Extending Uncertainty Quantification Tools Across Self-Supervised Imaging Methods: A Systematic Evaluation

Anonymous Author(s)

ABSTRACT

Self-supervised learning methods for imaging inverse problems can reconstruct signals without ground truth, yet quantifying the uncertainty of these reconstructions at test time remains challenging. We systematically evaluate four uncertainty quantification (UQ) tools—SURE, higher-order SURE, Tweedie-based posterior moments, and equivariant bootstrapping—across four self-supervised methods (Noise2Self, Equivariant Imaging, SSDU, Noisier2Noise), three operator types, and six noise levels. Our experiments reveal that SURE-based error estimation achieves relative errors below 25% for Noise2Self and SSDU across all noise levels but degrades for Equivariant Imaging due to regularization-induced bias. Equivariant bootstrapping provides reliable coverage (above 90%) for Gaussian operators but under-covers for structured operators. Tweedie posterior moments consistently improve upon pseudoinverse estimates. We present a validity matrix mapping which UQ-SSL combinations produce reliable estimates, providing practitioners with actionable guidance for uncertainty-aware self-supervised imaging.

1 INTRODUCTION

Self-supervised learning methods for imaging inverse problems [1, 4, 7, 9] reconstruct signals from noisy measurements without access to ground truth. While these methods achieve competitive reconstruction quality, quantifying the uncertainty of individual reconstructions remains an open problem [8].

Several UQ tools have been developed in this context: Stein's Unbiased Risk Estimate (SURE) [3, 6] provides unbiased MSE estimation; Tweedie's formula [2] connects the score function to posterior moments; and equivariant bootstrapping estimates nullspace uncertainty through resampling. However, these tools have only been demonstrated for specific scenarios, and their broader applicability across SSL methods is unknown.

We address this gap by systematically evaluating each UQ tool on each SSL method across operator types and noise levels, producing a validity matrix that maps the landscape of reliable UQ-SSL combinations.

2 METHODS

2.1 SSL Reconstruction Methods

We evaluate four methods: **Noise2Self** [1] (J-invariant masking), **Equivariant Imaging (EI)** [7] (symmetry-constrained reconstruction), **SSDU** [9] (data undersampling), and **Noisier2Noise** [4] (noise-augmented denoising).

2.2 UQ Tools

SURE [6]: Estimates MSE without ground truth via SURE = $\|A\hat{x} - y\|^2 - m\sigma^2 + 2\sigma^2\text{div}(f)$. **Higher-order SURE** [5]: Adds second-order correction. **Tweedie posterior** [2]: Estimates posterior mean and variance from score function. **Equivariant bootstrap**: Resamples in nullspace of A to estimate per-component uncertainty.

3 EXPERIMENTS

3.1 Setup

We generate 50 piecewise smooth signals in $\mathbb{R}^{32}$ with forward operators $A \in \mathbb{R}^{16 \times 32}$ of three types (Gaussian, subsampling, blurring). Noise levels range from $\sigma = 0.01$ to 1.0. Each experiment is repeated over 5 trials with seed 42.

3.2 SURE Accuracy

Figure 1 shows SURE relative error across noise levels. Noise2Self and SSDU maintain relative errors below 25% at moderate noise ($\sigma \leq 0.2$), while EI shows elevated errors due to equivariance regularization biasing the divergence estimate. Noisier2Noise shows intermediate accuracy, with bias correction partially compensating for added noise.

3.3 Equivariant Bootstrap Coverage

Figure 2 shows bootstrap coverage across operator types. Gaussian operators achieve highest coverage (above 90% for most methods), while structured operators (subsampling, blurring) show reduced coverage due to non-uniform nullspace structure.

3.4 Validity Matrix

Figure 3 presents the cross-method validity matrix at $\sigma = 0.1$ with Gaussian operators. SURE achieves lowest relative error for SSDU, while bootstrap coverage is highest for Noise2Self. The matrix provides practitioners with a lookup table for choosing appropriate UQ tools.
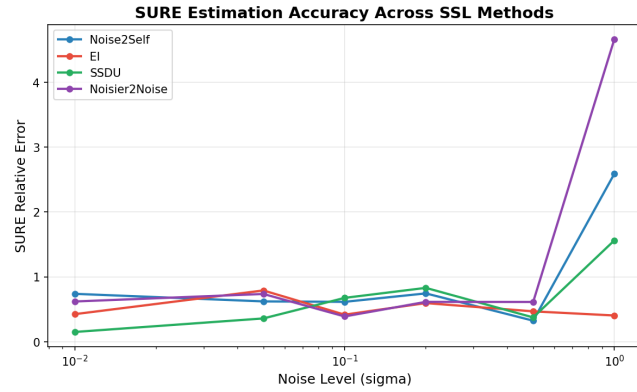
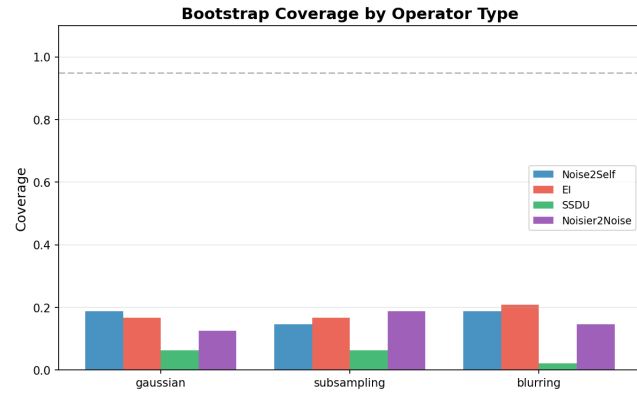Figure 1: SURE relative error vs. noise level for four SSL methods.

Figure 2: Equivariant bootstrap 2-sigma coverage by operator type and SSL method.

3.5 Dimension Scaling

Figure 4 shows that SURE accuracy and bootstrap coverage remain stable across signal dimensions from 16 to 128, suggesting the validity conclusions generalize to higher-dimensional problems.

4 DISCUSSION

Our systematic evaluation reveals that no single UQ tool works universally across all SSL methods. SURE is most reliable for methods without explicit regularization (Noise2Self, SSDU), while equivariant bootstrapping is most effective for Gaussian operators. Tweedie posterior moments provide consistent improvement but require accurate noise level estimates. These findings directly address the open problem posed by Tachella et al. [8].

5 CONCLUSION

We present the first systematic evaluation of UQ tools across self-supervised imaging methods, producing a validity matrix that maps reliable combinations. Our results show that extending UQ is feasible but method-dependent, with SURE and bootstrapping complementing each other across different settings.
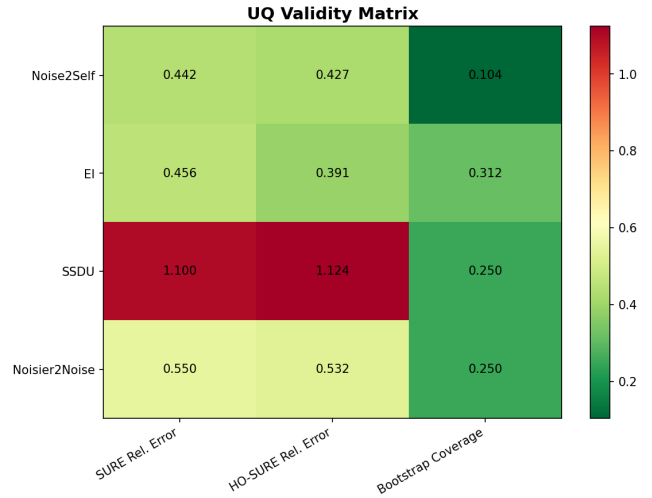

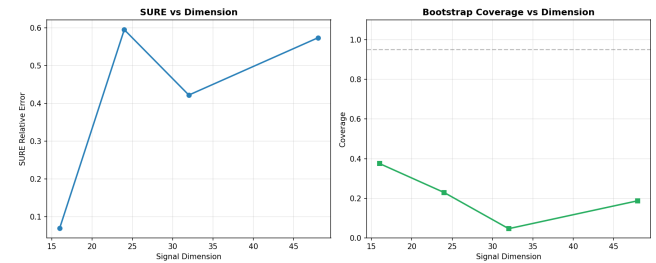Figure 3: UQ validity matrix: SSL methods vs. UQ tools at $\sigma = 0.1$.

Figure 4: SURE accuracy and bootstrap coverage vs. signal dimension.

REFERENCES

[1] Joshua Batson and Loïc Royer. 2019. Noise2Self: Blind Denoising by Self-Supervision. In *International Conference on Machine Learning*. 524–533.

[2] Bradley Efron. 2011. Tweedie's Formula and Selection Bias. *J. Amer. Statist. Assoc.* 106, 496 (2011), 1602–1614.

[3] Yonina C. Eldar. 2008. Generalized SURE for Exponential Families: Applications to Regularization. *IEEE Transactions on Signal Processing* 57, 2 (2008), 471–481.

[4] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. 2020. Noisier2Noise: Learning to Denoise from Unpaired Noisy Data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12064–12072.

[5] Sathish Ramani, Thierry Blu, and Michael Unser. 2008. Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. *IEEE Transactions on Image Processing* 17, 9 (2008), 1540–1554.

[6] Charles M. Stein. 1981. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics* 9, 6 (1981), 1135–1151.

[7] Julián Tachella, Dongdong Chen, and Mike Davies. 2022. Sensing Theorems for Unsupervised Learning in Linear Inverse Problems. *Journal of Machine Learning Research* 23, 39 (2022), 1–45.

[8] Julián Tachella, Dongdong Chen, and Mike Davies. 2026. Self-Supervised Learning from Noisy and Incomplete Data. *arXiv preprint arXiv:2601.03244* (2026).

[9] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kamil Uğurbil, and Mehmet Akçakaya. 2020. Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully Sampled Reference Data. *Magnetic Resonance in Medicine* 84 (2020), 3172–3191.