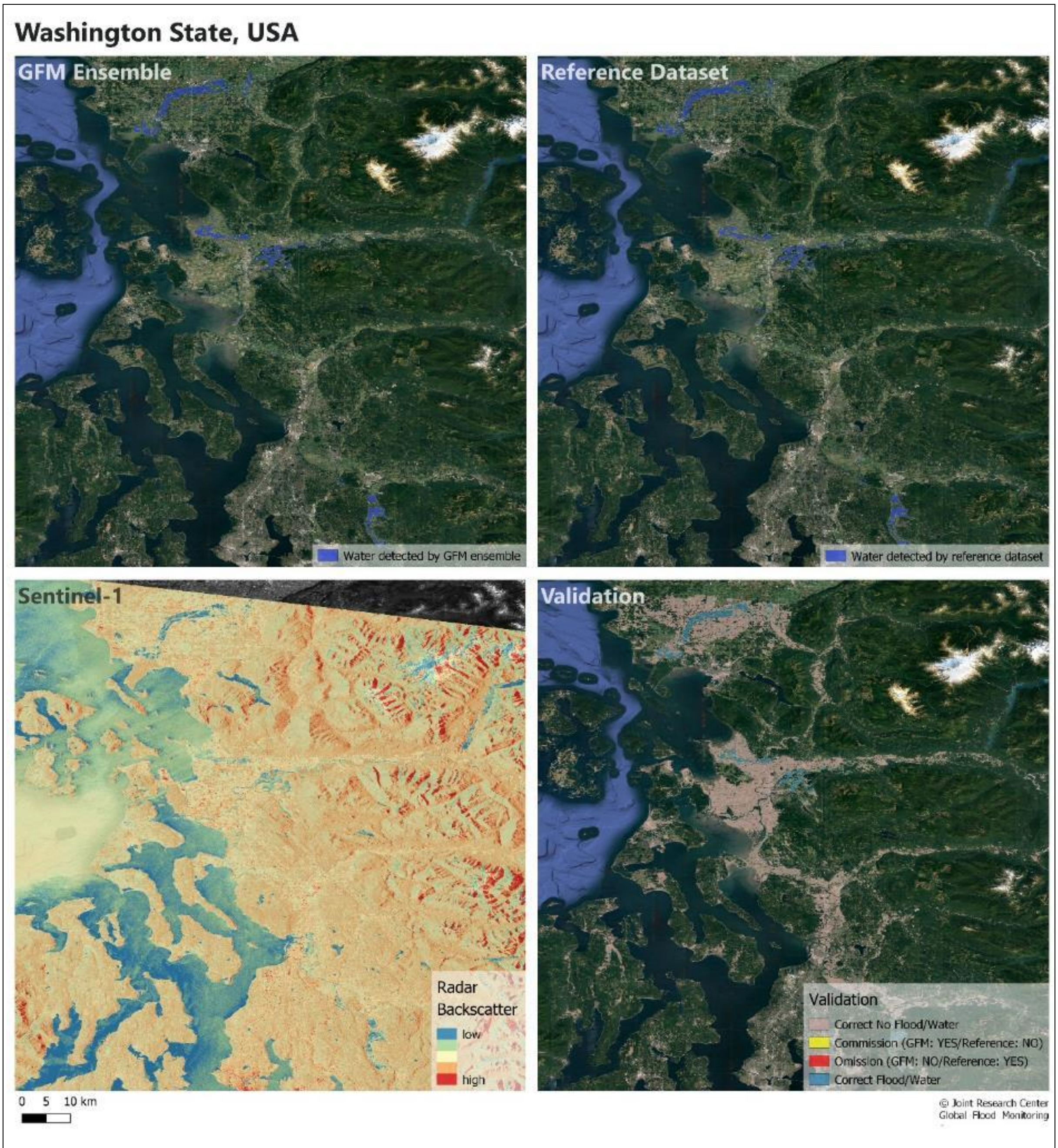


Figure 16: Flood event in USA (16.11.2021) - SAR-based water detection results from the GFM Ensemble (top-left) and Reference Dataset (top-right), Sentinel-1 backscatter image (bottom-left), and omission / commission errors (bottom-right).

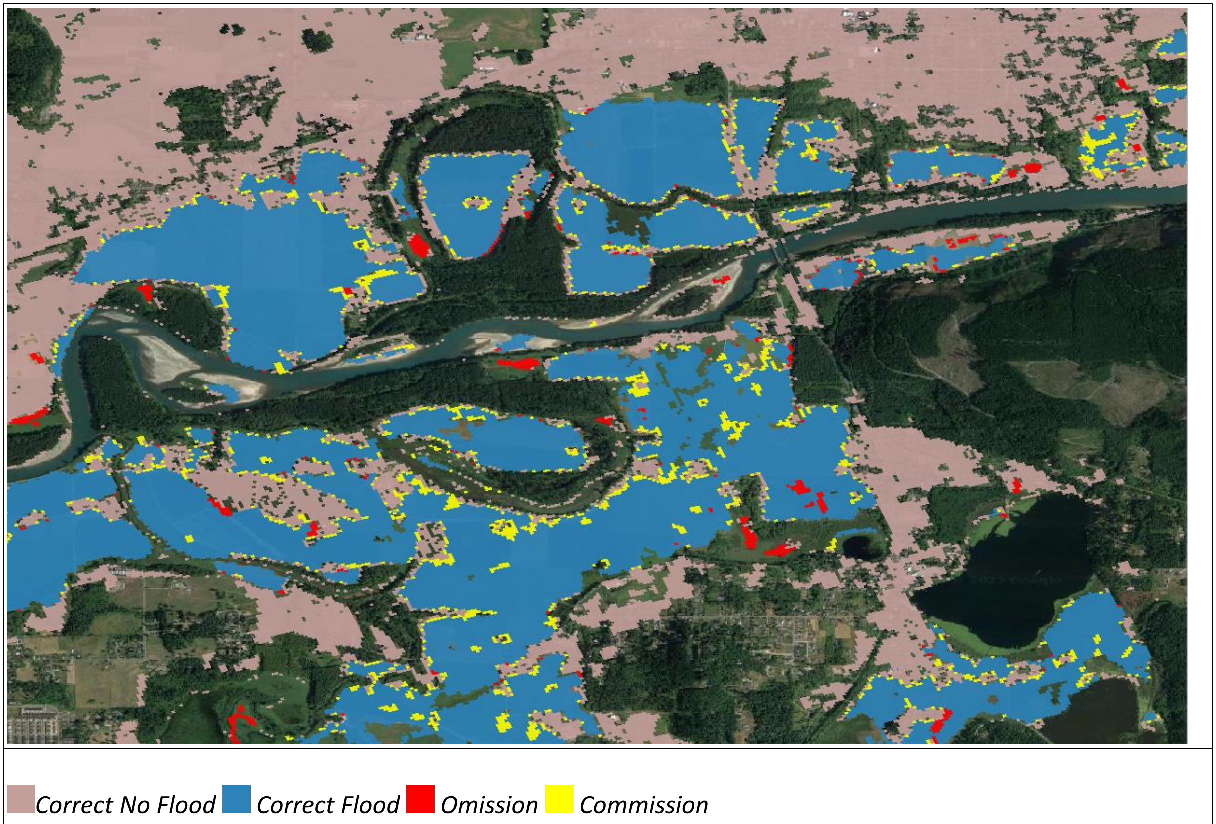


Figure 17: Flood event in USA (16.11.2021) - difference map between the GFM output layer Observed Water Extent and the Reference Dataset, showing omission / commission errors, and flood and no flood agreement.

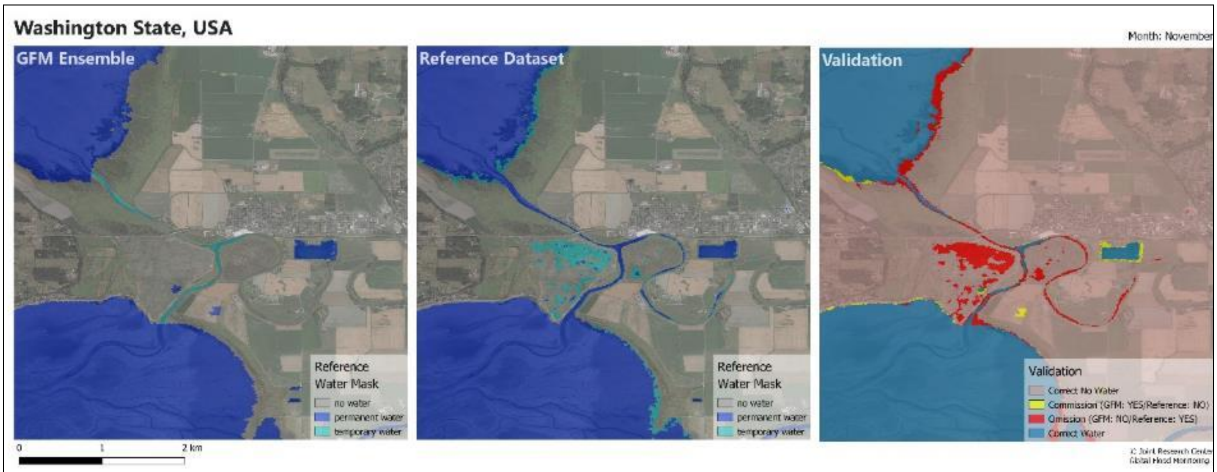


Figure 18: Flood event in USA (16.11.2021) – monthly Reference Water Mask for November from the GFM Ensemble (left) and Reference Dataset (middle), and omission / commission errors for permanent and seasonal water (right). Omission errors along the coast are clear.

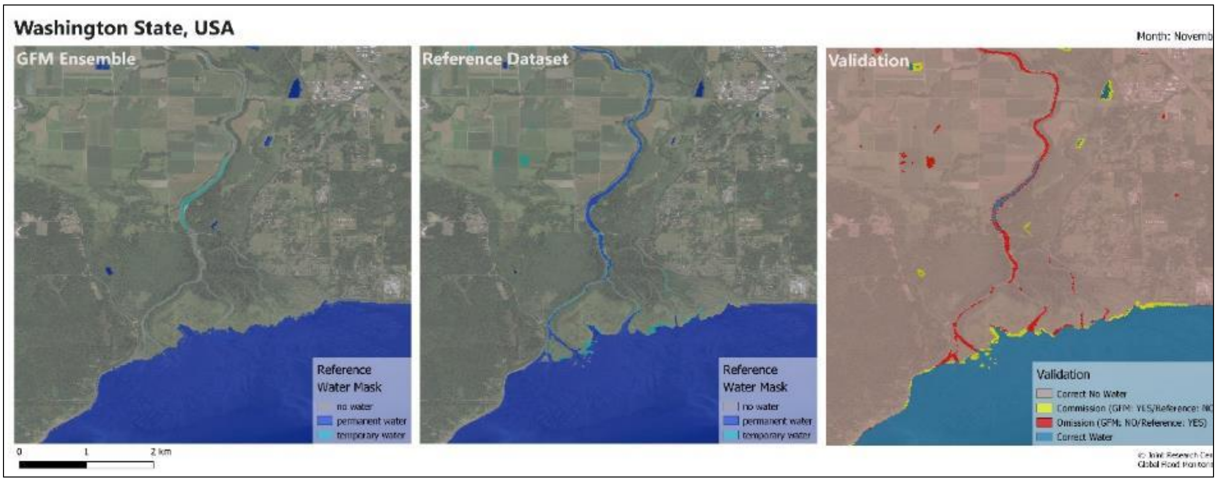
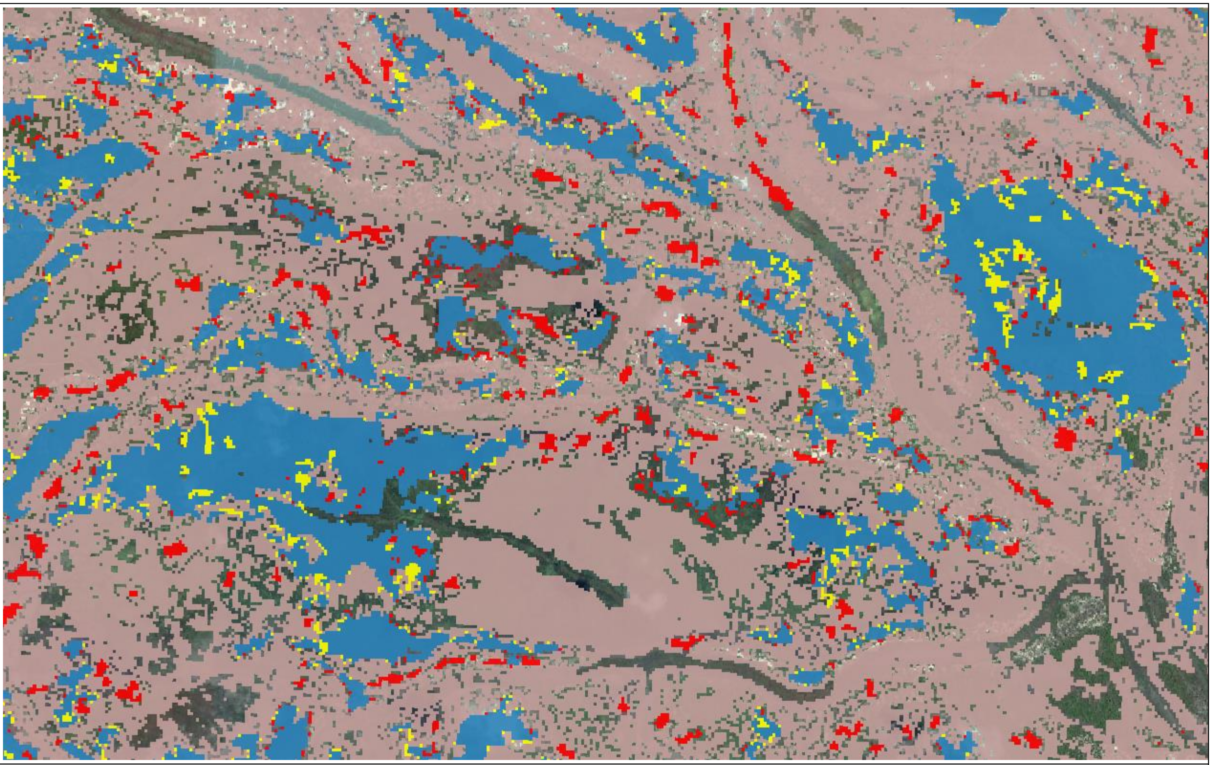


Figure 19: Flood event in USA (16.11.2021) – monthly Reference Water Mask for November from the GFM Ensemble (left) and Reference Dataset (middle), and omission / commission errors for permanent and seasonal water (right). Omission errors along the river are clear.

5.3.2 Discussion of results for flood event in South Sudan (22.11.2021)

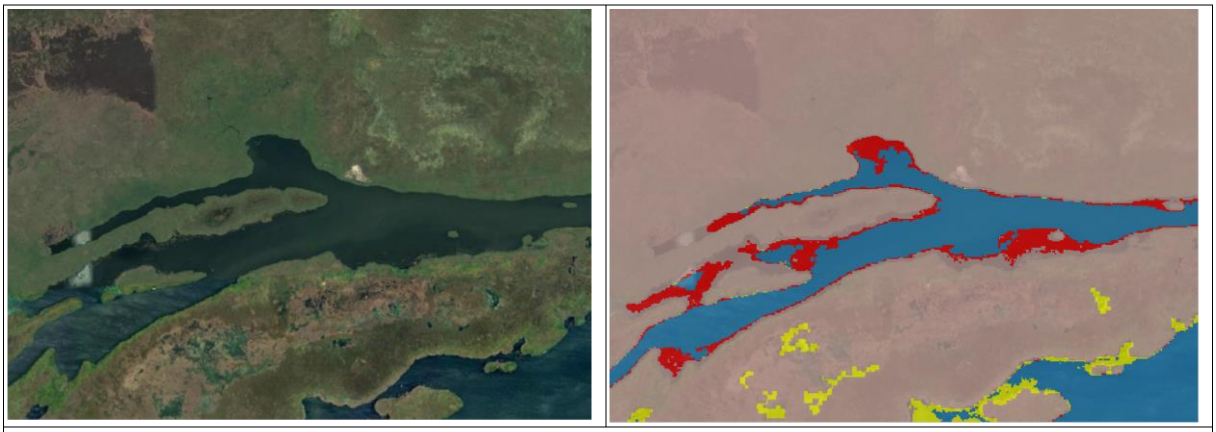
As can be seen in Figure 20, for the flood event in South Sudan, there is good agreement between the flood extent captured by the GFM Product and the independent Reference Dataset. Only smaller patches are over- and under-estimated by the GFM Ensemble compared with the Reference Dataset (with a tendency to under-estimate the flooded area, as suggested by a bias of 0.88, as shown in Table 17).

Similar to the Use Case in USA, for the flood event in South Sudan both the GFM Product and the independent Reference Dataset are largely in agreement regarding the permanent water class of the Reference Water Mask. Again, some differences arise due to water being classified as permanent water in the Reference Dataset, but as seasonal (monthly) in the GFM Ensemble, and vice versa. This is illustrated in Figure 21 and Figure 22 below. In Figure 21, the map on the right highlights (in red) those areas that are permanent water in the Reference Dataset, but not in the GFM Product. However, these areas are mostly captured in the monthly (seasonal) water of the GFM Product’s Reference Water Mask. This is illustrated in Figure 22, which shows the GFM Product’s reference (permanent and seasonal) water for four selected months (i.e. January, April, August and December).



Correct No Flood Correct Flood Omission errors Commission errors

Figure 20: Flood event in South Sudan (22.11.2021) - difference map between the GFM output layer Observed Water Extent and the Reference Dataset, showing omission / commission errors, and flood and no flood agreement.



Correct No Permanent Water Correct Permanent Water Omission errors Commission errors

Figure 21: Flood event in South Sudan (22.11.2021) - difference map between the permanent water class of the GFM Ensemble and the Reference Dataset (right), showing omission / commission errors, and water and no water agreement along a river area.

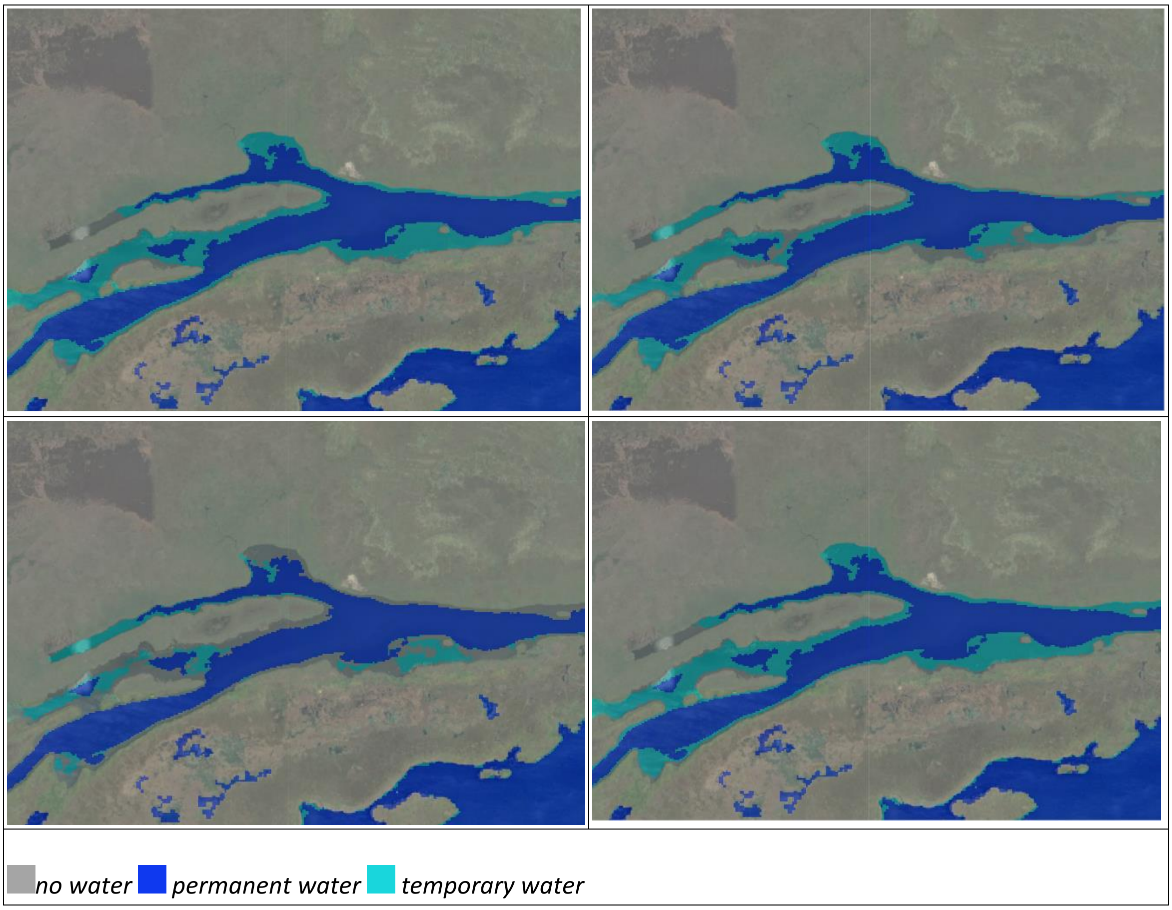


Figure 22: Flood event in South Sudan (22.11.2021) – examples of the monthly Reference Water Masks for January (top-left), April (top-right), August (bottom-left), and December (bottom-right) generated by the GFM Ensemble. These areas are mostly permanent water in the Reference Dataset.

5.3.3 Discussion of results for flood event in Australia (02.03.2022)

Again for the flood event in Australia (02.03.2022), good results were achieved. As can be seen in Figure 23 and Figure 24, the GFM Product and the independent Reference Dataset produce similar results, with the GFM Product underestimating the flood extent for smaller patches, compared with the Reference Dataset. Regarding the reference (i.e. permanent and seasonal) water extent, as can be seen in Figure 25, the GFM Product tends to overestimate slightly the extent of the permanent water compared with the Reference Dataset. As can be seen, rivers are well captured by both datasets, with the GFM Product again slightly overestimating.

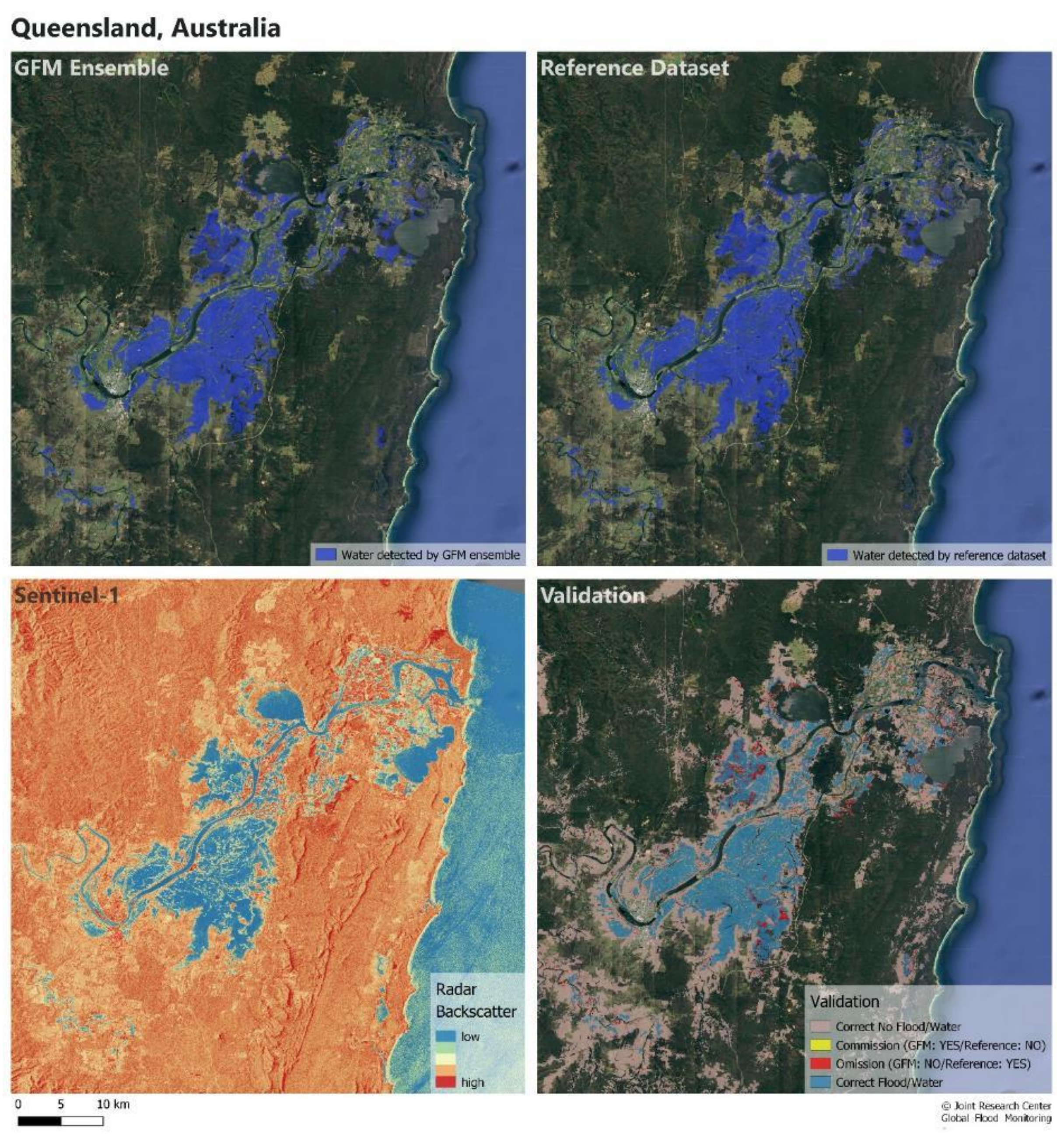


Figure 23: Flood event in Australia (02.03.2022) - SAR-based water detection results from the GFM Ensemble (top-left) and Reference Dataset (top-right), Sentinel-1 backscatter image (bottom-left), and omission / commission errors (bottom-right).

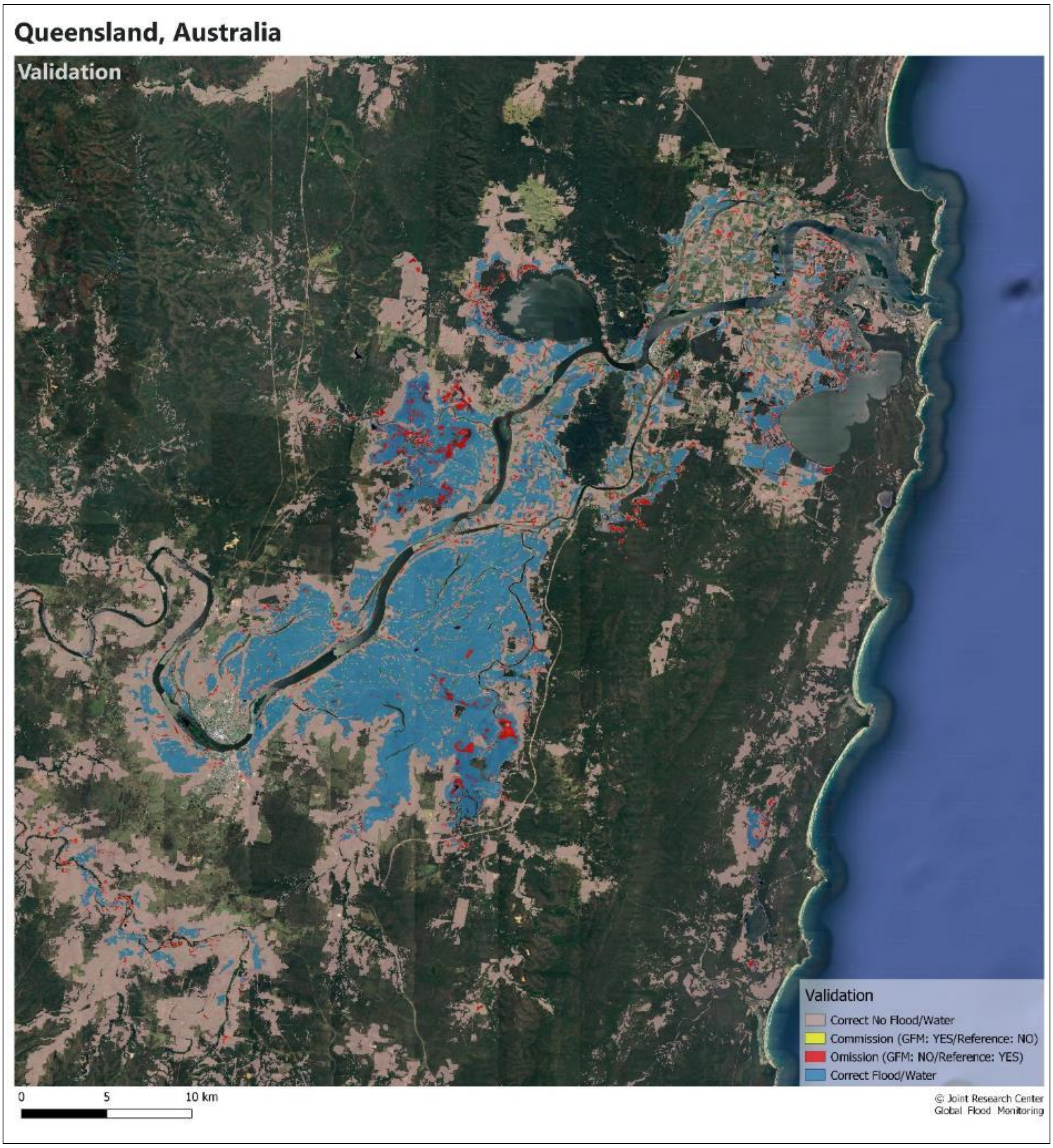


Figure 24: Flood event in Australia (02.03.2022) – Example highlighting the high level of agreement between the GFM Ensemble and the Reference Dataset (blue), and the slight underestimation of the GFM Ensemble compared with the Reference Dataset (red).

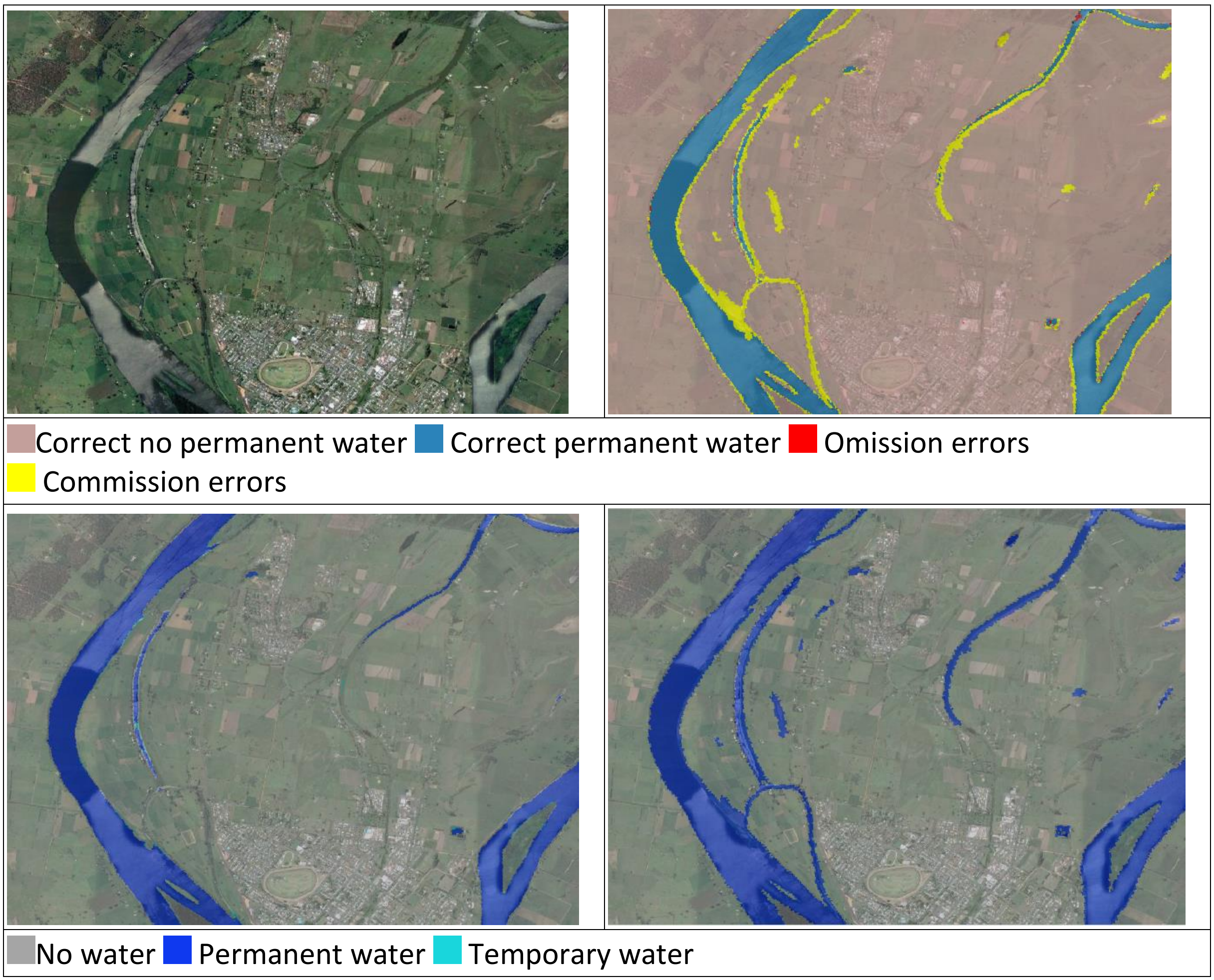


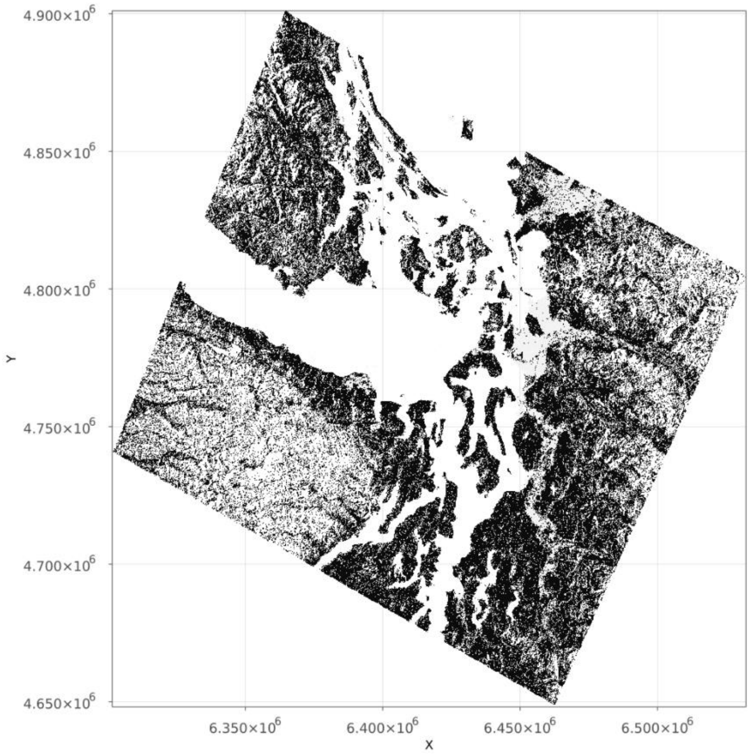
Figure 25: Flood event in Australia (02.03.2022) – Difference map (top-right) between the monthly Reference Water Masks for March computed using the Reference Dataset (bottom-left) and the GFM Ensemble (bottom-right), showing commission errors along a river.

5.4 Plausibility analysis of the Exclusion Mask for the three Use Cases

This Section provides a brief analysis of the general plausibility of the GFM Product output layer Exclusion Mask, for some Use Case areas selected for the Phase II accuracy assessment. As explained earlier in Section 2, the Exclusion Mask denotes areas where SAR-based water detection and flood mapping are not feasible, and is created by combining information layers describing five types of “static” ground surface characteristics: **no sensitivity** (e.g. urban areas, dense vegetation); water look-alikes (e.g. flat impervious areas, sand surfaces) due to **low backscatter**; areas where the Sentinel-1 signals are affected by **topographic distortions**; **radar shadows**; and areas with **low coverage** of Sentinel-1 observations.

Regarding the flood event in the USA (16.11.2021), the main characteristics of the Exclusion Mask for this Use Case area are summarized in Table 23. As can be seen in Figure 26, in this case the “no-sensitivity” layer of the Exclusion Mask also includes flood-prone areas such as agricultural areas along rivers and vegetated river-banks. The same issue is also highlighted in the case of the flood event in Australia (02.03.202), as described below.

Table 23: Flood event in USA (16.11.2021) – Overview of the GFM Product output layer Exclusion Mask for this Use Case area. Note that some of the Exclusion Mask’s sub-layers overlap.



Area of Interest	70,297	km² (total)
Excluded area	12,987.0	km² (18.5%)
No sensitivity area	11,060.4	km² (15.7%)
Low backscatter area	19.9	km² (0.03%) (Not including pixels from the sea - 255 in Exclusion Mask)
Topographic distortion area	10.898.9	km² (15.5%)
Radar shadow area	10.5	km² (0.01%)

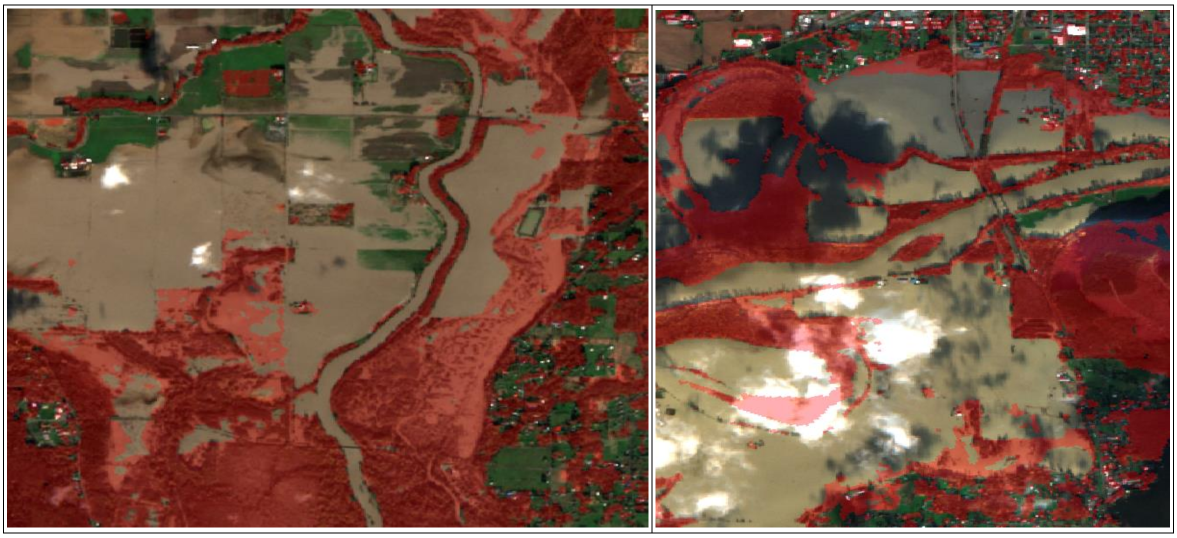


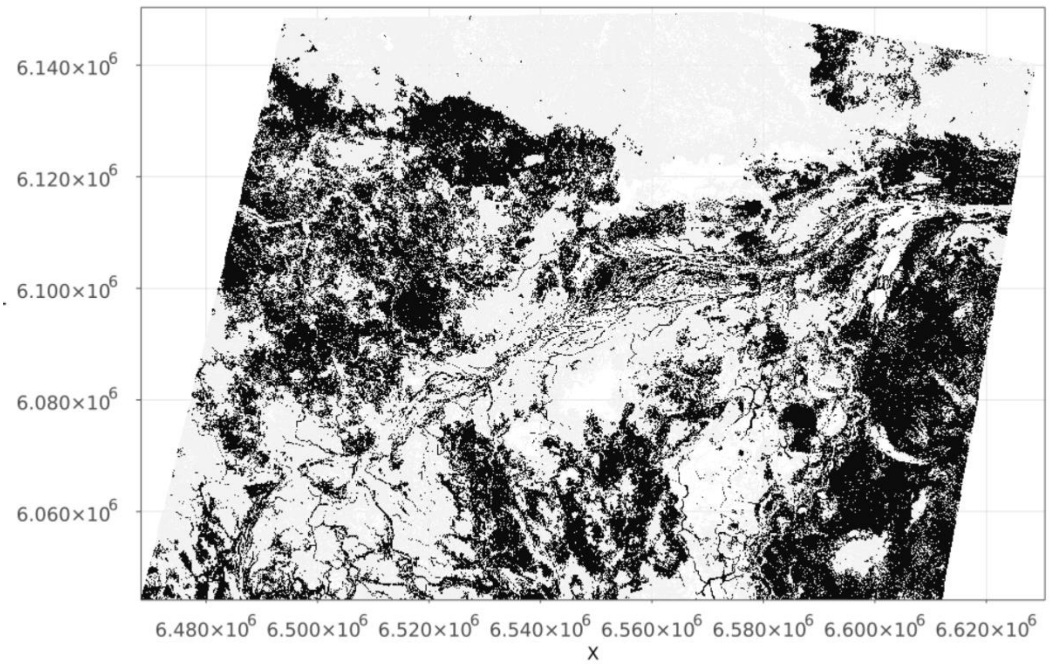
Figure 26: Flood event in USA (16.11.2021) - Sentinel-2 image on 16.11.2021 of the Use Case area, overlaid with the Exclusion Mask's "no-sensitivity" layer (in red), which in this example excludes vegetated areas along a river.

Regarding the flood event in South Sudan (22.11.2021), the main characteristics of the Exclusion Mask for this Use Case area are summarized in Table 24. For the defined area in South Sudan, the Exclusion Mask seems plausible, and the various input layers do not show any inconsistencies.

Regarding the flood event in Australia (02.03.2022), the main characteristics of the Exclusion Mask for this Use Case area are summarized in Table 25. For the defined Use Case in Australia, the "topographic distortion" layer of the Exclusion Mask seems plausible. However, as can be seen in Figure 27 and Figure 28 below, the "no sensitivity" layer of the Exclusion Mask seems to be very strict, and covers agricultural areas near rivers, as well as river-beds and river-banks along the river, thereby masking out flood-prone areas.

Generally speaking, for all of the considered areas of interest, the "topographic distortion" and the "low backscatter" layers of the GFM Product output layer Exclusion Mask, appear to be reasonable and plausible for all areas of interest, and do not exclude flood-prone areas. The highest uncertainty, however, seems to be in the Exclusion Mask's "no sensitivity" layer, as was shown above in the Use Case areas for the flood events in USA (16.11.2021) and Australia (02.03.2022), where agricultural areas and river-banks, for example, are partially excluded from the flood detection even though these are flood-prone areas.

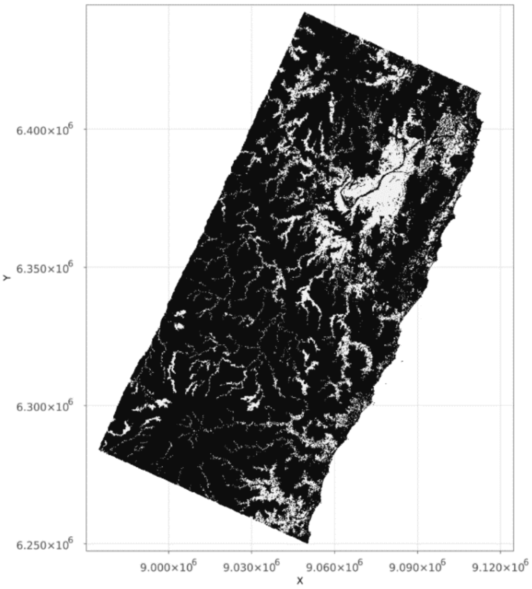
Table 24: Flood event in South Sudan (22.11.2021) – Overview of the GFM Product output layer Exclusion Mask for this Use Case area. Note that some of the Exclusion Mask's sub-layers overlap.



Area of Interest	27,514.0	km² (total)
Excluded area	5,552.4	km² (20.2%)
No sensitivity area	5,683.4	km² (20.7%) Note: Initial no-sens.
Low backscatter area	1.3	km² (0.00%)

Topographic distortion area	0.053	km^2 (0.0002%)
Radar shadow area	NA	

Table 25: Flood event in Australia (02.03.2022) – Overview of the GFM Product output layer Exclusion Mask for this Use Case area. Note that some of the Exclusion Mask's sub-layers overlap.



Area of Interest	16,067.0	km² (total)
Excluded area	11,622.1	km² (72.3%)
No sensitivity area	8,872.49	km² (55.2%)
Low backscatter area	247.87	km² (1.5%)
Topographic distortion area	9,408.93	km² (58.6%)
Radar shadow area	119.72	km² (0.75%)

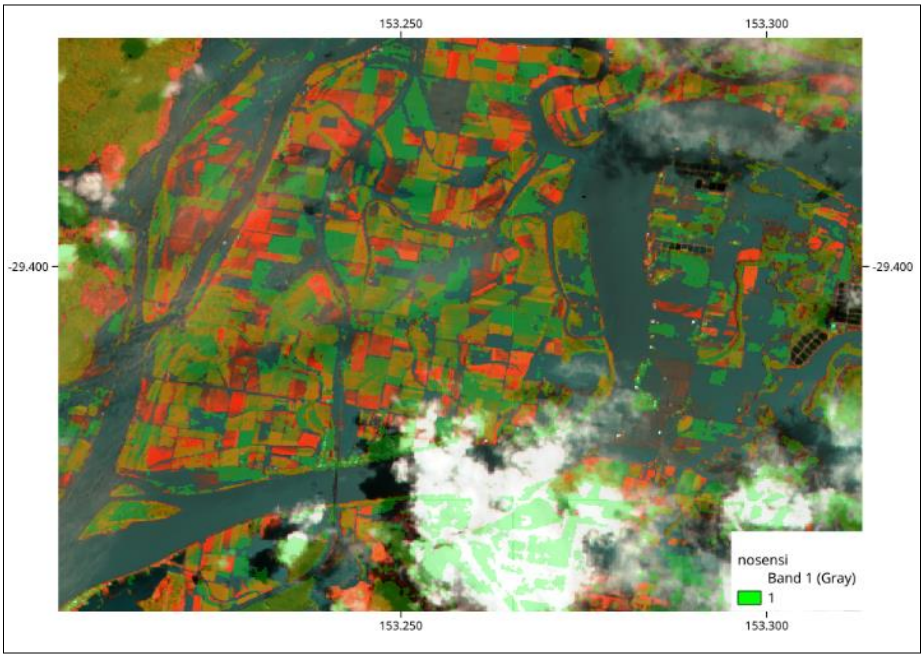


Figure 27: Flood event in Australia (02.03.2022) –Sentinel-2 image on 01.03.2022 of the Use Case area, overlaid with the Exclusion Mask's "no-sensitivity" layer (in green), which in this example covers agricultural fields, therefore excluding typical flood-prone areas.

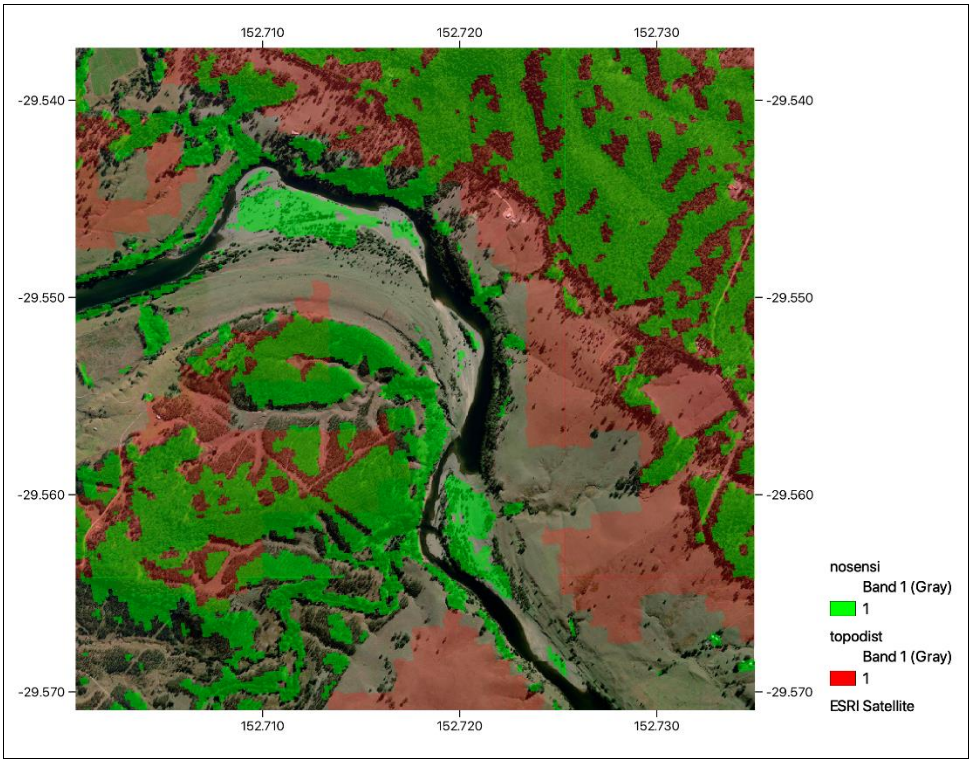


Figure 28: Flood event in Australia (02.03.2022) –Sentinel-2 image on 01.03.2022 of the Use Case area, overlaid with the Exclusion Mask's "no-sensitivity" layer (in green), which in this example excludes vegetated areas along a river.

6 Conclusions and Outlook

This report presents the first results of the independent quality assessment of the new Global Flood Monitoring (GFM) Product of the Copernicus Emergency Management Service, based on a three-month period of pre-operational testing. The initial (i.e. Phase I) quality assessment focused on the GFM Product's main output layer - namely Observed Flood Extent - which provides a continuous monitoring of floods worldwide. The second (i.e. Phase II) quality assessment also included the validation of the GFM Product output layer Reference Water Mask, which delineates permanent and seasonal water. In the Phase II quality assessment, the general plausibility of the GFM Product output layer Exclusion Mask, which delineates areas where SAR-based water mapping is not feasible (due to a variety of static ground surface characteristics), was also considered.

The quality assessment applied appropriate sampling procedures in order to assess systematically the thematic quality of the GFM Product, based on a representative set of Use Cases of worldwide flood events. The output quality of the service is measured using the mandatory Key Performance Indicator 3 (KPI-3), i.e. Product Quality (or thematic accuracy). While the validation exercise has been performed internally by the consortium responsibole for implementing and operating the GFM Product, the independence of the process was ensured through validation by a partner in the consortium that was not involved in the actual methodological development or product generation.

In the Phase I quality assessment, six selected Use Cases of worldwide flood events have been evaluated with regard to the GFM Product output layer Observed Flood Extent. Although for the six Use Cases in the Phase I assessment, the targeted thematic accuracy of 70-80% (based on the Critical Success Index or CSI) was not reached, a visual analysis of all selected flood events showed promising results and accurate flood mapping, especially for the flood events in Belgium (CSI = 69.3 %) and China (CSI = 65.8 %).

In the Phase II quality assessment, three additional Use Cases of worldwide flood events were assessed. This assessment, which also evaluated the GFM Product output layer Reference Water Mask, showed overall good results for both Observed Flood Extent as well as the reference water mask, with CSI values exceeding the targeted value. Evaluation of the seasonal water part of the reference water mask resulted in low CSI values for most areas and all considered months. This is mainly due to the fact that the seasonal water extent – which is computed as monthly observed water extent minus permanent water extent - is small in area, and so pixel-based "noise" has a big impact on the summary statistics.

Throughout the set-up phase of the GFM Product, various technical improvements were continuously made key elements of the SAR-based flood and water detection methodology. However, it was not within the scope of this pre-operational product and service quality assessment report to consider the effects on thematic accuracy of specific improvements in the pre-operational GFM Product methodology.

It should be noted that, in accordance with the Technical Specifications (TS), the pre-operational quality assessment which is described in this report was applied to the validation and assessment of the flood mapping algorithms used in the near real-time product generation during the set-up phase of the GFM Product. As required by the TS, for each year of the operational running of the GFM Product, the thematic accuracy of the output layers of global flood information, which are continuously generated in NRT, are constantly monitored and reported - together with other performance-related aspects of the service and product delivery (e.g. timeliness of production, stability of service) - on a quarterly basis, and published each year in an annual quality assessment report.

Finally, as was mentioned earlier in Section 2, in addition to the continuous, NRT generation of output layers of global flood information, the GFM Product is also being used to process the entire time-series of Sentinel-1 SAR imagery (covering the period 2015-2021), in order to generate an archive of observed worldwide flood events. The thematic accuracy of the generated archive of flood events will also be assessed based on a visual interpretation of sampling locations, and reported in the annual quality assessment reports.

Bibliography

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 (15), 3673. https://doi.org/10.3390/rs14153673

Card, D.H. (1982). Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogrammetric Engineering and Remote Sensing, 48, 431–439.

Chini M., Hostache R., Giustarini L., Matgen P. (2017), A Hierarchical Split-Based Approach for parametric thresholding of SAR images: flood inundation as a test case, IEEE Transactions on Geoscience and Remote Sensing, 55 (12), 6975-6988. https://doi.org/10.1109/TGRS.2017.2737664

Bradley, D. and Roth, G. (2007), "Adaptive Thresholding using the Integral Image." Journal of Graphics Tools 12 (2), https://doi.org/10.1080/2151237X.2007.10129236

European Commission (2020), Provision of an Automated, Global, Satellite-based Flood Monitoring Product for the Copernicus Emergency Management Service. Call for Tenders JRC/IPR/2020/OP/0551, Open Procedure, Tender specifications - Part 2: Technical specifications of 10/03/2020. https://etendering.ted.europa.eu/cft/cft-display.html?cftId=6052

Frantz, D., Haß, E., Uhl, A., Stoffels, J., and Hill, J. (2018), Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. Remote sensing of environment, 215, 471-481.

Hartigan, J. A. and Hartigan, P. M. (1985), The Dip Test of Unimodality. Ann. Statist. 13 (1) 70 – 84.

Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., and Riffler, M. (2019), A highly automated algorithm for wetland detection using multi-temporal optical satellite data. Remote Sensing of Environment, 224, 333-351.

Martinis, S., Twele, A., Kersten, J. (2015), A fully automated TerraSAR-X based flood service. ISPRS Journal of Photogrammetry and Remote Sensing, 104, 203-212. http://dx.doi.org/10.1016/j.isprsjprs.2014.07.014

Martinis, S., Twele, A., and Voigt, S. (2009), Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data, Nat. Hazards Earth Syst. Sci., 9, 303–314, https://doi.org/10.5194/nhess-9-303-2009

Matgen, P., Martinis, S., Wagner, W., Freeman, V., Zeil, P., and McCormick, N., (2020), Feasibility assessment of an automated, global, satellite-based flood-monitoring product for the Copernicus Emergency Management Service. JRC Technical Report. EUR 30073 EN. Publications Office of the European Union, Luxembourg. ISBN 978-92-76-10254-0. 47p. https://doi.org/10.2760/653891

Metzger, M.J. (2018), The Global Environmental Stratification: A high-resolution bioclimate map of the world [dataset]. The University of Edinburgh. https://doi.org/10.7488/ds/2354

Olofsson, P., Foody, G.M., Stehman, S.V., and Woodcock, C.E. (2013), Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment, 129, 122–131.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014), Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42-57.

Otsu, N. (1979), A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J. (2008), HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment, 112(9), 3469–3481.

Roberts, D., Mueller, N. and Mcintyre, A., (2017), "High-Dimensional Pixel Composites From Earth Observation Time Series." IEEE Transactions on Geoscience and Remote Sensing, Nov, 55(11). https://doi.org/10.1109/TGRS.2017.2723896

Salamon, P., McCormick, N., Reimer, C., Clarke, T., Bauer-Marschallinger, B., Wagner, W., Martinis, S., Chow, C., Böhnke, C., Matgen, P., Chini, M., Hostache, R., Molini, L., Fiori, E., Walli, A., (2021), The new, systematic global flood monitoring product of the Copernicus Emergency Management Service. International Geoscience and Remote Sensing Symposium (IGARRS) 2021. Paper WE4.O-6.4. https://doi.org/10.1109/IGARSS47720.2021.9554214

Twele, A., Cao, W., Plank, S. and Martinis, S. (2016), Sentinel-1-based flood mapping: a fully automated processing chain, International Journal of Remote Sensing, 37:13, 2990-3004.

Wagner, W., Freeman, V., Cao, S., Matgen, P., Chini, M., Salamon, P., McCormick, N., Martinis, S., Bauer-Marschallinger, B., Navacchi, C., Schramm, M., Reimer, C., and Briese, C., (2020), Data processing architectures for monitoring floods using Sentinel-1. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3-2020, 641–648. https://doi.org/10.5194/isprs-annals-V-3-2020-641-2020

Wilks, D. S. (2011), Statistical methods in the atmospheric sciences (Vol. 100). Academic Press.

List of Figures

List of Tables

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union