Calibrating Free Postcode Boundaries from OpenStreetMap

A Transferable Seed-Density Accuracy Curve (NL, DK, BE), with Italian and Swiss Tests

Ian Ronk
Independent researcher, Amsterdam, The Netherlands
ianronk0@gmail.com

August 2026

Abstract

Postcode boundary polygons are unavailable free of charge for much of Europe: several countries — the Netherlands, Denmark, Belgium, Switzerland, Finland and Norway among them — publish authoritative layers; elsewhere they are sold or absent. Voronoi tessellation of OpenStreetMap (OSM) address points is the natural estimator, but how many address points are needed — and whether the answer transfers across countries — has not been established. We calibrate a single OSM-Voronoi pipeline against national references in two countries (NL CBS PC4 and DK DAGI postnumre, 5,160 polygons), fit a seed-density-to-IoU curve, and test out-of-sample transfer to held-out Belgium (bpost, 1,188 polygons). The asymptote is robust across functional forms (mean matched IoU saturates at $a \approx 0.76$–0.82), while the fitted 0.7-IoU threshold is form-sensitive ($\approx$ 40–110 seeds); per-postcode scatter is wide (point-level $R^2 \approx 0.26$), so the curve calibrates the population mean, not individual polygons. The shape transfers: Belgium reaches mean matched IoU 0.618 (81% coverage) near the curve, though held-out error is roughly double the calibration error. Applied to Italy — where no free authoritative intra-city postcode layer exists — the pipeline produces 4,209 estimated CAP polygons; Milan at 855 seeds per CAP reaches mean IoU 0.783 (an earlier pipeline configuration), and the 2,903 single-CAP municipalities at median 19 seeds land where the curve predicts (mean IoU 0.50). Below saturation the binding constraint is address density — inverse-density power weighting does not improve on standard Voronoi — and each released polygon carries its seed count as a fitness-for-use flag. The transferability claim is scoped to the three flat, high-OSM-completeness countries tested; a Swiss evaluation, where the asymptote falls to 0.64, marks the boundary of the regime.

Keywords: Voronoi diagram, postal code, OpenStreetMap, spatial tessellation, IoU, OSM-derived geometry.

1 Introduction

Postcode-level spatial analyses — choropleths, areal joins, neighbourhood statistics — require postcode boundary *polygons* rather than points. Although the points themselves are abundant — OSM `addr:postcode` tags and national open-address datasets carry them in the millions — a free authoritative polygon layer exists in only a minority of European countries, among them the Netherlands (CBS PC4), Denmark (DAGI postnumre), Belgium (bpost), Switzerland (swisstopo PLZ), Finland (Statistics Finland Paavo) and Norway (Kartverket postnummerområder). A second regime has free but non-authoritative polygons traced directly in OSM as `boundary=postal_code` relations — essentially complete for Germany and Austria (OpenStreetMap contributors, 2026d) — where tracing supersedes estimation. In the remaining third regime — Italy, Spain, Ireland, intra-city France — the boundaries are sold by the national postal authority or a licensed vendor, or simply not published at all. The estimation question of this paper is confined to this third regime.

Voronoi tessellation of the address points is the established estimator for those missing boundaries, not a new idea: it is the documented production method behind Ordnance Survey's commercial Code-Point with Polygons product (Thiessen polygons over address points, sold as "notional" boundaries; Ordnance Survey, 2024), it has been run end-to-end on OSM addresses for German postcode areas with reference validation (Hauck, 2011), and open implementations exist for Great Britain (Longair, 2021) and Switzerland (Metashapes, 2017), the latter explicitly proposing Italy as a target. The peer-reviewed lineage of Voronoi region approximation runs back to Alani et al. (2001). What none of this prior work establishes is the question a user of such a layer actually faces: **how many address points are needed for a given accuracy, and does the answer transfer across countries?**

1.1 Motivating context: Italy

Italy illustrates the problem concisely. Its 41 officially-zoned cities (Poste Italiane, 2026b) are divided into multiple *Codici di Avviamento Postale* (CAP) zones, but the boundary layer is sold by Poste Italiane under a per-city commercial licence (Poste Italiane, 2026a); no free authoritative source exists. The only freely available shapes are community traces such as the CAP DI MILANO uMap for Milan (38 polygons, 2017), and OSM-Voronoi reconstructions such as Zornade's *CAP subcomunali Italia* (Zornade, 2025). Zornade demonstrates clear demand for such a layer but does not establish a required address-count threshold or cross-country transferability; we use Zornade's parcel-based v2 layer as a non-circular reference in §5.5, and leave a complete accuracy assessment of the Zornade layers themselves for future work.

1.2 Contribution

The contribution is accordingly not the pipeline but its *calibration*. We calibrate a single OSM-Voronoi pipeline against free authoritative references in two countries (NL CBS PC4 and DK DAGI postnumre, 5,160 reference polygons combined, 5,140 evaluated per-postcode pairs), fit an empirical seed-density-to-IoU relation, and test whether it transfers to a third country held out from the curve fitting (Belgium bpost, 1,188 polygons). The curve can be read as an empirical test of the Poisson–Voronoi approximation theory of Khmaladze and Toronjadze (2001) and Heveling and Reitzner (2009) — which predicts approximation error scaling with boundary length and point intensity — under conditions the theory does not cover: clustered non-Poisson seeds, label noise, and import-driven completeness variation. We then apply the pipeline to Italian CAPs and use the calibrated curve to bound expected accuracy. Released GeoJSON polygons for NL, DK, CH and IT carry per-polygon seed counts as a coverage quality flag.

2 Theory

2.1 Voronoi diagrams

Given a finite set of seed points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^2$, the Voronoi cell of seed p_i is

$$V_i = \{x \in \mathbb{R}^2 : \|x - p_i\| \leq \|x - p_j\| \; \forall\, j \neq i\}.$$

The collection $\{V_i\}$ partitions the plane up to a measure-zero boundary (Aurenhammer, 1991). We compute the diagram using SciPy's `scipy.spatial.Voronoi`, which wraps the Qhull library (Barber et al., 1996) via a lifted convex-hull construction in $O(n \log n)$ (Edelsbrunner and Seidel, 1986).

2.2 Power (weighted) diagrams

Uneven seed density pushes boundaries toward sparse areas: dense postcodes get many small cells while sparse ones get few large ones. The power diagram of Aurenhammer (1987) corrects this by assigning a weight w_i to each seed, replacing Euclidean distance with the power distance

$$d_{\text{pow}}(x, p_i) = \|x - p_i\|^2 - w_i^2.$$

The bisector between two seeds remains a straight line, but its position shifts away from the heavier seed, so sparsely-surrounded seeds can be given larger weights to claim more territory. The power diagram is computed in $O(n \log n)$ via a 3D convex-hull lifting. The headline pipeline uses the standard ($w_i \equiv 0$) construction; we assess whether inverse-density weighting improves on it in §5.6.

2.3 IoU as a polygon overlap metric

IoU measures the fractional overlap between two polygons: for estimated polygon A and reference polygon B,

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \in [0, 1],$$

the standard region-overlap measure (Jaccard index, Jaccard, 1901; the GIS-native treatment of region similarity is Winter, 2000; used as IoU in computer vision, Rezatofighi et al., 2019). We adopt IoU > 0.7 as the success threshold throughout — a computer-vision convention, and two of its properties matter for reading the results. IoU is area-relative: a large rural postcode can score 0.9 with kilometre-scale boundary errors, while a small urban CAP needs near-street precision to reach the same score. And the per-postcode seed count we regress on conflates address density with polygon area — part of the seed-to-IoU relation is plausibly a polygon-size effect rather than a pure density effect; we flag this as a limitation rather than attempt to separate the two.

3 Data

Three types of input feed the pipeline: OSM address points as seeds, authoritative postcode-boundary polygons for calibration and validation, and administrative boundaries as clip targets. Table 1 lists all datasets.

OSM `addr:postcode` points were downloaded in May 2026 via the Overpass API (OpenStreetMap contributors, 2026c) for the Netherlands, Belgium, Italy and Switzerland, and extracted from a Geofabrik PBF extract (Geofabrik GmbH, 2026) for Denmark (17.3 M addresses combined). The Netherlands (CBS PC4, 4,071 polygons (Centraal Bureau voor de Statistiek, 2024)) and Denmark (DAGI postnumre, 1,089 polygons (Klimadatastyrelsen / DAWA, 2026)) supply the calibration references. Belgium (bpost-sourced `georef-belgium-postal-codes`, 2021 vintage, geometry attributed to NGI–IGN, redistributed via Opendatasoft's georef repository; 1,188 polygons (bpost / Opendatasoft, 2021)) is held out from the curve fitting; one Belgian tile enters the (k, τ) operating-point sweep (§4.1), so the transfer is blind on the curve but not on the operating point. For Italy, ISTAT 2024 comune boundaries (Istituto Nazionale di Statistica (ISTAT), 2024) serve as the municipal-CAP reference, and the CAP DI MILANO uMap (uMap contributor, 2017) (community-produced, no authoritative commercial reference can be redistributed) provides an indicative reference for the multi-CAP Milan regime.

Figures 1 and 2 show Milan as a representative sample. Coverage is highly non-uniform — central CAPs carry roughly 3,900 addresses while peripheral ones fall below 100 — and the same pattern, at country-specific levels, holds across all four countries; we return to its consequences for the low-seed regime in §5.3. One completeness asymmetry deserves explicit note: the Dutch OSM address layer is a systematic import of the national BAG register (OpenStreetMap contributors, 2026b), ∼96–99% complete by the REGIO-OSM housenumber evaluation (regio-osm.de, 2026) — data imports shape OSM completeness patterns more generally (Zielstra et al., 2013; Barron et al., 2014) — so the NL calibration partly measures Voronoi-versus-reference error on quasi-authoritative input. Belgium, whose OSM addresses are mapper-mediated rather than bulk-imported (the Flemish CRAB address import is executed manually by the community; OpenStreetMap contributors, 2026a), is therefore the more demanding test of the curve.

Dataset	Role	Size
OSM `addr:postcode` (2026)	Seeds (NL/DK/BE/IT)	17.3 M
CBS Postcode4 (2024)	NL calibration reference	4,071
DAGI postnumre / DAWA (2026)	DK calibration reference	1,089
bpost (`georef-belgium-postal-codes`) (2021)	BE held-out reference	1,188
ISTAT 2024 comune boundaries (2024)	IT municipal-CAP reference	3,460
swisstopo PLZ (2026)	CH out-of-regime reference	3,190
CAP di Milano uMap (2017)	Milan indicative reference	38
Comune di Milano boundary (2024)	Italian clip target	1

Table 1: Datasets used in calibration, validation and application. NL and DK provide calibration; BE is held out; IT national uses ISTAT comuni for the municipal-CAP subset; Milan's uMap serves as an indicative reference for the multi-CAP regime.

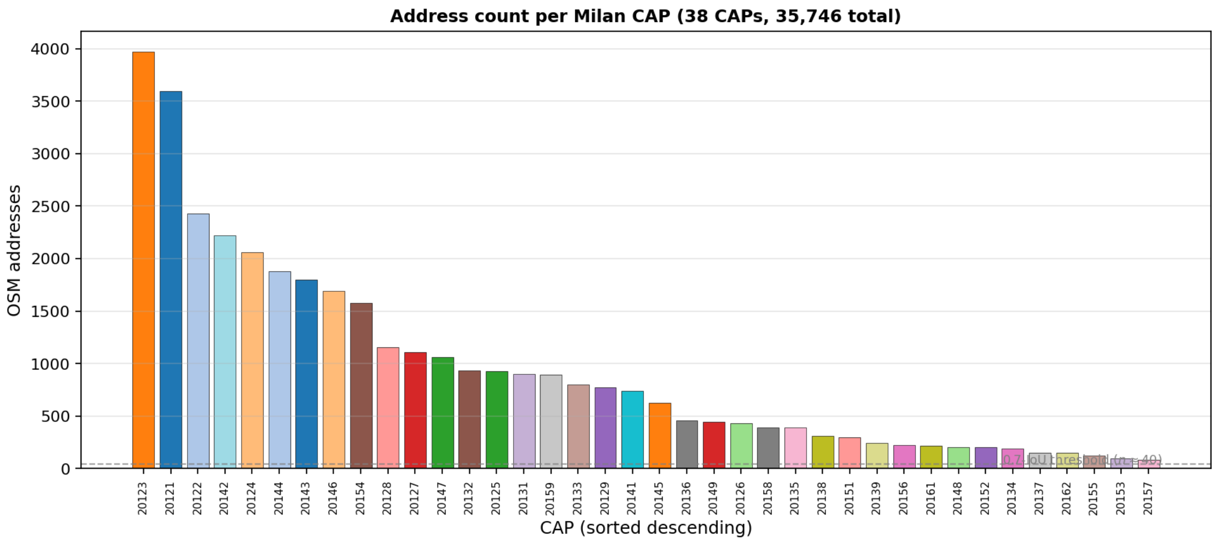

Figure 1: OSM address count per Milan CAP, sorted descending. The dashed line marks the fitted 0.7-IoU seed-count threshold (≈ 40 seeds under the pooled fit; §4.2).

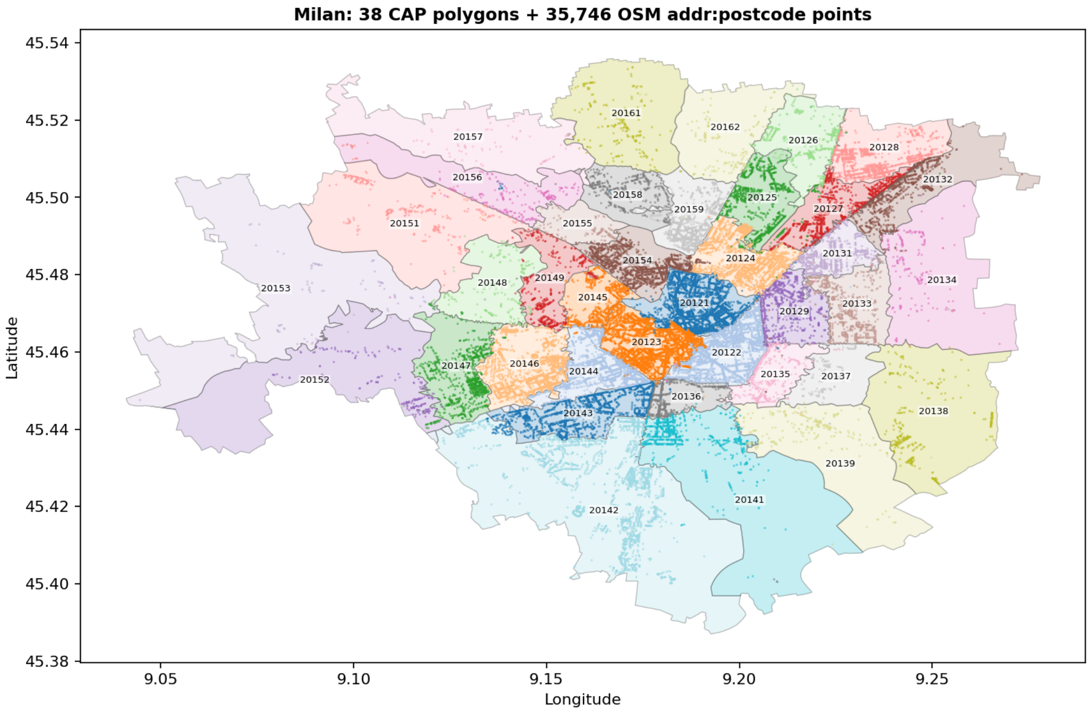

Figure 2: Milan: the 38 uMap CAP polygons with 35,746 OSM `addr:postcode` points overlaid (one colour per CAP). The figure uses a later, larger OSM extract than the pipeline run of §5.4 (33,476 input addresses); the counts differ for that reason.

4 Methods

4.1 The full pipeline

Each OSM address point is a Voronoi seed. The pipeline runs in four steps:

1. **kNN outlier removal.** Flag a point as an outlier if fewer than a fraction τ of its k nearest neighbours share its postcode label:

$$\text{outlier}(i) \iff \frac{1}{k} \sum_{j \in N_k(i)} \mathbb{1}\left[\,\text{pc}(j) = \text{pc}(i)\right] < \tau.$$

   Flagged points are removed before tessellation.
2. **Point-level Voronoi.** A SciPy Voronoi over the remaining points produces one cell per address; each cell inherits its seed's postcode label.
3. **Consolidation.** Cells are dissolved by postcode label into one (Multi)Polygon per postcode. Fragments are then: *(i)* exploded into single-part pieces; *(ii)* the largest fragment per postcode is kept as canonical; *(iii)* remaining fragments are reassigned to the neighbour sharing the longest border; *(iv)* residual slivers are filled with the nearest postcode.
4. **Tiled country runs.** The bounding box is divided into tiles (NL 6×4, DK 3×3, BE 2×2, IT 5×4) and the pipeline runs per tile against the reference polygons whose centroid lies in that tile. The clip target for each tile is the union of its in-tile reference polygons, simplified at $0.0002°$ (≈ 15–$20\,\text{m}$) with a topology clean-up. Per-postcode IoU is collected across all tiles and aggregated nationally. Two ways the reference enters beyond clipping deserve naming: seeds whose postcode is not in the reference list are dropped before tessellation (2.5% of NL and CH seeds; a deployment with no reference cannot perform this step), and interior tile seams place short segments of true reference boundary into the estimate — worth +0.9 IoU points on the BE matched mean (tiled vs. untiled reruns, `iou_results_country_BE_untiled_*.csv`);

the same mechanism operates in the other tiled runs. Calibration IoU is therefore slightly optimistic relative to a fully reference-free deployment; the release metadata records both roles.

Coordinate system. All computations run in unprojected geographic coordinates (EPSG:4326). At 41–57°N this imposes a country-specific ≈1.3–1.8× north–south/east–west anisotropy on the effective distance metric, which the empirical calibration absorbs; the reported IoU values are unaffected (IoU is invariant under the locally linear degree-to-metre map). The calibration therefore measures the pipeline as built — degree-space and metric-space tessellations are not geometrically identical — and metric or geodesic re-projection is listed as future work.

(k, τ) selection. We run a 3×3 sweep over $k \in \{3, 5, 10\}$ and $\tau \in \{0.3, 0.5, 0.7\}$ on representative tiles from NL (calibration), BE and IT (Fig. 3). $(k, \tau) = (3, 0.3)$ is the best cell in all three; NL is nearly flat (spread 0.042), while BE and IT are more sensitive but share the same best cell. Because a Belgian tile entered this preprocessing sweep, Belgium is held out from the curve *fitting* (§4.2) but not from the operating-point check; we flag the distinction here rather than claim a fully blind transfer, and the seed-density curve itself — the object the transfer test evaluates — is fit on NL and DK alone. The sweep informed the operating point used for the BE run and the canonical tuned Italian release; the NL/DK calibration runs, the released NL, DK and CH layers, the Milan run, and the Swiss evaluation all predate it and use the legacy $(5, 0.5)$ setting (§5.1), as recorded in the per-country release metadata.

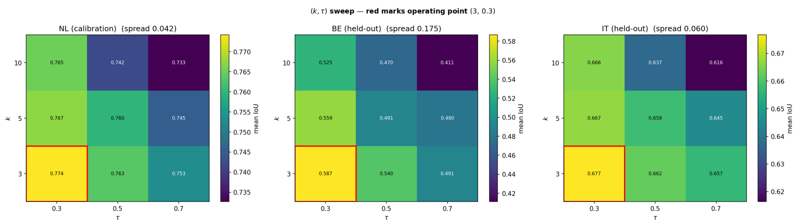

Figure 3: (k, τ) sensitivity sweep on representative NL (calibration), BE and IT tiles. Red marks the operating point $(3, 0.3)$.

4.2 The seed-density-to-IoU relation

Per-postcode (seed count, IoU) pairs from NL and DK form the calibration scatter. We log-bin by seed count and fit the bin medians with a saturating exponential

$$\mathrm{IoU}(n) = a(1 - e^{-n/b}),$$

where a is the asymptote and b the characteristic (e-folding) scale: IoU reaches $0.632\,a$ at $n = b$ and half the asymptote at $b \ln 2 \approx 11$ seeds. We headline this two-parameter form for interpretability — the asymptote and the characteristic scale map directly onto the two quantities a user cares about (the best attainable overlap and the seed count at which returns diminish) — and for its consistency with the diminishing-returns intuition of Poisson–Voronoi approximation (§1). The exponential tracks the bin medians well (bin-median $R^2 = 0.86$); per-postcode scatter is wide ($R^2 = 0.26$, RMSE 0.16), so the curve models central tendency, not individual postcodes.

Functional form. The saturating exponential is not, however, the best-fitting family. Fitting four families to the pooled NL+DK data and comparing on an AIC-style penalized point-level residual criterion (`r3_curve_uncertainty.py`; Table 2), the Michaelis–Menten and three-parameter Hill forms fit better, with the saturating exponential trailing by $\Delta\mathrm{AIC} \approx 390$. Because each family is fit to the log-binned medians rather than by point-level maximum likelihood, the criterion is a penalized point-level RSS rather than a calibrated AIC: magnitudes are indicative only (visibly so — the nested three-parameter exponential scores worse than its two-parameter special case, which cannot happen under point-level ML), the point-level RMSE column carries the same ordering, and spatial dependence among neighbouring postcodes further inflates nominal differences. The choice of family matters for the *threshold* but not for the asymptote: the asymptote is stable across all four forms ($a = 0.76$–0.82, the better-fitting forms sitting at the upper end), whereas the fitted 0.7-IoU crossing ranges from ≈ 40 seeds (saturating exponential) to ≈ 85–110 (Michaelis–Menten, Hill). We therefore retain the saturating exponential as an interpretable summary but treat the asymptote as a family-robust *band* (0.76–0.82) and the threshold as form-dependent, not a point estimate.

Family	Asymptote a	Point RMSE	ΔAIC
Hill (3-param)	0.817	0.152	0.0
Michaelis–Menten	0.801	0.153	39.1
Saturating exp. (headline)	0.763	0.158	387.3
3-param saturating exp.	0.755	0.159	447.9

Table 2: Curve-family comparison on pooled NL+DK point-level data ($n = 5{,}140$): each family is fit to the log-binned medians and the AIC-style criterion is evaluated on the per-postcode residuals (ΔAIC relative to the best family; a penalized point-level RSS, not a calibrated AIC — parameters are not point-level ML, so magnitudes are indicative and the RMSE column carries the same ordering). The asymptote we report is stable across families (0.76–0.82); the headlined saturating exponential is favoured for interpretability rather than fit. Source: `r3_curve_uncertainty.py`.

Pooling. For the scale parameter b, a hierarchical fit — shared asymptote, country-specific scale (the `hierarchical` block of `per_country_fits.json`: $a = 0.748$, $b_{\mathrm{NL}} = 27.6$, $b_{\mathrm{DK}} = 15.0$) — fits better than the pooled single curve ($\Delta\mathrm{AIC} = 149.5$ in favour of country-specific b); fitting each country fully independently reverses the b ordering ($b_{\mathrm{NL}} = 15.4$, $b_{\mathrm{DK}} = 18.9$) because the independent fits trade asymptote against scale. The parametrizations agree on the asymptote band. We nonetheless report the pooled curve as a deliberate single-shape simplification — the object the transfer test needs is one portable curve, not a per-country family — while noting that b is country-dependent and that the pooled fit is, by AIC, a simplification rather than the preferred model.

4.3 Validation and application

Belgium contributes nothing to the curve fit; its only upstream appearance is as one tile in the (k, τ) sweep of §4.1, which we carry as a partial-blindness caveat on the transfer test. We run the pipeline against the bpost reference (1,188 polygons) as the out-of-sample test of the fitted curve. For Italy, the selected (k=3, τ=0.3) pipeline is applied to a national OSM snapshot (3.05 M addresses); the municipal-CAP subset (3,460 CAPs mapping 1:1 to ISTAT comuni) provides an authoritative reference. Milan is scored against the CAP DI MILANO uMap (uMap contributor, 2017) (indicative reference; the Poste Italiane CAP Zone is commercial) clipped to the official comune boundary (Comune di Milano, 2024).

5 Results

5.1 Seed-count vs IoU: calibration on NL and DK

NL (n=4,070) and DK (n=1,070) trace the same saturating shape (Fig. 4). The canonical fit ($a = 0.763$, $b = 16.06$) reaches IoU 0.5 at $n{\approx}17$ seeds, IoU 0.7 at $n{\approx}40$, and asymptotes near 0.76. Per-postcode scatter is wide ($R^2{\approx}0.26$): raw bin medians do not cross 0.7 until $n{\approx}300$, so the 40-seed crossing is a fitted-curve threshold, not a raw-data observation. Both countries are predominantly saturated (median seeds $\approx 1{,}600$ in NL and ≈ 590 in DK, both far above the $\approx$ 40-seed threshold), and their mean IoU reflects this (NL 0.733, DK 0.675). Per-country asymptotes are compatible — NL CI $[0.73, 0.76]$, DK $[0.75, 0.81]$ — so we treat the calibration as a single shape with asymptote ≈ 0.76. One caveat on these intervals: the cluster bootstrap resamples postcodes as independent units, and spatial dependence among neighbouring postcodes makes the intervals anti-conservative — overlap is therefore a compatibility check, not a formal test (a spatial block bootstrap is future work; the pooled two-country bootstrap is degenerate for b and we do not report it as an interval).

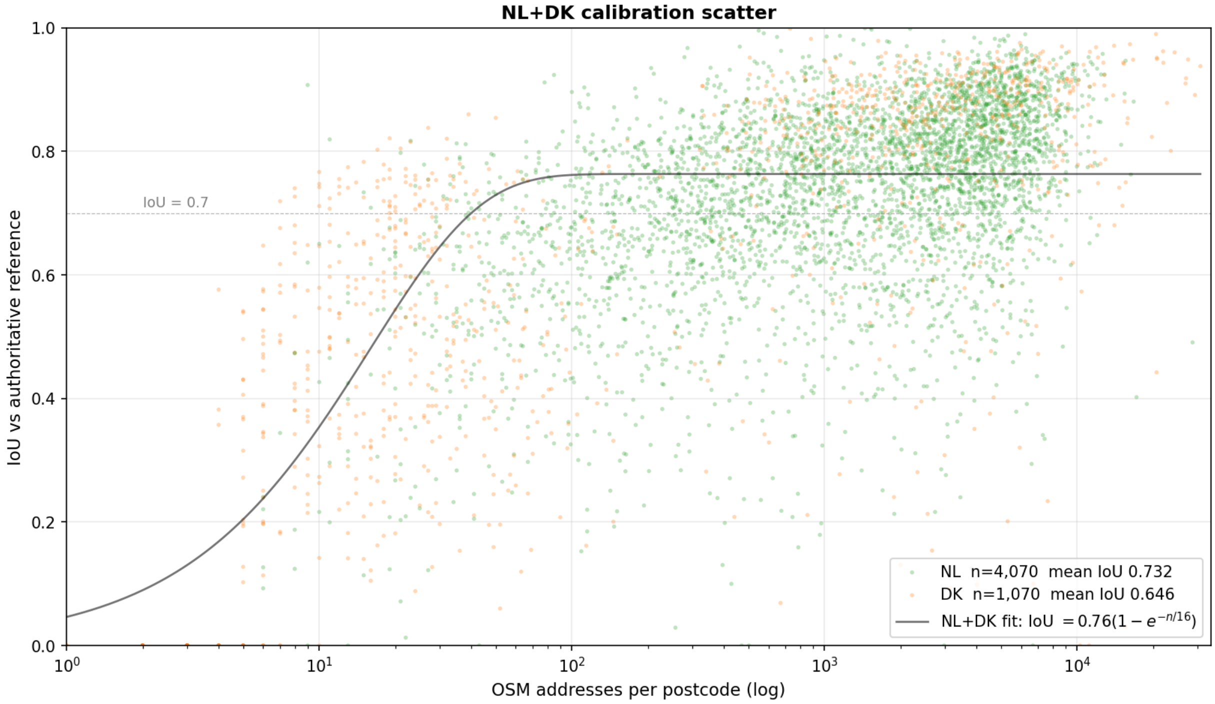

Figure 4: Per-postcode seed count vs IoU in NL (green) and DK (orange), with the saturating fit (black). NL and DK also appear as light-colour background in Figures 5 and 6.

Country	Unique PCs	Seeds/PC	Cov.	IoU (matched)	All-PC IoU	> 0.7 rate
NL	4,071	2,399	99.9%	0.733	0.732	66%
DK	1,070	2,365	95.8%	0.675	0.646	53%

Table 3: Calibration results per country. These runs — and hence the fitted curve — are the frozen legacy ($k = 5, \tau = 0.5$) runs, predating the operating-point selection of §4.1 (reported as committed; the NL sweep is nearly flat, spread 0.042, bounding the expected difference). Seeds/PC is the mean seed count per evaluated postcode. For DK, 19 of the 1,089 DAGI reference polygons drop out before evaluation, leaving 1,070 evaluated postcodes; coverage (95.8%) and the > 0.7 rate use the 1,070 denominator, while IoU (matched) averages matched polygons only (matched-only > 0.7 rate: 55%). All-PC IoU scores unmatched postcodes as 0. One NL polygon additionally drops from the seed–IoU scatter, giving the fit $n = 4{,}070 + 1{,}070 = 5{,}140$.

Baseline. To show what the multi-seed pipeline buys over a trivial alternative, we compare it against a one-centroid-per-postcode Voronoi — a single seed at each reference centroid, the minimal way to turn known postcode points into a tessellation (`run_country_centroid_baseline.py`; Table 4). The full pipeline improves on this baseline in every country, by +0.28 (NL) and +0.30 (DK) on the calibration pair and by smaller margins on the held-out and out-of-regime cases (CH is the out-of-regime Swiss evaluation, reported in §6). The margin is smallest on the out-of-regime

Swiss case and on Belgium, the organically-mapped held-out country, where the centroid baseline also matches more postcodes (1,046 vs 957) at lower IoU — i.e. on the most demanding test the address-density machinery trades a little coverage for a $\approx$ 10-point quality gain.

Country	**Centroid Voronoi**	**Full pipeline**	**Δ**
NL	0.456	0.733	+0.277
DK	0.373	0.675	+0.302
BE	0.519	0.618	+0.099
CH	0.489	0.549	+0.060

Table 4: Mean matched IoU of a one-centroid-per-postcode Voronoi baseline versus the full kNN + Voronoi + consolidation pipeline. The multi-seed pipeline improves on the trivial baseline everywhere; the margin is smallest on Switzerland (out of regime) and Belgium (the organically-mapped held-out case). Source: `run_country_centroid_baseline.py`.

5.2 Out-of-sample test on Belgium

Running the pipeline on Belgium against the bpost reference (1,188 polygons) produces 957 matched polygons (81% coverage).[1] The reference is a 2021 vintage while the OSM snapshot is May 2026; Belgian postcode boundaries change rarely, but this five-year gap is part of the held-out error. Mean matched IoU is 0.618 (median 0.636); the > 0.7 rate is 40%, against 66% in NL. Counting evaluated-but-unmatched postcodes as zero, the mean over all 1,127 evaluated reference polygons is 0.525 (`be_holdout_summary.json`) — the same matched-vs-all convention split reported for the other countries. One provenance note: the Belgian run uses the selected $(3, 0.3)$ operating point while the calibration curve is fit on the legacy $(5, 0.5)$ NL/DK runs (§5.1), so the transfer test compares across operating points as committed. For completeness, the frozen $(5, 0.5)$ Belgian run scores mean matched IoU 0.576 at 69% coverage (818/1,188; `iou_results_country_BE.csv`) — 4.2 IoU points and 12 coverage points below the selected operating point. The partial-blindness caveat of §4.1 is also bounded from the other side: the NL and IT sweep tiles each independently select $(3, 0.3)$, so a Belgium-blind selection would have produced the same 0.618 headline. The held-out RMSE on matched postcodes against the calibration fit is 0.306, bias +0.071 (residuals are IoU minus prediction, so the positive bias says the curve slightly under-predicts Belgium; both statistics are matched-only, `be_holdout_summary.json`). Held-out error is roughly double the calibration RMSE of 0.158 — the shape transfers; the per-postcode precision degrades.

Figure 5 overlays the Belgian matched postcodes on the NL/DK calibration curve: the saturating shape transfers out of sample.

[1] The bpost reference duplicates one postcode (0612), whose second row carries a zero-area geometry and is never evaluated, so the result CSVs carry 1,187 rows; coverage and counts here use the 1,188-polygon reference as denominator.

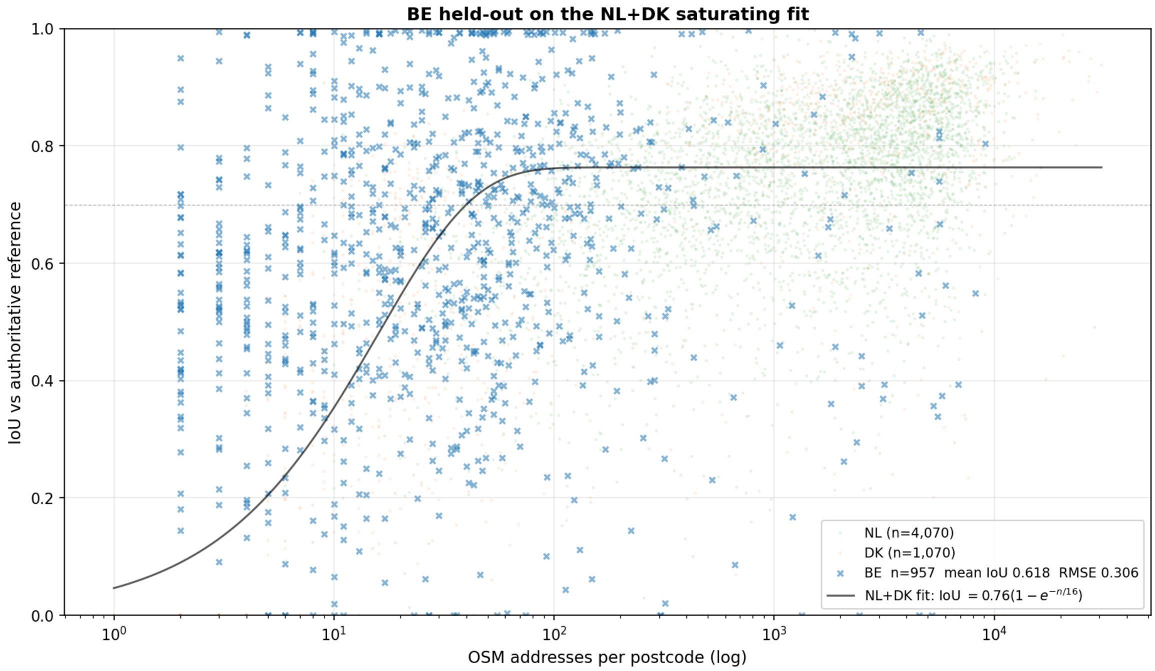

Figure 5: Belgium matched postcodes (blue) overlaid on the NL/DK calibration curve (black), with NL and DK scatter in light colour.

5.3 Curve validation at low seed density: Italian municipal CAPs

The single-CAP comuni — 3,460 of Italy's CAPs map 1:1 to an ISTAT comune, whose boundary is the postcode delivery zone by definition — provide a free reference that NL and DK never test: the low-seed regime. Most Italian comuni are small; at median 19 seeds per CAP the pipeline operates well below the ≈ 40-seed threshold. This extends the calibration curve's tested range downward and checks whether it extrapolates correctly outside NL and DK.

Applied to a national OSM snapshot (3.05 M addresses), the pipeline produces 4,209 CAP polygons. The municipal subset (3,460 pairs) matches 2,903 comuni (84% coverage), with mean IoU 0.500 (median 0.514). The curve predicts this: RMSE against the NL+DK fit is 0.242 with near-zero bias (+0.004). Figure 6 shows the result: the Italian scatter follows the calibration shape, consistent with the curve extrapolating correctly into the unsaturated regime.

The stated problem (§1) — sub-comunal boundaries in the 41 officially-zoned multi-CAP cities (Poste Italiane, 2026b) — is not covered by ISTAT comuni. For those ≈613 polygons, no free authoritative reference exists except Milan. Milan is therefore the deployment target, addressed in Section 5.4.

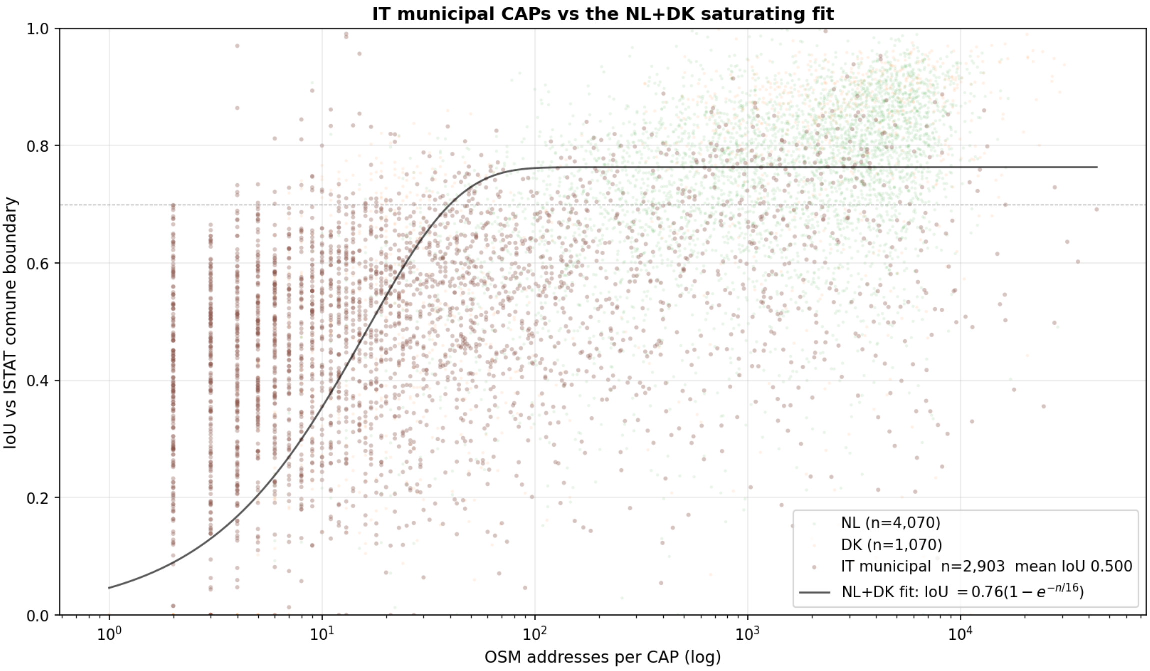

Figure 6: Italian municipal CAPs ($n = 2{,}903$, brown) against the calibration curve (black), with NL and DK scatter in light colour.

5.4 Milan: pipeline performance

The pipeline runs on 32,509 post-kNN seeds clipped to the official Comune di Milano boundary and scored against the 38-polygon CAP DI MILANO uMap layer. Table 5 and Figure 7 report the results. Mean IoU is 0.783 (median 0.798); 30/38 CAPs (79%) clear the 0.7 threshold and 19/38 (50%) clear 0.8. One provenance note: this run predates the operating-point selection and used $(k, \tau) = (5, 0.5)$, not the later $(3, 0.3)$. The $(5, 0.5)$ setting was originally chosen on Milan itself and then held fixed for the country runs, so Milan is not a tuning-blind test. Committed nearby-parameter variants score within half a point (0.783–0.789) — the headline is not sensitive to the setting — but the table reports the $(5, 0.5)$ artifact as committed, not a $(3, 0.3)$ re-run. The headline figure is the `"standard"` block of `data/results/milan_weighted_summary.json` (mean 0.7834, 30/38 $>$ 0.7; the similarly named `data/results/iou_results_it_milano.csv` is an unclipped national-snapshot export, mean 0.26, not this result). The score is reference-dependent: scored against Zornade's independently produced CAP layer (mixed cadastral/OSM sources; Zornade, 2025) rather than the uMap, the same Milan output agrees at mean IoU 0.63 under the comune-boundary clip — the gap reflects disagreement between community references and clipping conventions, not a change in the estimate; the envelope-matched comparison against the cadastral v2 layer gives 0.76 (§5.5).

Mean IoU	Median IoU	Min IoU	Max IoU
0.783	0.798	0.541	0.936

Table 5: Milan agreement with the CAP DI MILANO uMap (indicative reference). Full pipeline (kNN + Voronoi + consolidation), 38 CAPs.

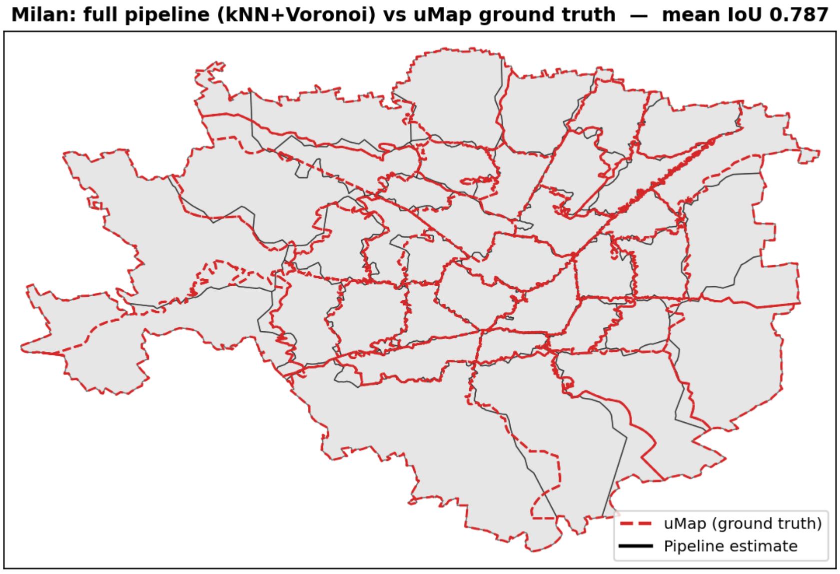

Figure 7: Milan: estimated CAP polygons (filled gray) against the uMap reference (red dashed).

5.5 Non-circular cross-check: cadastral-parcel CAP zones

Every multi-CAP reference so far is community-traced or OSM-adjacent, inviting a circularity concern: OSM-derived estimates scored against OSM-informed references. Zornade's v2 layer (Zornade, 2026) rebuilds intra-city CAP zones from *cadastral parcels* (Agenzia delle Entrate), an independently constructed, non-OSM reference. Scoring the frozen pipeline against it in the four largest multi-CAP cities, with estimate and reference clipped to the same envelope (`multicity_parcel_xcheck.py`), Milano (687 median seeds per CAP) reaches mean IoU 0.764 — on the calibrated asymptote (0.763) — while the sparser cities fall below the curve mean: Roma (295 median seeds) 0.632, Napoli (26) 0.520, Genova (49) 0.498, residuals -0.09 to -0.23 against the pooled fit. Read together, the four cities show the asymptote holding against a non-OSM reference at saturation, while below saturation intra-city CAPs land *below* the curve — consistent with the area-relativity caveat of §2 (small urban zones need near-street precision to reach a given IoU) — so for dense multi-CAP cities the curve is best read as an optimistic mean, not a lower bound.

5.6 Side experiment: power-diagram weighting does not improve on standard Voronoi

We test inverse-density weighting — assigning larger weights to rural seeds so they expand against urban neighbours — across Milan, Amsterdam and the Italian national pipeline by sweeping the weight scale α under two schemes (per-address and per-postcode). The best weighted variant gains at most 0.5–0.6 pp mean IoU over standard Voronoi in a narrow operating window and degrades quickly outside it; for the IT national run there is no gain at any α. The weighted-versus-standard comparisons are made within the weighted implementation ($\alpha = 0$ versus $\alpha > 0$), whose unweighted baseline differs from the headline SciPy pipeline by city (-4.3 pp on Amsterdam, $+0.2$ pp on Milan), so the like-for-like contrast is internal to one implementation. The null is plausibly because postal boundaries already bias polygons toward high-delivery seeds, so weighting in the opposite direction overshoots — a mechanism we state as hypothesis, not test. This is a negative result for the inverse-density family; volume-proxy weighting (POI count, building-footprint area, or residential-address density) is the natural next step and is left for future work.

6 Discussion

Seed density is the binding constraint. Per-country asymptotes are compatible (NL 0.74, DK 0.79; overlapping bootstrap CIs), and the fitted 0.7-crossing lies between ≈40 seeds (pooled saturating-exponential fit) and ≈85–110 once the hierarchical fit ($b_{\mathrm{NL}} \approx 28$; for scale, the independent NL fit gives $b = 15.4$ with bootstrap CI $[9.8, 25.1]$ — no bootstrap was run on the hierarchical parametrization) and the better-fitting Michaelis–Menten/Hill forms (§4.2, Table 2) are admitted — a practical decision rule encoded in the per-polygon seed count, with honest uncertainty on the cut-point. Belgium clusters near the calibration curve with the same saturating slope; the median Italian municipal CAP at 19 seeds lands where the curve predicts (residual bias +0.004), though with wide per-CAP scatter (RMSE 0.242) — the curve calibrates the mean, not individual predictions. Against the non-OSM cadastral reference (§5.5), saturated Milano sits on the asymptote while the sparser intra-city cities fall below the curve mean — in the dense multi-CAP regime the curve is an optimistic mean.

The stability regime has a known boundary, and we report it rather than let the title imply universality. A Swiss evaluation with the same pipeline (legacy $(5, 0.5)$ setting) against the swisstopo PLZ layer (n = 3,141; `per_country_fits.json`) fits an asymptote of 0.644 — well below the NL/DK asymptote CIs — with mean IoU 0.486, and a reference-noise check (independent OSM-traced PLZ vs swisstopo, mean IoU 0.93 on the 75 comparable polygons) confirms the Swiss reference layer is sound. Swiss PLZ are drawn over mountainous, sparsely addressed terrain where boundaries follow ridgelines rather than address geography — plausibly, not demonstrably, the mechanism. The cross-country claim of this paper is therefore scoped: the curve is stable across three flat, high-OSM-completeness Western European countries, and Switzerland marks where that regime ends.

The asymptote stops short of 1 plausibly because Voronoi mediatrices are smooth equidistance curves while real boundaries follow streets and administrative lines — a mechanism with partial support at best: shape complexity (Polsby–Popper inverse) alone explains 24% of residual variance in DK but only 3% in NL.

7 Conclusion

A single OSM-Voronoi pipeline calibrated on 5,140 NL and DK postcode observations produces a seed-density-to-IoU curve whose asymptote is stable across functional forms (mean matched IoU saturating at $a \approx 0.76$–0.82) and whose 0.7-threshold is form-sensitive (≈ 40–110 seeds). The shape transfers to Belgium, held out from the curve fitting (957/1,188 matched, mean IoU 0.618, held-out RMSE roughly double the calibration RMSE). Italy — at a median 19 seeds per municipal CAP — lands where the curve predicts (mean IoU 0.50); Milan at 855 seeds per CAP reaches 0.783. The transferability claim is scoped to the three flat, high-OSM-completeness countries tested; the Swiss evaluation (asymptote 0.644, §6) marks the boundary of the regime. Below saturation, the binding constraint on postcode estimation from OSM addresses is address density rather than the one algorithmic refinement we tested — inverse-density power weighting does not improve on standard Voronoi, while street snapping (future work) plausibly would raise the asymptote. The per-polygon seed count predicts the *population-mean* reliability of an estimate (the curve's point-level R^2 is only 0.26, so an individual polygon remains high-variance around the predicted mean), and is released with every polygon as the single most useful fitness-for-use flag a downstream user has in the absence of a reference layer.

Future work.

1. **Street snapping.** Snap Voronoi mediatrices to the nearest OSM road centreline, converting smooth equidistance curves into street-following boundaries. The geometry mechanism (§6) suggests this would close much of the asymptote-below-1 gap for densely-seeded postcodes.

2. **Additional countries.** The calibration covers two central/northern European countries with similar OSM-completeness profiles. Extending to Southern Europe, non-European references (US, Japan), and countries with different postal-authority constructions would test whether the (a, b) band is universal or region-specific.
3. **Metric-space tessellation.** The pipeline tessellates in unprojected EPSG:4326; re-running in per-country projected CRS (or with geodesic distances) would remove the latitude-dependent anisotropy the calibration currently absorbs, and quantify its geometric effect directly.
4. **Volume-proxy weighting.** The inverse-density power diagram (§5.6) is a negative result. Weighting seeds by a volume proxy — POI count, building-footprint area, or residential-address density — weights with the postal generative direction rather than against it, and may shift the asymptote.

8 Reproducibility

Data and code availability. The pipeline code, the released postcode-boundary polygons (NL, DK, IT, CH), and the redistributable reference layers are archived at [**Zenodo DOI to be minted on submission**] and developed at [**repository URL**]. The released GeoJSON is an OpenStreetMap derivative and is distributed under the Open Database License (ODbL 1.0), © OpenStreetMap contributors; community reference layers (CAP DI MILANO uMap, Zornade) retain their own terms and are attributed accordingly. Not redistributed: the bpost reference layer (rebuilt from the public source by a committed script), the commercial reference layers (Poste Italiane CAP Zone), and the CodAvvPostale CAP shapefile, whose licence and provenance we could not verify. Paths below are relative to the repository root.

Artefacts.

- *Source*: the manuscript under `paper/`.
- *OSM data*: Overpass API dumps (NL, BE, IT, CH) and a Geofabrik PBF extract (DK), May 2026 snapshots, stored under `data/raw/` in the archived deposit. Re-downloading returns current data; the archived snapshots are the canonical source.
- *Authoritative references*: CBS Postcode4 (NL), DAGI postnumre (DK), ISTAT comuni (IT), one folder per country under `data/reference/`. The bpost reference (BE) is not redistributed; it is rebuilt from the public source by `experiments/scripts/fetch_be_reference.py`.
- *Pipeline scripts*: country pipeline runner, Italian national runner, and the (k, τ) sweep, all in `experiments/scripts/`.
- *Analyses*: bootstrap CIs on the curve fit (`r3_curve_uncertainty.py`) and the residual-vs-geometry regression (`r7_completeness_vs_geometry.py`).
- *Parcel cross-check*: `multicity_parcel_xcheck.py`; its per-city CSV is committed under `data/results/`.
- *Per-country IoU results*: one CSV per country in `data/results/iou_results_country_*.csv`.
- *Polygon releases*: NL, DK, IT and CH GeoJSON under `paper/release/` (no BE release: a free bpost-sourced reference layer already circulates; CH is released despite the free swisstopo layer as the out-of-regime demonstrator, with its accuracy caveats documented). The canonical Italian release is `it_estimated_postcodes_v1_tuned.geojson` (4,209 polygons, the $(3, 0.3)$ run quoted in this paper); the un-suffixed `it_estimated_postcodes_v1.geojson` is the earlier $(5, 0.5)$ run (3,647 polygons) and is retained for provenance only.

Hyperparameters. Operating point $k = 3$, $\tau = 0.3$ (selected on the sweep of §4.1: NL, BE and IT tiles, with NL nearly flat), used for the BE run and the canonical tuned Italian release. The NL/DK calibration runs, the released NL, DK and CH GeoJSONs, the Milan run, and the Swiss evaluation predate the sweep and use the legacy $(5, 0.5)$ setting — originally tuned once on Milan and held fixed — as reported in §5.1 and §5.4 and recorded in each `*_release_summary.json`.

Per-tile seed floor 100 addresses, boundary simplification 0.0002°, fragment threshold $< 0.5\%$ of postcode area, tile grids: NL 6×4, DK 3×3, BE 2×2, IT 5×4.

AI usage statement

Claude (Anthropic) was used for writing and editing assistance, statistical-analysis support, and code development and debugging, including audit passes over the manuscript and analysis scripts, in all cases under the author's direction and review. The study design, data collection, results, and conclusions are the author's own; all reported numbers were verified against the committed analysis artifacts.

References

Alani, H., Jones, C. B., and Tudhope, D. (2001). Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, 15(4):287–306.

Aurenhammer, F. (1987). Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96.

Aurenhammer, F. (1991). Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483.

Barron, C., Neis, P., and Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6):877–895.

bpost / Opendatasoft (2021). georef-belgium-postal-codes. https://public.opendatasoft.com/explore/assets/georef-belgium-postal-codes/. Belgian postal-code polygons (2021 vintage, geometry attributed to NGI–IGN), redistributed via Opendatasoft's georef repository; retrieved May 2026.

Bundesamt für Landestopografie swisstopo (2026). Amtliches ortschaftenverzeichnis mit PLZ und perimeter. https://www.swisstopo.admin.ch/de/amtliches-ortschaftenverzeichnis. Swiss PLZ polygon layer (AMTOVZ), retrieved May 2026.

Centraal Bureau voor de Statistiek (2024). CBS postcode4 (PC4) statistical layer. Released via PDOK, https://www.pdok.nl/.

Comune di Milano (2024). Confini amministrativi del comune di milano (dataset `ds2841`, GeoJSON). Open Data portal, https://dati.comune.milano.it/dataset/ds2841-confini-amministrativi-del-comune-di-milano.

Edelsbrunner, H. and Seidel, R. (1986). Voronoi diagrams and arrangements. *Discrete & Computational Geometry*, 1(1):25–44.

Geofabrik GmbH (2026). OpenStreetMap data extracts. https://download.geofabrik.de/. PBF extracts by region, retrieved May 2026.

Hauck, C. (2011). Automatisierte Generierung von Postleitzahlgebieten aus OpenStreetMap-Daten unter Verwendung von Open Source GIS Software. Studienarbeit, Technische Universität Dresden, https://d-nb.info/1067728880/34.

Heveling, M. and Reitzner, M. (2009). Poisson–Voronoi approximation. *Annals of Applied Probability*, 19(2):719–736.

Istituto Nazionale di Statistica (ISTAT) (2024). Confini delle unità amministrative a fini statistici. https://www.istat.it/it/archivio/222527. Generalized comune boundaries, 1 January 2024 edition.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et du jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579.

Khmaladze, E. and Toronjadze, N. (2001). On the almost sure coverage property of Voronoi tessellation: the $\mathbb{R}^1$ case. *Advances in Applied Probability*, 33(4):756–764.

Klimadatastyrelsen / DAWA (2026). Postnumre as GeoJSON. https://api.dataforsyningen.dk/postnumre.

Longair, M. (2021). Open data GB postcode unit boundaries. Voronoi-based open dataset and `postcodes-mapit` service, https://longair.net/blog/2021/08/23/open-data-gb-postcode-unit-boundaries/.

Metashapes (2017). Reconstruction of postal code areas using OpenStreetMap. Blog post, Swiss OSM-Voronoi postcode reconstruction, https://metashapes.com/blog/reconstruction-postal-code-areas-using-openstreetmap/.

OpenStreetMap contributors (2026a). AGIV CRAB import. OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/AGIV_CRAB_Import. Mapper-mediated import of the Flemish CRAB address register; retrieved July 2026.

OpenStreetMap contributors (2026b). BAGimport. OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/BAGimport. Systematic import of the Dutch Basisregistratie Adressen en Gebouwen into OpenStreetMap; retrieved July 2026.

OpenStreetMap contributors (2026c). Overpass API. https://overpass-api.de/. `nwr[addr:postcode]` country queries, retrieved May 2026.

OpenStreetMap contributors (2026d). Tag:boundary=postal_code. OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/Tag:boundary%3Dpostal_code. Community-traced postal-code boundary relations; widely used in Germany, Austria and Belgium; retrieved July 2026.

Ordnance Survey (2024). Code-point with polygons: technical specification. Product documentation; Thiessen polygons generated from ADDRESS-POINT/PAF, https://www.ordnancesurvey.co.uk/products/code-point-polygons.

Poste Italiane (2026a). CAP Zone: dati integrabili in sistemi gis. https://business.poste.it/professionisti-imprese/prodotti/cap-zone-dati-integrabili-in-sistemi-geografici-gis.html.

Poste Italiane (2026b). Elenco dei centri di distribuzione delle città divise in zone CAP. https://business.poste.it/business/files/1476469283476/elenco-centri-distribuzione-citta-divise-in-zone-cap.pdf. 41 zoned cities: 27 original plus 14 added March 2009.

regio-osm.de (2026). Housenumber evaluations for OpenStreetMap. https://regio-osm.de/. Per-municipality OSM housenumber completeness evaluation; retrieved July 2026.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). Generalized Intersection over Union: a metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

uMap contributor (2017). CAP di Milano. OpenStreetMap uMap, https://umap.openstreetmap.fr/en/map/cap-di-milano_169260.

Winter, S. (2000). Location similarity of regions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(3):189–200.

Zielstra, D., Hochmair, H. H., and Neis, P. (2013). Assessing the effect of data imports on the completeness of OpenStreetMap – a United States case study. *Transactions in GIS*, 17(3):315–334.

Zornade (2025). CAP subcomunali italia: dataset gratuito di 9,440 zone postali. https://zornade.com/blog/cap-subcomunali-dataset-italia-download-gratuito-mappe-gis/.

Zornade (2026). CAP subcomunali italia v2: ricostruzione da particelle catastali. https://zornade.com/blog/cap-subcomunali-italia-v2-2026-catasto-download-dataset-gis. Parcel-based reconstruction of 576 intra-city CAP zones across 37 cities (cadastral parcels from Agenzia delle Entrate plus five sources); ODbL. Published 13 March 2026.