High-Resolution Flood Monitoring Based on Advanced Statistical Modeling of Sentinel-1 Multi-Temporal Stacks

Rosa Colacicco 1,*, Alberto Refice 2, Raffaele Nutricato 3, Fabio Bovenga 2, Giacomo Caporusso 2, Annarita D’Addabbo 2, Marco La Salandra 1, Francesco Paolo Lovergine 2, Davide Oscar Nitti 3 and Domenico Capolongo 1

1 Department of Earth and Geoenvironmental Sciences, University of Bari, 70125 Bari, Italy; marco.lasalandra@uniba.it (M.L.S.); domenico.capolongo@uniba.it (D.C.)
2 Institute for Electromagnetic Sensing of the Environment, National Research Council (IREA CNR), 70126 Bari, Italy; alberto.refice@cnr.it (A.R.); fabio.bolvergine@cnr.it (F.B.); caporusso.g@irea.ba.cnr.it (G.C.); daddabbo.a@irea.cnr.it (A.D.); francesco.lovergine@cnr.it (F.P.L.)
3 GAP srl c/o Department of Physics “M. Merlin”, University of Bari, 70125 Bari, Italy; raffaele.nutricato@gapsrl.eu (R.N.); davide.nitti@gapsrl.eu (D.O.N.)

* Correspondence: rosa.colacicco@uniba.it

Abstract: High-resolution flood monitoring can be achieved relying on multi-temporal analysis of remote sensing SAR data, through the implementation of semi-automated systems. Exploiting a Bayesian inference framework, conditioned probabilities can be estimated for the presence of floodwater at each image location and each acquisition date. We developed a procedure for efficient monitoring of floodwaters from SAR data cubes, which adopts a statistical modelling framework for SAR backscatter time series over normally unflooded areas based on Gaussian processes (GPs), in order to highlight flood events as outliers, causing abrupt variations in the trends. We found that non-parametric time series modelling improves the performances of Bayesian probabilistic inference with respect to state-of-the-art methodologies using, e.g., parametric fits based on periodic functions, by both reducing false detections and increasing true positives. Our approach also exploits ancillary data derived from a digital elevation model, including slopes, normalized heights above nearest drainage (HAND), and SAR imaging parameters such as shadow and layover conditions. It is here tested over an area that includes the so-called Metaponto Coastal Plain (MCP), in the Basilicata region (southern Italy), which is recurrently subject to floods. We illustrate the ability of our system to detect known (although not ground-truthed) and smaller, undocumented inundation events over large areas, and propose some consideration about its prospective use for contexts affected by similar events, over various land cover scenarios and climatic settings.

Keywords: flood monitoring; flood mapping; remote sensing; satellite imagery; Sentinel-1; Bayesian framework; time series analysis; Basilicata region

1. Introduction

Changing climate can exacerbate disaster risk in a variety of ways, e.g., increasing frequency and intensity of extreme events. Floods are among the most impacting environmental hazards; their costs in terms of human lives, infrastructure damage or loss, and agricultural impact can be enormous and continue to increase [1–3]. Such events are conditioned by multidimensional, interrelated factors such as extent and duration of rainfall, type of interested river basin, and soil moisture percentage before rainfall, among many others [4]. Understanding the phenomenon through monitoring turns out to be very important in order to design effective solutions, while respecting socioeconomic and environmental constraints [5]. Flood risk assessment and management have also been addressed by the European Union Flood Directive (2007/60/EC, 2007), which requires, for...
each member state, the redaction of hazard maps, the detection of risk areas, and the setup of effective monitoring and alerting systems.

Satellite remote sensing is a tremendous tool to observe and study the Earth, its land surface, the oceans, the atmosphere, and its dynamics from space, and has vital importance for the effective and sustainable future management of the planet [6]. The potential of satellite images for flood monitoring and mapping has been known for at least 40 years. Recently, the steadily increasing flow of high-resolution and open-access data, with high reliability, improved revisit time, and global geographical coverage, such as those arising from the European Space Agency’s constellation of the Sentinel satellites, is leading to a paradigm shift in this field of research. A recent review article [7] shows how the European Copernicus program with its Sentinel satellite constellation can be useful in various phases of flood management. This also triggers the development of new tools such as cloud-based platforms, e.g., Google Earth Engine [8–10] and the openEO platform [11], and different algorithms to extract information and knowledge about the effects and temporal developments of flood events.

Since natural disaster scenarios, and in particular floods, are typical examples of complex situations, different factors have to be considered to provide accurate and robust assessments [12]. New methodologies have been proposed to model the behavior of different ‘objects’ in remotely sensed images, in order to associate them to flood or no-flood conditions, producing, for example, “uncertain” flood maps based on fuzzy set theory [13,14] or adopting a Bayesian framework (Bf).

In particular, over the past decade, the Bf has been widely used for forecasting [15], detecting [16], modelling [17], monitoring, and thus mapping flood events [18–23]. The framework involves the estimation of posterior probabilities of occurrence of an event conditioned by some observations, through application of the Bayes equation, as a function of probabilities of each observation, conditioned on the presence or absence of the event (called likelihoods) and prior absolute probabilities of the same events. The results of this approach, which are probability maps, offer a series of advantages with respect to other, deterministic methodologies, e.g., the possibility of interpreting the computed probabilities in terms of uncertainty, or confidence of the occurrence of an event.

Among satellite products, Synthetic Aperture Radar (SAR) data allow researchers to track flood events in all-weather and day/night conditions [24,25]. Several approaches for the detection of flooding with SAR sensors have been developed, depending on spatial and temporal resolution, as well as on the number of images used as input, e.g., single-image image thresholding (SI) [26], change detection (CD) [27,28], or time series analysis (TSA) [29–31]. These three approaches involve increasing quantities of data and correspondingly increasing levels of accuracy. SI algorithms rely on the principle that calm, open-water surfaces act basically as mirrors for the impinging e.m. radiation, leading to very low backscatter toward the sensor. Dark image areas therefore generally correspond to water, although several look-alikes are typically present, caused, e.g., by smooth surfaces such as road tarmacs, or due to geometric effects such as radar shadows. See, for instance, the several dark patches appearing in the SAR intensity image corresponding to a flood event in Figure 4, many of which do not actually correspond to water, as from the flood map in Figure 3, and as explained in the paper later on. Moreover, flooded, open-water areas are often roughened by wind, which may increase backscatter. Similarly, especially over complex land cover such as agricultural terrains or urban areas, the presence of water may induce backscatter increase rather than decrease, due to multiple interactions of the radar waves with the objects on the ground, e.g., vegetation stems or buildings. These phenomena may cause underdetection. The CD approach reduces some of the above-mentioned problems, as it involves the comparison of a co-event image with at least one pre-event image. The choice of the pre-event image, however, can be critical, as often backscatter levels change continuously over the land surface, so that additional assumptions are often necessary in order to isolate backscatter changes due to floods. For these reasons, recent works have recurred to TSA, in which radar backscatter conditions over
Normally unflooded terrain are assumed to follow some predetermined trend, estimated through proper temporal modelling of backscatter levels. In this approach, flood events are expected to behave as abrupt and short-lived changes in the temporal dimension, which can be detected as unmodeled outliers, or “spikes” in the backscatter time series. The TSA approach has been recently proven to be the one yielding the best results in terms of accuracy [32], although objective assessment of such algorithms is admittedly difficult.

Regarding the temporal modeling, in spite of the overwhelmingly large amount of terrain types and ground environments potentially present over SAR imagery, the most used type of temporal model for the radar backscatter is a sum of harmonic terms with periods corresponding to a few integer fractions of a year (typically up to three), optionally added to a linear trend [33]. Harmonic models actually have a long history of applications in many time series studies of remotely sensed data, both optical and radar. Moreover, the parameters of such models have a set of desirable properties, such as their direct interpretability and the relative ease of estimation (which basically consists of a linear regression). In fact, the harmonic temporal parameters estimated globally over SAR backscatter data stacks have been recently announced to be available as a side product of a global flood service [34], and constitute a useful dataset to be further studied. Nevertheless, as better explained in the following, the choice of the appropriate model for the backscatter temporal trend can be crucial for the performance of the subsequent detection steps, which are based on the above-mentioned assumption that flood events cause short-lived “spikes” in the backscatter time series, with respect to a smooth trend. If such a smooth trend is not properly accounted for, residual variations, not due to flood events, may leak into the fit residuals, and thus cause false alarms. In some cases, unprecise fits may even over/undershoot and reduce the departure of outliers from the mean of the residuals, thus causing false negatives. In view of all this, the possibility of using different types of temporal models should be considered.

This article describes results of the implementation of a flood detection methodology from SAR data stacks, within a semi-automated system based on a Bf. Probabilities (i.e., confidences) of the presence of floodwaters are estimated pixel-wise at each acquisition time through the Bayes equation, using ancillary data such as terrain heights to assign prior probabilities, and estimating likelihoods from the data. In particular, likelihoods for the SAR backscatter distributions conditioned by the presence of floodwaters are estimated over permanent water bodies, while likelihoods for unflooded conditions are derived through the modeling and fit of SAR backscatter time series with smooth functions, aided by Gaussian processes (GPs).

GPs allow us to nicely fit a wide class of temporal trends with nonparametric functions, by specifying a kernel, i.e., an autocorrelation function in time. Several kernel types or classes can be postulated (e.g., exponential, Gaussian, etc.), with “hyperparameters” which can be optimized iteratively (“learned”) from the data. Although the setup of such hyperparameters requires some care, we find the framework is quite robust and provides good results on our data through use of a single class of kernels.

Results are obtained over an area of the Basilicata region that includes the so-called Metaponto Coastal Plain (MCP, southern Italy). The area is recurrently subjected to floods, so we focus on some recent relatively large events which are known to have caused problems to the infrastructures, agriculture, or both. Although no independent data are available for the assessment, we analyze some examples of local geomorphological features connected to the presence of floodwaters on the ground, which provide evidence supporting an improved performance obtained through the use of the GP-based temporal modelling approach in the likelihood estimation for unflooded conditions, with respect to the one which uses harmonic functions.

We then further illustrate the potential of our flood monitoring methodology, by extending the analysis to detect and characterize other minor events, which are not reported in news or other records, but which are nevertheless known to have had some impact on the landscape; we argue that the possibility of performing such detailed analyses constitutes a
considerable added value for high-resolution flood maps such as those produced by our system, which could thus be of aid in the management of areas subject to flood risks on a small, field-sized scale.

2. Materials and Methods

2.1. Study Area

The selected study area, shown by the green rectangle in Figure 1, includes most of the MCP (red polygon) [35], one of the larger coastal plains in Italy, on average 4 km wide and 35 km long, facing the Ionian Sea. The largest portion of the study area falls in the Basilicata region (small portions of Puglia and Calabria are also included), covering about 2100 km².

![Figure 1.](image)

The plain, oriented NE–SW, has an elevation of about 30 m above sea level (a.s.l.) in the southern part and 18 m a.s.l. in the northern part [36]. Due to the precipitation dynamics, characterized by a mean annual rainfall of about 100 mm, the climate of this area can be classified as semi-arid [37,38]. In particular, the MCP is marked by the presence of sandy and silty meandering fluvial systems [39], with some artificial and often channelized reaches, forming, at the mouth, wave-dominated deltas (deflected deltas, sensu [40]) in a microtidal regime [41]. Five main rivers, Bradano, Basento, Cavone, Agri, and Sinni, reach the Gulf of Taranto with directions roughly perpendicular to the coastline. These are Mediterranean-type streams, characterized by dry summers and winter floods. Fluvial terraces occur along rivers beds, distributed in different orders at different elevations [42].

The area of the MCP has been recognized as the most prone to flood hazards in the Basilicata region, with about 68,000 inhabitants exposed [43]. In the last two decades, indeed, because of climate and land use change [44], intrinsic lithological [45] and geomorphological properties [46], and human activities [36], it has been affected by a recurrence of flooding events, not exceptionally intense on an absolute scale, which caused significant economic damage to infrastructures, agricultural and tourist activities, and archaeological heritage (Figure S1). Of great significance, indeed, is the presence within the area of some
geological and archaeological sites of interest, such as the Special Nature Reserve of the Montalbano Jonico Badlands [47] and the Metaponto Archaeological Park, located north of the modern residential area of Metaponto, which represents the heart of the ancient city [48]. The land cover in the inland part of the area is mainly agricultural, characterized by cereal and vegetable crops and fruit plantations, while pine forest stands occupy large areas near river mouths and along the coast [45].

Qualitative and quantitative analyses of rainfall dynamics conducted in [36] revealed substantial changes in mean and extreme values over the last 60 years, with a very interesting change in the rainfall metric regime during the last climatic normal 1990–2019. These results confirm previous work [49], which showed a significant upward trend in the pluviometric regime of the Basilicata Region since 2000. Furthermore, hydrometric and erosion/deposition analyses of these main rivers show that floodings, particularly in the MCP, are caused by extreme rainfall events and anthropic activities, such as large-scale hydraulic reclamations and the lack of cleaning of artificial channels [36].

2.2. Major Flood Events from 2014

Table 1 summarizes the precise dates of occurrences, included river basins, and related descriptions of the major flood events that have affected the area from October 2014 to October 2020 (acquisition period of the images used in this study).

Table 1. Summary table of major flood events in the studied area, from October 2014 to October 2020. For every event, the specific dates, affected river basins, a short description, and information of cumulative rainfall, for the entire event, are shown.

<table>
<thead>
<tr>
<th>Year</th>
<th>Dates</th>
<th>River Basins</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25–28 March</td>
<td>Basento</td>
<td>Basento floods in the countryside of Pisticci, in several areas. The Tinchi-Bernalda provincial road, in Contrada Accio, is closed. Cumulative rainfall measured in Pisticci Scalo: 14.60 mm.</td>
</tr>
<tr>
<td>2018</td>
<td>18–23 October</td>
<td>Basento</td>
<td>The Pisticci area is particularly affected: flooding, road closures, and disruption of rail traffic. A total of 100 hectares of land are flooded in Scanzano Jonico as water overflowed from drainage canals clogged with reeds and weeds. Cumulative rainfall measured in Pisticci Scalo: 66.40 mm.</td>
</tr>
<tr>
<td>2020</td>
<td>26–27 March</td>
<td>Sinni, Agri, Basento</td>
<td>Flooding of the Sinni River in the countryside of Tursi—San Donato: 70.6 mm.</td>
</tr>
</tbody>
</table>

The table has been filled by referring to different sources, from the scientific literature [36,50,51] to the web; the service ‘EvAlMet—Eventi Alluvionali e precipitazioni meteoriche eccezionali del Metapontino’ (Alluvial Events and Exceptional Meteoric Precipitation in the Metapontino) consists of a WebGIS that provides the possibility of querying a database containing information of all the flood events that have occurred in the Metaponto area from 1959 to 2015 [52]. For each event, the triggering causes, the induced effects, the damage recorded in each affected municipality, and the record of the rainfall stations are presented. In addition, hydrological reports provided by the ‘Centro Funzionale Decentrato’ of the Basilicata Civil Defense [53] and several local news sites were consulted [54–57].

2.3. Sentinel-1 Data

Sentinel-1, which continues the C-band SAR Earth Observation heritage of ESA’s ERS-1, ERS-2, ENVISAT, and Canada’s RADARSAT-1 and RADARSAT-2 [58], is the first
mission under the EU Copernicus Program. Initially, it was a constellation of two identical satellites, Sentinel-1A (S-1A) and Sentinel-1B (S-1B), with near-polar/sun-synchronous orbit, launched from Kourou on a Soyuz rocket on 3 April 2014 and 25 April 2016, respectively. As a constellation of two satellites orbiting 180° apart, the mission covered the whole globe every 6 days (ascending/descending and overlap) from April 2016 to December 2021, when Sentinel-1B experienced an anomaly related to the instrument electronics power supply provided by the satellite platform, leaving it unable to deliver radar data. Before and after that period, S-1A acquired data and still continues to do so with a temporal resolution of 12 days.

In this study, the input of the implemented algorithm consists of a Sentinel-1 image stack. Specifically, the stack includes 302 images, acquired in ascending geometry, in Interferometric Wide Swath Mode (IW), in VV polarization, from 15 October 2014 to 19 October 2020, with an almost constant temporal resolution of 6 days. Images were cropped according to the study area and appropriate pre-processing was carried out, including calibration, thermal noise removal, and geocoding, using the ESA SNAP software tool.

In order to reduce the noise associated with the speckle effect [59], the non-local BM3D filter was applied to each SAR image [60], using the MuLoG (MUlti-channel LOgarithm with Gaussian denoising) general scheme [61,62]. By constructing the stack, amplitude information on single pixels can be studied as data time series.

2.4. Bayesian Framework

By means of Bayesian inference, we derive the posterior probabilities for a given pixel and a given acquisition date to belong to the flood or the non-flood class. These can be interpreted as a probabilistic confidence measure. The procedure follows steps already described in recent literature [23,63]. With $\sigma_0^F(t), t = 1, \ldots, N$, we denote the backscattering coefficients at pixel $P$ in the time series of the $N$ calibrated and coregistered SAR images. The presence of floodwater ($F$) at $P$ at a certain time $t$ is modelled as a conditional probability, $p(F|\sigma_0)$. Omitting the $P$ and $t$ indices for simplicity, the posterior probability $p(F|\sigma_0)$ is expressed through the Bayes equation as a function of prior absolute and conditional probabilities [22]:

$$p(F|\sigma_0) = \frac{p(\sigma_0|F)p(F)}{p(\sigma_0|F)p(F) + p(\sigma_0|\overline{F})p(\overline{F})},$$

where $p(\sigma_0|F)$ and $p(\sigma_0|\overline{F})$ are the so-called likelihoods, i.e., the probabilities for the observed backscatter conditioned by the presence or, respectively, the absence of floodwater on the surface. $p(F)$ and $p(\overline{F}) = 1 − p(F)$ indicate the prior probability of, respectively, flood and no flood.

In order to limit effects due to slopes and to the side-looking geometry acquisition of the SAR sensor, which could result in very low backscatter on shadowed areas, the priors $p(F)$ and $p(\overline{F})$ are modelled by using ancillary external information layers, which assign low or zero probability of being observed as flooded to certain areas on the basis of geomorphological information. For this study, we used the Height Above the Nearest Drainage (HAND) index [23], a mask for the layover effect and the slope information (Figure 2).

The HAND index, used for landscape characterization and hydrological analysis, normalizes topography according to the local relative heights found along the drainage network [64]. By excluding all pixels featuring a HAND index above a certain value, the impact of topography-related misclassification can be reduced significantly. Similar considerations apply to pixels falling on terrain with large slopes, as well as in radar shadow or layover areas. For this study, the 10 m Tinitaly Digital Elevation Model (DEM) by the Istituto Nazionale di Geofisica e Vulcanologia (INGV) [65] has been processed to obtain the HAND heights, the slopes, and the shadow/layover mask. Indicating HAND heights with $HD$, slope with $S$, and the layover/shadow binary mask with $G$, we set
where periodic trends. Typically, trends fluctuate with varying amplitudes and/or periods, enabling understanding and analysis of cyclical phenomena in the data.

The $p(F^0|F)$ and $p(F^0|\overline{F})$ likelihoods are modelled as Gaussian distributions. Their parameters are estimated from the data. The likelihood for flooded conditions, $p(F^0|F)$, is estimated once for all pixels over permanent water areas. In our case, it was estimated on an image patch covering the internal area of the San Giuliano dam, in the Matera province, taking into account only pixels constantly covered by water. Sloped shorelines have been excluded, as they could be submerged only in some periods of time.

For what concerns the likelihood of no-flood conditions, we estimate it on each pixel, relying on the assumption that the extreme variability of backscattering conditions on the ground for land areas cannot be treated in a spatially uniform manner. We therefore postulate that the backscatter time series on each pixel can be approximated by a smooth function of time, plus a white, Gaussian-distributed noise signal: $\sigma^0_F(t) = \overline{\sigma^0_F}(t) + \epsilon$. By subtracting from each trend a good approximation of the smooth temporal model, the parameters of the Gaussian distribution of the residual $\epsilon$ can be easily derived, and used in place of the original backscatter observable, to derive a likelihood of the kind $p(\epsilon|\overline{F})$. Such residuals should exhibit consistently smaller dispersions than the original temporal trends. Since we expect, as mentioned earlier, that flood events appear as (negative) short-lived anomalies, or “spikes” in the backscatter time series of normally non-flooded areas, they will be more evident as outliers in the $\epsilon$ residual time series. Consequently, these outliers in the residual distributions, corresponding to flood events, will have more chances of being assigned low no-flood likelihoods, $p(\epsilon|\overline{F})$, and a correspondingly higher confidence for the presence of flood in the final posterior probability $p(F^0|\sigma^0)$, as obtained from (1).

In most of the related work existing in the literature, the smooth backscatter variations in time over land areas are modelled as parametric functions, such as low-degree polynomials or sums of harmonic terms. For instance, as proposed by [29], a possible backscatter time series model is:

$$\overline{\sigma^0_F}(t) = \overline{c} + \sum_{i=1}^{k} \left[ c_i \cos \left( \frac{2\pi i t}{n} \right) + s_i \sin \left( \frac{2\pi i t}{n} \right) \right],$$

where $\overline{\sigma^0_F}$ is the average of $\sigma^0_F$ in time, $n$ the number of time units per year (e.g., 365.2425 if $t$ is in days), and $k$ the number of harmonic terms. In most cases (see, e.g., [26] and the discussion therein), $k = 3$. The $c_i$ and $s_i$ parameters can be estimated using, e.g., least-squares regression. This type of modelling allows for the identification and quantification of periodic patterns, enabling understanding and analysis of cyclical phenomena in the data.

However, many radar backscatter temporal series do not seem to follow strictly periodic trends. Typically, trends fluctuate with varying amplitudes and/or periods, according to various long-term drifts, as well as other aperiodic phenomena due to, e.g., climate change or land use/land cover variations. Although this is not considered to be a hindrance in applications aimed at, e.g., land classification, where the important
information to be gained from the temporal analysis is in the general trend (goodness of fit) and in the parameters of the fitted functions, in the case at hand, the time series analysis is aimed basically at removing any smooth variation of the analyzed quantity, so as to leave a residual characterized by a Gaussian statistics as much as possible, with “abnormal” values (registered during flood events) sticking out of the residue series as outliers; in this respect, any unfitted residual smooth fluctuation with respect to a given model which locally does not follow adequately the temporal trend could lead to false alarms for the identification of floodwater, or, on the contrary, even to missed identifications in cases of over/undershooting of the fitted trends. For some examples of these types of behavior, the reader can look at the sample backscatter time series depicted as blue circles in Figure 3f–h, which are then properly described and analyzed in the related discussion in the following sections.

**Figure 3.** (a) Probability flood maps obtained with the two types of modelling over a portion of the Basento river basin, near Pisticci Scalo, for the event of 14 March 2016, indicated in (b) within the complete study area. Areas having \( p(F|\sigma) \geq 0.5 \) in the maps obtained by modelling the backscatter time series through harmonic functions (Equation (2)), through GPs, and in both cases are shown in blue, red, and purple colors, respectively; (c–e) enlarged areas highlighting examples of detected flood pixels only in the harmonic fit case (e), in both cases (d), and only in the GP modelling case (e), respectively. In each case, a particular pixel is highlighted with a green star, and the corresponding backscatter time series in the 3 cases are shown in panels (f–h), respectively. The orange vertical line shows, in the plots, the results corresponding to the mentioned event of March 2016. (i) indicates the area shown in Figure 4. Base map images: Google Satellite. Further examples of occurrence of over- and underdetection are indicated in panel a by cyan circles and pink squares, respectively.
A Matern 3/2 function (3) was used as the kernel for our modelling [66]:

$$k_{v=3/2}(r) = \sigma \left( 1 + \frac{\sqrt{3}r}{l} \right) \exp \left( -\frac{\sqrt{3}r}{l} \right),$$

(3)

where $r = |t_i - t_j|$ is here the time difference (in days) between any two acquisitions, $l$ is a characteristic length-scale in time, and $\sigma$ is the signal (i.e., the SAR backscatter) standard deviation. In the fit process, the $l$ and $\sigma$ parameters are estimated from the data, starting from initial values corresponding to the standard deviation of the input time intervals of each image with respect to the starting acquisition date (in days), and the standard deviation of the backscatter data (in dB), respectively. Through repeated trial and error, we found that this kernel class gives good results in most of the cases explored, so that it can be applied “blindly” to all the backscatter time series for the image pixels corresponding to land areas not permanently covered by water and not excluded by the previously described masks. Of course, more sophisticated procedures could be adopted, such as the use of land cover-specific kernels or initial parameters. This is left to future studies.

3. Results

Our final product is a stack of maps, one for each acquisition date, containing the probability, or confidence, of the presence of floodwater, conditioned on the SAR and the
ancillary data through Equation (1). By analyzing the stack, it is possible to obtain several pieces of information; the most straightforward is the identification of the pixels likely affected by flooding in each map, which can be detected by applying thresholds to the probabilities (e.g., 0.5, or higher values if a higher confidence is required). In the following, we propose some evaluations of this kind of product obtained on our test data.

3.1. Flood Map Analysis and Methodology Comparison for the Events in Table 1

In order to evaluate the product maps, we show in Figure S2 (in the Supplementary Material to avoid cluttering the main text) the flood maps (Figure S2a,c,e,g), obtained by applying a threshold of 0.5, on the dates closest to the events shown in Table 1. Panels on the right of the figure display zoomed details of the areas mainly affected by each event, with indication of toponyms for easier identification; in particular, the areas near the towns of Pisticci (Figure S2b,d,f), Scanzano Jonico (Figure S2f), and Tursi (Figure S2h) are shown. Although no independent data are available for any of these events to quantitatively assess the results, expert geomorphological analysis, together with high-resolution optical imagery of the affected areas from, e.g., Google Earth, allows us to state the good quality of the maps. In particular, the map temporally closest to the maximum of the 2016 event, acquired on 14 March 2016 (Figure S2a,b), can be seen to affect a large area around the Basento river, in accordance with the reports summarized in Table 1, particularly in the countryside of Pisticci (Pisticci Scalo and Ponte Accio localities [68]). The March 2018 event (Figure S2c,d) also affected the Basento basin, in particular, as mentioned, in the countryside of Pisticci, although with a smaller impact on the area, similarly to the October 2018 event (Figure S2e,f). This event also affected the area surrounding the town of Scanzano Jonico, between the Agri and Sinni rivers. For the March 2020 event (Figure S2g,h), the maps show the results obtained in particular for the area near the town of Tursi, close to the Sinni river.

To illustrate the advantages of the proposed innovative approach to model each time series through Gaussian processes (GPs), we compare the obtained maps with those obtained from the same data, using the harmonic model in (2) to estimate the backscatter likelihoods in no-flood conditions (Figure 3). Most of the differences between the two results are due to the different performance of the temporal fits, which, in the case of GP regression, allow us to reduce the occurrence of “spurious” oscillations in the residuals of the time series and thus render such residuals more “regular” and Gaussian-distributed, thus improving the detection of flood events as low no-flood probability outliers in the residual time series, leading to higher confidence in the posterior flood probabilistic detection (Equation (1)). Some examples of local conditions where the improvements are particularly clear are displayed as zoom-in maps in Figure 3c–e. For each example, plots of time series are shown for a particular pixel (Figure 3f–h), with their fitted models $\hat{c}_p^2$, and the residuals $\epsilon$. We do not show the details of the derived statistical distributions involved in the derivation of the final posterior probabilities, as these show negligible differences in the two cases. In fact, in our case, the most important difference in the derivation of the final probabilities lies in the different way in which the time series of backscatter data over land areas are modelled and how well the residual time series approximate a white, Gaussian-distributed noise sequence, with events sticking out as outliers. We decided to show the results obtained for the date of 14 March 2016 (indicated also in the plots with an orange vertical line) in the maps, because the hydro-meteorological event and its dynamics are well-known in the literature.

The area in blue in Figure 3c shows a field which is not likely to have been flooded, since it appears to be slightly higher in elevation than the adjacent river course area; here, the harmonic model methodology gives a false alarm, caused by a residual time series which has a consistent trough in the period spanning the event. The harmonic fit does not accommodate sufficiently well the natural variation in backscatter (Figure 3f), likely due to, e.g., land cover variations due to crop cycles, or changes in soil moisture. From Figure 4, indeed, and in particular in the area highlighted by the rectangle labelled as ‘c’, it can be seen that the dB values are not as low as those of the surrounding flooded areas.
Nevertheless, the GP model better fits this natural variation, leaving smaller residuals with respect to the harmonic model. Consequently, the pixel (as well as the other pixels in the area) are correctly detected as unflooded in the GP fit version.

Figure 3d shows an area where both methods agree in giving a flood indication. This appears justified by its proximity to the main watercourse (Basento river) and the topography of the area. Pixels are clearly dark in the corresponding SAR image, shown in Figure 4 (the magenta rectangle labelled ‘d’ delimits the same area as in Figure 3d) and the corresponding backscatter residue time series in Figure 3g show significant outliers in both cases.

In the case highlighted in the rectangle ‘e’ in Figure 4, a portion of a field “closing” a C-shaped area is reported as flooded in the GP fit-derived probability map and not in the harmonic fit-derived one; in the sample backscatter time series in Figure 3h, residues of the GP fit (red stars) are slightly below those of the harmonic fit (blue diamonds) in the date of the flood (corresponding to the yellow vertical line). This causes the pixel to have a higher probability of being flooded in the GP fit case than in the harmonic fit case. The ground evidence supports the notion that the red-colored area is actually a low-lying part of the field, so it could have been flooded at the time of the event.

Other instances of over- and underdetection in the case of the harmonic fit, which are instead correctly treated in the GP-fit proposed methodology, have been identified in the maps, which we do not report here for brevity. Examples of such instances are indicated in Figure 3a as cyan circles and pink squares, respectively, for over- and underdetection.

Returning to the case of Figure 3e, as can be seen from a zoom of the SAR image that corresponds to the yellow box in Figure 3a for the same date of acquisition, the area under examination (labeled ‘c’ in Figure 4) is characterized by pixels with backscatter values not as low as those of other nearby flooded regions. Therefore, to better support the previous claims, further evidence was sought from both Google Earth imagery and ground data.

The detail panel in Figure 5a shows in the background a Google Earth satellite image from 15 April 2016; effects on the ground, consisting of ‘scars’ in the vegetation cover, most likely due to water passage and accumulation, are indicated by black arrows. Often, in fact, floods contribute to ‘flatten’ some vegetation/crop types. In addition, such water stagnation promotes deposition of suspended sediment. The geometry of some ground features can be correlated with our results, suggesting the presence of floodwater on the acquisition date. In order to evaluate the detailed topography of the area, a photogrammetric survey was carried out, producing a high-resolution (4 cm/px) DEM (Figure 5b), using a DJI Inspire 2 unmanned aerial vehicle (UAV) [69], equipped with a DJI Zenmuse X5S sensor [70]. The survey was performed at a flight altitude of 75 m above ground level at take-off location, using a nadir-looking camera, a single-grid path flight mission using an overlap, and a sidelap of the collected images of 80% and 70%, respectively. The images were acquired as RGB with a Ground Sampling Distance (GSD) of 1.6 cm/px, setting the ISO sensitivity value to 100, the f-stop to f/9, and the exposure time and the focal length of the camera to 1/800 s and 15 mm, respectively. The acquired dataset was processed according to the Structure from Motion (SfM) principle, as described in [71], starting from the computation of descriptors and tie points between overlapping images, up to the bundle adjustment and the generation of depth maps, needed for DEM construction.

The topographic profile (Figure 5d), the trace of which is shown in Figure 5c, indicates a slightly concave geometry on the center-left part, which suggests the possible accumulation of water. The lack of detected floodwater in the central part of the profile, on the other hand, is probably related to both the slightly higher elevation and the presence of thicker vegetation, typical of this type of land cover. However, we could not find clear indication of its actual characteristics in 2016. It could, for instance, be hypothesized that the height of the vegetation itself influenced the backscatter, possibly producing a double-bounce effect. We also note that the small height differences involved in this example are not sufficient to exceed height or slope thresholds as set in the prior probabilities, nor to cause radar shadow or layover.
3.2. Flood Map Evaluation

In order to provide reliable evidence of the capabilities of the proposed methodologies, as well as to highlight the potential of the produced flood maps, we performed an analysis of our probabilistic flood map stacks in both the spatial and the temporal domain, by calculating the extension of surfaces detected as flooded over fixed sub-areas throughout the temporal axis. This allowed us to assess the method’s ability to detect the known events reported in Table 1, as well as its potential to reveal other events, characterized by minor extents and/or not reported in the existing literature or in the news. Note that we refrain from comparing flooded areas in the two time series fit approaches, as these are cumulative figures which mix the effects of over- and underdetections. In view of the mentioned absence of independent validation data, differences in such figures would be difficult to evaluate objectively.

First, we selected three subset regions (Figure 6), of equal area. Subset n°1 covers part of the Basento River, near Pisticci Scalo; subset n°2 is an area near Montalbano Jonico, which includes part of the Agri River; subset n°3 delimits an area near Policoro, including part of the Sinni river. These areas are usually most affected by flood events.

Figure 5. (a) High-resolution satellite image, from Google Earth, acquired on 15 April 2016, one month after the event of 11–18 March 2016, reported in the first row of Table 1. Black arrows indicate evidence in the field, which can be traced back to the accumulation of water; (b) DEM (4 cm/px) obtained by UAV photogrammetric survey; (c) results compared with the DEM of the area (perimeter in light blue), and trace, in green, of the topographic profile shown in (d). On the profile, flood information is indicated, both for harmonic (in red) and GP (blue) modelling results. Base map: Google Satellite.
Donato pixels, related to the dates for which the numbered peaks were identified, and the estimated pixels with probability of being inundated greater than 90% (i.e., threshold of 0.9). This higher threshold was chosen to reduce the uncertainty in the total area values, allowing us to highlight only the largest events. By plotting the total number of such pixels against the dates of acquisition, it is possible to isolate some particular events, which can be identified as peaks in the sequence. We then considered each of the highest peaks, indicated in the three plots in Table 2, showing the total number of flooded areas in km² as peaks in the sequence. We then considered each of the highest peaks, indicated in Figure 6, it is possible to visualize the obtained flood probability maps, related to the dates of the mentioned peaks, for each chosen subset. We summarize the information contained in the three plots in Table 2, showing the total number of flooded areas in km² (considering that each pixel has the size of 100 m²).

For each of the three subsets, we calculated, in each obtained map, the total number of pixels with probability of being inundated greater than 90% (i.e., threshold of 0.9). This higher threshold was chosen to reduce the uncertainty in the total area values, allowing us to highlight only the largest events. By plotting the total number of such pixels against the dates of acquisition, it is possible to isolate some particular events, which can be identified as peaks in the sequence. We then considered each of the highest peaks, indicated in Figure 7 with orange dots, inspecting the corresponding flood probability maps, also in conjunction with daily rainfall data, in order to gain some insight on the corresponding events. Daily rainfall data were downloaded from the ALSIA website [72] from three meteorological stations: 'Pisticci—Pisticci Scalo' (Cod. PI3) for the first subset, 'Tursi—San Donato' (Cod. MSD) for the second, and 'Policoro—Az. Pantanelli UNIBA' (Cod. PO3) for the third one.

A selected number of peaks were identified for each area, numbered and related to specific dates. In Figure S3, it is possible to visualize the obtained flood probability maps, related to the dates of the mentioned peaks, for each chosen subset. We summarize the information contained in the three plots in Table 2, showing the total number of flooded pixels, related to the dates for which the numbered peaks were identified, and the estimated flooded area in km².
Figure 7. Comparison between rainfall data (y-axis on the left), expressed in mm/day, and number of pixels with probability of being inundated greater than 90% (y-axis on the right), for each acquisition date (x-axis), for the three chosen subset regions in Figure 6. The black dotted lines correspond to the dates of peaks, for each subset. For each date and subset, the estimated flooded area is indicated in km². Numbers and dates in orange identify detected flood events.

Table 2. Number of pixels (orange dots in Figure 7) with probability of being inundated > 90%, related to dates of peaks, for each subset. For each date and subset, the estimated flooded area is indicated in km². Bold numbers indicate events that could be considered as minor floodings, due to correlation with rainfall data. The March 2016 event, which shows an evident peak, is highlighted in red. The event, known in the literature, is also present and described in Table 1.

<table>
<thead>
<tr>
<th>Dates</th>
<th>Subset n°1</th>
<th>Subset n°2</th>
<th>Subset n°3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N° Pixels</td>
<td>Area (km²)</td>
<td>N° Pixels</td>
</tr>
<tr>
<td>08/11/2014</td>
<td>416</td>
<td>0.042</td>
<td>1857</td>
</tr>
<tr>
<td>31/01/2015</td>
<td>12,971</td>
<td>1.297</td>
<td>389</td>
</tr>
<tr>
<td>11/08/2015</td>
<td>11</td>
<td>0.001</td>
<td>48</td>
</tr>
<tr>
<td>14/03/2016</td>
<td>23,991</td>
<td>2.399</td>
<td>2833</td>
</tr>
<tr>
<td>26/03/2016</td>
<td>1271</td>
<td>0.127</td>
<td>2152</td>
</tr>
<tr>
<td>22/03/2018</td>
<td>7300</td>
<td>0.730</td>
<td>8397</td>
</tr>
<tr>
<td>03/02/2019</td>
<td>2138</td>
<td>0.214</td>
<td>30</td>
</tr>
<tr>
<td>04/05/2019</td>
<td>530</td>
<td>0.053</td>
<td>458</td>
</tr>
<tr>
<td>06/11/2019</td>
<td>32</td>
<td>0.003</td>
<td>1387</td>
</tr>
<tr>
<td>24/11/2019</td>
<td>3081</td>
<td>0.308</td>
<td>6828</td>
</tr>
<tr>
<td>22/04/2020</td>
<td>837</td>
<td>0.084</td>
<td>1357</td>
</tr>
<tr>
<td>07/10/2020</td>
<td>182</td>
<td>0.018</td>
<td>646</td>
</tr>
</tbody>
</table>

4. Discussion

Regarding the flood map extent results on our test site, we chose to display the six events that, according to Table 2, have the highest peaks in the plots in Figure 7: 31 January 2015 (a), 11 August 2015 (b), 14 March 2016 (c), 22 March 2018 (d), 24 November 2019 (e), and 22 April 2020 (f). From Figures 7 and S3 and Table 2, it is possible to state the following considerations:

- Not all the events in Table 1 are clearly identifiable in the plots; in fact, we can only spot 1: the March 2016 event (in red). This may be due to the fact that the events have limited extensions and also because the acquisitions do not correspond to the peak phases of the events, which is the case for the 14 March 2016 scene. As can also be seen from Figure S2, the maps for the other three dates do not show any relevant flood...
extents, for the whole area (satellite acquisitions 1 or 2 days after the peak of the event). On the other hand, the methodology makes it possible to identify minor events not easily found in the literature.

- Of all the events in Table 1, only for the one on 14 March 2016 is there a high peak related to the flood extent matching with a rainfall peak, and this occurs only on the plot for subset region n°1, located near the Pisticci Scalo area. A slightly lower peak is visible on the plot for subarea n. 2, while for the third region, no orange peak is visible in correspondence to peaks in the rainfall data.

- The only two dates where we find flood extent peaks on all three subareas are 22 March 2018 and 24 November 2019. For what concerns the former, in the morning of 22 March 2018, the entire Basilicata region was affected by a severe cold storm, resulting in widespread snowfall [73,74]. Since the local acquisition time of Sentinel 1 on that date is 16:47, when the snowfall event was finished and practically all the snow had melted, the peaks visible on that date on all three subareas (numbered 3, 4, and 2, respectively, in the three plots) could be due to the extensive water pools left after snow melting. Regarding the date of 24 November 2019, we have clear agreement with rainfall peaks in all three subsets.

- Considerable variation can be seen in the detected flood extents which result after each rainfall event, either within each subarea or among different subareas for the same events. In particular, not all the strongest rainfall events correspond to detected peaks of flood extents, and vice versa. This heterogeneity arises in part from the fact that the subareas are located on different river basins (Basento, Agri, Sinni), which respond differently to rainfall, both in time and in space [75]. For instance, the part of the Basento river falling within subset n°1 crosses a hilly area, with slopes oriented towards the MCP. The second subset is located in the third section of the Agri river, where the slopes gradually decrease and the alluvial plain of the watercourse widens considerably. The third subset, on the other hand, is in the final part of the Sinni river, in the MCP; although a complete geomorphological and/or hydrological analysis of such features is out of the scope of the present work, it is clear that the kind of information made available by the analysis of the produced flood probability map stacks is of unprecedented resolution, so that it could constitute a new and precious tool of analysis of the landscape responses to meteorological events, a crucial task in, e.g., climate change studies.

- According to plots and Table 2, the most relevant flood events occur in the months of November and March. This is in agreement with the analysis conducted by Bentivenga et al. [36], on the seasonality of flood events, up to 2019, in the same survey area.

We also observe that these minor events, as confirmed by direct oral testimonies of local citizens met during field surveys, have produced, over time, damages to small and large infrastructures, as well as to economic activities mainly of an agricultural type.

We finally remark that our proposed methodology, whose schematic flow chart is shown in the paper graphical abstract, is general and applicable to different geological and climatic contexts. The slight additional computational cost of the use of Gaussian process temporal fits, with respect to more straightforward parametric models, is largely compensated by the shown improved performances in floodwater detection, which appear convincingly supported by geomorphic analyses, in spite of the lack of independent ground truth. The methodology is, furthermore, highly scalable, as each pixel or small portion of land is treated in a virtually independent way. This allows for the possibility of applying parallel computing solutions, thus drastically decreasing processing times, as well as exploring efficient computing solutions to update the temporal models [76], towards the goal of obtaining near real-time results.

5. Conclusions

A methodology is proposed to derive flood probability maps from multitemporal SAR data. We use a Bayesian framework to calculate conditional posterior probabilities
of observing floodwaters from prior probability distributions inferred from ancillary geometric, topographic, and geomorphic information, and likelihoods estimated from the SAR data. The result, a stack of flood posterior probability maps, conditioned on the data, can be interpreted in terms of confidence of the observation. The approach is based on the detection of statistical anomalies with respect to slow-changing time trends of backscatter intensities, modeled through Gaussian processes. By analyzing the final probability maps, e.g., through thresholding, it is possible to visualize the areas where a river flood event is most likely to have occurred, at full resolution both in time and space.

An area in the Basilicata region, which includes the so-called Metaponto Coastal Plain (southern Italy), recurrently subject to flooding, was chosen as a test, and investigated through the application of our methodology to Sentinel-1 images acquired between October 2014 and October 2020. The results show good performance for known events throughout the time frame. The use of the above-mentioned non-parametric models for the temporal fit of slow-varying trends shows evidence of improved performance with respect to maps obtained from the same data using parametric models consisting of sums of harmonic terms. In fact, as is most often the case with flood monitoring activities, independent data are not available for a quantitative assessment. Nevertheless, evidence for the evaluation is here based on expert interpretation of geomorphic signatures of floodwaters left on the ground, supported, in one case, by a subsequent Lidar ground campaign.

We have, in addition, evaluated the results of our proposed methodology in terms of time series of flooded area extents over selected portions of the territory close to meteorological stations, near three different water courses (Basento, Agri, Sinni). This allowed us to identify the effects of other minor meteorological events, unreported in the news but supported by the analysis of precipitations, as peaks in the flood extent distributions.

The proposed methodology is suitable for further improvements and upgrades. One possibility is, e.g., to include the comparison of the results with anemometric data, in order to evaluate the backscatter response on water surfaces affected by windy phenomena. Another possible improvement is to include information coming from land use/land cover, both in the time series modelling and in the validation phase. These are all promising directions for future work.

Such products could be useful to different stakeholders such as public entities, insurance agencies, and private companies, for all phases of risk management, starting from proper territorial planning, all the way to post-emergency phases. These maps, indeed, turn out to be important tools to improve the flooding hazard assessments in riverine environments and for a proper management of water resources.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16020294/s1, Figure S1: Flooding in Basilicata region; Figure S2: Flood maps of known events; Figure S3: Flood maps of detected events in Table 2 [77,78].


Funding: This work was supported by the Italian Ministry of Education, University and Research, under POC Research and Innovation 2014–2020 for ‘Innovative doctorates with industrial characterisation’ (R.C. PhD course), and D.D. 2261 (6.9.2018), Programma Operativo Nazionale Ricerca e Innovazione (PON R&I) 2014–2020, Project OT4CLIMA. Furthermore, this study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan—NRRP, Mission 4, Component 2, Investment 1.3—D.D. 1243 2/8/2022, PE0000005).

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: Research performed in the framework of project GEORES—Applicativo GEOspaziale a supporto del miglioramento della sostenibilità ambientale e RESilienza ai cambiamenti
climatrici nelle aree urbane", funded by the Italian Space Agency (ASI), Agreement n. 2023-42-HH.1.0. R.C. thanks Francesco Zucca (Department of Earth and Environmental Sciences, University of Pavia, 27100 Pavia, Italy) and Monica Bini (Department of Earth Sciences, University of Pisa, 56126 Pisa, Italy) for their invaluable help.

Conflicts of Interest: The authors declare no conflicts of interest.

References


15. Han, S.; Coulibaly, P. Bayesian Flood Forecasting Methods: A Review. J. Hydrol. 2017, 551, 340–351. [CrossRef]


36. Bentivenga, M.; Giano, S.I.; Piccarreta, M. Recent Increase of Flood Frequency in the Ionian Belt of Basilicata Region, Southern Italy: Human or Climatic Changes? Water 2020, 12, 2062. [CrossRef]


42. Gioia, D.; Bavusi, M.; Di Leo, P.; Giammatteo, T.; Schiattarella, M. Geoarchaeology and Geomorphology of the Metaponto Area, Ionian Coastal Belt, Italy. J. Maps 2020, 16, 117–125. [CrossRef]

43. Lapietra, I.; Rizzo, A.; Colacisco, R.; Dellino, P.; Capolongo, D. Evaluation of Social Vulnerability to Flood Hazard in Basilicata Region (Southern Italy). Water 2023, 15, 1175. [CrossRef]


45. Boenzi, F.; Caldara, M.; Capolongo, D.; Dellino, P.; Piccarreta, M.; Simone, O. Late Pleistocene–Holocene Landscape Evolution in Fossa Bradanica, Basilicata (Southern Italy). Geomorphology 2008, 102, 297–306. [CrossRef]


55. La Gazzetta del Mezzogiorno. Available online: https://lagazzettademezzogiorno.it/ (accessed on 8 April 2023).


70. DJI Zenmuse X5. Available online: https://www.dji.com/it/zenmuse-x5s/info#specs (accessed on 1 September 2023).


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.