# On Equivariant Model Selection through the Lens of Uncertainty

Putri van der Linden, Alexander Timans, Dharmesh Tailor, Erik Bekkers

*Amsterdam Machine Learning Lab (AMLab), University of Amsterdam*

TPM @ uai2025

## Motivation

A growing number of equivariant and non-equivariant models are being developed which encode geometric inductive biases to various extents.

We want a *lightweight*, *post-hoc* way to compare model classes and assess the fit of their inductive biases with minimal alterations. Can uncertainty help?

## Uncertainty-Guided Model Selection

*Tasks.* Classification on ModelNet40 and Regression on QM9.

*Model Catalogue.* Four variants of *Rapidash* [1]:

- ***Invariant*** — invariant message passing
- ***Equivariant*** — equivariant layers
- ***Augment*** — SO(3)-augmented training
- ***Plain*** — fully unconstrained

**Which one to choose** ?

*Metrics.* We compare frequentist, Bayesian, and calibration-based measures to naive error-based evaluation*:*

- Conformal prediction interval size
- ECE & Brier score (calibration)
- Test data likelihood
- Bayesian marginal likelihood

Recall *Bayesian model selection:* $p(D\,|\,M) \;=\; \int p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; d\theta$

We use *Last Layer Laplace* [2] as a lightweight model-agnostic option to compute the marginal likelihood for pretrained models. Then we obtain

$$\log p(D\,|\,M) \approx \underbrace{\log p(D\,|\,\theta_*, M)}_{\text{Data fit}} - \underbrace{\left[\frac{1}{2}\log\left|\frac{1}{2\pi}\mathbf{H}_*\right| - \log p(\theta_*\,|\,M)\right]}_{\text{Model complexity}}$$

## Experiments

| | | Train data $D_{train}$ | | | | Test data $D_{test}$ | |
|---|---|---|---|---|---|---|---|
| **Target** | **Model** $M$ | **MAE ↓** | **LogLik ↑** | **Complexity ↓** | **Log- MargLik ↑** | **MAE ↓** | **LogLik ↑** |
| $\mu$ | *Invariant* | 0.0025 | -101084 | 767 | -101851 | 0.0204 | 22064 |
| | *Equivariant* | 0.0083 | -101091 | 723 | **-101814** | **0.0145** | **22940** |
| | *Augment* | 0.0048 | -101086 | 799 | -101886 | 0.0254 | 20826 |
| | *Plain* | 0.0038 | -101086 | 798 | -101884 | 0.0296 | 19622 |
| $\alpha$ | *Invariant* | 0.0102 | -101097 | 1530 | **-102628** | 0.0613 | 768 |
| | *Equivariant* | 0.0290 | -101176 | 1515 | -102691 | **0.0522** | **9014** |
| | *Augment* | 0.0153 | -101112 | 1521 | -102633 | 0.0679 | -3732 |
| | *Plain* | 0.0106 | -101100 | 1564 | -102664 | 0.0888 | -19273 |
| $\epsilon_{HOMO}$ | *Invariant* | 0.2540 | -101083 | 1211 | -102295 | 23.4848 | 20586 |
| | *Equivariant* | 2.9681 | -101084 | 1243 | -102327 | **21.3705** | **20989** |
| | *Augment* | 0.7900 | -101083 | 1149 | -102233 | 27.4825 | 19461 |
| | *Plain* | 0.2288 | -101083 | 1148 | **-102231** | 33.6994 | 17578 |

Tab: QM9 results. Predictive error (via the mean absolute error) and data fit via the log-likelihood (LogLik), Bayesian model complexity, and the overall log-marginal likelihood (Log-MargLik)

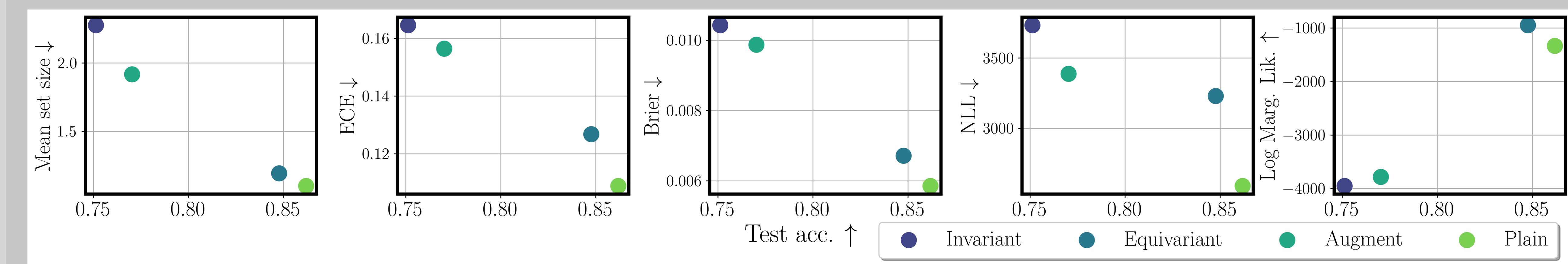


Fig: Uncertainty-based measures on ModelNet40. 'NLL' refers to the negative log-likelihood of the model's direct softmax output, while 'log Marg Lik' refers to the Bayesian notion.

## Outlook

- Frequentist and calibration based metrics align directly with performance
- Naive Marg. Lik. via Laplace selects models inconsistently; does not seem to pick up on last-layer feature differences induced by geometric constraints.
- **How to design flexible priors informed by equivariant representations for symmetry-aware Bayesian model selection ?**

## References

[1] Probing Equivariance and Symmetry Breaking in Convolutional Networks. S. Vadgama, M. M. Islam, D. Buracas, C. Shewmake, A. Moskalev, E. Bekkers. 2025

[2] Laplace Redux -- Effortless Bayesian Deep Learning. E. Daxberger, A. Kristiadi, A. Immer, R. Eschenhagen, M. Bauer, P. Hennig. NeurIPS 2021

AMLAB Amsterdam Machine Learning Lab

University of Amsterdam