Defining Operational Conditions for Safety-Critical AI-Based Systems from Data

Johann M. Christensen
Institute for AI Safety and Security
German Aerospace Center (DLR)
Sankt Augustin, Germany
johann.christensen@dlr.de

Elena Hoemann
Institute for AI Safety and Security
German Aerospace Center (DLR)
Sankt Augustin, Germany
elena.hoemann@dlr.de

Frank Köster
Institute for AI Safety and Security
German Aerospace Center (DLR)
Sankt Augustin, Germany
frank.koester@dlr.de

Sven Hallerbach
Institute for AI Safety and Security
German Aerospace Center (DLR)
Sankt Augustin, Germany
sven.hallerbach@dlr.de

Abstract

Artificial Intelligence (AI) has been on the rise in many domains, including numerous safety-critical applications. However, for complex systems in the real world, defining the underlying environmental conditions in which the AI-based system must operate—the Operational Design Domain (ODD)—is extremely challenging. This often results in an incomplete description of the ODD, which contrasts with the requirements of many domains for certifying AI-based systems. Traditionally, the ODD is created in the early stages of the development process, drawing on sophisticated expert knowledge and related standards. This paper presents a novel method for defining the ODD a posteriori from previously collected data using a multidimensional kernel-based representation. This approach is validated through both synthetic benchmarks and a real-world aviation use case. Moreover, the paper defines similarity of two ODDs if they generate the same outputs up to Lebesgue-null input sets and proves convergence in volume of the calibrated representation under the stated assumptions. The novel, safety-by-design, deterministic kernel-based ODD representation is derived fully automatically, given documented assurance inputs, permutation-stable, bounded by construction, and, under affine-equivariant per-dimension normalization, invariant to the choice of units. Utilizing the proposed ODD representation supports future certification of data-driven, safety-critical AI-based systems.

1 Introduction

The rapid adoption of Artificial Intelligence (AI) across a wide range of application domains has fundamentally changed how complex systems are designed and operated. In safety-critical domains, AI-based systems are increasingly responsible for perception, decision-making, and control [1–9]. While these capabilities enable unprecedented levels of autonomy and performance, they also introduce new challenges for safety assurance, validation, and certification [10, 11]. Research into the safety of AI-based systems has only recently begun to catch up with their practical deployment. Especially in domains with stringent safety requirements, e.g., aviation, safety considerations are not merely desirable but mandatory to enable certification and operational approval [12–15]. Current approaches to ensuring the safety of AI-based systems predominantly rely on reactive measures, such as extensive testing, runtime monitoring, and post hoc failure analysis [16–18]. Although effective,

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

these approaches are often time-consuming, costly, and difficult to scale. In contrast, safety-by-design methodologies aim to integrate safety considerations directly into the system development process, favoring proactive design measures over reactive mitigation. This paradigm shift has given rise to the research field of AI engineering, which seeks systematic, certifiable methods for designing AI-based systems from the outset. A central concept in safety-by-design AI engineering is the Operational Design Domain (ODD) [16, 19, 20]. The ODD specifies the set of operational conditions under which an AI-based system is intended to function safely [19, 21–23]. By explicitly defining these conditions, the ODD provides a foundation for system design and verification, and has become a key instrument in multiple regulatory frameworks. However, despite its importance, defining an ODD remains challenging [24, 25]. Traditionally, ODDs are created manually by domain experts early in the development process [24, 22, 26]. While expert-defined ODDs are effective when system behavior and environmental conditions are well understood, they become increasingly difficult to construct for complex real-world systems. Prior research characterizes ODDs from data in an aeronautical context [27]; however, the approach provides neither a deterministic, order-independent representation nor a formal framework for ODD similarity—both of which support the certification use cases addressed in this work (cf. Appendix J). Data-driven ODDs promise several advantages: they can capture complex, implicit parameter dependencies, naturally reflect real operational conditions, and be updated as new data become available. This work addresses this gap by introducing a deterministic, kernel-based framework for data-driven ODD construction. Instead of manually creating an ODD and iteratively optimizing it, the proposed method constructs the ODD directly from data using samples as anchor points for analytically defined kernel functions. The resulting ODD representation is continuous and bounded and, for a fixed calibration procedure, permutation-stable and uniquely determined by the available data. Importantly, the method remains applicable even in sparse-data regimes, making it suitable for both early development stages and mature systems. This work aims to provide a mathematically rigorous framework for deriving ODDs solely from data and supporting safety-by-design assurance activities. This work formalizes ODDs as mathematical structures, introduces a kernel-based affinity representation, and proposes a procedure for parameter selection and handling of out-of-distribution samples that runs fully automatically, given documented assurance inputs. By construction, the resulting ODD representation is bounded, deterministic, and interpretable, enabling its use not only for modeling and monitoring but also as a foundation for future certification of AI-based systems.

2 State of the Art

The ODD concept has gained acceptance in the automotive and aviation domains, with other domains now following [6, 28–30]. ODDs are intended to represent the real-world conditions under which automated systems, including AI-based systems, are expected to function, but necessarily remain approximations of those conditions [19, 21–23]. Originating from the automotive domain, the ODD is loosely defined as the specific conditions, such as environmental, geographic, and time-of-day restrictions, under which an autonomous driving system is expected to function [21, 22]. The definition of the European Union Aviation Safety Agency (EASA) is similar, extending it to any AI-based system and, by extension, to any artificial intelligence/machine learning (AI/ML) constituent [23]. Moreover, the ODD is split into two parts: the taxonomy and the ontology [19, 31]. The taxonomy describes the overall set of parameters and their applicable ranges, e.g., the car's current speed or the aircraft's altitude. The ontology, however, describes the interrelations among those parameters and how certain parameters influence one another. An example here might be the minimum possible turning radius of a car, depending on its speed, or the landing speed of an aircraft, depending on current weather and visibility. As such, the ODD plays a critical role in safety-by-design AI engineering methodologies to ensure compliance with applicable standards and regulations. However, creating the ODD is not trivial. While the ranges of some parameters can be quite evident, others, especially their interconnections, are almost impossible to define beforehand, for example, if they depend on complex processes like weather.

This clear commitment to ODDs as a standardized framework for describing operational conditions requires a comprehensive representation used not only for modeling and information exchange but also for monitors that can detect whether the system still operates within the ODD [19, 23, 25]. Although some ODD representations use structured representations such as tables or ASAM's YAML-based OpenODD format [20, 32, 33], extensive mathematical representations remain a subject of ongoing research. Most works represent the ODD as an n-dimensional space, with different

approaches to constraining it, e.g., convex polytopes, which provide a clean representation but cannot model nonlinear ODD ontologies [34]. Others provide an extensive framework for describing the ODD in great detail, but the ontology representation cannot handle constraints imposed by parameter combinations [35]. Research has also been conducted to split the ODD into *micro ODDs* (μODDs), which are used to improve handling and cover potentially complex ODDs [36, 37]. Thus, while ODDs can contain abstract definitions of the environment, they are, in general, currently defined by a parameter space that describes the overall taxonomy.

Constructing a membership region from sampled in-distribution (ID) points also connects the ODD-from-data problem to one-class classification and statistical support estimation. Classical estimators such as the one-class support vector machine [38], support vector data description [39], and kernel-density superlevel sets [40] likewise enclose a sampled distribution. The proposed affinity is globally a noisy-OR aggregation and, when overlap terms are small, is closely related to a variable-bandwidth kernel-density estimate through the per-anchor widths of Equation (9); scalable density-matrix approximations to kernel-density estimation also exist [41, 42]. Unlike the other estimators, the construction in this work targets the ODD-specific combination of a deterministic, permutation-stable membership region, an explicit out-of-distribution (OOD) consistency constraint with a proven termination bound, and a bounded affinity suitable for formal safety argumentation. Neural OOD detectors are excluded because incorporating learned detectors would introduce a separate training and assurance problem—which, in a safety-by-design setting, would itself require another ODD—beyond the scope of this work. The quantitative comparison with representative one-class baselines is summarized in Appendix A.

From the ODD, scenarios can be derived that provide a high-level description of operating conditions defined through a subset of parameter ranges [43, 44]. For example, the scenarios driving and flying at night in rainy weather would define the time of day to be between 6 p.m. and 6 a.m. and the precipitation and cloud coverage to be greater than zero. Sometimes, this is further split into logical and concrete scenarios, where logical scenarios define the aforementioned ranges, while concrete scenarios specify initial values for each parameter [45]. Nevertheless, a scenario is always a series of snapshots, called scenes [43]. Each scene is therefore a unique set of ODD parameter values, together with data from all sensors, given the current environmental conditions.

3 Mathematical Representation of ODDs

As previously described, multiple frameworks exist for structuring the ODD. However, all those representations lack a comprehensive yet generic framework for representing the ontology, as well as a means to compare two ODDs and determine whether they express the same problem. Both are solved with the following mathematical representation of an ODD.

Definition 3.1 (ODD Structures)**.** Let the ODD be defined as the structure $\mathcal{O} = (X, R_1^{\mathcal{O}}, \ldots, R_r^{\mathcal{O}}, f^{\mathcal{O}}, \Omega^{\mathcal{O}})$, where $X \subseteq \mathbb{R}^n$ is the domain, also called taxonomy, $R_i^{\mathcal{O}} = \{\boldsymbol{x} \in X \mid R_i(\boldsymbol{x}) = 1\} \subseteq X$ is the set of points satisfying the i-th predicate, also called ontology, and $f^{\mathcal{O}} \colon X \times \Omega^{\mathcal{O}} \to Y$, with the set of admissible noise values $\Omega^{\mathcal{O}}$ and the codomain Y, is the (stochastic) interpretation function. Thus, the signature is defined as $\sigma = \{R_1, \ldots, R_r, f, \Omega\}$ and the arities as $\mathrm{ar}(R_i) = 1$, $\mathrm{ar}(f) = n + \dim(\Omega)$, and $\mathrm{ar}(\Omega) = 0$.

Definition 3.2 (Similarity of ODD Structures)**.** Let $X_j \subseteq \mathbb{R}^{n_j}$ be the taxonomy of $\mathcal{O}_j$, $j \in \{1, 2\}$. Let $\mathcal{L}^m$ denote the m-dimensional Lebesgue measure. Two ODD structures $\mathcal{O}_1$ and $\mathcal{O}_2$ are considered similar, written $\mathcal{O}_1 \sim \mathcal{O}_2$, if and only if their generated datasets agree after discarding Lebesgue-null subsets of their input regions:

$$\begin{aligned} \mathcal{O}_1 \sim \mathcal{O}_2 \overset{\text{def}}{\iff} \exists Z_1 \subseteq \mathcal{R}^{\mathcal{O}_1}, Z_2 \subseteq \mathcal{R}^{\mathcal{O}_2} \colon \mathcal{L}^{n_1}(Z_1) = \mathcal{L}^{n_2}(Z_2) = 0, \\ f^{\mathcal{O}_1}\left((\mathcal{R}^{\mathcal{O}_1} \setminus Z_1) \times \Omega^{\mathcal{O}_1}\right) = f^{\mathcal{O}_2}\left((\mathcal{R}^{\mathcal{O}_2} \setminus Z_2) \times \Omega^{\mathcal{O}_2}\right). \end{aligned} \tag{1}$$

Here, $\mathcal{R}^{\mathcal{O}}$ is the conjunction of all relational predicates in $\mathcal{O}$, defined as $\mathcal{R}^{\mathcal{O}} \equiv \bigcap_{i=1}^{r} R_i^{\mathcal{O}}$.

This image-based notion is data-centric and intentionally abstracts away semantic differences, output frequencies, and safety weighting; refinements that compare induced output distributions or restrict comparison to hazardous outputs are left to future work.[1]

[1] As an example, given the definition of similarity, let two ODDs be defined as $\mathcal{O}_1 = (\mathbb{R}^2, \{\boldsymbol{x} \in X \mid x_1 \geq 0\}, \{\boldsymbol{x} \in X \mid x_2 \geq 0\}, \{\boldsymbol{x} \in X \mid x_1 + x_2 \leq 2\}, f^{\mathcal{O}_1}(\boldsymbol{x}) = x_1 + x_2, \Omega^{\mathcal{O}_1} = \{0\})$ and $\mathcal{O}_2 = (\{x \in \mathbb{R} \mid 0 \leq x \leq 2\}, f^{\mathcal{O}_2}(x) = x, \Omega^{\mathcal{O}_2} = \{0\})$. Thus, both generated output images equal $[0, 2]$, and therefore, $\mathcal{O}_1 \sim \mathcal{O}_2$.

4 The Kernel-Based Representation

Polytope-based ODD representations [34] (cf. Appendix B) work well for expert-defined ODDs but are unsuitable for data-driven construction: convex hulls may enclose regions outside the true ODD, and neural network–based representations are not order-independent, introducing avoidable uncertainty [46]. Instead, the proposed method assigns a kernel function to each anchor point and derives the ODD directly from data, without expert knowledge. The resulting representation is continuous and bounded, permutation-stable, and uniquely determined by the data under the calibration procedure of Section 4.3—in both sparse and dense regimes.

4.1 Formal Definition of the Kernel-Based ODD

The following subsection introduces a formal definition, followed by parameterization and algorithmic construction of the data-driven ODD. Let $X \subseteq \mathbb{R}^n$ denote the parameter space of the ODD, where each point $\boldsymbol{x} \in X$ represents a concrete operational condition described by n continuous parameters.

Definition 4.1 (Samples and Anchor Points)**.** Let $\mathcal{D} = \mathcal{D}_{\mathrm{ID}} \cup \mathcal{D}_{\mathrm{OOD}} \subseteq X$ be a finite dataset, where $\mathcal{D}_{\mathrm{ID}} \subseteq X$ denotes the set of in-distribution (ID) samples that are considered part of the ODD, and $\mathcal{D}_{\mathrm{OOD}} \subset X$ denotes the set of out-of-distribution (OOD) samples that are explicitly not part of the ODD. Thus, $\mathcal{D}_{\mathrm{ID}} \cap \mathcal{D}_{\mathrm{OOD}} = \emptyset$. Before anchor selection, the ID input records are de-duplicated using resolution cells fixed by documented per-dimension acquisition resolutions. Each occupied cell retains the observed record nearest its mean; the mean is accumulated over lexicographically ordered records, and ties are broken lexicographically. Merged record identifiers and visit counts are retained as provenance and coverage evidence but do not affect the ODD geometry. Cross-class conflicts must be resolved by the surrounding data-management process so that $\mathcal{D}_{\mathrm{ID}} \cap \mathcal{D}_{\mathrm{OOD}} = \emptyset$; the OOD samples themselves are not merged, and all remain subject to the consistency constraint of Definition 4.5. The anchor set $\mathcal{A} \subseteq \mathcal{D}_{\mathrm{ID}}$ is the finite (sub)set of retained ID samples used to parameterize the kernel-based ODD representation. Retained anchors and OOD samples are indexed in lexicographic order; geometrically identical OOD records are interchangeable because their provenance does not affect the construction. In general, all retained ID samples are used, i.e., $\mathcal{A} \equiv \mathcal{D}_{\mathrm{ID}}$.

Definition 4.2 (Local Affinity Function)**.** For each anchor point $\boldsymbol{x}_i \in \mathcal{A}$, $i = 1, \ldots, N$, a local affinity function $\alpha_i \colon X \to [0, 1]$ is defined using a positive-definite kernel.

Definition 4.3 (Global ODD Affinity Function)**.** The kernel-based ODD representation is defined as a global affinity function $\alpha \colon X \to [0, 1]$ constructed by superposition of all local affinity functions:

$$\alpha(\boldsymbol{x}) = 1 - \prod_{i=1}^{N} \left(1 - \alpha_i(\boldsymbol{x})\right). \tag{2}$$

The value $\alpha(\boldsymbol{x})$ represents the degree to which the operational condition $\boldsymbol{x}$ belongs to the ODD. The superposition is exactly the noisy-OR aggregation of the local affinities. Its deviation from their sum is bounded by

$$0 \leq \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}) - \alpha(\boldsymbol{x}) \leq \sum_{i<j} \alpha_i(\boldsymbol{x})\alpha_j(\boldsymbol{x}). \tag{3}$$

Definition 4.4 (Threshold-Based ODD Membership)**.** Given the global affinity function α, let $\zeta \in (0, 1)$ be the affinity threshold. For the deployed construction, ζ is obtained by the split-conformal calibration of Appendix A.6 from a held-out ID set and a documented false-exclusion budget ε. A sample $\boldsymbol{x} \in X$ is considered to be *inside* the data-driven ODD if and only if

$$\alpha(\boldsymbol{x}) \geq \zeta. \tag{4}$$

Definition 4.5 (OOD Consistency Constraint)**.** Assume $\mathcal{A} \cap \mathcal{D}_{\mathrm{OOD}} = \emptyset$. Let $\xi \in (0, \zeta)$, with ζ the membership threshold of the data-driven ODD, denote a predefined maximum allowed affinity for OOD samples; $\xi < \zeta$ guarantees that every observed OOD sample lies strictly outside the data-driven ODD. The kernel-based ODD representation shall satisfy

$$\alpha(\boldsymbol{x}) \leq \xi \quad \forall \boldsymbol{x} \in \mathcal{D}_{\mathrm{OOD}}. \tag{5}$$

If this constraint is violated, the most contributing kernels must be adjusted, while keeping the anchor points fixed, until the constraint is satisfied.

In practice, at each iteration, a globally most-violated OOD sample $\boldsymbol{x}^* \in \arg\max_{\boldsymbol{x}\in\mathcal{D}_{\text{OOD}}} \alpha(\boldsymbol{x})$ is selected, with the lowest canonical OOD-sample index selected on an exact tie. A kernel with the largest contribution to $\alpha(\boldsymbol{x}^*)$, i.e., $i^* \in \arg\max_i \alpha_i(\boldsymbol{x}^*)$, is then identified, again with the lowest canonical anchor index selected on an exact tie, and its covariance matrix is scaled by a fixed factor $c \in (0, 1)$. This process is repeated until all OOD samples satisfy $\alpha(\boldsymbol{x}) \leq \xi$. Its termination is proven in Appendix C. Together with canonical indexing and exact nearest-neighbor search, these tie rules make the construction permutation-stable in exact arithmetic. Repeated runs on identical inputs return identical output.

The previous definitions, together with the notation introduced in Section 3, are now used to define the kernel-based ODD.

Definition 4.6 (Kernel-based Operational Design Domain)**.** A kernel-based ODD is defined as the structure

$$\mathcal{O} = (X, R_{\text{M}}^{\mathcal{O}}, f^{\mathcal{O}}, \Omega^{\mathcal{O}}, a_1^{\mathcal{O}}, \ldots, a_N^{\mathcal{O}}), \tag{6}$$

where $R_{\text{M}}^{\mathcal{O}} = \{\boldsymbol{x} \in X \mid \alpha(\boldsymbol{x}) \geq \zeta\}$ is the sole membership predicate and $a_i^{\mathcal{O}} \in \mathcal{A}$ are the anchor points with $\text{ar}(a_i^{\mathcal{O}}) = 0$.

The fixed-parameter coverage and outer-halo behavior of the kernel-based representation, together with the two-sided convergence in volume of the calibrated construction, are analyzed in Appendix D.

4.2 Choice of Kernel

The affinity function must be continuous, bounded in $[0, 1]$, approach 0 for $\|\boldsymbol{x} - \boldsymbol{x}_i\|_2 \to \infty$, and scale to arbitrarily many dimensions. These requirements are satisfied by standard kernels such as the Laplacian kernel [47], the leptokurtic Cauchy–Lorentz kernel [48], the cubic spline kernel [49], and the radial basis function (RBF) kernel [50], used in this work. Its kernel function is defined as

$$\alpha_i(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{x}_i)\right), \tag{7}$$

for the i-th anchor point, where $\boldsymbol{x}_i \in \mathbb{R}^n$ is the anchor point, and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{n\times n}$ is the symmetric, positive-definite free parameter matrix. The multiplicative superposition in Definition 4.3 ensures the global affinity remains bounded in $[0, 1]$ even when anchor points coincide. Among the kernels above, the RBF kernel is selected for two reasons: its parameter matrix directly encodes per-dimension squared length scales, matching the nearest-neighbor parameterization of Section 4.3 and the diagonal approximation analyzed in Appendix E, and its logarithm is a quadratic form, permitting the exact, numerically stable log-space evaluation of Appendix F.1. The affinity superposition itself is kernel-agnostic; other kernels satisfying the stated requirements can be substituted, but their parameter calibration and termination analysis must be adapted accordingly.

4.3 Choosing the Kernel Parameters

Almost every kernel has a free parameter matrix. Defining all those parameters by hand would undermine the purpose of an automated, data-driven ODD derivation framework. Moreover, for the RBF kernel, N n-dimensional anchor points would lead to up to Nn^2 parameters to define. To allow automated parameter definition, some assumptions and simplifications must be made. First, to avoid implicit bias, the parameter ranges should be normalized. This would ensure that all ranges are again of the same order of magnitude. Thus, their corresponding matrix entries are also roughly of the same order of magnitude, reducing the likelihood that extreme matrix entries make the affinity function unstable. Second, the coordinate components of nearby anchor points are assumed to be locally independent, while their relationships are governed only by global effects. For the Gaussian RBF kernel, this assumption yields an axis-aligned factorization over the parameter dimensions and hence a diagonal $\boldsymbol{\Sigma}_i$ [51]. This construction combines the local-independence assumption with the principle of local stationarity, under which a process may behave as stationary within sufficiently small regions although its structure varies globally [52]. Under this assumption, the diagonal construction captures the relevant axis-aligned local scaling without introducing unsupported local cross-coordinate terms. Appendix E gives an exact error bound relative to a corresponding hypothetical full-covariance kernel when the assumption is violated; the discrepancy depends on the omitted off-diagonal structure and need not vanish solely because anchor density increases. The geometric density of distinct

retained operational states provides evidence of local support, whereas repeated observations within the documented acquisition resolution provide coverage evidence only and are removed before constructing the geometry. The corresponding parameter in the $\mathbf{\Sigma}$ matrix should thus increase with the density of distinct nearby anchors, resulting in a wider local kernel influence. Therefore, the parameters depend on nearest-neighbor distances between de-duplicated anchors; genuinely distinct nearby anchors still amplify one another through Equation (2). Thus, the entries of the $\mathbf{\Sigma}$ matrix can be defined as

$$\sigma_{kk}^{(i)} = \kappa \exp(-\eta d_{ik}^*), \quad k = 1, \ldots, n, \tag{8}$$

where $\sigma_{kk}^{(i)}$ are the diagonal entries of $\mathbf{\Sigma}_i$, κ is the maximum diagonal entry, and η is the decay factor. In global mode, $j_i^* := \min \arg\min_{j \neq i} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$, where the minimum selects the lowest canonical anchor index, and $d_{ik}^* := |x_{ik} - x_{j_i^* k}|$; in per-dimension mode, $d_{ik}^* := \min_{j \neq i} |x_{ik} - x_{jk}|$. This approach reduces the number of manually defined parameters from Nn^2 to 2, or $2n$ if κ and η are defined per dimension.

This, however, can lead to unstable matrix entries when anchor points are spread far apart, as the corresponding $\sigma_{kk}^{(i)}$ will be rather small and, in turn, the corresponding entries in $\mathbf{\Sigma}^{-1}$ can become numerically unstable. Here, a possible approach is to define a lower limit for Equation (8). This would lead to

$$\sigma_{kk}^{(i)} = (\kappa - \lambda) \exp(-\eta d_{ik}^*) + \lambda, \quad k = 1, \ldots, n, \tag{9}$$

where $\lambda \ll \kappa$ is the lower bound of $\sigma_{kk}^{(i)}$, specified relative to the maximum value, $\lambda = \lambda_{\text{rel}} \kappa$ with $\lambda_{\text{rel}} = \exp(-10)$, which bounds the condition number of $\mathbf{\Sigma}^{-1}$ by $1/\lambda_{\text{rel}}$ and, unlike an absolute lower bound, preserves the scale invariance established below.

While Equations (8) and (9) reduce the parameter count, the remaining choice of κ and η is crucial. The exponent of Equation (9) must be dimensionless and $\sigma_{kk}^{(i)}$ is a variance, so η carries the unit of an inverse length and κ that of a squared length. Fixing numerical values such as $\kappa = \eta = 1$ therefore implicitly selects a unit of length, and the resulting ODD changes under the very normalization recommended above. To resolve this, the parameters are calibrated from the data itself. Let the full-space nearest-neighbor distances and their median be $d_i^* := \min_{j \neq i} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$ and $\tilde{d} := \text{median}_i\, d_i^*$, respectively. The calibrated parameters are defined as

$$\eta = \frac{\gamma}{\tilde{d}}, \qquad \kappa = \left(s\tilde{d}\right)^2, \qquad \lambda = \lambda_{\text{rel}} \kappa, \tag{10}$$

with dimensionless constants γ and s, with defaults $\gamma = 1$ and the scheduling rule $s = s_N = \sqrt{\ln N}$, where N is the number of anchor points that instantiate kernels. Equation (9) then depends on the data only through the dimensionless ratio $d_{ik}^*/\tilde{d}$, scaled by the squared local length unit $\tilde{d}^2$: densely supported anchor points receive the maximum kernel width $s\tilde{d}$, i.e., s median neighbor gaps, wide enough to bridge the gaps between neighboring anchor points but no wider, while isolated anchor points receive exponentially narrower kernels, limiting their influence. The calibration is defined with respect to the global nearest neighbor: per-dimension distances measure the spacing of one-dimensional projections, which shrinks with the inverse number of anchor points and vanishes entirely for quantized parameters, and therefore provides no meaningful local length scale. At least two anchor points are required for the calibration to be defined. The schedule s_N, under which Theorem D.5 establishes two-sided convergence, is used for all calibrated experiments unless stated otherwise.

Proposition 4.7 (Scale Invariance of the Calibrated ODD)**.** *Let $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with $\boldsymbol{A} = \text{diag}(a_1, \ldots, a_n)$, $a_k > 0$, be an affine re-expression of the parameters. The affinity function derived with Equations* (9) *and* (10) *satisfies $\alpha'(T(\boldsymbol{x})) = \alpha(\boldsymbol{x})$ for all $\boldsymbol{x}$ whenever T is a similarity transformation, i.e., $a_1 = \cdots = a_n$. Moreover, under a per-dimension affine normalization with equivariant defining bounds, the full derivation pipeline is invariant under arbitrary per-dimension affine re-expressions of the raw data.*

The proof is given in Appendix D.1. In particular, under such a normalization, the derived ODD does not depend on the choice of units—a property the fixed parameterization $\kappa = \eta = 1$ does not possess.

4.4 Handling Out-of-Distribution Data

Often, the data used to define the data-driven ODD include not only ID samples as anchor points but also OOD samples explicitly outside the ODD. Here, it is important to ensure that these points are not later classified as part of the ODD. However, most kernels, especially the RBF kernel, cannot be zero anywhere; instead, a maximum affinity score can be defined that, at the locations of OOD samples, cannot be violated. Inspired by the Development Assurance Levels [53], depending on the classification of failure conditions of the AI-based systems, different maximum values can be defined.[2] After calculating the kernel parameters purely on ID samples, the global affinity function $\alpha(\boldsymbol{x})$ has to be sampled at all OOD sample locations. If the affinity is above the defined threshold, the most influential kernels have to be tuned such that $\alpha(\boldsymbol{x})$ stays below the threshold. This adjustment constrains the affinity only at observed OOD locations; unsafe regions absent from the OOD set are not guaranteed to be excluded, so the constraint complements rather than replaces expert-specified exclusions (cf. Section 6.1).

4.5 Creation of the Kernel-Based ODD

All previously mentioned steps are combined in Algorithm 1, which defines how to derive the ODD from data. First, the available records are collected and split into ID and OOD classes, and a held-out calibration set is reserved from the ID records. The remaining ID records are then de-duplicated using the documented acquisition resolutions, before anchor subsampling and kernel calibration. The held-out calibration records do not instantiate kernels and retain their original sampling frequencies, while every OOD sample remains in the consistency check. Next, a kernel is chosen, and its parameters are calibrated from the retained ID anchors as described in Section 4.3. After the OOD consistency adjustment, ζ is calibrated from the held-out ID set for the documented false-exclusion budget ε, as described in Appendix A.6. For L input ID records, deterministic cell construction and representative selection require at most $O(Ln \log L)$ preprocessing time and $O(Ln)$ space. The naïve algorithm (cf. Algorithm 1) runs in $O(N^2 n + N_{\mathrm{adj}} MNn + N_{\mathrm{cal}} Nn)$ time and requires $O(Nn)$ space with N retained anchors, M OOD samples, N_{adj} OOD-correction steps, and N_{cal} calibration records. Each subsequent query evaluates in $O(Nn)$ time. The individual terms and optional theoretical accelerations are discussed in Appendix H.

5 Validation

As part of this work, the autoSAFE tool was developed and open-sourced[3] to validate the proposed method for deriving the ODD $\mathcal{O}$ from data. Implementation details are given in Appendix F; hardware, runtimes, normalization, and default parameterizations are documented in Appendix A.1, with experiment-specific variations reported alongside their results.

5.1 Synthetic Benchmarks

The synthetic validation covers analytically specified half-space, annular, two-blob, banana-shaped, five-dimensional constrained, and rotated-band ODDs, including Monte Carlo configurations from two to twelve dimensions. The experiments examine anchor-count and parameter sensitivity, boundary halos, OOD adjustment, membership-threshold calibration, sampling density and duplicates, anchor-subset and permutation stability, covariance structure, and truncated-affinity scalability. Complete setups, hyperparameters, and results are reported in Appendices A, E and H. The two-dimensional suite compares autoSAFE with kernel density estimation (KDE) [40], a Gaussian mixture model (GMM) selected using the Bayesian information criterion (BIC) [54, 55], k-nearest-neighbor (k-NN) distance, a one-class support vector machine (OC-SVM) [38], Isolation Forest [56], and the convex hull; at $N = 1000$, autoSAFE remains within 0.015 48 area under the precision–recall curve (AUPR)

[2] Here, "fully automated" means execution given documented assurance inputs. The acquisition-resolution metadata used for de-duplication are likewise documented assurance inputs obtained from the recording-system specification, not parameters tuned against evaluation performance. The parameters κ, η, and λ are calibrated from data through Equation (10) using fixed γ and λ_{rel} and the scheduling rule $s_N = \sqrt{\ln N}$. The membership threshold ζ is conformally calibrated for a documented false-exclusion budget ε (cf. Appendix A.6), while $c \in (0, 1)$ is a fixed algorithmic step-size parameter affecting adjustment granularity and convergence rate. The OOD affinity bound ξ remains a documented assurance decision because the Development Assurance Levels prescribe no numerical affinity bound for a data-driven, kernel-based ODD.

[3] GitHub repository: `https://github.com/DLR-KI/autosafe`

of the strongest baseline and substantially outperforms the convex hull on the non-convex domains. Across the remaining tests, calibrated bandwidths control exterior expansion as coverage grows, all 72 OOD-adjustment configurations satisfy the specified affinity bound, conformal false-exclusion rates track their targets, and permutation-induced affinity differences remain at the level of floating-point rounding; in the sparse five-dimensional case, the conformal threshold keeps its target false-exclusion rate, but the region at a fixed threshold strongly depends on s.

5.2 Real-World Use Case

In aviation, one of the most important tasks of pilots during en route travel is avoiding collisions with other aircraft. The next-generation system, Advanced Collision Avoidance System X (ACAS X), is currently under development and expected to contain an AI-based system [57–60]. Preliminary implementations split the system into a vertical and a horizontal component, called the Vertical Collision Avoidance System (VCAS) and the Horizontal Collision Avoidance System (HCAS), respectively [61–63]; see Appendix I for details. Based on parameter ranges required for safe operation and previous experiments from other works [8, 64], a dataset of all observed state vectors has been created for each component, with 622 110 anchor points for VCAS and 385 925 for HCAS. Next, they were used to create the data-driven ODD $\mathcal{O}$. Once the data-driven ODD had been created, 10^6 validation samples were created: one half drawn uniformly from the parameter ranges defined in Tables 6 and 7, expanded by 10 % per continuous dimension, and one half by perturbing randomly selected anchor points with isotropic Gaussian noise of per-dimension standard deviation $3\tilde{d}/\sqrt{n}$, so that the expected displacement is comparable to the calibrated kernel scale. The discrete values of s_{adv} are not perturbed and only take their valid values. As before, for each validation sample, the affinity value and its membership in the known ODD and in the anchor-derived reference regions of Table 4 were calculated. Subsequently, the confusion matrices were computed for different affinity thresholds, and precision and recall vs. affinity threshold plots were generated. The thresholds are spaced adaptively toward both decision extremes and are evaluated directly through the log-survival rule of Equation (39); Appendix A.5 gives the exact grid. The resulting plots for VCAS and HCAS are shown in Figure 8, together with the split-conformal threshold $\hat{\zeta}_{0.1}$ for a false-exclusion budget of 10 % (cf. Appendix A.6), which is 0.9923 and 0.9861, for VCAS and HCAS, respectively. Precision remains above 0.99 at every threshold, whereas accepting every validation sample would only reach the shares of samples inside the known ODD, 0.789 for VCAS and 0.738 for HCAS. Between the smallest positive threshold and $\hat{\zeta}_{0.1}$, recall decreases from 0.633 to 0.592 and from 0.655 to 0.461, for VCAS and HCAS, respectively. Recall is lower than in the Monte Carlo example because the recorded states do not visit large parts of the known ODD, so many validation samples inside it lie far from every anchor point. Because the known VCAS ODD is available here, the curves are evaluated directly against that reference. In applications without independently specified ODD membership information, affinity thresholds instead require held-out validation data with the relevant labels. A separate VCAS OOD experiment selected 20 000 ID anchors uniformly at random without replacement from the 559 918 records outside an excluded region; here, anchor subsampling means that only the selected ID records instantiate kernels, while all 62 192 records in the excluded region remain OOD constraints. The experiment used 200 000 evaluation samples and terminated after 4389 adjustment steps, modifying 235 kernels and reducing the maximum observed OOD affinity to 0.009 97 for $\xi = 0.01$. This result is explicitly scoped to the subsampled VCAS construction; complementary full-data HCAS reference-agreement, seven-reference aviation, and synthetic and real-data anchor-subset results are reported in Tables 4 and 5 and Figure 7.

6 Discussion

The presented results demonstrate that the proposed kernel-based approach can effectively approximate an underlying ODD from data alone, given labeled ID samples and the calibration procedure of Section 4.3. This capability is crucial for legacy systems or complex environments where an analytic description of the ODD is unavailable or incomplete. A key finding is that agreement with a reference ODD can be quantified on held-out data through the affinity-threshold sweep.

The convex hull is generally inadequate as the final ODD representation. Real-world operational parameters often form non-convex, manifold-like structures with *holes* or concavities (e.g., specific combinations of altitude and speed that are physically impossible or unsafe). A convex hull heedlessly

encloses these void regions, potentially deeming the system safe in conditions where no data exist, and safety cannot be verified. In contrast, the region represented by the proposed kernel-based method follows the anchor points and largely excludes non-convex void regions, up to a boundary layer whose width depends on the local anchor spacing. On held-out points from the excluded VCAS region, the convex hull has a false-positive rate of 0.547, compared with 0.042 for autoSAFE without OOD samples and 0.0055 after adding the region's points as OOD constraints (cf. Table 3). At the fixed threshold $\zeta = 0.5$, the intersection over union (IoU) on the two-blob and banana ODDs is 0.855 and 0.775 for autoSAFE, compared with 0.488 and 0.338 for the convex hull; on the annulus, autoSAFE's IoU of 0.892 versus 0.898 for the hull is accompanied by a smaller mean false-positive boundary distance of 0.144 versus 0.358, distinguishing a thin exterior halo from errors inside the excluded hole (cf. Table 2). With the calibrated parameters of Equation (10), the affinity boundary layer scales with the local anchor spacing, so the representation is conservative in sparsely covered regions and tightens around the data as coverage increases (cf. Section 6.1 and Theorem D.5).

The OOD consistency adjustment was exercised on synthetic half-space and annular ODDs with $N = 1000$ anchors and $M \in \{10, 100, 1000\}$ uniformly sampled OOD points. To isolate the OOD-adjustment behavior, the 72 tested configurations use the fixed evaluation threshold $\zeta = 0.6$ and cover $\xi \in \{0.01, 0.1, 0.3\}$ and $c \in \{0.5, 0.8, 0.9, 0.99\}$. In every configuration, the adjustment terminated with all observed OOD affinities below ξ, after 0–32 551 iterations; across the grid, a single kernel was rescaled at most 553 times. Collateral effects were evaluated on separately sampled held-out points from the analytically specified in-distribution region, not on the anchors. At most 5.0 % of the held-out points that passed ζ before adjustment fell below it afterward. Across the same evaluation, recall decreased by at most 0.047, while precision increased by up to 0.090 as the over-inflated boundary tightened.

A held-out conformal calibration set determines ζ from data: across the tested error budgets, the empirical false-exclusion rate closely follows the target. For a documented false-exclusion budget ε, ζ is selected from a held-out ID calibration set so that, under exchangeability, a future ID sample falls below the threshold with probability at most ε (cf. Appendix A.6).

Furthermore, the continuous nature of the affinity function $\alpha(\boldsymbol{x})$ introduces a built-in mechanism for outlier detection and runtime monitoring. Unlike rigid geometric boundaries that provide a binary inside/outside classification, the affinity score offers a measure of proximity to known safe conditions. As the system state drifts away from the anchor points, $\alpha(\boldsymbol{x})$ decays smoothly. This property enables the definition of graded warning zones, allowing the system to trigger degradation modes or alert operators before the hard safety boundary is crossed. This *soft* boundary is a distinct advantage of the kernel-based representation over traditional Boolean logic ODD definitions.

6.1 Limitations

The presented kernel-based representation does have some limitations. First, the affinity depends on the selected kernel parameters; for the RBF kernel, these are κ, η, and λ. The calibration in Equation (10) removes dependence on coordinate scale and reduces the remaining choice to the dimensionless constants γ and λ_{rel}. For the tested two-dimensional ODDs with $N = 1000$, the calibrated parameter grid produces an absolute AUPR range of at most 0.0252. On the linear ODD, fixed pairs of κ and η on normalized data rank the validation points similarly well, with AUPR values of 0.966–0.995, but their level sets at $\zeta = 0.5$ reach an IoU of only 0.189–0.294, compared with 0.859–0.902 for s_N, because fixed values implicitly select a unit of length (cf. Figure 5). The sparse five-dimensional constrained ODD is more sensitive: its AUPR stays within 0.825–0.909 across the calibrated grid, but at $\zeta = 0.5$ the IoU reaches at most 0.029 for s_N, compared with up to 0.591 for $s = 1$. This highlights both the need for the conformal threshold calibration of Appendix A.6 and the potential value of a dimension-dependent scheduling rule as discussed in Appendix A.4. Second, the affinity-based representation has two density regimes. In sparsely covered regions, it can under-approximate the intended ODD, whereas near densely covered boundaries, fixed kernel parameters can over-approximate it through accumulated kernel tails. Proposition D.1 proves eventual interior coverage and bounds the represented region by an outer halo, but it does not establish two-sided set convergence at fixed parameters. The calibrated parameterization instead guarantees that the global containment radius shrinks to zero and that the symmetric-difference volume converges to zero under the assumptions of Theorem D.5. The measured fixed and calibrated regimes are shown in Figure 3. Third, the derived ODD reflects where operational data are dense, not

which operating conditions are safe in a causal or system-level sense. Operating conditions that are unsafe yet present in the recorded data remain inside the ODD, and valid conditions absent from the data are excluded; establishing the safety of the intended functionality within the ODD remains a separate assurance activity (cf. ISO 21448 [65] and UL 4600 [66]). Furthermore, the validation in Section 5 draws its samples from the same generating processes as the anchor data; robustness to distribution shift between development and operation, and to mislabeled ID or OOD samples, has not been evaluated. The data-driven ODD should therefore be read as one reproducible evidence artifact within a broader safety case (cf. Section 6.2), not as certification-level safety evidence on its own. Resolution-cell de-duplication depends on reliable acquisition metadata. A cell width larger than the actual recording resolution can merge genuinely distinct operational states, whereas a smaller width can preserve information-free multiplicity. The selected resolutions and resulting merge counts must therefore be documented as part of the data-assurance record. Nevertheless, this approach is not meant to deprecate outright the development of ODDs based on expert knowledge. It is meant to support current best practices in ODD development and to bolster development in areas with little to no expert knowledge.

6.2 Certification Impact

The proposed method provides a reproducible evidence artifact for the overall assurance and certification of AI-based systems. A well-defined ODD serves as the basis for the development of any AI-based system, whether in automotive [19, 22] or aviation [23]. Given the deterministic and explainable approach provided in this work, the derived ODD delineates the operational limits within which the safe deployment of an AI-based system can be argued. This, for example, feeds into EASA's learning assurance and its requirement for a continuous safety assessment. Moreover, the kernel-based representation can be used as a monitoring tool to detect whether the system remains within the defined boundaries. Thus, the proposed kernel-based approach supplements the proactive development of safety-by-design AI-based systems and the continuous monitoring of deployed systems, functioning as an early-detection mechanism before the system leaves its operational boundaries. Appendix J details selected objectives of the EASA concept paper [23] that the presented framework supports.

7 Conclusion

This work showed that, although a fully truthful data-driven reconstruction of an ODD is not always possible, the proposed approach can sufficiently approximate the underlying ODD. This work also presents a formal, data-centric definition of ODD similarity, necessary for future data-driven ODD development. Importantly, the presented method yields an ODD representation that is deterministic, permutation-stable, and conservative in sparsely covered regions (cf. Remark D.2), addressing key safety-by-design requirements on the path toward certification. By using a kernel-based representation, the framework provides a mathematically rigorous means to define operational boundaries that are interpretable and tractable. The novel, safety-by-design, deterministic kernel-based affinity representation of ODDs is derived fully automatically, given documented assurance inputs, via an algorithm that is permutation-stable, bounded by construction, and, under affine-equivariant per-dimension normalization, invariant to the choice of units. Unlike prior data-centric approaches [27], the proposed method makes no assumption about the parametric form of the ODD boundary, is permutation-stable with respect to sample ordering, and produces an affinity function that is bounded by construction and amenable to formal safety argumentation. Future work will focus on optimizing the kernel hyperparameter selection process. While this work focused on diagonal kernel parameter matrices to enhance robustness and interpretability, selectively extending the framework to model cross-dimensional dependencies could further improve expressiveness in tightly coupled systems. Such extensions would require careful constraints to preserve determinism and certifiability. Moreover, integrating temporal information through time-dependent kernels could enable the representation of evolving or context-dependent ODDs. Finally, future work should explore tighter integration of kernel-based ODDs into certification workflows. This includes linking affinity thresholds to formal safety requirements, leveraging the representation for runtime assurance monitors, and studying how data-driven ODD updates can be performed in a controlled and certifiable manner. Such developments would further strengthen the role of data-driven ODDs in deploying safety-critical AI-based systems.

Acknowledgments and Disclosure of Funding

This work received no external funding. The authors declare no competing interests.

References

[1] F. Constans, "Go around decision aid in the approach and landing phases for a large transport aircraft," in *2nd European Conference for Aerospace Sciences*, 07 2007, pp. 1–9.

[2] K. D. Julian, M. J. Kochenderfer, and M. P. Owen, "Deep neural network compression for aircraft collision avoidance systems," *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 3, pp. 598–608, 03 2019.

[3] D. Ziakkas, M. Wayne Suckow, and K. Pechlivanis, "The artificial intelligence (ai) certification challenges in future single pilot operations (sipo)," in *Human Factors in Transportation*, vol. 95. AHFE International, 07 2023, pp. 629–635.

[4] H. Bello, D. Geißler, L. Ray, S. Müller-Divéky, P. Müller, S. Kittrell, M. Liu, B. Zhou, and P. Lukowicz, "Towards certifiable ai in aviation: landscape, challenges, and opportunities," 09 2024.

[5] E. Erdogan and B. Baspinar, "Explainable and fast-converging artificial intelligence solution to control a nonlinear aircraft model in air combat," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2024, pp. 1–5.

[6] L. Greiner-Fuchs and M. Cichon, "A framework to define operational design domains for automated train operations," *IEEE Access*, vol. 12, pp. 167 022–167 039, 11 2024.

[7] U. Mandal, G. Amir, H. Wu, I. Daukantas, F. L. Newell, U. Ravaioli, B. Meng, M. Durling, K. Hobbs, M. Ganai, T. Shim, G. Katz, and C. Barrett, "Safe and reliable training of learning-based aerospace controllers," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2024, pp. 1–10.

[8] T. Stefani, J. M. Christensen, E. Hoemann, A. Anilkumar Girija, F. Köster, T. Krüger, and S. Hallerbach, "Applying model-based system engineering and devops on the implementation of an ai-based collision avoidance system," in *34th Congress of the International Council of the Aeronautical Sciences*. CEAS, 09 2024, pp. 1–12. [Online]. Available: https://www.icas.org/icas_archive/icas2024/data/preview/icas2024_0869.htm

[9] M. R. Schlichting, V. Rasmussen, H. Alazzeh, H. Liu, K. Jafari, A. F. Hardy, D. M. Asmar, and M. J. Kochenderfer, "Leraat: Llm-enabled real-time aviation advisory tool," in *ECAI 2025*. Amsterdam, The Netherlands: IOS Press, 10 2025, pp. 5167–5170.

[10] H. Forsberg, J. Linden, J. Hjorth, T. Manefjord, and M. Daneshtalab, "Challenges in using neural networks in safety-critical applications," in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 10 2020, pp. 1–7.

[11] F. Tambon, G. Laberge, L. An, A. Nikanjam, P. S. N. Mindom, Y. Pequignot, F. Khomh, G. Antoniol, E. Merlo, and F. Laviolette, "How to certify machine learning based safety-critical systems? a systematic literature review," *Automated Software Engineering*, vol. 29, no. 2, pp. 1–74, 11 2022.

[12] F. Mamalet, E. Jenn, G. Flandin, H. Delseny, C. Gabreau, A. Gauffriau, B. Beaudouin, L. Ponsolle, L. Alecu, H. Bonnin, B. Beltran, D. Duchel, J.-B. Ginestet, A. Hervieu, S. Pasquet, K. Delmas, C. Pagetti, J.-M. Gabriel, C. Chapdelaine, S. Picard, M. Damour, C. Cappi, L. Gardès, F. D. Grancey, B. Lefevre, S. Gerchinovitz, and A. Albore, "White paper machine learning in certified systems," IRT Saint Exupéry ; ANITI, Toulouse, France, Research Report, 03 2021. [Online]. Available: https://hal.science/hal-03176080

[13] S. M. Katz, K. D. Julian, C. A. Strong, and M. J. Kochenderfer, "Generating probabilistic safety guarantees for neural network controllers," *Machine Learning*, vol. 112, no. 8, pp. 2903–2931, 08 2023.

[14] M. Rabe, S. Milz, and P. Mader, "Development methodologies for safety critical machine learning applications in the automotive domain: A survey," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Piscataway, NJ, USA: IEEE, 06 2021, pp. 129–141.

[15] J. M. Christensen, T. Stefani, A. Anilkumar Girija, E. Hoemann, A. Vogt, V. Werbilo, U. Durak, F. Köster, T. Krüger, and S. Hallerbach, "Formulating an engineering framework for future ai certification in aviation," *Aerospace*, vol. 12, no. 6, pp. 1–27, 05 2025.

[16] I. Colwell, B. Phan, S. Saleem, R. Salay, and K. Czarnecki, "An automated vehicle safety concept based on runtime restriction of the operational design domain," in *2018 IEEE Intelligent Vehicles Symposium*. Piscataway, NJ, USA: IEEE, 06 2018, pp. 1910–1917.

[17] C. Sun, "Operational design domain monitoring and augmentation for autonomous driving," Ph.D. dissertation, UWSpace, 12 2022. [Online]. Available: https://hdl.handle.net/10012/18964

[18] C. Torens, S. Gupta, N. Roy, J. Sprockhoff, and U. Durak, "From operational design domain to runtime monitoring of ai-based aviation systems," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2024, pp. 1–9.

[19] The British Standards Institution, *Operational design domain (ODD) taxonomy for an automated driving system (ADS) – Specification*, The British Standards Institution Publicly Available Specification PAS 1883:2020, 08 2020. [Online]. Available: https://www.bsigroup.com/en-GB/insights-and-media/insights/brochures/pas-1883-operational-design-domain-odd-taxonomy-for-ads-specification/

[20] F. Werner, J. M. Christensen, T. Stefani, F. Köster, E. Hoemann, and S. Hallerbach, "Formulating a learning assurance-based framework for ai-based systems in aviation," *Aerospace*, vol. 13, no. 2, pp. 1–30, 02 2026.

[21] SAE International, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE International Surface Vehicle Recommended Practice J3016, 04 2021.

[22] International Organization for Standardization (ISO), *Road vehicles — Test scenarios for automated driving systems — Specification for operational design domain*, ISO/TC 22/SC 33 International Standard ISO 34 503:2023, 08 2023.

[23] European Union Aviation Safety Agency (EASA), "Easa concept paper: Guidance for level 1 & 2 machine learning applications," European Union Aviation Safety Agency (EASA), Cologne, Germany, Tech. Rep., 04 2024. [Online]. Available: https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2

[24] M. Gyllenhammar, R. Johansson, F. Warg, D. Chen, H.-M. Heyn, M. Sanfridson, J. Söderberg, A. Thorsén, and S. Ursing, "Towards an operational design domain that supports the safety argumentation of an automated driving system," in *10th European Congress on Embedded Real Time Software and Systems*, 01 2020, pp. 1–10. [Online]. Available: https://hal.science/hal-02456077

[25] C. Cappi, N. Cohen, M. Ducoffe, C. Gabreau, L. Gardes, A. Gauffriau, J.-B. Ginestet, F. Mamalet, V. Mussot, C. Pagetti, and D. Vigouroux, "How to design a dataset compliant with an ml-based system odd?" in *12th European Congress on Embedded Real Time Software and Systems*, 06 2024, pp. 1–10. [Online]. Available: https://hal.science/hal-04614554

[26] T. Stefani, A. Anilkumar Girija, R. Mut, S. Hallerbach, and T. Krüger, "From the concept of operations towards an operational design domain for safe ai in aviation," in *Deutscher Luft- und Raumfahrtkongress 2023*. Bonn, Germany: Deutsche Gesellschaft für Luft- und Raumfahrt - Lilienthal-Oberth e.V., 09 2023, pp. 1–8. [Online]. Available: https://elib.dlr.de/197957/

[27] F. Kaakai, S. Adibhatla, G. Pai, and E. Escorihuela, "Data-centric operational design domain characterization for machine learning-based aeronautical products," in *Computer Safety, Reliability, and Security*, ser. Lecture Notes in Computer Science, J. Guiochet, S. Tonetta, and F. Bitsch, Eds., vol. 14181. Cham, Switzerland: Springer Nature Switzerland, 09 2023, pp. 227–242.

[28] S. Huseynzade, "Creation of an operational design document for the autonomous shipping industry," Master's thesis, Mälardalen University, Västerås, Sweden, 10 2023. [Online]. Available: https://urn.fi/URN:NBN:fi-fe20231005138964

[29] M. Adedjouma, B. Botella, J. Ibanez-Guzman, K. Mantissa, C.-M. Proum, and A. Smaoui, "Defining operational design domain for autonomous systems: A domain-agnostic and risk-based approach," in *2024 19th Annual System of Systems Engineering Conference*. Piscataway, NJ, USA: IEEE, 06 2024, pp. 166–171.

[30] T. Stefani, J. M. Christensen, A. Anilkumar Girija, S. Gupta, U. Durak, F. Köster, T. Krüger, and S. Hallerbach, "Automated scenario generation from operational design domain model for testing ai-based systems in aviation," *CEAS Aeronautical Journal*, vol. 16, no. 1, pp. 197–212, 01 2025.

[31] R. Trypuz, P. Kulicki, and M. Sopek, "Ontology of autonomous driving based on the sae j3016 standard," *Semantic Web*, vol. 15, no. 5, pp. 1837–1862, 10 2024.

[32] Verein zur Förderung der internationalen Standardisierung von Automatisierungs- und Meßsystemen (ASAM) e.V., "Asam openodd: Concept paper," ASAM e.V., Höhenkirchen-Siegertsbrunn, Germany, Tech. Rep., 10 2021. [Online]. Available: https://www.asam.net/index.php?eID=dumpFile&t=f&f=4544&token=1260ce1c4f0afdbe18261f7137c689b1d9c27576

[33] ——, "Asam openodd base standard 1.0.0 specification," 04 2025. [Online]. Available: https://publications.pages.asam.net/standards/ASAM_OpenODD/ASAM_OpenODD/latest/specification/index.html

[34] V. Nenchev, "One stack, diverse vehicles: Checking safe portability of automated driving software," in *2025 IEEE/SICE International Symposium on System Integration*. Piscataway, NJ, USA: IEEE, 01 2025, pp. 764–769.

[35] A. Shakeri, "Formalization of operational domain and operational design domain for automated vehicles," in *2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion*. Piscataway, NJ, USA: IEEE, 07 2024, pp. 990–997.

[36] P. Koopman, B. Osyk, and J. Weast, "Autonomous vehicles meet the physical world: Rss, variability, uncertainty, and proving safety," in *Computer Safety, Reliability, and Security*, ser. Lecture Notes in Computer Science, A. Romanovsky, E. Troubitsyna, and F. Bitsch, Eds., vol. 11698. Cham, Switzerland: Springer International Publishing, 09 2019, pp. 245–253.

[37] A. Salvi, G. Weiss, M. Trapp, F. Oboril, and C. Buerkle, "Fuzzy interpretation of operational design domains in autonomous driving," in *2022 IEEE Intelligent Vehicles Symposium*. Piscataway, NJ, USA: IEEE, 06 2022, pp. 1261–1267.

[38] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 07 2001.

[39] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 01 2004.

[40] L. Devroye and G. L. Wise, "Detection of abnormal behavior via nonparametric estimation of the support," *SIAM Journal on Applied Mathematics*, vol. 38, no. 3, pp. 480–488, 06 1980.

[41] F. A. González, A. Gallego, S. Toledo-Cortés, and V. Vargas-Calderón, "Learning with density matrices and random features," *Quantum Machine Intelligence*, vol. 4, no. 2, pp. 1–17, 12 2022.

[42] J. A. Gallego, J. F. Osorio, and F. A. Gonzalez, "Fast kernel density estimation with density matrices and random fourier features," in *Advances in Artificial Intelligence – IBERAMIA 2022*, ser. Lecture Notes in Computer Science, A. C. Bicharra Garcia, M. Ferro, and J. C. Rodríguez Ribón, Eds., vol. 13788. Cham, Switzerland: Springer International Publishing, 11 2022, pp. 160–172.

[43] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and substantiating the terms scene, situation, and scenario for automated driving," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Piscataway, NJ, USA: IEEE, 09 2015, pp. 982–988.

[44] P. Weissensteiner, G. Stettinger, S. Khastgir, and D. Watzenig, "Operational design domain-driven coverage for the safety argumentation of automated vehicles," *IEEE Access*, vol. 11, pp. 12 263–12 284, 2023.

[45] X. Zhang, S. Khastgir, J.-K. Tiele, K. Takenaka, T. Hayakawa, and P. Jennings, "ODD and Behavior Based Scenario Generation for Automated Driving Systems," *IEEE Access*, vol. 12, pp. 10 652–10 663, 2024.

[46] M. J. Kochenderfer, *Algorithms for Validation*, S. M. Katz, A. L. Corso, and R. J. Moss, Eds. Cambridge, MA, USA: MIT Press, 2026.

[47] M. Rupp, "Machine learning for quantum mechanics in a nutshell," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1058–1073, 08 2015.

[48] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York, NY, USA: Wiley, 01 2009, vol. 2.

[49] J. J. Monaghan, "Smoothed particle hydrodynamics," *Annual Review of Astronomy and Astrophysics*, vol. 30, no. 1, pp. 543–574, 09 1992.

[50] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 2nd ed., ser. Springer Series in Statistics. New York, NY, USA: Springer New York, 02 2009.

[51] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. New York, NY, USA: Springer, 2006.

[52] M. Fuentes, "Spectral methods for nonstationary spatial processes," *Biometrika*, vol. 89, no. 1, pp. 197–210, 03 2002.

[53] SAE International, *Guidelines for Development of Civil Aircraft and Systems*, SAE International Aerospace Recommended Practice ARP4754B, 12 2023.

[54] G. J. McLachlan and D. Peel, *Finite Mixture Models*, ser. Wiley Series in Probability and Statistics. New York, NY, USA: Wiley, 09 2000.

[55] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 03 1978.

[56] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. Piscataway, NJ, USA: IEEE, 12 2008, pp. 413–422.

[57] European Organization for Civil Aviation Equipment (EUROCAE), *Minimum Operational Performance Standards for Airborne Collision Avoidance System X (ACAS X) (ACAS Xa with ACAS Xo)*, European Organization for Civil Aviation Equipment (EUROCAE) Std. ED-256, 10 2018.

[58] ——, *Minimum Operational Performance Standard (MOPS) for ACAS Xu*, European Organization for Civil Aviation Equipment (EUROCAE) Std. ED-275, 12 2020.

[59] RTCA, Inc., *Minimum Operational Performance Standards for Airborne Collision Avoidance System X (ACAS X) (ACAS Xa and ACAS Xo)*, RTCA, Inc. Std. DO-385, 10 2018. [Online]. Available: https://standards.globalspec.com/std/14222352/RTCA%20DO-385

[60] ——, *Minimum Operational Performance Standards for Airborne Collision Avoidance System Xu (ACAS Xu)*, RTCA, Inc. Std. DO-386, 12 2020. [Online]. Available: https://standards.globalspec.com/std/14360425/rtca-do-386-volume-1-2

[61] G. Katz, C. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer, "Reluplex: An efficient SMT solver for verifying deep neural networks," in *Computer Aided Verification*, ser. Lecture Notes in Computer Science, R. Majumdar and V. Kunčak, Eds., vol. 10426. Cham, Switzerland: Springer, 07 2017, pp. 97–117.

[62] K. D. Julian and M. J. Kochenderfer, "Guaranteeing safety for neural network-based aircraft collision avoidance systems," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2019, pp. 1–10.

[63] D. Manzanas Lopez, T. T. Johnson, S. Bak, H.-D. Tran, and K. L. Hobbs, "Evaluation of neural network verification methods for air-to-air collision avoidance," *Journal of Air Transportation*, vol. 31, no. 1, pp. 1–17, 01 2023.

[64] J. M. Christensen, A. Anilkumar Girija, T. Stefani, U. Durak, E. Hoemann, F. Köster, T. Krüger, and S. Hallerbach, "Advancing the ai-based realization of acas x towards real-world application," in *2024 IEEE 36th International Conference on Tools with Artificial Intelligence*. Piscataway, NJ, USA: IEEE, 10 2024, pp. 57–64.

[65] International Organization for Standardization (ISO), *Road vehicles — Safety of the intended functionality*, ISO/TC 22/SC 32 International Standard ISO 21 448:2022, 2022.

[66] Underwriters Laboratories, *Standard for Evaluation of Autonomous Products*, Underwriters Laboratories Standard UL 4600, 03 2023.

[67] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 09 1949.

[68] A. N. Angelopoulos and S. Bates, "Conformal prediction: A gentle introduction," *Foundations and Trends in Machine Learning*, vol. 16, no. 4, pp. 494–591, 03 2023.

[69] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Palo Alto, CA, USA: AAAI Press, 08 1996, pp. 226–231.

[70] L. C. Evans, *Measure Theory and Fine Properties of Functions*, 2nd ed., ser. Textbooks in Mathematics. NY, USA: Chapman and Hall/CRC, 03 2025.

[71] D. Eppstein, M. S. Paterson, and F. F. Yao, "On nearest-neighbor graphs," *Discrete & Computational Geometry*, vol. 17, no. 3, pp. 263–282, 04 1997.

[72] C. McDiarmid, "On the method of bounded differences," in *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference*, ser. London Mathematical Society Lecture Note Series, J. Siemons, Ed. Cambridge, UK: Cambridge University Press, 08 1989, vol. 141, pp. 148–188.

[73] R. Durrett, *Probability: Theory and Examples*, 5th ed., ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press, 04 2019, no. 49.

[74] M. Penrose, *Random Geometric Graphs*, ser. Oxford Studies in Probability. Oxford, UK: Oxford University Press, 05 2003, no. 5.

[75] M. Mächler, "Accurately computing $\log(1 - \exp(-|a|))$ assessed by the rmpfr package," 09 2015. [Online]. Available: https://CRAN.R-project.org/package=Rmpfr/vignettes/log1mexp-note.pdf

[76] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., ser. Other Titles in Applied Mathematics. Philadelphia, PA, USA: SIAM, 01 2002, no. 80.

[77] Verein zur Förderung der internationalen Standardisierung von Automatisierungs- und Meßsystemen (ASAM) e.V., "Asam openlabel 1.0.0," 11 2021. [Online]. Available: https://www.asam.net/standards/detail/openlabel/

[78] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 09 2020.

[79] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 03 2020.

[80] I. Filippidis, S. Dathathri, S. C. Livingston, N. Ozay, and R. M. Murray, "Control design for hybrid systems with tulip: The temporal logic planning toolbox," in *2016 IEEE Conference on Control Applications*. Piscataway, NJ, USA: IEEE, 09 2016, pp. 1030–1041.

[81] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 07 2021.

[82] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou, "The faiss library," *IEEE Transactions on Big Data*, vol. 12, no. 2, pp. 346–361, 04 2026.

[83] M. Damour, F. De Grancey, C. Gabreau, A. Gauffriau, J.-B. Ginestet, A. Hervieu, T. Huraux, C. Pagetti, L. Ponsolle, and A. Clavière, "Towards certification of a reduced footprint acas-xu system: A hybrid ml-based solution," in *Computer Safety, Reliability, and Security*, ser. Lecture Notes in Computer Science, vol. 12852. Cham, Switzerland: Springer International Publishing, 09 2021, pp. 34–48.

[84] K. D. Julian, J. Lopez, J. S. Brush, M. P. Owen, and M. J. Kochenderfer, "Policy compression for aircraft collision avoidance systems," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2016, pp. 1–10.

[85] J. M. Christensen, W. Zaeske, J. Beck, S. Friedrich, T. Stefani, A. A. Girija, E. Hoemann, U. Durak, F. Köster, T. Krüger, and S. Hallerbach, "Towards certifiable ai in aviation: A framework for neural network assurance using advanced visualization and safety nets," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference*. Piscataway, NJ, USA: IEEE, 09 2024, pp. 1–9.

[86] European Union Aviation Safety Agency (EASA), "Easa concept paper: Guidance for safety-related artificial intelligence applications," European Union Aviation Safety Agency (EASA), Cologne, Germany, Tech. Rep., 06 2026. [Online]. Available: https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-proposed-issue-3

A Additional Validation Results

A.1 Experimental Setup

All experiments were conducted on a system with an Intel Core i9-13900K, an NVIDIA RTX 4090, and 64 GB of RAM using Python 3.14.3. The original Monte Carlo experiments completed in under 5 min, the additional validation experiments individually required between 9 s and 15 min, and the aviation experiments approximately 11 h–14 h each. The original Monte Carlo experiments used the manual parameterization $\kappa = 1$, $\eta = 1$, and $\lambda_{\text{rel}} = \exp(-10)$ on the raw parameter space. Unless stated otherwise in the corresponding experiment description, the calibrated synthetic experiments used per-dimension normalization to $[-1, 1]$, $\gamma = 1$, the scheduling rule $s = s_N = \sqrt{\ln N}$, and $\lambda_{\text{rel}} = \exp(-10)$. Threshold-free metrics such as the AUPR rank the validation points by the log-survival score $-S(\boldsymbol{x})$ of Equation (38), which orders them exactly as α does but stays distinguishable where α rounds to one in finite precision. For the aviation experiments, normalization was based on the known ODD definitions derived from prior research [62] and summarized in Tables 6 and 7.

A.2 Monte Carlo Validation

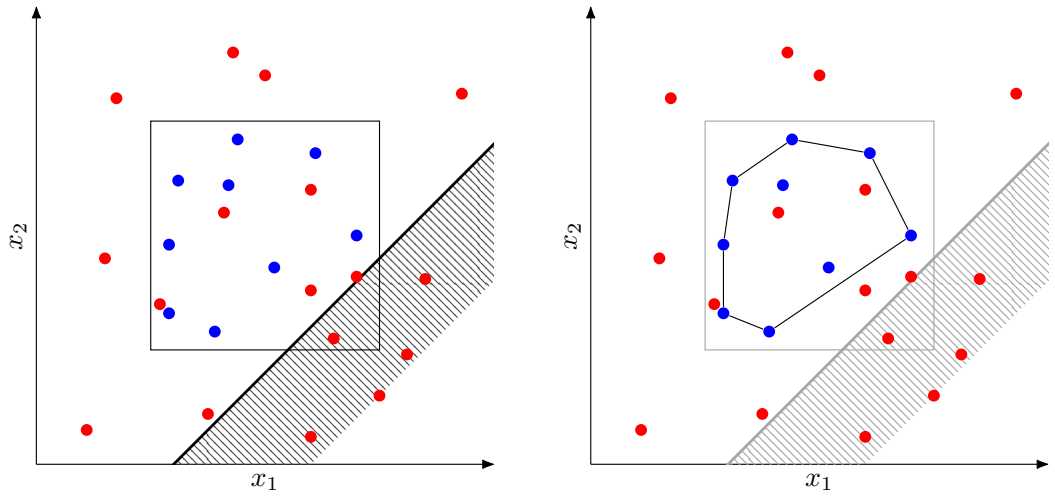
Monte Carlo methods (MCM) [67] have proven useful for numerical problem solving. Here, MCM is used to generate synthetic data for analytically specified ODDs by sampling anchor points uniformly within each ODD and validation points uniformly from a larger enclosing domain, thereby enabling affinity-threshold membership to be compared directly with known ODD membership. The experiments cover ODD structures from two to twelve dimensions. The two-dimensional anchor-count sweep described below additionally isolates the effect of data coverage from the sparsest tested case to dense sampling.

A representative two-dimensional case is depicted in Figure 1a, where the blue dots are the anchor points, the red dots are the validation samples, the black rectangle is the polytope X, and the black line is the boundary specified by R_1. For this setup, the data-driven ODD $\mathcal{O}$ was derived according to Algorithm 1, although with manual parameterization, across anchor counts ranging from the sparsest tested configuration to dense sampling. More precisely, X was defined as $[-5, 5] \times [-5, 5]$, $R_1^{\mathcal{O}} = \{\boldsymbol{x} \in X \mid x_2 \geq x_1 - 3\}$, and the tested anchor counts were $N \in \{3, 5, 10, 25, 30, 50, 100, 300, 1000, 3000, 10\,000\}$. For each anchor count, five independently seeded runs were performed, each using a newly sampled anchor set and 100 000 validation points. The large validation set was chosen to provide coverage inside and outside the analytically specified ODD, including regions close to the anchors. The validation points were drawn uniformly from $[-10, 10] \times [-10, 10]$, whose side lengths are twice those of X.

Next, for all validation points and all anchor point configurations, the affinity and whether each point lies within the original ODD were calculated. Based on the data, the confusion matrices were computed for different affinity thresholds. The evaluation uses an adaptive log-space threshold grid whose exact construction is given in Appendix A.5. Here, the affinity threshold determines whether a validation sample point is classified as inside the data-driven ODD. Validation samples with affinity values at or above the affinity threshold ζ are considered within the data-driven ODD; those below it are not. Given the results, precision and recall vs. affinity threshold plots were created (cf. Figure 2). With these fixed kernel parameters, increasing the number of anchor points decreases precision while increasing recall, consistent with the outer halo of Proposition D.1; the calibrated parameters avoid this expansion (cf. Figure 3).

Moreover, the sharp edges where the affinity is either zero or one follow from Equation (2): the RBF kernel cannot be exactly zero, so no validation sample attains an affinity of exactly zero, while an affinity of exactly one would, in exact arithmetic, require a validation sample to land precisely on an anchor point. In finite-precision arithmetic, however, samples deep inside the ODD can saturate to an affinity of exactly one; the reported results therefore use the log-space rule of Equation (39). Given the precision and recall vs. affinity threshold plots, one can derive an affinity threshold for the use case by balancing precision and recall within overall constraints, such as required safety levels.

The underlying ODD is known for this synthetic validation. However, it will generally not be available in a real-world application. Therefore, the data-driven ODD is also compared with the convex hull of all anchor points as a geometric reference. This approach is illustrated in Figure 1b, where the original ODD structure is still drawn, albeit faintly. The same steps as before were repeated, and



(a) Setup for comparing the data-driven ODD $\mathcal{O}$ to the underlying original ODD. The black rectangle represents the polytope X and the black line depicts the inequality R_1.

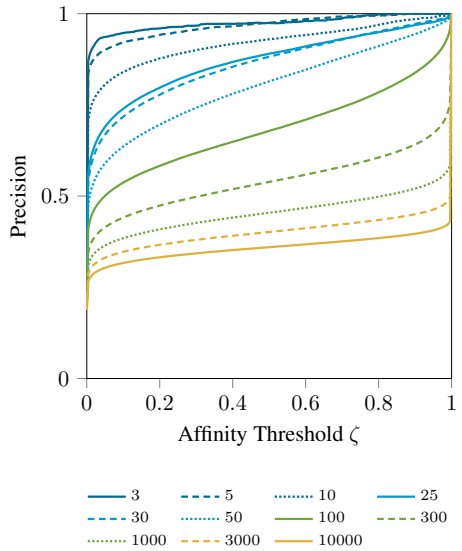
(b) Setup for comparing the data-driven ODD $\mathcal{O}$ to the convex hull of all anchor points.

Figure 1: Setup for the two-dimensional example of the Monte Carlo method. The blue dots are the anchor points from which the data-driven ODD $\mathcal{O}$ is derived. The red dots are the validation samples to compare the data-driven ODD $\mathcal{O}$ to the underlying original ODD.

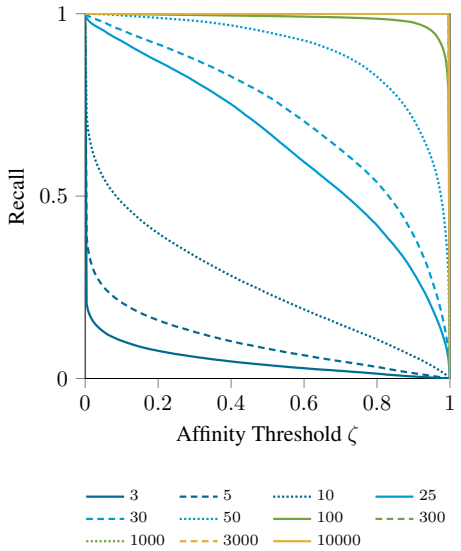
corresponding precision and recall vs. affinity threshold curves were computed using convex-hull membership as the reference. The per-N analysis compares the convex hull–referenced curves with the underlying ODD–referenced curves on a shared threshold grid. For two membership curves q^{ODD} and q^{ref}, with q denoting precision or recall, the curve agreement is $R^2 = 1 - \sum_j (q_j^{\mathrm{ref}} - q_j^{\mathrm{ODD}})^2 / \sum_j (q_j^{\mathrm{ODD}} - \bar{q}^{\mathrm{ODD}})^2$, where the underlying-ODD curve is the target.

The precision-curve R^2 is negative for $N \leq 50$, e.g., –3.06 at $N = 25$, becomes positive at $N = 100$, and reaches 0.971 at $N = 1000$ and 0.999 65 at $N = 10\,000$, where the recall-curve R^2 reaches 1.000 (cf. Figure 4). In this setting, the convex hull is therefore a usable proxy reference only beyond approximately 100 anchors. The corresponding area under the precision–recall curve (AUPR) reaches $0.999\,02 \pm 0.000\,11$, whereas at $N = 5$, 25, and 50 it is 0.734 ± 0.056, 0.903 ± 0.020, and 0.949 ± 0.012, respectively. These results quantify both the sparse-data regime and the observed improvement toward the underlying ODD over the tested anchor-count range. At $N = 1000$, autoSAFE is within 0.015 48 AUPR of the strongest tested baseline throughout the two-dimensional suite and substantially exceeds the convex hull on the non-convex annulus and two-blob domains (cf. Table 1).

The boundary-halo experiment in Figure 3 complements the threshold-sweep results by directly measuring exterior expansion at $\zeta = 0.5$ as the anchor count increases. The fixed-parameter halo grows from 1.027 to 3.426, whereas the calibrated halo shrinks to 0.102. These measurements illustrate why calibration is needed to control exterior expansion under denser sampling.

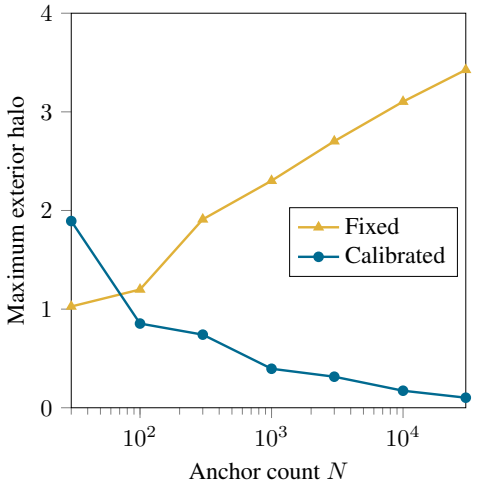


(a) Mean precision against known ODD membership for each tested anchor count.

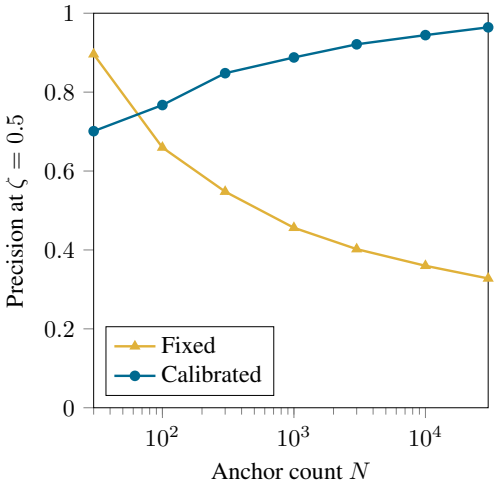


(b) Mean recall against known ODD membership for each tested anchor count.

Figure 2: Precision and recall vs. affinity threshold for the Monte Carlo anchor-count sweep. Each curve is the mean over five seeds for one anchor count, exposing the sparse-data regime and the progression with increasing coverage.



(a) Maximum distance from a point classified as in the ODD at $\zeta = 0.5$ to the known ODD boundary as the number of anchor points increases.



(b) Precision of the represented region at $\zeta = 0.5$ for the same anchor counts and parameterizations.

Figure 3: Measured exterior halo and precision as the number of anchor points increases; both horizontal axes use a logarithmic scale. The fixed-parameter halo grows from 1.027 to 3.426, whereas the calibrated halo shrinks to 0.102.

A.3 One-Class Baselines

Table 1 compares the represented region with standard one-class estimators on synthetic domains with known membership. The baselines are KDE, GMM selected using BIC, k-NN distance, OC-SVM, Isolation Forest, and the convex hull. Given support vector data description yields, for kernels with constant self-similarity such as the RBF kernel, the same decision function as the OC-SVM with correspondingly chosen regularization, it is not separately evaluated [38, 39]. The comparison does not claim superior density estimation; it shows that the proposed assurance-oriented construction remains competitive on the two-dimensional suite while retaining its behavior on non-convex supports. All two-dimensional domains use $X = [-5, 5]^2$: the linear ODD satisfies $x_2 \geq x_1 - 3$, the annulus ODD satisfies $1 \leq x_1^2 + x_2^2 \leq 16$, the two-blob ODD is the union of disks of radius 1.5 centered at $(-2.5, -2.5)$ and $(2.5, 2.5)$, and the banana ODD satisfies $|x_2 - 0.4x_1^2 + 2| \leq 1$. The five-dimensional constrained ODD is the intersection of $X = [-5, 5]^5$, $x_2 \geq x_1 - 3$, and $\sum_{k=1}^{5} x_k^2 \leq 60$. For each domain, $100\,000$ validation points are sampled uniformly from the taxonomy box enlarged by a factor of two about its center. KDE selects the greatest training log-likelihood among Scott's rule, Silverman's rule, and bandwidths 0.05, 0.1, 0.2, and 0.5. The GMM selects 1, 2, 4, or 8 components by BIC. The OC-SVM selects the greatest mean training decision margin over $\nu \in \{0.01, 0.05, 0.1\}$ and kernel coefficients $\gamma \in \{\text{scale}, 1, 10\}$. The k-NN baseline uses $k = 5$, Isolation Forest uses 200 estimators, and every experiment uses random seed 42.

Table 1: AUPR against the known region for $N = 1000$ anchors. autoSAFE ranks the validation points by the log-survival score $-S(\boldsymbol{x})$. Values are rounded to three decimal places; boldface indicates the method with the highest AUPR in each dataset.

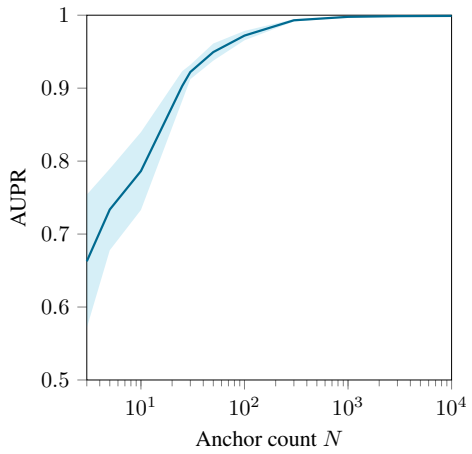
Method	Linear	Annulus	Two Blobs	Banana	5D ODD	Order Deviation
autoSAFE	0.992	0.990	0.985	0.970	0.856	$5.8 \cdot 10^{-15}$
KDE	0.996	0.995	0.996	**0.986**	0.692	—
GMM (BIC)	**0.998**	**0.998**	**0.997**	0.983	**0.882**	—
k-NN distance	0.991	0.991	0.984	0.970	0.854	—
OC-SVM	0.995	0.817	0.997	0.337	0.868	—
Isolation Forest	0.952	0.918	0.972	0.821	0.675	—
Convex hull	0.988	0.901	0.482	0.332	0.632	—

Table 2: Fixed-threshold comparison against the known non-convex regions for $N = 1000$ anchors. autoSAFE is evaluated at $\zeta = 0.5$, while convex-hull membership is binary. The mean false-positive boundary distance is approximated in the original coordinate space by the distance from each false-positive validation point to its nearest validation point with the opposite true label; lower values are better.

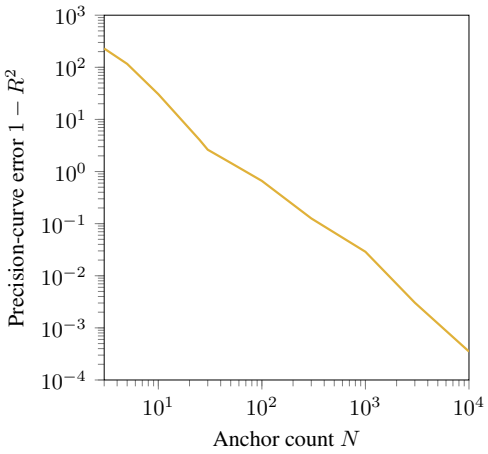
ODD	Method	IoU	False-positive rate	Mean false-positive boundary distance
Annulus	autoSAFE	0.892	0.016	0.144
Annulus	Convex hull	0.898	0.009	0.358
Two blobs	autoSAFE	0.855	0.006	0.107
Two blobs	Convex hull	0.488	0.037	1.043
Banana	autoSAFE	0.775	0.012	0.113
Banana	Convex hull	0.338	0.080	1.311

The five-dimensional constrained ODD is the most challenging case of the suite: autoSAFE ranks third, behind GMM and OC-SVM, and its region at the fixed threshold of $\zeta = 0.5$ strongly over-approximates the true region, with a false-positive rate of 0.864, as the informative thresholds lie very close to one (cf. Section 6.1). The order-deviation value is the maximum absolute affinity difference across five permutations of the ID anchors and was measured only for autoSAFE; the dashes denote unavailable measurements for the other methods.

A.4 Anchor Count and Parameter Sensitivity



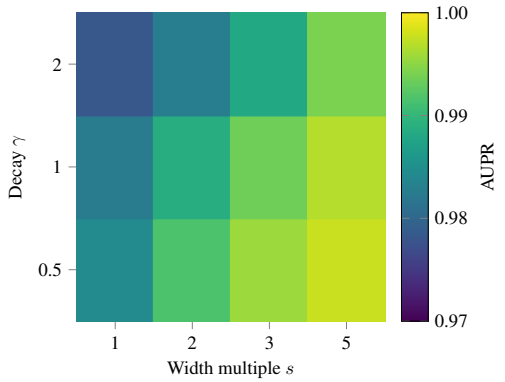
(a) Mean AUPR against the known ODD as the number of anchors increases. The shaded region shows one standard deviation across five random seeds.



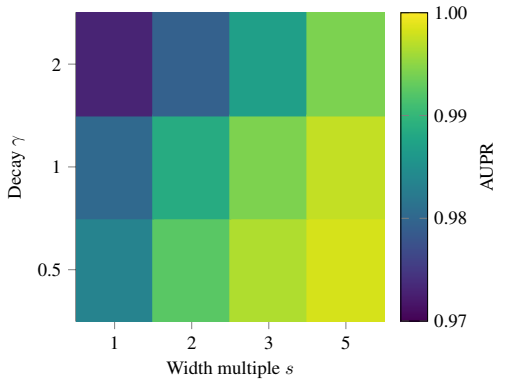
(b) Disagreement, measured as $1 - R^2$, between the precision–affinity-threshold curves obtained using the known ODD and the convex hull as references.

Figure 4: AUPR and precision-curve agreement for the two-dimensional linear ODD as the number of anchors increases. The AUPR reaches $0.999\,02 \pm 0.000\,11$ and the precision-curve agreement reaches $R^2 = 0.999\,65$ at $N = 10\,000$.

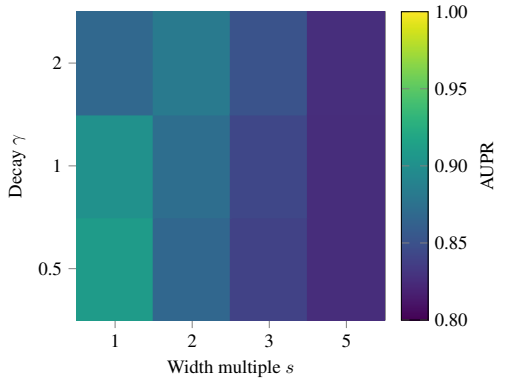
The sensitivity of the sparse five-dimensional ODD to s suggests that the width multiple should decrease with the dimension n: the number of anchors within a fixed multiple of $\tilde{d}$ grows rapidly with n, and with it the accumulated noisy-OR tail at any fixed s. A dimension-dependent constant $c_n > 0$ in the scheduling rule, $s_{n,N} = c_n s_N$, leaves Theorem D.5 unchanged because n is fixed as N grows. For $c_n = n^{-1/2}$ and $\gamma = 1$, the IoU of the five-dimensional ODD at $\zeta = 0.5$ rises from 0.022–0.029 without the correction factor c_n to 0.217–0.289 with the correction factor c_n for anchor counts from 100 to 10 000, and its AUPR at $N = 1000$ from 0.850 to 0.906. On the two-dimensional ODDs at $N = 1000$, the IoU changes from 0.894 and 0.870 to 0.901 and 0.917, whereas at $N = 100$ the annulus IoU decreases from 0.772 to 0.679. A stronger scaling of $c_n = n^{-1}$ narrows the two-dimensional kernels excessively, with IoU values of 0.709 and 0.772 at $N = 1000$. Since even $c_n = n^{-1/2}$ remains below the five-dimensional IoU of 0.645 reached at $s = 0.75$, a principled and justified choice of c_n is left to future work.



(a) The two-dimensional linear ODD is the intersection of $X = [-5, 5]^2$ with the half-space $x_2 \geq x_1 - 3$.



(b) The two-dimensional annulus ODD is the ring $1 \leq x_1^2 + x_2^2 \leq 16$ within $X = [-5, 5]^2$, which excludes the central hole.



(c) The five-dimensional constrained ODD is the intersection of $X = [-5, 5]^5$, the half-space $x_2 \geq x_1 - 3$, and the ball $\sum_{k=1}^{5} x_k^2 \leq 60$.

Figure 5: AUPR sensitivity to the dimensionless calibration constants s and γ at $N = 1000$. Across the full grid, the AUPR remains within 0.978–0.998 for the linear ODD and within 0.973–0.998 for the annulus ODD. For the same anchor count, the five-dimensional constrained ODD ranges over 0.825–0.909 and reaches its maximum at $\gamma = 0.5$ and $s = 1$. The scheduling rule $s_N \approx 2.63$ at this anchor count yields AUPR values of 0.986–0.994, 0.984–0.995, and 0.846–0.860 for the three ODDs.

A.5 Adaptive Threshold Grid

Let $L \geq 2$ be the requested number of threshold pairs and let $S = \ln(1 - \zeta)$ be the corresponding log-survival threshold. The grid always contains the endpoint pairs $(\zeta, S) = (0, 0)$ and $(1, -\infty)$. Approximately one quarter of the remaining $L - 2$ entries are assigned to a lower logarithmic edge, one quarter to a linear middle interval, and one half to an upper logarithmic edge. For the lower edge, t is spaced linearly from 1 to d_{lo} and $\zeta = 10^{-t}$, where d_{lo} is the larger of 16 and $-\log_{10} \zeta_{\min}^{+}$ for the smallest positive observed affinity $\zeta_{\min}^{+}$. The middle section is linear in ζ from 0.1 to 0.9. For the upper edge, t is spaced linearly from 1 to d_{hi}, $S = -t \ln 10$, and $\zeta = 1 - \exp(S)$, where d_{hi} is the larger of 16 and $|S_{\min}|/\ln 10$ for the most negative finite observed log-survival. Duplicate junction pairs are removed, and all decisions use the stored value of S directly through Equation (39).

A.6 Distribution-Free Calibration of the Membership Threshold

Let $Q(\boldsymbol{x}) := -S(\boldsymbol{x}) = -\ln(1 - \alpha(\boldsymbol{x}))$ be the log-space affinity score and let $Q_{(1)} \leq \cdots \leq Q_{(N_{\mathrm{cal}})}$ be the ordered scores of a held-out ID calibration set of size N_{cal} that is exchangeable with a future ID sample. For a target false-exclusion budget $\varepsilon \in (0, 1)$, define $k := \lfloor \varepsilon(N_{\mathrm{cal}} + 1) \rfloor$, $\hat{t} := Q_{(k)}$, and

$\hat{\zeta} := 1 - \exp(-\hat{t})$. The construction is certifiable only when $k \geq 1$, which requires $N_{\text{cal}} \geq 1/\varepsilon - 1$. By the split-conformal rank argument [68],

$$\mathbb{P}\left(\alpha(\boldsymbol{x}_{N_{\text{cal}}+1}) < \hat{\zeta}\right) = \mathbb{P}\left(Q(\boldsymbol{x}_{N_{\text{cal}}+1}) < \hat{t}\right) \leq \frac{k}{N_{\text{cal}}+1} \leq \varepsilon. \tag{11}$$

The calibration set is reserved before anchor de-duplication, is disjoint from the anchor set, and retains the operational record frequencies required for exchangeability with a future ID record. Its scores are computed using the final representation after anchor de-duplication and any OOD adjustment. The required ordering $\xi < \hat{\zeta}$ is checked explicitly when the two thresholds are used together. The conformal guarantee controls false exclusion only; false inclusion is addressed separately by the outer-halo analysis and the observed-OOD consistency constraint in Appendix D and Definition 4.5. Calibrating on Q preserves the ordering of α while retaining distinctions that finite-precision affinity values can lose near one. For each of the linear, annular, and five-dimensional constrained domains, total ID pools of size $N_{\text{pool}} \in \{500, 1000, 5000\}$ were split equally into $N = N_{\text{cal}} \in \{250, 500, 2500\}$ anchors and calibration records, with fresh sets of 5000 ID and 5000 out-of-region test samples per repetition. Across 20 repetitions, the observed false-exclusion-rate ranges are 0.0063–0.0110, 0.0493–0.0523, and 0.0985–0.1035 for $\varepsilon = 0.01$, 0.05, and 0.10, respectively, while 93.6 %–99.8 % of out-of-region points remain excluded.

A.7 Non-Convex Exclusion and Anchor-Set Robustness

The excluded VCAS region is defined by $|h| \leq 300\,\text{ft}$ and $\tau \leq 8\,\text{s}$; it represents low vertical separation shortly before the closest point of approach (CPA).

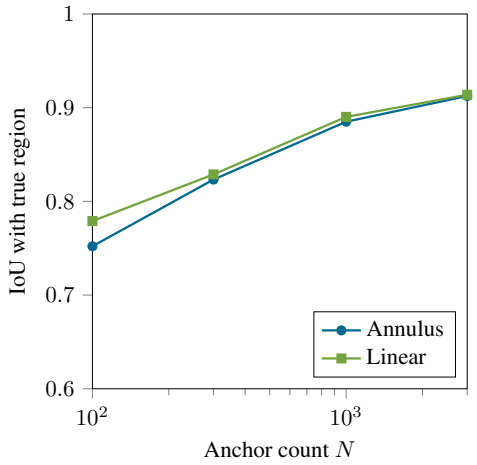
Table 3: False-positive rate on held-out points from the excluded VCAS region. Relative changes use autoSAFE without OOD constraints as the baseline; lower values are better.

Method	False-positive rate	Relative change from autoSAFE
autoSAFE	0.0420	—
autoSAFE with OOD constraints	0.0055	86.9 % improvement
Convex hull	0.5470	1202.4 % worsening

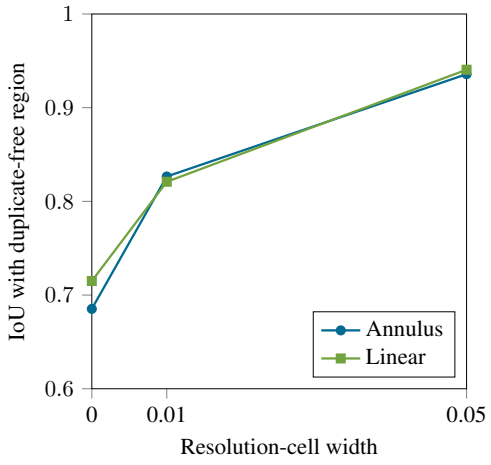
The sampling-density and resolution-cell de-duplication probes are shown in Figure 6. Duplicates affect the calibrated construction in two ways: an anchor present $m > 1$ times raises its local contribution from α_i to $1 - (1 - \alpha_i)^m > \alpha_i$ away from the anchor itself, which the probe reproduces exactly, and repeated records drive nearest-neighbor distances toward zero, which shifts the median distance $\tilde{d}$ and thereby every calibrated kernel width. On the annulus, one additional copy of 10 % of the anchors already lowers the IoU with the duplicate-free region to 0.971; in the de-duplication probe, duplicates lower it to 0.685 for the calibrated construction, compared with 0.820 for fixed parameters, while resolution-cell de-duplication restores it to 0.936.

The calibrated full-data aviation evaluations used all 622 110 VCAS anchors and all 385 925 HCAS anchors, with 10^6 evaluation samples for each use case. At the smallest positive evaluated affinity threshold, precision and recall against the known ODD were 1.000 and 0.633 for VCAS and 0.995 and 0.655 for HCAS, compared with prevalences of 0.789 and 0.738 in the respective evaluation samples. Because the recorded operational states do not fill the known ODDs, recall below one reflects regions of the known ODD that no anchor point represents rather than a reconstruction error of $\mathcal{O}$. The full-data aviation runs additionally compare the autoSAFE membership decisions with seven reference regions fitted to the same anchors, including the core support from density-based spatial clustering of applications with noise (DBSCAN) [69]. Consequently, the IoU values in Table 4 quantify agreement between representations and are not baseline-versus-ground-truth accuracy measurements. For the production-scale data, the global envelope identified as `hull_single` by the evaluation artifact is a radial approximation based on the mean and maximum anchor radius, rather than an exact convex hull.

For the VCAS subset experiment, two independently seeded sets of 5000 anchors were selected from the complete anchor pool and evaluated on independently sampled sets of 200 000 points. The known VCAS ODD is the only reference in this experiment that does not itself change with the selected



(a) IoU with the known annulus and linear ODDs as the number of anchors increases.



(b) IoU with the duplicate-free reference regions as the candidate resolution-cell width increases.

Figure 6: Effect of sampling density and resolution-cell de-duplication. Increasing the density of distinct anchors raises the IoU with the known region from approximately 0.75–0.78 to 0.91; in the duplicate probe, collapsing records within shared resolution cells restores the IoU with the duplicate-free region from 0.685 to 0.936. These results motivate the deterministic ingestion rule in Algorithm 1: record multiplicity is retained as coverage and provenance metadata but does not instantiate additional kernels.

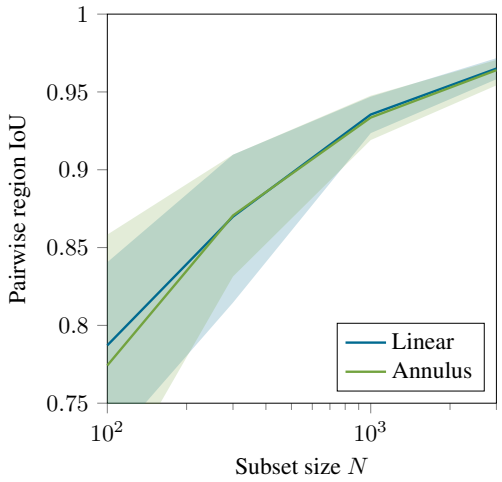
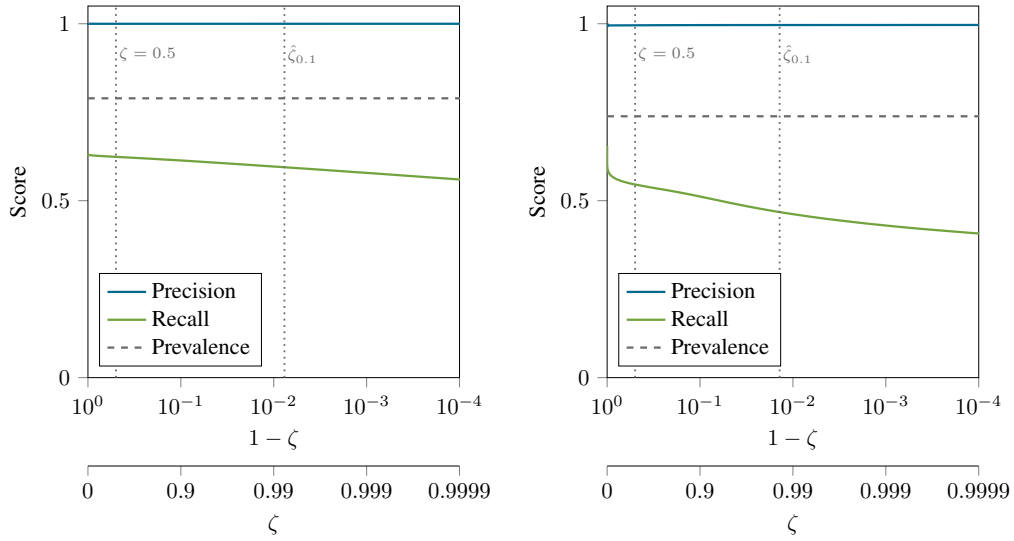


Figure 7: Mean pairwise region IoU across ten independently selected anchor subsets as the subset size increases; the shaded regions span the minimum and maximum of the 45 pairwise comparisons at each size. The mean rises from 0.77–0.79 at $N = 100$ to approximately 0.96 at $N = 3000$, where every pair exceeds 0.95. The narrowing range shows that the represented region becomes less sensitive to which anchors are selected as the subset grows.

anchor subset. As shown in Table 5, the precision and IoU of both subsets remain within 0.002 of the full-data evaluation despite the approximately 124-fold reduction in anchor count.



(a) Precision and recall of the VCAS ODD derived from all 622 110 anchor points relative to the affinity threshold. Precision stays above 0.999 at every threshold, while recall decreases from 0.633 to 0.558 at $1 - \zeta = 10^{-4}$.

(b) Precision and recall of the HCAS ODD derived from all 385 925 anchor points relative to the affinity threshold. Precision stays above 0.99 at every threshold, while recall decreases from 0.655 to 0.406 at $1 - \zeta = 10^{-4}$.

Figure 8: Precision and recall of the data-driven ODD $\mathcal{O}$ against the known ODDs of VCAS and HCAS [62] as the affinity threshold ζ approaches one, with $1 - \zeta$ on a logarithmic scale above the corresponding ζ. The dotted lines mark $\zeta = 0.5$ and the split-conformal threshold $\hat{\zeta}_{0.1}$, and the dashed line marks the share of evaluation samples inside the known ODD. Recall below one reflects regions of the known ODD that the recorded states never visit.

Table 4: IoU between autoSAFE and anchor-derived reference regions at the smallest positive evaluated affinity threshold, $\zeta = 1 \cdot 10^{-16}$, using 10^6 evaluation samples per use case.

Reference region	VCAS	HCAS
Global radial envelope	0.522	0.550
Clustered hull reference	0.534	0.657
k-NN support	0.737	0.539
k-means boundaries	0.684	0.762
KDE superlevel set	0.250	0.174
Clustered KDE superlevel sets	0.250	0.184
DBSCAN core support	0.786	0.573

Table 5: Agreement with the known VCAS ODD at the smallest positive evaluated affinity threshold. The subset runs use independently drawn anchor and evaluation sets; the values therefore measure end-to-end stability rather than paired pointwise agreement.

Run	Anchors	Evaluation samples	Prevalence	Precision	Recall	IoU
VCAS, subset seed 1	5000	200 000	0.789	0.998	0.633	0.632
VCAS, subset seed 2	5000	200 000	0.789	0.998	0.634	0.633
VCAS, complete dataset	622 110	1 000 000	0.789	1.000	0.633	0.633

A.8 Permutation Stability

Across 20 permutations of the ID and OOD sample orderings, the derived covariance matrices were identical, while the evaluated affinities differed by at most $1.3{\cdot}10^{-15}$; the residual stems solely from the finite-precision summation order of the batched evaluation (cf. Appendix F.1).

B High-Dimensional ODD Representation

To model the ODD structure $\mathcal{O}$, especially in higher dimensions, a robust framework is required. To model the parameter space X, polytopes have proven effective [34]. In general, given the complexity of the real world, the ODD polytope is not convex. While polytopes can be concave, modeling the ODD polytope as a union of convex polytopes is helpful, enabling the use of convex optimization tools on each component polytope. Thus, the ODD polytope $\mathcal{P}$ can be defined as the union of a set of J unique convex non-intersecting polytopes $\mathcal{C}_j$ as

$$\mathcal{P} = \bigcup_{j=1}^{J} \mathcal{C}_j, \tag{12}$$

with the half-space definition of $\mathcal{C}$ using the componentwise inequality

$$\mathcal{C}_j = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}_j \boldsymbol{x} \leq \boldsymbol{b}_j\}, \tag{13}$$

where $\boldsymbol{A}_j$ is the matrix containing the normal vectors to the j-th hyperplane defining the half-space and $\boldsymbol{b}_j$ is the right-hand side vector that determines the distance of the hyperplane from the origin.

C Proof of Termination

Proposition C.1 (Termination of Global OOD Consistency Adjustment)**.** *The global OOD consistency adjustment procedure in Algorithm 1 terminates in a finite number of steps, ensuring $\alpha(\boldsymbol{x}) \leq \xi$ for all $\boldsymbol{x} \in \mathcal{D}_{\text{OOD}}$.*

Proof. If $\mathcal{D}_{\text{OOD}} = \emptyset$, the claim is immediate. Otherwise, let $\boldsymbol{\Sigma}_i^{(0)}$ be the initial covariance of kernel i, and define

$$q_i^{\min} := \min_{\boldsymbol{z} \in \mathcal{D}_{\text{OOD}}} (\boldsymbol{z} - \boldsymbol{x}_i)^\top \left(\boldsymbol{\Sigma}_i^{(0)}\right)^{-1} (\boldsymbol{z} - \boldsymbol{x}_i) > 0. \tag{14}$$

The minimum is positive because the OOD set is finite, the covariance is positive-definite, and $\mathcal{A} \cap \mathcal{D}_{\text{OOD}} = \emptyset$, which follows from $\mathcal{A} \subseteq \mathcal{D}_{\text{ID}}$ and $\mathcal{D}_{\text{ID}} \cap \mathcal{D}_{\text{OOD}} = \emptyset$.

Let t_i denote the number of times kernel i has been scaled. Then $\boldsymbol{\Sigma}_i^{(t_i)} = c^{t_i} \boldsymbol{\Sigma}_i^{(0)}$, and for every $\boldsymbol{z} \in \mathcal{D}_{\text{OOD}}$,

$$\alpha_i^{(t_i)}(\boldsymbol{z}) = \exp\left(-\frac{c^{-t_i}}{2} (\boldsymbol{z} - \boldsymbol{x}_i)^\top \left(\boldsymbol{\Sigma}_i^{(0)}\right)^{-1} (\boldsymbol{z} - \boldsymbol{x}_i)\right) \tag{15}$$

$$\leq \exp\left(-\frac{q_i^{\min} c^{-t_i}}{2}\right). \tag{16}$$

Set $\xi_N := 1 - (1 - \xi)^{1/N}$, and define the finite per-kernel bound

$$K_i := \max\left\{0, \left\lceil \frac{\ln\left(\frac{-2\ln(\xi_N)}{q_i^{\min}}\right)}{\ln\left(\frac{1}{c}\right)} \right\rceil\right\}. \tag{17}$$

After K_i selections of kernel i, its local affinity is at most ξ_N at every OOD point.

It remains to show that the algorithm cannot select kernel i more than K_i times. Whenever the loop has not terminated, its selected OOD point $\boldsymbol{z}^*$ satisfies $\alpha(\boldsymbol{z}^*) > \xi$. If all local affinities at $\boldsymbol{z}^*$ were at most ξ_N, then

$$\alpha(\boldsymbol{z}^*) = 1 - \prod_{i=1}^{N} (1 - \alpha_i(\boldsymbol{z}^*)) \leq 1 - (1 - \xi_N)^N = \xi, \tag{18}$$

a contradiction. Thus some local affinity at $\boldsymbol{z}^*$ exceeds ξ_N, and the dominant kernel selected by the algorithm also exceeds ξ_N. A kernel that has already been selected K_i times cannot satisfy this condition and is never selected again while a violation remains. Consequently, each kernel is selected at most K_i times, and the total number of adjustments satisfies

$$N_{\mathrm{adj}} \leq \sum_{i=1}^{N} K_i < \infty. \tag{19}$$

This argument is unaffected by ties: any maximizer of either selection step satisfies the same bound. The loop therefore terminates after finitely many adjustments, and its stopping condition gives $\alpha(\boldsymbol{z}) \leq \xi$ for every $\boldsymbol{z} \in \mathcal{D}_{\mathrm{OOD}}$. □

D Asymptotic Similarity of the Kernel-Based ODD

This section distinguishes the behavior of fixed kernel parameters from that of the calibrated parameters in Equation (10). For a set $E \subseteq \mathbb{R}^n$, $d(\boldsymbol{x}, E) := \inf_{\boldsymbol{y} \in E} \|\boldsymbol{x} - \boldsymbol{y}\|_2$ denotes the Euclidean point-to-set distance, $B(\boldsymbol{x}, r)$ the open Euclidean ball, ∂E the boundary, and $\mathrm{int}(E)$ the interior.

D.1 Proof of Scale Invariance

This subsection proves Proposition 4.7.

Equivariance condition. Let $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with $\boldsymbol{A} = \mathrm{diag}(a_1, \ldots, a_n)$, $a_k > 0$, and let α' denote the affinity built on the transformed anchor points $\{T(\boldsymbol{x}_i)\}$ with covariance matrices $\boldsymbol{\Sigma}_i'$. The Mahalanobis form satisfies

$$\begin{aligned} &(T(\boldsymbol{x}) - T(\boldsymbol{x}_i))^\top \left(\boldsymbol{A}\boldsymbol{\Sigma}_i\boldsymbol{A}^\top\right)^{-1} (T(\boldsymbol{x}) - T(\boldsymbol{x}_i)) \\ &= (\boldsymbol{x} - \boldsymbol{x}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x} - \boldsymbol{x}_i). \end{aligned} \tag{20}$$

Consequently, $\alpha'(T(\boldsymbol{x})) = \alpha(\boldsymbol{x})$ for all $\boldsymbol{x}$ if $\boldsymbol{\Sigma}_i' = \boldsymbol{A}\boldsymbol{\Sigma}_i\boldsymbol{A}^\top$, i.e., if $\sigma_{kk}'^{(i)} = a_k^2 \sigma_{kk}^{(i)}$ for all i and k.

Similarity transformations. Let $a_1 = \cdots = a_n = a > 0$. Both nearest-neighbor modes are preserved under T, and the distances satisfy $d_i^{*\prime} = a d_i^*$, $d_{ik}^{*\prime} = a d_{ik}^*$, and $\tilde{d}' = a\tilde{d}$. The calibrated parameters of Equation (10) consequently satisfy $\eta' = \eta/a$, $\kappa' = a^2\kappa$, and $\lambda' = a^2\lambda$. It follows that

$$\sigma_{kk}'^{(i)} = \left(a^2\kappa - a^2\lambda\right) \exp\left(-\frac{\eta}{a} a d_{ik}^*\right) + a^2\lambda = a^2 \sigma_{kk}^{(i)}, \tag{21}$$

which is precisely the equivariance condition above. The relative specification $\lambda = \lambda_{\mathrm{rel}}\kappa$ is essential because an absolute lower bound would not transform as a squared length.

Per-dimension affine re-expressions. For general $\boldsymbol{A} = \mathrm{diag}(a_1, \ldots, a_n)$, the Euclidean nearest-neighbor assignment may change, so the preceding argument does not apply directly. Assume that the complete derivation pipeline first normalizes every parameter by a per-dimension affine transform whose defining bounds re-express equivariantly with the data. Re-expressing the raw data by T re-expresses the normalization bounds identically, so the normalized anchor set and every quantity derived from it remain unchanged. The normalization therefore makes the complete pipeline invariant under arbitrary per-dimension affine re-expressions of the raw data.

D.2 Fixed-Parameter Behavior

Proposition D.1 (Fixed-parameter behavior)**.** *Let the reference membership region $R^* \subset \mathbb{R}^n$ be compact with positive finite volume and a boundary of Lebesgue measure zero. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be independent and uniformly distributed on R^*, let α_N be the global affinity constructed from the first N anchors through Equation* (9)*, and, for $N \geq 2$, let $R_N := \{\boldsymbol{x} : \alpha_N(\boldsymbol{x}) \geq \zeta\}$, where $\zeta \in (0, 1)$, $\kappa > 0$, $\eta \geq 0$, and $0 < \lambda \leq \kappa$ are held fixed as N grows. Then, almost surely, Lebesgue-almost every $\boldsymbol{x} \in \mathrm{int}(R^*)$ belongs to R_N for all sufficiently large N. Moreover, for every $N \geq 2$,*

$$R_N \subseteq (R^*)^{+\rho_N}, \qquad \rho_N := \sqrt{2\kappa \ln\left(\frac{N}{\zeta}\right)}, \tag{22}$$

where $(R^*)^{+r} := \{\boldsymbol{x} : d(\boldsymbol{x}, R^*) \leq r\}$ denotes the closed outward dilation of R^* by radius r.

Here, "almost surely" means with probability one over the random sequence of anchor points; it does not express uncertainty about which conclusion holds once that sequence is fixed.

Proof. Fix $\boldsymbol{x} \in \operatorname{int}(R^*)$ and choose $\varepsilon > 0$ such that $B(\boldsymbol{x}, \varepsilon) \subset R^*$ and $\exp(-\varepsilon^2/(2\lambda)) \geq \zeta$. With probability one, an infinite i.i.d. sample from R^* hits this positive-volume ball; from the first such hit onward there is an anchor $\boldsymbol{x}_i$ with $\|\boldsymbol{x} - \boldsymbol{x}_i\|_2 \leq \varepsilon$. Since every diagonal entry of $\boldsymbol{\Sigma}_i$ is at least λ,

$$\alpha_i(\boldsymbol{x}) \geq \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|_2^2}{2\lambda}\right) \geq \zeta, \tag{23}$$

and hence $\alpha_N(\boldsymbol{x}) \geq \alpha_i(\boldsymbol{x}) \geq \zeta$ for every subsequent N. Applying Fubini's theorem [70] to the sampling event and using the null boundary gives the first claim simultaneously for Lebesgue-almost every interior point.

For the outer bound, fix $\boldsymbol{x}$ at distance $\delta = d(\boldsymbol{x}, R^*)$ from R^*. Every anchor is at least δ away, and every covariance eigenvalue is at most κ, so

$$\alpha_i(\boldsymbol{x}) \leq \exp\left(-\frac{\delta^2}{2\kappa}\right). \tag{24}$$

The union bound for the noisy-OR aggregation gives

$$\alpha_N(\boldsymbol{x}) \leq \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}) \leq N \exp\left(-\frac{\delta^2}{2\kappa}\right). \tag{25}$$

If $\delta > \rho_N$, the right-hand side is smaller than ζ; therefore $\boldsymbol{x} \notin R_N$, which proves Equation (22). □

Remark D.2 (Two fixed-parameter regimes). The first part establishes eventual interior coverage for Lebesgue-almost every point, consistent with the sparse-sample under-approximation observed empirically. The outer radius in Equation (22), however, grows like $\sqrt{\ln N}$, so it does not establish two-sided set convergence at fixed κ and λ. Dense noisy-OR superposition can instead expand the boundary unless bandwidths shrink with N. The OOD adjustment guarantees exclusion only at the finite OOD samples to which it is applied; it does not change this global fixed-parameter conclusion.

D.3 Calibrated Kernel Parameters

For the following calibrated result, let $R^* \subset \mathbb{R}^n$ be compact with positive finite volume V and a Lipschitz boundary, and let v_n denote the volume of the n-dimensional unit ball. The Lipschitz boundary implies an interior volume condition: there exist $r_0 > 0$ and $0 < b \leq 1$ such that $\mathcal{L}^n(B(\boldsymbol{x}, r) \cap R^*) \geq b \cdot v_n \cdot r^n$ for all $\boldsymbol{x} \in R^*$ and $r \leq r_0$. For the first N anchors, let $\tilde{d}_N$ denote the median full-space nearest-neighbor distance. Under the calibrated parameters of Equation (10), $\kappa_N = (s_N \tilde{d}_N)^2$ and $\lambda_N = \lambda_{\text{rel}} \kappa_N$ shrink as N grows, while $\eta_N = \gamma / \tilde{d}_N$ grows. The following lemmas establish the rates needed for two-sided convergence in measure.

Lemma D.3 (Scaling of the median nearest-neighbor distance)**.** *There exist constants $0 < c_1 \leq c_2 < \infty$, depending only on n and R^*, such that almost surely $c_1 N^{-1/n} \leq \tilde{d}_N \leq c_2 N^{-1/n}$ for all sufficiently large N.*

Proof. Let $F_N(t)$ denote the fraction of anchor points whose nearest-neighbor distance is at most $tN^{-1/n}$. Conditional on $\boldsymbol{x}_i = \boldsymbol{x}$, the remaining $N - 1$ points avoid $B(\boldsymbol{x}, tN^{-1/n})$ with probability $\left(1 - \mathcal{L}^n(B(\boldsymbol{x}, tN^{-1/n}) \cap R^*)/V\right)^{N-1}$. The interior volume condition and the unit-ball upper bound therefore give

$$1 - \exp\left(-\frac{b \cdot v_n \cdot t^n}{V}\right) + o(1) \leq \mathbb{E}F_N(t) \leq 1 - \exp\left(-\frac{v_n \cdot t^n}{V}\right) + o(1). \tag{26}$$

Thus, t_1 can be chosen small enough that $\limsup_N \mathbb{E}F_N(t_1) \leq 0.25$, and t_2 can be chosen large enough that $\liminf_N \mathbb{E}F_N(t_2) \geq 0.75$. In fixed dimension, a point can be the nearest neighbor of at most τ_n other points, where τ_n depends only on n [71]. Replacing one anchor can therefore change $F_N(t)$ by at most $(2\tau_n + 1)/N$. McDiarmid's bounded-differences inequality [72] gives

$$\mathbb{P}\left(|F_N(t) - \mathbb{E}F_N(t)| \geq \frac{1}{8}\right) \leq 2 \exp\left(-\frac{N}{32(2\tau_n + 1)^2}\right), \tag{27}$$

which is summable in N. The Borel–Cantelli lemma [73] implies that almost surely $F_N(t_1) < 0.5 < F_N(t_2)$ for all sufficiently large N. Consequently, $t_1 N^{-1/n} \le \tilde{d}_N \le t_2 N^{-1/n}$; related nearest-neighbor laws are given in prior research [74]. □

Lemma D.4 (Local rates at a fixed interior point)**.** *Fix $\boldsymbol{x} \in \operatorname{int}(R^*)$, let $\boldsymbol{x}_{(1),N}$ and $\boldsymbol{x}_{(2),N}$ be its nearest and second-nearest anchors, and let $r_{(2),N}(\boldsymbol{x}) := \|\boldsymbol{x} - \boldsymbol{x}_{(2),N}\|_2$. There exists $C > 0$ such that almost surely $r_{(2),N}(\boldsymbol{x}) \le (C \ln \ln N/N)^{1/n}$ for all sufficiently large N.*

Proof. The distance $r_{(2),N}(\boldsymbol{x})$ is non-increasing in N. Let $N_k = 2^k$ and $r_k = (a \ln k/N_k)^{1/n}$, with $a > 0$ to be chosen. For sufficiently large k, $B(\boldsymbol{x}, r_k) \subseteq R^*$, so $q_k := \mathcal{L}^n(B(\boldsymbol{x}, r_k))/V = v_n \cdot a \ln k/(V \cdot N_k)$. The probability that this ball contains at most one of the N_k anchors satisfies

$$(1-q_k)^{N_k} + N_k \cdot q_k (1-q_k)^{N_k - 1} \le (1 + N_k \cdot q_k) \exp(-(N_k - 1)q_k) \le C_a (\ln k) k^{-v_n \cdot a/V}. \tag{28}$$

This bound is summable in k when $v_n \cdot a/V \ge 3$. The Borel–Cantelli lemma [73] gives $r_{(2),N_k}(\boldsymbol{x}) \le r_k$ almost surely for all sufficiently large k. For $N_k \le N < N_{k+1}$, monotonicity gives $r_{(2),N}(\boldsymbol{x}) \le (2a \ln k/N)^{1/n}$, and $\ln k \le \ln \ln N + 1$ proves the claim. □

For sets E and F, $E \triangle F := (E \setminus F) \cup (F \setminus E)$ denotes their symmetric difference.

The asymptotic results below concern the retained anchor sequence under continuous sampling. Under their assumptions, exact duplicate observations occur with probability zero, so exact-record de-duplication is almost surely the identity. Resolution-cell merging for quantized acquisition is a finite-sample preprocessing operation and is not covered by the asymptotic statement.

Theorem D.5 (Asymptotic ODD Similarity, calibrated parameters)**.** *Let $n \ge 2$, let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be independent and uniformly distributed on R^*, and calibrate the parameters as in Equation* (10) *with $s_N = \sqrt{\ln N}$, $\gamma > 0$, and $\lambda_{\mathrm{rel}} \in (0,1)$. For every $\zeta \in (0,1)$ and $N \ge 2$, let α_N be the global affinity constructed from the first N anchors and let $R_N := \{\boldsymbol{x} : \alpha_N(\boldsymbol{x}) \ge \zeta\}$ be its unadjusted membership region. Then, almost surely, Lebesgue-almost every $\boldsymbol{x} \in \operatorname{int}(R^*)$ belongs to R_N for all sufficiently large N, every fixed $\boldsymbol{x}$ with $d(\boldsymbol{x}, R^*) > 0$ lies outside R_N for all sufficiently large N, and*

$$\mathcal{L}^n(R_N \triangle R^*) \to 0. \tag{29}$$

Thus, for a shared interpretation function and noise space, the represented ODD approaches the similarity criterion of Definition 3.2 in the symmetric-difference measure.

Proof. Interior points. Fix $\boldsymbol{x} \in \operatorname{int}(R^*)$. By Lemma D.4, almost surely $\|\boldsymbol{x} - \boldsymbol{x}_{(1),N}\|_2 \le r_{(2),N}(\boldsymbol{x}) \le (C \ln \ln N/N)^{1/n}$ eventually. The nearest-neighbor distance of $\boldsymbol{x}_{(1),N}$ satisfies $d^*_{(1),N} \le \|\boldsymbol{x}_{(1),N} - \boldsymbol{x}_{(2),N}\|_2 \le 2(C \ln \ln N/N)^{1/n}$. For either nearest-neighbor mode, the coordinatewise distance used for each diagonal entry is at most $d^*_{(1),N}$. By Lemma D.3, $\eta_N \cdot d^*_{(1),N} \le t_N := (2\gamma C^{1/n}/c_1)(\ln \ln N)^{1/n}$. Every diagonal entry of $\boldsymbol{\Sigma}_{(1),N}$ therefore satisfies

$$\sigma^{(1,N)}_{kk} \ge (1 - \lambda_{\mathrm{rel}}) c_1^2 s_N^2 N^{-2/n} \exp(-t_N). \tag{30}$$

It follows that

$$\alpha_{(1),N}(\boldsymbol{x}) \ge \exp\left(-\frac{C^{2/n} (\ln \ln N)^{2/n} \exp(t_N)}{2(1-\lambda_{\mathrm{rel}}) c_1^2 \ln N}\right) \to 1. \tag{31}$$

Indeed, the logarithm of the exponent magnitude is $(2/n) \ln \ln \ln N + t_N - \ln \ln N + O(1)$, which tends to $-\infty$ because $n \ge 2$. Hence $\alpha_N(\boldsymbol{x}) \ge \alpha_{(1),N}(\boldsymbol{x}) \ge \zeta$ eventually.

Exterior points. Let $\delta = d(\boldsymbol{x}, R^*) > 0$. Every kernel satisfies $\sigma^{(i)}_{kk} \le \kappa_N$, so the noisy-OR union bound and Lemma D.3 give

$$\alpha_N(\boldsymbol{x}) \le N \exp\left(-\frac{\delta^2}{2\kappa_N}\right) \le N \exp\left(-\frac{\delta^2 N^{2/n}}{2c_2^2 \ln N}\right) \to 0. \tag{32}$$

Moreover, the represented region is globally contained in a shrinking dilation because

$$\rho_N^{\mathrm{cal}} := \sqrt{2\kappa_N \ln\left(\frac{N}{\zeta}\right)} = s_N \tilde{d}_N \sqrt{2\ln\left(\frac{N}{\zeta}\right)} \to 0. \tag{33}$$

Convergence in measure. The preceding steps give pointwise indicator convergence for every fixed point outside ∂R^*, almost surely. Fubini's theorem [70] makes this conclusion simultaneous for Lebesgue-almost every point, and the boundary is Lebesgue-null by the Lipschitz assumption. The shrinking outer containment supplies a common compact superset, so the dominated convergence theorem [70] yields $\mathcal{L}^n(R_N \triangle R^*) \to 0$. □

E Justification of Diagonal Sigma-Matrices

E.1 Theoretical Justification

The deployed covariance matrices are diagonal but anisotropic: each coordinate receives its own local scale according to Equation (9). This choice represents axis-aligned changes in local sampling density while retaining linear storage and evaluation cost in the input dimension. It does not, however, represent rotated anisotropy or correlations between coordinates.

A diagonal approximation to a full local covariance is statistically justified only when the off-diagonal population correlations are negligible, i.e., when the coordinate components of nearby anchor points are approximately uncorrelated in the fixed working coordinate system. Local uniformity by itself is insufficient: a locally elongated distribution may have persistent cross-coordinate correlation, even with infinitely many samples. Consequently, the diagonal form is an explicit accuracy–efficiency trade-off whose discrepancy from a corresponding full-covariance structure is quantified below. This trade-off concerns finite-sample accuracy, not consistency: Theorem D.5 assumes a Lipschitz boundary but no independence between coordinates, so the calibrated diagonal construction also converges in volume for coupled parameters.

E.2 Approximation Error Bound

Let $\mathbf{\Sigma}_i^{\mathrm{full}}$ denote a symmetric positive-definite full covariance and let $\mathbf{\Sigma}_i^{\mathrm{diag}} = \mathrm{diag}(\mathbf{\Sigma}_i^{\mathrm{full}})$ be its diagonal projection. This comparison isolates covariance structure; $\mathbf{\Sigma}_i^{\mathrm{full}}$ is not an estimator used by the deployed method. For $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_i$, define

$$q_i^{\mathrm{full}}(\boldsymbol{x}) = \boldsymbol{v}^\top \left(\mathbf{\Sigma}_i^{\mathrm{full}}\right)^{-1} \boldsymbol{v}, \qquad q_i^{\mathrm{diag}}(\boldsymbol{x}) = \boldsymbol{v}^\top \left(\mathbf{\Sigma}_i^{\mathrm{diag}}\right)^{-1} \boldsymbol{v}. \tag{34}$$

The resolvent identity gives the exact relation

$$\left(\mathbf{\Sigma}_i^{\mathrm{full}}\right)^{-1} - \left(\mathbf{\Sigma}_i^{\mathrm{diag}}\right)^{-1} = \left(\mathbf{\Sigma}_i^{\mathrm{full}}\right)^{-1} \left(\mathbf{\Sigma}_i^{\mathrm{diag}} - \mathbf{\Sigma}_i^{\mathrm{full}}\right) \left(\mathbf{\Sigma}_i^{\mathrm{diag}}\right)^{-1}. \tag{35}$$

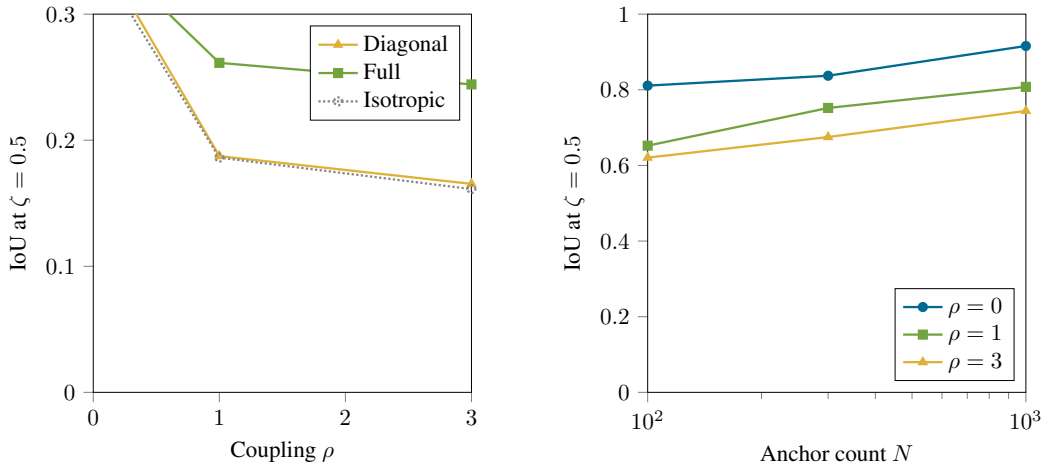
The mean value theorem applied to $t \mapsto \exp(-t/2)$, followed by the spectral norm bound for a quadratic form, therefore yields

$$\begin{aligned} \left|\alpha_i^{\mathrm{full}}(\boldsymbol{x}) - \alpha_i^{\mathrm{diag}}(\boldsymbol{x})\right| \le \frac{1}{2} \left\|\left(\mathbf{\Sigma}_i^{\mathrm{full}}\right)^{-1}\right\|_2 \left\|\mathbf{\Sigma}_i^{\mathrm{diag}} - \mathbf{\Sigma}_i^{\mathrm{full}}\right\|_2 \\ \left\|\left(\mathbf{\Sigma}_i^{\mathrm{diag}}\right)^{-1}\right\|_2 \|\boldsymbol{v}\|_2^2 \max\left\{\alpha_i^{\mathrm{full}}(\boldsymbol{x}), \alpha_i^{\mathrm{diag}}(\boldsymbol{x})\right\}. \end{aligned} \tag{36}$$

Thus, the local discrepancy is controlled by the omitted off-diagonal structure and by the conditioning of both covariance matrices. It vanishes for an axis-aligned covariance, but it need not vanish merely because more locally correlated samples are observed. For the global noisy-OR affinity, a telescoping product bound gives

$$\left|\alpha^{\mathrm{full}}(\boldsymbol{x}) - \alpha^{\mathrm{diag}}(\boldsymbol{x})\right| \le \sum_{i=1}^{N} \left|\alpha_i^{\mathrm{full}}(\boldsymbol{x}) - \alpha_i^{\mathrm{diag}}(\boldsymbol{x})\right|, \tag{37}$$

so Equation (36) can be summed over the kernels that contribute materially at $\boldsymbol{x}$. The covariance probe in Figure 9 uses the rotated-band region $\{(x_1, x_2) \in [-5, 5]^2 : |x_2 - \rho x_1| \le 0.6\}$, where ρ controls the rotational coupling.



(a) Controlled covariance-structure comparison at $N = 100$. The case $\rho = 0$ is axis-aligned, whereas $\rho = 1$ and $\rho = 3$ rotate the narrow ODD relative to the working coordinates.

(b) IoU of the deployed diagonal construction as the anchor count increases. All three curves improve with additional anchors, reaching 0.916, 0.808, and 0.744 at $N = 1000$.

Figure 9: Quantitative covariance-structure probe under increasing rotational coupling. At $N = 100$, full and diagonal covariance produce nearly identical IoU for the axis-aligned case, while full covariance raises the IoU from 0.187 to 0.261 at $\rho = 1$ and from 0.165 to 0.244 at $\rho = 3$. The measured gap quantifies the expected loss when the local-independence assumption is violated, while the increasing curves of the deployed construction show that greater anchor density still improves the represented region at every tested coupling strength.

F Implementation Details

F.1 Numerical Evaluation of the Affinity

The product in Equation (2) is exact in real arithmetic, but direct finite-precision evaluation can absorb small local contributions, underflow the survival product, and round the final subtraction to an affinity of exactly one. The precise breakpoints depend on the chosen arithmetic; the failure is therefore not specific to one floating-point format. For the full VCAS run with $N = 622\,110$ anchors (cf. Section 5.2), the elementary bound $\alpha(\boldsymbol{x}) \geq 1 - \exp(-\sum_i \alpha_i(\boldsymbol{x}))$ leads to exponentially small survival probabilities where the accumulated local affinity is large; in direct finite-precision evaluation, $\alpha(\boldsymbol{x}) = 1$ for 26.3 % of the evaluation samples, causing ordinary affinity thresholds to lose distinctions within that saturated subset.

Therefore, the affinity is evaluated in log-space. Defining the log-survival function

$$S(\boldsymbol{x}) = \ln\left(1 - \alpha(\boldsymbol{x})\right) = \sum_i \ln\left(1 - \alpha_i(\boldsymbol{x})\right) \tag{38}$$

turns the product into a sum of same-signed terms. Each term is computed with a stable branchwise scheme [75], and the accumulated rounding error is bounded by standard summation results [76]. Since $t \mapsto \ln(1 - t)$ is strictly decreasing, the membership rule transforms exactly:

$$\alpha(\boldsymbol{x}) \geq \zeta \iff S(\boldsymbol{x}) \leq \ln(1 - \zeta). \tag{39}$$

All reported results use Equation (39).

F.2 Software Interface

The Python-based tool provides a sophisticated interface for various applications, including Monte Carlo–based sampling for validation (cf. Appendix A.2) and the provision of real-world data in different formats (cf. Section 5.2). The tool can parse data from a wide range of sources, including CSV files, NumPy-like arrays, dataframes, and, most importantly, the JSON-based ASAM OpenLABEL format [77]. The latter is an annotation format commonly used in the automotive domain, designed to

facilitate the labeling and tracking of objects across multiple scenes. At ingestion, the ID records that define the ODD geometry are first de-duplicated using documented acquisition-resolution metadata while preserving merged source identifiers and visit counts. The retained records are then converted into an internal kernel-based representation that enables efficient querying of the affinity value for one or more sample points. Thus, it can be integrated into current development methodologies and used as an online tool to determine whether new, unseen samples fall within the ODD, thereby enhancing the safety of already deployed AI-based systems. This is achieved using a set of libraries, such as NumPy [78] and SciPy [79] for efficient data handling, polytope [80] for representing ODD polytopes (cf. Section 3), and Faiss [81, 82] for fast nearest-neighbor searches.

G The autoSAFE Algorithm

Algorithm 1 Automated Kernel-Based ODD Derivation for RBF Kernels

Input:
  ID record collection $\widetilde{\mathcal{D}}_{\mathrm{ID}}$ and OOD sample set $\mathcal{D}_{\mathrm{OOD}}$
  Documented acquisition-grid metadata $(\boldsymbol{q}, \boldsymbol{o})$
  Calibration constants γ and λ_{rel}, scheduling rule $s_N = \sqrt{\ln N}$
  OOD affinity threshold ξ
  Covariance shrink factor $c \in (0, 1)$
  False-exclusion budget $\varepsilon \in (0, 1)$ and calibration-set size $N_{\mathrm{cal}} \geq 1/\varepsilon - 1$
  Optional anchor count N
  Documented random seed for the calibration split and the optional anchor subsampling
Calibration Split:
Reserve N_{cal} held-out records $\mathcal{D}_{\mathrm{cal}}$, selected uniformly at random from $\widetilde{\mathcal{D}}_{\mathrm{ID}}$, and remove them from $\widetilde{\mathcal{D}}_{\mathrm{ID}}$
Deterministic ID Anchor De-duplication:
Partition $\widetilde{\mathcal{D}}_{\mathrm{ID}}$ into acquisition-resolution cells defined by $(\boldsymbol{q}, \boldsymbol{o})$
for each occupied cell C **do**
  Sort the records in C lexicographically and compute their mean $\bar{\boldsymbol{x}}_C$ in that order
  Retain the lexicographically first minimizer of $\sum_{k=1}^{n}((x_k - \bar{x}_{Ck})/q_k)^2$ over $\boldsymbol{x} \in C$
  Store the merged source identifiers and record count as provenance metadata
end for
Index the retained ID representatives and OOD samples in lexicographic order
Set $\mathcal{D}_{\mathrm{ID}}$ to the canonically indexed ID representatives
Set anchor points $\mathcal{A} \leftarrow \mathcal{D}_{\mathrm{ID}}$, or, if N is given, select N of them uniformly at random without replacement, retaining their canonical order
Kernel Parameter Estimation:
Compute the full-space nearest-neighbor distances and their median $\tilde{d}$
Set $\eta \leftarrow \gamma/\tilde{d}$, $\kappa \leftarrow (s_N\tilde{d})^2$, and $\lambda \leftarrow \lambda_{\mathrm{rel}}\kappa$
for each anchor point $\boldsymbol{x}_i \in \mathcal{A}$ **do**
  Compute the selected global or per-dimension exact nearest-neighbor distances d^*_{ik}, $k = 1, \ldots, n$ {lowest canonical anchor index resolves global-mode ties}
  for each dimension k **do**
    Set $\sigma^{(i)}_{kk} \leftarrow (\kappa - \lambda)\exp(-\eta d^*_{ik}) + \lambda$
  end for
  Define kernel $\alpha_i(\boldsymbol{x})$ using $\boldsymbol{\Sigma}_i$
end for
OOD Consistency Check (cf. Definition 4.5): {requires $\mathcal{A} \cap \mathcal{D}_{\mathrm{OOD}} = \emptyset$}
while $\exists \boldsymbol{x}_{\mathrm{OOD}} \in \mathcal{D}_{\mathrm{OOD}}$ s.t. $\alpha(\boldsymbol{x}_{\mathrm{OOD}}) > \xi$ **do**
  $\boldsymbol{x}^* \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{D}_{\mathrm{OOD}}} \alpha(\boldsymbol{x})$ {globally most-violated OOD point}
  $i^* \leftarrow \arg\max_i \alpha_i(\boldsymbol{x}^*)$ {dominant kernel}
  Adjust $\boldsymbol{\Sigma}_{i^*} \leftarrow c \cdot \boldsymbol{\Sigma}_{i^*}, \quad c \in (0, 1)$
end while
Membership Threshold Calibration (cf. Appendix A.6):
Compute $Q(\boldsymbol{x}) = -\ln(1 - \alpha(\boldsymbol{x}))$ for all $\boldsymbol{x} \in \mathcal{D}_{\mathrm{cal}}$ and sort them in ascending order
Set $k \leftarrow \lfloor \varepsilon(N_{\mathrm{cal}} + 1) \rfloor$, $\hat{t} \leftarrow Q_{(k)}$, and $\hat{\zeta} \leftarrow 1 - \exp(-\hat{t})$
Verify $\xi < \hat{\zeta}$
Output: Deterministic kernel-based ODD affinity function $\alpha(\boldsymbol{x})$ and membership threshold $\hat{\zeta}$

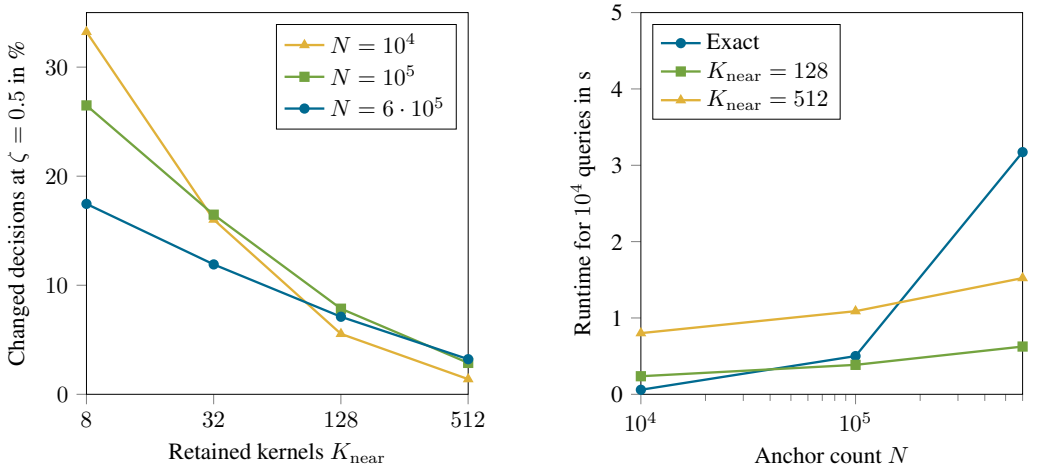
H Computational Scalability

The pseudocode in Algorithm 1 has total construction cost $O(N^2 n + N_{\mathrm{adj}} MNn + N_{\mathrm{cal}} Nn)$ and model-storage cost $O(Nn)$. Here, N is the number of anchors, n is the parameter-space dimension, M is the number of OOD samples, and N_{adj} is the finite number of covariance-adjustment steps bounded in Proposition C.1. The term $N^2 n$ arises because the direct nearest-neighbor construction

compares every pair of anchors in n dimensions. Each OOD adjustment evaluates all N kernels at all M OOD samples, which accounts for MNn operations per step and $N_{\text{adj}}MNn$ in total. Calibrating $\hat{\zeta}$ evaluates all N kernels at the N_{cal} held-out records, which accounts for $N_{\text{cal}}Nn$ operations. After construction, an exact affinity query evaluates all local affinities and the noisy-OR aggregation in $O(Nn)$ time.

An exact nearest-neighbor index with $O(n \log N)$ query cost in fixed dimension can replace the pairwise search, reducing the first term to $O(Nn \log N)$ under the index's stated assumptions without changing the mathematical construction. A separate approximation retains only the K_{near} kernels whose anchors are nearest in normalized Euclidean distance during affinity evaluation, yielding $O(n \log N + K_{\text{near}}n)$ time per query when paired with such an index. Because the RBF contributions are positive, truncation always underestimates the exact noisy-OR affinity, but the omitted tail and the resulting membership decisions depend on K_{near}, the anchor density, and the kernel widths. The quantitative experiment in Figure 10 therefore reports affinity error, decision changes, runtime, and host-side working-set memory rather than treating truncation as exact.

For $K_{\text{near}} = 512$, the maximum absolute affinity error increases from 0.018 at $N = 1000$ to 0.434 at $N = 6 \cdot 10^5$; at the largest anchor count, the mean and 99th-percentile errors are 0.0303 and 0.257. For $K_{\text{near}} = 128$ at $N = 6 \cdot 10^5$, the corresponding maximum, mean, and 99th-percentile errors are 0.696, 0.0687, and 0.503. At this anchor count, the anchor and inverse-diagonal model arrays shared by both evaluators occupy 48 MB, while the Python-traced temporary working set of the truncated implementation peaks at 184 MB for $K_{\text{near}} = 128$ and 737 MB for $K_{\text{near}} = 512$. These working-set measurements exclude JAX/XLA backend allocations used by the exact evaluator and therefore do not establish a measured memory advantage over exact evaluation. Truncation is therefore studied only as an optional acceleration; the implementation and all other reported results use exact evaluation.



(a) Fraction of changed membership decisions at $\zeta = 0.5$ as the number of retained kernels increases for three anchor counts.

(b) Wall-clock runtime for 10 000 queries under exact evaluation and two truncation levels as the number of anchors increases.

Figure 10: Accuracy and runtime of K_{near}-kernel truncation on the five-dimensional constrained ODD with 10 000 queries. At $N = 6 \cdot 10^5$, $K_{\text{near}} = 512$ reduces runtime from 3.17 s to 1.52 s with 3.22 % changed decisions at $\zeta = 0.5$, whereas $K_{\text{near}} = 128$ takes 0.63 s with 7.11 % changed decisions. The measured crossover is approximately $2.3 \cdot 10^5$ anchors for $K_{\text{near}} = 512$; below it, the fixed costs of nearest-neighbor lookup and selected-kernel gathering plausibly outweigh the kernel evaluations avoided by truncation.

I Description of Aviation Use Cases

In aviation, one of the most important tasks of pilots during en route travel is avoiding collisions with other aircraft. The next generation of collision avoidance systems is expected to improve

performance significantly; however, it has become apparent that the proposed system cannot run on current state-of-the-art avionics hardware because its memory requirements are too high [58, 83]. Preliminary minimal implementations split the system into a horizontal and a vertical component, called the Horizontal Collision Avoidance System (HCAS) and the Vertical Collision Avoidance System (VCAS), respectively, using neural networks for policy compression [84, 62, 2]. Thus, a stringent safety argumentation will be required for an AI-based collision avoidance system [85]. Given that the standards clearly define the operational conditions, an appropriate ODD can be derived and later compared with the ODD derived by autoSAFE. Therefore, this use case is a prime example for further study in this work. However, in line with previous research [62], only a subset of the full operational conditions will be examined.

I.1 Horizontal Collision Avoidance System

For HCAS, the decision depends on the range and bearing angle to the intruder, the relative heading angle, the ownship's and intruder's speed, the time to CPA, and the advisory issued in the previous time step. A geometric overview of this problem is given in Figure 11. Furthermore, the parameter

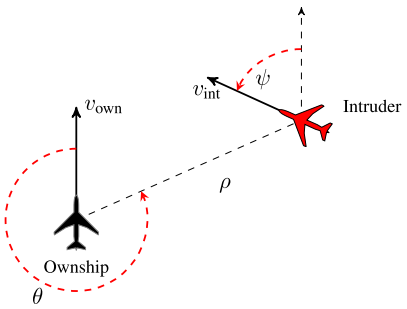


Figure 11: Geometry of the horizontal collision avoidance scenario for HCAS, from [62]. The black ownship is trying to avoid the (malicious) red intruder by diverting either to the left or to the right.

ranges are shown in Table 6.

Table 6: Parameter ranges defining the taxonomy X^* in the reported HCAS evaluation. The ranges are a subset of the original HCAS implementation [62].

Variable	Description	Range
ρ	Range to intruder	$[0\,\text{ft}, 50\,000\,\text{ft}]$
θ	Bearing angle	$[-180°, 180°]$
ψ	Relative heading	$[-180°, 180°]$
v_{own}	Ownship speed	$[840\,\text{ft}\,\text{s}^{-1}, 870\,\text{ft}\,\text{s}^{-1}]$
v_{int}	Intruder speed	$[760\,\text{ft}\,\text{s}^{-1}, 870\,\text{ft}\,\text{s}^{-1}]$
τ	Time to CPA	$[0\,\text{s}, 80\,\text{s}]$
s_{adv}	Previous advisory	$\{0, 1, 2, 3, 4\}$

I.1.1 Known HCAS ODD Structure

Following Definition 3.1, the known HCAS ODD is defined as the structure

$$\mathcal{O}^*_{\text{HCAS}} = \left(X^*, f^{\mathcal{O}^*}, \Omega^{\mathcal{O}^*}\right), \tag{40}$$

where the taxonomy $X^* \subset \mathbb{R}^7$ is the hyperrectangle

$$\begin{aligned} X^* = &[0\,\text{ft}, 50\,000\,\text{ft}] \times [-180°, 180°] \times [-180°, 180°] \\ &\times [840\,\text{ft}\,\text{s}^{-1}, 870\,\text{ft}\,\text{s}^{-1}] \times [760\,\text{ft}\,\text{s}^{-1}, 870\,\text{ft}\,\text{s}^{-1}] \\ &\times [0\,\text{s}, 80\,\text{s}] \times \{0, 1, 2, 3, 4\}, \end{aligned} \tag{41}$$

spanned by $(\rho, \theta, \psi, v_{\text{own}}, v_{\text{int}}, \tau, s_{\text{adv}})$ as specified in Table 6 and following prior research [62]. No additional relational constraints beyond the parameter ranges are imposed for this use case, i.e., $r = 0$ and $\mathcal{R}^{\mathcal{O}^*} = X^*$. The interpretation function $f^{\mathcal{O}^*} : X^* \to Y$ is the deterministic HCAS policy from [62], mapping each state vector $\boldsymbol{x} \in X^*$ to an advisory $y \in Y = \{0, 1, 2, 3, 4\}$. Since the policy is deterministic, the noise space is trivial; thus $\Omega^{\mathcal{O}^*} = \{0\}$.

I.1.2 Special Variables

The previous advisory $s_{\text{adv}} \in \{0, 1, 2, 3, 4\}$ is a discrete categorical variable. For the kernel computation, its codes are retained as a scalar coordinate with unit spacing; the kernel metric therefore treats numerical code separation as geometric distance. The bearing angle θ and the relative heading ψ are periodic; the validation samples and the known-ODD membership respect this periodicity, whereas the kernel metric treats both as ordinary coordinates, so that states on either side of $\pm 180^\circ$ count as distant. All seven parameters are retained, giving a seven-dimensional effective parameter space.

I.2 Vertical Collision Avoidance System

For VCAS, the decision depends on the relative altitude between the ownship and the intruder, their corresponding vertical rates, the time to CPA, and the advisory issued in the previous time step. A geometric overview of this problem is given in Figure 12. Furthermore, the parameter ranges are

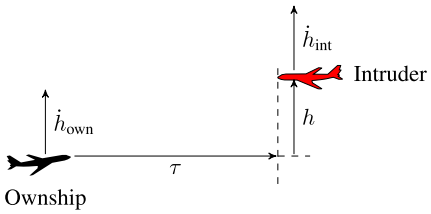


Figure 12: Geometry of the vertical collision avoidance scenario for VCAS, from [62]. The black ownship is trying to avoid the (malicious) red intruder by diverting through climbing or descending.

shown in Table 7.

Table 7: Parameter ranges of the nominal five-dimensional state space in the reported VCAS evaluation. The ranges are a subset of the original VCAS implementation [62].

Variable	Description	Range
h	Relative altitude	$[-3000\,\text{ft}, 3000\,\text{ft}]$
$\dot{h}_{\text{own}}$	Ownship vert. rate	$[-3000\,\text{ft min}^{-1}, 3000\,\text{ft min}^{-1}]$
$\dot{h}_{\text{int}}$	Intruder vert. rate	$0\,\text{ft min}^{-1}$
τ	Time to CPA	$[0\,\text{s}, 40\,\text{s}]$
s_{adv}	Previous advisory	$\{0, 1, 2, \ldots, 8\}$

I.2.1 Known VCAS ODD Structure

Following Definition 3.1, the known VCAS ODD is defined as the structure

$$\mathcal{O}^*_{\text{VCAS}} = \left(X^*, f^{\mathcal{O}^*}, \Omega^{\mathcal{O}^*}\right), \tag{42}$$

where the taxonomy $X^* \subset \mathbb{R}^4$ is the product space

$$X^* = [-3000\,\text{ft}, 3000\,\text{ft}] \times [-3000\,\text{ft min}^{-1}, 3000\,\text{ft min}^{-1}] \times [0\,\text{s}, 40\,\text{s}] \times \{0, 1, 2, \ldots, 8\}, \tag{43}$$

spanned by $(h, \dot{h}_{\text{own}}, \tau, s_{\text{adv}})$ as specified in Table 7 and following prior research [62]. No additional relational constraints beyond the parameter ranges are imposed for this use case, i.e., $r = 0$ and $\mathcal{R}^{\mathcal{O}^*} = X^*$. The interpretation function $f^{\mathcal{O}^*} : X^* \to Y$ is the deterministic VCAS policy from [62], mapping each state vector $\boldsymbol{x} \in X^*$ to an advisory $y \in Y = \{0, 1, \ldots, 8\}$. Since the policy is deterministic, the noise space is trivial; thus $\Omega^{\mathcal{O}^*} = \{0\}$.

I.2.2 Special Variables

Two parameters require specific treatment. First, the previous advisory $s_{\mathrm{adv}} \in \{0, 1, \ldots, 8\}$ is a discrete categorical variable; it is encoded as a continuous coordinate with unit spacing between adjacent values, which is appropriate because adjacent advisory values reflect operationally similar prior commands. Second, the intruder vertical rate $\dot{h}_{\mathrm{int}}$ takes the single value $0\,\mathrm{ft\,min}^{-1}$ across all states in the dataset, yielding zero variance in that dimension. A zero-variance dimension produces a degenerate covariance matrix entry and contributes no discriminative information to the kernel; it is therefore excluded from the kernel computation and from X^* above. This reduces the effective parameter space from the nominal 5D to 4D.

J EASA Objectives Addressed by autoSAFE

The operational domain (OD) describes the aircraft- or system-level operating conditions, while the ODD applies at the AI/ML constituent level. The quoted objectives below abbreviate data quality requirements as DQRs. Table 8 relates properties of autoSAFE to selected objectives of the EASA concept paper, Issue 02 [23]; the corresponding identifiers in the proposed Issue 03 [86] are shown in parentheses. The mapping identifies potential contributions to an assurance case, not compliance with an objective, and each contribution remains use-case specific.

Table 8: Selected objectives from the EASA concept paper to which autoSAFE can contribute. Issue 02 identifiers are followed by their counterparts in proposed Issue 03 [23, 86].

Objective	**Excerpt**	**Rationale**
DA-03 (SURK-DA-03)	"The applicant should define the set of parameters pertaining to the AI/ML constituent ODD, and trace them to the corresponding parameters pertaining to the OD when applicable."	The framework represents the ODD taxonomy and its interdependencies through an explicit membership function derived from documented operational records. Traceability from anchors to the OD depends on the provenance maintained for those records by the surrounding data-management process.
DA-04 (SU-DA-04)	"The applicant should capture the DQRs for all data required for training, testing, and verification of the AI/ML constituent [. . .]"	The affinity provides a per-sample measure of support relative to the constructed ODD, and low-affinity regions expose gaps in the recorded coverage. This supplies quantitative evidence toward completeness and representativeness requirements, but it does not establish every required aspect of data quality.
DA-07 (SURK-DA-07)	"The applicant should validate each of the requirements captured under Objectives DA-02, DA-03, DA-04, DA-05 and the architecture captured under Objective DA-06."	The similarity notion of Definition 3.2 and the held-out evaluation in Section 5 quantify discrepancies between an independently specified ODD and the data-driven representation. This evidence can focus subject-matter-expert review, but it does not replace that review.
DM-07 (SU-DM-06)	"The applicant should ensure verification of the data, as appropriate, throughout the data management process so that the data management requirements (including the DQRs) are addressed."	Under the canonical ordering and tie rules, the construction is permutation-stable in exact arithmetic, making the ODD-derivation transformation reproducible. Unambiguous traceability to source and intermediate data remains the responsibility of the surrounding data-management process.

Continued on next page

Table 8: Selected objectives from the EASA concept paper to which autoSAFE can contribute. Issue 02 identifiers are followed by their counterparts in proposed Issue 03 [23, 86]. (Continued)

Objective	Excerpt	Rationale
DM-08 (SURK-DA-12, SU-DA-13)	"The applicant should perform a data verification step to confirm the appropriateness of the defined ODD and of the data sets used for the training, validation and verification of the ML model."	**Optional, conditional support.** When independently specified ODD membership information is available, precision and recall quantify agreement with that reference, and held-out training, validation, or verification samples can be evaluated against the constructed support. Without independent reference membership, autoSAFE does not by itself establish the appropriateness of the final ODD or every aspect of dataset appropriateness.
LM-07-SL (S-LM-04; SU-LM-05)	"The applicant should account for the bias-variance trade-off in the model family selection and should provide evidence of the reproducibility of the model training process."	The RBF kernel family is fixed and justified in Section 4, and calibration reduces its remaining constants to γ and λ_{rel} (cf. Section 4.3). The ODD derivation has no random initialization or stochastic optimization and is reproducible for identical documented inputs; this supports the construction itself rather than the separate training of an AI/ML model.
LM-11 (SU-LM-05)	"The applicant should provide an analysis on the stability of the learning algorithms."	The derivation contains no stochastic model-training loop, so instability from random initialization or stochastic optimization does not apply. Permutation stability establishes repeatability for identical inputs, while the parameter and subset experiments characterize selected input changes; neither result proves stability under arbitrary perturbations, label errors, or distribution shift.
LM-12 (SU-LM-12)	"The applicant should perform and document the verification of the stability of the trained model, covering the whole AI/ML constituent ODD."	The explicit ODD representation and affinity can guide the selection of nominal, boundary, and low-coverage verification cases. autoSAFE does not evaluate the trained model's outputs under perturbations and therefore supplies only partial support for this objective.

Continued on next page

Table 8: Selected objectives from the EASA concept paper to which autoSAFE can contribute. Issue 02 identifiers are followed by their counterparts in proposed Issue 03 [23, 86]. (Continued)

Objective	**Excerpt**	**Rationale**
SA-02 (CRA-02)	"The applicant should identify which data needs to be recorded for the purpose of supporting the continuous safety assessment."	Operational input vectors and their affinities are candidate records for later monitoring of ODD assumptions. The framework does not define data-integrity, retention, or retrieval mechanisms and does not establish distribution-shift detection.
SA-03 (CRA-03)	"In preparation of the continuous safety assessment, the applicant should define metrics, target values, thresholds and evaluation periods to guarantee that design assumptions hold."	The bounded affinity and membership threshold provide an ODD-membership metric and decision boundary that can contribute to monitoring. Safety target values, evaluation periods, and any relation between affinity and system-level safety-margin erosion require separate assurance decisions.
EXP-19 (restructured in Issue 03 to: EXP-04, SA-05/RA-10)	"Information concerning unsafe AI-based system operating conditions should be provided to the end user to enable them to take appropriate corrective action in a timely manner."	Affinity thresholds can supply an input to a separately designed warning or fallback function as operational support decreases. The affinity does not determine whether a condition is intrinsically unsafe, implement end-user communication, or guarantee timely corrective action.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the contributions—a deterministic, kernel-based, data-driven ODD framework—and the scope is bounded to the specific conditions under which the method is valid.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 6.1 discusses limitations regarding kernel parameter sensitivity and conservative under-approximation.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Proposition C.1 and the results in Appendices D and E are stated with full assumptions.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The experimental settings, ODD definitions, sample counts, and calibration choices are specified in Section 5 and Appendix A; the source code, data, and experiment configurations accompany the public release.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The public repository linked in Section 5 provides the source code, data, documentation, and experiment configurations used by this work.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The relevant calibration constants, thresholds, sample counts, and data splits are stated in Section 5 and Appendix A, and the exact configurations are included in the open-source release.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Sampling variability is reported where relevant: Figure 4 shows one standard deviation across five random seeds, Figure 7 reports the range across subset comparisons, and Appendix A.6 reports ranges across 20 repetitions.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The compute resources available for the experiments, together with an upper bound of the execution time, are given in Section 5 and Appendix H.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work involves no human subjects and develops a safety-assurance method with no foreseeable ethical concerns; Section 6.1 explicitly cautions that data coverage is not system safety and confines the method to one evidence artifact within a broader safety case.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The potential benefit is more reproducible evidence about operational coverage for safety-critical AI systems, while the principal risk is misplaced confidence when unsafe or valid conditions are absent from the data; both are discussed in Sections 6 and 6.1.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: No high risk models are developed as part of this work.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The major third-party libraries are cited in Appendix F, and their licenses are recorded in the accompanying repository. The HCAS and VCAS parameterization is attributed to prior research [62].

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The autoSAFE software, data formats, and experiment workflows are documented in the public repository linked in Section 5.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: No crowdsourcing experiments or research with human subjects have been conducted as part of this work.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No research with human subjects has been conducted as part of this work.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, a declaration is not required.

Answer: [N/A]

Justification: