Mapping Flood-Exposed Road Segments from Freely Available Satellite Imagery During The 2022 Flood Season

Oladimeji Basit Alaka[1,*]; Olayemi Bolaji[2]; Ridwan Tiamiyu[3]
[1]Rivetline Concept Limited, Lagos, Nigeria
[2]Leedigital Research and Innovation, Ibadan, Nigeria
[3]Fountain Construction Company, Nigeria
[*]Corresponding author: basitoladimeji@gmail.com
D.O.I: 10.56201/ijgem.v8.no2.2022.pg117.143

Abstract
Free synthetic aperture radar maps flood water through cloud and is increasingly used to infer which roads are cut. We ask what that inference is worth, using Sentinel-1 backscatter over two contrasting Nigerian study areas across the 2022 season, with 53,158 road segments and 14,878 km of Open Street Map network. Skill is measured against a hydrological reference of impassable segments, built by driving a calibrated rainfall-runoff and terrain model with catchment rainfall. On the Kwara floodplain, at 20percent canopy, detection reaches an F_1 of 0.356 and an area under the receiver operating curveof 0.746, modest but usable for screening. On the Osun basin, at 76 percent canopy, no tile of any scene is bimodal and the method makes no claim across fifteen acquisitions. That null measures the detector's blindness, not the incidence of flooding: the same season put 1,045 mm of rain on the catchment and the reference calls 566 Osun segment-days impassable. Recall falls 26.0 percentage points across built-up land against the roughly 15 points reported for urban radar; the gradient across road class instead tracks the embankment and pass ability ladders of the reference model. The sharpest failure is geometric: across the 473 Kwara segments whose 30 m context holds a mapped watercourse or that carry a bridge tag, precision is 0.008 against 0.438 elsewhere, and only 8 of them carry the tag, so the mechanism is proximity to water rather than bridge decks. Within one basin, F_1 varies 18.8 times more across spatial blocks than across reference replicates. We publishcalibrated closure sets with their confidence rather than a binary map.

Keywords: *flood mapping, Sentinel-1, road network exposure, detection skill, canopy, validation, Nigeria*

1 Introduction

The 2022 flood season displaced more than a million people across Nigeria and cut road links in states that had not flooded in living memory. Agencies responding to that kind of event need to know which roads are usable, and the appeal of free satellite radar is obvious: it sees through cloud, its coverage is global and systematic, and it costs nothing(Twele et al. 2016;DeVries et al. 2020). The flood mapping literature is mature and its accuracy is reasonably well characterized. Thresholding methods from Otsu (1979)through the split-based approach of Martin is et al. (2009)and the hierarchical refinement of Chini et al. (2017)a restandard, Landuyt et al. (2019)assesses them systematically, and Bernhofen et al. (2018)compares flood models and

observations collectively. Water extent is a solved enough problem that global products exist(Pekel et al. 2016;Tellman et al. 2021).

Roads are not water extent, and the step from one to the other is where the useful question lives. A flooded pixel beside a road does not close the road. A road under canopy is invisible whether or not it is flooded (Tsyganskaya et al. 2018). A road in a built-up area sits in a radar environment dominated by double-bounce returns that mimic and mask water(Mason et al. 2010,2014;Giustarini et al. 2013). And a road that runs beside or over a watercourse sits next to water the sensor will find on any date, flood or no flood.

We therefore do not ask whether radar can map floods. We ask what its road closure inferences are worth, where they fail, and whether the failures are predictable enough to work around. That is a validation question, and the validation literature is clear about how to answer it(Congalton 1991;Olofsson et al. 2014;Wing et al. 2017), including the argument of Pontius and Millones (2011)against reporting agreement statistics that conflate quantity and allocation error.

The setting is not a neutral backdrop for that question but part of what makes it answerable. The two basins differ sharply, one carrying 20 percent canopy and 0.6 percent built-up land and the other 76 percent canopy and 6 percent built-up, and both are observed by the same relative orbit on the same dates. Working against that, the nominal revisit is 12 days, with one 24-day gap in our series, so flash flooding on smaller rivers can begin and recede unobserved and that bounds what any space borne method can see. The road data is volunteered(Barrington-Leigh and Millard-Ball 2017;Camboim et al. 2015;Herfort et al. 2021), so the network itself carries completeness and attribute uncertainty. And the reference set of impassable segments against which skill is measured is derived from a calibrated rainfall-and-terrain model rather than from a field record of closures. What the paper adds is a detection skill measured for road closure rather than for water extent, and measured separately by road class, canopy, built-up fraction and height above drainage, so that the failure modes are located rather than averaged away. One of those measurements is a complete failure, and it is reported as one: in the canopy-heavy basin the method produces no detections at all, and the mechanism behind that is worth more to a reader than a low score would be. The output is published as calibrated closure probabilities carrying their confidence rather than as a binary map, because a downstream user needs the uncertainty more than the classification. Section 2 reviews the mapping, validation and road-impact literatures. Section 3 sets out the study inputs and how the reference set was built. Section 4 describes the processing chain and the skill design. Section 5 reports overall and stratified skill and the failure modes, and Section 6 the sweeps and the variance decomposition. Section 7 takes up what follows.

2 Related Work

2.1 Mapping water with radar

Radar flood mapping rests on the fact that smooth open water returns very little energy to the sensor, so inundation appears as a dark mode in the backscatter histogram. Otsu (1979)and Kittler and Illingworth (1986)supply the thresholding machinery, Lee (1980)the speckle filtering, and Martinis et al. (2009)the split-based tile selection that finds the sub-scenes where a bimodal histogram actually exists. Chini et al. (2017)refines the approach hierarchically, Twele et al. (2016)builds an operational Sentinel-1 chain,Shen et al. (2019)a rapid variant, and DeVries et al. (2020)implements the family at scale on cloud infrastructure. Landuyt et al. (2019)evaluates these methods against one another. The split-based step matters more in this paper than usual, because it is the step that fails completely in one of our two study areas.

2.2 Optical alternatives and why they do not rescue the problem

Optical indices are simpler and better established(McFeeters 1996;Xu 2006;Feyisa et al. 2014), and Sentinel-2 supplies them freely(Drusch et al. 2012). Their limitations cloud, and Wilson and Jetz (2016)shows that tropical cloud climatology removes exactly the dates a flood study needs. We retrieved optical scenes and report their availability in Section 4 rather than assuming it.

2.3 Validating a classification honestly

Congalton (1991)and Olofsson et al. (2014)set out accuracy assessment practice, Cohen (1960)supplies the agreement statistic, and Pontius and Millones (2011)argues against relying on it because it conflates quantity disagreement with allocation disagreement. We follow that advice and report precision, recall, F1and the critical success index alongside the receiver operating curve rather than leading with a single agreement number.

Wing et al. (2017)and Bernhofen et al. (2018)establish the accuracy envelopes that flood extent methods achieve, which is what our own water extent result is benchmarked against. Trigg et al. (2016)examines the credibility of continental flood models.

2.4 From water to roads

The transport-impact literature is where the step this paper examines gets made. Pregnolato et al. (2016)and Pregnolato et al. (2017)relate inundation depth to vehicle disruption and establish that shallow water reduces speed long before it stops traffic, Yin et al. (2016)models pluvial flooding on an urban road network ,Kermanshah and Derrible (2017)treats network robustness under flooding, Arrighi et al. (2019)preparedness, and Koks et al. (2019)multi-hazard exposure of global infrastructure.

Every one of these requires a closure input, and most take it from a flood extent product without characterizing what that step costs. Our contribution is to characterize it.

2.5 Terrain, and the urban and vegetation problems

Height above nearest drainage(Rennó et al. 2008;Nobre et al. 2011;Nobre et al. 2016)is the standard terrain normalization, resting on elevation data whose quality bounds it(Farr et al. 2007;Yamazaki et al. 2017;Hawker et al. 2019).Alfieri et al. (2013)gives the continental forecasting context. Two physical problems limit radar specifically. Under vegetation, water is either invisible or produces enhanced double-bounce return rather than a dark mode(Tsyganskaya et al. 2018;Chapman et al. 2015). In built-up areas, corner reflections between walls and ground both mimic and mask flooding(Mason et al. 2010,2014;Giustarini et al. 2013).Giustarini et al. (2013)reports roughly a 15-point accuracy penalty in urban settings, which is the figure our own built-up results compared against.

2.6 Nigerian flooding

Nkeki et al. (2013)analyzes the Niger-Benue flood system, Adelekan (2010)and Agbola et al. (2012)the urban cases, Douglas et al. (2008)the African picture, and Nkwunonwo et al. (2020)and Echendu (2020)the modeling and policy context.

2.7 What is missing

No study we located reports flood detection skill for *road segments* stratified by the physical conditions that govern radar performance, and none reports what happens on the roads that run beside or across mapped watercourses, where standing water is present whether or not the road is

cut. The flood mapping literature validates water extent; the transport literature consumes closure sets without validating them.

3 Study Inputs and Reference Construction

The imagery is Sentinel-1 radiometrically terrain-corrected gamma-nought backscatter at 10 m, relative orbit 103 ascending, on 22 dates per study area, 15 in the flood season and 7 in the dry base line. Kwara spans two frames on each date and Osun one, so the retrieval is 44 scenes for Kwara and 22 for Osun, 66 in total, from the Microsoft Planetary Computer. The season dates run from 8June to 5 December 2022 at the nominal 12-day revisit, with one 24-day gap between 19 August and 12September that falls on the rising limb of the Niger flood. Optical cover is Sentinel-2 Level 2Asurface reflectance across 112 dates from the public mirror. Terrain comes from the Copernicus DEMGLO-30, canopy and built-up fraction from ESA World Cover 2021, rainfall from CHIRPS across 306 days and four bounding boxes, and the network from OpenStreetMap, 53,158 road segments totaling14,878 km. Table 1 lists each with its license and vintage.

Table 1. Sources used in this study, with provider, vintage, spatial resolution and license.

Product	Provider	Vintage	Resolution	License	Volume
Sentinel-1A GRD, terrain corrected	ESA / Microsoft Planetary Computer	2022	10 m	CC-BY 4.0	66 scenes
Sentinel-2 L2A surface reflectance	ESA / AWS Open Data (Element 84)	2022	10–20 m	CC-BY-SA 3.0 IGO	112 dates
Copernicus DEM GLO-30	ESA / AWS Open Data	2011–2015	30 m	Copernicus free license	5 tiles
ESA WorldCover v200	ESA / AWS Open Data	2021	10 m	CC-BY 4.0	2 tiles
CHIRPS v2.0 africa_daily	Climate Hazards Center, UCSB	2022	0.05°	public domain	306 days, 4 boxes
OpenStreetMap highways	OSM contributors, Overpass API	retrieved 2022	vector	ODbL 1.0	42,505 ways
OpenStreetMap waterways	OSM contributors, Overpass API	retrieved 2022	vector	ODbL 1.0	1,119 ways

Note: All seven sources are freely available without registration or payment, and each was reached through a public catalogue endpoint that required no account: the Planetary Computer search interface with an anonymous read-only token for Sentinel-1, the Earth Search index of the AWS Open Data mirror for Sentinel-2, open buckets for the elevation and land cover tiles, the Climate Hazards Center archive for rainfall, and the Overpass API for the two OpenStreetMap layers. Scene identifiers, acquisition times and orbits are held in the archived scene inventories, and every response is cached so that the processing chain reruns without a network.

Two intended sources are recorded as unresolved rather than absent, a distinction that matters for anyone repeating this: Google Earth Engine was unavailable to this project and the CopernicusData Space endpoint returned not-found for the collections we needed. The emergency mapping products of the Copernicus service, UNOSAT, the Dartmouth Flood Observatory and the Global Flood Database were sought and not obtained within this study.

The reference set of which segments were impassable on which dates was derived from a

documented rainfall-and-terrain model. CHIRPS catchment rainfall drives a two-store linear reservoir; storage is converted to flood stage above the main-stem drainage network by a power law whose scale, storage threshold and routing lag are fitted; stage is differenced against each segment's freeboard, which is its height above the channel band plus its formed embankment; and a segment is called impassable when the resulting depth exceeds a class-dependent threshold. Every skill figure in this paper is therefore an agreement between an imagery-derived classification and that hydrological reference, and what it establishes is the structure of detector failure across strata rather than accuracy against a field record of closures.

Osun's reference takes Kwara's fitted scale and storage-threshold quantile, driven by Osun's own catchment rainfall, with a lag of zero set from catchment scale. The transfer is forced: the Osun radar detects no standing water all season and the peak flooded share of its trustworthy stratum is0.16 percent, so calibrating its reference against its own radar would return a reference of zero closures and let the detector score perfectly on a basin where it is blind. The absolute level of every Osun reference number therefore follows from Kwara's fit, and the tables and benchmark rows that carry one note the transfer. Osun's own imagery is the thinner of the two on every count: no bimodal tile on any of fifteen acquisitions, a trustworthy stratum of 7.2 percent of the study are aagainst Kwara's 40.8 percent, and a mean Sentinel-2 clear-sky fraction of 19.5 percent. Two further parameters of the reference model are set by road class. Embankment heights areas signed by road class, from 1.10 m for trunk down to 0.05 m for a track, and the depth at which a segment becomes impassable is assigned by road class as well, from 0.45 m for trunk down to 0.15 m for a track. Their joint effect is that the reference calls a falling share of segment-days impassable as road class rises, from 5.94 percent on tracks to 0.04 percent on trunk roads. Section 5 reports skill by road class against that base rate.

Bugs found and fixed. Three are worth recording because each produced plausible wrong output. A cloud-optimized GeoTIFF decimation factor assumed meter units on a degree-referenced grid. Flat-area routing seeded on river banks, which capped flow accumulation at 6 km^{2}on the Niger, a river draining two million. And a float32 cast destroyed the tie-break gradient inthe drainage algorithm.

4 Methods

4.1 The two study areas

The design rests on a contrast. Table 2 and Figure 1 describe the two areas: the Kwara floodplain of the Niger, at 20 percent canopy and 0.6 percent built-up, and the Osun River basin, at 76 percent canopy and 6 percent built-up. They are observed by the same relative orbit on the same dates with the same processing chain, so any difference in performance is attributable to the surface rather than to the acquisition.

Table 2. The two study areas: terrain, land cover, road network and the imagery available over the 2022 flood season.

Quantity	Kwara	Osun
Study area (km^2)	6,305	5,077
Elevation, median (m)	131	363
Slope, median (deg)	1.93	3.63
HAND, grid median (m)	8.6	14.8
Tree cover (% of area)	20.2	76.0
Built-up (% of area)	0.6	6.1

Quantity	Kwara	Osun
Cropland (% of area)	35.3	1.4
Road segments	11,280	41,878
Network length (km)	4,127	10,751
Median segment length (m)	397	242
trunk (km)	26.8	151.3
primary (km)	198.2	329.7
secondary (km)	20.9	464.6
tertiary (km)	306.2	787.7
local (km)	2,095.3	8,506.1
track (km)	1,479.7	511.6
HAND, segment median (m)	3.6	14.9
Segments with mapped water within 30 m	472	214
Segments carrying an OSM bridge tag	8	13
Mean tree cover on segments	0.111	0.318
Mean built-up on segments	0.078	0.504
Sentinel-1 season dates	15	15
Sentinel-1 baseline dates	7	7
Sentinel-2 season dates	75	37
Sentinel-2 mean clear-sky (%)	33.1	19.5
Sentinel-2 dates above 50% clear	23	7
Season rainfall, Jun–Nov (mm)	936	1,045

*Note:*HAND is height above nearest drainage, derived from the Copernicus DEM GLO-30 by depression filling, D8 routing and flow accumulation. Elevation, slope and the grid median of HAND are taken over every cell of the analysis grid; the segment median of HAND is taken over segments, and the two differ because the network is not distributed like the terrain. Tree cover and built-up fraction are ESA WorldCover 2021 classes 10 and 50. Class 50 covers buildings, roads and other built structures, so a road's own surface scores as built-up: this is why the mean built-up fraction on Osun segments is 0.504 while built-up land is 6.1 percent of the Osun study area, and the two numbers are not alternative estimates of the same thing. Segment attributes are measured within 30 m of the centerline. Mapped water within 30 m and an OSM bridge tag are separate counts and overlap only partly; the stratum used in Table 4 is their union. Clear-sky fraction is computed from the Sentinel-2 scene classification layer over the study area itself, not from the tile-level metadata figure.

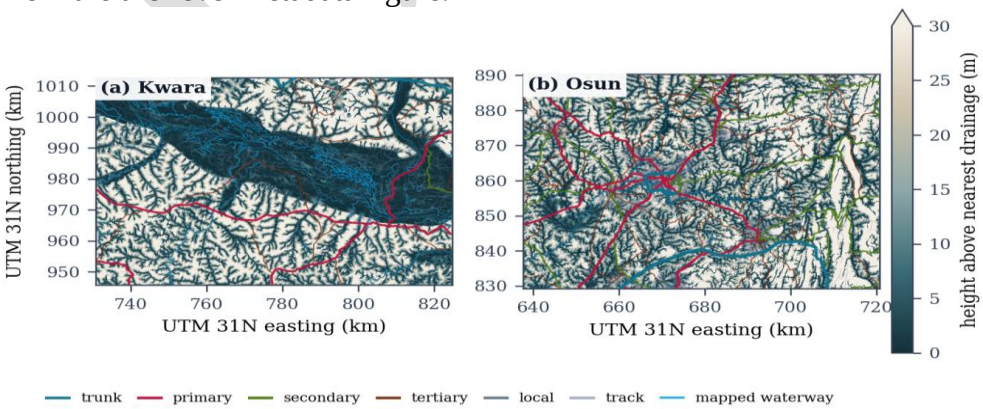

Figure 1. The two study areas, observed by the same relative orbit on the same dates. Each panel

maps height above nearest drainage over the analysis grid, with the OpenStreetMap road network drawn by class and the mapped drainage overlaid. (a) is the Kwara floodplain of the Niger, 20 percent canopy and 0.6 percent built-up land, whose network runs across a low flat plain; (b) is the Osun River basin, 76 percent canopy and 6 percent built-up, whose network sits higher above its drainage. Coordinates are UTM zone 31N.

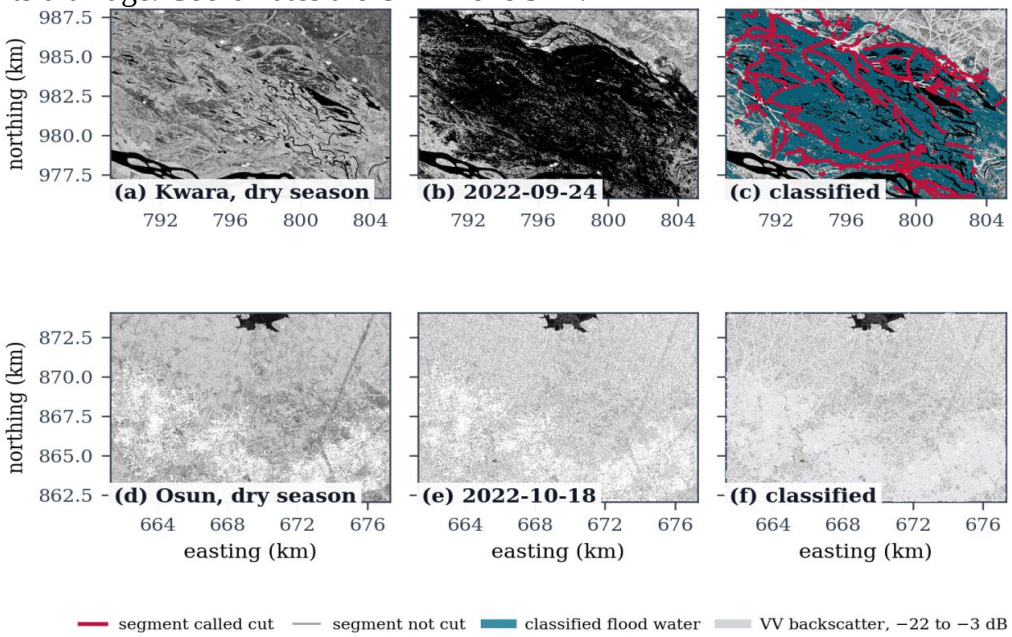

Figure 2. Sentinel-1 VV backscatter over a sub-window of each study area, before and during the flood season, with the classified inundation and the road network. Top row, Kwara: (a) the dry-season median, (b) the seasonal peak of 24 September 2022, (c) the same scene with classified flood water shaded and segments called cut drawn in crimson. Bottom row, Osun: (d) the dry-season median, (e) 18 October 2022, (f) the classified result, which carries no shaded water and no cut segment. The dark mode that thresholding depends on is plainly present in (b) and absent in (e). Coordinates are UTM zone 31N. Figure 2 shows what that contrast looks like in the imagery itself. Over Kwara the seasonal peak darkens most of the floodplain and the classified water follows the low ground. Over Osun the peak scene is visually indistinguishable from the dry-season baseline, and no water is shaded, which is the failure this paper spends most of its length characterizing.

Optical imagery does not rescue the problem, and Figure 3 reports its availability rather than assuming it. Of 75 Sentinel-2 acquisitions over Kwara in the season, 23 exceed 50percent clear sky, and of 37 over Osun, 7 do. The clear dates cluster in November, after the Niger peak has passed.

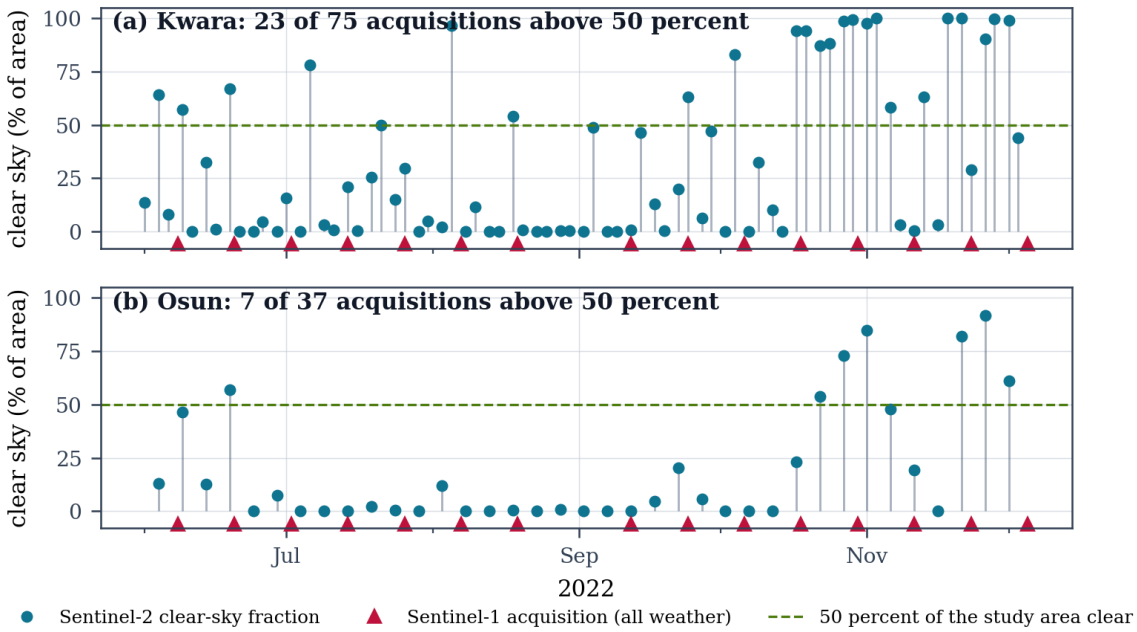

Figure 3. What the optical sensor could see. Each stem is one Sentinel-2 acquisition, plotted at the fraction of the study area its scene classification layer calls clear; the dashed line marks 50 percent; the triangles along the base mark the Sentinel-1 acquisitions, which are unaffected by cloud. (a) Kwara, 23 of 75 acquisitions above 50 percent clear. (b) Osun, 7 of 37. Clear dates concentrate in November, after the Kwara flood peak of late September.

4.2 Processing chain

Backscatter is speckle-filtered, tiles are tested for bimodality following
Martinis et al. (2009), a threshold is selected where bimodality exists, and the resulting
water mask is refined against terrain using height above nearest drainage (Nobre et al. 2011). Road segments are intersected with the water mask under a buffer, and aper-segment water fraction is converted to a closure probability that is then calibrated.

The bimodality test is the step to watch. A date on which fewer than three tiles of the scene are bimodal yields no data-driven threshold and falls back to a fixed and deliberately conservative −20 dB. That is a designed safeguard against thresholding noise, and in one of our two areas it fires on every scene. The fallback is conservative rather than empty: on the two Kwara dates whereat fired it still classified 0.18 and 1.01 $km^2$of water and called one segment cut, and across the fifteen Osun dates it classified between 0.00 and 0.02 km^2on seven of them. What it never produced in Osun was a single segment above the decision cut.

4.3 Skill assessment and stratification

Skill is reported as precision, recall, F1 and the critical success index at segment level, with a receiver operating curve across the decision threshold. Following Pontius and Millones (2011)we do not lead with an agreement coefficient.

The stratification is the substance. Skill is computed separately by road class, by canopy fraction, by built-up fraction, by height above nearest drainage, and separately for segments with mapped water inside their 30 m context, because each of these corresponds to a physical mechanism that should affect radar performance in a predictable direction.

The stratifies differ in what a gradient across them can be attributed to. Canopy fraction and

built-up fraction are read from ESA World Cover and enter nothing in the reference model, so a gradient across them is a property of the detector. Height above nearest drainage and proximity to mapped water enter only through the terrain that also drives the detector's own mask. Road class enters directly, through the embankment and pass ability ladders set out in Section 3, so its gradient has to be read against the base rate in each class.

Metrics within a stratum are computed on the confusion cells pooled across the twenty reference replicates rather than by averaging twenty per-replicate ratios. A thinly populated stratum can hold no reference positive in some replicates, which makes recall undefined there and poisons an average while leaving precision finite, and the result is a table that prints an F1 with no recall beside it. Pooling is defined whenever the pooled cells are, and where it is not defined every quantity describing the positive class is left blank together.

4.4 Calibration and the output for downstream use

Closure probabilities are calibrated so that a stated confidence corresponds to an observed frequency, and the calibration is assessed by Brier skill score. The result is published as a closure set carrying per-date water fraction and calibrated confidence per segment rather than as a binary map, with a documented schema, because a downstream network study needs the uncertainty.

4.5 Topology of the delivered closure sets

The published closure set is a segment attribute table with a time series attached. It is not a routable network, and any analysis that needs routing must rebuild the topology from the geometry rather than from the delivered node identifiers.

Each segment record carries node from and node to, integer identifiers formed by snapping the segment's two endpoints to a 5 m grid. The processing chain splits every Open Street Mapway into analysis segments of roughly 400 m before anything else happens, so the only points that ever-become segment endpoints are the cut points of a way. Two consecutive pieces of one-way share an endpoint and therefore share an identifier; two different ways meeting at a junction almost never do, because the junction sits in the interior of at least one of them and an interior vertex is never hashed. The identifiers chain a way. They do not build a network.

The consequence is large enough to name in numbers. A graph assembled from the published identifiers puts Kwara into 2,237 components whose largest holds 75.0 km of 4,127.2, or 1.82percent of the network, and Osun into 17,969 components whose largest holds 115.4 km of 10,751.0, or 1.07 percent. No connectivity, betweenness, shortest-path or cut result is meaningful on that graph, and this paper reports none. The repair costs nothing and needs no new download. The cached Overpass responses were retrieved with full geometry, so every way carries its ordered list of OpenStreetMap node identifiers. Joining ways wherever they share a node identifier, and mapping the result onto these segments through the way identifier each carries, gives 62 connected components in Kwara and 124 in Osun. The delivered file now carries the measured component counts and this recipe in its own metadata, under a revised schema version, so that a consumer holding the first release is forced to re-read before proceeding.

4.6 Variance reporting

Two variance components are computed and reported separately throughout, never pooled. The replicate component is the spread of a statistic across twenty independent realizations of the reference model, which differ in the per-segment elevation error the reference carries, with the study area held fixed. The block component is the spread of the same statistic across six spatial

blocks within one study area with the reference replicate held fixed. Neither is a between-basin variance: with two study areas there is no such estimate to be had, and the paper does not claim one.

5 Results

5.1 One area works modestly, the other not at all

Table 3 and Figure 4 report overall skill.

On the Kwara floodplain, segment-level detection reaches an F1 of 0.356, with precision 0.378, recall 0.337, a critical success index of 0.217 and an area under the receiver operating curve of0.746. That is modest. It is enough to screen a network for attention and not enough to dispatch on.

Table 3. Detection skill at the operating point, with the two variance components reported separately.

Area	Precision	Recall	F1	CSI	MCC	κ	Positives
Kwara	0.378	0.337	0.356	0.217	0.336	0.335	5,709
replicate SD (n=20)	0.0135	0.0077	0.0097	0.0072	0.0101	0.0099	
area SD (6 blocks)	0.1622	0.2066	0.1834	0.1193	0.1705	0.1726	
Osun	–	0.000	0.000	0.000	–	0.000	566
replicate SD (n=20)	–	0.0000	0.0000	0.0000	–	0.0000	
area SD (6 blocks)	–	0.0000	0.0000	0.0000	–	0.0000	

Note: The operating point is a split-based backscatter threshold selected on each date's own imagery, after Martinis et al. (2009): the scene is tiled, each tile is tested for the bimodality that open water produces, and Otsu's split is taken on the pooled histogram of the most bimodal tiles. A date with fewer than three bimodal tiles falls back to a fixed-20dB, which is conservative rather than empty: it fired on 2 of 15 Kwara dates and on all 15 Osun dates and still mapped a trace of water on each. The rest of the operating point is a 30 m buffer from the centerline and a decision cut of 25% of buffered cells classified as water. CSI is the critical success index and MCC is Matthews' correlation coefficient. The replicate standard deviation is the spread across 20 independent realizations of the reference model with the area held fixed; the area standard deviation is the spread across spatial blocks with the reference replicate held fixed. The two are different kinds of uncertainty and must not be added. Precision, recall, F1, CSI, MCC and$\kappa$are means over the 20 replicates; Positives is the count on the baseline replicate alone, so it does not average to the mean reference rate. Skill here is agreement with the hydrological reference of Section 3; Osun's stage law within that reference is transferred from Kwara.

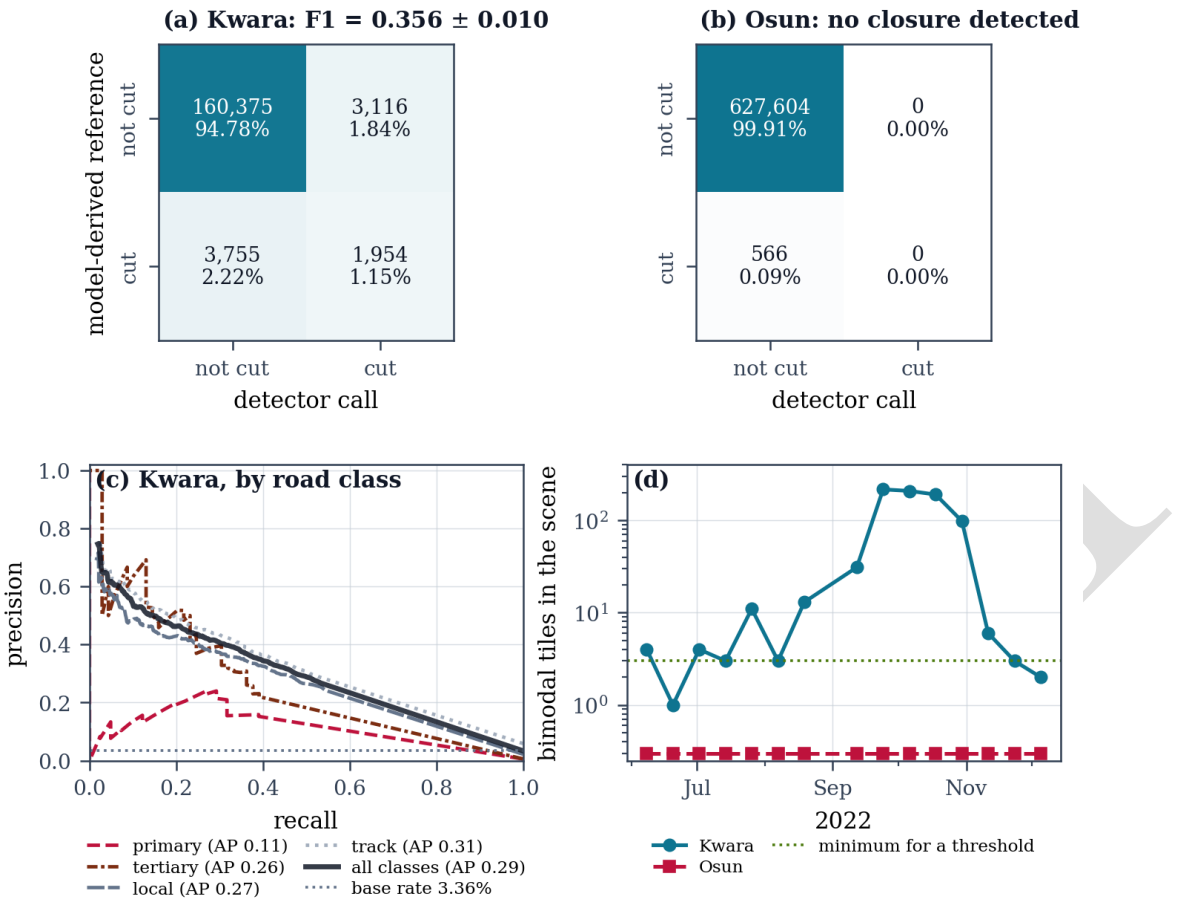

Figure 4. Segment-level agreement with the hydrological reference, and the diagnostic that governs it. (a) and (b) are the confusion tables on the baseline replicate, Kwara and Osun, in counts and as a percentage of all segment-days; Osun's detector column is empty. (c) is the precision-recall curve for Kwara by road class, obtained by sweeping the decision cut on the water fraction, with average precision per class and the 3.36 percent reference base rate marked. (d) is the count of bimodal tiles per scene for both areas on a log axis, against the minimum of three that a data-driven threshold requires: Kwara crosses it on 13 of 15 dates and Osun on none.

On the Osun basin the method returns no detection across all fifteen acquisitions. This is not allowed score. No tile of any scene passed the bimodality test, so no threshold was ever selected from the data. Panel (d) of Figure 4 shows the diagnostic directly: Kwara's bimodal tile count rises from 4 in June to 216 at the September peak, while Osun's is zero on every date.

Two things about that null need stating precisely. First, the conservative −20 dB fallback didmap a trace of water in Osun, between 0.00 and 0.02 km^{2}on seven of the fifteen dates, which is why the Osun water-extent agreement reported in Section 6 is a small number rather than an undefined one. What is exactly zero is the number of segments that reached the decision cut. Second, and more important, the null is a statement about the detector and not about the basin. The hydrological reference calls 566 Osun segment-days impassable over the same season, peaking at 177 segments on 18 October, on a catchment that received 1,045 mm of rain between June and November. The correct reading is that the method made no claim in Osun, not that Osun did not flood. Three mechanisms explain the blindness and they compound. At 76 percent canopy, most of the basin's surface is not visible to the sensor in the first place(Tsyganskaya et al. 2018). Ianthe built-up parts, double-bounce returns raise the bright mode and fill the

histogram valley that thresholding depends on(Mason et al. 2014). And a 12-day revisit, with a 24-day gaping this series, cannot see flash flooding on a basin of this size, which recedes between acquisitions. This is a method with a domain, not a method with an accuracy. Reporting pooled figure across both areas would produce a number describing neither.

5.2 Where skill goes within the area that works

Table 4 and Figure 5 report the stratification, and the gradients are steep.

By road class, F1 falls from 0.368 on tracks to 0.344 on local roads, 0.233 on tertiary and 0.097on primary. Kwara's trunk roads carry no reference positive at all on the baseline replicate and only 9across all twenty, and its secondary roads carry 154 positives across twenty replicates against zero detections, so their recall is 0.000 and their precision is undefined. Both are reported that way rather than as a number computed on a handful of cases.

The ordering runs opposite to operational need, and the temptation is to conclude that the method is least reliable on the roads that matter most. That conclusion is not available from this design. The reference rate falls monotonically with road class, from 5.94 percent of segment-days on tracks to 0.04 percent on trunk roads, because embankment heights rise with class in the reference model and so do the depths at which a segment is called impassable, both set out in Section 3. F1 is sensitive to the base rate, so a monotone fall in F1 across stratified whose base rate falls monotonically tracks the base rate at least as much as the detector. What the road-class rows do establish is narrower: the detector found almost nothing on the higher classes, whether because those roads genuinely flood less, or because a formed embankment keeps the carriageway above the water the radar sees beside it, or because the reference model's ladders say so. This design cannot separate the three.

Table 4. Detection skill disaggregated by road class, tree cover, built-up cover, segment length, height above nearest drainage and the presence of mapped water within 30 m of the centerline.

Stratum	Segments	Ref. rate (%)	Precision	Recall	F1	F1 replicate SD
Kwara, Niger floodplain						
Road class						
trunk	67	0.04	–	0.000	0.000	–
primary	496	0.40	0.092	0.102	0.097	0.0682
secondary	52	0.99	–	0.000	0.000	–
tertiary	767	0.74	0.457	0.156	0.233	0.0633
local	5,694	2.12	0.338	0.350	0.344	0.0151
track	4,204	5.94	0.404	0.338	0.368	0.0098
Tree cover						
0-10%	8,105	3.66	0.392	0.339	0.364	0.0109

Stratum	Segments	Ref. rate (%)	Precision	Recall	F1	F1 replicate SD
10-30%	1,705	2.46	0.331	0.353	0.341	0.0162
30-60%	923	2.46	0.324	0.345	0.334	0.0311
>60%	547	3.12	0.364	0.254	0.299	0.0247
Built-up cover						
<5%	8,835	4.03	0.383	0.348	0.365	0.0094
5-20%	1,044	1.42	0.292	0.201	0.238	0.0352
>20%	1,401	0.51	0.152	0.089	0.112	0.0425
Segment length						
<150 m	750	3.43	0.380	0.331	0.354	0.0290
150-300 m	904	4.05	0.399	0.376	0.388	0.0189
300-450 m	8,941	3.15	0.377	0.330	0.352	0.0110
>450 m	685	5.06	0.363	0.363	0.363	0.0197
HAND						
<2 m	3,913	8.88	0.412	0.352	0.380	0.0099
2-5 m	2,513	1.23	0.131	0.174	0.150	0.0221
5-10 m	1,249	0.00	0.000	–	–	–
>10 m	3,605	0.00	–	–	–	–
Mapped water within 30 m						
within 30 m	473	0.12	0.008	0.701	0.016	0.0095
beyond 30 m	10,807	3.50	0.438	0.337	0.381	0.0105
Osun, Osun River basin						
Road class						
trunk	401	0.00	–	0.000	0.000	–
primary	868	0.01	–	0.000	0.000	–
secondary	1,182	0.07	–	0.000	0.000	0.0000

Stratum	Segments	Ref. rate (%)	Precision	Recall	F1	F1 replicate SD
tertiary	1,985	0.04	–	0.000	0.000	0.0000
local	35,959	0.09	–	0.000	0.000	0.0000
track	1,483	0.28	–	0.000	0.000	0.0000
Tree cover						
0-10%	17,932	0.07	–	0.000	0.000	0.0000
10-30%	7,659	0.12	–	0.000	0.000	0.0000
30-60%	6,182	0.16	–	0.000	0.000	0.0000
>60%	10,105	0.08	–	0.000	0.000	0.0000
Built-up cover						
<5%	11,317	0.11	–	0.000	0.000	0.0000
5-20%	3,746	0.12	–	0.000	0.000	0.0000
>20%	26,815	0.08	–	0.000	0.000	0.0000
Segment length						
<150 m	13,918	0.13	–	0.000	0.000	0.0000
150-300 m	9,406	0.09	–	0.000	0.000	0.0000
300-450 m	16,810	0.07	–	0.000	0.000	0.0000
>450 m	1,744	0.09	–	0.000	0.000	0.0000
HAND						
<2 m	1,588	2.21	–	0.000	0.000	0.0000
2-5 m	3,796	0.11	–	0.000	0.000	0.0000
5-10 m	7,587	0.00	–	–	–	–
>10 m	28,907	0.00	–	–	–	–
Mapped water within 30 m						
within 30 m	221	0.01	–	0.000	0.000	–
beyond 30 m	41,657	0.09	–	0.000	0.000	0.0000

Note: Reference rate is the share of segment-days the hydrological reference calls impassable in that stratum; precision and recall are not comparable across strata without it. F1 is base-rate sensitive, and the reference rate falls with road class by construction, because the reference model's embankment heights and pass ability depths rise with class, so the road-class gradient has to be read against it. Metrics come from confusion cells pooled over the 20 reference replicates, and every quantity describing the positive class is blanked together where recall is undefined. Cover fractions are ESA World Cover 2021 within 30 m of the centerline; class 50 includes road surfaces, so the built-up stratified is partly the road itself. The water stratum is every segment whose 30 m context holds a mapped waterway or that carries an OSM bridge tag; only 8 of the 473 in Kwara and 13 of the 221 in Osun carry the tag. Osun's stage law within the reference is transferred from Kwara.

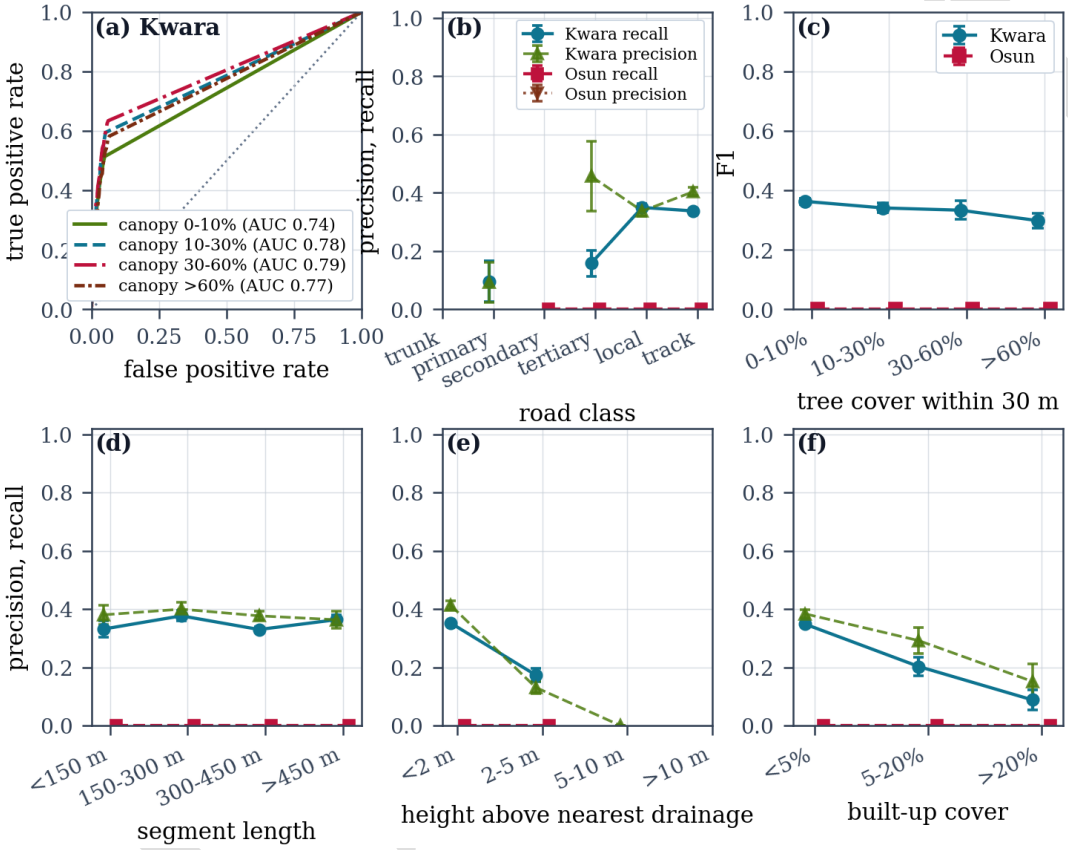

Figure 5. Skill by stratum, both study areas. (a) is the only receiver operating panel in the set: Kwara's curves by canopy fraction, obtained by sweeping the decision cut, with the area under each curve. The remaining panels plot precision and recall at the operating point with their replicate standard deviations, against (b) road class, (d) segment length, (e) height above nearest drainage and (f) built-up cover; (c) plots F1 against tree cover. Osun sits on zero throughout, which is the flat lower series in every panel. There is no panel for the mapped-water stratum, which Table 4 carries instead. The canopy and built-up gradients are the detector's own. By canopy, F1 falls from 0.364 to 0.299 and recall from 0.339 to 0.254 between the low and high strata. By built-up fraction, F1 falls from0.365 to 0.112, a drop of 25.4 percentage points, and recall from 0.348 to 0.089, a drop of 26.0percentage points. The comparable published figure, a roughly 15-point urban penalty inGiustarini et al. (2013), is an accuracy and recall penalty, so the recall drop is the like-for-like comparison and it is the larger of the two. Our built-up stratum is more mixed than a dedicated urban study's would be, and the ESA World Cover caveat noted with Table 2 applies to it.

Above 2 m of height above nearest drainage, skill collapses, and the terrain refinement is the obvious reason. The reading has a limit, though. Between 2 and 5 m the detector still fires on 1.6percent of segment-days and precision is 0.131, so what happens at elevation is not that detections are suppressed but that the ones that survive are mostly wrong. Above 5 m the reference calls nothing impassable and the comparison stops being informative in either direction.

5.3 Roads next to water

The sharpest failure in this study is not statistical but geometric.

The stratum is every segment whose 30 m context holds a mapped watercourse together with every segment carrying an OpenStreetMap bridge tag, 473 of them in Kwara. Across it precision is 0.008against 0.438 elsewhere, while recall over the same stratum is 0.701 against 0.337. Pooled over the twenty reference replicates the stratum holds 117 true positives against 14,043 false ones and 50 false negatives. The detector is not missing these segments; it is calling roughly one in ten of their segment-days cut, against a reference rate of 0.12 percent. Almost every positive is false, and the reason is that the water the sensor sees is really there and is there on every date. The two rates come from opposite sides of the same geometry. The reference treats a segment with a mapped crossing as bridged and holds it passable until the water reaches an assumed 3.5 m deck freeboard, which is why its positive rate in this stratum is 0.12 percent while the detector calls9.98 percent of the same segment-days cut. A different freeboard would move the printed precision of 0.008 up or down without touching the mechanism, which is that water inside the 30 m context ofa road is present flood or no flood. The stratum is not what the phrase "bridge problem" would suggest. Of the 473 Kwara segments in it, 8 carry an OpenStreetMap bridge tag, and of the 221 in Osun, 13 do. For roughly 98 percent of the stratum the segment is not a deck over a river at all; it is a road whose 30 m context happens to contain a mapped watercourse, which in practice usually means a road running beside a stream. The failure mechanism is proximity to mapped water, and the small tagged-bridge subset isa special case of it rather than the whole of it. This is not fixable by better thresholding, because the water is really there. It requires masking segments with mapped water in their context out of the closure inference, or treating them under a separate rule, and any study that consumes a flood extent product to infer road closures without doing so will report a large number of confident false closures. Whether those segments are also the ones whose closure would matter most to the network is a question this paper does not answer: no connectivity, betweenness or cut analysis was performed, and Section 4.5 explains why the published closure set does not support one.

6 Robustness Checks

6.1 Sweeps

Table 5 and Figure 6 report the sweeps.

The backscatter threshold matters most. Across−22to−10 dB, F1 ranges from 0.130 to 0.365,so the headline figure sits at the top of a wide range and a poorly chosen threshold loses most of the available skill. The road buffer barely matters. Across 10 to 120 m, F1 ranges from 0.343 to 0.376. That is worth knowing because buffer width is the parameter practitioners most often argue about, and here it is close to irrelevant.

The decision cut also matters less than its prominence suggests. The parameter is the water fraction cut, the share of buffered cells that must classify as water before a segment is called cut, and it is not a calibrated probability. Swept from 0.05 to 0.80 it moves F1 between 0.172

and0.370, with the maximum of 0.370 at a cut of 0.10 against 0.356 at this paper's operating value of0.25. Panel (c) of Figure 6 shows why the curve is flat over its useful range: precision and recall cross close to the operating point, so tightening the cut buys precision at almost exactly the recall it costs.

Table 5. Sensitivity of detection skill to the three parameters the analyst chooses: the backscatter threshold, the buffer between water and centerline, and the fraction of buffered cells required before a segment is called cut.

Parameter	Swept over	Operating value	F1 at operating	F1 range	Best value	Best F1
Kwara, Niger floodplain						
VV threshold	−22to−10 dB	split-based,−20.0to −12.4dB	0.356	0.130–0.365	−11.0dB	0.365
Buffer distance	10 to 120 m	30 m	0.356	0.343–0.376	120 m	0.376
Decision cut	0.05 to 0.80	0.25	0.356	0.172–0.370	0.10	0.370
Osun, Osun River basin						
VV threshold	−22to−10 dB	split-based,−20.0to −20.0dB	0.000	0.000–0.000	−22.0dB	0.000
Buffer distance	10 to 120 m	30 m	0.000	0.000–0.000	10 m	0.000
Decision cut	0.05 to 0.80	0.25	0.000	0.000–0.000	0.05	0.000

*Note:*F1 at operating is the mean over reference replicates at this paper's operating point. F1 range is the full span of the sweep, so a narrow range means the result does not depend on the choice. Each parameter is swept with the other two held at their operating values.

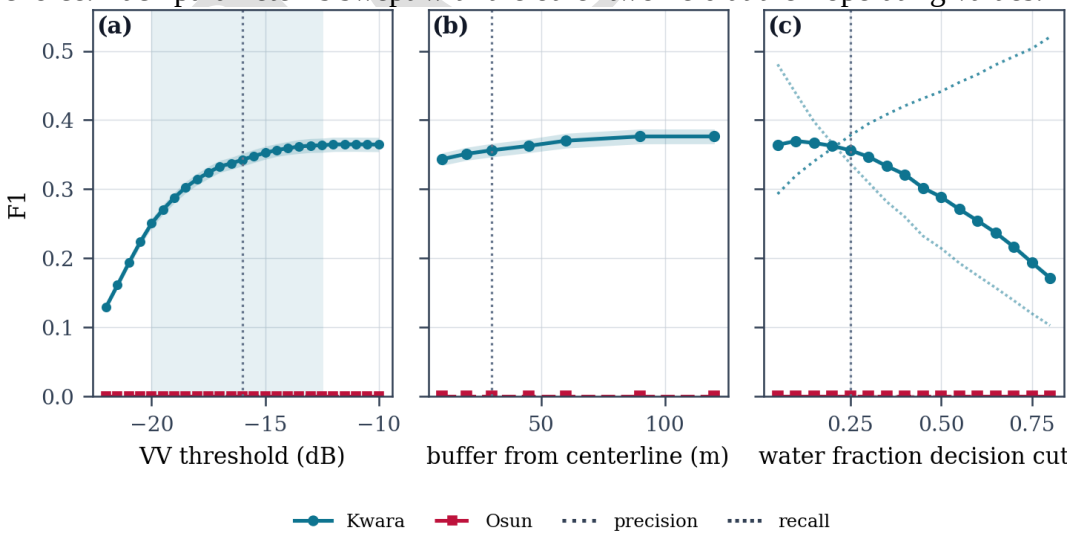

Figure 6. Sweeps over the three parameters the analyst chooses, F1 against each, with the operating value marked by a dotted vertical line and the replicate spread shaded. (a) the VV

backscatter threshold, with the shaded band giving the range the split-based selection actually chose across the season. (b) the buffer from the centerline. (c) the water fraction decision cut, with precision and recall added as light dotted lines to show where they cross. Kwara is the upper series in each panel and Osun the flat lower one.

6.2 Replicate variance against block variance

The two are a factor of 18.8 apart. Holding the area fixed and varying the reference across its twenty replicates gives a standard deviation on F1 of 0.0097. Holding the reference replicate fixed and varying across six spatial blocks within the Kwara study area gives 0.1834.

The comparison is between two sources of uncertainty that a reader might otherwise assume are of the same size, and they are not: which part of a basin is examined moves the answer far more than which replicate of the reference it is scored against. It follows that a skill figure should be reported with the surface conditions it was measured under, and that published accuracy figures for flood detection are better read as properties of a validation site than of a method.

What does not follow is a number for how much skill varies between basins. The block component is within-area spread, not a between-area one, and with two study areas there is no between-area variance to estimate. Our two areas do differ by more than the difference between modest performance and none, which points the same way, but it is one contrast and not a variance component.

6.3 Benchmarks

Eighteen quantities are recorded against published values in Table 6. Seven of them have a published numeric range to sit inside or outside, and four of those fall inside it and three below. Four more are checked against a published claim that states a direction and attaches no level to it, so the most that can be said is that the direction holds: the recall losses across built-up land and across closed canopy in Kwara, and the mean clear-sky fraction in each area. Placing any of the four inside an interval would be reporting an interval the source does not give. The remaining seven are marked not comparable, and they divide into three kinds. Four are road-pass ability scores, for which no published envelope exists at all. One is Osun’s automatic water threshold, which was never selected. Two are Osun degradation statistics, differences between detection rates in an area where the detector made no detection, so they are zero by arithmetic rather than by measurement and grading them in either direction would read a null as a result.

The Kwara water-extent critical success index of 0.404 sits inside the 0.30 to 0.90 envelope Bernhofen et al. (2018)reports for flood extent methods generally, but below the0.45 to 0.70 that source gives for Nigerian floodplains specifically, which is the tighter and more relevant of the two. The comparison also needs a caveat the number does not carry on its face. The stage law underlying the reference is calibrated to the radar’s own total flooded *area*, so any agreement on area is circular. What is not circular is *where* that area falls, which the elevation model decides alone, and the 0.404 is a cell-level agreement rather than anareal one. That is the part of the check that carries information. Three quantities fall below their envelope. Osun’s water-extent index of 0.039 is one of them, and it is a score rather than a null: the fallback threshold classified a trace of water and the comparison is real, it is simply very poor. The other two are the modeled seasonal stage ranges,1.33 m in Kwara and 1.10 m in Osun, both smaller than the 3.91 m floodplain vertical error Hawker et al. (2019)reports for elevation models of this generation. That failure was acted on rather than noted. The vertical uncertainty was carried into the reference as a per-segment error varied across replicates, and swept from zero to the full published value. At

the operating value of 1.2 m, F1 is 0.358; at the full 3.91 m it is 0.296, a fall of 0.062, or 17percent of the operating value, and with no elevation error at all it would be 0.405. The loss falls almost entirely on recall while precision barely moves, which is what a vertical error should do: it scatters segments across the pass ability threshold rather than relocating the flood. The sweep also moves the reference positive rate from 3.09 percent to 4.82 percent of segment-days, so the reference model's base rate is itself a function of an assumption about elevation error, and the absolute skill figures should be read with that in mind.

The road pass ability rows carry no verdict at all, and that absence is the paper's point restated. No published envelope exists for road closure detection skill, so there is nothing to compare against. This paper is an attempt to start one.

Table 6. Benchmarking the quantities this study produces against published values for comparable systems.

Quantity	This study	Published range	Source	Verdict
Water-extent CSI at the seasonal peak, Kwara	0.404	0.30–0.90 overall; 0.45–0.70 for Nigerian floodplains	Bernhofen et al. 2018	inside
Automatic VV water threshold, Kwara	−13.1 dB (−14.8 to−12.4)	−22to−12dB for C-band VV open water	Twele et al. 2016; Martinis et al. 2009	inside
Dry-season median VV over land, Kwara	−11.5 dB	−12to−5dB for vegetated and cropped land	Twele et al. 2016	inside
Water-extent CSI at the seasonal peak, Osun†	0.039	0.30–0.90 overall; 0.45–0.70 for Nigerian floodplains	Bernhofen et al. 2018	below
Automatic VV water threshold, Osun	never selected	−22to−12dB for C-band VV open water	Twele et al. 2016	not comparable
Dry-season median VV over land, Osun	−7.8 dB	−12to−5dB for vegetated and cropped land	Twele et al. 2016	inside
Recall loss, open to built-up, Kwara	26.0 pp	about 15 pp (rural 90% to urban 75%)	Giustarini et al. 2013	direction only
Recall loss, open to closed canopy, Kwara	8.5 pp	a detection floor, not a small percentage penalty	Chapman et al. 2015; Tsyganskaya et al. 2018	direction only
Recall loss, open to built-up, Osun†	0.0 pp	about 15 pp (rural 90% to urban 75%)	Giustarini et al. 2013	not comparable
Recall loss, open to closed canopy, Osun†	0.0 pp	a detection floor, not a small percentage penalty	Chapman et al. 2015; Tsyganskaya et al. 2018	not comparable

Quantity	This study	Published range	Source	Verdict
Seasonal stage range against DEM error, Kwara	1.33 m	SRTM vertical RMSE 3.91 m on floodplains	Hawker et al. 2019	below
Seasonal stage range against DEM error, Osun†	1.10 m	SRTM vertical RMSE 3.91 m on floodplains	Hawker et al. 2019	below
Sentinel-2 mean clear-sky fraction, Kwara	33.1%	cloud is the binding constraint on optical flood mapping in the wet tropics	Wilson and Jetz 2016; Landuyt et al. 2019	direction only
Sentinel-2 mean clear-sky fraction, Osun	19.5%	cloud is the binding constraint on optical flood mapping in the wet tropics	Wilson and Jetz 2016; Landuyt et al. 2019	direction only
Road-passability CSI, Kwara	0.217	no published envelope exists	–	not comparable
Road-passability F1, Kwara	0.356	no published envelope exists	–	not comparable
Road-passability CSI, Osun†	0.000	no published envelope exists	–	not comparable
Road-passability F1, Osun†	0.000	no published envelope exists	–	not comparable

*Note:*Every published range in this table was read from the source named beside it rather than from a secondary summary, and where a quantity falls outside its range the text states what was done about it. A verdict of inside or below is available only where the source states a numeric range; where it states a direction and attaches no level, the check is directional and the verdict is direction only. The published literature validates flood maps against water extent; this study validates against road passability, so the comparison is indicative rather than like for like, and that is the point the paper makes.†marks a row whose absolute level rests on a transferred calibration: Osun's stage law is taken from Kwara because the Osun radar detects no standing water all season and cannot calibrate its own reference. Water-extent agreement is cell-level rather than areal, for the reason given in the text. Degradation rows for an area in which the detector made no detection at all are marked not comparable, because a difference between two zero detection rates is zero by arithmetic and is not a measured penalty. Road-passability rows rest on the hydrological reference of Section 3.

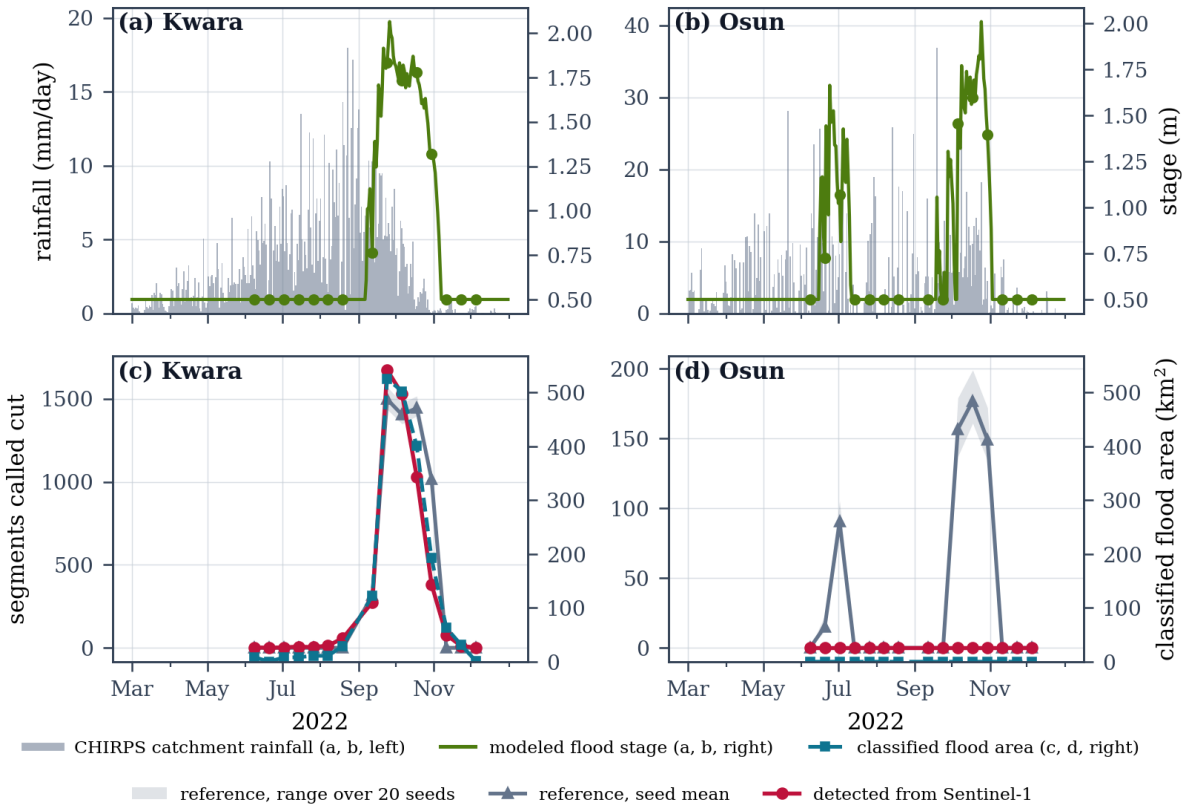

Figure 7. The 2022 season in both areas. (a) and (b) give CHIRPS daily catchment rainfall on the left axis and the modeled flood stage that drives the reference on the right, for Kwara and Osun; the markers on the stage line are the Sentinel-1 acquisition dates. (c) and (d) give the number of segments called cut on the left axis, by the reference as a replicate mean with its range over 20 replicates shaded and by the Sentinel-1 detector, with the classified flood area on the right. Kwara's detected and reference series track each other through the September peak. Osun's reference rises twice while its detected series stays at zero all season, which is the null this paper reports. Figure 7 places both areas on the same time base and makes the shape of the season visible. In Kwara the detector and the reference rise and fall together through the September peak, which is the agreement the F1 of 0.356 summarizes. In Osun the reference rises twice, in early July and again through October, while the detected series never leaves zero. The gap between those towlines is the whole of the Osun result.

7 Discussion

7.1 What follows for anyone using flood products on roads

The audience here is anyone building a road closure layer from satellite flood mapping, which increasingly includes national emergency agencies, and in Nigeria the National Emergency Management Agency and the state emergency management agencies of Kwara and Osun.

First, mask segments with mapped water inside their buffer out of the closure inference, or handle them under a separate rule. This is a one-line fix and it removes the study's largest source of confident false positives. Note that it is a broader class of road than the word bridge suggests:473 Kwara segments qualify and 8 of them carry a bridge tag, so the rule has to key on proximity to mapped water rather than on bridge tagging, which is sparse in both areas.

Second, do not apply a single accuracy figure across a region. Within Kwara alone, the spread of F1across six spatial blocks is 18.8 times its spread across twenty reference replicates. A skill

figure should be reported with the surface conditions it was measured under.

Third, publish probabilities rather than binary maps. Our Kwara closure set carries calibrated confidence with a Brier skill score of 0.153, and a downstream user can threshold it for their own tolerance for false alarms. The Osun set has a Brier skill score of 0.000, which is itself the honest signal that there is nothing in it to threshold. A binary map hides both facts.

Fourth, treat canopy as a domain boundary rather than as a difficulty. In the Osun basin the correct output was no claim, and a method that had forced a threshold anyway would have produced a confident and entirely spurious map. The Osun null measures the detector's blindness under canopy, not the incidence of flooding in the Osun basin, where 1,045 mm of rain fell between June and November and the hydrological reference calls 566 segment-days impassable.

Fifth, do not treat a published segment table as a network without checking that it is one. Ours was not, and said it was; Section 4.5 gives the measured consequence and the repair.

7.2 Transferability

Whether any of this carry to a third basin is settled by the paper's own numbers rather than by an argument. The chain needs nothing that is not free and nothing that is not global, so it can be run wherever Sentinel-1 acquires, which for the operational land modes is most of the populated world. What it returns there is the open question. Inside the Kwara study area alone, F1 moves 18.8 times more across six spatial blocks than across twenty realizations of the reference, and across the two basins it moves between a usable screening tool and no claim at all. A figure that unstable within one basin of 6,305 $km^2$is a description of a surface, and carrying the 0.356elsewhere would mean carrying the Niger floodplain with it.

The design does travel, even where the numbers do not. Any study reporting flood detection skill for roads can and should report it by canopy, by built-up fraction and by proximity to mapped water, because those are the physical mechanisms and they are all available from open data. Road class is worth reporting too, but only alongside the reference rate in that class, because otherwise the gradients uninterpretable. That costs nothing and it converts a pooled number into a statement about where the method can be believed.

7.3 Limitations

Osun's stage law is transferred from Kwara rather than fitted, for the reason given in Section 3, so the absolute level of Osun's reference follows from Kwara's fit; every Osun number derived from it, including the water-extent index of 0.039 and the seasonal stage range of 1.10 m, is marked in Table 6. Three parameters of the reference model are set rather than fitted here: embankment height and pass ability depth, both assigned by road class, and the 3.5 m deck freeboard at watercourse crossings.

The revisit is 12 days, with one 24-day gap between 19 August and 12 September 2022 that falls on the rising limb of the Niger flood, so this study speaks to sustained inundation only, and it uses ascending relative orbit 103 alone, which holds incidence angle and look direction constant and leaves their effects unmeasured.

8 Conclusion

We asked what free satellite radar is worth for inferring road closures, using Sentinel-1 across two contrasting Nigerian basins in the 2022 season with 53,158 road segments.

The answer is that the method has a domain rather than an accuracy. On the open Niger floodplain it reaches an F1 of 0.356 against the hydrological reference, with an area under the

curve of 0.746, enough for screening. On the canopy-heavy Osun basin it makes no claim at all across fifteen acquisitions, because no scene is bimodal and no threshold is ever selected. That null describes the detector and not the basin: the same season put 1,045 mm of rain on the Osun catchment and the hydrological reference calls 566 segment-days impassable there. Reporting an average of the two areas would describe neither.

Within the working domain, skill is stratified by surface condition. F1 falls from 0.364 to 0.299across canopy and from 0.365 to 0.112 across built-up land, both declines being properties of the detector. Skill also falls across road class, from 0.368 on tracks to 0.097 on primary roads, but the reference rate falls with class as the embankment and pass ability thresholds rise, so this design cannot separate that ordering from a real effect.

The most actionable result concerns roads next to water. Across the 473 Kwara segments whose 30 context holds a mapped watercourse or that carry a bridge tag, precision is 0.008 against 0.438elsewhere while recall rises to 0.701: the detector fires on roughly one segment-day in ten there, against a reference rate of 0.12 percent. Only 8 of those segments carry the tag, so the mechanisms proximity to mapped water rather than a deck over a river. Any pipeline inferring closures from flood extent will generate confident false closures on that class of road unless it masks them or handles them separately. Whether those are also the network’s most important links is a question we did not measure and do not claim.

Within one basin, the spread of F1 across six spatial blocks is 18.8 times its spread across twenty reference replicates, which means published flood detection accuracies are better read as properties of their validation sites than of their methods. We therefore publish calibrated closure probabilities rather than a binary map, with the measured limits of the published file stated in its own metadata so that a study consuming it inherits the uncertainty rather than a claim of truth.

What would improve on this is ground observation. A modest record of which segments were actually impassable, on a few dates, in both basins, would convert every skill figure here from internal consistency into accuracy, and it is the one input that no amount of imagery can substitute for.

References

Adelekan, Ibidun O. 2010."Vulnerability of Poor Urban Coastal Communities to Flooding inLagos,Nigeria."*Environment and Urbanization*22 (2): 433–50.https://doi.org/10.1177/0956247810380141.

Agbola, Babatunde S., Owolabi Ajayi, Olalekan J. Taiwo, and Bolanle W. Wahab. 2012."TheAugust2011 Flood inIbadan,Nigeria: Anthropogenic Causes and Consequences."*International Journal of Disaster Risk Science*3 (4): 207–17.https://doi.org/10.1007/s13753-012-0021-3.

Alfieri, L., P. Burek, E. Dutra, et al. 2013."GloFAS— Global Ensemble Streamflow Forecasting and Flood Early Warning."*Hydrology and Earth System Sciences*17 (3): 1161–75.https://doi.org/10.5194/hess-17-1161-2013.

Arrighi, C., M. Pregnolato, R. J. Dawson, and F. Castelli. 2019."Preparedness Against Mobility Disruption by Floods."*Science of The Total Environment*654: 1010–22.https://doi.org/10.1016/j.scitotenv.2018.11.191.

Barrington-Leigh, Christopher, and Adam Millard-Ball. 2017."The World's User-Generated Road Map Is More Than 80% Complete."*PLOS ONE*12 (8): e0180698.https://doi.org/10.1371/journal.pone.0180698.

Bernhofen, Mark V., Charlie Whyman, Mark A. Trigg, et al. 2018."A First Collective Validation of Global Fluvial Flood Models for Major Floods inNigeriaandMozambique."*Environmental Research Letters*13 (10): 104007.https://doi.org/10.1088/1748-9326/aae014.

Camboim, Silvana Philippi, João Vitor Meza Bravo, and Claudia Robbi Sluter. 2015."An Investigation into the Completeness of, and the Updates to,OpenStreetMapData in a Heterogeneous Area inBrazil."*ISPRS International Journal of Geo-Information*4 (3): 1366–88.https://doi.org/10.3390/ijgi4031366.

Chapman, Bruce, Kyle McDonald, Masanobu Shimada, Ake Rosenqvist, Ronny Schroeder, and Laura Hess. 2015."Mapping Regional Inundation with SpaceborneL-BandSAR."*Remote Sensing*7 (5): 5440–70.https://doi.org/10.3390/rs70505440.

Chini, Marco, Renaud Hostache, Laura Giustarini, and Patrick Matgen. 2017."A Hierarchical Split-Based Approach for Parametric Thresholding ofSARImages: Flood Inundation as a Test Case."*IEEE Transactions on Geoscience and Remote Sensing*55 (12): 6975–88.https://doi.org/10.1109/TGRS.2017.2737664.

Cohen, Jacob. 1960."A Coefficient of Agreement for Nominal Scales."*Educational and Psychological Measurement*20 (1): 37–46.https://doi.org/10.1177/001316446002000104.

Congalton, Russell G. 1991."A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data."*Remote Sensing of Environment*37 (1): 35–46.https://doi.org/10.1016/0034-4257(91)90048-B.

DeVries, Ben, Chengquan Huang, John Armston, Wenli Huang, John W. Jones, and Megan W. Lang. 2020."Rapid and Robust Monitoring of Flood Events UsingSentinel-1andLandsatData on theGoogle Earth Engine."*Remote Sensing of Environment*240: 111664.https://doi.org/10.1016/j.rse.2020.111664.

Douglas, Ian, Kurshid Alam, Maryanne Maghenda, Yasmin McDonnell, Louise McLean, and Jack Campbell. 2008."Unjust Waters: Climate Change, Flooding and the Urban Poor inAfrica."*Environment and Urbanization*20 (1): 187–205.https://doi.org/10.1177/0956247808089156.

Drusch, M., U. Del Bello, S. Carlier, et al. 2012."Sentinel-2:ESA's Optical High-Resolution

Mission forGMESOperational Services."*Remote Sensing of Environment*120: 25–36.https://doi.org/10.1016/j.rse.2011.11.026.

Echendu, Adaku Jane. 2020."The Impact of Flooding onNigeria's Sustainable Development Goals (SDGs)."*Ecosystem Health and Sustainability*6 (1): 1791735.https://doi.org/10.1080/20964129.2020.1791735.

Farr, Tom G., Paul A. Rosen, Edward Caro, et al. 2007."TheShuttle Radar Topography Mission."*Reviews of Geophysics*45 (2).https://doi.org/10.1029/2005RG000183.

Feyisa, Gudina L., Henrik Meilby, Rasmus Fensholt, and Simon R. Proud. 2014."Automated Water Extraction Index: A New Technique for Surface Water Mapping UsingLandsatImagery."*Remote Sensing of Environment*140: 23–35.https://doi.org/10.1016/j.rse.2013.08.029.

Giustarini, Laura, Renaud Hostache, Patrick Matgen, Guy J.-P. Schumann, Paul D. Bates, and David C. Mason. 2013."A Change Detection Approach to Flood Mapping in Urban Areas UsingTerraSAR-X."*IEEE Transactions on Geoscience and Remote Sensing*51 (4): 2417–30.https://doi.org/10.1109/TGRS.2012.2210901.

Hawker, Laurence, Jeffrey Neal, and Paul Bates. 2019."Accuracy Assessment of theTanDEM-X90Digital Elevation Modelfor Selected Floodplain Sites."*Remote Sensing of Environment*232: 111319.https://doi.org/10.1016/j.rse.2019.111319.

Herfort, Benjamin, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. 2021."The Evolution of Humanitarian Mapping Within theOpenStreetMapCommunity."*Scientific Reports*11 (1): 3037.https://doi.org/10.1038/s41598-021-82404-z.

Kermanshah, Amirhassan, and Sybil Derrible. 2017."Robustness of Road Systems to Extreme Flooding: Using Elements ofGIS, Travel Demand, and Network Science."*Natural Hazards*86 (1): 151–64.https://doi.org/10.1007/s11069-016-2678-1.

Kittler, J., and J. Illingworth. 1986."Minimum Error Thresholding."*Pattern Recognition*19 (1): 41–47.https://doi.org/10.1016/0031-3203(86)90030-0.

Koks, E. E., J. Rozenberg, C. Zorn, et al. 2019."A Global Multi-Hazard Risk Analysis of Road and Railway Infrastructure Assets."*Nature Communications*10 (1): 2677.https://doi.org/10.1038/s41467-019-10442-3.

Landuyt, Lisa, Alexandra Van Wesemael, Guy J.-P. Schumann, Renaud Hostache, Niko E. C. Verhoest, and Frieke M. B. Van Coillie. 2019."Flood Mapping Based on Synthetic Aperture Radar: An Assessment of Established Approaches."*IEEE Transactions on Geoscience and Remote Sensing*57 (2): 722–39.https://doi.org/10.1109/TGRS.2018.2860054.

Lee, Jong-Sen. 1980."Digital Image Enhancement and Noise Filtering by Use of Local Statistics."*IEEE Transactions on Pattern Analysis and Machine Intelligence*PAMI-2 (2): 165–68.https://doi.org/10.1109/TPAMI.1980.4766994.

Martinis, S., A. Twele, and S. Voigt. 2009."Towards Operational Near Real-Time Flood Detection Using a Split-Based Automatic Thresholding Procedure on High ResolutionTerraSAR-XData."*Natural Hazards and Earth System Sciences*9 (2): 303–14.https://doi.org/10.5194/nhess-9-303-2009.

Mason, D. C., L. Giustarini, J. Garcia-Pintado, and H. L. Cloke. 2014."Detection of Flooded Urban Areas in High ResolutionSynthetic Aperture RadarImages Using Double Scattering."*International Journal of Applied Earth Observation and Geoinformation*28: 150–59.https://doi.org/10.1016/j.jag.2013.12.002.

Mason, D. C., R. Speck, B. Devereux, G. J.-P. Schumann, J. C. Neal, and P. D. Bates. 2010."Flood Detection in Urban Areas UsingTerraSAR-X."*IEEE Transactions on Geoscience and Remote Sensing*48 (2): 882–94.https://doi.org/10.1109/TGRS.2009.2029236.

McFeeters, S. K. 1996."The Use of theNormalized Difference Water Index(NDWI) in the Delineation of Open Water Features."*International Journal of Remote Sensing*17 (7): 1425–32.https://doi.org/10.1080/01431169608948714.

Nkeki, Felix Ndidi, Philip John Henah, and Vincent Nduka Ojeh. 2013."Geospatial Techniques for the Assessment and Analysis of Flood Risk Along theNiger-BenueBasin inNigeria."*Journal of Geographic Information System*5 (2): 123–35.https://doi.org/10.4236/jgis.2013.52013.

Nkwunonwo, U. C., M. Whitworth, and B. Baily. 2020."A Review of the Current Status of Flood Modelling for Urban Flood Risk Management in the Developing Countries."*Scientific African*7: e00269.https://doi.org/10.1016/j.sciaf.2020.e00269.

Nobre, A. D., L. A. Cuartas, M. Hodnett, et al. 2011."HeightAbovetheNearest Drainage— a Hydrologically Relevant New Terrain Model."*Journal of Hydrology*404 (1-2): 13–29.https://doi.org/10.1016/j.jhydrol.2011.03.051.

Nobre, Antonio Donato, Luz Adriana Cuartas, Marcos Rodrigo Momo, Dirceu Luís Severo, Adilson Pinheiro, and Carlos Afonso Nobre. 2016."HANDContour: A New Proxy Predictor of Inundation Extent."*Hydrological Processes*30 (2): 320–33.https://doi.org/10.1002/hyp.10581.

Olofsson, Pontus, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. 2014."Good Practices for Estimating Area and Assessing Accuracy of Land Change."*Remote Sensing of Environment*148: 42–57.https://doi.org/10.1016/j.rse.2014.02.015.

Otsu, Nobuyuki. 1979."A Threshold Selection Method from Gray-Level Histograms."*IEEE Transactions on Systems, Man, and Cybernetics*9 (1): 62–66.https://doi.org/10.1109/TSMC.1979.4310076.

Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S. Belward. 2016."High-Resolution Mapping of Global Surface Water and Its Long-Term Changes."*Nature*540 (7633): 418–22.https://doi.org/10.1038/nature20584.

Pontius, Robert Gilmore, and Marco Millones. 2011."Death toKappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment."*International Journal of Remote Sensing*32 (15): 4407–29.https://doi.org/10.1080/01431161.2011.552923.

Pregnolato, Maria, Alistair Ford, Craig Robson, Vassilis Glenis, Stuart Barr, and Richard Dawson. 2016."Assessing Urban Strategies for Reducing the Impacts of Extreme Weather on Infrastructure Networks."*Royal Society Open Science*3 (5): 160023.https://doi.org/10.1098/rsos.160023.

Pregnolato, Maria, Alistair Ford, Sean M. Wilkinson, and Richard J. Dawson. 2017."The Impact of Flooding on Road Transport: A Depth-Disruption Function."*Transportation Research Part D: Transport and Environment*55: 67–81.https://doi.org/10.1016/j.trd.2017.06.020.

Rennó, Camilo Daleles, Antonio Donato Nobre, Luz Adriana Cuartas, et al. 2008."HAND, a New Terrain Descriptor UsingSRTM-DEM: Mapping Terra-Firme Rainforest Environments inAmazonia."*Remote Sensing of Environment*112 (9): 3469–81.https://doi.org/10.1016/j.rse.2008.03.018.

Shen, Xinyi, Emmanouil N. Anagnostou, George H. Allen, G. Robert Brakenridge, and Albert J. Kettner. 2019."Near-Real-Time Non-Obstructed Flood Inundation Mapping Using Synthetic Aperture Radar."*Remote Sensing of Environment*221: 302–15.https://doi.org/10.1016/j.rse.2018.11.008.

Tellman, B., J. A. Sullivan, C. Kuhn, et al. 2021."Satellite Imaging Reveals Increased Proportion of Population Exposed to Floods."*Nature*596 (7870): 80–86.https://doi.org/10.1038/s41586-021-03695-w.

Trigg, M. A., C. E. Birch, J. C. Neal, et al. 2016."The Credibility Challenge for Global Fluvial Flood Risk Analysis."*Environmental Research Letters*11 (9): 094014.https://doi.org/10.1088/1748-9326/11/9/094014.

Tsyganskaya, Viktoriya, Sandro Martinis, Philip Marzahn, and Ralf Ludwig. 2018."SAR-Based Detection of Flooded Vegetation — a Review of Characteristics and Approaches."*International Journal of Remote Sensing*39 (8): 2255–93.https://doi.org/10.1080/01431161.2017.1420938.

Twele, André, Wenxi Cao, Simon Plank, and Sandro Martinis. 2016."Sentinel-1-Based Flood Mapping: A Fully Automated Processing Chain."*International Journal of Remote Sensing*37 (13): 2990–3004.https://doi.org/10.1080/01431161.2016.1192304.

Wilson, Adam M., and Walter Jetz. 2016."Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions."*PLOS Biology*14 (3): e1002415.https://doi.org/10.1371/journal.pbio.1002415.

Wing, Oliver E. J., Paul D. Bates, Christopher C. Sampson, Andrew M. Smith, Kris A. Johnson, and Tyler A. Erickson. 2017."Validation of a 30 m Resolution Flood Hazard Model of the ConterminousUnited States."*Water Resources Research*53 (9): 7968–86.https://doi.org/10.1002/2017WR020917.

Xu, Hanqiu. 2006."Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery."*International Journal of Remote Sensing*27 (14): 3025–33.https://doi.org/10.1080/01431160600589179.

Yamazaki, Dai, Daiki Ikeshima, Ryunosuke Tawatari, et al. 2017."A High-Accuracy Map of Global Terrain Elevations."*Geophysical Research Letters*44 (11): 5844–53.https://doi.org/10.1002/2017GL072874.

Yin, Jie, Dapeng Yu, Zhane Yin, Min Liu, and Qing He. 2016."Evaluating the Impact and Risk of Pluvial Flash Flood on Intra-Urban Road Network: A Case Study in the City Center ofShanghai,China."*Journal of Hydrology*537: 138–45.https://doi.org/10.1016/j.jhydrol.2016.03.037.