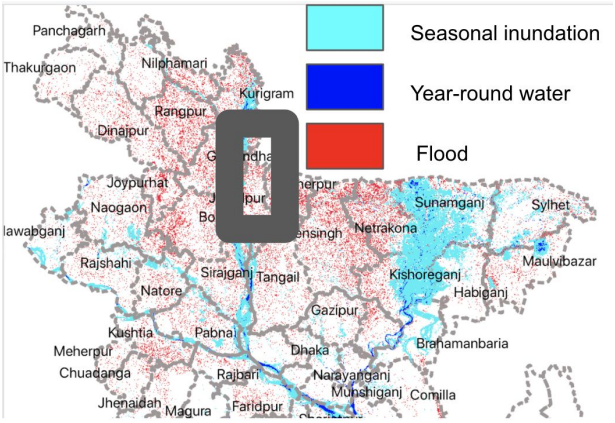
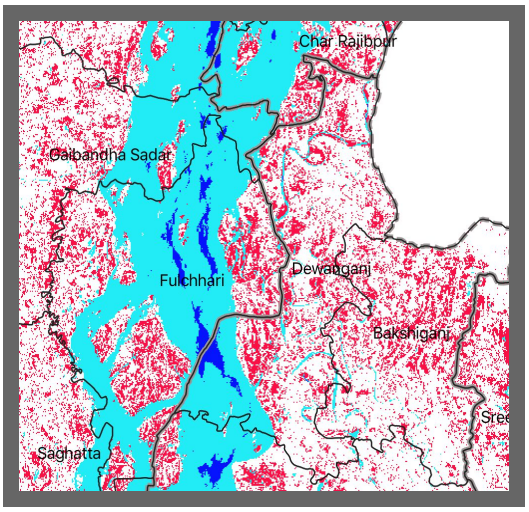
Metadata document for the Flood GeoTiff for the 2015 flood in Bangladesh:
For FFWC
Data analysis by Max Goodman and Beth Tellman

Data: 6 single band 10m resolution geotiff images for flooding for August 17, 2017 (all geotiff images mosaiced together will cover the country) using the Sentinel-1 satellite. Inundation is defined as water area not frequently present (less than 25% of observations) in the previous year (Nov 2015 - Nov 2016), which we assume to be a year in which damaging floods did not occur. Flooded area per upazila in Sirajganj explained 8% of the variation in crop damage for the year 2017 using this method (see details below for how the choice of "baseline" non flood years influences the algorithm result). NOTE: we have not validated the accuracy of this algorithm for Bangladesh, but the paper the algorithm is based on reports accuracy of 87-93% from validation data in Houston and Mozambique.

Values*:
0: Non-Water
1: Flooding VV Band (Identified as water in < 25% of the baseline images)
2: Flooding VH Band (Identified as water in < 25% of the baseline images)
3: Flooding in both VH and VV Bands (High Confidence Flooding) (Identified as water in < 25% of the baseline images)
10: prior inundation; no flag
11: prior inundation; VV only (Identified as water in > 25% of the baseline images)
12: prior inundation; VH only (Identified as water in > 25% of the baseline images)
13: prior inundation; VV + VH (Identified as water in > 25% of the baseline images)
20: Permanent Open Water (Identified as water in > 80% of the baseline images)

*VH and VV and two different bands in the Sentinel-1 imagery





Example map using these data with prior inundation in light blue, flooding in red, and permanent water in dark blue for August 2017.

Method:

Synthetic Aperture Radar (SAR) has a critical advantage over visual-light imagery for our application in Bangladesh: it can detect the presence of surface water yet is unaffected by cloud cover (DeVries et al., 2020). The European Space agency's Sentinel-1 satellites have been collecting SAR data at a ~10m resolution with a 6-day return time since 2014. Our Google Earth Engine algorithm for processing sentinel-1 SAR backscatter was adapted slightly from the algorithm published by DeVries et al. in *Remote Sensing of Environment*, 2020. It relies on the principle that water provides reduced radar backscatter compared to dry land. For each pixel in the SAR image collection, the algorithm considers the scale of the backscatter from the target (flooded) day, as well the time series of backscatter values across a user-defined baseline period (e.g. period where there is no inundation) (Figure 1). This baseline period is intended to represent normal or non-flooded conditions. From each time series, the algorithm extracts a mean and a standard deviation, which it compares against the target-day value in order to calculate a z-score (how many standard deviations away from the baseline mean the target value is). Larger positive or negative z-scores indicate greater likelihood that there was a true difference in ground conditions between the target day and the baseline. Because flooding reduces radar backscatter, only significant negative z-scores indicate flooding. A z-score of negative 2.5 was applied as the flooding significance threshold. This algorithm was run in parallel for both VV and VH radar polarization signatures, and, consistent with DeVries et al. (2020), only pixels that met the threshold in both polarizations were considered to be flooded with high confidence.

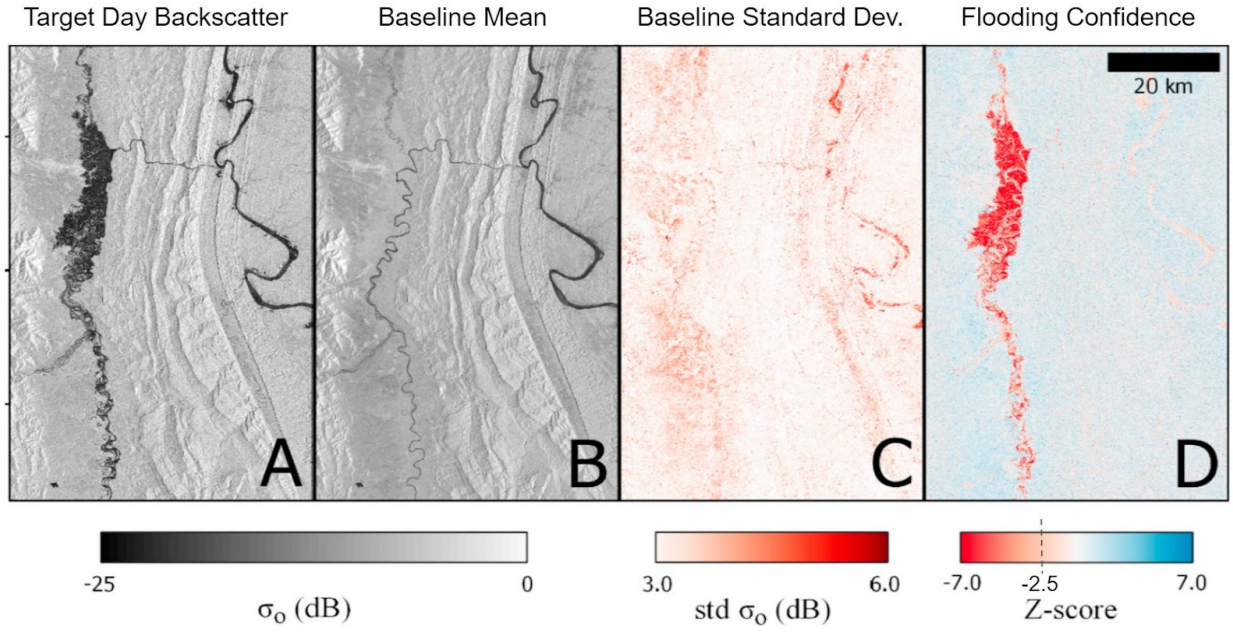


Figure 1. Steps of SAR backscatter analysis. Credit DeVries et al.

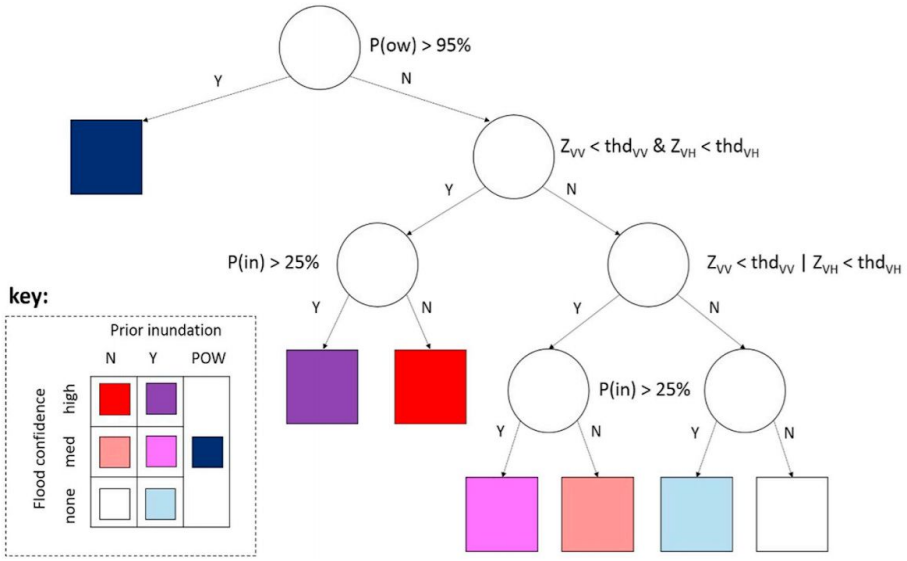


Fig. 4. Classification tree and colour key indicating floodwater classes included in this study. The open water probability, P(ow), and total inundation probability, P (in), were derived from historical Landsat data. Permanent open water (POW) was assigned to all pixels with a P(ow) > 95%. Other flood classes were derived from VV and VH Z-scores (Z_{VV} and Z_{VH}, respectively) with an applied threshold (thd_{VV} and thd_{VH}, respectively).

Figure 2. Devries et al 2020 flood algorithm description. We used thresholds of -2.5 for the VV and VH band for this study (thd vv and thd vh= -2.5)

With required functions available from DeVries at:

- https://github.com/bendv/s1flood

DeVries et al. used historical Landsat imagery (including JRC monthly history) to discriminate between seasonally inundated areas and abnormal flooding (using a threshold of 25% of images showing water). While making this distinction would be very valuable in the application of index insurance to Bangladesh, we were concerned about this method for the same reason we chose not to use JRC directly as a proxy for flooded area: cloud interference. Areas that flooded seasonally during the monsoon would often be cloud-obscured, limiting the percentage of past Landsat images displaying water. Sure enough, when applying this method to Bangladesh, we found that chars and other areas along the course of the Brahmaputra were identified by the algorithm as “previously inundated” due to the river’s historical shifts, while the Haor Basin in Sylhet, which floods annually, was not. To approximate the same Bonna/Barsha distinction on our outputs using SAR-only data, we initiated the algorithm with two different baselines. The first took the 2015-2016 dry season as a baseline, to try and capture the total area flooded at the target date. The second took the full year of 2016 as a baseline, so that flooding which occurred in both 2016 and 2017 would be discounted and only flooding uniquely associated with the disaster year would be displayed. The 2015-2016 year was chosen because the Sentinel-1 data available on GEE only goes back to 2014, and the flooding during the summer of 2016 was less destructive than in the summers of 2015 or 2017.

We applied zonal statistics in Google Earth Engine (through the reduce.Regions tool) across 11 upazilas on the Lower Brahmaputra (Jamuna) to calculate flooded area. This action counted the number of pixels in each upazila with high flooding confidence and multiplied this sum by the average square-meter area of the pixels, to produce an estimate for flooded area

within each upazila. 4 sets of flooded area estimates were synthesized, using the two baselines (dry season 2015-16 and all-year 2016) and the two target dates (August 17th, 2017 and August 23rd 2015). These estimates were compared against each other and then plotted against upazila-level BBS damage data representing households damaged and acres of crop lost. Different damage data availability in the two years prevented a one-to-one comparison: 2015 data were available only for Sirajganj upazilas, and crop damage was reported in only two upazilas. Area sum results are presented here, but their comparison to BBS damage is reported under the section "cross-comparison results."

Table 1. Algorithm Outputs for Jamuna-Adjacent Upazillas

Upazila	Flooded Area (Sqkm) Via DeVries Algorithm			
	Dry-Season Baseline (Nov 2015 - Mar 2016)		All-Year Baseline (Nov 2015 - Nov 2016)	
	Aug. 23, 2015	Aug. 17, 2017	Aug. 23, 2015	Aug. 17, 2017
Bhuapur	14.06	56.20	0.34	10.95
Nagarpur	9.07	38.85	0.25	4.39
Tangail Sadar	13.14	31.56	0.47	5.72
Dhunat	11.27	19.06	0.58	5.72
Sariakandi	30.28	104.69	0.81	29.89
Bera	13.84	38.09	0.40	4.20
Kazipur	28.51	90.71	0.92	26.13
Belkuchi	6.35	28.42	0.18	6.06
Chauhali	14.80	35.91	0.37	4.96
Shahjadpur	14.52	112.22	0.34	9.05
Sirajganj Sadar	15.64	48.63	0.53	11.65

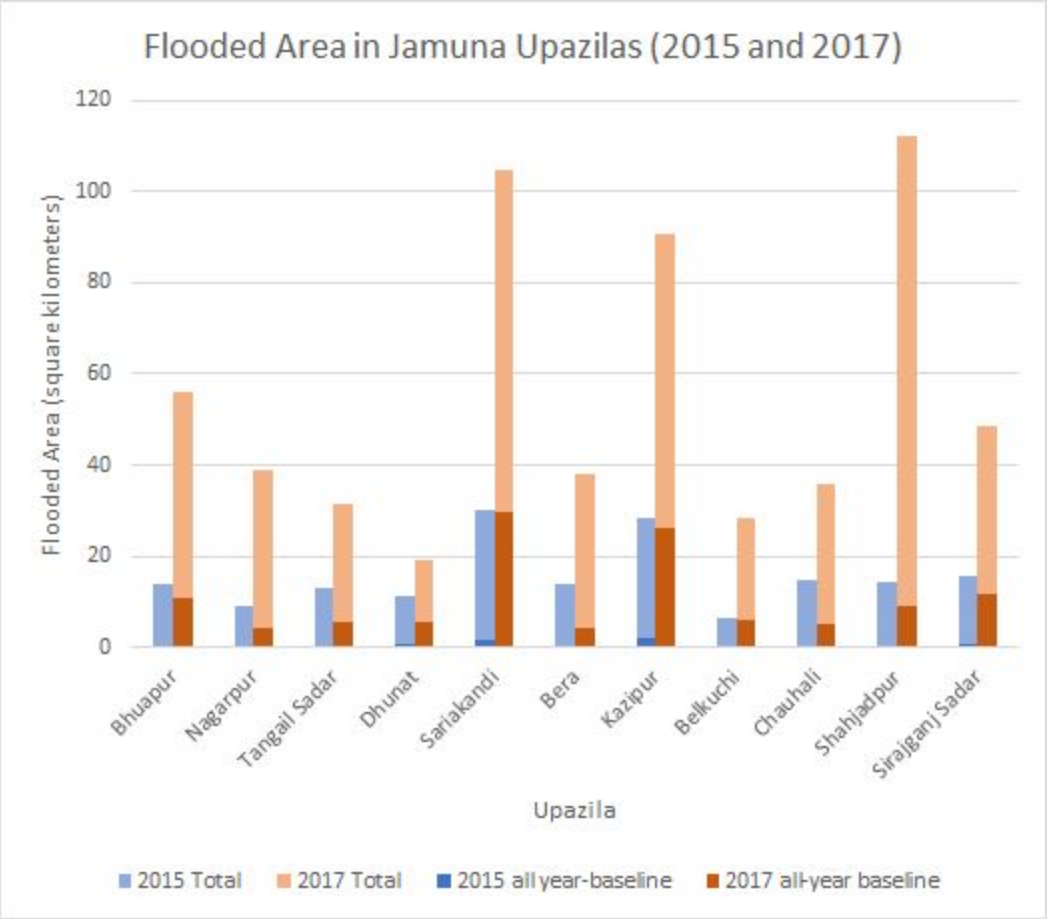


Figure 3. Barplot of flooded area per upazila (data form Table 1).

As expected, 2017 displayed a much greater total flooded area than 2015. Kazipur and Sariakandi, both of which contain large and numerous char-islands, experienced high total flooded areas in both years. Shahjadpur, interestingly, experienced the largest flooded area in 2017, but had a middling flood rank in 2015. Most notably, the all-year baseline reported reasonably-sized flooded areas for 2017, but reported less than one square kilometer of abnormal flooding for every upazila for 2015. This outcome has a couple of possible explanations: (1) summer flooding in 2015 was not much more extreme than in 2016, and the absence of BBS data doesn't validate 2016 as a non-disaster year (2) flood timing rather than flood area is the differentiating factor between 2015 and 2016, so damaging floods can't be differentiated purely on the basis of abnormal location.

Using the dry-season baseline, SAR flooded area explained ~4% of the variation in the number households affected and showed no relationship at all to acreage of crop damage (Figure 4). Using the full-year baseline increased the predictive power of flood area damage to ~6% on households affected and ~8% on crop damage (Figure 5). This result validates the idea that SAR-derived indexes would be improved by the discrimination between bonna and barsha flooding, and should use a baseline period that includes a non-disaster rainy season. Nonetheless, because of the small sample size, none of the regressions are statistically

significant at a P <0.05 level, and we can therefore not yet definitively conclude that flooded area is a meaningful index for assessing damage either to households or agriculture.

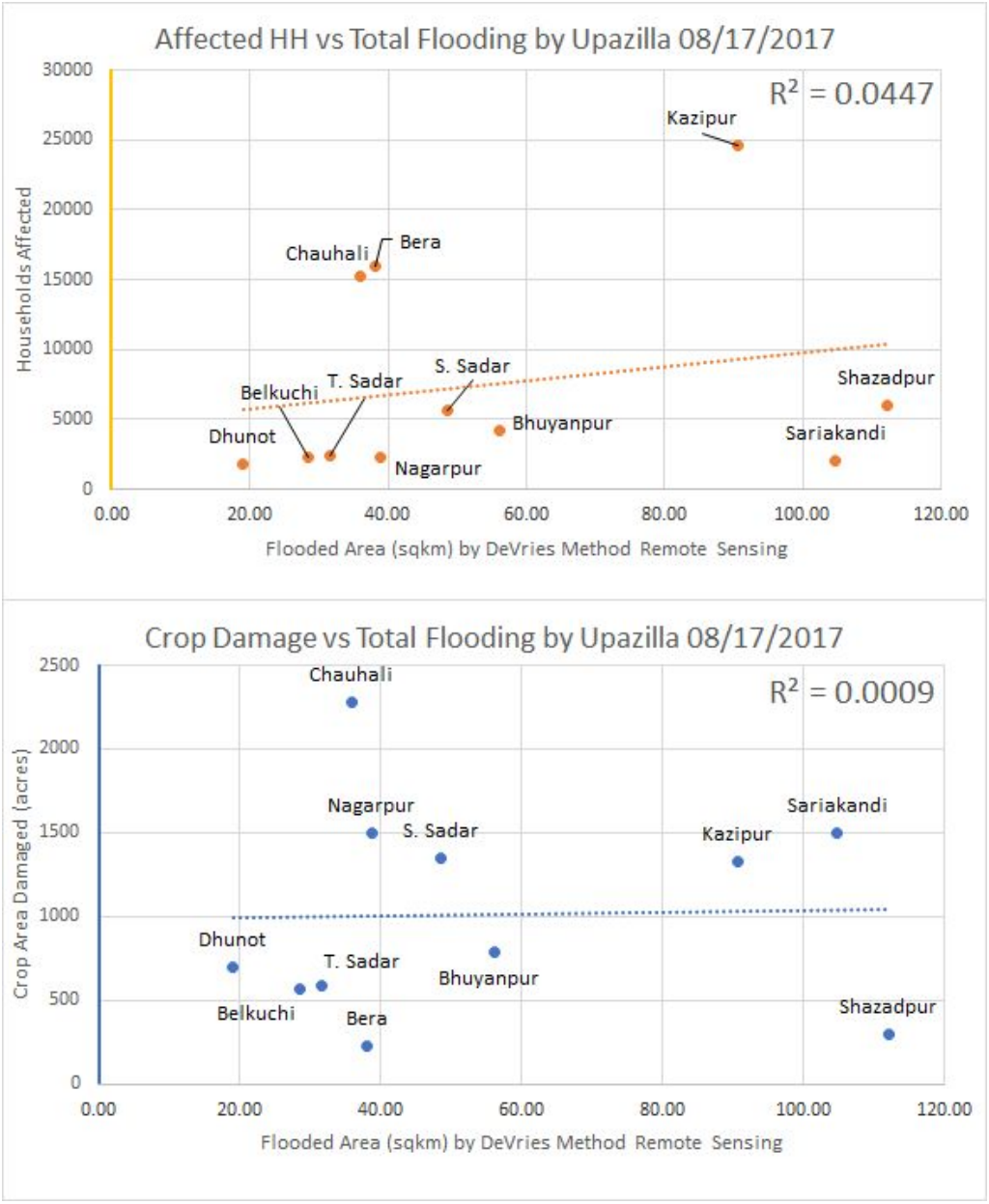


Figure 4. Regression plots of Crop Damage and Flooding-Affected Households by flooded area using DeVries Method SAR. Baselined in the dry season to represent total flooded area.

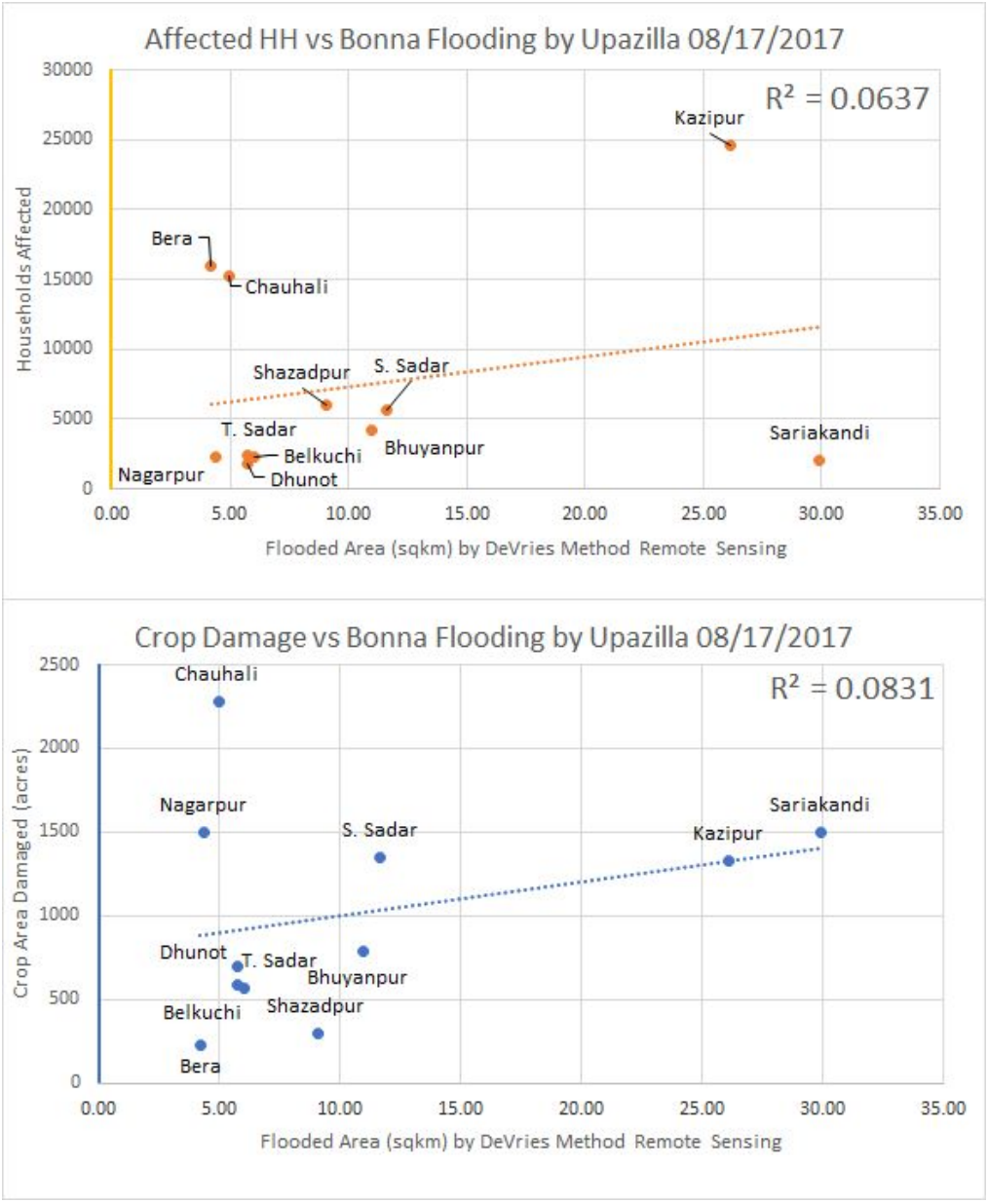


Figure 5. Regression plots of Crop Damage and Flooding-Affected Households by flooded area using DeVries Method SAR. Baselined over the entirety 2016 to represent abnormally flooded area (bonna).

References:

DeVries, Ben, Chengquan Huang, John Armston, Wenli Huang, John W. Jones, and Megan W. Lang. “Rapid and Robust Monitoring of Flood Events Using Sentinel-1 and Landsat Data on the Google Earth Engine.” *Remote Sensing of Environment* 240 (April 2020): 111664. https://doi.org/10.1016/j.rse.2020.111664.