Reply to Reviewer Comments

Comment:

First of all, sorry I am a bit late with this and I hope my comments are still useful. I think they have also been highlighted in part and in other words by the other reviewer and the authors already posted some comments. The case study presented is interesting but what is lacking is some detail on the actual data processed, some kind of validation (although difficult here) but then some cross validation at least with other available maps processed either from the same Sentinel-1 by other organizations or maps processed from other EO imagery. It would also be useful to put the processing and used processing chain in context with other fully automated methods used - there are now many of those, for example HASARD by LIST or the chain used at DLR ZKI or indeed maps from the Dartmouth Flood Observatory. Such maps can then also be used to cross-validate and get a good idea about the sensitivity/uncertainty in the presented flood maps. I am not sure about the nature and scope of brief communications in NHESS but at this point the presented paper reads like a story or account of the event rather than a brief communication of science and assessment of results.

Response:

We would like to thank the reviewer for his/her constructive comments. We agree with the reviewer: despite being the RAPID algorithm fully validated for historical cases, validation (or more specifically cross-validation) for this event was missing in the original version of the manuscript. In the revised manuscript we are now including a comparison between the RAPID and the EMS product (which is based on the same SAR observation) in Figure 2, in the new Table 1, and in text, and discuss differences between the two products.

In particular, we added this paragraph at the end of the Methodology section:

"The RAPID system has been quantitatively validated in past studies against manually derived flood maps using (overall, user, producer) agreement scores, representing (accuracy, true positive rate, precision) parameters of the confusion matrix. Specifically, for Hurricane Harvey, RAPID was validated against the DFO comprehensive flood map of August 30, 2017 (Shen et al., 2019) and against the USGS DSWE Northwestern flood map of June 25, 2019 (Yang et al, 2019). RAPID yielded consistently high agreement scores for Harvey (93%, 75%, 77%) and the Northwestern flood (96%, 84%, 76%). For Hurricane Dorian, we are presenting a comparison between RAPID and the Copernicus Emergency Management Service (EMS) first estimate maps (available at https://emergency.copernicus.eu/mapping/list-of-components/EMSR385/FEP/ALL), both derived from the Sentinel-1 SAR observations. EMS flooding maps are not available for the entire SAR images, but only for the Abaco Islands on September 2, 2019, and for Grand Bahama on September 4, 2019."

We also added this paragraph towards the end of the Results section:

"The agreement (overall, user, producer) scores between RAPID and EMS flooding maps for the Abaco Islands on September 2 and September 4, derived from the confusion matrix shown in Table 1, were (77%, 90%, 41%) and (89%, 61%, 86%), respectively. The high overall and user agreement scores for the September 2 flooding are also depicted in the flood maps of Figure 2 indicating a very good overlap of the two products over the coast of Great Abaco, while the relatively low producer agreement comes from the lack of flood detection by the EMS algorithm over the multiple near-sea-surface-elevation islands, located in the front of the western coast of Great Abaco. The relatively low user agreement score between the two products on September 4 is due to the fact that RAPID classifies some non-flooded areas within the EMS flooded boundary, which are expected to occur as a consequence of the flood recession."

We replaced Figure 2 with the following:
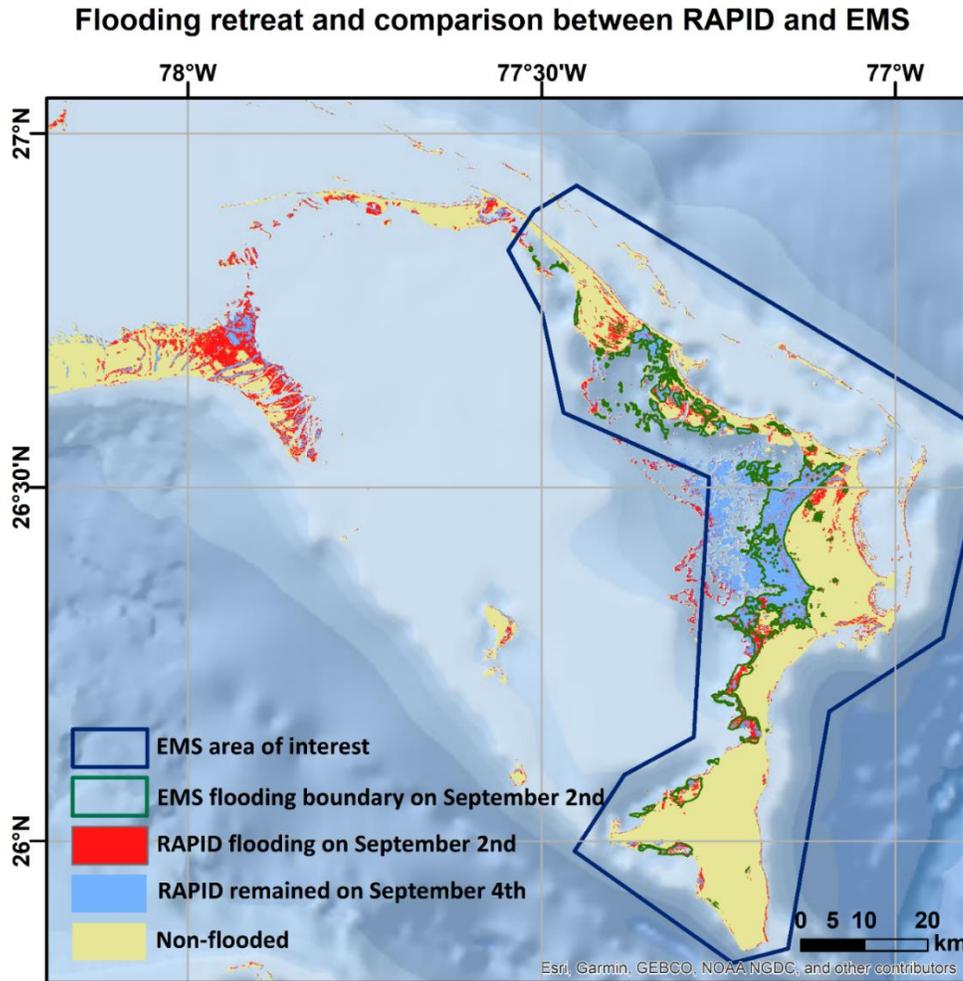
Figure 2: Ocean background from World Ocean Base map (ESRI et al. 2014; list of contributors available at:
http://downloads.esri.com/esri_content_doc/da/WorldOcean_ContributorsDA64.pdf). Flooded and non-flooded areas on September 2 and September 4, 2019 derived from the RAPID algorithm that processed SAR data from the Sentinel-1 overpasses, and flooded boundary from EMS.

And we added Table 1:

Confusion Matrix		September 2 – Great Abaco		September 4 – Grand Bahama	
		EMS		EMS	
		Flooded	Non-flooded	Flooded	Non-flooded
RAPID	Flooded	2,274,927 (14.5%)	3,318,143 (21.1%)	1,880,609 (13.2%)	32,989 (2.3%)
	Non-flooded	260,335 (1.7%)	9,847,017 (62.7%)	1,219,786 (8.6%)	10,710,519 (75.9%)

Table 1: Confusion matrix between RAPID and EMS flooding products for September 2, 2019 overpass over Great Abaco (left) and for September 4, 2019 overpass over Grand Bahama (right). For each matrix, number and percentage of pixels is reported.

Beyond the additions implemented in the paper, we also visually compared, in Figure (a), the VV-pol SAR images on July 4 (dry condition, left) and on September 2 2019 (peak flooding, right). This visual and subjective comparison will not be included in the paper due to space limitations. From the comparison, however, it is evident that all the flat islands in oval 1 in Figure (a), are flooded. Large areas in oval 2 (zoomed in the bottom images for facilitating the comparison) are also partially flooded, despite they are not as dark as other areas. The same occurs for the other red circles. These flooded or partially flooded areas are not captured by the EMS algorithm (Figure 2), probably due to the use of fixed
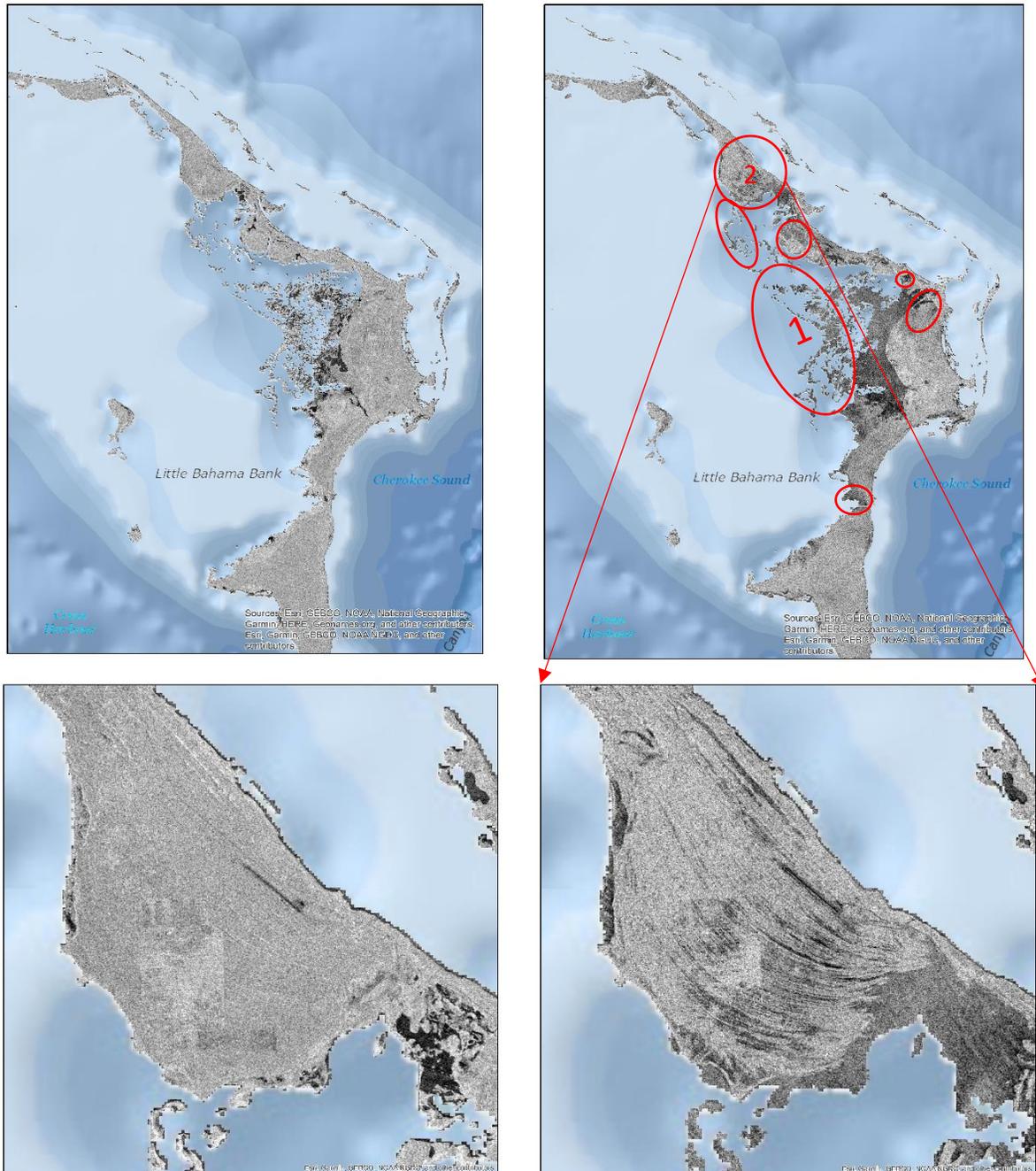
Figure a: VV-polarized SAR images (in dB scale) over Bahamas on July 4, 2019 (left column) and on September 2, 2019 (right column).

thresholds. In RAPID, the segmentation threshold is instead optimized individually for each image to reach the best goodness of fit of the theoretical water-class distribution.

In order to put the RAPID processing chain in context with other automated methods, we introduced the following sentence: *Only a few SAR-based flood delineation methods (e.g. Horritt et al. 2003, Martinis et al. 2009, Matgen et al., 2011, Giustarini et al. 2012, Lu et al. 2014, Chini et al. 2017, Cian et al. 2018) have the potential to be fully automated (Shen et al. 2019b)."* All these references have been extensively discussed in Shen et al. 2019b, which reads as follows:

"Toward automation, Matgen et al. (2011) developed the M2a algorithm to determine the threshold that makes the non-water pixels (below the threshold) best fit a gamma distribution—a theoretical distribution of any given class in a SAR image. They then extended flooded areas using RGA from detected water pixels using a larger threshold—99 percentile of the "water" backscatter gamma distribution—arguing that flood maps resulting from region growing should include all "open water" pixels connected to the seeds. Then they applied a change detection technique to backscattering to reduce over-detection within the identified water bodies caused by water-like surfaces, as well as to remove permanent water pixels. Based on the same concept, Giustarini et al. (2013) developed an iterative approach to calibrate the segmentation threshold, distribution parameter, and region growing threshold (M2b). They applied the same segmentation threshold to the dry reference SAR image to obtain the permanent water area. They claimed, however, that if the intensity distribution of the SAR image were not bimodal, the automated threshold determination might not work.

Lu et al. (2014) used a changed detection approach, first to detect a core flood area that contained a more plausible but incomplete collection of flood pixels, and then to derive the statistical curve of the water class to segment water pixels. The major advantage of this approach is that a bimodal distribution is not compulsory. In practice, a non-bimodal distribution often occurs. The change detection threshold might be difficult to determine and globalize.

Assuming even prior probability of flooded and non-flooded conditions, Giustarini et al. (2016) computed probabilistic flood maps that characterize the uncertainty of flood delineation. The probability reported in this study, however, related to the uncertainty neither in extent nor in time. Rather, it was the uncertainty of a SAR image classification based on backscattering.

Taking advantage of big earth observation (EO) data, the two most recent studies—Cian et al. (2018) and Shen et al. (2019)—implemented full automation of inundation retrieval. With the CD principle underpinning both methods, they employed multiple dry references instead of one supported by operational satellite SAR data for multiple years.

Cian et al. (2018) developed two CD-based flood indices, the Normalized Difference Flood Index (NDFI) and the Normalized Difference Vegetated Flood Index (NDVFI), assuming a number of revisits for each pixel in dry conditions was available."

About automation of the entire processing chain, on the LIST website, at https://www.list.lu/en/news/list-contributes-to-monitor-mozambique-floods-with-satellite-imagery/ is written that: *"The project partners intend to develop a fully automated tool - based on HASARD ® - that could generate different flood risk maps, with no human intervention, as soon as a flood disaster occurs."* Therefore, according to the information written on the website, HASARD triggering by LIST is currently not automated.

The DLR ZKI needs to be activated too, and activation for Hurricane Dorian does not appear on the website: https://activations.zki.dlr.de/viewer/#/en/georss

With the addition of the comparison between RAPID and EMS, the discussion of those results, and the addition of references related to other automated methods, we hope to have addressed the reviewer's concerns about the science scope of this brief communication.