When Does Learned Inertial Odometry Fail?

Anonymous authors
Paper under double-blind review

Abstract

Learned inertial odometry estimates velocity directly from IMU sequences and reduces the drift caused by inertial integration, but its cross-domain generalization remains limited. Existing studies mainly focus on improving target-domain accuracy, while aggregate metrics provide little insight into why failures occur. We analyze target-domain failures by relating them to source-domain motion coverage, source prediction error, and model internal states. Our results show that source motion coverage alone does not determine prediction failure. Some failures occur under motion conditions that are rare in the source domain, while others arise under source-covered conditions near source low-error motion regions. For the latter, failed predictions tend to remain biased toward the source-domain velocity associated with similar motion. This indicates that motion–velocity relations learned from source data may become invalid after domain shift. We further show that different failure types favor different target-domain adaptation data, and that motion and internal-state information provide complementary cues for failure detection. These results link the data regularities that enable direct velocity prediction to the cross-domain failures of learned inertial odometry.

1 Introduction

Learned inertial odometry exploits motion patterns in data to reduce the drift caused by long-term integration of low-cost IMU measurements, but its performance remains sensitive to unseen motion and sensor conditions (Zhao et al., 2025b;a). RIDI uses learned human-motion velocity to assist inertial estimation and correct acceleration errors (Yan et al., 2018). IONet estimates motion states from IMU windows instead of continuous integration (Chen et al., 2018). RoNIN further directly regresses velocity from IMU sequences (Herath et al., 2020).

Notably, these models can predict velocity without an explicitly given initial velocity. Their predictions therefore cannot be explained as simply learning conventional inertial integration. It remains unclear what additional information they exploit from the training data, and how the resulting data dependence relates to their cross-domain generalization.

Relationships learned from the training data may no longer hold when test conditions change, leading to cross-domain degradation. Existing studies typically evaluate generalization using aggregate trajectory errors on target domains (Cao et al., 2022; Zhao et al., 2025a). Such metrics reveal whether performance degrades, but provide limited insight into which learned relationships remain valid in the target domain. Learned inertial odometry is fundamentally a regression problem, and prior work on cross-domain regression shows that prediction risk depends on more specific source–target distribution relations rather than on an overall distribution gap alone (Pathak et al., 2022; Eyre et al., 2024). Understanding cross-domain failure therefore requires examining where performance degradation occurs in the input space and whether the data dependencies learned in the source domain still hold under those conditions.

We start from failed target-domain samples and progressively trace the possible sources of cross-domain failure. We first identify target samples that are insufficiently covered by the source domain and test whether input novelty alone can explain prediction failure. For failed samples that remain within the source coverage, we then examine model performance under similar motion conditions in the source domain. Large source-domain errors indicate that the difficulty already exists during training, whereas target failures near source low-error motion regions motivate examining differences between source and target velocity conditions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

For the latter case, we further compare the label relations and prediction biases between source and target samples under similar input conditions, and examine how these failures are reflected in the model's internal features.

- **Source motion coverage alone does not determine target-domain failure.** Insufficient source coverage does not necessarily lead to failure, while many failures also occur under motion conditions already covered by the source domain.
- **Target predictions can retain a source-domain velocity tendency.** When similar motion corresponds to different velocities in the target domain, failed predictions often remain biased toward the source-domain velocity response.
- **Different failure types benefit from different target-domain adaptation data.** Adaptation data selected for a specific failure type improve that type most effectively, with limited transfer to other failure types.

2 Characterizing Cross-Domain Failure

2.1 Motion and Failure Description

Prediction error can vary across input conditions, so aggregate performance alone may not fully characterize model failure. Let $X = \{(\boldsymbol{\omega}_t, \mathbf{a}_t)\}_{t=1}^{N}$ denote an IMU window containing N samples, where $\boldsymbol{\omega}_t = [\omega_{x,t}, \omega_{y,t}, \omega_{z,t}]^\top$ and $\mathbf{a}_t = [a_{x,t}, a_{y,t}, a_{z,t}]^\top$ are the gyroscope and accelerometer measurements at time step t. We represent each window by five motion functionals,

$$\boldsymbol{\Phi}(X) = [\phi_1, \phi_2, \phi_3, \phi_4, \phi_5]^\top \in \mathbb{R}^5, \tag{1}$$

which capture motion intensity, temporal variation, and rotational characteristics:

$$\begin{aligned}
\phi_1 &= \frac{1}{N}\sum_{t=1}^{N} \|\boldsymbol{\omega}_t\|_2^2, & \phi_2 &= \frac{1}{N}\sum_{t=1}^{N} \|\mathbf{a}_t - \bar{\mathbf{a}}\|_2^2, \\
\phi_3 &= \frac{1}{\Delta t}\sqrt{\frac{1}{N-1}\sum_{t=2}^{N} \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}\|_2^2}, & \phi_4 &= \frac{1}{\Delta t}\sqrt{\frac{1}{N-1}\sum_{t=2}^{N} \|\mathbf{a}_t - \mathbf{a}_{t-1}\|_2^2}, \\
\phi_5 &= \frac{\sum_{t=1}^{N} \omega_{z,t}^2}{\sum_{t=1}^{N} \|\boldsymbol{\omega}_t\|_2^2 + \epsilon}, & \bar{\mathbf{a}} &= \frac{1}{N}\sum_{t=1}^{N} \mathbf{a}_t.
\end{aligned} \tag{2}$$

Here, $\Delta t = 0.01\,\mathrm{s}$ is the sampling interval. The features ϕ_1–ϕ_5 respectively describe rotational intensity, centered acceleration variation, temporal changes in angular velocity and acceleration, and the axial composition of rotational motion.

For the i-th IMU window, let $\hat{\mathbf{v}}_i$ and $\mathbf{v}_i$ denote the predicted and reference velocity vectors, respectively. We define the prediction error as the relative velocity error

$$e_i = \frac{\|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_2}{\max\left(\|\mathbf{v}_i\|_2, v_{\min}\right)}, \tag{3}$$

where $v_{\min} = 1\,\mathrm{m/s}$ for all carriers and evaluations. Prediction failure is defined with respect to the error distribution of the corresponding source-domain model. Specifically,

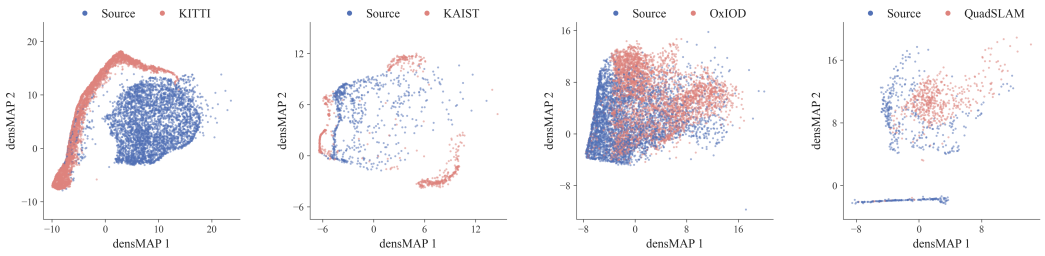
$$F_i = \mathbb{I}(e_i > \tau_S), \qquad \tau_S = Q_{0.90}\left(e^{\text{SourceVal}}\right), \tag{4}$$

where $Q_{0.90}(\cdot)$ denotes the 90th percentile of the relative prediction errors on the source validation set. Thus, $F_i = 1$ indicates that the prediction error exceeds the range observed for 90% of the source-validation windows. Both $v_{\min}$ and τ_S are determined without using target-domain samples.

Each window X_i is thus represented by the triplet $(\boldsymbol{\Phi}(X_i), e_i, F_i)$, yielding

$$\mathcal{D} = \{(\boldsymbol{\Phi}(X_i), e_i, F_i)\}_{i=1}^{M}, \tag{5}$$

where M is the number of windows in the dataset. We estimate the relationship between motion and prediction error on source-domain data and evaluate it on held-out source trajectories. We then use this relationship to distinguish shifts in the target motion distribution from changes in prediction error under comparable motion.



(a) Car (b) Drone (c) Human (d) Dog

Figure 1: Source and target motion distributions across four carriers.

Table 1: LiLY failure rates by source coverage.

		Failure rate (%)	
Carrier	Target	Source-sparse	Source-covered
Car	KITTI	74.51	84.87
Drone	KAIST	1.14	1.19
Human	OxIOD	97.73	93.78
Dog	QuadSLAM	73.94	35.27

2.2 Motion-Space Structure and Domain Shift

We first examine shifts in the motion distribution from the source domain to four public target datasets: KITTI, KAIST, OxIOD, and QuadSLAM. For each source–target pair, we jointly project the five-dimensional motion representations into two dimensions using densMAP. Figure 1 shows the resulting distributions.

Figure 1 shows that target samples partly overlap with Source while also extending into lower-density regions. We quantify this in the original five-dimensional motion space using a Gaussian KDE fitted on SourceTrain. Defining

$$D(\mathbf{\Phi}) = -\log \hat{p}_S(\mathbf{\Phi}), \tag{6}$$

we set the coverage threshold as the trajectory-equal SourceTrain $Q_{0.95}$,

$$\tau_{\text{cov}} = Q_{0.95}(D_S). \tag{7}$$

Target samples with $D(\mathbf{\Phi}) \leq \tau_{\text{cov}}$ are treated as source-covered and the remainder as source-sparse.

3 Understanding Prediction Failures under Motion Shift

We examine how far source motion coverage explains target-domain prediction failure. For source-sparse samples, we analyze failure rates across speed ranges to understand when limited coverage is associated with large errors and when predictions remain accurate. For source-covered failures, we identify those occurring in source high-error regions and separate the remaining cases as residual failures. We then examine whether the model's internal states help characterize these residual failures.

Figure 2 shows the source and OxIOD motion distributions, with LiLY prediction errors in source-sparse and source-covered regions enlarged in panels (b) and (c), respectively.

Table 2: LiLY errors by speed in source-sparse regions.

Target	Speed (m/s)	N	Failure (%)	Median EPE (m/s)
KITTI	$[0, 0.1)$	0	—	—
	$[0.1, 1)$	0	—	—
	$[1, \infty)$	153	74.51	4.147
KAIST	$[0, 0.1)$	65	0.00	0.077
	$[0.1, 1)$	110	1.82	0.297
	$[1, \infty)$	0	—	—
OxIOD	$[0, 0.1)$	86	9.30	0.069
	$[0.1, 1)$	6196	98.66	0.862
	$[1, \infty)$	1059	99.43	1.166
QuadSLAM	$[0, 0.1)$	28	14.29	0.098
	$[0.1, 1)$	601	76.54	0.650
	$[1, \infty)$	315	74.29	0.643

N: windows per ground-truth speed interval; —: empty interval. Failure follows Sec. 2.1.

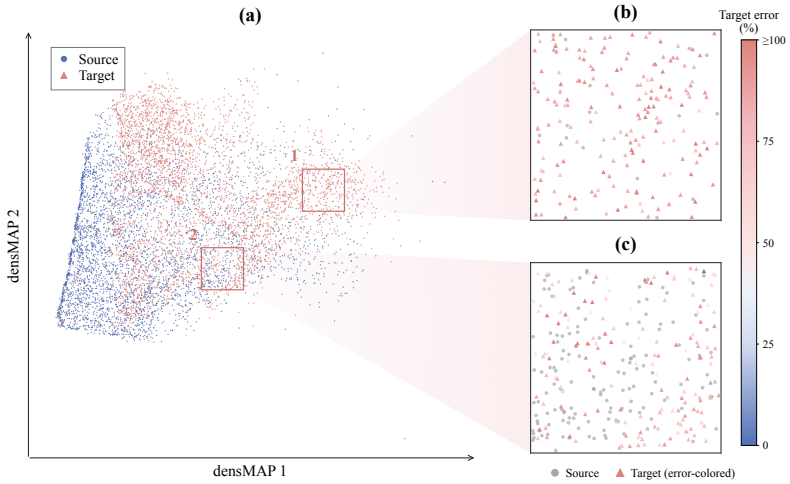


Figure 2: Motion coverage and prediction error on OxIOD.

3.1 Non-failure under source-sparse motion

Failure is prevalent among source-sparse windows in KITTI, OxIOD, and QuadSLAM, reaching 97.73% on OxIOD, but only 1.14% on KAIST (Table 1). This contrast shows that limited source coverage alone does not determine prediction failure.

We examine speed as a possible explanation for KAIST's lower failure rate. Its source-sparse windows have a median ground-truth speed of $0.162\,\mathrm{m/s}$. Among the 175 windows, 65 have speeds below $0.1\,\mathrm{m/s}$, and none are classified as failures. These windows account for 37.57% of the 173 source-sparse non-failures.

To examine whether KAIST's lower failure rate persists within comparable speed ranges, we group source-sparse windows into $[0, 0.1)$, $[0.1, 1)$, and $[1, \infty)\,\mathrm{m/s}$. We report failure rates alongside the median endpoint error (EPE), where $\mathrm{EPE}_i = \|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_2$ measures the absolute velocity error of each window in $\mathrm{m/s}$.

Table 2 shows that source-sparse windows below $0.1\,\mathrm{m/s}$ generally have low failure rates and median EPE below $0.1\,\mathrm{m/s}$. Such windows are more common in KAIST, accounting for 37.14% of its source-sparse samples.

Table 3: Relative prediction errors in selected high-error motion regions.

			Motion region					SourceTrain			SourceVal		
Source	Model	Region	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	N	Median (%)	Mode (%)	N	Median (%)	Mode (%)
Car	LiLY	$\mathcal{R}_{242}$	H	H	H	H	M	482	18.99	13.15	133	25.06	16.37
	Tartan	$\mathcal{R}_{133}$	M	M	H	H	L	168	68.50	68.97	104	62.71	64.71
Drone	LiLY	$\mathcal{R}_5$	L	L	L	M	M	283	34.11	24.75	110	25.47	19.65
	Tartan	$\mathcal{R}_5$	L	L	L	M	M	283	58.75	54.73	110	34.31	32.38
Human	LiLY	$\mathcal{R}_{242}$	H	H	H	H	M	1377	28.49	24.33	219	30.01	19.05
	Tartan	$\mathcal{R}_{243}$	H	H	H	H	H	1348	21.63	15.92	124	19.61	14.59
Dog	LiLY	$\mathcal{R}_{203}$	H	M	M	M	M	213	44.00	52.08	111	30.96	27.76
	Tartan	$\mathcal{R}_{203}$	H	M	M	M	M	213	35.80	29.90	111	30.80	16.63

L/M/H: low/medium/high SourceTrain-defined intervals; N: IMU windows.

KAIST's lower failure rate also persists within $[0.1, 1)\,\mathrm{m/s}$: 1.82% of windows fail, compared with 98.66% on OxIOD and 76.54% on QuadSLAM. Its median EPE is also lower. Thus, the difference extends beyond KAIST's larger proportion of near-stationary windows. Limited source coverage can coincide with small prediction errors, and its association with failure varies across target datasets.

3.2 Failures under Source-Covered Motion

We next examine prediction failures that occur under source-covered motion conditions. A possible explanation is that some target failures simply occur in motion conditions that are already difficult for the model in the source domain. We therefore characterize how prediction error varies across the joint source motion space.

Using SourceTrain only, each motion functional ϕ_j is divided into low, medium, and high intervals according to its empirical one-third and two-third quantiles. The Cartesian product of the five discretized functionals yields $3^5 = 243$ disjoint motion regions, denoted by $\{\mathcal{R}_k\}$. The resulting partition is fixed and applied unchanged to SourceVal and target-domain samples.

For each occupied region $\mathcal{R}_k$, we summarize the relative prediction-error distribution by its median,

$$e_k^{\text{med}} = \text{median}\,\{e_i \mid \mathbf{\Phi}(X_i) \in \mathcal{R}_k\}\,, \tag{8}$$

and by the mode of its estimated density,

$$e_k^{\text{mode}} = \arg\max_e \hat{p}_k(e), \tag{9}$$

where $\hat{p}_k(e)$ is the error density within $\mathcal{R}_k$, estimated by Gaussian KDE with Scott's bandwidth. The median summarizes the typical error, and the mode locates the density peak. The relative error e_i is defined in Sec. 2.1, with $v_{\min} = 1\,\mathrm{m/s}$ for all carriers.

After partitioning the source motion space, we select the region with the highest SourceTrain median error for each model–carrier setting among regions with at least 100 windows in each split and available mode estimates. Table 3 reports these regions and their SourceVal statistics. Their errors generally remain high on held-out source trajectories. Complete statistics for all occupied regions are provided in Appendix C.

For the subsequent analysis, let n_k^S and $e_{k,S}^{\text{med}}$ denote the number of SourceTrain inference windows and their median relative error in region $\mathcal{R}_k$. A region is labeled high-error when

$$n_k^S \geq 30, \qquad e_{k,S}^{\text{med}} > \text{median}\{e_i \mid X_i \in \text{SourceTrain}\}. \tag{10}$$

Regions with at least 30 windows and a median error at or below the same reference are labeled low-error; regions with fewer than 30 windows are left unconfirmed. These labels use SourceTrain only. SourceVal statistics assess transfer of regional error patterns to held-out trajectories.

We then assign each source-covered target failure to its source-defined motion region and measure the fraction occurring in source high-error regions. The corresponding proportions are reported in the high-error-region row of Table 4.

Table 4: Breakdown of source-covered target failures.

Category	KITTI	KAIST	OxIOD	QuadSLAM
Covered failures	**3,462**	**6**	**23,938**	**79**
High-error region	6 (0.17%)	1 (16.67%)	13,179 (55.05%)	77 (97.47%)
Motion and speed	0	0	301 (1.26%)	0
Motion only	2,966 (85.67%)	0	8,884 (37.11%)	0
High-error match	0	1 (16.67%)	196 (0.82%)	2 (2.53%)
Unconfirmed only	0	0	1 (< 0.01%)	0
No match	490 (14.15%)	4 (66.67%)	1,377 (5.75%)	0

Entries are counts (% of covered failures). *Motion and speed* and *Motion only* denote low-error-region neighbors with and without a speed match. Other matches use motion distance.

The breakdown in Table 4 shows that many covered failures occur in source high-error motion regions for some targets. For OxIOD, 55.05% of covered failures fall into regions that are already associated with large prediction errors in Source, while this proportion reaches 97.47% for QuadSLAM. These failures occur under motion conditions where the source model already makes large errors, consistent with the source-domain difficulty continuing on Target.

However, this explanation is incomplete. A large number of covered failures still occur near source samples from low-error motion regions. This is particularly evident for KITTI, where 85.67% are associated with low-error source neighbors under similar motion but different speed. OxIOD also retains a substantial residual subset, with 37.11% of covered failures showing the same pattern. We therefore next focus on these failures to determine why similar, source-covered motion conditions can still lead to markedly different prediction outcomes in the target domain.

For the remaining source-covered failures outside the source high-error regions, we examine whether a comparable SourceTrain sample from a low-error motion region exists. For each target failure X_i^T, we search the corresponding carrier's SourceTrain reference set for neighbors satisfying

$$d_\Phi(i,j) = \left\| \widetilde{\mathbf{\Phi}}(X_i^T) - \widetilde{\mathbf{\Phi}}(X_j^S) \right\|_2 \leq \delta_\Phi(i), \tag{11}$$

where $\widetilde{\mathbf{\Phi}}$ is standardized using SourceTrain statistics. Matching thresholds are calibrated on Source and fixed for Target; Appendix D.1 details the calibration. Neighbors are labeled by their SourceTrain region's error category. When a neighbor from a low-error region is available, we also compare ground-truth speeds:

$$d_v(i,j) = \left| \|\mathbf{v}_i^T\|_2 - \|\mathbf{v}_j^S\|_2 \right|. \tag{12}$$

A target failure is classified as motion-and-speed matched if a low-error-region neighbor satisfies both $d_\Phi(i,j) \leq \delta_\Phi(i)$ and $d_v(i,j) \leq \delta_v(i)$, where $\delta_v(i)$ is calibrated on Source using the corresponding global or speed-bin-specific reference. If low-error-region motion neighbors exist but none meets the speed criterion, the failure is classified as motion matched but speed mismatched. Low-error-region matches take priority over high-error-region matches. The remaining failures are grouped by whether they have high-error-region neighbors, only unconfirmed-region neighbors, or no motion neighbors meeting the threshold. Table 4 reports the category counts, and Fig. **??** shows their proportions.

For KITTI and OxIOD, 85.67% and 37.11% of covered failures, respectively, have neighbors from source low-error motion regions with similar motion but different ground-truth speeds. These cases show that similar motion representations can correspond to different speed conditions across domains.

We next examine whether predictions in motion-matched, speed-mismatched cases remain closer to the source velocity reference than to the target ground truth. For each covered failure, we construct a source reference from the matched neighbors and compare the prediction's Euclidean distance to the two references.

Velocity vectors are expressed in the body/device-frame coordinates used by the model labels. Let $\mathbf{v}_i^S$ denote the coordinate-wise median of the ground-truth velocity vectors of all matched SourceTrain

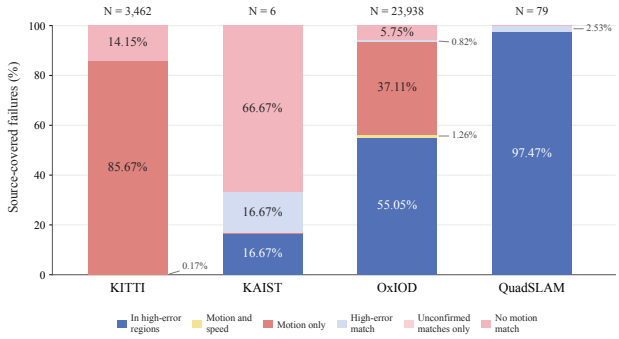


Figure 3: Composition of LiLY source-covered failures.

Table 5: LiLY prediction proximity to source and target velocities.

		Closer reference (%)	
Target	N	Source	Target
KITTI	2,966	100.00	0.00
OxIOD	8,884	95.15	4.85

N: motion-matched, speed-mismatched covered failures.

neighbors from low-error regions. We compute

$$d_i^S = \|\hat{\mathbf{v}}_i - \mathbf{v}_i^S\|_2, \qquad d_i^T = \|\hat{\mathbf{v}}_i - \mathbf{v}_i\|_2. \tag{13}$$

A prediction is classified as source-closer if $d_i^S < d_i^T$ and target-closer if $d_i^T < d_i^S$. No distance ties occurred. Table 5 reports the percentage of each case.

As shown in Table 5, all selected KITTI failures are closer to the source reference for LiLY. On OxIOD, the proportion is 95.15%. This pattern is consistent with predictions retaining a source-domain velocity tendency when similar motion corresponds to different target velocities. Its occurrence in both datasets supports examining the learned motion–velocity relation as a potential source of these failures. We next investigate whether the model's internal states reflect this discrepancy.

3.3 Velocity Information Retained under Prediction Failure

We next examine whether velocity information remains accessible in the learned representation across failure types. We hypothesize that some failures reflect a mismatch between the latent representation and the original output mapping, rather than a loss of velocity information.

We evaluate readout recovery on OxIOD for three categories defined by the original model: source-sparse failures, source-covered failures in high-error regions, and source-covered failures in low-error regions. The latter two categories use each target window's own source-defined region. We freeze the representation and fit an affine ridge velocity readout using 1,600 windows from two target calibration sessions. Five-fold evaluation holds out complete sessions, with the same test windows across calibration conditions. Recovery is defined as

$$\text{Recovery}_c = \frac{1}{|\mathcal{F}_c|} \sum_{i \in \mathcal{F}_c} \mathbb{I}\big(e_i^{\text{cal}} \le \tau_S\big)\,, \tag{14}$$

where $\mathcal{F}_c$ contains the original failures in category c, e_i^{cal} is the held-out error after readout fitting, and τ_S is the original SourceVal-Q90 threshold. We vary the failure composition of the calibration sessions to test category-specific recovery; Appendix E.1 gives the selection and fitting details.

Figure 4 shows that prediction failures remain substantially recoverable from the frozen latent representation.

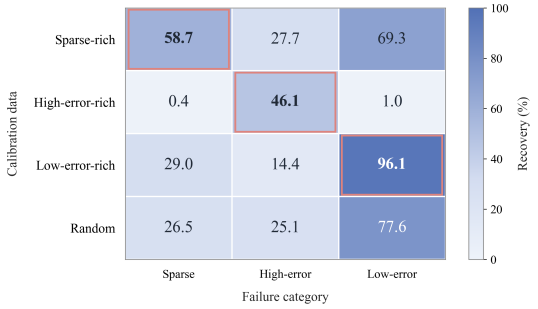


Figure 4: LiLY recovery rates by calibration data and failure category on OxIOD.

For LiLY on OxIOD, each failure category achieves its highest pooled recovery rate when calibration sessions are enriched with the same category. Recovery reaches 96.1% for covered failures in source low-error regions. Calibration on other categories yields lower recovery and sometimes worsens performance.

These results show that readout adaptation depends on the failure composition of the calibration data. The motion-based categories identify differences in recovery behavior and provide a basis for selecting calibration sessions suited to the failures being addressed.

4 Cross-Domain Failure Detection

4.1 Failure Detection Formulation

Following the failure definition in Sec. 2, we examine whether input motion and model internal states provide complementary information for target-domain failure detection. Labeled target sessions are used to train the detectors, which are evaluated on held-out sessions.

We consider two single-source detectors. The *Motion-only* detector uses the five-dimensional motion representation $\mathbf{\Phi}_i \in \mathbb{R}^5$ defined in Sec. 2. The *Internal-only* detector uses features $\mathbf{z}_i$ from the frozen odometry model. For LiLY (Yao et al., 2022), these concatenate the 20-dimensional latent vector, 512-dimensional activations immediately before the final velocity layer, and four head-response summaries, giving 536 dimensions. Appendix F.1 defines the features and fitting procedure. Both detectors use logistic regression:

$$s_M = f_M(\mathbf{\Phi}_i), \qquad s_I = f_I(\mathbf{z}_i), \tag{15}$$

where s_M and s_I denote the failure scores from motion and internal features, respectively.

We further construct a *Combined* detector using score-level fusion. The base detectors produce probabilities. Their training-session out-of-fold probabilities are centered by their respective means, $\tilde{s}_{M,i} = s_{M,i} - \bar{s}_M$ and $\tilde{s}_{I,i} = s_{I,i} - \bar{s}_I$, and combined with their interaction as

$$\mathbf{q}_i = [\tilde{s}_{M,i}, \tilde{s}_{I,i}, \tilde{s}_{M,i}\tilde{s}_{I,i}], \tag{16}$$

and a second logistic regression produces the final failure score,

$$s_C = f_C(\mathbf{q}_i). \tag{17}$$

The fusion model is trained on these out-of-fold features. The base detectors are then refitted on all training sessions, and their held-out predictions use the same centering means.

4.2 Failure Detection Evaluation

We evaluate failure detection on unseen target sessions using a leave-one-session-out (LOSO) protocol. Each detector is trained on the remaining target sessions and tested on the held-out session under the same split. All feature standardization and inner-fold fitting use training sessions only. We

Table 6: Pooled out-of-session failure detection with LiLY.

		AUROC			AUPRC		
Target	Fail. (%)	Motion	Internal	Combined	Motion	Internal	Combined
KITTI	84.50	0.971	0.929	**0.978**	0.993	0.983	**0.994**
OxIOD	94.66	0.841	0.803	**0.871**	0.985	0.982	**0.987**
QuadSLAM	66.52	0.670	0.772	**0.791**	0.770	0.842	**0.846**

Table 7: Failure detection at a calibration FPR of 10%. Each entry reports recall / test FPR (%).

Target	Motion-only	Internal-only	Combined
KITTI	95.4 / 10.4	72.7 / 11.1	**96.8 / 10.4**
OxIOD	44.4 / 9.9	30.7 / 8.0	**55.1 / 8.8**
QuadSLAM	22.7 / 9.2	28.3 / 7.4	**34.6 / 9.7**

report LiLY in the main text and provide Tartan results in Appendix F.2. KAIST is omitted from the main comparison because only eight LiLY failures are available.

Table 6 reports the pooled out-of-session AUROC and AUPRC. Continuous scores from all held-out sessions are concatenated before evaluation.

For LiLY, motion features separate KITTI failures with an AUROC of 0.971, whereas internal features are more informative on QuadSLAM (0.772 versus 0.670 for motion). The latter comparison shows that internal states expose failure patterns less evident in the five motion features. Combining both sources gives the highest AUROC across the three targets, supporting their complementary roles in failure detection. Session-bootstrap intervals are reported in Appendix F.

We further evaluate the same LOSO predictions at a fixed operating point. For each fold, the decision threshold is the 90th percentile of calibration scores for non-failure windows in the training sessions and then applied unchanged to the held-out session. Table 7 reports the resulting recall and actual test FPR.

At the fixed operating point, motion information already provides high recall on KITTI, reaching 95.4%. Combining motion and internal features further increases recall to 96.8%. On OxIOD and QuadSLAM, the Combined detector improves recall from the best single-source results of 44.4% and 28.3% to 55.1% and 34.6%, respectively. These results further show that motion and internal features contribute complementary information to failure detection across different target domains.

5 Conclusion

This work studies cross-domain failure in learned inertial odometry by tracing how target-domain errors relate to source-domain motion coverage, source prediction behavior, and model internal states. Our results show that source motion coverage alone does not determine failure. Some failures extend prediction difficulties already present in the source domain, while others occur under source-covered conditions near source low-error motion regions. In the latter case, failed predictions often remain biased toward the source-domain velocity response, indicating that source-domain motion–velocity relations continue to shape predictions after domain shift. We further find that different failure types benefit from different target-domain adaptation data, and that motion and internal-state information provide complementary cues for failure detection on unseen target sessions.

These findings suggest a broader view of learned inertial odometry. Its ability to predict velocity without an explicitly given initial velocity relies on regularities learned from data, and the same regularities can become a source of failure when they no longer hold in the target domain. Cross-domain degradation should therefore be understood not only as a change in input distribution, but also as a change in whether the predictive relations learned from the source remain valid.

Reproducibility statement

We provide the experimental settings required to reproduce our results, including dataset preprocessing and splits, window construction, failure definition, motion descriptors, source-coverage estimation, motion-matching procedures, statistical tests, model configurations, and session-level evaluation protocols. Additional implementation details and hyperparameters are provided in the appendix and supplementary material. All reported results are obtained using fixed evaluation procedures defined from the source data unless otherwise stated.

AI Use Disclosure

In this work, we used generative AI tools to assist with language editing and polishing, and with literature retrieval and discovery. We did not use generative AI tools for research ideation or execution, generating synthetic datasets, or proving mathematical claims. We reviewed all AI-assisted work and take responsibility for the final content of this work.

References

Xiya Cao, Caifa Zhou, Dandan Zeng, and Yongliang Wang. RIO: Rotation-equivariance supervised learning of robust inertial odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6614–6623, 2022.

Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. IONet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 6468–6476, 2018. doi: 10.1609/aaai.v32i1.12102.

Benjamin Eyre, Elliot Creager, David Madras, Vardan Papyan, and Richard Zemel. Out of the ordinary: Spectrally adapting regression for covariate shift. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 12701–12722, 2024.

Sachini Herath, Hang Yan, and Yasutaka Furukawa. RoNIN: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3146–3152. IEEE, 2020. doi: 10.1109/ICRA40945.2020.9196860.

Royina Karegoudra Jayanth, Yinshuang Xu, Ziyun Wang, Evangelos Chatzipantazis, Kostas Daniilidis, and Daniel Gehrig. EqNIO: Subequivariant neural inertial odometry. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/6554c4151c2ccbd36da2d55c015b2039-Abstract-Conference.html.

Wenxin Liu, David Caruso, Eddy Ilg, Jing Dong, Anastasios I. Mourikis, Kostas Daniilidis, Vijay Kumar, and Jakob Engel. TLIO: Tight learned inertial odometry. *IEEE Robotics and Automation Letters*, 5(4):5653–5660, 2020. doi: 10.1109/LRA.2020.3007421.

Reese Pathak, Cong Ma, and Martin Wainwright. A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 17517–17530, 2022.

Elan Rosenfeld, Pradeep Ravikumar, and Andrej Risteski. Domain-adjusted regression or: ERM may already learn features sufficient for out-of-distribution generalization. *arXiv preprint arXiv:2202.06856*, 2022. URL https://arxiv.org/abs/2202.06856.

Hang Yan, Qi Shan, and Yasutaka Furukawa. RIDI: Robust IMU double integration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 621–636, 2018.

William Yang, Byron Zhang, and Olga Russakovsky. ImageNet-OOD: Deciphering modern out-of-distribution detection algorithms. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/b739cbaa0e94bfd7669d9573e9235411-Abstract-Conference.html.

Weiran Yao, Guangyi Chen, and Kun Zhang. Learning Latent Causal Dynamics. arXiv:2202.04828, 2022. URL https://arxiv.org/abs/2202.04828.

Shibo Zhao, Sifan Zhou, Raphael Blanchard, Yuheng Qiu, Wenshan Wang, and Sebastian Scherer. Tartan IMU: A light foundation model for inertial positioning in robotics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22520–22529, 2025a.

Shibo Zhao, Sifan Zhou, Yuchen Zhang, Ji Zhang, Chen Wang, Wenshan Wang, and Sebastian Scherer. Resilient odometry via hierarchical adaptation. *Science Robotics*, 10(109):eadv1818, 2025b. doi: 10.1126/scirobotics.adv1818.

A Related Work

A.1 Learned Inertial Odometry

Learning-based inertial odometry has evolved from using learned quantities to assist conventional inertial propagation toward directly estimating motion states from IMU sequences. RIDI learns velocity information to correct acceleration errors before integration (Yan et al., 2018), while IONet replaces continuous integration with window-based motion-state estimation (Chen et al., 2018). RoNIN further formulates inertial odometry as direct velocity regression from IMU sequences (Herath et al., 2020). Beyond direct regression, TLIO predicts relative displacement together with its uncertainty and tightly integrates these learned measurements into an EKF (Liu et al., 2020).

A.2 Cross-Domain Generalization in Learned Inertial Odometry

Cross-domain generalization has become a central problem in learned inertial odometry. RIO introduces rotation-equivariant supervision and test-time training to adapt the model to unseen data (Cao et al., 2022). EqNIO incorporates inertial symmetries directly into the network through equivariant canonicalization, reducing sensitivity to changes in reference frames and motion directions (Jayanth et al., 2025). More recently, TartanIMU uses multi-platform pretraining together with lightweight and test-time adaptation to improve transfer across different robotic platforms (Zhao et al., 2025a). These methods improve target-domain performance through stronger structural priors, broader training data, or adaptation.

A.3 Distribution Shift and Failure Analysis

Distribution shift in regression depends on more than the overall distance between source and target domains. Pathak et al. (Pathak et al., 2022) characterize regression under covariate shift through local source–target distribution mismatch, while Eyre et al. (Eyre et al., 2024) relate out-of-domain regression risk to the spectral structure of source and target data. Other studies further suggest that prediction failure does not necessarily imply the loss of task-relevant representations. DARE shows that features learned by empirical risk minimization can remain useful under distribution shift, while adaptation of the final predictor may recover target-domain performance (Rosenfeld et al., 2022). More broadly, recent OOD studies distinguish distribution shift from actual prediction failure and show that the two need not coincide (Yang et al., 2024). These findings motivate analyzing cross-domain degradation at the failure level rather than through aggregate domain differences alone. We examine how source motion coverage, regional prediction errors, and learned representations relate to target-domain failures in inertial odometry.

B Experimental Details and Definitions

B.1 Data and evaluation windows

We use four source-carrier reference sets and their corresponding target datasets. Table 8 reports the trajectories and windows used for regional error analysis and target evaluation. SourceTrain and SourceVal contain disjoint trajectories. Target groups follow the trajectory or recording identifiers in the evaluation caches and are kept intact when splitting calibration and evaluation data.

Table 8: Source and target evaluation data.

	SourceTrain		SourceVal			Target	
Source carrier	Trajectories	Windows	Trajectories	Windows	Dataset	Groups	Windows
Car	44	23,533	12	7,348	KITTI	152	4,232
Drone	289	14,213	48	2,327	KAIST	10	678
Human	26	29,270	7	7,738	OxIOD	17	32,866
Dog	36	14,915	13	6,301	QuadSLAM	8	1,168

SourceTrain window counts refer to the non-overlapping windows with model predictions used for regional error statistics. The dense SourceTrain motion references used for coverage and neighbor matching contain 939,604 Car, 557,249 Drone, 1,169,786 Human, and 595,196 Dog windows.

Motion features are computed from one-second physical IMU windows sampled at 100 Hz. SourceTrain regional errors use non-overlapping windows evaluated with the model state reset for each window. SourceVal uses the cached ten-second contexts, excluding padded windows through the validity mask. The dense SourceTrain motion reference uses one-second windows with a 25 ms stride. The temporal-change features use finite differences divided by $\Delta t = 0.01\,\mathrm{s}$.

B.2 Frozen models

LiLY (Yao et al., 2022) uses the frozen shared checkpoint selected at epoch 71 by validation L1 error. Its six IMU channels are normalized using training statistics computed with equal weight across the four carriers and equal weight across trajectories within each carrier. The normalization record contains 395 source-training trajectories. TartanIMU (Zhao et al., 2025a) uses the released unified checkpoint with a shared ResNet–LSTM backbone and carrier-specific velocity heads. We apply the head corresponding to each source carrier. Both models use the same cached physical IMU windows and reference velocities.

Table 9: Frozen-model inputs.

Model	Evaluation rate	Channel order
LiLY	100 Hz	Acceleration, gyroscope
TartanIMU	40 Hz	Gyroscope, acceleration

LiLY receives the normalized 100 Hz inputs. TartanIMU receives physical IMU inputs resampled from 100 to 40 Hz by polyphase filtering. Velocity predictions and references use the body/device-frame coordinates of the cached model labels.

The checkpoint SHA256 prefixes are `424ca2dbf13f5803` for LiLY and `ca8cc3e15a42ec91` for TartanIMU. The encoder and original velocity head remain frozen for failure characterization. Readout recovery fits an affine velocity map on frozen features; failure detection fits logistic regressions on frozen motion and internal features.

B.3 Error thresholds and motion standardization

All experiments use $v_{\min} = 1\,\mathrm{m/s}$. Failure thresholds are the corresponding model/carrier SourceVal 90th percentiles of relative velocity error. Table 10 reports $100\tau_S$; the dimensionless threshold in the failure definition is obtained by dividing these values by 100.

Table 10: SourceVal failure thresholds (%).

Source carrier	LiLY	TartanIMU
Car	72.59931	54.21193
Drone	87.22629	123.26026
Human	36.23965	26.07838
Dog	33.64501	29.03272

For regional statistics and motion-neighbor matching, the five motion features are standardized by the corresponding non-overlapping SourceTrain windows' means and population standard deviations. The one-third and two-third SourceTrain quantiles define the region boundaries, which are applied to SourceVal and Target. Coverage uses the frozen source Gaussian KDE and its SourceTrain 95th-percentile sparsity threshold, with equal total weight per source trajectory. These motion-feature statistics are separate from the raw-IMU normalization used by LiLY.

C Additional Motion-Space Analysis

C.1 Complete SourceTrain and SourceVal regional statistics

Tables 11–14 report every region containing at least one SourceTrain or SourceVal inference window: 201 Car, 165 Drone, 183 Human, and 95 Dog regions. Each row contains both models' relative-error median and estimated mode, reported as percentages; smaller values indicate lower errors. Window counts are shared by the two models. Regions empty in both splits are omitted.

The five-letter motion code lists the low (L), medium (M), or high (H) interval of $\phi_1, \ldots, \phi_5$ in that order. With $c_j \in \{0, 1, 2\}$ denoting L, M, and H, the region identifier is $1{+}81c_1{+}27c_2{+}9c_3{+}3c_4{+}c_5$. Values equal to a quantile boundary enter the higher interval. The outer intervals extend to infinity to include validation windows beyond the observed SourceTrain range.

Medians use all windows in each region. Modes are estimated for regions with at least 30 windows using a Gaussian KDE with Scott's bandwidth; the density peak is found numerically over the observed error range. For constant errors, the mode equals the median. A dash denotes an empty split for the median or fewer than 30 windows for the mode. All values use $v_{\min} = 1\,\mathrm{m/s}$. SourceTrain and SourceVal follow the inference-context settings in Appendix B.1.

Table 11: Complete regional error statistics for Car (errors in %).

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
1	LLLLL	1268	7.12	4.34	6.79	5.20	444	8.07	4.97	5.21	4.18
2	LLLLM	1718	2.89	1.80	2.79	1.94	715	4.47	3.01	3.95	3.23
3	LLLLH	182	6.02	4.47	5.96	5.10	50	9.53	7.71	7.41	6.52
4	LLLML	165	11.47	7.53	15.64	8.63	36	19.30	9.22	4.86	5.16
5	LLLMM	187	11.56	9.36	28.84	10.91	36	8.53	7.27	7.21	7.00
6	LLLMH	19	8.74	—	9.93	—	12	6.66	—	5.86	—
7	LLLHL	3	113.97	—	98.69	—	0	—	—	—	—
8	LLLHM	11	7.86	—	63.57	—	1	96.62	—	67.90	—
10	LLMLL	89	22.47	16.29	16.66	12.67	17	22.08	—	14.73	—
11	LLMLM	72	10.61	7.70	10.27	9.30	56	13.88	10.14	8.96	8.17
12	LLMLH	23	7.09	—	8.25	—	2	13.69	—	6.54	—
13	LLMML	132	18.76	12.84	28.39	13.57	41	32.90	17.84	25.08	12.73
14	LLMMM	558	6.33	4.97	25.52	14.08	173	9.53	7.78	16.39	11.17
15	LLMMH	37	6.72	4.54	11.50	9.89	10	17.28	—	12.78	—
16	LLMHL	0	—	—	—	—	1	49.99	—	42.19	—
17	LLMHM	20	5.08	—	36.64	—	11	7.19	—	32.02	—

Continued on next page.

Table 11 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
19	LLHLL	2	8.09	—	70.06	—	0	—	—	—	—
20	LLHLM	2	25.94	—	21.92	—	1	50.68	—	38.74	—
21	LLHLH	0	—	—	—	—	1	19.95	—	9.75	—
22	LLHML	93	5.82	6.64	71.46	68.83	58	173.42	191.59	71.31	67.00
23	LLHMM	56	7.36	6.41	50.98	13.49	43	10.68	11.00	22.94	11.85
24	LLHMH	5	6.56	—	5.59	—	3	4.29	—	19.17	—
25	LLHHL	20	6.11	—	69.07	—	37	97.76	195.54	70.80	67.24
26	LLHHM	6	9.55	—	67.85	—	13	9.86	—	57.86	—
27	LLHHH	1	25.13	—	77.37	—	0	—	—	—	—
28	LMLLL	99	5.87	5.36	6.92	6.07	54	7.88	6.45	5.65	5.83
29	LMLLM	159	9.46	7.53	6.78	6.20	58	10.89	8.61	5.28	4.49
30	LMLLH	5	9.17	—	5.52	—	0	—	—	—	—
31	LMLML	200	7.02	5.74	6.96	5.95	19	5.51	—	6.25	—
32	LMLMM	241	14.79	9.66	13.11	9.91	37	16.91	11.92	10.56	9.26
33	LMLMH	13	14.47	—	8.37	—	5	32.17	—	7.89	—
34	LMLHL	49	10.91	6.28	39.62	28.84	8	61.56	—	15.49	—
35	LMLHM	80	19.78	12.54	48.31	56.53	24	31.84	—	12.54	—
36	LMLHH	2	27.88	—	26.04	—	6	12.40	—	13.26	—
37	LMMLL	20	9.60	—	11.08	—	30	8.36	7.62	4.74	4.45
38	LMMLM	11	7.26	—	7.61	—	19	16.06	—	7.14	—
40	LMMML	456	7.53	5.64	9.90	8.30	130	8.65	7.84	7.33	6.23
41	LMMMM	563	9.19	6.64	13.94	8.49	184	12.59	9.77	11.03	8.51
42	LMMMH	4	15.18	—	16.97	—	1	19.92	—	53.52	—
43	LMMHL	133	16.53	13.40	29.69	13.97	41	29.46	14.91	21.81	16.38
44	LMMHM	385	8.92	6.87	25.41	15.25	97	10.42	7.59	17.08	12.59
45	LMMHH	1	18.30	—	67.02	—	0	—	—	—	—
49	LMHML	8	26.30	—	11.95	—	6	27.50	—	69.41	—
50	LMHMM	28	7.32	—	21.04	—	16	20.25	—	23.40	—
52	LMHHL	60	15.04	13.16	41.08	17.11	64	60.81	13.37	55.19	59.33
53	LMHHM	72	12.87	10.13	29.51	14.97	107	34.55	17.61	24.90	14.85
54	LMHHH	1	8.50	—	6.72	—	0	—	—	—	—
55	LHLLL	1	73.67	—	3.07	—	0	—	—	—	—
56	LHLLM	0	—	—	—	—	1	26.53	—	27.00	—
58	LHLML	1	56.45	—	73.05	—	1	56.58	—	2.29	—
59	LHLMM	1	44.77	—	48.14	—	0	—	—	—	—
61	LHLHL	2	64.20	—	0.97	—	6	48.93	—	8.90	—
62	LHLHM	0	—	—	—	—	6	13.80	—	8.15	—
64	LHMLL	0	—	—	—	—	1	74.83	—	31.67	—
67	LHMML	76	4.37	4.10	6.89	6.60	16	5.35	—	4.77	—
68	LHMMM	14	4.38	—	7.38	—	1	35.80	—	6.89	—
70	LHMHL	93	12.67	12.11	18.00	15.24	29	44.95	—	48.24	—
71	LHMHM	148	13.71	11.78	14.57	12.97	28	16.69	—	15.86	—
76	LHHML	1	66.35	—	20.68	—	0	—	—	—	—
79	LHHHL	157	13.67	10.54	14.08	12.25	88	33.39	18.76	20.39	15.08
80	LHHHM	91	14.57	12.15	20.69	17.15	43	14.30	13.91	12.43	11.07
82	MLLLL	157	17.72	11.79	14.78	12.20	15	24.34	—	17.16	—
83	MLLLM	280	25.30	11.31	19.72	12.59	60	25.32	16.31	17.16	14.65
84	MLLLH	379	9.92	8.06	8.65	6.86	127	8.82	7.56	6.87	5.66
85	MLLML	4	11.76	—	19.82	—	0	—	—	—	—
86	MLLMM	21	21.61	—	35.75	—	2	30.99	—	14.67	—
87	MLLMH	95	9.60	6.49	9.08	7.94	32	7.66	6.44	6.19	5.87
90	MLLHH	1	3.33	—	9.66	—	2	30.59	—	23.90	—
91	MLMLL	175	18.24	12.23	12.35	11.14	1	22.75	—	9.89	—

Continued on next page.

Table 11 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
92	MLMLM	65	28.50	18.53	24.61	21.28	4	11.22	—	14.52	—
93	MLMLH	159	8.22	6.97	8.25	7.04	37	8.38	7.74	8.47	6.78
94	MLMML	6	5.12	—	8.29	—	1	131.56	—	29.64	—
95	MLMMM	40	11.65	10.13	10.96	12.06	11	60.28	—	15.15	—
96	MLMMH	209	7.33	5.72	9.24	7.70	84	9.04	5.31	10.33	8.24
97	MLMHL	1	5.51	—	25.29	—	0	—	—	—	—
98	MLMHM	0	—	—	—	—	1	19.63	—	32.67	—
99	MLMHH	1	8.17	—	28.58	—	0	—	—	—	—
101	MLHLM	1	20.65	—	17.84	—	0	—	—	—	—
102	MLHLH	1	7.20	—	11.14	—	0	—	—	—	—
103	MLHML	12	4.60	—	76.30	—	0	—	—	—	—
104	MLHMM	8	14.19	—	41.11	—	8	12.74	—	6.89	—
105	MLHMH	91	7.19	5.66	9.38	7.71	46	10.14	6.36	7.39	6.07
106	MLHHL	38	3.36	2.77	68.90	67.15	2	6.06	—	89.45	—
107	MLHHM	1	5.44	—	27.15	—	2	50.10	—	39.49	—
108	MLHHH	22	4.39	—	8.35	—	13	10.00	—	5.52	—
109	MMLLL	70	15.23	12.18	16.29	14.90	10	22.58	—	12.49	—
110	MMLLM	156	19.24	12.54	15.85	11.94	61	11.72	9.53	12.24	9.74
111	MMLLH	107	10.90	8.71	7.82	6.45	40	11.23	8.80	4.78	4.12
112	MMLML	37	22.10	17.43	24.44	21.92	6	31.70	—	31.98	—
113	MMLMM	151	31.48	19.56	33.87	25.32	17	26.37	—	16.47	—
114	MMLMH	179	12.37	9.63	12.35	10.53	40	9.57	7.60	6.84	5.36
115	MMLHL	3	37.93	—	136.15	—	0	—	—	—	—
116	MMLHM	26	25.95	—	52.82	—	0	—	—	—	—
117	MMLHH	12	18.90	—	21.86	—	21	9.85	—	9.30	—
118	MMMLL	516	16.50	11.29	11.60	10.08	7	15.23	—	8.40	—
119	MMMLM	154	21.70	16.52	17.25	15.04	24	12.46	—	5.91	—
120	MMMLH	29	13.34	—	10.86	—	18	11.12	—	4.57	—
121	MMMML	106	15.65	12.07	16.49	13.99	43	16.66	10.63	12.02	9.07
122	MMMMM	295	12.63	9.03	12.96	10.47	108	16.14	11.99	11.24	8.87
123	MMMMH	308	9.83	7.39	10.50	8.39	69	14.50	10.60	12.21	10.14
124	MMMHL	13	17.31	—	18.19	—	1	2.05	—	16.55	—
125	MMMHM	59	15.57	11.31	27.03	18.33	30	21.13	18.48	17.91	14.20
126	MMMHH	59	23.09	13.27	24.57	21.61	16	20.02	—	24.89	—
127	MMHLL	85	14.65	11.96	11.36	11.70	0	—	—	—	—
128	MMHLM	13	26.49	—	16.50	—	0	—	—	—	—
130	MMHML	152	14.82	12.04	11.32	10.82	2	34.19	—	15.08	—
131	MMHMM	94	17.49	11.15	16.54	12.66	23	26.45	—	12.07	—
132	MMHMH	75	8.00	5.40	12.77	11.15	25	13.66	—	9.72	—
133	MMHHL	168	4.11	4.67	68.50	68.97	104	22.29	10.68	62.71	64.71
134	MMHHM	153	12.65	8.94	23.29	12.87	167	37.51	19.73	37.53	14.59
135	MMHHH	101	8.10	6.77	11.75	10.47	109	11.77	8.79	12.45	10.19
136	MHLLL	1	43.47	—	8.16	—	0	—	—	—	—
137	MHLLM	4	20.57	—	17.99	—	0	—	—	—	—
138	MHLLH	1	11.90	—	2.85	—	0	—	—	—	—
139	MHLML	12	26.55	—	19.01	—	1	22.05	—	44.55	—
140	MHLMM	25	34.62	—	26.81	—	0	—	—	—	—
141	MHLMH	2	28.71	—	21.48	—	1	21.27	—	11.86	—
142	MHLHL	6	27.18	—	42.17	—	0	—	—	—	—
143	MHLHM	1	136.32	—	56.90	—	0	—	—	—	—
144	MHLHH	0	—	—	—	—	1	24.60	—	3.56	—
145	MHMLL	11	21.77	—	10.75	—	2	12.17	—	12.41	—
146	MHMLM	5	22.05	—	23.64	—	0	—	—	—	—

Continued on next page.

Table 11 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
147	MHMLH	0	—	—	—	—	2	22.82	—	11.28	—
148	MHMML	99	7.51	6.38	13.91	11.19	57	11.29	10.47	13.52	9.92
149	MHMMM	120	14.59	11.00	16.19	11.43	53	13.73	11.69	11.43	6.82
150	MHMMH	22	13.97	—	8.62	—	5	7.45	—	6.52	—
151	MHMHL	83	12.21	8.67	17.51	12.73	18	12.20	—	17.71	—
152	MHMHM	201	17.79	12.01	20.09	14.18	46	10.21	9.96	15.38	10.43
153	MHMHH	34	11.79	10.02	15.29	12.11	10	10.65	—	10.71	—
154	MHHLL	2	15.31	—	4.45	—	0	—	—	—	—
157	MHHML	134	17.08	12.33	13.42	12.48	9	23.98	—	7.86	—
158	MHHMM	34	21.89	18.63	17.42	15.94	4	36.79	—	12.53	—
159	MHHMH	5	8.13	—	5.57	—	0	—	—	—	—
160	MHHHL	1567	8.37	5.97	16.48	12.35	248	14.65	11.82	15.57	13.32
161	MHHHM	537	16.03	11.92	19.48	14.93	232	18.18	13.87	14.31	12.12
162	MHHHH	50	17.91	12.29	17.80	13.34	17	19.62	—	15.48	—
163	HLLLL	20	21.59	—	11.90	—	6	10.69	—	5.32	—
164	HLLLM	45	33.95	18.20	26.70	16.83	6	43.52	—	22.47	—
165	HLLLH	866	9.26	8.05	12.48	8.96	356	6.69	6.21	9.54	6.46
167	HLLMM	1	20.87	—	27.85	—	0	—	—	—	—
168	HLLMH	143	10.45	8.91	14.34	9.39	69	7.82	7.30	8.13	7.73
171	HLLHH	0	—	—	—	—	2	7.28	—	8.97	—
172	HLMLL	5	24.91	—	11.84	—	1	52.05	—	32.36	—
173	HLMLM	15	40.04	—	29.35	—	2	17.68	—	19.73	—
174	HLMLH	149	8.70	7.83	12.88	9.06	41	8.54	8.13	10.52	8.88
175	HLMML	1	13.13	—	5.10	—	0	—	—	—	—
176	HLMMM	1	3.74	—	5.93	—	0	—	—	—	—
177	HLMMH	114	7.37	6.58	11.47	7.43	36	8.25	7.48	11.46	8.21
180	HLMHH	0	—	—	—	—	1	3.15	—	18.41	—
183	HLHLH	1	46.99	—	19.73	—	2	85.91	—	23.29	—
186	HLHMH	36	5.13	4.60	8.11	6.94	34	8.08	8.57	10.03	9.32
189	HLHHH	11	5.43	—	13.61	—	6	7.08	—	5.08	—
190	HMLLL	26	19.04	—	19.73	—	7	10.57	—	18.40	—
191	HMLLM	27	18.20	—	16.57	—	8	13.37	—	11.57	—
192	HMLLH	223	7.43	6.53	8.51	7.38	99	7.10	6.36	6.47	5.87
193	HMLML	1	25.49	—	8.15	—	0	—	—	—	—
194	HMLMM	11	34.56	—	31.89	—	0	—	—	—	—
195	HMLMH	273	9.00	7.36	10.01	8.47	74	6.20	5.71	6.38	5.39
198	HMLHH	11	20.84	—	10.25	—	42	6.36	5.59	6.06	5.32
199	HMMLL	30	25.40	16.55	17.04	14.30	5	16.65	—	17.35	—
200	HMMLM	48	29.71	19.02	20.27	10.08	11	30.44	—	20.40	—
201	HMMLH	186	10.27	8.75	10.62	7.59	84	6.50	6.80	5.52	4.65
202	HMMML	21	8.00	—	12.68	—	2	59.73	—	35.64	—
203	HMMMM	26	14.25	—	19.36	—	10	20.65	—	13.21	—
204	HMMMH	770	8.25	6.62	9.54	7.55	169	10.81	7.26	10.60	7.45
205	HMMHL	3	3.84	—	4.99	—	0	—	—	—	—
206	HMMHM	2	14.58	—	9.41	—	0	—	—	—	—
207	HMMHH	57	13.77	12.04	17.39	8.93	8	45.51	—	32.25	—
209	HMHLM	8	35.97	—	46.31	—	1	98.41	—	83.98	—
210	HMHLH	25	21.41	—	25.62	—	1	32.80	—	12.86	—
211	HMHML	6	34.52	—	37.78	—	0	—	—	—	—
212	HMHMM	17	24.51	—	37.48	—	1	99.94	—	71.38	—
213	HMHMH	145	8.65	7.06	9.29	7.95	62	11.72	10.12	10.89	9.37
215	HMHHM	5	22.41	—	19.07	—	0	—	—	—	—
216	HMHHH	212	8.16	6.68	9.85	8.23	206	9.84	8.46	9.90	8.18

Continued on next page.

Table 11 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
217	HHLLL	8	24.84	—	37.09	—	2	25.72	—	24.47	—
218	HHLLM	7	32.09	—	21.31	—	2	35.12	—	15.15	—
219	HHLLH	2	27.27	—	27.77	—	0	—	—	—	—
220	HHLML	4	27.51	—	13.11	—	0	—	—	—	—
221	HHLMM	19	37.90	—	23.02	—	1	10.87	—	21.57	—
222	HHLMH	19	17.65	—	12.26	—	3	14.55	—	6.96	—
224	HHLHM	1	67.49	—	13.22	—	0	—	—	—	—
225	HHLHH	1	11.07	—	13.95	—	4	6.20	—	9.23	—
226	HHMLL	24	20.49	—	17.21	—	2	10.69	—	25.88	—
227	HHMLM	18	32.26	—	27.83	—	6	13.44	—	14.77	—
228	HHMLH	24	15.18	—	14.97	—	20	3.89	—	8.09	—
229	HHMML	70	17.03	11.93	22.77	20.34	21	21.99	—	14.02	—
230	HHMMM	55	22.50	17.94	27.38	21.06	18	29.43	—	21.84	—
231	HHMMH	319	10.38	8.23	9.96	8.31	56	8.46	7.94	8.03	7.74
232	HHMHL	33	26.15	19.68	22.36	19.54	8	12.35	—	16.75	—
233	HHMHM	37	25.67	16.93	33.70	26.79	8	18.43	—	14.46	—
234	HHMHH	231	13.13	9.20	12.46	9.46	27	10.23	—	10.22	—
235	HHHLL	18	34.21	—	28.77	—	2	44.11	—	14.49	—
236	HHHLM	18	38.34	—	22.44	—	3	9.45	—	38.56	—
237	HHHLH	29	18.39	—	27.13	—	1	88.89	—	18.11	—
238	HHHML	109	25.17	21.10	32.01	21.53	3	11.73	—	27.88	—
239	HHHMM	134	21.68	14.79	23.60	16.64	8	13.82	—	31.21	—
240	HHHMH	355	7.37	5.06	7.77	6.50	51	6.44	3.95	6.89	5.69
241	HHHHL	879	12.15	8.24	18.33	13.57	251	15.05	11.00	16.12	14.57
242	HHHHM	482	18.99	13.15	23.98	18.00	133	25.06	16.37	16.12	14.53
243	HHHHH	1408	9.57	6.47	10.35	6.69	384	10.67	8.88	10.48	8.43

Table 12: Complete regional error statistics for Drone (errors in %).

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
1	LLLLL	984	20.14	18.08	28.28	22.83	37	0.24	1.77	1.25	2.96
2	LLLLM	392	21.57	16.18	38.03	38.69	108	44.33	37.57	79.02	107.02
3	LLLLH	264	17.31	14.14	33.86	30.99	72	31.71	12.77	39.79	32.81
4	LLLML	252	28.11	22.84	34.84	28.93	35	31.48	36.52	33.01	34.87
5	LLLMM	283	34.11	24.75	58.75	54.73	110	25.47	19.65	34.31	32.38
6	LLLMH	334	23.95	18.79	43.27	35.23	116	19.12	15.30	25.94	21.20
7	LLLHL	12	18.14	—	23.70	—	0	—	—	—	—
8	LLLHM	2	31.01	—	46.79	—	2	87.02	—	85.42	—
9	LLLHH	10	50.49	—	38.49	—	2	104.86	—	101.81	—
10	LLMLL	67	23.89	10.71	39.18	48.77	7	11.99	—	56.70	—
11	LLMLM	52	21.01	15.95	47.10	40.92	10	67.06	—	106.97	—
12	LLMLH	6	27.08	—	40.17	—	1	46.01	—	151.61	—
13	LLMML	225	0.03	1.06	2.50	2.74	33	0.02	1.32	1.03	1.05
14	LLMMM	95	32.73	20.38	54.91	38.44	10	46.34	—	50.02	—
15	LLMMH	20	22.98	—	53.58	—	5	16.60	—	21.56	—
16	LLMHL	71	5.00	3.63	3.10	3.83	12	37.58	—	28.15	—
17	LLMHM	65	21.17	15.69	26.32	22.65	1	20.34	—	58.20	—
18	LLMHH	9	18.85	—	59.83	—	0	—	—	—	—
22	LLHML	3	16.83	—	5.39	—	0	—	—	—	—

Continued on next page.

Table 12 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
23	LLHMM	1	7.84	—	90.46	—	0	—	—	—	—
25	LLHHL	4	11.82	—	14.00	—	0	—	—	—	—
28	LMLLL	63	18.24	11.50	37.31	32.50	10	15.22	—	32.07	—
29	LMLLM	16	14.43	—	35.42	—	3	58.92	—	46.30	—
31	LMLML	49	33.36	24.20	57.99	50.62	21	39.98	—	40.92	—
32	LMLMM	26	24.29	—	67.32	—	3	69.14	—	45.66	—
33	LMLMH	1	35.43	—	9.02	—	1	9.80	—	18.41	—
34	LMLHL	5	28.27	—	41.96	—	7	53.66	—	46.92	—
35	LMLHM	14	27.84	—	28.60	—	3	101.62	—	114.21	—
36	LMLHH	4	74.51	—	83.57	—	0	—	—	—	—
37	LMMLL	16	31.72	—	25.76	—	12	23.54	—	28.60	—
38	LMMLM	6	7.64	—	64.20	—	2	77.54	—	45.91	—
40	LMMML	49	25.28	17.86	30.86	22.04	23	31.24	—	37.04	—
41	LMMMM	8	12.75	—	54.21	—	2	40.17	—	83.80	—
43	LMMHL	438	29.90	25.68	39.83	32.16	43	39.92	35.20	41.71	37.57
44	LMMHM	392	19.98	14.67	28.78	22.95	4	74.99	—	45.79	—
45	LMMHH	14	26.46	—	36.69	—	0	—	—	—	—
46	LMHLL	2	1.89	—	2.42	—	1	0.04	—	0.18	—
47	LMHLM	2	5.35	—	49.06	—	0	—	—	—	—
49	LMHML	6	0.09	—	1.33	—	1	0.03	—	16.02	—
50	LMHMM	3	5.36	—	98.05	—	0	—	—	—	—
52	LMHHL	292	30.58	27.69	44.85	31.77	47	55.33	51.23	63.38	86.19
53	LMHHM	64	30.27	23.19	48.05	44.39	10	54.32	—	98.25	—
54	LMHHH	0	—	—	—	—	1	92.51	—	82.64	—
58	LHLML	1	27.13	—	7.44	—	0	—	—	—	—
64	LHMLL	3	33.47	—	15.80	—	0	—	—	—	—
67	LHMML	8	9.75	—	13.32	—	1	1.06	—	10.30	—
70	LHMHL	16	4.87	—	4.72	—	0	—	—	—	—
71	LHMHM	3	17.53	—	13.02	—	0	—	—	—	—
73	LHHLL	2	5.14	—	3.92	—	0	—	—	—	—
76	LHHML	1	15.23	—	16.70	—	0	—	—	—	—
79	LHHHL	78	21.73	17.30	35.20	27.54	12	39.79	—	55.73	—
80	LHHHM	5	26.67	—	21.05	—	3	64.10	—	71.01	—
83	MLLLM	15	14.99	—	28.49	—	0	—	—	—	—
84	MLLLH	290	17.63	14.71	48.18	36.90	231	63.82	54.35	85.46	75.05
86	MLLMM	10	17.96	—	46.42	—	2	44.73	—	134.06	—
87	MLLMH	586	15.04	11.13	56.74	65.87	118	51.12	48.13	83.78	72.53
90	MLLHH	35	18.29	14.99	63.40	66.71	4	52.58	—	97.70	—
92	MLMLM	22	16.62	—	33.19	—	6	31.09	—	82.78	—
93	MLMLH	42	14.83	14.80	43.17	40.82	55	70.34	68.90	95.12	94.79
95	MLMMM	14	10.25	—	30.99	—	2	58.70	—	35.01	—
96	MLMMH	312	15.07	10.28	67.11	72.32	44	38.36	31.65	71.23	49.33
98	MLMHM	6	33.07	—	52.97	—	0	—	—	—	—
99	MLMHH	149	17.32	12.88	71.40	71.83	5	37.17	—	69.46	—
100	MLHLL	1	9.82	—	74.22	—	0	—	—	—	—
101	MLHLM	3	6.51	—	24.04	—	3	32.72	—	55.87	—
102	MLHLH	0	—	—	—	—	1	31.55	—	113.95	—
104	MLHMM	3	12.42	—	32.69	—	1	28.02	—	35.03	—
105	MLHMH	0	—	—	—	—	1	34.14	—	50.31	—
106	MLHHL	0	—	—	—	—	5	19.96	—	20.25	—
109	MMLLL	32	14.56	12.07	46.17	43.26	6	35.57	—	45.27	—
110	MMLLM	127	14.20	12.19	42.77	31.38	3	21.39	—	66.10	—
111	MMLLH	167	14.45	12.80	50.32	34.89	8	50.10	—	131.57	—

Continued on next page.

Table 12 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
112	MMLML	13	33.64	—	41.85	—	6	36.24	—	40.47	—
113	MMLMM	104	13.75	12.42	34.79	27.99	6	67.72	—	68.34	—
114	MMLMH	209	14.82	12.93	45.09	31.08	36	50.35	45.04	73.13	50.89
115	MMLHL	1	46.59	—	64.74	—	1	71.99	—	48.90	—
116	MMLHM	5	20.77	—	32.75	—	0	—	—	—	—
117	MMLHH	24	11.90	—	68.93	—	1	53.34	—	73.17	—
118	MMMLL	54	20.67	13.39	35.36	30.17	24	34.48	—	40.01	—
119	MMMLM	120	13.83	11.87	33.20	21.68	19	20.79	—	41.24	—
120	MMMLH	79	16.27	13.19	34.77	25.80	13	48.16	—	140.95	—
121	MMMML	61	34.46	22.82	40.33	30.59	20	41.27	—	41.73	—
122	MMMMM	105	19.46	13.55	44.34	33.16	37	26.16	23.68	54.11	45.07
123	MMMMH	157	14.40	11.92	57.60	32.86	39	35.61	33.34	68.25	64.60
124	MMMHL	47	30.40	24.43	43.16	35.05	13	47.53	—	55.00	—
125	MMMHM	170	20.67	13.39	40.00	29.14	10	31.79	—	54.44	—
126	MMMHH	279	14.75	12.02	39.79	32.92	11	24.48	—	54.67	—
127	MMHLL	11	15.47	—	32.29	—	6	15.42	—	31.67	—
128	MMHLM	7	26.68	—	76.00	—	14	13.91	—	47.09	—
130	MMHML	16	21.39	—	32.15	—	12	21.25	—	43.85	—
131	MMHMM	12	9.02	—	26.12	—	43	17.24	15.16	38.56	35.17
132	MMHMH	2	9.02	—	19.40	—	6	38.59	—	74.44	—
133	MMHHL	163	25.94	19.63	37.77	26.78	49	30.23	20.28	55.33	37.55
134	MMHHM	213	24.39	14.33	43.13	32.00	59	42.72	28.36	78.69	66.70
135	MMHHH	45	10.15	8.33	52.21	44.42	13	29.13	—	103.39	—
136	MHLLL	65	18.06	18.71	31.50	29.47	6	26.59	—	39.18	—
137	MHLLM	14	8.12	—	48.38	—	0	—	—	—	—
138	MHLLH	4	12.32	—	19.60	—	0	—	—	—	—
139	MHLML	10	20.52	—	22.96	—	1	26.75	—	39.33	—
140	MHLMM	8	7.47	—	17.36	—	3	46.84	—	29.28	—
141	MHLMH	1	7.41	—	12.49	—	0	—	—	—	—
145	MHMLL	121	12.65	9.33	28.72	23.91	32	34.13	18.53	42.85	42.96
146	MHMLM	51	9.56	8.25	31.44	24.69	10	33.63	—	47.51	—
147	MHMLH	2	9.46	—	15.40	—	1	31.58	—	36.06	—
148	MHMML	60	14.59	13.17	26.13	20.11	60	41.46	35.93	44.18	38.52
149	MHMMM	39	8.79	8.45	40.06	27.61	18	34.61	—	49.49	—
150	MHMMH	5	28.62	—	25.22	—	2	30.76	—	33.72	—
151	MHMHL	37	21.57	15.70	27.82	24.21	10	34.34	—	31.90	—
152	MHMHM	18	9.26	—	25.04	—	1	26.70	—	60.71	—
153	MHMHH	1	18.50	—	54.39	—	0	—	—	—	—
154	MHHLL	14	8.91	—	25.71	—	2	48.75	—	50.27	—
155	MHHLM	1	7.14	—	3.07	—	1	12.13	—	14.81	—
157	MHHML	56	28.55	24.58	36.81	29.59	12	40.86	—	44.42	—
158	MHHMM	7	7.21	—	16.79	—	9	15.59	—	26.20	—
160	MHHHL	330	17.31	12.76	29.90	22.63	49	29.89	29.93	34.38	31.52
161	MHHHM	174	11.37	9.21	34.53	29.66	29	47.73	—	70.31	—
162	MHHHH	8	18.65	—	83.85	—	1	53.10	—	80.23	—
165	HLLLH	33	13.78	9.26	34.10	28.66	23	41.36	—	81.11	—
168	HLLMH	31	13.95	13.89	47.77	40.91	11	48.93	—	75.09	—
171	HLLHH	1	11.10	—	112.15	—	0	—	—	—	—
174	HLMLH	18	8.84	—	68.54	—	2	15.79	—	73.71	—
177	HLMMH	14	7.94	—	24.16	—	5	28.65	—	30.43	—
180	HLMHH	2	16.64	—	44.31	—	0	—	—	—	—
191	HMLLM	3	5.25	—	35.65	—	0	—	—	—	—
192	HMLLH	107	13.00	11.38	59.11	60.72	6	38.46	—	89.48	—

Continued on next page.

Table 12 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
194	HMLMM	5	9.48	—	31.15	—	0	—	—	—	—
195	HMLMH	85	14.46	12.63	48.57	41.13	0	—	—	—	—
198	HMLHH	3	3.73	—	60.28	—	0	—	—	—	—
200	HMMLM	57	9.57	8.13	21.10	16.26	1	10.20	—	50.55	—
201	HMMLH	170	12.66	11.13	35.06	28.70	14	19.66	—	130.47	—
203	HMMMM	32	9.71	9.48	26.27	23.31	2	13.24	—	54.41	—
204	HMMMH	106	12.29	11.85	47.15	33.39	25	27.29	—	65.15	—
206	HMMHM	9	11.64	—	23.97	—	0	—	—	—	—
207	HMMHH	136	9.00	8.62	50.45	33.49	1	31.38	—	128.22	—
208	HMHLL	2	56.80	—	84.80	—	1	68.73	—	28.46	—
209	HMHLM	79	12.51	11.98	30.55	25.68	8	8.79	—	34.76	—
210	HMHLH	16	11.84	—	25.34	—	3	9.19	—	23.52	—
211	HMHML	5	36.52	—	58.84	—	0	—	—	—	—
212	HMHMM	29	12.16	—	33.67	—	13	20.17	—	60.81	—
213	HMHMH	9	8.23	—	35.98	—	9	31.40	—	52.41	—
214	HMHHL	48	15.13	14.16	25.61	22.92	8	22.80	—	26.26	—
215	HMHHM	68	9.16	8.22	25.16	22.68	6	30.51	—	55.47	—
216	HMHHH	75	11.16	11.32	33.44	25.92	0	—	—	—	—
218	HHLLM	3	11.24	—	49.02	—	0	—	—	—	—
219	HHLLH	12	16.31	—	36.23	—	0	—	—	—	—
221	HHLMM	1	24.07	—	25.53	—	1	37.17	—	24.60	—
222	HHLMH	22	12.97	—	38.00	—	2	20.85	—	75.21	—
226	HHMLL	11	10.03	—	25.36	—	11	26.00	—	41.98	—
227	HHMLM	108	10.38	9.42	24.55	18.99	5	32.66	—	43.56	—
228	HHMLH	169	12.09	9.74	30.01	20.79	2	22.10	—	80.26	—
229	HHMML	7	11.89	—	30.46	—	8	43.22	—	41.75	—
230	HHMMM	98	10.36	9.41	19.39	14.43	33	26.08	24.75	49.03	42.28
231	HHMMH	181	9.52	7.96	25.51	19.84	31	19.58	18.60	56.49	53.61
232	HHMHL	11	12.50	—	19.65	—	0	—	—	—	—
233	HHMHM	23	10.47	—	21.97	—	6	22.20	—	46.18	—
234	HHMHH	71	11.24	11.13	22.36	17.47	8	13.07	—	46.33	—
235	HHHLL	203	8.77	7.57	20.61	17.17	14	47.80	—	38.96	—
236	HHHLM	472	11.80	11.39	24.17	17.63	0	—	—	—	—
237	HHHLH	158	11.08	10.22	26.36	20.06	0	—	—	—	—
238	HHHML	276	11.06	9.71	27.19	19.44	48	53.55	45.67	49.33	45.13
239	HHHMM	519	10.44	9.19	23.48	18.54	4	82.41	—	156.69	—
240	HHHMH	162	10.90	10.01	22.36	17.48	0	—	—	—	—
241	HHHHL	436	13.00	11.60	25.45	21.39	51	44.03	61.65	42.11	30.27
242	HHHHM	554	11.10	10.06	22.92	17.50	14	41.21	—	58.73	—
243	HHHHH	98	10.73	10.05	28.32	21.07	0	—	—	—	—

Table 13: Complete regional error statistics for Human (errors in %).

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
1	LLLLL	1610	1.02	0.81	1.22	0.95	423	0.88	0.64	1.05	0.74
2	LLLLM	2596	1.01	0.77	1.16	0.86	782	0.79	0.58	0.95	0.65
3	LLLLH	2571	0.91	0.67	1.08	0.78	864	0.73	0.57	0.86	0.66
4	LLLML	116	1.88	1.93	2.14	2.16	34	1.24	1.34	1.44	1.70
5	LLLMM	303	1.54	1.34	1.77	1.53	54	1.16	1.51	1.58	1.80

Continued on next page.

Table 13 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
6	LLLMH	345	1.69	1.53	1.97	1.72	67	0.79	0.85	0.93	0.95
7	LLLHL	4	1.17	—	1.82	—	0	—	—	—	—
8	LLLHM	11	1.18	—	1.83	—	0	—	—	—	—
9	LLLHH	12	1.99	—	2.17	—	1	0.60	—	0.72	—
10	LLMLL	45	1.12	1.05	1.45	1.28	18	0.82	—	1.04	—
11	LLMLM	66	1.80	1.65	2.10	1.91	18	1.34	—	1.49	—
12	LLMLH	76	1.59	1.43	1.84	1.49	35	0.79	0.73	0.94	0.76
13	LLMML	56	1.92	2.02	2.37	2.22	16	2.10	—	1.91	—
14	LLMMM	181	1.89	1.67	2.15	1.87	43	2.69	1.41	2.21	1.37
15	LLMMH	167	1.76	1.68	1.77	1.76	47	1.28	1.71	1.82	1.69
16	LLMHL	9	1.92	—	1.39	—	2	0.52	—	0.92	—
17	LLMHM	14	2.05	—	2.22	—	4	1.79	—	1.56	—
18	LLMHH	45	2.15	1.64	2.27	1.85	9	0.31	—	0.50	—
24	LLHMH	1	3.60	—	3.17	—	0	—	—	—	—
26	LLHHM	0	—	—	—	—	3	0.50	—	0.78	—
28	LMLLL	42	7.40	5.64	6.97	5.24	8	5.02	—	4.69	—
29	LMLLM	116	7.14	5.30	6.72	4.57	32	5.68	5.15	6.49	4.71
30	LMLLH	139	5.02	3.07	5.02	3.34	82	5.21	4.52	5.51	4.28
31	LMLML	33	7.34	5.83	5.20	4.54	3	3.33	—	11.57	—
32	LMLMM	115	12.15	7.22	10.24	7.22	11	7.90	—	8.97	—
33	LMLMH	124	11.19	7.05	10.06	7.43	17	7.76	—	12.37	—
35	LMLHM	3	23.99	—	11.19	—	0	—	—	—	—
36	LMLHH	5	32.22	—	13.57	—	0	—	—	—	—
37	LMMLL	3	2.63	—	1.98	—	0	—	—	—	—
38	LMMLM	14	5.79	—	4.07	—	6	4.47	—	4.43	—
39	LMMLH	22	3.47	—	3.88	—	8	3.36	—	4.96	—
40	LMMML	45	5.06	4.39	8.23	4.39	10	14.85	—	19.86	—
41	LMMMM	222	6.92	4.89	7.38	5.35	34	7.46	3.82	6.81	5.32
42	LMMMH	247	9.67	5.43	8.57	5.85	52	5.90	4.11	6.29	4.89
43	LMMHL	11	3.19	—	5.26	—	1	25.83	—	18.47	—
44	LMMHM	101	11.73	7.42	10.11	7.16	22	26.55	—	19.35	—
45	LMMHH	106	12.45	7.35	11.42	9.54	11	13.24	—	12.16	—
49	LMHML	1	1.60	—	1.66	—	0	—	—	—	—
51	LMHMH	4	1.20	—	1.30	—	3	0.48	—	0.55	—
52	LMHHL	2	1.21	—	1.60	—	1	1.44	—	1.43	—
53	LMHHM	9	2.96	—	2.88	—	3	0.49	—	0.39	—
54	LMHHH	23	7.14	—	5.37	—	5	1.20	—	1.71	—
59	LHLMM	3	11.10	—	10.91	—	0	—	—	—	—
60	LHLMH	6	27.06	—	12.80	—	0	—	—	—	—
62	LHLHM	1	9.54	—	5.74	—	2	19.09	—	7.22	—
63	LHLHH	1	42.66	—	15.49	—	0	—	—	—	—
67	LHMML	2	66.02	—	22.17	—	0	—	—	—	—
68	LHMMM	18	21.77	—	16.61	—	1	15.79	—	9.51	—
69	LHMMH	14	22.54	—	13.87	—	2	26.41	—	18.11	—
70	LHMHL	6	70.59	—	27.32	—	19	17.88	—	12.95	—
71	LHMHM	33	17.61	17.27	15.75	15.63	86	16.98	15.90	15.94	15.78
72	LHMHH	48	25.19	21.05	12.84	11.25	52	14.14	13.15	13.82	12.63
80	LHHHM	3	9.19	—	4.13	—	2	11.30	—	6.97	—
81	LHHHH	7	22.45	—	24.07	—	4	21.27	—	23.03	—
82	MLLLL	410	2.08	1.76	2.02	1.71	117	2.16	1.83	2.06	1.65
83	MLLLM	174	2.73	2.44	3.05	2.21	51	3.83	2.90	4.22	2.80
84	MLLLH	70	3.01	2.35	3.58	2.85	22	4.67	—	4.01	—
85	MLLML	43	3.54	2.75	4.16	2.99	6	7.12	—	7.87	—

Continued on next page.

Table 13 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
86	MLLMM	23	7.54	—	7.87	—	3	6.76	—	5.87	—
87	MLLMH	11	8.01	—	7.00	—	0	—	—	—	—
88	MLLHL	1	2.54	—	1.61	—	0	—	—	—	—
89	MLLHM	1	1.21	—	2.18	—	0	—	—	—	—
91	MLMLL	201	2.32	1.81	2.38	1.90	79	2.18	1.88	2.27	1.90
92	MLMLM	82	2.75	2.50	2.76	2.67	37	2.70	2.04	2.43	2.22
93	MLMLH	39	2.06	1.73	2.32	2.02	34	1.04	1.16	1.36	1.37
94	MLMML	140	2.69	2.48	2.87	2.60	47	2.23	1.71	2.34	1.99
95	MLMMM	103	3.58	2.74	3.90	3.04	23	4.01	—	4.03	—
96	MLMMH	57	2.17	2.02	2.49	2.23	11	2.34	—	2.67	—
97	MLMHL	3	2.82	—	1.46	—	1	12.36	—	6.60	—
98	MLMHM	3	7.87	—	7.40	—	0	—	—	—	—
99	MLMHH	4	2.17	—	2.58	—	1	3.40	—	0.85	—
100	MLHLL	1	0.84	—	1.26	—	0	—	—	—	—
103	MLHML	3	1.05	—	0.92	—	1	1.31	—	1.03	—
105	MLHMH	6	1.45	—	1.12	—	2	0.37	—	0.50	—
106	MLHHL	0	—	—	—	—	1	1.71	—	0.78	—
107	MLHHM	0	—	—	—	—	1	0.83	—	1.18	—
109	MMLLL	140	6.07	3.43	5.70	4.12	53	8.67	7.34	8.17	5.96
110	MMLLM	158	6.04	3.04	6.18	4.15	66	8.70	4.80	8.03	5.22
111	MMLLH	222	3.78	2.93	4.60	3.43	102	6.66	3.80	6.34	3.41
112	MMLML	64	12.24	7.79	9.58	6.88	14	18.13	—	19.23	—
113	MMLMM	67	10.51	7.07	9.48	7.37	18	15.87	—	11.30	—
114	MMLMH	87	8.94	5.41	8.38	6.90	12	8.32	—	9.49	—
115	MMLHL	2	34.65	—	26.76	—	1	16.83	—	12.14	—
116	MMLHM	1	5.13	—	21.53	—	0	—	—	—	—
117	MMLHH	2	3.85	—	8.49	—	0	—	—	—	—
118	MMMLL	145	4.97	2.97	5.72	3.31	61	5.58	4.17	5.25	4.02
119	MMMLM	202	5.16	3.35	4.95	3.25	58	6.10	4.06	6.73	3.54
120	MMMLH	177	4.78	3.44	5.10	3.58	45	4.27	3.48	5.33	3.11
121	MMMML	772	11.03	4.48	9.67	5.16	182	12.18	6.69	9.35	5.84
122	MMMMM	1059	11.18	5.45	10.56	5.90	219	11.72	6.09	10.33	6.28
123	MMMMH	956	9.91	4.36	9.10	4.45	190	8.28	4.82	7.38	4.10
124	MMMHL	177	18.43	13.29	13.64	8.18	66	21.78	16.09	15.91	15.95
125	MMMHM	267	17.20	14.40	12.97	9.94	78	25.13	23.10	13.64	11.19
126	MMMHH	253	17.15	10.15	12.59	9.41	44	25.53	23.52	14.15	12.16
127	MMHLL	2	0.88	—	2.85	—	2	1.29	—	0.78	—
128	MMHLM	1	18.41	—	21.14	—	0	—	—	—	—
129	MMHLH	0	—	—	—	—	2	1.65	—	3.11	—
130	MMHML	91	2.83	2.41	3.36	2.52	29	2.37	—	1.97	—
131	MMHMM	113	5.02	3.45	4.88	3.73	24	4.98	—	4.86	—
132	MMHMH	158	4.48	3.64	4.75	4.00	34	1.45	1.96	1.75	2.21
133	MMHHL	74	8.24	4.69	8.91	5.02	22	13.83	—	12.31	—
134	MMHHM	184	9.75	5.53	8.58	5.86	51	22.39	18.61	15.61	14.84
135	MMHHH	193	6.65	4.75	8.09	5.94	37	10.65	5.58	8.79	3.75
136	MHLLL	4	8.80	—	4.61	—	1	14.13	—	15.80	—
137	MHLLM	3	12.88	—	11.23	—	1	20.69	—	26.98	—
138	MHLLH	8	8.90	—	7.22	—	10	15.50	—	10.51	—
139	MHLML	5	16.68	—	15.70	—	2	19.37	—	7.36	—
140	MHLMM	6	14.56	—	8.39	—	1	49.39	—	16.81	—
141	MHLMH	18	22.23	—	19.52	—	3	12.23	—	10.12	—
143	MHLHM	0	—	—	—	—	1	10.76	—	12.26	—
145	MHMLL	1	7.95	—	3.86	—	0	—	—	—	—

Continued on next page.

Table 13 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
146	MHMLM	3	5.41	—	3.53	—	3	9.85	—	9.09	—
147	MHMLH	7	8.81	—	7.11	—	7	6.34	—	4.48	—
148	MHMML	93	19.35	15.85	16.27	15.31	11	32.78	—	20.23	—
149	MHMMM	192	17.22	13.07	14.26	12.47	26	16.82	—	14.54	—
150	MHMMH	234	18.19	13.65	15.34	10.38	35	20.01	14.83	15.24	12.71
151	MHMHL	126	25.20	23.19	17.23	14.71	131	25.34	21.31	17.81	16.85
152	MHMHM	305	21.65	18.22	16.47	15.10	96	27.18	22.07	19.20	16.89
153	MHMHH	365	22.50	18.10	15.72	14.24	58	25.53	24.30	16.25	13.63
157	MHHML	8	8.07	—	10.70	—	1	14.92	—	20.14	—
158	MHHMM	13	27.37	—	17.68	—	2	15.91	—	11.79	—
159	MHHMH	15	16.12	—	8.05	—	1	12.13	—	28.57	—
160	MHHHL	116	20.92	17.18	17.19	14.69	25	26.97	—	20.04	—
161	MHHHM	563	24.85	21.70	17.86	15.15	77	24.15	20.57	18.59	17.00
162	MHHHH	729	25.14	19.96	19.00	17.00	77	24.79	23.58	16.06	14.42
163	HLLLL	25	3.55	—	3.05	—	2	3.52	—	3.09	—
164	HLLLM	2	6.04	—	6.70	—	1	9.46	—	12.05	—
166	HLLML	1	1.74	—	0.98	—	0	—	—	—	—
172	HLMLL	74	3.23	2.71	2.80	2.49	23	3.39	—	3.57	—
173	HLMLM	2	8.82	—	13.20	—	2	6.17	—	7.18	—
174	HLMLH	1	9.47	—	6.01	—	0	—	—	—	—
175	HLMML	33	6.04	4.89	5.11	3.62	13	4.25	—	5.72	—
176	HLMMM	0	—	—	—	—	1	11.92	—	14.46	—
178	HLMHL	3	7.45	—	5.96	—	0	—	—	—	—
181	HLHLL	5	4.84	—	3.56	—	2	4.31	—	4.01	—
182	HLHLM	1	8.72	—	6.67	—	0	—	—	—	—
184	HLHML	6	3.40	—	4.35	—	2	4.57	—	1.73	—
185	HLHMM	1	3.51	—	4.04	—	0	—	—	—	—
190	HMLLL	20	10.44	—	11.10	—	3	13.30	—	16.16	—
191	HMLLM	6	7.13	—	6.55	—	2	10.04	—	15.51	—
192	HMLLH	2	10.28	—	19.28	—	0	—	—	—	—
193	HMLML	7	14.22	—	20.56	—	1	22.72	—	18.06	—
194	HMLMM	2	34.51	—	36.21	—	0	—	—	—	—
199	HMMLL	145	5.93	4.24	5.87	4.02	64	5.49	4.16	5.80	4.12
200	HMMLM	44	6.38	4.99	6.66	5.45	11	10.71	—	10.35	—
201	HMMLH	12	6.95	—	11.93	—	11	5.36	—	6.51	—
202	HMMML	569	13.86	7.32	11.58	6.77	135	16.69	9.38	12.71	7.65
203	HMMMM	181	13.31	6.93	11.86	6.89	47	16.05	7.83	9.66	7.23
204	HMMMH	102	8.21	5.58	8.52	5.65	28	9.15	—	11.09	—
205	HMMHL	75	17.78	14.18	14.08	9.39	41	25.35	25.38	15.00	13.39
206	HMMHM	30	17.91	14.25	14.81	9.56	8	28.51	—	14.16	—
207	HMMHH	3	8.06	—	9.80	—	2	9.76	—	10.37	—
208	HMHLL	26	3.46	—	3.17	—	17	3.15	—	2.87	—
209	HMHLM	4	3.97	—	3.02	—	3	2.65	—	1.68	—
210	HMHLH	1	6.89	—	5.15	—	1	1.88	—	0.71	—
211	HMHML	750	8.86	4.70	8.12	4.13	297	8.05	4.64	7.38	4.17
212	HMHMM	168	7.05	4.64	6.98	4.69	48	5.36	3.96	4.93	4.10
213	HMHMH	77	4.46	4.07	4.07	3.77	28	5.88	—	4.07	—
214	HMHHL	393	17.26	7.90	13.20	8.12	167	21.39	18.71	15.55	10.09
215	HMHHM	117	14.19	9.82	11.75	8.58	46	22.22	11.98	13.53	11.09
216	HMHHH	68	8.87	6.35	8.00	6.33	17	14.72	—	10.66	—
217	HHLLL	2	9.89	—	8.93	—	1	27.70	—	14.75	—
218	HHLLM	3	8.65	—	16.65	—	0	—	—	—	—
219	HHLLH	3	13.94	—	24.46	—	1	1.88	—	5.70	—

Continued on next page.

Table 13 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
220	HHLML	2	12.03	—	15.03	—	0	—	—	—	—
221	HHLMM	2	9.05	—	16.33	—	0	—	—	—	—
222	HHLMH	4	19.29	—	12.02	—	1	30.47	—	9.31	—
226	HHMLL	5	9.10	—	7.22	—	4	10.46	—	11.48	—
227	HHMLM	4	10.75	—	9.04	—	3	9.26	—	4.67	—
228	HHMLH	20	6.25	—	6.21	—	4	5.50	—	6.43	—
229	HHMML	203	17.99	14.95	18.23	18.05	38	32.20	30.89	19.56	18.00
230	HHMMM	167	15.16	11.37	13.59	10.89	32	16.36	16.23	15.30	13.71
231	HHMMH	267	13.57	7.88	14.36	8.04	48	12.32	9.63	9.88	7.33
232	HHMHL	122	23.51	21.03	18.34	16.50	55	27.87	20.62	21.68	21.59
233	HHMHM	85	24.25	18.42	18.66	14.68	22	31.14	—	21.35	—
234	HHMHH	92	26.03	23.19	19.29	18.64	17	21.99	—	21.43	—
235	HHHLL	0	—	—	—	—	1	25.94	—	22.71	—
238	HHHML	463	14.44	7.75	12.19	7.89	127	13.95	9.79	13.61	9.21
239	HHHMM	228	15.02	6.00	11.18	7.12	51	9.15	6.88	8.02	6.42
240	HHHMH	183	10.70	6.77	9.71	6.71	55	7.98	5.98	7.70	6.16
241	HHHHL	2221	24.01	11.23	18.37	12.59	599	26.85	26.45	20.23	16.98
242	HHHHM	1377	28.49	24.33	21.41	14.43	219	30.01	19.05	20.48	17.24
243	HHHHH	1348	27.96	20.83	21.63	15.92	124	27.03	27.31	19.61	14.59

Table 14: Complete regional error statistics for Dog (errors in %).

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
1	LLLLL	738	0.41	0.51	0.42	0.52	170	1.00	1.08	0.97	0.66
2	LLLLM	2384	0.22	0.27	0.22	0.27	793	0.49	0.62	0.47	0.40
3	LLLLH	1713	0.18	0.19	0.18	0.17	265	0.57	0.71	0.55	0.53
4	LLLML	19	2.19	—	1.84	—	5	5.60	—	5.93	—
5	LLLMM	6	2.65	—	2.64	—	3	2.03	—	1.92	—
6	LLLMH	1	2.68	—	2.68	—	1	1.10	—	1.08	—
10	LLMLL	5	2.64	—	1.94	—	0	—	—	—	—
11	LLMLM	2	3.41	—	3.41	—	0	—	—	—	—
13	LLMML	6	3.10	—	2.39	—	3	1.11	—	1.11	—
14	LLMMM	0	—	—	—	—	1	2.63	—	2.62	—
15	LLMMH	0	—	—	—	—	1	0.59	—	0.59	—
28	LMLLL	6	1.31	—	1.37	—	3	24.17	—	24.19	—
29	LMLLM	10	0.89	—	1.07	—	1	1.53	—	1.58	—
30	LMLLH	6	1.05	—	1.02	—	0	—	—	—	—
31	LMLML	11	5.02	—	5.01	—	3	3.16	—	2.99	—
32	LMLMM	3	1.41	—	1.76	—	2	2.89	—	2.88	—
33	LMLMH	2	1.03	—	1.03	—	0	—	—	—	—
38	LMMLM	4	0.97	—	1.16	—	1	2.36	—	2.66	—
40	LMMML	30	2.49	2.42	1.96	2.06	21	2.79	—	2.86	—
41	LMMMM	24	2.82	—	2.09	—	30	2.07	2.49	2.00	1.78
42	LMMMH	2	1.15	—	1.13	—	0	—	—	—	—
82	MLLLL	11	5.28	—	5.28	—	3	12.53	—	8.03	—
83	MLLLM	3	5.32	—	5.16	—	1	0.55	—	0.69	—
84	MLLLH	48	5.09	3.93	2.41	2.08	1	2.06	—	1.87	—
85	MLLML	3	3.50	—	3.69	—	0	—	—	—	—
86	MLLMM	0	—	—	—	—	1	4.19	—	3.56	—

Continued on next page.

Table 14 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
87	MLLMH	1	3.96	—	1.77	—	0	—	—	—	—
91	MLMLL	17	2.57	—	2.25	—	2	12.83	—	3.22	—
92	MLMLM	7	2.93	—	2.91	—	2	5.35	—	3.45	—
93	MLMLH	0	—	—	—	—	2	2.52	—	2.22	—
94	MLMML	1	1.73	—	1.71	—	0	—	—	—	—
95	MLMMM	1	2.37	—	1.93	—	0	—	—	—	—
96	MLMMH	5	5.77	—	4.72	—	1	1.36	—	1.36	—
109	MMLLL	2	10.50	—	11.49	—	0	—	—	—	—
110	MMLLM	1	1.44	—	1.65	—	1	2.19	—	2.31	—
111	MMLLH	1	3.76	—	3.45	—	1	3.52	—	3.51	—
112	MMLML	1	33.84	—	33.83	—	4	27.05	—	26.94	—
113	MMLMM	1	3.06	—	3.04	—	0	—	—	—	—
118	MMMLL	7	2.90	—	2.32	—	1	5.24	—	6.23	—
119	MMMLM	5	3.00	—	2.90	—	1	5.63	—	7.07	—
120	MMMLH	1	1.16	—	1.24	—	1	10.43	—	3.35	—
121	MMMML	714	7.06	5.11	8.19	5.70	714	6.30	4.29	6.67	5.25
122	MMMMM	1023	32.51	8.22	25.86	9.80	477	15.65	8.61	12.61	8.03
123	MMMMH	518	19.33	7.24	18.96	8.49	174	11.38	7.75	10.37	7.86
124	MMMHL	203	5.89	4.67	9.62	7.12	74	9.09	6.26	10.75	9.39
125	MMMHM	48	6.23	5.37	8.14	6.52	24	10.72	—	8.35	—
126	MMMHH	29	5.58	—	6.87	—	20	6.03	—	5.80	—
130	MMHML	41	13.53	8.26	17.46	8.95	19	10.00	—	6.72	—
131	MMHMM	41	21.94	12.80	17.49	13.18	6	8.21	—	7.44	—
132	MMHMH	6	8.03	—	6.93	—	0	—	—	—	—
133	MMHHL	61	5.45	4.51	8.55	7.54	13	5.77	—	7.27	—
134	MMHHM	22	6.99	—	8.84	—	7	11.72	—	12.22	—
135	MMHHH	6	10.30	—	12.61	—	2	3.10	—	4.95	—
148	MHMML	218	3.54	2.91	5.37	4.65	101	4.53	3.54	5.19	4.60
149	MHMMM	23	5.31	—	7.12	—	20	5.82	—	5.10	—
150	MHMMH	2	11.58	—	18.11	—	3	8.87	—	7.57	—
151	MHMHL	232	5.50	4.38	8.66	7.79	38	9.21	7.65	10.69	10.08
152	MHMHM	26	8.58	—	8.77	—	10	6.97	—	6.27	—
153	MHMHH	8	7.25	—	13.24	—	3	9.26	—	9.38	—
157	MHHML	147	4.97	3.95	5.77	4.88	39	5.12	4.58	7.70	5.03
158	MHHMM	18	7.50	—	10.14	—	13	9.39	—	7.64	—
160	MHHHL	1275	6.76	4.60	10.16	8.52	347	9.78	7.48	8.95	7.53
161	MHHHM	187	9.33	7.42	11.53	9.58	64	7.41	5.77	7.76	5.87
162	MHHHH	7	9.55	—	8.21	—	6	8.03	—	12.56	—
174	HLMLH	0	—	—	—	—	3	14.83	—	3.90	—
177	HLMMH	1	4.79	—	3.58	—	0	—	—	—	—
190	HMLLL	1	1.94	—	7.35	—	1	11.69	—	11.42	—
193	HMLML	0	—	—	—	—	1	12.37	—	11.96	—
199	HMMLL	0	—	—	—	—	1	31.00	—	22.91	—
200	HMMLM	0	—	—	—	—	2	6.64	—	14.71	—
201	HMMLH	0	—	—	—	—	6	16.87	—	9.75	—
202	HMMML	42	20.16	14.38	20.11	12.64	24	18.11	—	12.55	—
203	HMMMM	213	44.00	52.08	35.80	29.90	111	30.96	27.76	30.80	16.63
204	HMMMH	1311	21.50	8.28	21.16	9.38	765	11.62	7.39	12.19	8.35
205	HMMHL	3	32.75	—	5.20	—	1	26.75	—	19.32	—
206	HMMHM	4	13.18	—	18.32	—	1	13.50	—	23.30	—
207	HMMHH	135	7.58	5.68	9.57	6.86	98	8.16	7.43	8.64	7.50
211	HMHML	84	21.52	13.24	19.64	12.87	15	17.30	—	15.18	—
212	HMHMM	159	43.22	56.29	31.76	26.05	62	25.59	13.11	23.52	15.22

Continued on next page.

Table 14 continued.

		SourceTrain					SourceVal				
			LiLY		Tartan			LiLY		Tartan	
Region	Motion	N	Med.	Mode	Med.	Mode	N	Med.	Mode	Med.	Mode
213	HMHMH	107	16.25	9.70	16.21	11.41	30	18.63	13.89	16.70	11.97
214	HMHHL	2	11.70	—	20.73	—	2	24.81	—	20.98	—
215	HMHHM	4	17.39	—	16.53	—	1	12.67	—	12.31	—
216	HMHHH	77	10.88	9.62	11.88	8.89	18	7.97	—	8.31	—
229	HHMML	7	53.58	—	53.59	—	11	36.71	—	33.78	—
230	HHMMM	15	65.33	—	55.42	—	9	31.71	—	49.86	—
231	HHMMH	33	9.19	9.23	11.70	10.48	37	7.48	5.95	7.81	6.92
232	HHMHL	3	47.31	—	13.42	—	3	46.37	—	25.81	—
233	HHMHM	5	8.63	—	4.72	—	1	15.23	—	35.97	—
234	HHMHH	38	9.33	8.18	9.68	7.93	30	15.99	12.80	14.14	11.81
238	HHHML	33	12.45	10.71	19.55	16.41	21	17.30	—	14.41	—
239	HHHMM	37	18.95	11.77	29.59	18.94	32	30.66	27.18	29.69	26.74
240	HHHMH	61	5.80	3.03	8.35	2.68	29	13.06	—	10.10	—
241	HHHHL	1049	11.64	8.52	16.60	12.07	643	11.04	9.29	10.94	8.56
242	HHHHM	695	16.83	12.33	18.02	12.15	311	13.63	11.60	11.51	9.08
243	HHHHH	853	17.00	11.72	16.65	12.83	533	16.15	12.74	14.28	11.17

D Extended Cross-Domain Failure Decomposition

D.1 Source-reference matching

Each target is matched to SourceTrain from the corresponding source carrier. Region labels use SourceTrain inference errors: regions containing at least 30 windows are high-error if their median exceeds the overall SourceTrain median and low-error otherwise; smaller regions are unconfirmed. Neighbor searches use the dense SourceTrain motion reference, standardized by the frozen SourceTrain means and standard deviations.

Thresholds are calibrated using SourceVal non-failures outside high-error regions, queried against SourceTrain neighbors outside high-error regions. Candidate speed bins use quartiles of the dense SourceTrain ground-truth speeds. Adjacent bins are merged until each contains at least 100 calibration windows from at least three trajectories. A bin is merged with the adjacent bin having fewer windows, with ties resolved to the left. Motion thresholds are the 95th percentiles of nearest-neighbor distances. Bin-specific thresholds are used when their maximum-to-minimum ratio is at least two; otherwise, a global threshold is used. Within each retained bin, the speed threshold is the 95th percentile of the minimum absolute speed difference among neighbors satisfying the motion threshold, computed over calibration windows with eligible neighbors. These rules are applied separately to each model and carrier.

Target windows first receive their own region label. Covered failures in high-error regions form the first category. The remaining covered failures are matched to source neighbors, with low-error-region neighbors taking priority, followed by high-error-region neighbors, unconfirmed-region neighbors, and no eligible neighbor. A low-error-region match is speed-matched when at least one eligible neighbor satisfies the speed threshold. Velocity-proximity analysis uses the same speed-mismatched failures and all their eligible low-error-region neighbors. Predictions and ground-truth vectors retain the body/device coordinates of the model labels.

D.2 Failure counts and matching results

Table 15 reports failure counts in source-sparse and source-covered windows for both models. All results use $v_{\min} = 1\,\mathrm{m/s}$ and the corresponding model–carrier SourceVal-Q90 threshold.

Table 15: Failure rates by source coverage.

		Source-sparse			Source-covered		
Target	Model	N	Failures	Rate (%)	N	Failures	Rate (%)
KITTI	LiLY	153	114	74.51	4,079	3,462	84.87
KITTI	Tartan	153	153	100.00	4,079	3,498	85.76
KAIST	LiLY	175	2	1.14	503	6	1.19
KAIST	Tartan	175	1	0.57	503	2	0.40
OxIOD	LiLY	7,341	7,174	97.73	25,525	23,938	93.78
OxIOD	Tartan	7,341	7,236	98.57	25,525	24,148	94.61
QuadSLAM	LiLY	944	698	73.94	224	79	35.27
QuadSLAM	Tartan	944	851	90.15	224	101	45.09

N is the number of windows in each coverage group. Rates use the corresponding group count as the denominator.

The complete LiLY partition is reported in Table 4. For TartanIMU, Table 16 reports source high-error-region membership and the motion-matched, speed-mismatched subset on KITTI and OxIOD; the other residual failures are aggregated.

Table 17 compares target predictions with the matched source velocity reference and target ground truth. The source reference is the coordinate-wise median over all eligible SourceTrain neighbors in low-error regions. Each prediction is assigned to the reference with smaller Euclidean distance.

Table 16: TartanIMU covered-failure groups.

Target	Covered failures	High-error region	Speed-mismatched	Other residual
KITTI	3,498	6 (0.17%)	3,481 (99.51%)	11 (0.31%)
OxIOD	24,148	13,788 (57.10%)	8,906 (36.88%)	1,454 (6.02%)

The last three columns partition the covered failures and report counts (percentages). High-error region refers to the target window's own source-defined region. Speed-mismatched windows lie outside high-error regions and have motion-matched source low-error-region neighbors but no speed match. Other residual combines the remaining neighbor categories.

Table 17: Prediction proximity to source and target velocities.

			Closer reference	
Target	Model	N	Source	Target
KITTI	LiLY	2,966	2,966 (100.00%)	0 (0.00%)
KITTI	Tartan	3,481	3,481 (100.00%)	0 (0.00%)
OxIOD	LiLY	8,884	8,453 (95.15%)	431 (4.85%)
OxIOD	Tartan	8,906	8,897 (99.90%)	9 (0.10%)

N counts motion-matched, speed-mismatched covered failures. The final two columns report counts (percentages). No equal-distance ties occurred.

Table 18 lists the source-calibrated thresholds used in these matching analyses. Speed bins are selected by ground-truth speed magnitude. Motion distances use standardized five-dimensional features.

Table 18: Motion and speed matching thresholds.

Model	Source carrier	Target	Speed interval (m/s)	δ_Φ	δ_v (m/s)
LiLY	Car	KITTI	All speeds	0.179687	0.070305
LiLY	Drone	KAIST	All speeds	0.348488	0.686769
LiLY	Human	OxIOD	$[0, 0.028847)$	0.051775	0.006016
LiLY	Human	OxIOD	$[0.028847, 0.111200)$	0.086057	0.023496
LiLY	Human	OxIOD	$[0.111200, \infty)$	0.230005	0.039174
LiLY	Dog	QuadSLAM	$[0, 0.006271)$	0.006351	0.002219
LiLY	Dog	QuadSLAM	$[0.006271, 0.209130)$	0.081526	0.016962
LiLY	Dog	QuadSLAM	$[0.209130, \infty)$	0.117969	0.083031
Tartan	Car	KITTI	$[0, 0.170048)$	0.195533	0.036827
Tartan	Car	KITTI	$[0.170048, 0.428544)$	0.184390	0.058643
Tartan	Car	KITTI	$[0.428544, 0.826540)$	0.134890	0.207082
Tartan	Car	KITTI	$[0.826540, \infty)$	0.436365	0.050321
Tartan	Human	OxIOD	$[0, 0.028847)$	0.050078	0.006591
Tartan	Human	OxIOD	$[0.028847, 0.111200)$	0.085018	0.024475
Tartan	Human	OxIOD	$[0.111200, \infty)$	0.253551	0.015919

Values are rounded to six decimal places; matching uses the stored full-precision values. A global threshold applies when the source calibration selects no speed-specific thresholds. TartanIMU rows cover the two carrier references used in the velocity-proximity analysis.

E Internal-State and Target-Domain Adaptation Analysis

E.1 Readout recovery protocol

The 17 OxIOD sessions are divided into five groups. In each fold, two calibration sessions are selected from the non-test groups. Eligible sessions contain at least 800 windows. For each enriched condition, sessions are ranked by the fraction of their original failures belonging to the corresponding category, and the two highest-ranked sessions are selected. The random/mixed condition selects two eligible sessions at random. We uniformly sample 800 windows without replacement from all windows in each selected session, yielding 1,600 calibration windows per condition, including failures and non-failures.

The readout uses the frozen 20-dimensional LiLY latent vector (128-dimensional Tartan features in the additional experiment). Features are standardized using calibration statistics, and an affine ridge readout with regularization parameter $\alpha = 1$ is fitted to ground-truth velocity. Failure labels, source-region labels, and the SourceVal-Q90 threshold remain fixed. The three reported categories comprise source-sparse failures, covered failures whose own regions are high-error, and covered failures whose own regions are low-error. Covered failures in unconfirmed regions are recorded separately. Each window is evaluated once on a held-out session; recovery rates pool windows within each category. The all-windows sensitivity analysis uses all windows from the same selected sessions.

E.2 Recovery by failure category

Table 19 reports the OxIOD results for the fixed-budget protocol in Appendix E.1. We use C_1 for source-sparse failures, C_2 for covered failures in source high-error regions, and C_3 for covered failures in source low-error regions. The region categories are assigned to each target window from its own motion representation.

Table 19: Readout recovery on OxIOD.

Calibration condition	C_1	C_2	C_3
LiLY: $N_1 = 7,174$, $N_2 = 13,179$, $N_3 = 9,958$			
Sparse-rich	**4,210 (58.7)**	3,653 (27.7)	6,903 (69.3)
High-error-region-rich	29 (0.4)	**6,081 (46.1)**	103 (1.0)
Low-error-region-rich	2,078 (29.0)	1,892 (14.4)	**9,574 (96.1)**
Random/mixed	1,899 (26.5)	3,302 (25.1)	7,732 (77.6)
Tartan: $N_1 = 7,236$, $N_2 = 13,788$, $N_3 = 9,518$			
Sparse-rich	**2,322 (32.1)**	2,865 (20.8)	3,025 (31.8)
High-error-region-rich	98 (1.4)	**4,138 (30.0)**	14 (0.1)
Low-error-region-rich	1,029 (14.2)	1,995 (14.5)	**8,546 (89.8)**
Random/mixed	764 (10.6)	2,593 (18.8)	5,487 (57.6)

N_c is the number of original failures in category C_c. Entries give recovered window counts and recovery percentages in parentheses. Recovery means that the held-out readout error is at or below the fixed, model-specific SourceVal-Q90 threshold. Each condition uses 1,600 calibration windows per fold. Bold entries mark the highest recovery rate in each model–category pair.

For both models, each category has its highest pooled recovery rate under calibration enriched in that category. In C_3, low-error-region-rich calibration recovers 9,574 of 9,958 LiLY failures (96.1%) and 8,546 of 9,518 Tartan failures (89.8%), compared with 77.6% and 57.6% under random/mixed calibration. Using all windows from the same selected sessions preserves the best condition for each category; the corresponding $C_1/C_2/C_3$ recovery rates are 62.8/45.3/96.4% for LiLY and 36.0/29.3/90.1% for Tartan.

E.3 Error changes across categories

Table 20 reports paired error changes and newly introduced failures. For category C_c, the error change is $\Delta e_c = 100 \text{ median}_{i \in C_c}(e_i - e_i^{\text{cal}})$ in percentage points, where e_i^{cal} is the held-out readout error. Positive values indicate lower error after readout fitting. New failures are original non-failures whose readout error exceeds the same fixed threshold.

Table 20: Error changes and new failures on OxIOD.

	Median paired reduction (pp)			
Calibration condition	C_1	C_2	C_3	New failures
LiLY: 1,754 original non-failures				
Sparse-rich	+58.4	-45.2	+53.6	1,518 (86.5)
High-error-region-rich	-46.4	+33.6	+0.0	1,062 (60.5)
Low-error-region-rich	-311.1	-393.7	+71.4	1,489 (84.9)
Random/mixed	-34.7	-35.0	+65.6	1,552 (88.5)
Tartan: 1,482 original non-failures				
Sparse-rich	+47.7	+6.9	+47.8	1,099 (74.2)
High-error-region-rich	-49.0	+14.9	-0.4	937 (63.2)
Low-error-region-rich	-46.0	-14.8	+73.8	1,397 (94.3)
Random/mixed	-3.1	+11.1	+63.3	1,343 (90.6)

Error reductions use the original category counts in Table 19. New failures are reported as counts, with percentages of original non-failures in parentheses. All values pool the same held-out predictions used for the recovery rates.

With low-error-region-rich calibration, LiLY reduces C_3 error by a median of 71.4 percentage points while increasing C_1 and C_2 errors by 311.1 and 393.7 points. The same readout turns 84.9% of its original non-failures into failures. The recovery gains therefore depend on the evaluated category and coexist with deterioration elsewhere.

F Additional Failure Detection Results

F.1 Detector features and fitting

For LiLY, the internal feature vector concatenates the frozen encoder factors $[\mathbf{z}_c, \mathbf{z}_s, \mathbf{z}_d] \in \mathbb{R}^{20}$ with dimensions 4, 4, and 12, the final hidden activations $\mathbf{h} \in \mathbb{R}^{512}$, and four head-response summaries. The hidden vector concatenates the 256-dimensional second-ReLU outputs of the horizontal and vertical velocity branches, immediately before their final linear layers. Let W be the combined velocity-output matrix, including its output scaling, and let B have orthonormal columns spanning the row space of W. The four summaries are

$$\mathbf{u} = \left[\|\mathbf{h}\|_2,\ \sum_j \mathbb{I}(h_j > 0),\ \frac{\|B^\top \mathbf{h}\|_2}{\max(\|\mathbf{h}\|_2, 10^{-20})},\ \frac{\|W\mathbf{h}\|_2}{\max(\|\mathbf{h}\|_2, 10^{-20})} \right]. \tag{18}$$

Tartan uses its frozen 128-dimensional features entering the velocity head.

In each outer leave-one-session-out fold, the remaining sessions form the training set. Motion-only and Internal-only each fit a StandardScaler followed by unweighted logistic regression with $C = 1$, at most 3,000 iterations, and tolerance 10^{-6}. Each scaler is fitted only to the corresponding training subset. Within the outer training set, GroupKFold with $\min(4, K)$ folds, where K is the number of training sessions, generates out-of-fold base probabilities. Their means define the score centering, and an unscaled logistic regression with $C = 1$ and tolerance 10^{-8} fits the two centered scores and their product. The base models are refitted on the complete outer training set before predicting the held-out session. Inner folds with one training class use that class's empirical probability; outer folds with only one training class are marked unavailable.

At the 10% calibration-FPR operating point, the threshold is the empirical 90th percentile of non-failure calibration scores, and a score strictly above it predicts failure. Base-detector calibration

scores are inner-fold out-of-fold predictions. Combined calibration scores are predictions from the fitted fusion model on the base out-of-fold features. Each threshold is then applied to the held-out session, for which recall and actual FPR are reported. Rank metrics pool the held-out probabilities. Confidence intervals use 1,000 resamples of whole sessions with fixed fitted predictions and shared resamples across detectors.

F.2 Detection results across models and targets

Table 21 reports pooled LOSO results for both models and all four targets. Failure labels use $v_{\text{min}} = 1\,\text{m}/\text{s}$ and the corresponding SourceVal 90th percentile. Each target window is evaluated once on a held-out session.

Table 21: LOSO failure-detection performance.

Model	Target	Detector	Failures / N	AUROC [95% CI]	AUPRC [95% CI]
LiLY	KITTI	Motion	3,576 / 4,232	0.971 [0.957, 0.983]	0.993 [0.989, 0.996]
		Internal		0.929 [0.903, 0.952]	0.983 [0.970, 0.990]
		Combined		0.978 [0.967, 0.988]	0.994 [0.990, 0.997]
LiLY	KAIST	Motion	8 / 678	0.649 [0.403, 0.898]	0.080 [0.011, 0.258]
		Internal		0.987 [0.976, 0.998]	0.401 [0.250, 0.902]
		Combined		0.579 [0.355, 0.760]	0.089 [0.008, 0.349]
LiLY	OxIOD	Motion	31,112 / 32,866	0.841 [0.763, 0.896]	0.985 [0.974, 0.991]
		Internal		0.803 [0.760, 0.848]	0.982 [0.975, 0.988]
		Combined		0.871 [0.804, 0.920]	0.987 [0.978, 0.993]
LiLY	QuadSLAM	Motion	777 / 1,168	0.670 [0.631, 0.712]	0.770 [0.686, 0.832]
		Internal		0.772 [0.736, 0.806]	0.842 [0.785, 0.880]
		Combined		0.791 [0.763, 0.820]	0.846 [0.774, 0.887]
Tartan	KITTI	Motion	3,651 / 4,232	0.994 [0.987, 0.997]	0.999 [0.997, 1.000]
		Internal		0.982 [0.970, 0.991]	0.996 [0.992, 0.999]
		Combined		0.991 [0.981, 0.998]	0.997 [0.993, 1.000]
Tartan	KAIST	Motion	3 / 678	0.185 [0.002, 0.398]	0.003 [0.002, 0.007]
		Internal		0.961 [0.904, 1.000]	0.378 [0.021, 1.000]
		Combined		0.129 [0.025, 0.282]	0.003 [0.002, 0.008]
Tartan	OxIOD	Motion	31,384 / 32,866	0.888 [0.848, 0.922]	0.989 [0.984, 0.993]
		Internal		0.933 [0.903, 0.958]	0.995 [0.992, 0.998]
		Combined		0.929 [0.898, 0.957]	0.994 [0.991, 0.997]
Tartan	QuadSLAM	Motion	952 / 1,168	0.702 [0.614, 0.838]	0.861 [0.725, 0.956]
		Internal		0.900 [0.840, 0.941]	0.971 [0.958, 0.981]
		Combined		0.887 [0.837, 0.925]	0.954 [0.935, 0.973]

AUPRC denotes average precision. Intervals are the 2.5th and 97.5th percentiles of 1,000 whole-session bootstrap resamples of the fixed held-out predictions, shared across detectors. Resamples missing a required class are omitted: 969 are valid for Tartan on KAIST and 1,000 for the other model–target pairs.

Table 22 reports recall and actual test FPR at thresholds calibrated for a 10% FPR on the training sessions.

Table 22: Failure detection at 10% calibration FPR.

Model	Target	Detector	Recall (%) [95% CI]	Test FPR (%) [95% CI]
LiLY	KITTI	Motion	95.4 [91.4, 97.5]	10.4 [5.3, 16.1]
		Internal	72.7 [64.0, 78.9]	11.1 [4.3, 19.9]
		Combined	96.8 [94.1, 98.3]	10.4 [5.7, 15.5]
LiLY	KAIST	Motion	12.5 [0.0, 50.0]	11.8 [0.6, 28.5]
		Internal	100.0 [100.0, 100.0]	7.5 [3.9, 10.5]
		Combined	100.0 [100.0, 100.0]	12.8 [1.9, 28.4]
LiLY	OxIOD	Motion	44.4 [24.7, 63.3]	9.9 [5.6, 14.1]
		Internal	30.7 [24.8, 36.8]	8.0 [6.1, 9.9]
		Combined	55.1 [39.4, 69.9]	8.8 [6.8, 11.0]
LiLY	QuadSLAM	Motion	22.7 [18.6, 27.0]	9.2 [7.6, 10.8]
		Internal	28.3 [17.4, 36.0]	7.4 [4.2, 11.1]
		Combined	34.6 [29.8, 39.3]	9.7 [7.7, 11.4]
Tartan	KITTI	Motion	99.0 [97.8, 99.6]	10.7 [4.7, 18.0]
		Internal	98.6 [97.2, 99.4]	10.3 [6.0, 15.4]
		Combined	99.7 [99.2, 99.9]	10.2 [5.2, 16.3]
Tartan	KAIST	Motion	0.0 [0.0, 0.0]	15.0 [3.4, 26.2]
		Internal	100.0 [100.0, 100.0]	10.7 [5.9, 14.8]
		Combined	33.3 [0.0, 100.0]	7.6 [2.9, 13.9]
Tartan	OxIOD	Motion	60.0 [43.6, 74.8]	10.2 [6.1, 13.7]
		Internal	75.5 [70.0, 80.7]	9.6 [5.4, 13.5]
		Combined	78.3 [65.0, 88.8]	10.3 [5.3, 15.1]
Tartan	QuadSLAM	Motion	10.6 [7.1, 13.8]	6.9 [3.2, 9.5]
		Internal	63.3 [56.1, 69.6]	8.8 [4.2, 16.1]
		Combined	63.9 [54.9, 73.1]	8.8 [4.0, 16.1]

Thresholds are selected within each training split and applied to its held-out session. Counts are the same as in Table 21. Intervals use the same whole-session resamples. Tartan–KAIST recall uses 969 valid resamples; all other recall intervals and all FPR intervals use 1,000.