Int. j. adv. multidisc. res. stud. 2023; 3(6):3071-3084

Received: 10-11-2023
Accepted: 20-12-2023

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Combining Field Survey and Open Geospatial Data for Transport Infrastructure Decision Support in Data-Scarce Regions

[1] Ridwan Tiamiyu, [2] Oladimeji Basit Alaka, [3] Olayemi Bolaji, [4] Oluwabusola Ajayi

[1] Fountain Construction Company, Nigeria
[2] Rivetline Concept Limited, Lagos, Nigeria
[3] Leedigital Research and Innovation, Ibadan, Nigeria
[4] Bisaj Engineering Limited, Osogbo, Nigeria

Corresponding Author: **Ridwan Tiamiyu**

Abstract

Transport decision support in regions without a road inventory is built on open geospatial layers whose reliability is usually unknown. We ask what that costs, and what a field survey is worth against it. Reading the error budget of thirteen prior studies out of their own result files rather than their prose, 23 of 42 paper-layer pairs carry no measured error at all, and every one of the thirteen studies has at least one. Where error has been measured it is poor: an agreement coefficient of 0.041 between two readings of a volunteered surface layer, and a flood detector at $F_1 = 0.356$ that returns nothing under canopy. Taking a rural rehabilitation decision that funds the most urgent decile of 7,068 links, and drawing 1,000 truth ensembles from those measured kernels, 325 of the 707 funded links, 46 percent and 594 km, are links a perfectly informed agency would not have funded. Two results follow. Propagating the kernels the program has already measured, with no fieldwork whatever, recovers 34.2 percent of the value of perfect information; it is the largest single return in the study and it is free. And survey effort should go to market and health access before surface condition, and not to served population at any budget, whose entire value at a complete census is no larger than the Monte Carlo noise of the calculation. A hundred visits chosen by decision uncertainty are worth roughly 629 chosen at random. In a worked case on 800 links, a survey costing between GBP 3,902 and 12,697 cuts misdirected funding from 55 links in 80 to 32. The layer ranking holds in 26 of 30 sensitivity variants, and the four that disagree all move the decision rule rather than the data.

Keywords: Value of Information, Volunteered Geographic Information, Error Propagation, Rural Roads, Survey Design, Decision Support, Nigeria

1. Introduction

An agency prioritizing rural road rehabilitation in a region with no road inventory has a well-populated toolbox and one missing number. Network geometry, gridded population, facility registers and satellite flood extents are all free, global and immediately usable (OpenStreetMap contributors, n.d.; World Bank 2019; Iimi *et al*. 2016) [42, 60, 29]. What none of them carries is a statement of how wrong it is in the place where it is about to be used.

The consequences of that gap are not hypothetical, and they are quantifiable, because a program of thirteen studies preceding this one measured some of them. Two readings of the same volunteered surface data, taken from two independent tag families, agree at a Cohen's coefficient of 0.041 (Cohen 1960) [16]. A Sentinel-1 flood detector reaches $F_1 = 0.356$ in open floodplain and returns no detections at all under canopy. A national health facility register carries 10.5 percent of its entries unconfirmed as operational. Those are the layers the toolbox supplies, and they are the layers a funding decision is about to be made on.

This paper asks the question that follows: given a fixed field budget, which layer should be surveyed, and what does a day of fieldwork buy? We answer it as a value-of-information analysis (Zhang *et al*. 2021; Canessa *et al*. 2015; Bhattacharjya *et al*. 2010) [61, 14, 8] run over the program's own measured error, and we deliver a documented workflow applied to a worked case. The contribution is an error budget and a decision calculation, not a review of the thirteen studies.

The calculation is possible here for reasons that will not hold everywhere. The error estimates are measurements made in those studies for their own purposes, not our estimates of our own layers. The decision is a real one with a real objective, funding the most urgent decile of a candidate list. And the survey can be costed, because measured productivity figures for road condition enumeration in comparable African settings exist (Workman 2018, 2017) [59, 58].

We build an error budget across thirteen studies by reading their result files rather than their prose, and report how much of the program's evidential base rests on layers whose error nobody has measured. We then compute the value of information by layer and by survey size, identify where the curves cross and saturate, and show that the ranking is driven by how much of each layer's error survives the decision rule rather than by how large that error is. Last, we apply the workflow end to end to a worked case and report the recommendation before and after survey, with costs.

Section 2 places the work in the value-of-information and data-quality literatures. Section 3 builds the error budget. Section 4 gives the decision model and the information calculation. Section 5 reports the results, Section 6 the worked case, and Section 7 what follows.

2. Related Work

2.1 Value of information

The value of information asks what a measurement is worth before it is taken, by comparing the expected quality of a decision made with it against one made without (Canessa *et al*. 2015; Zhang *et al*. 2021) [14, 61]. In civil engineering the tradition is structural: inspection and monitoring of bridges and components (Thöns 2018; Quirk *et al*. 2018; Memarzadeh and Pozzi 2016b, 2016a) [54, 45, 39, 38]. In health economics the expected value of sample information gives the machinery for asking how large a study should be (Ades *et al*. 2004; Heath *et al*. 2017) [1, 26]. Spatially, Bhattacharjya *et al*. (2010) [8] give the treatment closest to what we need, for decisions over a set of correlated locations.

What is largely missing is the application to geospatial input layers in transport planning. We located no study computing the value of a field survey against the measured error of an open geospatial layer for a network prioritization decision. That absence is the gap this paper occupies and it also means we have no template; the design here borrows the sample-size logic from remote sensing accuracy assessment (Foody 2009) [22] rather than from a transport precedent.

2.2 Data quality in volunteered and open geospatial data

The quality literature for volunteered geographic information is mature on completeness and positional accuracy (Haklay 2010; Senaratne *et al*. 2017; Barron *et al*. 2014; Barrington-Leigh and Millard-Ball 2017; Minghini and Frassinelli 2019) [25, 51, 6, 5, 40] and increasingly framed as fitness for use rather than accuracy in the abstract (Goodchild and Li 2012; Campalani *et al*. 2022; Wilson *et al*. 2017) [24, 13, 57]. Coverage is known to be uneven in ways that correlate with income and attention (Mahabir *et al*. 2017; Scholz *et al*. 2018; Brovelli *et al*. 2017) [36, 50, 11]. For African transport specifically, the mapping of informal networks has its own literature (Klopp and Cavoli 2019; Williams *et al*. 2015; Lawal and Anyiam 2019) [31, 56, 34].

Attribute quality, and specifically road surface condition, is the weak point. Efforts to infer it from imagery (Cadamuro *et al*. 2019; Jean *et al*. 2016) [12, 30], from smartphone sensors (Sandamal and Pasindu 2022; Thegeya *et al*. 2022) [48, 52] and from unpaved-road rating scales (Eaton *et al*. 1987; Mbabazi 2019; Sayers *et al*. 1986) [20, 37, 49] all exist. The nearest available work predicts road quality from satellite imagery at 94.0 percent binary paved-against-unpaved accuracy, but validates against a continental infrastructure diagnostic rather than against OpenStreetMap tags (Brewer *et al*. 2021) [10]. A citable comparison of OpenStreetMap surface tags against field verification does not exist, a point we return to in Section 7.4.

2.3 Error propagation and decision consequence

Propagating input error through a spatial model is long established (Heuvelink *et al*. 1989; Crosetto and Tarantola 2001) [27, 18] and has been applied to spatial multi-criteria analysis (Feizizadeh *et al*. 2014; Ligmann-Zielinska and Jankowski 2014) [21, 35]. The composite indicator literature makes the parallel argument that a ranking is a joint function of data, weights and aggregation (Nardo *et al*. 2008; Saisana *et al*. 2005; Becker *et al*. 2017; Paruolo *et al*. 2013) [41, 47, 7, 43], and the sensitivity tradition in multi-criteria analysis concentrates on weights (Chen *et al*. 2013; Deluka-Tibljas *et al*. 2013) [15, 19]. Hosseini and Smadi (2021) [28] is the nearest asset-management precedent, relating prediction accuracy to decision quality in pavement management.

The step this paper takes beyond that literature is to price the error rather than only propagate it, and to price it against a survey that could be commissioned tomorrow.

2.4 Agreement statistics, and a caution about them

Agreement between two readings is conventionally Cohen's coefficient (Cohen 1960) [16] with the interpretive bands of Landis and Koch (1977) [32], and the accuracy-assessment tradition uses the same statistic (Congalton 1991) [17]. Foody (2020) [23] argues it is unsuitable for accuracy assessment. We use it only as it was used upstream, as a reliability measure between two readings; a reliability figure stands in for an accuracy figure throughout, because no accuracy figure exists, and Section 7.4 sets out what that costs.

3. The Error Budget Across Thirteen Studies

3.1 How it was built

Thirteen studies precede this one. For each, we enumerated the layers its headline depends on and asked whether the study measured that layer's error. Every cell was resolved at run time from a stated key in the study's own result file, not from its prose, because a prose summary and the result file behind it do not always agree. The budget is 42 paper-layer rows and is reported in Table 1.

Table 1: Error budget across the thirteen preceding studies: how many of the data layers each headline rests on carry an error the study itself measured

Study	Headline quantity	Layers	Measured	Not measured	Layers with no measured error
Concrete fatigue	Fatigue reduction factor, river gravel	3	2	1	Aggregate property anchors
Drainage condition	Inter-rater agreement, drainage condition	3	2	1	OpenStreetMap corridor geometry
Health access	Population within 30 min of a facility	5	2	3	WorldPop raster; OpenStreetMap road geometry; boarding time
Delivery routing	On-time delivery rate, deterministic plan	3	1	2	OpenStreetMap network geometry; delivery instances
Essential trips	Walkable fallback by income interaction	2	1	1	Google Community Mobility series
Crash clustering	Top-decile hotspot overlap under displacement	4	2	2	OpenStreetMap corridor geometry; relative risks
Emergency response	Survival lost by optimizing coverage	3	0	3	Health facility register; chain times; crash surface
Motor park loading	Elasticity of delay to dwell mean, in-lane	3	1	2	Vehicle dwell times; OpenStreetMap corridor and width tags
Flood detection	Sentinel-1 flood detection F_1	4	2	2	Reference impassability labels; segment topology
Flood rerouting	Settlement pairs severed by the worst link	3	1	2	Origin-destination flows; network topology
Coverage and condition	Population failing the road condition test	3	2	1	WorldPop raster
Willingness to pay	Willingness to pay for the shared service	2	1	1	NBS transport fare series
Rank stability	Rank agreement under layer perturbation	4	2	2	WorldPop raster; access speed model
All		42	19	23	55% of layers

Note. Each row is one study, named for its subject and identified by the quantity its headline reports; the thirteen are analyses of Nigerian road infrastructure, spanning pavement materials, drainage, health and market access, freight routing, crash clustering, emergency response, roadside parking, flood detection and rerouting, network coverage, fare willingness to pay and rank stability. Every numeric cell was built from each study's own result files rather than from its prose, resolving against a stated key, and the resolution is verified at run time. A layer counts as measured when the study reports a number for that layer's error, not when it merely reports uncertainty in its own output. Every one of the thirteen studies carries at least one unmeasured layer. The rightmost column names every unmeasured layer rather than only the first, so the two recurring gaps can be counted off it: OpenStreetMap road or corridor geometry is used with no positional error measured in five studies and the WorldPop gridded population raster in three.

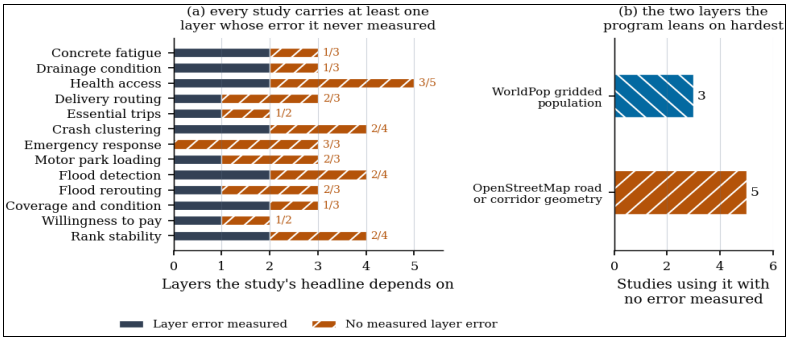

Fig 1: The error budget across the thirteen preceding studies. Panel (a) gives, for each study, how many of the layers its headline depends on carry an error the study itself measured (slate, solid) against how many do not (orange, hatched), with the unmeasured fraction labeled; every study has at least one unmeasured layer and none has a clean sheet. Panel (b) counts the studies using each of the two most-relied-upon layers with no error measured: OpenStreetMap road or corridor geometry in five, WorldPop gridded population in three

3.2 What it shows

Nineteen of the 42 layers carry an error the study itself measured. Twenty-three do not, which is 55 percent. Every one of the thirteen studies carries at least one layer whose error it never measured, and the number with a clean sheet is zero.

The gaps are not scattered at random. OpenStreetMap road or corridor geometry is used with no positional error measured in five studies, and the WorldPop gridded population raster in three. The two layers the program leans on hardest are the two whose error it has never quantified. Fig 1 abve shows the distribution.

Four studies did measure their layer error, and those measurements are what the rest of this paper reuses. Table 2 states each with its provenance: an agreement coefficient of 0.041 between two readings of the OpenStreetMap surface layer, $F_1 = 0.356$ for a Sentinel-1 flood detector with zero recall under canopy, 10.5 percent of a health facility register not confirmed operational, and a segment file described as routable that leaves 2,237 components.

Table 2: The measured error estimates this study reuses, and how much of each layer's error is common to every link rather than specific to one

Layer	From	Statistic	Value	Sample	Common (%)	After scoring (%)
Surface condition	Condition study	Cohen's κ, two independent readings of the same OpenStreetMap ways	0.041	155 ways	0.008	0.027
Flood exposure	Flood study	F_1 of the Sentinel-1 detector, Niger floodplain	0.356	169,200 segment-days	0.008	0.004
Flood exposure	Flood study	Recall in the canopied Osun basin	0.000	628,170 segment-days	—	—
Served population	Prioritization study	Assumed multiplicative log standard deviation	0.21	no published bound	0.257	4.175
Market and health access	Access, prioritization	Span between the two published speed tables	1.5× either way	no Nigerian measurement	35.88	1.46
Health facility register	Prioritization study	Share of the eHealth Africa register not confirmed operational	0.105	1,607 facilities	—	—

Note. Every entry in the Value column is a measurement one of those studies reports rather than a quantity set here. The two rightmost columns are estimated from the 2,000-draw prior ensemble: they decompose each layer's departure from its open reading into a component shared by every link in a draw and a component specific to the link. The final column repeats the decomposition after the scoring rule rescales the criterion, which is what the decision actually sees. Two rows carry a dash because the quantity is a rate over a stratum rather than a per-link layer.

The budget is a finding about the program rather than a criticism of any study in it. A study that does not measure a layer's error is not thereby wrong; it is silent on something a downstream user needs. An error budget makes that silence visible and countable, and 55 percent is the count.

What the budget is made of shapes everything downstream. The measured layers are not a random sample of the layers: they are the ones somebody had a reason to check, which usually meant suspecting them. The condition layer was measured because its tags looked unreliable, and the flood detector because a null result demanded explanation. That selection runs against this paper's argument rather than for it, since the unmeasured layers may well be better than the measured ones, and it is the reason we frame the audit as a count of silences rather than as an estimate of how wrong the program is.

Measurement is also cheap where it happened. All four of the measured errors were obtained at the desk, from material the program already held. The condition agreement came from reading the same volunteered source twice through two different tag families; the detector skill came from comparing against a hydrological reference; the register share came from a status field the register already carried. Each cost a day of analysis at most. That is worth stating next to the 55 percent, because the obstacle to closing those gaps is evidently not cost.

4. Methods

The truth ensemble propagates the measured error kernels through 1,000 draws; the Monte Carlo replicate check repeats the whole calculation on five independent ensembles of 300 draws each; the sensitivity analysis runs 30 variants over eight of the analyst-set quantities.

4.1 The decision

The decision is the one a rural roads agency actually faces: fund the most urgent decile of a candidate list. Let L index the 7,068 candidate rural links from the prioritization study, which span the two windows analyzed here, the Niger floodplain window in Kwara and the Osun basin, each carrying five indicators, and let Z denote the layer values. A ranking $r(Z)$ is produced by weighted summation over normalized indicators, and the decision funds the set $F(Z)$ of the 707 highest-ranked links.

Write $F^{\star} = F(Z^{\star})$ for the set a perfectly informed agency would fund, where $Z^{\star}$ is the truth. The value of a decision is the share of achievable urgency it captures,

$$V(F) = \frac{\sum_{i \in F} u_i^{\star} - \sum_{i \in F^{open}} u_i^{\star}}{\sum_{i \in F^{\star}} u_i^{\star} - \sum_{i \in F^{open}} u_i^{\star}} \quad (1)$$

With $u_i^{\star}$ the true urgency of link i and $F^{open} = F(Z^{open})$ the set an agency funds by ranking the open layers at face value. Equation 1 is scaled so that perfect information scores one and the open layers read at face value score zero, which makes every value-of-information figure in this paper a share of what perfect information would add to the list the agency already has, rather than an unanchored number.

Two scales are therefore in play and the paper keeps them apart. The raw capture:

$$C(F) = \frac{\sum_{i \in F} u_i^{\star}}{\sum_{i \in F^{\star}} u_i^{\star}}$$

Is the share of achievable urgency a funded list realizes, and on it perfect information scores one while a list that funds no urgency at all scores zero; a random allocation of the same size is one point on that scale and not its origin. Equation 1 is C rescaled to put its zero at $C(F^{open})$. Raw captures are quoted only where they are named as such, and every share, curve and table entry labeled a share of the value of perfect information is on the scale of Equation 1.

4.2 The truth ensemble

The true layer values $Z^{\star}$ are not known, and the analysis has to carry that uncertainty rather than assume it away. What the program supplies is a set of measured error kernels, and those kernels, applied to the open reading, define a posterior over the true state. The ensemble is how that posterior is propagated through the ranking: each draw is one state of the network admissible under the errors the program has measured, and every expectation reported below is an expectation over that uncertainty. This is the standard construction in a value-of-information analysis, where the quantity being priced is exactly a reduction in uncertainty about an unknown true state. We draw:

$$Z^{\star(m)} \sim P(Z^{\star} \mid Z^{open}), m = 1, \ldots, 1000 \quad (2)$$

Using only the four measured kernels of Table 2: the condition cross-tabulation reproducing $\kappa = 0.041$, the flood detector's per-area confusion counts, the register's non-operational share, and the population and access kernels the prioritization study carried. Every quantity reported below is an expectation over this ensemble.

Three decision bases are compared against it. *Open layers* ranks on Z^{open} read at face value. *Measured kernels* ranks on the posterior mean, which uses the error the program has already measured and requires no fieldwork whatever. *Perfect information* ranks on $Z^{\star}$ and is the upper bound by construction.

4.3 Surveying, and the posterior update

A survey visits a set $S \subset L$ of links and records one or more layers. For each layer we apply the conjugate update its kernel admits: Dirichlet for the condition grade, Beta for the flood state, and a discrete update over the catchment radius for population. The access layer has no tractable exact posterior, because it compounds speed multipliers, a market threshold and per-facility operational status; we use a hierarchical Gaussian approximation and cross-check it by applying the same approximation to the three layers where the exact update is available. The largest disagreement is 0.0723, on condition, where the approximation is conservative; on access itself the two differ by 0.0169.

Links are chosen by decision uncertainty, that is by how close a link sits to the funding boundary across the ensemble, rather than at random. Section 5.4 reports what that choice is worth.

4.4 The variance decomposition

To explain the shape of the resulting curves we decompose each layer's error into a component common to every link in a draw and a component specific to the link, at two points: on the layer's own scale, and after the scoring rule has been applied. The difference between those two is the paper's mechanism and is reported in Section 5.2.

4.5 Costing

Survey effort is expressed in links visited and in kilometers, and costed at the per-kilometer productivity and price figures measured for road condition enumeration in comparable African settings (Workman 2018, 2017) [59, 58]. No published cost or productivity figure exists for enumerating served population or for recording seasonal impassability, so all layers are priced at the condition-survey rate.

4.6 Sampling noise is not geography

Replicate variance is the Monte Carlo noise of the truth ensemble, estimated by repeating the calculation on five independent ensembles. *Area variance* is the difference between the two study windows. They are reported separately in Section 6.2, in separate table columns, and are never added.

4.7 Analyst-set quantities

Fourteen quantities are set by the analyst, six inherited from the prioritization study and eight introduced here. Table 3 lists them all. The inherited six are the weight vector, the market threshold, the grade multipliers, the link frame, the population error scale and the flood date-independence assumption. The eight introduced here are the funded share, the value function of Equation 1, survey fidelity, survey noise, the survey unit, the assumption that the decision maker knows the kernels, the scoring bands, and the access update.

Table 3: Every analyst-set quantity in the model, and what happens to the two claims when it moves

Analyst-set quantity	Reported value	From	Status	Free correction under the sweep	Top layer at $n = 100$
Criterion weight vector	equal, 0.2 each	Prioritization study	inherited	0.177 to 0.555	access 8, condition 1, population 1
Market-center population threshold	5,000 persons	Prioritization study	inherited	swept inside every draw	—
Speed multipliers for condition grades 3, 4, 5	0.80, 0.60, 0.40 of the unpaved class speed	prioritization, from access study	inherited	0.354 to 0.354	access 1
Stroke cut at 3.0 km and 300 m floor	3,000 m cut, 300 m floor	Prioritization study	inherited	elsewhere, see note	—
Multiplicative population error, log sd	0.21	Prioritization study	inherited	0.331 to 0.344	access 4
Flood dates treated as independent	15 independent acquisition dates	prioritization, from flood study	inherited	0.347 to 0.364	access 2
Funded share of the candidate set	0.10 (the top decile, 707 of 7,068 links)	this study	new	0.334 to 0.375	access 2
Benefit of funding a link	its equal-weight priority score under the true layers	this study	new	-0.099 to 0.340	access 3, condition 2
A field visit resolves the surveyed link exactly	perfect on the visited link	this study	new	0.341 to 0.341	access 3
Observation noise in the imperfect arm, log sd	0 in the reported specification	this study	new	elsewhere, see note	—
One survey unit is one link visit	1 link = 1 unit for every layer	this study	new	elsewhere, see note	—
The decision maker knows the measured error kernels	yes; the condition cross-tabulation and the detector confusion counts are priors	this study	new	elsewhere, see note	—
Fixed scoring band for each criterion	1st and 99th percentile of the calibration ensemble's true values	this study	new	0.328 to 0.530	access 3
Access layer posterior update	hierarchical Gaussian on log travel time,	this study	new	elsewhere, see note	—

	common plus idiosyncratic				

Note. Six rows are inherited from the prioritization study and eight are new here, and 30 variants over 8 of the 14 keys were run. The market threshold is drawn afresh inside every truth draw rather than swept as a variant, and four keys are handled elsewhere: the link frame is the prioritization study's construction and is carried unchanged, the survey unit is swept in the costed worked case, the naive reference on every row is itself the sweep of what the decision maker knows, and the access posterior update is cross-checked against the benchmark set. The free correction is the share of the value of perfect information a decision maker captures by applying the measured error kernels without going into the field. Every row of this table, the reported specification included, is computed on the sweep's own 300-draw ensemble, where the reported specification gives 0.341; the 1,000-draw headline ensemble of Table 4 puts the same quantity at 0.342 and the difference is replicate noise. Market and health access is the layer worth resolving first in 26 of 30 variants; the four that disagree are named in the summary file and all four move either the weight vector or the aggregation rule. The spread across this table is parameter variance and is a different quantity from the replicate variance in Table 6.

Table 4: Share of the value of perfect information captured, by layer surveyed and by number of links visited, under two ways of choosing which links to visit

Survey design	Layer resolved	$n = 0$	$n = 25$	$n = 100$	$n = 400$	$n = 1200$	$n = 3500$	$n = 7068$
Random	Surface condition	0.342	0.343	0.345	0.353	0.377	0.449	0.558
Random	Flood exposure	0.342	0.343	0.346	0.356	0.381	0.456	0.573
Random	Served population	0.342	0.343	0.344	0.345	0.345	0.346	0.347
Random	Market and health access	0.342	0.370	0.373	0.381	0.402	0.476	0.623
Random	All four at one visit	0.342	0.372	0.380	0.405	0.472	0.673	1.000
Targeted	Surface condition	0.342	0.345	0.363	0.442	0.556	0.558	0.558
Targeted	Flood exposure	0.342	0.343	0.344	0.358	0.464	0.538	0.573
Targeted	Served population	0.342	0.344	0.344	0.345	0.347	0.347	0.347
Targeted	Market and health access	0.342	0.384	0.402	0.476	0.591	0.623	0.623
Targeted	All four at one visit	0.342	0.390	0.425	0.563	0.850	0.968	1.000

Note. 1,000 truth draws, 7,068 candidate links, a funded decile of 707 links. The floor is the ranking an agency gets from the open layers read at face value and the ceiling is perfect information; $n = 0$ is therefore the value of applying the measured error kernels with no fieldwork at all, which is 0.342. Targeted means visiting the links whose funding status is least certain under the prior, which a decision maker can compute before going anywhere. Replicate variance is reported separately in Table 6 and is never added to anything here.

Section 6.3 reports 30 variants over eight of them and which conclusions survive them.

5. Results

5.1 What layer error costs a funding decision

See Table 4 reports the three decision bases against perfect information.

Ranking on the open layers at face value realizes 0.8896 of the urgency a perfectly informed agency would fund, against 0.6612 for a random allocation of the same size. Those are raw captures. Equation 1 puts its zero at that 0.8896, so every share reported from here on is a share of the 0.1104 that perfect information would add. In links, 325 of the 707 funded, 46 percent, carrying 594 km, are links a perfectly informed agency would not have funded. That is the size of the problem, and it is not a rounding error on a weight-sensitivity analysis.

Applying the measured error kernels, and going nowhere, lifts the raw capture to 0.9275 and cuts the expected count of misdirected links from 324.8 to 268.4. In the units of Equation 1 that recovers 34.2 percent of the value of perfect information and saves 56 misdirected links, for the cost of reading four numbers out of result files that already existed. It is the largest single return in this study and it is free.

The correction does not make the layers better; the open reading and the posterior mean are built from exactly the same observations. What it does is stop the decision treating a grade of three as a fact when the kernel says a grade of three is very nearly a draw from the marginal. The ranking that results is more cautious about links whose position rests on an unreliable indicator and correspondingly more confident about links whose position rests on several agreeing ones. That is shrinkage, and shrinkage is worth 34.2 percent of perfect information here because the layers are poor enough for the raw reading to be badly overconfident.

The corollary is that the correction is worth less the better the layers are, and worth nothing at all if they are perfect. An agency reading this should not expect a third of perfect information from applying kernels to good data. They should expect the return to scale with how bad the data is, which is a reason to measure the badness first.

It is free because the kernels are not new measurements. They are measurements the program already made, sitting in result files, and the correction consists of ranking on the posterior mean rather than on the raw reading. Any program that has measured any layer's error can do this, and the calculation is an afternoon.

5.2 Why the curves have the shape they have

The two rightmost columns of Table 2 report the decomposition and Fig 2 shows it.

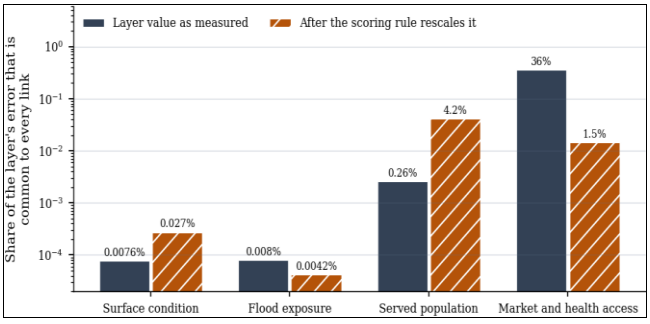

Fig 2: The share of each layer's error that is common to every link, on a logarithmic axis, measured on the layer's own scale (slate, solid) and again after the scoring rule has rescaled it (orange, hatched). Market and health access falls from 36 percent to 1.5 percent, because an additive score with a band fixed in advance is invariant to a shift shared by the whole network. Served population moves the other way, from 0.26 to 4.2 percent. For a ranking decision almost all layer error is idiosyncratic, which is why no cheap global calibration exists

On the layer's own scale the access layer looks highly calibratable: 35.9 percent of its error is a level shift shared by the whole network, the kind of thing a handful of measured trips would settle. The decision does not see it. An additive score with a band fixed in advance is invariant to a shift common to every link, so that 35.9 percent collapses to 1.46 percent by the time it reaches a rank. Condition moves the other way, from 0.0076 percent to 0.0275 percent, and population from 0.257 to 4.175 percent.

The conclusion is general rather than local. For a ranking decision, essentially all layer error is idiosyncratic, because a ranking is invariant to whatever the layers share. There is no cheap global calibration to be had, and value therefore accrues with links visited rather than with a one-off correction. That is a property of ranking decisions and not of this network, and it is why the survey curves below rise slowly and steadily rather than jumping.

It also explains an asymmetry that would otherwise look odd. Served population has the second largest common component after the scoring rule, at 4.2 percent, and is nonetheless the layer least worth surveying. Common variance is not the only thing that fails to reach a ranking: a layer whose error is large but whose indicator barely discriminates between candidate links cannot move the order much either way. Population enters this composite through a logarithm precisely because it is heavily skewed, and the transform that keeps a handful of periurban links from dominating the score also flattens the differences the survey would resolve. The layer is important to the decision in the sense that the agency cares about people; it is unimportant to the ranking, which is what the survey budget is actually buying.

5.3 Which layer to survey

Table 4 gives the value captured against links visited, by layer, and Fig 3 plots it.

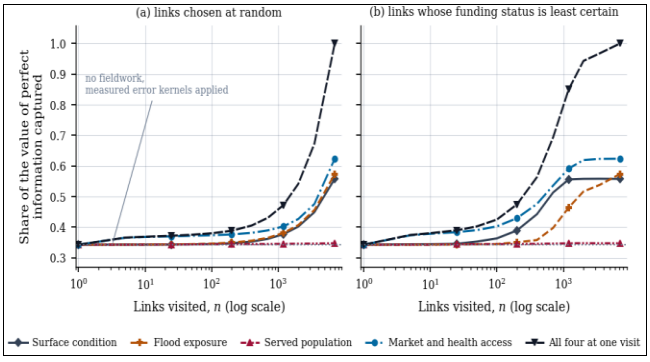

Fig 3: Share of the value of perfect information captured against links visited, on a logarithmic axis, for each layer alone and for all four recorded at one visit. Panel (a) samples links at random; panel (b) targets the links whose funding status is least certain. The horizontal reference near 0.34 is the no-fieldwork floor reached by applying the measured error kernels alone. Served population (crimson, triangles) is flat against that floor at every survey size in both panels. Targeting lifts every curve and lifts the all-four curve furthest

The ranking at a hundred visits is market and health access first and surface condition second. Flood exposure and served population are not separated: Table 4 prints both at 0.344, and they are indistinguishable at any precision worth printing, since the gap between them is about a twentieth of the replicate standard deviation. The margins matter more than the order. Against the no-fieldwork floor of 0.342, a hundred visits recording access add 0.060, condition 0.021, flood 0.002 and population 0.002. Recording all four layers at each of the hundred links adds 0.083.

Served population is the striking case. Resolving that layer entirely, on all 7,068 links, adds 0.0047 of the value of perfect information, which the replicate standard deviation of 0.0046 reported in Section 6.2 cannot be told apart from. At the error the program attributes to it, that layer contributes essentially nothing to this decision and should not be surveyed at any budget. The error scale for that layer, a log standard deviation of 0.21, was revisited after the benchmark check, and Section 6.3 reports the result at four times it.

Fig 4 shows the marginal value per visit. Two crossings occur. Condition leads flood until 5,192 visits and flood leads above it, the only crossing among the four under targeted sampling, because the flood layer only comes into its own once the survey is large enough to reach links the detector missed entirely. Flood overtakes population at 81 visits.

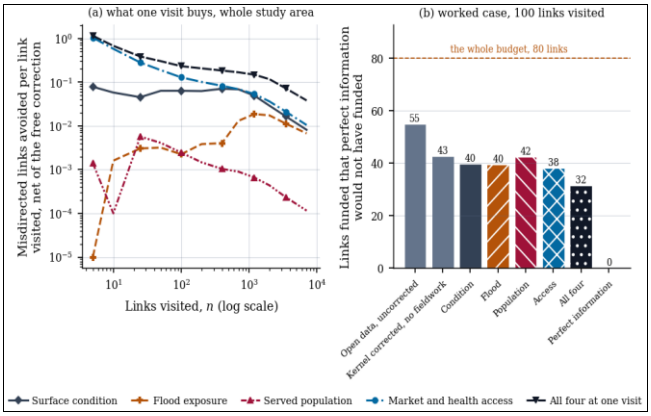

Fig 4: Panel (a) gives the marginal return to one more visit across the whole study area, as misdirected links avoided per link visited net of the free correction, on logarithmic axes. Access and the all-four curve dominate throughout; the flood curve rises between roughly 100 and 1,000 visits as the survey begins reaching links the detector missed. Panel (b) gives the worked case: links funded that perfect information would not have funded, under each decision basis, against the whole budget of 80 links marked by the dashed line

The curves saturate at very different points. Access reaches 95 percent of its full value at 2,000 visits, condition at 1,200, and flood not until the full 7,068. Beyond roughly two thousand visits, a third of the network, further survey of the two layers worth surveying stops paying. No saturation point is quoted for population, because the entire curve spans about one replicate standard deviation and its shape is not interpretable.

5.4 Targeting is worth more than sample size

At a hundred visits, choosing links by decision uncertainty captures 0.425 against 0.380 for a random sample of the same size. Reaching the targeted hundred-visit result by

random sampling takes roughly 629 visits. The same field budget spent on the right links is worth several times what it is worth spent at random, and identifying the right links requires no fieldwork, only the posterior the program's own kernels already define.

The factor of roughly six is itself an artifact of where the uncertainty sits, and it should be read as a property of this decision rather than a general multiplier. A decision funding the top decile concentrates its uncertainty in a narrow band around the boundary, so a targeted sample can find that band immediately. A decision with a diffuse objective, or one funding a much larger share of the network, would have its uncertainty spread more evenly and would gain correspondingly less from targeting. What generalizes is the direction and the mechanism, not the six.

One visit should record everything. Resolving all four layers at each visited link captures 0.425 at a hundred visits against 0.402 for the best single layer, and 0.850 at 1,200 visits against 0.591. A field team standing on a link can record its surface, whether it was cut last season, who it serves and how long the trip to market takes, and the analysis says the marginal cost of the extra three is well repaid.

6. The Worked Case

Table 5 reports the case and Fig 5 shows it.

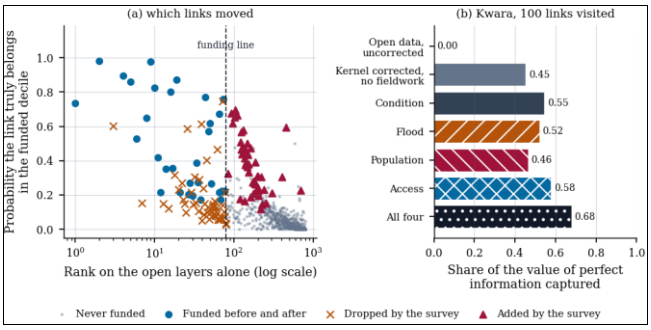

Fig 5: The worked case in the Niger floodplain window. Panel (a) plots each link's probability of truly belonging in the funded decile against its rank on the open layers alone, with the funding line at rank 80: links dropped by the survey (orange crosses) cluster below the line's left, and links added (crimson triangles) cluster above it to the right, so the survey moves work across the boundary in both directions. Panel (b) gives the share of the value of perfect information captured under each decision basis, from the uncorrected open data at zero to all four layers surveyed at 0.68

Table 5: Worked case, the Niger floodplain window in Kwara State: what an open-data funding list is worth, and what a hundred link visits buy

Decision basis	Links misdirected	Kilometers	Value captured	Links changed	Realized value
Open layers, read at face value	54.9	117	0.000	0	0.804
Measured error kernels, no fieldwork	42.7	154	0.454	47	0.903
Surface condition surveyed	39.8	135	0.545	46	0.920
Flood exposure surveyed	39.6	136	0.520	44	0.907
Served population surveyed	42.4	152	0.465	46	0.899
Market and health access surveyed	38.2	117	0.579	46	0.929
All four at one visit surveyed	31.6	87	0.678	49	0.948
Perfect information	0.0	0	1.000	55	1.000

Note. 800 candidate rural links carrying 1510 km; the decision funds the most urgent 80 of them. Columns two to four are expectations over 800 truth draws; columns five and six are one realized survey under one draw of the truth and are reported for concreteness only. Links misdirected counts funded links that a perfectly informed agency would not have funded. The survey visits the 100 links whose funding status is least certain, 423 km of road, which the measured productivity of a traditional condition survey team puts at 10.9 to 16.0 team-days and the measured per-kilometer cost at GBP 3902 to 12697 at 2016–17 prices. All four layers are priced at the condition-survey rate.

The case is the Niger floodplain window in Kwara: 800 candidate rural links carrying 1,510 km, with the decision funding 80. It is the harder of the two windows, because the flood layer carries what little signal it has here and market access is worst here.

Before. The open layers give a list of 80 links, of which 54.9 in expectation, 69 percent, carrying 117 km, are links a perfectly informed agency would not have funded.

The recommended survey. Visit the 100 links whose funding status is least certain under the prior and record all four layers. That is 423 km of road, between 10.9 and 16.0 team-days at measured productivity, and between GBP 3,902 and 12,697 at 2016 to 2017 prices.

After. Misdirected links fall to 31.6 and misdirected kilometers to 87, capturing 0.678 of the value of perfect information. Applying the kernels alone, with no survey, already reaches 0.454.

6.1 One row of that table needs reading carefully

The kernel correction funds fewer wrong links than the open reading, 54.9 down to 42.7, but the wrong links it does fund are longer, so misdirected kilometers rise from 117 to 154. Correcting the condition layer for its measured optimism pushes long rural links up the list and some of them do not belong there.

It changes the advice. Surveying brings both columns down together: access surveying returns kilometers to 117 and all four layers to 87. An agency budgeting in kilometers rather than in contracts should read the kilometer column and should not stop at the free correction, even though the free correction is the best value per unit of effort in the study.

In the single realized survey the funded list changes by 49 of its 80 links. That is one draw of the truth and one survey, reported for concreteness; the expectations above are the numbers to quote.

6.2 Replicate variance and area variance

Table 6 sets the two quantities side by side, in columns that are never combined.

Across five independent replicate ensembles at 700 visited links, the standard deviation of the mean share of perfect information captured is 0.0025 for access, 0.0046 for condition, 0.0046 for population and 0.0054 for flood. The calculation reproduces to the third decimal place.

Table 6: Replicate variance and area variance, reported separately

Layer resolved	Mean	MC s.d.	Osun basin	Niger floodplain	Area gap	Gap / MC s.d.
Surface condition	0.3550	0.0046	0.4089	0.5524	0.1434	31
Flood exposure	0.3597	0.0054	0.3823	0.7307	0.3484	65
Served population	0.3385	0.0046	0.3721	0.4642	0.0921	20
Market and health access	0.3854	0.0025	0.4170	0.6153	0.1983	78
All four at one visit	0.4273	0.0022	0.4628	0.9419	0.4791	216

Note. Share of the value of perfect information captured at $n = 700$ visited links under random sampling. Replicate variance, reported as the Monte Carlo standard deviation (MC s.d.), is the spread of the mean across five independent replicate truth ensembles of 300 draws each and is sampling noise. Area variance is the same quantity computed inside each study window with that window's own decile as the budget, and is geography. The two are different quantities and must never be added; the final column shows the area gap running from 20 times the replicate standard deviation for served population to 216 times when all four layers are recorded at one visit, so a reader who pooled them would be reporting geography as noise.

Across the four single layers, area variance is between twenty and seventy-eight times larger, and it is 216 times larger when all four are recorded at one visit. Computed inside each window against that window's own decile, the value captured runs 0.4089 in the Osun basin against 0.5524 in the Niger floodplain for condition, 0.3823 against 0.7307 for flood, 0.3721 against 0.4642 for population, and 0.4170 against 0.6153 for access. Fig 6 plots the two components side by side.

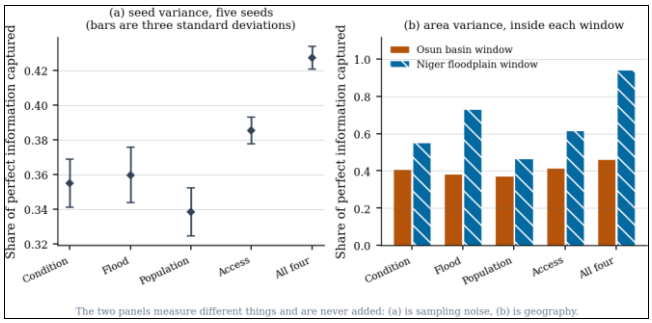

Fig 6: Replicate variance and area variance, plotted apart and never added. Panel (a) gives the mean share of perfect information captured across five independent replicate ensembles, with bars at three standard deviations; the spread is Monte Carlo sampling noise and is negligible. Panel (b) gives the same quantity computed inside each study window against that window's own decile, Osun basin against Niger floodplain. The gap between windows runs from twenty times the replicate standard deviation for served population to 216 times for all four layers recorded at one visit, and it is geography rather than noise

The two are never added. Replicate variance is Monte Carlo sampling noise and is negligible; area variance is geography and is the finding. Fieldwork is worth substantially more in the Niger floodplain than in the Osun basin, and an agency with one survey budget and two basins should spend it in the floodplain.

The reason the floodplain repays fieldwork better is visible in the layers rather than mysterious. It is the window where the flood indicator carries any signal at all, since the detector returned nothing in the basin, and it is the window where market access is worst and therefore most variable. Both give a survey more to correct. The basin window, by contrast, is closer to a case where the open layers are uninformative in the same direction everywhere, and a survey there spends most of its visits confirming an order the ranking already had.

That is not yet a rule about floodplains. Two windows do not establish which geographical features predict a high return to fieldwork; the return varies by a factor we can measure between two specific places and cannot yet attribute.

6.3 What survives moving the analyst-set quantities

Thirty variants over eight of the fourteen quantities are reported in Table 3 and Fig 7.

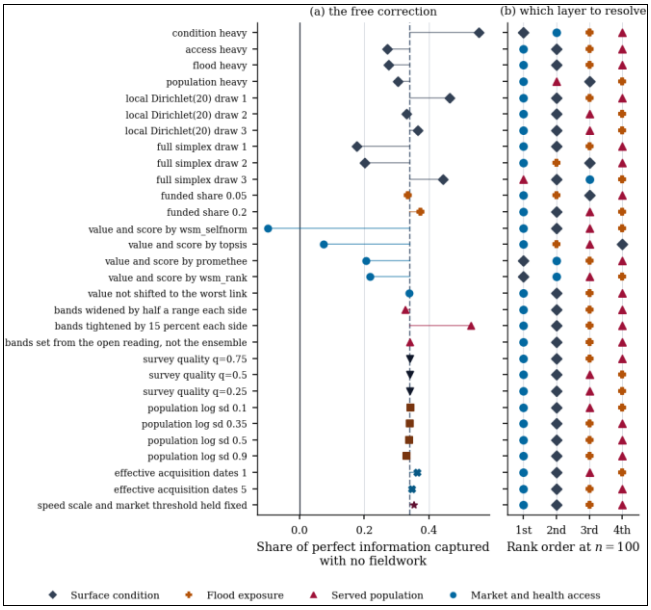

Fig 7: Sensitivity across 30 variants over eight of the analyst-set quantities. Panel (a) gives the value of the free kernel correction under each variant, against the reported specification marked by the dashed line at 0.341, the value the sweep's own 300-draw ensemble gives for the quantity the headline ensemble puts at 0.342; the single negative value is the variant that rescales each criterion on whatever matrix it holds, which is a renormalization artifact rather than a substantive reversal. Panel (b) gives the rank order of the four layers at 100 visits under each variant. Market and health access is first in 26 of 30, the four exceptions all move the weight vector or the aggregation rule, and served population is last or second to last in every variant

The free correction is not universally positive. It ranges from -0.099 to 0.555 against 0.341 for the reported specification on the sweep's own ensemble, the quantity the 1,000-draw headline ensemble puts at 0.342, and is negative in exactly one variant: the one that rescales each criterion on whatever matrix it is holding, which is the prioritization study's own rule. Under a self-normalizing rule the score is nonlinear in the layer values and undoes posterior shrinkage. That is a renormalization artifact rather than a substantive reversal, and a reader running the obvious analysis will hit it.

The layer ranking is not fully robust. Access is the layer to resolve first in 26 of 30 variants, 87 percent. The four that disagree all move either the weight vector or the aggregation rule: a condition-heavy weight vector, one draw from the

full simplex, PROMETHEE II, and rank normalization. The full four-layer order is identical in 47 percent of variants. The claim that survives without qualification is the weaker and more useful one: served population is last or second to last in every single variant.

6.4 Benchmarks

Twelve quantities were compared against published measurement. Two are inside by construction, two inside, one consistent in direction and physically expected, one outside and far below in the direction that motivates the study, one outside in the direction that required action, one at the edge, one consistent in direction but not like for like, and three have no usable comparator.

The measured F_1 of 0.356 sits inside the published range for Sentinel-1 flood water (Bonafilia *et al*. 2020) [9], and zero detections under closed canopy is physically expected rather than anomalous (Landuyt *et al*. 2020; Refice *et al*. 2020; Tsyganskaya *et al*. 2018) [33, 46, 55]. The agreement coefficient of 0.041 sits far below trained-rater agreement, which is the motivating fact of the paper rather than a failure of it. The comparators are routine visual bridge inspection (Phares *et al*. 2001) [44] and a pavement condition index variability study running 63 inspectors over 61 control sections (Andrei and Arabestani 2018) [2]; both report disagreement between trained raters that is an order of magnitude better than what two readings of the volunteered layer achieve. The register's 10.5 percent non-operational share sits at the edge above field-measured values of 2.2 to 6.0 percent (Bala *et al*. 2020) [3]. And the 46 percent misdirection rate sits inside the swing Nigerian coverage estimates show under a change of method, from 100 percent to 57 (Banke-Thomas *et al*. 2021) [4].

One check failed and was acted on. The assumed population error scale, a log standard deviation of 0.21, is too small against the only published measurement of gridded population error in Nigeria, which puts the WorldPop constrained product recovering a median 0.27 of field-referenced population (Thomson *et al*. 2021) [53]. That implies a log error near unity, four times the scale assumed here. An additional sweep arm at log standard deviation 0.90 was added and everything downstream rerun. Served population remains the layer least worth surveying and access remains the layer to resolve first; the free correction moves only from 0.341 to 0.331. The conclusion survives the correction and both numbers are reported. We note that the comparator measures urban informal settlements rather than rural catchments, so it is an upper bound rather than a like-for-like figure, but the direction is unambiguous.

Three quantities have no usable published comparator and each is stated as a gap: the share of decision loss that inspection removes in road asset management, where the quantified literature postdates this study's citable window; the accuracy of OpenStreetMap surface tags against field verification, where the relevant studies are later still; and any comparison of rank instability from layer error against rank instability from weight choice, for which no external precedent could be found.

7. Discussion

7.1 The workflow, stated as a procedure

The paper's practical output is a sequence an agency can follow.

Assemble the open layers and build the decision on them. Then, before commissioning anything, collect whatever error estimates exist for those layers, from the literature, from a duplicate reading of the same source, or from earlier work on the same layers, and rank on the posterior mean rather than the raw values. On this problem that step alone recovers a third of the value of perfect information and costs nothing.

Then decompose each layer's error into what is common to the network and what is specific to a link, and check how much of each survives the decision rule. For a ranking decision the common part mostly does not survive, which is why the calibration exercise an analyst instinctively reaches for is worth less than it appears.

Then compute the value of information by layer, target the survey at the links whose decision status is least certain, and record every layer at every visited link. Finally, report the survey size at which the curve saturates, so the agency knows when to stop.

Before trusting the value-of-information ranking, vary the decision rule, because the ranking here changed in four of thirty variants and every one of the four moved either the weight vector or the aggregation rule rather than the data. And before quoting any figure, separate the sampling noise of the calculation from the geographic variation between areas, because the second is at least twenty times the first in this study and pooling them would have hidden the only spatial finding the paper has. Both checks belong in the sequence and both are easy to leave out.

7.2 What generalizes

The numbers do not generalize. They depend on how badly these particular layers are measured, on a five-indicator additive rule, and on two river basins in southwestern Nigeria.

Three structural results should transfer. That a ranking decision is invariant to layer error common across the network, so global calibration buys little and value accrues with links visited, follows from the algebra of ranking rather than from this dataset. That targeted sampling beats random sampling by a large factor follows from the same place. And that propagating already-measured error is nearly free while collecting new error estimates is not is a fact about where the cost sits in any such program.

The most transferable finding may be the error budget itself. Fifty-five percent of the paper-layer pairs in a thirteen-study program carry no measured error, and the two most-used layers are among the least measured. We would expect a similar audit of any comparable program to return a similar number, and the audit is cheap.

There is a reasonable objection to that expectation. The program audited here is unusual in having measured any layer error at all, and it did so because several of its studies were designed around uncertainty as a subject. A program with different concerns might have measured nothing, in which case its budget would read worse, or might have inherited layers from a national inventory with published accuracy statements, in which case it would read much better. What we can say is that in a setting with no road inventory, where every layer is open data assembled by the analyst, the budget is likely to look like this one, and that is exactly the setting the value-of-information question matters most in.

7.3 Implications for practice

First, measure something. A single duplicate reading of an existing layer, which is what produced the agreement coefficient of 0.041, cost nothing and turned out to be the most valuable input to this entire analysis. Programs should generate at least one error estimate per layer as a matter of routine, because without one the value-of-information calculation cannot be run at all and the honest description of the ranking is that its reliability is unknown.

Second, spend the survey budget where the decision is uncertain, not where the network is convenient, and record every layer at every visit.

Third, do not survey a layer because it is central to the story. Served population is the layer this decision is nominally about, and it is the layer least worth measuring, because the decision is a ranking and the population error largely does not change the order.

7.4 Limitations

Of the fourteen analyst-set quantities listed in Section 4.7, eight are swept across 30 variants, and one conclusion, the full four-layer ordering, does not survive them intact.

A reliability figure is standing in for an accuracy figure. The condition treatment rests on agreement between two readings of the same volunteered source rather than on error against ground truth, and no citable comparison of OpenStreetMap surface tags against field verification exists to close the gap. The kernel therefore measures how far two defensible readings diverge, which is a lower bound on how far either diverges from the road.

The survey is costed at one rate. No published cost or productivity figure exists for enumerating served population or recording seasonal impassability, so all four layers are priced at the condition-survey rate. A survey recording all four almost certainly costs more per link than one recording condition alone, which would move the all-four curve right without changing its shape.

8. Conclusions

We audited the error budget of a thirteen-study transport research program by reading its result files, and found that 23 of 42 paper-layer pairs, 55 percent, carry no measured error at all, with every study carrying at least one such layer and the two most-used layers among the least measured.

Against a rural rehabilitation decision funding the most urgent decile of 7,068 links, and drawing 1,000 truth ensembles from the error the program did measure, 325 of the 707 funded links, 46 percent and 594 km, are links a perfectly informed agency would not have funded.

Two results follow for practice. Propagating the error kernels already measured, with no fieldwork, recovers 34.2 percent of the value of perfect information and is the largest single return in the study. And survey effort should go to market and health access first and to served population at no budget, whose entire value at a complete census is 0.0047, which the Monte Carlo noise of the calculation cannot be told apart from. A hundred visits chosen by decision uncertainty are worth roughly 629 chosen at random, and a visit should record every layer.

The mechanism behind all of it is that a ranking decision is invariant to the part of a layer's error that is common across the network. The access layer's error is 36 percent common on its own scale and 1.5 percent common once it reaches a rank. There is no cheap global calibration for a ranking, and value accrues with links visited.

In a worked case on 800 links in the Niger floodplain, a survey of 423 km costing between GBP 3,902 and 12,697 cuts misdirected funding from 55 links in 80 to 32 and changes 49 of the 80 funded. The free kernel correction alone cuts misdirected links but raises misdirected kilometers, because it pushes long rural links up the list, so an agency budgeting in kilometers should not stop there.

The recommendation we would most like to see adopted is the cheapest one. Measure something about every layer, even by reading the same source twice, because an unquantified layer cannot be defended, cannot be priced, and cannot be told apart from a good one.

We would add one caution to that recommendation, since it is the kind that invites box-ticking. An error estimate produced to satisfy a checklist, on a convenient sample, measuring whatever is easiest to measure, will produce a kernel that propagates confidently and wrongly. The condition figure this paper leans on is useful precisely because it is uncomfortable: it says two defensible readings of the same data barely agree, and it was not produced to reassure anyone. A program adopting this workflow should expect its first honest error estimate to be worse than it hoped, and should treat that as the measurement working.

9. References

1. Ades AE, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. Medical Decision Making. 2004; 24(2):207-227. Doi: https://doi.org/10.1177/0272989X04263162
2. Andrei Dan, Mohammad Arabestani. ASTM D6433 Pavement Condition Index Variability Study. Applied Pavement Technology, 2018.
3. Bala Umar, Olufemi Ajumobi, Abdulaziz Umar. Assessment of Health Service Delivery Parameters in Kano and Zamfara States, Nigeria. BMC Health Services Research. 2020; 20:874. Doi: https://doi.org/10.1186/s12913-020-05722-4
4. Banke-Thomas Aduragbemi, Kerry LM Wong, Francis I Ayomoh, Rokibat O Giwa-Ayedun, Lenka Benova. In Cities, It's Not Far, but it Takes Long: Comparing Estimates of Travel Times to Reach Health Facilities in Nigeria. BMJ Global Health. 2021; 6(1):e004318. Doi: https://doi.org/10.1136/bmjgh-2020-004318
5. Barrington-Leigh Christopher, Adam Millard-Ball. The World's User-Generated Road Map is More Than 80% Complete. PLoS One. 2017; 12(8):e0180698. Doi: https://doi.org/10.1371/journal.pone.0180698
6. Barron Christopher, Pascal Neis, Alexander Zipf. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS. 2014; 18(6):877-895. Doi: https://doi.org/10.1111/tgis.12073
7. Becker William, Michaela Saisana, Paolo Paruolo, Ine Vandecasteele. Weights and Importance in Composite Indicators: Closing the Gap. Ecological Indicators. 2017; 80:12-22. Doi: https://doi.org/10.1016/j.ecolind.2017.03.056
8. Bhattacharjya Debarun, Jo Eidsvik, Tapan Mukerji. The Value of Information in Spatial Decision Making. Mathematical Geosciences. 2010; 42(2):141-163. Doi: https://doi.org/10.1007/s11004-009-9256-y

9. Bonafilia Derrick, Beth Tellman, Tyler Anderson, Erica Issenberg. Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, 835-845. Doi: https://doi.org/10.1109/CVPRW50498.2020.00113
10. Brewer Eric, Jason Lin, Peter Kemper, John Hennin, Daniel Runfola. Predicting Road Quality Using High Resolution Satellite Imagery: A Transfer Learning Approach. PLoS One. 2021; 16(7):e0253370. Doi: https://doi.org/10.1371/journal.pone.0253370
11. Brovelli Maria Antonia, Marco Minghini, Monia Molinari, Peter Mooney. Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. Transactions in GIS. 2017; 21(2):191-206. Doi: https://doi.org/10.1111/tgis.12182
12. Cadamuro Gabriel, Aggrey Muhebwa, Jay Taneja. Street Smarts: Measuring Intercity Road Quality Using Deep Learning on Satellite Imagery. Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '19), 2019, 145-154. Doi: https://doi.org/10.1145/3314344.3332493
13. Campalani Piero, Massimiliano Pittore, Kathrin Renner. Assessing OpenStreetMap Roads Fitness-for-Use for Disaster Risk Assessment Applications in Developing Countries: The Case of Burundi. OSF Preprints, 2022. Doi: https://doi.org/10.31730/osf.io/rbhz4
14. Canessa Stefano, Gurutzeta Guillera-Arroita, José J Lahoz-Monfort, *et al*. When do we Need More Data? A Primer on Calculating the Value of Information for Applied Ecologists. Methods in Ecology and Evolution. 2015; 6(10):1219-1228. Doi: https://doi.org/10.1111/2041-210X.12423
15. Chen Yun, Jia Yu, Shahbaz Khan. The Spatial Framework for Weight Sensitivity Analysis in AHP-Based Multi-Criteria Decision Making. Environmental Modelling & Software. 2013; 48:129-140. Doi: https://doi.org/10.1016/j.envsoft.2013.06.010
16. Cohen Jacob. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960; 20(1):37-46. Doi: https://doi.org/10.1177/001316446002000104
17. Congalton Russell G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment. 1991; 37(1):35-46. Doi: https://doi.org/10.1016/0034-4257(91)90048-B
18. Crosetto Michele, Stefano Tarantola. Uncertainty and Sensitivity Analysis: Tools for GIS-Based Model Implementation. International Journal of Geographical Information Science. 2001; 15(5):415-437. Doi: https://doi.org/10.1080/13658810110053125
19. Deluka-Tibljas Aleksandra, Barbara Karleusa, Nevena Dragicevic. Review of Multicriteria-Analysis Methods Application in Decision Making about Transport Infrastructure. Journal of the Croatian Association of Civil Engineers. 2013; 65(7):619-631. Doi: https://doi.org/10.14256/JCE.850.2013
20. Eaton Robert A, Sonja Gerard, Ronald S Dattilo. A Method for Rating Unsurfaced Roads. Transportation Research Record (Washington, DC). 1987; 1106:34-38. https://onlinepubs.trb.org/Onlinepubs/trr/1987/1106v2/1106v2-004.pdf
21. Feizizadeh Bakhtiar, Piotr Jankowski, Thomas Blaschke. A GIS Based Spatially-Explicit Sensitivity and Uncertainty Analysis Approach for Multi-Criteria Decision Analysis. Computers & Geosciences. 2014; 64:81-95. Doi: https://doi.org/10.1016/j.cageo.2013.11.009
22. Foody Giles M. Sample Size Determination for Image Classification Accuracy Assessment and Comparison. International Journal of Remote Sensing. 2009; 30(20):5273-5291. Doi: https://doi.org/10.1080/01431160903130937
23. Foody Giles M. Explaining the Unsuitability of the Kappa Coefficient in the Assessment and Comparison of the Accuracy of Thematic Maps Obtained by Image Classification. Remote Sensing of Environment. 2020; 239:111630. Doi: https://doi.org/10.1016/j.rse.2019.111630
24. Goodchild Michael F, Linna Li. Assuring the Quality of Volunteered Geographic Information. Spatial Statistics. 2012; 1:110-120. Doi: https://doi.org/10.1016/j.spasta.2012.03.002
25. Haklay Mordechai. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. Environment and Planning B: Planning and Design. 2010; 37(4):682-703. Doi: https://doi.org/10.1068/b35097
26. Heath Anna, Ioanna Manolopoulou, Gianluca Baio. A Review of Methods for Analysis of the Expected Value of Information. Medical Decision Making. 2017; 37(7):747-758. Doi: https://doi.org/10.1177/0272989X17697692
27. Heuvelink Gerard BM, Peter A Burrough, Alfred Stein. Propagation of Errors in Spatial Modelling with GIS. International Journal of Geographical Information Systems. 1989; 3(4):303-322. Doi: https://doi.org/10.1080/02693798908941518
28. Hosseini Seyed Amir, Omar Smadi. How Prediction Accuracy can Affect the Decision-Making Process in Pavement Management Systems. Infrastructures. 2021; 6(2):28. Doi: https://doi.org/10.3390/infrastructures6020028
29. Iimi Atsushi, Farhad Ahmed, Edward Charles Anderson, *et al*. New Rural Access Index: Main Determinants and Correlation to Poverty. Policy Research Working Paper No. 7876. The World Bank, 2016. Doi: https://doi.org/10.1596/1813-9450-7876
30. Jean Neal, Marshall Burke, Michael Xie, Matthew Davis W, David B Lobell, Stefano Ermon. Combining Satellite Imagery and Machine Learning to Predict Poverty. Science. 2016; 353(6301):790-794. Doi: https://doi.org/10.1126/science.aaf7894
31. Klopp Jacqueline M, Clemence Cavoli. Mapping Minibuses in Maputo and Nairobi: Engaging Paratransit in Transportation Planning in African Cities. Transport Reviews. 2019; 39(5):657-676. Doi: https://doi.org/10.1080/01441647.2019.1598513
32. Landis J Richard, Gary G Koch. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977; 33(1):159-174. Doi: https://doi.org/10.2307/2529310
33. Landuyt Lisa, Niko EC Verhoest, Frieke MB, Van Coillie. Flood Mapping in Vegetated Areas Using an Unsupervised Clustering Approach on Sentinel-1 and -2 Imagery. Remote Sensing. 2020; 12(21):3611. Doi:

https://doi.org/10.3390/rs12213611
34. Lawal Olanrewaju, Felix E Anyiam. Modelling Geographic Accessibility to Primary Health Care Facilities: Combining Open Data and Geospatial Analysis. Geo-Spatial Information Science. 2019; 22(3):174-184. Doi: https://doi.org/10.1080/10095020.2019.1645508
35. Ligmann-Zielinska Arika, Piotr Jankowski. Spatially-Explicit Integrated Uncertainty and Sensitivity Analysis of Criteria Weights in Multicriteria Land Suitability Evaluation. Environmental Modelling & Software. 2014; 57:235-247. Doi: https://doi.org/10.1016/j.envsoft.2014.03.007
36. Mahabir Ron, Anthony Stefanidis, Arie Croitoru, Andrew T Crooks, Peggy Agouris. Authoritative and Volunteered Geographical Information in a Developing Country: A Comparative Case Study of Road Datasets in Nairobi, Kenya. ISPRS International Journal of Geo-Information. 2017; 6(1):24. Doi: https://doi.org/10.3390/ijgi6010024
37. Mbabazi Elia. Impact of Unpaved Road Condition on Rural Transport Services. Proceedings of the Institution of Civil Engineers - Municipal Engineer. 2019; 172(4):239-245. Doi: https://doi.org/10.1680/jmuen.18.00048
38. Memarzadeh Milad, Matteo Pozzi. Value of Information in Sequential Decision Making: Component Inspection, Permanent Monitoring and System-Level Scheduling. Reliability Engineering & System Safety. 2016a; 154:137-151. Doi: https://doi.org/10.1016/j.ress.2016.05.014
39. Memarzadeh Milad, Matteo Pozzi. Value of Information in Sequential Decision Making: Component Inspection, Permanent Monitoring and System-Level Scheduling. Reliability Engineering and System Safety. 2016b; 154:137-151. Doi: https://doi.org/10.1016/j.ress.2016.05.014
40. Minghini Marco, Francesco Frassinelli. OpenStreetMap History for Intrinsic Quality Assessment: Is OSM up-to-Date? Open Geospatial Data, Software and Standards. 2019; 4(1):9. Doi: https://doi.org/10.1186/s40965-019-0067-x
41. Nardo Michela, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffmann, Enrico Giovannini. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, 2008. Doi: https://doi.org/10.1787/9789264043466-en
42. OpenStreetMap contributors. OpenStreetMap, n.d. https://www.openstreetmap.org
43. Paruolo Paolo, Michaela Saisana, Andrea Saltelli. Ratings and Rankings: Voodoo or Science? Journal of the Royal Statistical Society: Series A (Statistics in Society). 2013; 176(3):609-634. Doi: https://doi.org/10.1111/j.1467-985X.2012.01059.x
44. Phares Brent M, Benjamin A Graybeal, Dennis D Rolander, Mark E Moore, Glenn A Washer. Reliability and Accuracy of Routine Inspection of Highway Bridges. Transportation Research Record: Journal of the Transportation Research Board. 2001; 1749(1):82-92. Doi: https://doi.org/10.3141/1749-13
45. Quirk Lorcan, José C Matos, Jimmy Murphy, Vikram Pakrashi. Visual Inspection and Bridge Management. Structure and Infrastructure Engineering. 2018; 14(3):320-332. Doi: https://doi.org/10.1080/15732479.2017.1352000
46. Refice Alberto, Marina Zingaro, Annarita D'Addabbo, Marco Chini. Integrating C- and L-Band SAR Imagery for Detailed Flood Monitoring of Remote Vegetated Areas. Water. 2020; 12(10):2745. Doi: https://doi.org/10.3390/w12102745
47. Saisana Michaela, Andrea Saltelli, Stefano Tarantola. Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2005; 168(2):307-323. Doi: https://doi.org/10.1111/j.1467-985X.2005.00350.x
48. Sandamal RMK, Pasindu HR. Applicability of Smartphone-Based Roughness Data for Rural Road Pavement Condition Evaluation. International Journal of Pavement Engineering. 2022; 23(3):663-672. Doi: https://doi.org/10.1080/10298436.2020.1765243
49. Sayers Michael W, Thomas D Gillespie, William DO Paterson. Guidelines for Conducting and Calibrating Road Roughness Measurements. World Bank Technical Paper No. 46. The World Bank, 1986. https://documents1.worldbank.org/curated/en/851131468160775725/pdf/multi-page.pdf
50. Scholz Stefan, Paul Knight, Melanie Eckle, Sabrina Marx, Alexander Zipf. Volunteered Geographic Information for Disaster Risk Reduction-the Missing Maps Approach and Its Potential Within the Red Cross and Red Crescent Movement. Remote Sensing. 2018; 10(8):1239. Doi: https://doi.org/10.3390/rs10081239
51. Senaratne Hansi, Amin Mobasheri, Ahmed Loai Ali, Cristina Capineri, Mordechai Haklay. A Review of Volunteered Geographic Information Quality Assessment Methods. International Journal of Geographical Information Science. 2017; 31(1):139-167. Doi: https://doi.org/10.1080/13658816.2016.1189556
52. Thegeya Aaron, Thomas Mitterling, Arturo Martinez, Joseph Albert Nino Bulan, Ron Lester Durante, Jayzon Mag-atas. Application of Machine Learning Algorithms on Satellite Imagery for Road Quality Monitoring: An Alternative Approach to Road Quality Surveys. {ADB} Economics Working Paper Series No. 675. Asian Development Bank, 2022. Doi: https://doi.org/10.22617/WPS220587-2
53. Thomson Dana R, Andrea E Gaughan, Forrest R Stevens, Greg Yetman, Peter Elias, Robert Chen. Evaluating the Accuracy of Gridded Population Estimates in Slums: A Case Study in Nigeria and Kenya. Urban Science. 2021; 5(2):48. Doi: https://doi.org/10.3390/urbansci5020048
54. Thöns Sebastian. On the Value of Monitoring Information for the Structural Integrity and Risk Management. Computer-Aided Civil and Infrastructure Engineering. 2018; 33(1):79-94. Doi: https://doi.org/10.1111/mice.12332
55. Tsyganskaya Viktoriya, Sandro Martinis, Philip Marzahn, Ralf Ludwig. SAR-Based Detection of Flooded Vegetation: A Review of Characteristics and Approaches. International Journal of Remote Sensing. 2018; 39(8):2255-2293. Doi: https://doi.org/10.1080/01431161.2017.1420938
56. Williams Sarah, Adam White, Peter Waiganjo, Daniel Orwa, Jacqueline Klopp. The Digital Matatu Project:

Using Cell Phones to Create an Open Source Data for Nairobi's Semi-Formal Bus System. Journal of Transport Geography. 2015; 49:39-51. Doi: https://doi.org/10.1016/j.jtrangeo.2015.10.005
57. Wilson Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, Tracy K Teal. Good Enough Practices in Scientific Computing. PLoS Computational Biology. 2017; 13(6):e1005510. Doi: https://doi.org/10.1371/journal.pcbi.1005510
58. Workman Robert. Final Trials Report. No. GEN2070A. TRL Ltd for AfCAP, 2017.
59. Workman Robert. The Use of Appropriate High-Tech Solutions for Road Network and Condition Analysis. GEN2070A Final Report. TRL Ltd for ReCAP/AfCAP, 2018.
60. World Bank. World: Measuring Rural Access, Update 2017/18. No. ACS26526. The World Bank, Transport Global Practice, 2019. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/543621569435525309/world-measuring-rural-access-update-2017-18
61. Zhang Wei-Heng, Da-Gang Lu, Jianjun Qin, Sebastian Thöns, Michael Havbro Faber. Value of Information Analysis in Civil and Infrastructure Engineering: A Review. Journal of Infrastructure Preservation and Resilience. 2021; 2(1):16. Doi: https://doi.org/10.1186/s43065-021-00027-0