Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

2nd Workshop on Explainable AI for Space - EASi

within IJCAI-ECAI 2026

AIRBUS

Attribution and Uncertainty Behavior of Learned Residual Gyro Correction for Gyro-Stellar Estimation

Mariela De Lucas Alvarez[1], Melvin Laux[1], Arthur de Freitas Precht[2], Maurice Martin[2], Edoardo Caroselli[2], Frank Kirchner[1], Alexander Fabisch[1]

Motivation & Question

Learned gyro-residual correction improves attitude estimation, but its behavior must be interpretable before safety-critical integration. We ask:

- When is the model reliable or uncertain?
- Do aleatoric and epistemic uncertainty respond differently under OOD sensor effects?
- Which sensors and time regions drive correction and uncertainty?

Model & Evaluation

A 1D CNN predicts gyro-residual correction from short gyro, STR1, and STR2 windows. An aleatoric head outputs the correction μ and variance σ^2, while a five-seed ensemble estimates epistemic uncertainty. Evaluation uses nominal ID logs and structured OOD perturbations: AR(1) correlated noise, vibration, drift and dropout.

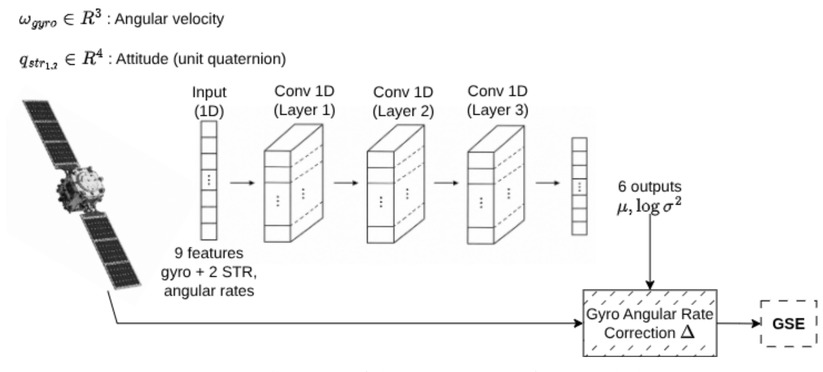


Figure 1: CNN residual-correction model.

Residual Correction

Table 1: Residual magnitude statistics (deg/s).

Metric	Raw	Naive	CNN	Improvement
Mean	6.61e-3	3.41e-3	6.22e-4	90.6%
RMSE	7.33e-3	4.95e-3	8.44e-4	88.5%
Median	5.91e-3	2.20e-3	4.62e-4	92.2%
p95	1.21e-2	1.03e-2	1.66e-3	86.2%

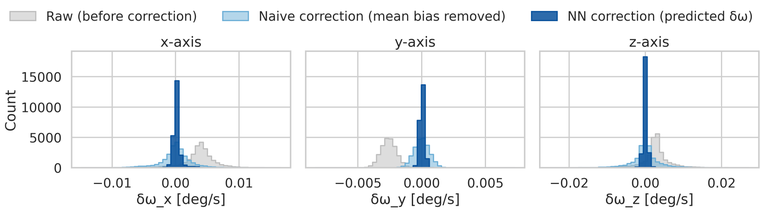
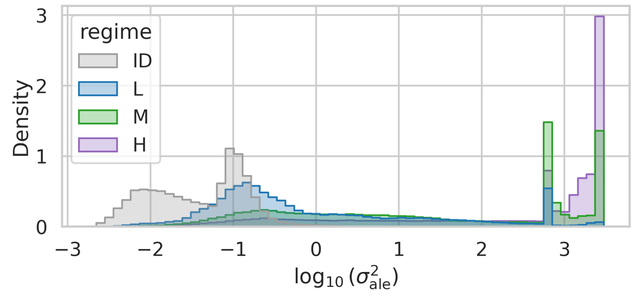


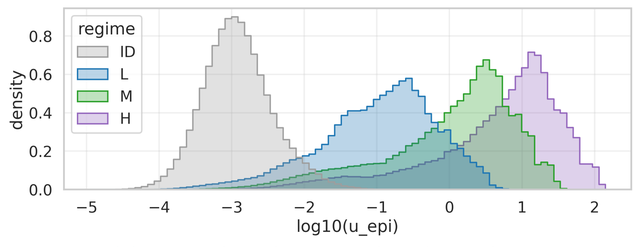
Figure 2: Residual distribution per-axis.

The learned correction reduces gyro residual error substantially compared with the raw gyro and naive baseline.

UQ Behavior - ID/ODD



(a) Predicted aleatoric variance $\log \sigma^2$.

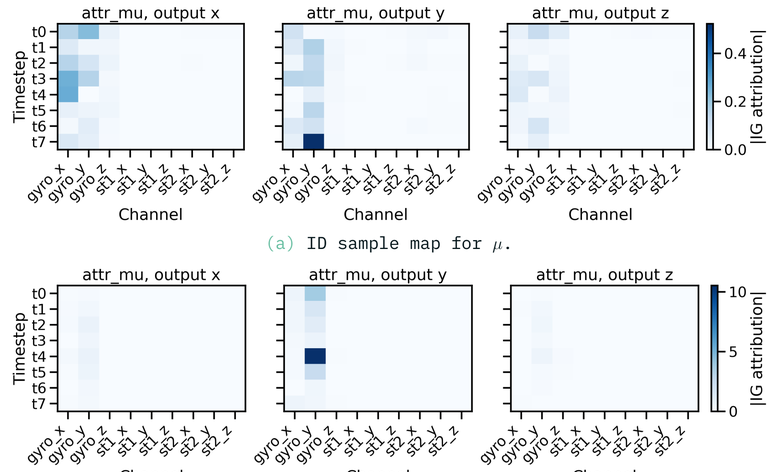


(b) $\log_{10}$ epistemic uncertainty by regime.

Figure 3: Aleatoric & epistemic uncertainty across regimes.

OOD perturbations generally increase uncertainty, but ID/OOD overlap shows that magnitude alone is insufficient for regime separation.

Representative sample IG



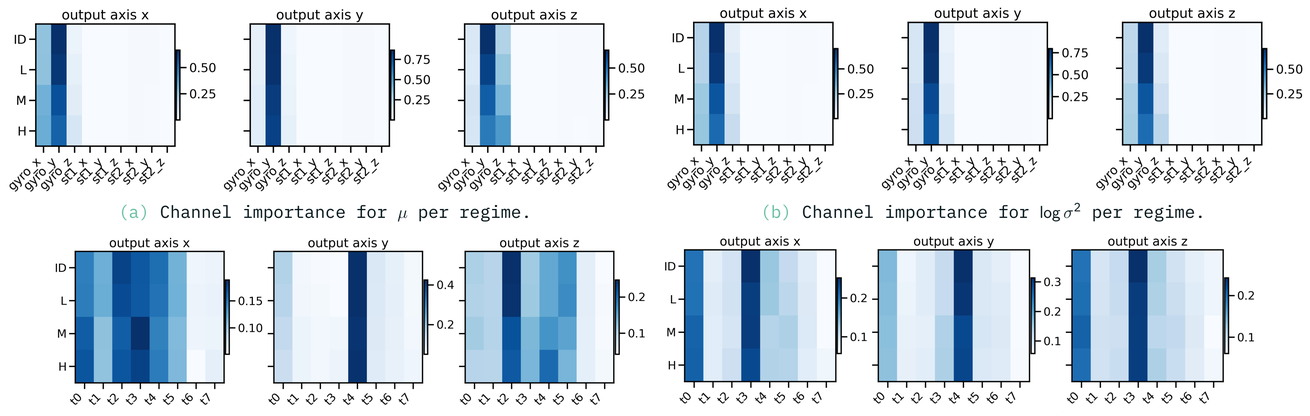
(a) ID sample map for μ.

(b) OOD-M sample map for μ.

Figure 4: Single-sample attribution.

Single-window IG shows which local sensor channels drive one correction and uncertainty estimate.

Attribution Behavior



(a) Channel importance for μ per regime.

(b) Channel importance for $\log \sigma^2$ per regime.

(c) Temporal importance for μ per regime.

(d) Temporal importance for $\log \sigma^2$ per regime.

Figure 5: Attribution separates correction drivers from uncertainty drivers

Correction and uncertainty rely on different channels and time regions, indicating that uncertainty is not merely a by-product of prediction error.

Takeaway

- Learned residual correction improves gyro error, but trustworthy integration requires understanding model behavior under perturbations.
- UQ indicates when reliability changes; attribution helps explain why by identifying the sensor channels and time regions driving correction and uncertainty.
- Together, they provide a diagnostic layer for learned components in hybrid attitude-estimation pipelines.

With funding from the: Federal Ministry of Research, Technology and Space

This work was funded by the European Aerospace Agency under the GSTP programme, activity GT1I-602SA "Artificial Intelligence techniques for spacecraft attitude control and estimation" (project acronym: AI4AOCS), contract number 4000145154/24/NL/MGu, lead by Airbus Defence and Space GmbH. This work was partially supported by the German Federal Ministry of Research, Technology and Space (BMFTR) under the Robotics Institute Germany (RIG).

Contact:
Mariela De Lucas Alvarez,
ORCID: 0000-0003-0846-4507,
mariela.de_lucas_alvarez@dfki.de

[1] Robotics Innovation Center - DFKI GmbH, Bremen, Germany