Notes on Intuitionistic Fuzzy Sets
Print ISSN 1310–4926, Online ISSN 2367–8283
2026, Volume 32, Number 3, 258–279
DOI: 10.7546/nifs.2026.32.3.258-279

Calibrated circular and elliptical regions for intuitionistic fuzzy assessments: A methodological benchmark of geometry coverage and boundary effects

Panagiotis Chountas [1,*] and Georgy Urumov [2]

[1] School of Computer Science and Engineering, University of Westminster
115 New Cavendish Street, London W1W 6UW, UK
e-mail: chountp@westminster.ac.uk

[2] School of Computer Science and Engineering, University of Westminster
115 New Cavendish Street, London W1W 6UW, UK
e-mail: G.Urumov1@westminster.ac.uk

* *Corresponding Author*

Received: 14 August 2026 **Revised:** 1 October 2026
Accepted: 7 October 2026 **Online First:** 6 November 2026

Abstract: Circular and elliptic intuitionistic fuzzy sets motivate region-based descriptions of membership and non-membership uncertainty. This methodological benchmark compares circles, axis-aligned ellipses and oriented ellipses on the admissible intuitionistic triangle. It combines established geometric identities and split-conformal calibration with a reproducible evaluation of coverage, admissible area and parameter variability. Six main scenarios and training-size and coverage-level sensitivity analyses comprise 4,200 independent synthetic datasets and 12,600 fits. The axis-aligned ellipse reduces mean admissible area by 53.4 percent in the main axial-anisotropy scenario; an oriented baseline captures corresponding gains for tilted variation. A full-versus-admissible-area diagnostic isolates the effect of clipping: in the boundary scenario, the oriented ellipse has a 28.5 percent mean-area saving relative to the circle before clipping and a 25.5 percent saving after clipping. Coverage uncertainty accompanies the efficiency comparisons. The contribution is an assessment-domain benchmark and methodological synthesis; it introduces

Copyright © 2026 by the Authors. This is an Open Access paper distributed under the terms and conditions of the Creative Commons Attribution 4.0 International License (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/

neither a new fuzzy-set algebra nor a new conformal method. Real expert behaviour and downstream decision performance remain outside its evidence.

Keywords: Intuitionistic fuzzy assessments, Circular regions, Elliptical regions, Split conformal calibration, Boundary truncation, Synthetic evaluation.
2020 Mathematics Subject Classification: 03E72, 62G15.

1 Introduction and contribution

An intuitionistic fuzzy assessment describes an object using a membership degree μ and a non-membership degree ν, with $\mu \geq 0$, $\nu \geq 0$ and $\mu + \nu \leq 1$ [3]. Membership expresses support for a specified property, while non-membership expresses opposition to it. The remainder $1 - \mu - \nu$ is the hesitation degree, representing the part not assigned to either judgement. For example, $(\mu, \nu) = (0.60, 0.20)$ leaves hesitation 0.20. These degrees are not automatically probabilities of mutually exclusive events; the interpretation depends on the assessment task. This distinction matters when a statistical model is later introduced for repeated assessments.

A single pair summarizes one assessment, but repeated evaluations of the same object can vary across observations. A geometric region can summarize the spread of such admissible pairs around a representative centre. Circular and elliptic intuitionistic fuzzy sets provide related geometric descriptions and associated operations [4, 5], with distance measures [6] and decision applications [10]. In the comparison adopted here, a circle allows equal spread in every direction; an axis-aligned ellipse permits different membership and non-membership spreads. An oriented ellipse additionally follows tilted variation between the two coordinates.

Figure 1 illustrates these geometric differences using the same coordinate scale in each panel. Equal semiaxes recover a circle; unequal semiaxes allow directional spread, and rotation changes the direction of the principal axes. The constructed points illustrate shape only.

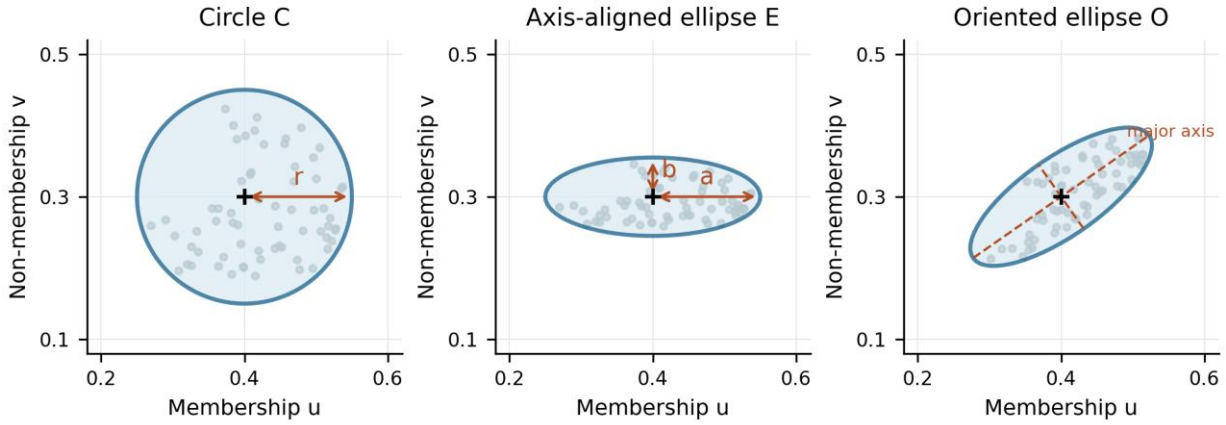


Figure 1. Circular, axis-aligned elliptic and oriented elliptic comparison regions. The centre is (0.40,0.30). The circle has radius 0.15; both ellipses have principal semiaxes 0.15 and 0.055, and the oriented ellipse is rotated by 35 degrees.

The choice of shape creates a practical trade-off. A region should cover new assessments reliably, but should also be compact enough to convey useful information. An excessively large region can achieve high coverage while distinguishing little about the likely assessment values. Conversely, a small region can appear precise while missing many observations. The paper therefore compares area together with held-out coverage at a shared calibration target. Greater geometric flexibility is useful only if its estimated shape improves this balance without introducing excessive instability.

Admissibility adds a further consideration. Every assessment must remain in the triangle defined by non-negative degrees whose sum does not exceed one. A circle or ellipse fitted near its boundary can extend into coordinates that are impossible under this definition. Intersecting the fitted shape with the triangle removes those coordinates and changes its area. Consequently, the area of the full shape can give a different impression of relative efficiency from the area that is actually admissible. This motivates the explicit boundary-clipping analysis in the benchmark.

Unequal spread, skewness, tail thickness and multimodality are different properties. A circle or ellipse specifies a set, rather than a probability law. Allowing two semiaxes represents axis-aligned anisotropy; allowing orientation also represents tilted linear variation. Neither extension exactly describes arbitrary asymmetric or disconnected supports. We therefore avoid a general claim that an elliptic representation handles all non-normal data.

Against this background, the research question is how axis flexibility, orientation and boundary clipping affect admissible area and predictive coverage when all methods share the same centre, calibration observations and test data. The geometric models summarize possible assessment coordinates; probability statements concern the stated sampling model for repeated observations. The experiments address this controlled comparison rather than the full algebra of circular or elliptic intuitionistic fuzzy sets.

1.1 Illustrative application to supplier reliability

Consider repeated assessments of one product supplier against a fixed reliability criterion. Each assessment records support μ and opposition ν; hesitation is $1 - \mu - \nu$. An illustrative pair (0.60,0.20) therefore expresses support 0.60, opposition 0.20 and hesitation 0.20. These values are degrees of judgement, not probabilities of delivery success or failure. Comparability requires a common elicitation scale and criterion; assessments of unrelated suppliers should not be pooled without a justified model.

Suppose positive evidence, such as delivery records, is interpreted consistently, while judgements about negative evidence, such as unresolved quality complaints, vary more. The assessment cloud may then be narrow in membership and wide in non-membership, as illustrated in Figure 2. An axis-aligned ellipse can accommodate these unequal spreads. If support and opposition also vary together, an oriented ellipse can follow the tilted variation. Both shapes in Figure 2 lie entirely inside Δ, so clipping removes nothing: this figure illustrates unequal directional spread. The separate boundary example in Figure 3 illustrates clipping. For supplier assessments near $\mu + \nu = 1$, a fitted shape can extend into coordinates with negative hesitation. Intersecting it with Δ removes those impossible coordinates. The retained area can decrease, but inclusion of any admissible assessment is unchanged. The benchmark therefore compares admissible areas after clipping, which can affect circular and elliptical regions differently.

The benchmark asks whether this flexibility yields a smaller admissible region while preserving comparable coverage of a new assessment under the stated sampling assumptions. In this application, compactness would mean a more precise summary of assessment variability; it would not establish that a supplier is reliable or that a procurement decision improves. Figure 2 uses constructed data only. No supplier records or expert responses were collected, and the example is a motivation for the synthetic study rather than an empirical case study.

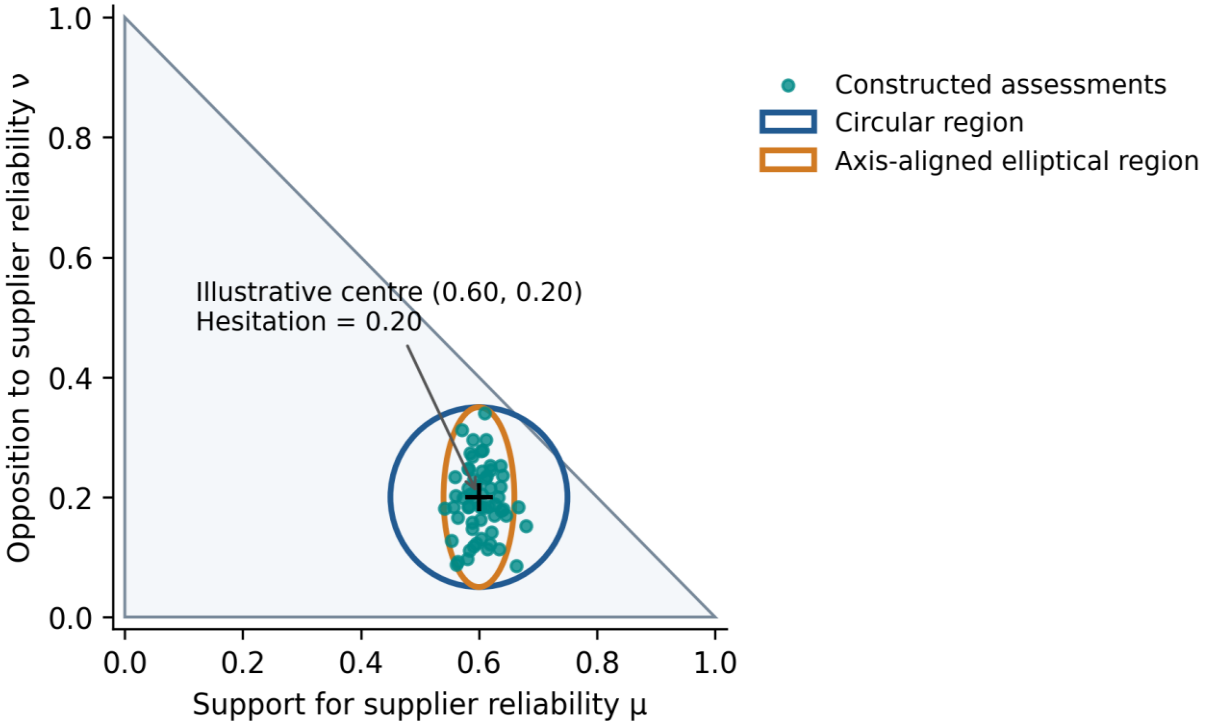


Figure 2. Illustrative supplier assessments using constructed data.

Sixty points are generated with seed 182 by adding independent normal deviations with standard deviations 0.025 and 0.065 to (0.60,0.20); all generated pairs are admissible. The circle has radius 0.15 and the ellipse has membership and non-membership semiaxes 0.06 and 0.15. Both shapes lie entirely inside Δ, so no area is removed by clipping. These fixed shapes illustrate directional spread; they are not calibrated regions and imply no coverage guarantee.

A worked supplier example with clipping considers a second assessment centre $(\mu, \nu) = (0.60, 0.35)$, leaving hesitation 0.05. Suppose the supplier has substantial supporting evidence and substantial opposing evidence, with little unassigned judgement. A circle of radius 0.12 and an axis-aligned ellipse with semiaxes 0.06 and 0.12 both extend above $\mu + \nu = 1$. Figure 3 shows their intersections with Δ. For example, (0.65, 0.40) lies in both full shapes but is removed because its implied hesitation is −0.05. The admissible point (0.60, 0.40), whose hesitation is zero, remains in both regions.

The full circle area is 0.04524, falling to 0.03098 after clipping; the ellipse area falls from 0.02262 to 0.01655. Thus clipping removes different proportions of the two shapes. These calculations illustrate why the study measures retained admissible area. They do not establish equal coverage: the shapes are fixed for explanation, whereas the benchmark fits and calibrates

regions and evaluates coverage on held-out assessments. This boundary example complements Figure 2, where both shapes are wholly admissible and clipping removes nothing.

1.2 Relationship to existing methods

The calibration rule is standard split-conformal prediction, rather than a new algorithm [2, 11]. Mahalanobis-based ellipsoidal uncertainty sets already appear in conformal robust optimization [8], and ellipsoidal conformal inference has also been developed for multi-target regression [9]. The oriented baseline used here belongs to that established geometric approach; its inclusion prevents attributing the effect of rotation to the axis-aligned model.

Henderson, Mazoyer and Gamboa [7] study conformal ellipsoids and establish asymptotic volume reductions relative to balls under ellipticity assumptions, with numerical examples beyond Gaussian data. Thus neither ball-versus-ellipsoid compactness nor heavy-tailed testing is a new research question here. Their regression scores and theoretical analysis differ from this unconditional, clipped assessment-region benchmark.

Amaral *et al.* [1] develop conformal prediction for compositional responses, including simplex-respecting constructions, coverage and region-size evaluation, and real-data applications. Their Dirichlet-regression setting and density-based regions differ from the covariance-based shapes used here. Nevertheless, simplex geometry alone cannot support a novelty claim for the present paper.

The map $(\mu, \nu) \mapsto (\mu, \nu, 1 - \mu - \nu)$ is a bijection from Δ to the closed three-component simplex. It motivates comparison with compositional-data methods, while preserving the semantic distinction: membership, non-membership and hesitation are not automatically probabilities or observed material proportions. Numerical area comparisons between different coordinate systems would also require a common reference measure.

The operational question addressed here is how much axis flexibility, orientation and support clipping change region efficiency for a shared assessment-sampling and calibration protocol. This is a chosen benchmark question, not an assertion that the broader literature lacks constrained prediction regions. The present covariance estimators do not minimize clipped area, and the study makes no claim that they are optimal.

The contribution is a specific assessment-domain benchmark, with incremental empirical findings rather than methodological novelty. It separates marginal spread from orientation, compares all three estimators on identical data, and measures how clipping changes their relative area efficiency. The elementary nesting and support-function arguments provide exposition, not new theorems. The simulation count controls Monte Carlo precision; it does not establish originality. No first-of-its-kind claim is made.

2 Assessment domain and comparison regions

For a fixed object x, an assessment $z = (\mu, \nu)$ obeys $\mu \geq 0$, $\nu \geq 0$ and $\mu + \nu \leq 1$; the hesitation degree is $1 - \mu - \nu$. All admissible pairs lie in the triangle Δ. Repeated observations in this paper represent independent assessments of one object or a defined homogeneous population. The analysis does not justify pooling unrelated alternatives or classes.

$$\Delta = \{(u,v) : u \geq 0,\ v \geq 0,\ u+v \leq 1\}. \tag{1}$$

For centre $c \in \Delta$ and positive parameters, define the full circle B_C and axis-aligned ellipse B_E. Their admissible comparison regions are intersections with Δ.

$$B_C(c,r) = \{z : \| z-c \|_2^2 \leq r^2\}, \quad R_C = B_C \cap \Delta, \tag{2}$$

$$B_E(c,a,b) = \{z : (u-\mu)^2/a^2 + (v-\nu)^2/b^2 \leq 1\}, \quad R_E = B_E \cap \Delta \tag{3}$$

The oriented model uses a symmetric positive-definite shape matrix Q. Here O is a statistical baseline, not a claim about an additional original E-IFS operation.

$$B_O(c,Q) = \{z : (z-c)^T Q^{-1} (z-c) \leq 1\}, \quad R_O = B_O \cap \Delta. \tag{4}$$

The benchmark restricts circle radius and ellipse principal semiaxes to (0, $\sqrt{2}$]. Zero-size or out-of-range fitted shapes return Δ as a separately recorded conservative fallback. This fallback is not labelled a nondegenerate ellipse. These definitions are local comparison conventions; references [4, 5] provide the foundational C-IFS and E-IFS definitions, while [6, 10] concern distances and an application, respectively.

2.1 Nesting and a controlled area benchmark

Proposition 1. *Within these nondegenerate geometric families, every circular region is an axis-aligned elliptic region, and every axis-aligned elliptic region is an oriented elliptic region.*

Proof. Set $a = b = r$ in (3) and set $Q = \mathrm{diag}(a^2, b^2)$ in (4). The full regions coincide in each case; intersection with Δ preserves equality. Thus, nesting concerns the feasible region classes. It does not imply that an estimator with more parameters attains lower test loss.

For an identical objective over nested feasible classes, the infimum loss in the larger class is no greater. This optimization fact does not guarantee that a particular training algorithm attains the infimum or generalizes better.

Proposition 2. *For deviations $D = \{(p,0), (-p,0), (0,q), (0,-q)\}$ about a fixed interior centre, $p, q > 0$, assume that all translated points lie in Δ. Then the least full-circle area is $\pi(\max(p, q))^2$ and the least full axis-aligned ellipse area is πpq.*

Proof. The circle requires $r \geq \max(p, q)$. The ellipse requires $a \geq p$ and $b \geq q$, with the lower bounds jointly attainable. Their area ratio is $\min(p,q)/\max(p,q)$, strictly below one exactly when $p \neq q$. The statement concerns full shapes at a fixed centre, not their clipped areas or fitted held-out regions.

2.2 Exact boundary containment

Proposition 3. For $c = (\mu, \nu)$, the full axis-aligned ellipse is contained in Δ if and only if $a \leq \mu$, $b \leq \nu$, and $\mu + \nu + \sqrt{a^2 + b^2} \leq 1$. For a circle, replace *a* and *b* by *r*.

Proof. The minimum horizontal and vertical coordinates are $\mu - a$ and $\nu - b$. The support function gives the maximum of $u + v$ as $\mu + \nu + \sqrt{a^2 + b^2}$. These three bounds are precisely the half-space inequalities defining Δ.

For a general Q, the corresponding conditions are:

$$\sqrt{Q_{11}} \le \mu, \ \sqrt{Q_{22}} \le \nu, \text{ and } \ \mu + \nu + \sqrt{Q_{11} + 2Q_{12} + Q_{22}} \le 1.$$

They follow from the same support-function argument. Full area πr^2, πab or $\pi\sqrt{\det(Q)}$ equals admissible area only when the relevant containment conditions hold.

As an exact illustration, $c = (0.40, 0.30)$, $p = 0.12$ and $q = 0.04$ give circle area 0.04524 and ellipse area 0.01508, a ratio of $1/3$. Proposition 3 confirms containment of both shapes. This four-point construction is an algebraic example, separate from the simulations.

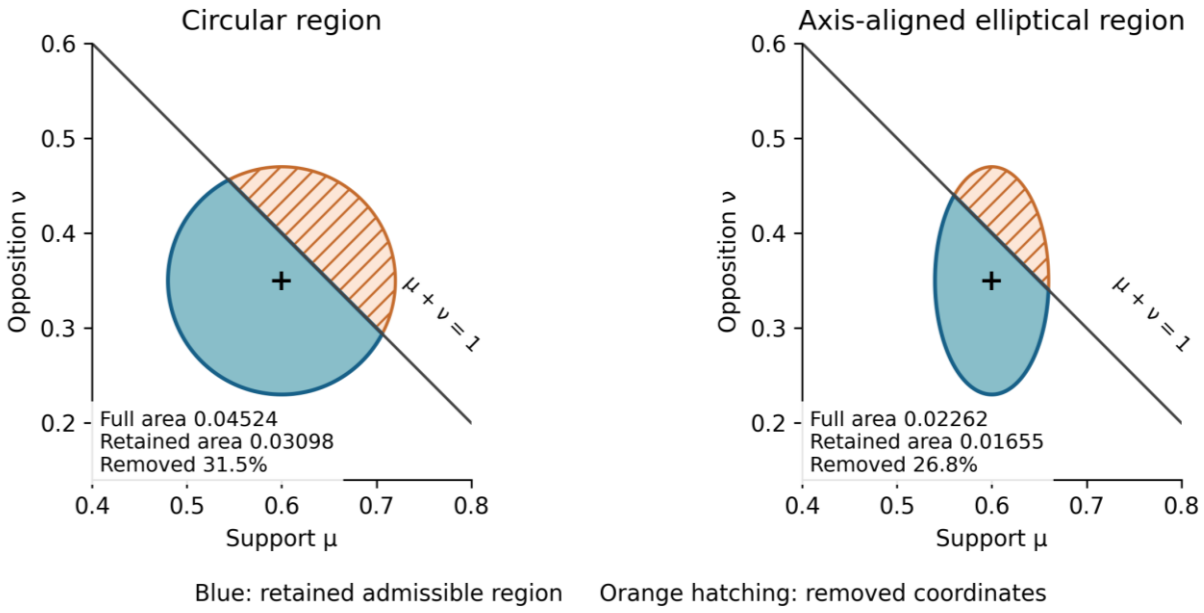


Figure 3. Constructed supplier example with boundary clipping.

Figure 3 applies this intersection operation to the supplier reliability example introduced in Section 1.1. Both constructed shapes cross $\mu + \nu = 1$; the blue portions remain admissible and the hatched portions are removed. The areas describe this worked application, rather than Table 9 averages. Both regions are centered at (0.60,0.35), with hesitation 0.05. The circle has radius 0.12; the axis-aligned ellipse has membership and non-membership semiaxes 0.06 and 0.12. Full and retained areas are 0.04524 and 0.03098 for the circle, and 0.02262 and 0.01655 for the ellipse. Clipping removes 31.5 and 26.8 percent, respectively. These fixed geometric constructions use no observed supplier data and provide no calibrated coverage claim.

3 Fitting calibration and statistical scope

3.1 Training and calibration algorithm

Use disjoint training, calibration and test sets. Estimate c by the training mean, which lies in Δ by convexity, and S by the unbiased training covariance. For C use $H = I$. For E use $H = \operatorname{diag}\left(\max(S_{11}, 10^{-12}), \max(S_{22}, 10^{-12})\right)$. For O retain the eigenvectors of S and replace each eigenvalue below 10^{-12} by 10^{-12}. This numerical regularization is fixed independently of test observations.

$$d_H(z) = \sqrt{\left((z-c)^T H^{-1}(z-c)\right)}, \quad k = \lceil (m+1)\gamma \rceil \tag{5}$$

For m calibration observations and target coverage γ, let t be the kth smallest calibration score. If $k > m$, return Δ. Otherwise set $Q = t^2 H$ and form $R = \{z : d_H(z) \le t\} \cap \Delta$. The circle has $r = t$, the axis-aligned ellipse has $a = t\sqrt{H_{11}}$ and $b = t\sqrt{H_{22}}$, and the oriented ellipse has principal semiaxes equal to the square roots of the eigenvalues of Q. A zero, nonfinite or out-of-range principal semiaxis triggers the recorded Δ fallback. No fallback occurred in the reported experiments.

3.2 Coverage interpretation

Under independent identically distributed assessment pairs and a training-frozen score function, calibration and new-observation scores are exchangeable. The rank argument gives $P\{Z_{new} \in R\} \ge \gamma$, with the usual conservative treatment of ties [2, 11]. This probability averages over the calibration sample and the new observation; under the iid (independent and identically distributed) setup it also holds conditional on the frozen training sample. It is not a guarantee for every realized calibration set, object subgroup or expert. Clipping to Δ does not remove a possible new observation under the assumed support, and replacing an invalid region by all of Δ cannot reduce coverage.

Distribution-free refers to the absence of a specified parametric data distribution, not the absence of sampling assumptions. Standard conformal calibration is already distribution-free under exchangeability. Robustness of covariance estimation, numerical stability and small area are different questions; the calibration theorem does not guarantee them.

3.3 Clustered assessments and resampling

Splitting whole experts or subjects avoids leakage, but alone does not make all their individual assessments exchangeable with a new point in the sense required by this procedure. A cluster-level guarantee needs a defined target, such as a new independent expert, and an appropriate cluster-level score and weighting. Alternatively, one assessment can be sampled by a fixed rule from each independent cluster. The present code and theorem concern iid pairs only. Temporal dependence similarly requires separate methods and assumptions; ellipsoidal conformal methods for dependent time series address a different problem [12].

For real-data evaluation, first assign unique independent unit identifiers to disjoint outer fitting and evaluation pools. Any bootstrap copies of a unit must remain in its assigned role. Resample within the fitting pool and split by original identifier into training and calibration roles, or resample an already fixed training/calibration partition separately. Resample the independent evaluation pool separately for evaluation uncertainty. Refit and recalibrate for procedure-level uncertainty; for a frozen fitted region, resample only its independent evaluation units. Do not bootstrap a pooled dataset and then randomly split its duplicate rows across roles.

This resampling estimates uncertainty for the specified evaluation design; it is not an additional conformal coverage theorem. Repeated random splits of a single dataset are dependent and must not be counted as independent experimental replications.

4 Experimental design and endpoints

The main experiment uses six configurations, with 200 independent generated datasets per configuration, training size n=200, calibration size $m = 199$ and test size 1,000. All three methods share the same observations within a repetition.

The random stream is NumPy `default_rng(20261026)`; Appendix A specifies generation order. A further 15 configurations vary training size or γ for the first three scenarios, giving 21 configurations, 4,200 datasets and 12,600 fits in total.

Each generator first samples in the plane and then rejects pairs outside Δ. Table 1 describes the generating law before conditioning. Rejection changes moments, dependence and mixture weights, particularly near the boundary; labels such as isotropic and tilted describe the intended preconditioning geometry. The synthetic data do not encode real expert behaviour.

Table 1. Generating laws before conditioning on Δ

Scenario	Centre	Gaussian scales or mixture	Interpretation
Isotropic	(0.40, 0.30)	σ = (0.045, 0.045)	Equal marginal spread
Anisotropic	(0.40, 0.30)	σ = (0.090, 0.025)	Unequal axial spread
Tilted	(0.40, 0.30)	Rotate independent scales (0.090, 0.025) by 45°	Oblique principal variation
Boundary	(0.48, 0.48)	σ = (0.070, 0.025)	Substantial simplex truncation
Contaminated	(0.40, 0.30)	0.9 $N(0, D^2)$ + 0.1 $N(0, 16D^2)$; D = diag(0.060, 0.025)	Nominal 10% broad component
Bimodal	(0.40, 0.30)	Equal offsets ±(0.090, 0.040), each with σ = (0.020, 0.020)	Separated mixture components

Sensitivity configurations use $n \in \{50, 100, 500\}$ at $\gamma = 0.90$, and $\gamma \in \{0.80, 0.95\}$ at $n = 200$. Calibration and test sizes remain 199 and 1,000. These are one-factor sensitivity analyses, not a full factorial experiment. Results at $n = 200$ and $\gamma = 0.90$ are the common baseline. The influence of calibration-sample size is not separately estimated.

4.1 Coverage compactness and uncertainty

Held-out coverage is the test fraction in R. Area is $S(R) = \text{area}(R)$, integrated after clipping to Δ. Pairwise differences are $\Delta S = S(M) - S(C)$ and $\Delta C = \text{coverage}(M) - \text{coverage}(C)$, for $M \in \{E, O\}$; direct $O-E$ differences are also reported for the main configurations. Ratios in sensitivity tables are ratios of mean areas, not means of per-repetition ratios. All repetitions enter every reported endpoint; no test-coverage filtering occurs.

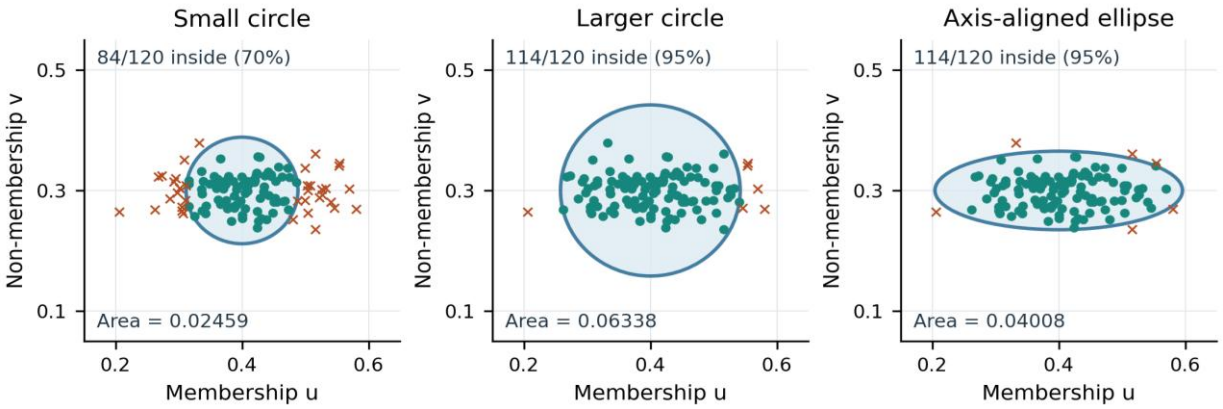


Figure 4. Coverage and compactness on a constructed sample of 120 admissible pairs.

The small circle contains 84 points (70 percent); the larger circle and axis-aligned ellipse each contain 114 points (95 percent). Filled points are included and crosses are excluded. Shapes were sized using this same sample for illustration, so these percentages are not held-out coverage or conformal guarantees. Areas are measured in Δ.

Figure 4 illustrates why region size and coverage must be considered jointly. Enlarging a circle increases the number of included points, while an ellipse can cover the same number with a smaller area when the sample varies mainly along one axis. These are in-sample constructions; the held-out evaluation protocol remains the basis for the reported results.

For each endpoint or paired difference x over $B = 200$ independent simulation repetitions, report the mean, Monte Carlo standard error $s_x / \sqrt{B}$, and a pointwise normal interval mean $\pm 1.96 s_x / \sqrt{B}$. Tables display the intervals, from which the standard errors can be recovered. These intervals quantify uncertainty in the simulation expectation, rather than the spread of individual fitted regions. The intervals are exploratory and are not multiplicity-adjusted.

For an explicit coverage comparison, use $\delta = 0.01$ as an illustrative tolerance and inspect whether the lower endpoint of the paired ΔC interval is above $-\delta$. This margin was chosen during revision after inspecting the earlier experiment, and has no elicited application-specific justification. It is therefore an exploratory compatibility check, not a preregistered or confirmatory noninferiority test. No formal claim of equal coverage follows from a nonsignificant difference.

The separate observed shortfall diagnostic is the fraction of repetitions with test coverage below $\gamma - 0.02$.

4.2 Stability and numerical implementation

Anisotropy is $\kappa =$ major/minor semiaxis. Report mean κ and across-repetition coefficients of variation CV $=$ sample SD/mean for the calibrated major and minor semiaxes. These variations combine shape estimation and calibration variability; they do not measure orientation stability or robustness against every contamination mechanism. Per-fit area, coverage, semiaxis and fallback arrays are generated by the appendix code.

For Q, write

$$h = \sqrt{Q_{11}},$$

$$\text{midpoint } (v \mid u) = v + (Q_{21} / Q_{11})(u - \mu),$$

and

$$\text{half-height} = \sqrt{(Q_{22} - Q_{21}^2 / Q_{11})\{1 - (u - \mu)^2 / Q_{11}\}}.$$

Intersect the resulting vertical interval with $[0, 1 - u]$, then integrate its nonnegative length over $u \in [\max(0, \mu - h), \min(1, \mu + h)]$. This formula handles all three models. SciPy quad uses absolute and relative tolerances 10^{-8}; reported integration error estimates above 10^{-7} stop execution. Integration warnings are treated as errors. Table 2 summarizes the evaluation criteria, the quantities reported for each criterion, and the limits of the study's scope.

Table 2. Evaluation scope

Criterion	ported here	Interpretation
Validity and area	Clipped-domain area; fallback count	Support respected by construction
Coverage	Means, paired intervals and shortfall frequencies	Marginal prediction performance
Anisotropy and stability	κ and semiaxis coefficients of variation	Estimated directional variation
Sample and coverage sensitivity	Training sizes and nominal γ	Selected design sensitivity
Computation	Algorithm and numerical tolerances	No runtime superiority claim
Tility	Not measured	Requires a genuine decision application

5 Results

Tables 3–5 give the main configurations at $\gamma = 0.90$. *C* denotes circle, *E* axis-aligned ellipse and *O* oriented ellipse. Coverage is a proportion in Table 3 and a percentage-point difference in Table 4. No numerical failure or Δ fallback occurred in any of the 12,600 fits. Table 4 reports *E*−*C*, *O*−*C* and *O*−*E*; the direct contrast uses the same independent repetitions and paired interval calculation.

Table 3. Main coverage and admissible area

Scenario	Model	Mean area	Mean coverage [95% MC interval]	Fraction below .88
Isotropic	*C*	0.02943	0.898 [0.895, 0.901]	0.220
Isotropic	*E*	0.02952	0.898 [0.895, 0.901]	0.220
Isotropic	*O*	0.02966	0.898 [0.895, 0.902]	0.225
Anisotropic	*C*	0.07172	0.899 [0.895, 0.902]	0.220
Anisotropic	*E*	0.03345	0.903 [0.900, 0.907]	0.190
Anisotropic	*O*	0.03352	0.903 [0.899, 0.906]	0.200
Tilted	*C*	0.06911	0.901 [0.897, 0.904]	0.160
Tilted	*E*	0.06909	0.900 [0.897, 0.904]	0.160
Tilted	*O*	0.03221	0.901 [0.898, 0.905]	0.185
Boundary	*C*	0.02132	0.901 [0.898, 0.904]	0.165
Boundary	*E*	0.01598	0.899 [0.896, 0.902]	0.205
Boundary	*O*	0.01589	0.901 [0.898, 0.904]	0.160
Contaminated	*C*	0.05022	0.901 [0.898, 0.905]	0.190
Contaminated	*E*	0.03342	0.902 [0.899, 0.905]	0.155
Contaminated	*O*	0.03396	0.902 [0.898, 0.905]	0.200
Bimodal	*C*	0.05139	0.901 [0.898, 0.904]	0.180
Bimodal	*E*	0.04352	0.900 [0.897, 0.904]	0.190
Bimodal	*O*	0.02424	0.899 [0.896, 0.902]	0.195

Mean area for *E* is 0.03345 versus 0.07172 for *C* in the anisotropic configuration, a 53.4 percent reduction. For tilted variation, *E* and *C* have almost identical areas; *O* reduces mean area from 0.06911 to 0.03221, also approximately 53.4 percent. These findings distinguish axis flexibility from orientation. In the isotropic configuration, the more flexible estimators have slightly larger mean regions, consistent with finite-sample estimation costs.

Table 4. Paired area and coverage differences

Scenario	Contrast	ΔArea [95% MC interval]	ΔCoverage in pp [95% MC interval]
Isotropic	$E - C$	0.000089 [–0.000010, 0.000189]	0.005 [–0.082, 0.091]
Isotropic	$O - C$	0.000229 [0.000096, 0.000361]	0.036 [–0.079, 0.151]
Isotropic	$O - E$	0.000139 [0.000044, 0.000235]	0.032 [–0.049, 0.112]
Anisotropic	$E - C$	–0.038276 [–0.039374, –0.037177]	0.456 [0.129, 0.782]
Anisotropic	$O - C$	–0.038207 [–0.039292, –0.037122]	0.413 [0.095, 0.731]
Anisotropic	$O - E$	0.000069 [–0.000060, 0.000198]	–0.043 [–0.139, 0.054]
Tilted	$E - C$	–0.000024 [–0.000118, 0.000070]	–0.022 [–0.054, 0.011]
Tilted	$O - C$	–0.036901 [–0.037827, –0.035975]	0.068 [–0.220, 0.357]
Tilted	$O - E$	–0.036877 [–0.037801, –0.035953]	0.090 [–0.199, 0.379]
Boundary	$E - C$	–0.005339 [–0.005561, –0.005117]	–0.205 [–0.464, 0.053]
Boundary	$O - C$	–0.005437 [–0.005672, –0.005202]	–0.013 [–0.288, 0.262]
Boundary	$O - E$	–0.000098 [–0.000196, 4.97e-7]	0.192 [0.032, 0.353]
Contaminated	$E - C$	–0.016803 [–0.017696, –0.015909]	0.067 [–0.154, 0.288]
Contaminated	$O - C$	–0.016261 [–0.017186, –0.015337]	0.035 [–0.180, 0.249]
Contaminated	$O - E$	0.000541 [0.000268, 0.000815]	–0.032 [–0.133, 0.068]
Bimodal	$E - C$	–0.007870 [–0.008193, –0.007546]	–0.054 [–0.379, 0.272]
Bimodal	$O - C$	–0.027143 [–0.027657, –0.026629]	–0.192 [–0.636, 0.252]
Bimodal	$O - E$	–0.019273 [–0.019684, –0.018863]	–0.139 [–0.528, 0.251]

The direct $O-E$ contrast in the tilted scenario gives mean area difference −0.03688 (95 percent Monte Carlo interval [−0.03780, −0.03595]) and coverage difference +0.090 percentage points (interval [−0.199,0.379]). In the anisotropic scenario, the $O-E$ area interval includes zero; in the isotropic and contaminated scenarios O has a larger mean area than E, with positive pointwise area intervals. These exploratory contrasts support a geometry-dependent benefit from orientation, rather than universal superiority of the oriented estimator.

All main pairwise coverage intervals have lower endpoints above −1 percentage point. This includes the direct $O-E$ contrasts. Thus each meets the illustrative retrospective compatibility criterion, without proving equivalence or a confirmatory noninferiority result. The anisotropic $E-C$ coverage difference is +0.456 percentage points, interval [0.129, 0.782]; its smaller area does not arise from lower mean coverage. The boundary $E-C$ difference is −0.205 percentage points, interval [−0.464, 0.053].

The oriented model also produces smaller mean regions for the bimodal generator. This can reflect efficient orientation of an envelope around both clusters; a convex ellipse still includes low-density space between them. This Gaussian mixture has positive density throughout the interior of Δ after conditioning, so its support is connected. Neither compactness nor calibrated coverage implies faithful representation of its separated high-density clusters.

Table 5. Anisotropy and calibrated semiaxis variability

Scenario	Model	Mean κ	CV major	CV minor
Isotropic	*C*	1.000	0.047	0.047
Isotropic	*E*	1.060	0.054	0.050
Isotropic	*O*	1.097	0.054	0.052
Anisotropic	*C*	1.000	0.065	0.065
Anisotropic	*E*	3.592	0.063	0.059
Anisotropic	*O*	3.603	0.064	0.060
Tilted	*C*	1.000	0.056	0.056
Tilted	*E*	1.030	0.058	0.056
Tilted	*O*	3.506	0.050	0.057
Boundary	*C*	1.000	0.063	0.063
Boundary	*E*	2.135	0.066	0.062
Boundary	*O*	2.228	0.069	0.063
Contaminated	*C*	1.000	0.101	0.101
Contaminated	*E*	2.168	0.105	0.114
Contaminated	*O*	2.208	0.105	0.118
Bimodal	*C*	1.000	0.029	0.029
Bimodal	*E*	2.060	0.037	0.034
Bimodal	*O*	5.036	0.062	0.045

Here E has mean $\kappa \approx 1.030$ in the tilted scenario, whereas O has $\kappa \approx 3.506$, showing how nearly equal marginal spreads conceal directional variation. Contamination produces greater calibrated semiaxis variation than the isotropic setting. This is a descriptive comparison across different generating laws, not a controlled robustness effect with every other aspect held constant.

6 Sensitivity analyses

Tables 6–8 report the additional configurations. Coverage entries list $C/E/O$ means. Paired coverage intervals are in percentage points. Ratio entries give E/C and O/C mean-area ratios. All intervals remain pointwise and exploratory. Calibration size is fixed, so changing n measures training-shape estimation rather than calibration precision.

Table 6. Isotropic sensitivity

n	γ	Coverage $C/E/O$	Area ratio E/C and O/C	ΔCoverage $E-C$ [95% interval]	ΔCoverage $O-C$ [95% interval]
50	0.90	0.899/0.900/0.900	1.017 / 1.029	0.07 [–0.06, 0.20]	0.08 [–0.07, 0.24]
100	0.90	0.898/0.899/0.898	1.008 / 1.010	0.07 [–0.04, 0.18]	–0.02 [–0.17, 0.14]
500	0.90	0.902/0.901/0.902	1.000 / 1.003	–0.03 [–0.11, 0.04]	–0.01 [–0.10, 0.09]
200	0.80	0.801/0.801/0.800	1.002 / 1.001	–0.05 [–0.16, 0.06]	–0.13 [–0.29, 0.02]
200	0.95	0.948/0.948/0.948	1.002 / 1.010	–0.03 [–0.11, 0.05]	0.04 [–0.06, 0.14]

Table 7. Anisotropic sensitivity

n	γ	Coverage $C/E/O$	Area ratio E/C and O/C	ΔCoverage $E-C$ [95% interval]	ΔCoverage $O-C$ [95% interval]
50	0.90	0.898/0.900/0.900	0.467 / 0.473	0.25 [–0.07, 0.57]	0.21 [–0.12, 0.54]
100	0.90	0.901/0.901/0.901	0.461 / 0.463	–0.01 [–0.31, 0.30]	–0.01 [–0.32, 0.30]
500	0.90	0.901/0.903/0.904	0.466 / 0.466	0.27 [–0.07, 0.61]	0.30 [–0.03, 0.64]
200	0.80	0.802/0.801/0.800	0.516 / 0.516	–0.09 [–0.49, 0.32]	–0.14 [–0.54, 0.27]
200	0.95	0.949/0.949/0.950	0.424 / 0.427	0.01 [–0.21, 0.22]	0.07 [–0.16, 0.29]

Table 8. Tilted sensitivity

n	γ	Coverage *C*/*E*/*O*	Area ratio *E*/*C* and O/C	ΔCoverage *E*−*C* [95% interval]	ΔCoverage *O*−*C* [95% interval]
50	0.90	0.898/0.898/0.898	1.004 / 0.473	0.03 [–0.03, 0.08]	–0.03 [–0.35, 0.29]
100	0.90	0.902/0.901/0.901	1.000 / 0.465	–0.03 [–0.07, 0.01]	–0.11 [–0.43, 0.22]
500	0.90	0.899/0.899/0.899	1.001 / 0.463	0.00 [–0.02, 0.02]	–0.07 [–0.39, 0.25]
200	0.80	0.803/0.803/0.803	1.002 / 0.522	–0.01 [–0.05, 0.04]	0.00 [–0.39, 0.40]
200	0.95	0.950/0.950/0.950	1.001 / 0.433	0.01 [–0.01, 0.03]	0.05 [–0.21, 0.31]

For isotropic data, the extra area associated with flexibility is larger at $n = 50$ than at $n = 500$. For anisotropic data, *E*/*C* ratios remain near 0.46–0.47 across the examined training sizes; *O* offers little further compactness in this axial setting. For tilted data, *O*/*C* ratios stay near 0.46–0.47 while *E*/*C* stays near one. At $\gamma = 0.80$ and 0.95 the area ratios change, so a single nominal coverage level does not characterize performance completely. These are observations for the tested settings rather than monotonicity or asymptotic claims.

6.1 Full shape area versus admissible area

For each fitted ellipse, full area is π times the major and minor semiaxes. For the circle the same expression gives πr^2. Let F_M and A_M denote the mean full and admissible areas for model M in a configuration. Table 9 reports removed area as $100(1 - A_M / F_M)$, together with F_E / F_C, A_E / A_C, F_O / F_C and A_O / A_C. These are ratios of means across the existing 200 paired repetitions, not per-fit percentage averages. Values indistinguishable from zero at numerical integration tolerance are displayed as zero. This diagnostic is descriptive; no new coverage claim or optimization guarantee is attached.

Table 9. Effect of support clipping on relative compactness

Scenario	Area removed % *C*/*E*/*O*	Full *E*/*C*	Admissible *E*/*C*	Full *O*/*C*	Admissible *O*/*C*
Isotropic	0.00/0.00/0.00	1.003	1.003	1.008	1.008
Anisotropic	0.00/0.00/0.00	0.466	0.466	0.467	0.467
Tilted	0.00/0.00/0.00	1.000	1.000	0.466	0.466
Boundary	14.20/13.71/10.58	0.745	0.750	0.715	0.745
Contaminated	0.00/0.00/0.00	0.665	0.665	0.676	0.676
Bimodal	0.00/0.00/0.02	0.847	0.847	0.472	0.472

In the boundary configuration, clipping removes 14.20, 13.71 and 10.58 percent of mean full area for C, E and O, respectively. The O/C ratio changes from 0.715 to 0.745, reducing the apparent saving from 28.5 to 25.5 percent; the corresponding E/C ratios are 0.745 and 0.750. Model ordering does not reverse in this benchmark. The result demonstrates a quantitative change in estimated efficiency, not evidence that clipping necessarily changes model rankings. For most other configurations, full and admissible mean areas are virtually identical; this limits how broadly the boundary effect should be interpreted.

7 Discussion and limitations

The results support a conditional geometric conclusion: axis flexibility helps when training data identify stable axial anisotropy, and orientation helps when principal variation is tilted. Added parameters can increase region size for nearly isotropic data. Nesting of geometric families alone cannot predict the behaviour of a covariance-based estimator at finite sample size.

Returning to the supplier example, a smaller admissible region at comparable coverage would summarize repeated reliability assessments more precisely. It would not directly rank suppliers, validate their reliability or establish better purchasing outcomes. An empirical study would need a defined prediction target, such as a future assessment of the same supplier, and expert and time identifiers to assess dependence and justify calibration. Comparing procurement outcomes would require an additional decision rule and outcome data. The present benchmark provides a geometric comparison to inform such a study, not an application validation.

The study does not replace the formal C-IFS and E-IFS theories. Region fitting, probabilistic calibration and fuzzy-set operations are distinct layers. Repeated assessments require a justified sampling interpretation before the probability statements apply. Real experts can be heterogeneous, dependent or inconsistent; the present simulation supplies no empirical evidence about such settings.

The work is positioned as a methodological benchmark or research note. The domain-specific value lies in the controlled comparison and the explicit distinction between full-shape and admissible-region efficiency. It does not replace prior ellipsoidal conformal methods [7, 8, 9] or prior prediction methods on a simplex [1]. Its empirical findings concern the present estimators and configurations, and should not be generalized to all E-IFS or C-IFS decision procedures.

Limitations include synthetic data only, fixed calibration size, a finite scenario grid, no application-derived coverage margin, and no multiplicity-adjusted confirmatory analysis. The covariance estimators are not designed to be robust. The oriented baseline does not address arbitrary nonlinear or disconnected structure. Parameter stability is reported, but runtime, memory, orientation variability and downstream utility are outside the present endpoints.

Standard split calibration already supplies distribution-free marginal coverage under the stated iid or suitable exchangeability assumptions. Future methodological work should instead examine robust centre and shape estimators, cluster-aware or temporal calibration, calibration-size sensitivity, and unions or mixtures for disconnected supports. Each extension requires its own evaluation and assumptions.

A possible extension is a training-stage shape-selection problem that directly targets admissible area rather than covariance fit or full-ellipse volume. For example, one could

minimize area ($B(c, Q) \cap \Delta$) subject to a stated training-coverage constraint and regularization, then freeze the selected score and calibrate its scale on untouched calibration data. The reported methods do not solve this optimization problem. Its numerical behaviour, finite-sample efficiency and relationship to existing minimum-volume-set methods would need investigation before it could support a novelty claim.

An application contribution would require genuinely repeated expert assessments with object, expert and time identifiers; a defined target such as a new expert or a new assessment; and a calibration procedure justified for that sampling structure. Evaluation would then compare region efficiency and actual decision outcomes under identical aggregation rules. No such dataset was used here, and no application result is inferred from the synthetic benchmark.

8 Conclusion

This benchmark applies established circular and elliptical geometry and standard split-conformal calibration to repeated admissible assessments. The synthetic experiments quantify how orientation and boundary clipping affect compactness while reporting coverage uncertainty. In the boundary configuration, clipping reduces the apparent oriented-versus-circular area saving from 28.5 to 25.5 percent, showing why full-shape area is not interchangeable with admissible area. This is an incremental benchmark result, not a new calibration theorem or algebraic generalisation. A stronger methodological or application contribution would require an implemented advance with appropriate prior-art comparisons or genuine assessment data and outcomes.

Ellipses can describe uncertainty more precisely than circles, while covering a similar proportion of new assessments—when the uncertainty spreads more in one direction than another. In one simulated setting, the ellipse used about 53% less area.

If the spread is tilted, rotating the ellipse helps. When uncertainty spreads equally in all directions, ellipses offer little benefit and can perform slightly worse.

The paper also shows why clipping matters: portions representing impossible assessments must be removed before comparing region sizes. This can change how large the ellipse's advantage appears.

These findings come from simulations; the supplier example explains the idea but does not validate it with real supplier data.

References

[1] Amaral, L. P., Cabezas, L. M. C., Ramos, T. R., & Pereira, G. H. G. A. (2026). *Conformal prediction for compositional data*. Preprint. Available online at: https://arxiv.org/abs/2511.18141

[2] Angelopoulos, A. N., & Bates, S. (2022). *A gentle introduction to conformal prediction and distribution-free uncertainty quantification.* Preprint. Available online at: https://arxiv.org/abs/2107.07511v6

[3] Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1), 87–96.

[4] Atanassov, K. T. (2020). Circular intuitionistic fuzzy sets. *Journal of Intelligent & Fuzzy Systems*, 39(5), 5981–5986.

[5] Atanassov, K. T. (2021). Elliptic intuitionistic fuzzy sets. *Comptes rendus de l'Académie bulgare des Sciences*, 74(6), 812–819.

[6] Atanassov, K. T., & Marinov, E. (2021). Four distances for circular intuitionistic fuzzy sets. *Mathematics*, 9(10), 1121.

[7] Henderson, I., Mazoyer, A., & Gamboa, F. (2025). *Adaptive inference with random ellipsoids through conformal conditional linear expectation.* Preprint. Available online at: https://arxiv.org/abs/2409.18508v2

[8] Johnstone, C., & Cox, B. (2021). Conformal uncertainty sets for robust optimization. *Proceedings of Machine Learning Research*, 152, 72–90.

[9] Messoudi, S., Destercke, S., & Rousseau, S. (2022). Ellipsoidal conformal inference for multi-target regression. *Proceedings of Machine Learning Research*, 179, 294–306.

[10] Traneva, V., & Tranev, S. (2023). An elliptic intuitionistic fuzzy model for franchisor selection. *Annals of Computer Science and Information Systems*, 37, 337–343.

[11] Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York.

[12] Xu, C., Jiang, H., & Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. *Proceedings of Machine Learning Research*, 235, 55076–55099.

Appendix A Reproducible simulation code

The complete script below produces the summary statistics and fitted-region results used in Tables 3–9. Run it in a writable directory. It creates `simulation_v3_summary.json` and `simulation_v3_raw.npz`. The raw array has axes configuration, repetition, model and metric; models are *C*/*E*/*O* and metrics are area, coverage, κ, major semiaxis, minor semiaxis and fallback indicator. The clipping_diagnostic entry supplies full and admissible mean areas, removed-area percentages, and ellipse-to-circle area ratios for Table 9. The *O* entry also reports `darea_vs_E` and `dcov_vs_E` for direct $O-E$ contrasts. Use the JSON statistics to recover reported means, intervals and coefficients of variation. The generator processes the six main configurations first, followed by the fifteen sensitivity configurations in the order shown in `run()`. Software used: NumPy 2.3.5; SciPy 1.17.0. The code fixes the random seed and numerical tolerances. Minor floating-point differences may occur across platforms.

```
"""Reproduce the manuscript benchmark. Run with NumPy and SciPy."""
import json
import warnings
from pathlib import Path
import numpy as np
from scipy.integrate import quad, IntegrationWarning

SEED, REPS, M, NTEST = 20261026, 200, 199, 1000
SCENARIOS = ["Isotropic", "Anisotropic", "Tilted",
             "Boundary", "Contaminated", "Bimodal"]
METHODS = ["C", "E", "O"]
rng = np.random.default_rng(SEED)

def sample(kind, n):
    chunks, count = [], 0
    while count < n:
        batch = max(100, 2 * (n - count))
        centre = np.array([.4, .3])
        sd = np.array([.045, .045])
        z = rng.normal(size=(batch, 2))
        if kind == "Anisotropic":
            sd = np.array([.09, .025])
        if kind == "Tilted":
            z = (z * [.09, .025]) @ (
                np.array([[1., 1.], [-1., 1.]]) / np.sqrt(2))
            sd = np.ones(2)
        if kind == "Boundary":
            centre, sd = np.array([.48, .48]), np.array([.07, .025])
        if kind == "Contaminated":
            z *= np.where(rng.random((batch, 1)) < .1, 4., 1.)
            sd = np.array([.06, .025])
        if kind == "Bimodal":
            sign = np.where(rng.random((batch, 1)) < .5, -1., 1.)
            z, sd = .02 * z + sign * [.09, .04], np.ones(2)
        x = centre + z * sd
        x = x[(x >= 0).all(axis=1) & (x.sum(axis=1) <= 1)]
        chunks.append(x)
        count += len(x)
    return np.concatenate(chunks)[:n]

def clipped_area(c, q):
    # q is the final shape matrix: (z-c)' inv(q) (z-c) <= 1.
    rx = np.sqrt(q[0, 0])
    lo, hi = max(0., c[0] - rx), min(1., c[0] + rx)
    conditional_var = q[1, 1] - q[1, 0]**2 / q[0, 0]
    def height(u):
        mid = c[1] + q[1, 0] / q[0, 0] * (u - c[0])
        half = np.sqrt(max(0., conditional_var *
                           (1. - ((u - c[0]) / rx)**2)))
        return max(0., min(1. - u, mid + half) - max(0., mid - half))
    value, error = quad(height, lo, hi, epsabs=1e-8,
                        epsrel=1e-8, limit=150)
    if error > 1e-7:
        raise RuntimeError("Area integration error above tolerance")
    return value

def fit_eval(train, cal, test, gamma):
```

```
    c = train.mean(axis=0)
    cov = np.cov(train, rowvar=False, ddof=1)
    eig, vectors = np.linalg.eigh(cov)
    full = (vectors * np.maximum(eig, 1e-12)) @ vectors.T
    diagonal = np.diag(np.maximum(np.diag(cov), 1e-12))
    result = []
    for base in [np.eye(2), diagonal, full]:
        inverse = np.linalg.inv(base)
        def score(z):
            delta = z - c
            return np.sqrt(np.maximum(
                0., np.einsum("ni,ij,nj->n", delta, inverse, delta)))
        k = int(np.ceil((len(cal) + 1) * gamma))
        t = np.sort(score(cal))[k - 1] if k <= len(cal) else np.inf
        if not np.isfinite(t):
            result.append([.5, 1., np.nan, np.nan, np.nan, 1.])
            continue
        q = t**2 * base
        axes = np.sqrt(np.linalg.eigvalsh(q))
        fallback = (not np.isfinite(axes).all()
                    or axes.min() <= 0 or axes.max() > np.sqrt(2))
        if fallback:
            result.append([.5, 1., np.nan, np.nan, np.nan, 1.])
        else:
            result.append([clipped_area(c, q),
                           np.mean(score(test) <= t),
                           axes.max() / axes.min(),
                           axes.max(), axes.min(), 0.])
    return np.array(result)

def interval(x):
    mean = float(np.mean(x))
    se = float(np.std(x, ddof=1) / np.sqrt(len(x)))
    return [mean, mean - 1.96 * se, mean + 1.96 * se, se]

def summarize(values):
    output = {}
    for j, name in enumerate(METHODS):
        v = values[:, j, :]
        valid = v[:, 5] == 0
        output[name] = {
            "area": interval(v[:, 0]), "coverage": interval(v[:, 1]),
            "kappa": float(np.nanmean(v[:, 2])),
            "cv_major": float(np.nanstd(v[:, 3], ddof=1) /
                              np.nanmean(v[:, 3])),
            "cv_minor": float(np.nanstd(v[:, 4], ddof=1) /
                              np.nanmean(v[:, 4])),
            "fallbacks": int(np.sum(~valid))}
        if j:
            output[name]["darea"] = interval(v[:, 0] - values[:, 0, 0])
            output[name]["dcov"] = interval(v[:, 1] - values[:, 0, 1])
    output["O"]["darea_vs_E"] = interval(values[:, 2, 0] - values[:, 1, 0])
    output["O"]["dcov_vs_E"] = interval(values[:, 2, 1] - values[:, 1, 1])
    full = np.pi * values[:, :, 3] * values[:, :, 4]
    f = full.mean(axis=0)
    a = values[:, :, 0].mean(axis=0)
    output["clipping_diagnostic"] = {
```

```
        "mean_full_area": f.tolist(),
        "mean_admissible_area": a.tolist(),
        "removed_percent": (100. * (1. - a / f)).tolist(),
        "full_ratios_EC_OC": [float(f[1]/f[0]), float(f[2]/f[0])],
        "admissible_ratios_EC_OC":
            [float(a[1]/a[0]), float(a[2]/a[0])]}
    return output

def run():
    configs = [(kind, 200, .90) for kind in SCENARIOS]
    configs += [(kind, n, gamma)
                for kind in SCENARIOS[:3]
                for n, gamma in [(50, .90), (100, .90), (500, .90),
                                 (200, .80), (200, .95)]]
    summaries, raw = [], []
    with warnings.catch_warnings():
        warnings.simplefilter("error", IntegrationWarning)
        for kind, n, gamma in configs:
            repetitions = []
            for _ in range(REPS):
                x = sample(kind, n + M + NTEST)
                repetitions.append(fit_eval(
                    x[:n], x[n:n+M], x[n+M:], gamma))
            values = np.array(repetitions)
            entry = {"scenario": kind, "n": n, "gamma": gamma,
                     "summary": summarize(values)}
            for j, name in enumerate(METHODS):
                entry["summary"][name]["below"] = float(
                    np.mean(values[:, j, 1] < gamma - .02))
            summaries.append(entry)
            raw.append(values)
    np.savez_compressed("simulation_v3_raw.npz", values=np.array(raw))
    Path("simulation_v3_summary.json").write_text(json.dumps(summaries, indent=2))
    return summaries

if __name__ == "__main__":
    results = run()
    print("Completed", len(results) * REPS, "datasets and",
          len(results) * REPS * len(METHODS), "fits")
    print("Fallbacks", sum(r["summary"][m]["fallbacks"]
                           for r in results for m in METHODS))
    for row in results[:6]:
        print(row["scenario"], {m: round(row["summary"][m]["area"][0], 5)
                                for m in METHODS})
```