Re-learning observed flood extents can improve remote sensing products

Fabio Brill[1,2], Stefan Schlaffer[3,4], Sandro Martinis[4], Kai Schröter[1], Heidi Kreibich[1]

[1]Section of Hydrology, GFZ [2]Institute for Environmental Science and Geography, University of Potsdam
[3]Department of Geodesy and Geoinformation, TU Wien [4]German Remote Sensing Data Center, DLR

Objective

Satellite-based flood masks often exhibit **high specificity**, meaning that detected flood is reliable, but **low sensitivity** especially in **urban and vegetated** terrain, meaning that the areas classified as non-flooded are less reliable. Reasons are **unavoidable** radar effects in these areas.

We consequently propose to treat these products as **positive and unlabelled (PU)** data, and re-learn the flood extent from other features – such as topography, hydrography, distance to buildings, and rainfall – via a **one-class classifier** (**OCC**), that can operate with only the class of interest being labelled. This allows us to treat the entire flood mask as training area. Using the **2017 Hurricane Harvey** flood in Houston as test case, we approach the research question:

"Given a satellite-based flood mask, where should we expect inundation in reality?"

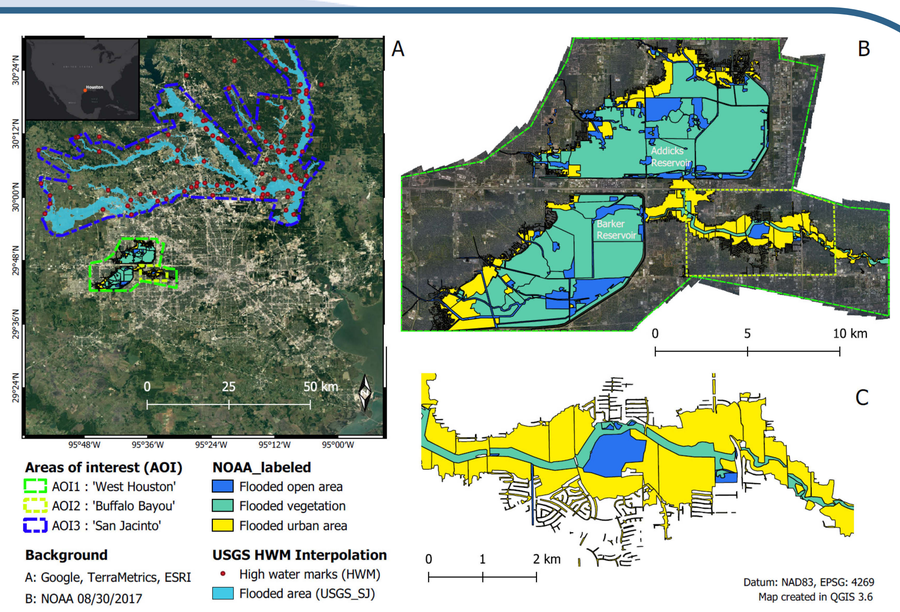

Figure 1. Areas of interest (AOI) and reference data produced from a 50 cm resolution aerial image

Table 1. Evaluation of the 3 initial satellite-based flood masks (details in Table 2) on the 3 AOIs. Metrics include percentages of detected open water, flooded vegetation and flooded urban areas, error bias, sensitivity, specificity, accuracy, Cohen's kappa for all classes as well as for vegetation and urban areas separately.

Product - AOI	%open	%veg.	%urban	EB	Sens.	Spec.	Acc.	κ	$\kappa_{veg.}$	κ_{urban}
EMSR_229 - 1/West Houston	32.06	1.16	0.43	0.001	0.06	0.999	0.63	0.07	0.01	0.01
EMSR_229 - 2/Buffalo Bayou	0	1.16	0	-	0	0	0	0	0	0
EMSR_229 - 3/San Jacinto	-	-	-	0.01	0.05	0.99	0.76	0.06	-	-
DLR_BN - 1/West Houston	69.01	19.60	41.36	0.03	0.32	0.99	0.73	0.34	0.24	0.51
DLR_BN - 2/Buffalo Bayou	3.53	6.93	23.27	0.04	0.21	0.99	0.82	0.28	0.06	0.31
DLR_CNN - 2/Buffalo Bayou	63.77	46.84	42.41	0.13	0.44	0.98	0.86	0.51	0.27	0.50

Methods

- Biased Support Vector Machine (**BSVM**, Liu et al. 2003)
- **MaxEnt** (Phillips et al. 2006)
- Regular **SVM** as **benchmark** model
- Postprocessing by **region-growing** to remove predictions that are not connected to the initial mask

Table 2. Flood masks used for training and validation

Floodmask	Data source	Date of image	Resolution	Usage
EMSR_229	Cosmo-SkyMed	31.08.2017	30 m	Training
DLR_BN	Sentinel-1	30.08.2017	15 m	Training
DLR_CNN	TerraSAR-X	01.09.2017	40 m (32 x 1.25)	Training
NOAA_labeled	Aerial image	30.08.2017	0.5 m	Validation
USGS_SJ	HWM	Maximum extent	3 m	Validation

Table 3. Datasets and features

Feature	Data source	Category
HAND_large_lake_river	NED + OSM	Topo
HAND_major_river	NED + OSM	Topo
HAND_small_stream_canal	NED + OSM	Topo
Dist_large_lake_river	OSM	Topo
Dist_major_river	OSM	Topo
Dist_small_stream_canal	OSM	Topo
Slope	NED	Topo
Curvature	NED	Topo
TWI	NED	Topo
TPI 11x11	NED	Topo
TPI 51x51	NED	Topo
TPI 101x101	NED	Topo
Rainfall_sum	NWS	Rain
Rainfall_acc	NWS + NED	Rain
Dist_to_buildings	Microsoft USBuildingFootprints	Buildings

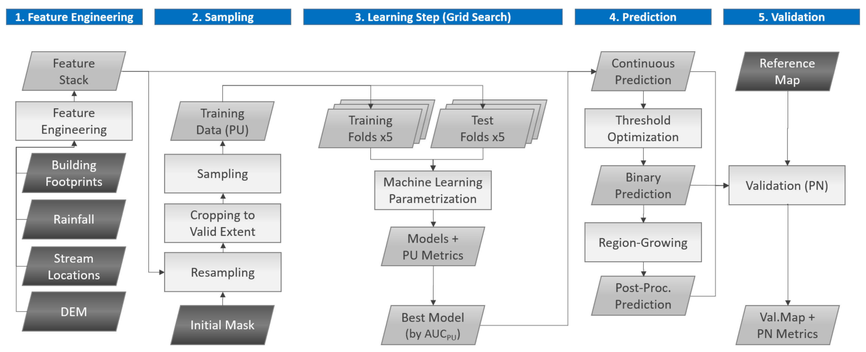

Figure 2. Flowchart of the extrapolation procedure

Results

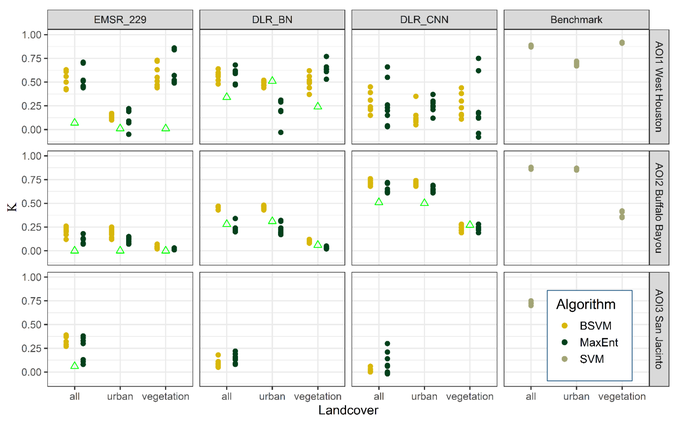

Figure 3. Kappa score obtained, the green triangle denotes the skill of the original products, if the product is defined on that AOI

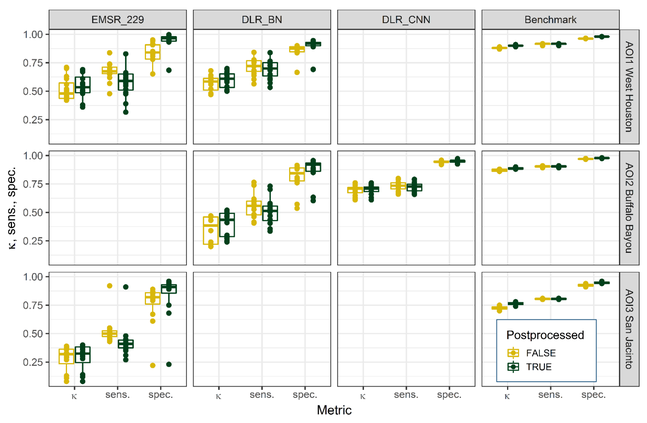

Figure 4. Effect of postprocessing on the metrics kappa, sensitivity and specificity. The range of the boxplots includes the results from both OCC algorithms (BSVM and MaxEnt).

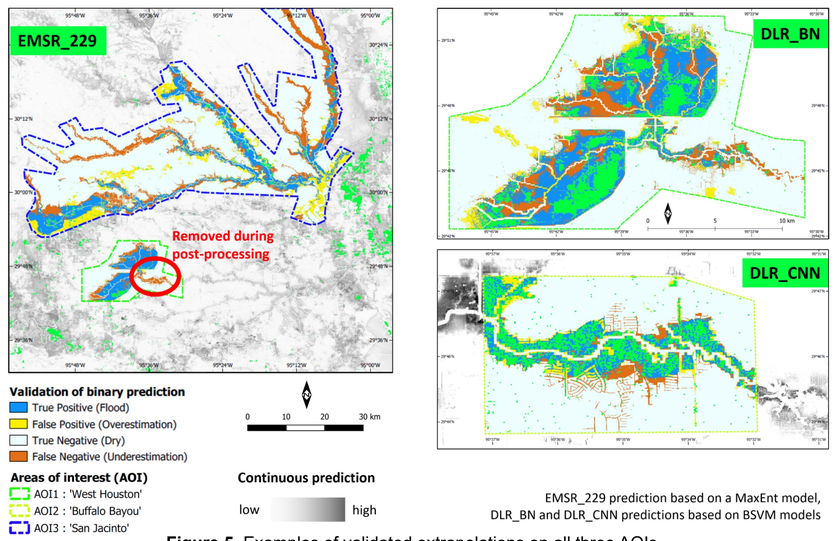

Figure 5. Examples of validated extrapolations on all three AOIs

Take-Home

- **All tested satellite-based flood masks could be considerably improved**, with obtainable increases in Kappa score of **0.2-0.7** in the best case (optimal threshold).
- The entire initial mask can be processed in one piece and the most important features can be derived from a **DEM and stream locations**
- Rainfall and building data did not consistently improve predictions, although some positive cases were observed
- Stability of the default classification threshold depends on the **representativeness of the initial mask**
- The approach could be tested for **data fusion with individual flood location samples** from within urban areas, e.g. from social media or street camera footage.

Brill, F.; Schlaffer, S.; Martinis, S.; Schröter, K.; Kreibich, H. Extrapolating Satellite-Based Flood Masks by One-Class Classification—A Test Case in Houston. *Remote Sens.* **2021**, *13*, 2042. https://doi.org/10.3390/rs13112042
Liu, B.; Dai, Y.; Li, X.; Lee, W.; Yu, P. Building text classifiers using positive and unlabeled examples. In Proceedings of the Third IEEE International Conference on Data Mining, Melbourne, FL, USA, 22 Nov. **2003**.
Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. Ecol. Model. **2006**, 190, 231–259

The RIESGOS project was funded by the German ministry of education and research (BMBF, 03G0876B)