Case study 1: Lokoja and Idah, Nigeria

How well do Geovatra Baseline's flood zones match floods that were actually observed?

> **Draft for review, 2026-11-13.** Every number comes from `case_study_nigeria.json`, made by `manage.py case_study_nigeria`. The scoring rules were fixed before any score was computed, and every score is reported, including those that look worse.

Findings

- **Our scoring reproduces the published study.** Scored by our code, the study's own JRC maps get CSI 0.777 at Lokoja and 0.691 at Idah; the paper reports 0.78 and 0.7. So our numbers can be set beside theirs.
- **Baseline's flood-prone land matches the 2012 flood about as well as the best global model in the study:** CSI 0.78 at Lokoja (good: above 0.7) and 0.694 at Idah (moderate), against 0.78 and 0.7 for the best published model, and 0.73 and 0.53 for a locally calibrated regional model. It caught 93 % and 99 % of the observed flood.
- **That agreement comes almost entirely from the flood model inside the zones.** The JRC v2 1-in-100-year map alone scores 0.779 and 0.694. The satellite water records and terrain don't raise the agreement score; what they add is the ranking below.
- **The zones rank flood likelihood in the right order, at both sites and for both floods.** Share of each zone that flooded in 2012: A 96.6 / 97.6 %, B 78.0 / 65.8 %, C 15.9 / 2.7 %, no zone 0.9 / 0.9 % (Lokoja / Idah). In the fully independent 2022 flood: A 61.1 / 44.5 %, B 23.1 / 16.5 %, C 3.4 / 0.5 %, no zone 0.1 / 0 %.
- **Idah is weaker, as in the published study.** On its flat, wide floodplain the flood-prone land is +42 % larger than the 2012 flood (Lokoja: +12 %). Part of this is expected: the zones show a 1-in-100-year flood and 2012 was roughly a 1-in-50-year flood.
- **Zone C is right to be left out of "flood-prone".** Adding it lowers CSI to 0.625 and 0.575: it adds much more land that didn't flood than land that did.
- **Zone A alone is not the flood.** It caught only 23 % and 14 % of the 2012 flood. Zone A means "strong evidence", not "the area that floods".
- **Against the 2022 flood seen by radar, the zones contained 90 % and 96 % of it, but CSI is low (0.311 and 0.195)**: radar saw 502 and 672 km² of flood water, much less than the 1-in-100-year zones. Radar sees water only on the days it passes, and not under trees or in towns.

Method

Areas and references

The analysis areas of Bernhofen et al. (2018): Lokoja (4,154 km², the Niger-Benue confluence, narrow confined floodplains) and Idah (4,861 km², a flat, wide floodplain). Baseline (workflow baseline/1.0.0, unchanged) was run on exactly these areas.

- **2012 flood**: the maximum flood extent from MODIS (Dartmouth Flood Observatory, 15 Sept-29 Oct 2012, merged), as used by the study; 1,352 km² at Lokoja and 2,408 km² at Idah. From the study's published inputs (Research Data Leeds, doi:10.5518/340, CC BY 4.0).
- **October 2022 flood**: flood water seen by Sentinel-1 radar (dark in at least one scene of October 2022 and not in most dry-season scenes of February-March 2023). No Baseline input covers 2022, so this test is fully independent.

Scores (as in the study)

CSI (critical success index) = hits / (hits + misses + false alarms), from 0 to 1; above 0.7 is usually called good and below 0.5 poor. **Hit rate** (recall) = share of the observed flood that the zones contain. **Precision** = share of the zones that flooded. **Bias** = zone area / observed area - 1 (positive: larger than the flood). Scored on the study's 1/1200° (about 90 m) grid, inside the analysis areas; our 30 m zones were brought to that grid by the most common value per cell.

Rules fixed before scoring

Primary score: flood-prone zones A + B plus permanent water against the 2012 flood (a river channel is under water in a flood, and the published models include channels). Also reported, whatever they show: A + B, A, A + B + C, the JRC v2 model alone, terrain alone, and the zones against the 2022 radar flood.

A known overlap

Zone A uses water seen by satellites in 1984-2020 (JRC Global Surface Water), which includes 2012. Zone A is therefore not fully independent of the 2012 reference, and its high share flooded in 2012 is partly circular. The JRC v2 flood model, the terrain and the 2022 test don't have this overlap, and the ranking of the zones holds in 2022 too.

Results

Against the 2012 flood

Map	Lokoja CSI	hit rate	precision	bias	Idah CSI	hit rate	precision	bias
A + B + permanent water (primary)	**0.78**	**0.929**	**0.83**	**0.119**	**0.694**	**0.989**	**0.699**	**0.415**
A + B	0.728	0.864	0.822	0.052	0.668	0.952	0.691	0.378
A	0.226	0.228	0.966	-0.764	0.14	0.14	0.976	-0.856
A + B + C	0.625	0.921	0.661	0.393	0.575	0.96	0.589	0.63
Terrain alone (HAND at or below 7 m / 10 m)	0.616	0.893	0.665	0.344	0.573	0.948	0.592	0.602
JRC v2 model alone, 1-in-100-year	0.779	0.927	0.83	0.117	0.694	0.989	0.699	0.414

The study's six models, scored by our code (CSI)

Model	Lokoja 25-yr	Lokoja 100-yr	Idah 25-yr	Idah 100-yr
JRC	0.777	0.781	0.691	0.702
CaMa-Flood	0.679	0.759	0.478	0.583
CIMA-UNEP	0.681	0.722	0.582	0.633
ECMWF	0.666	0.655	0.623	0.617
GLOFRIS	0.432	0.519	0.409	0.456
SSBN	0.64	0.725	0.518	0.654

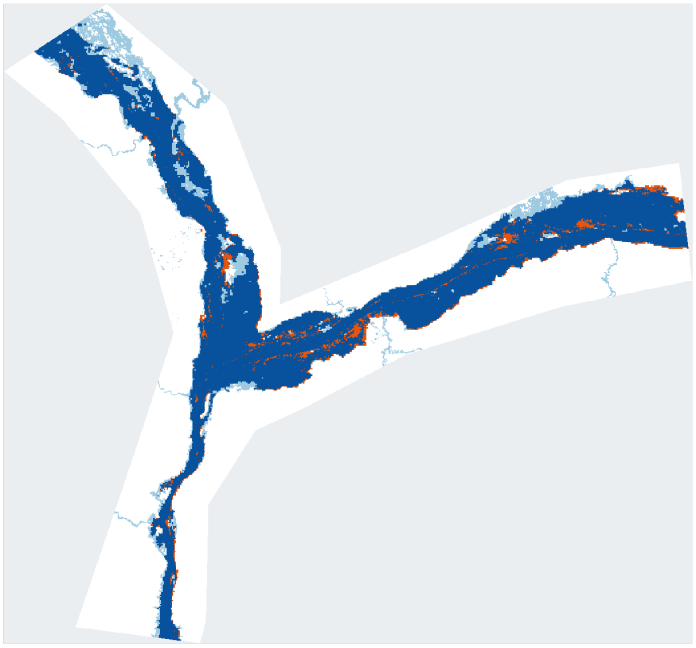
SSBN's shorter return period is 20 years, not 25 (as in the study). The paper reports the best model (JRC) at 0.78 and 0.7, and a regional model calibrated with local data at 0.73 and 0.53 (a slightly different area).

Against the 2022 flood (radar, independent)

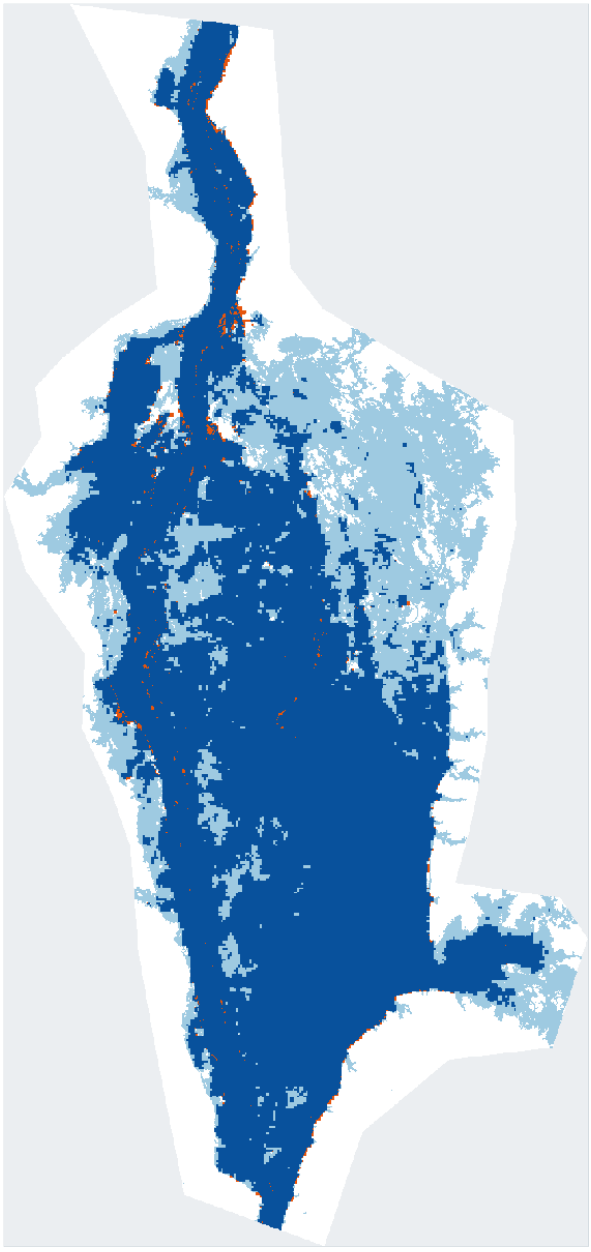
Zones	Lokoja CSI	hit rate	precision	Idah CSI	hit rate	precision
A + B	0.311	0.905	0.321	0.195	0.964	0.196
A	0.324	0.408	0.611	0.188	0.246	0.445
A + B + C	0.247	0.936	0.251	0.166	0.968	0.167

Where the zones and the 2012 flood agree

both flooded, not in the zones in the zones, not flooded



Lokoja: the confined valleys fit closely; misses are mostly thin strips along the river edges.



Idah: the zones spread east across the flat floodplain beyond the 2012 flood.

Limits

- Two areas on one river system, one large flood each year tested. The results show the method can work on large rivers; they don't show how it does on small rivers, in cities, in deltas or in dry regions.
- MODIS sees about 250 m cells and misses water under trees and in towns; radar sees only the days it passes. Neither reference is the true flood.
- The zones mark a 1-in-100-year flood; 2012 was roughly 1-in-50 and 2022 smaller still, so some "false alarms" are land that floods only in larger floods.
- Zone A overlaps with the 2012 reference (see Method).
- The JRC v2 maps (2024) may have been checked against floods like 2012 by their makers; we can't rule that out.
- Found and fixed during this study: the engine placed its elevation and water-history inputs half a cell (up to 15 m) too far north-west. The numbers here come from the corrected engine (data.fetch_dem and data.fetch_water_occurrence 1.2.0); before the fix the scores differed by at most 0.005 in CSI.

Reproduce it

Runs 0acf83dc-ad71-4870-9b34-378e0ceee809 (Lokoja) and cf87573d-816c-4389-a9bf-ad74bb7e4e3e (Idah), workflow case-study/nigeria 1.0.0 (baseline/1.0.0 plus the 2022 radar step, which no module reads), code 63353c0. Command: `scripts/baseline.sh case_study_nigeria --export /export/cases/nigeria`. Reference: Bernhofen, M. V. et al. (2018), A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique, Environmental Research Letters 13, 104007, doi:10.1088/1748-9326/aae014. Data: Bernhofen et al. (2018), Research Data Leeds, doi:10.5518/340 (CC BY 4.0).