# Sparsity Promoting Scale Mixture Priors Fit Large Imaging Datasets

ALEXANDER STRANG

UNIVERSITY OF CALIFORNIA, BERKELEY. STATISTICS.

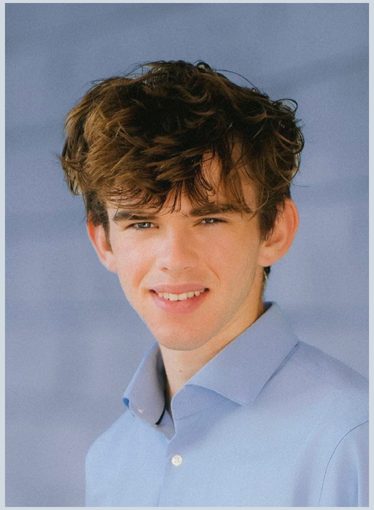
ICERM, BAYESIAN INVERSE PROBLEMS AND UQ (2026)

Brandon Marks
Stanford, MS Stats

Yash Dave
Stanford, MS Stats

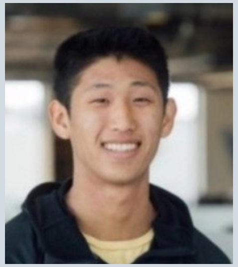
Hannah Chung

Zixun Wang



Michael Murphy

Riya Patwa



Simon Cha

Henry Liev

Ben Rabin

Max Yen

Jiaxin Li

Johan Phan

Enoch Mun

Andres Gutierrez

Sebastian Hoffman

# Introduction

MODEL, PROPERTIES, MOTIVATION

# Hierarchical Model

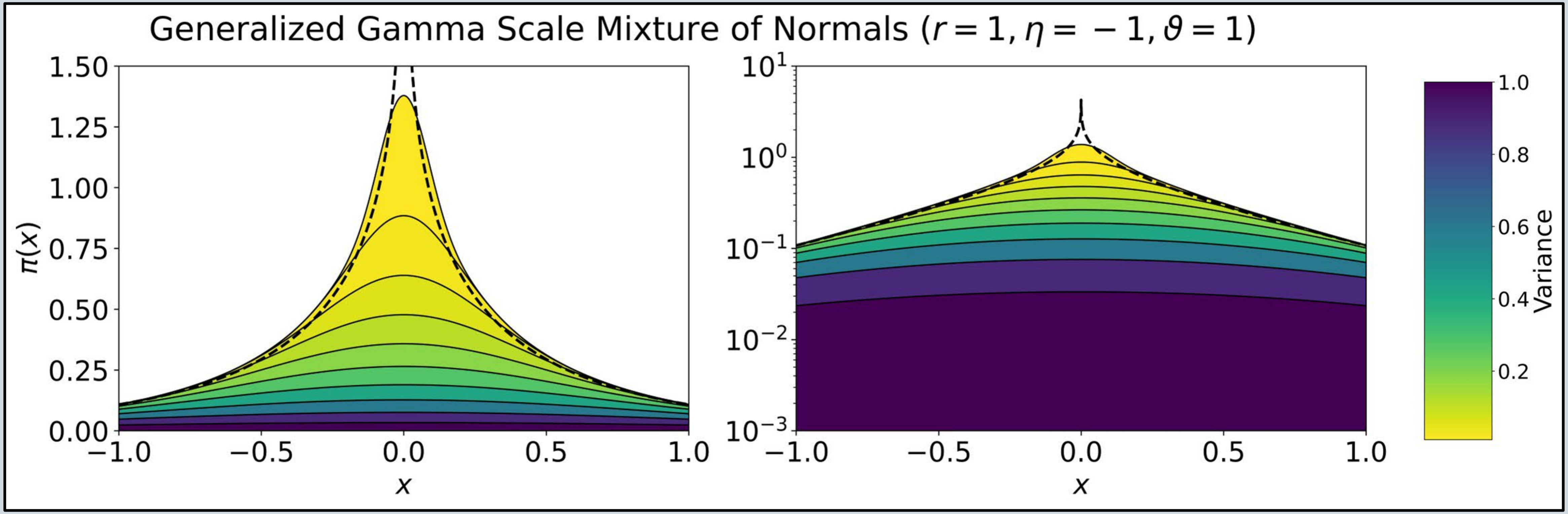
Bayesian inverse problems require a prior model.

We consider the **hierarchical model**:

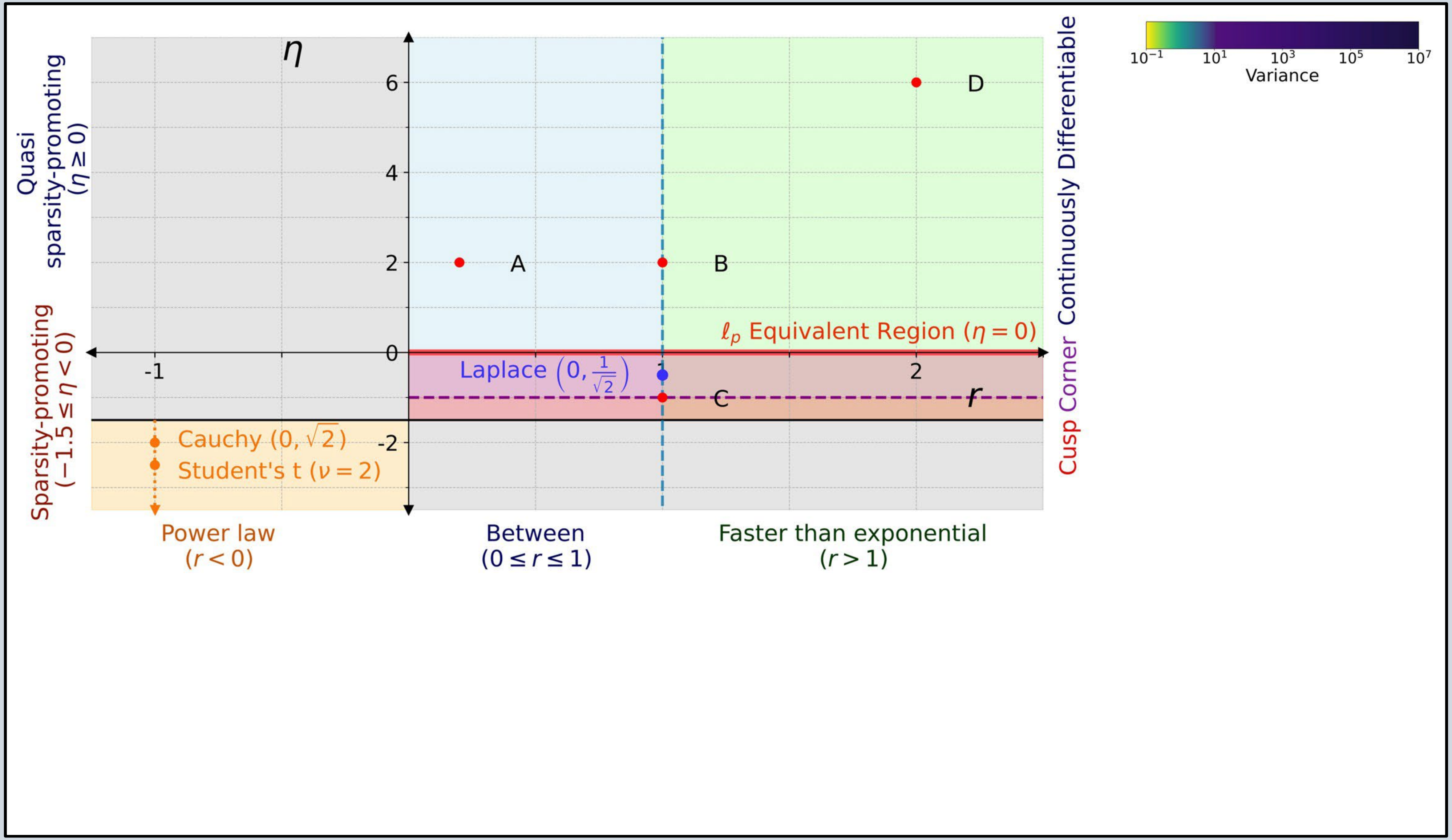
$$\textbf{(1)}\ \Theta_j \sim \text{Generalized Gamma}(r_j, \beta_j, \vartheta_j) \quad \to \quad \textbf{(2)}\ X_j | (\Theta_j = \theta) \sim \text{Normal}(0, \theta^2)$$

This model promotes sparsity when $\eta = r\,\beta\ - \frac{3}{2} \leq 0.$

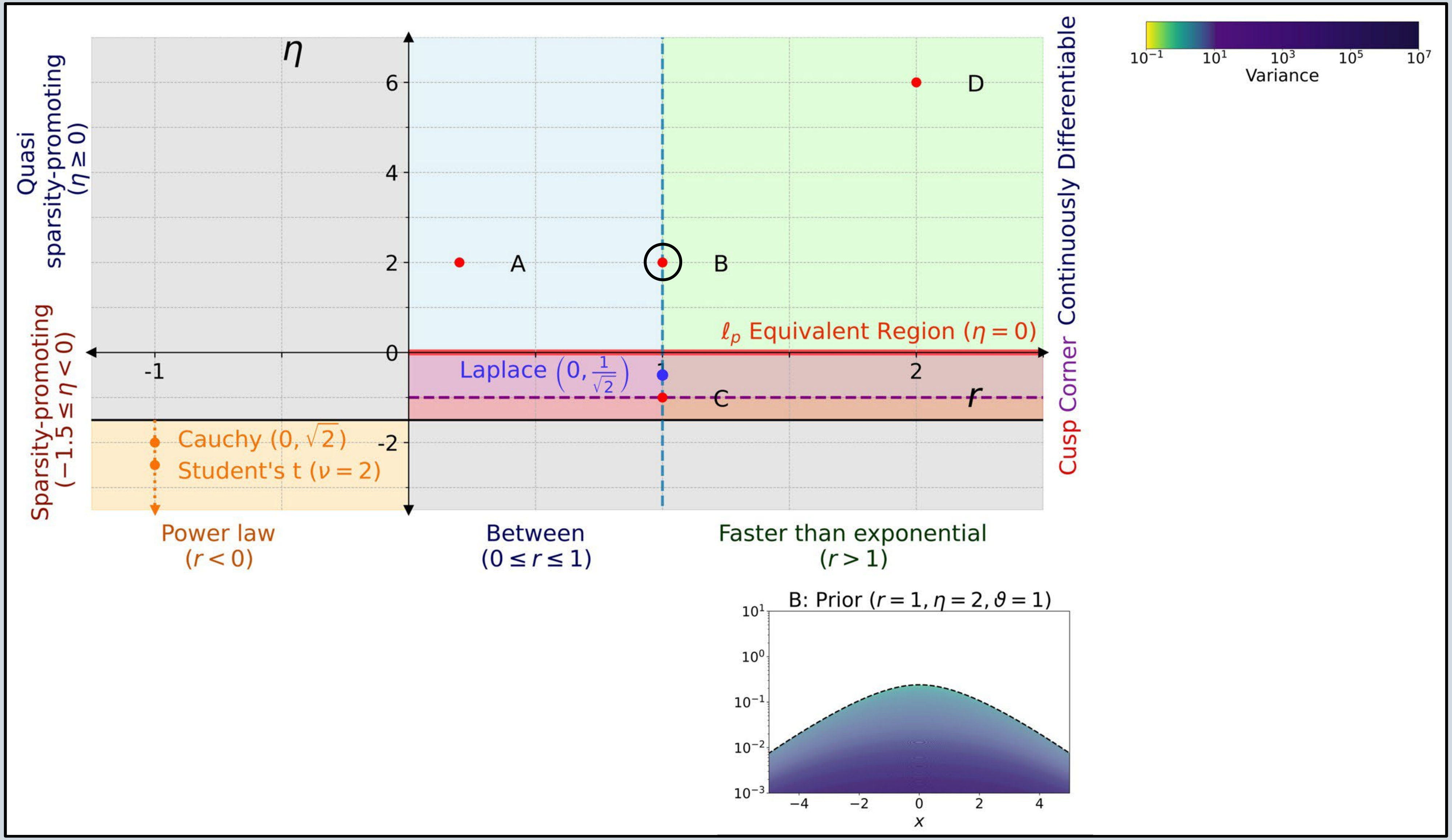
This model has **two shape parameters ($\boldsymbol{r}, \boldsymbol{\eta}$)** and **one scale parameter ($\boldsymbol{\vartheta}$)**



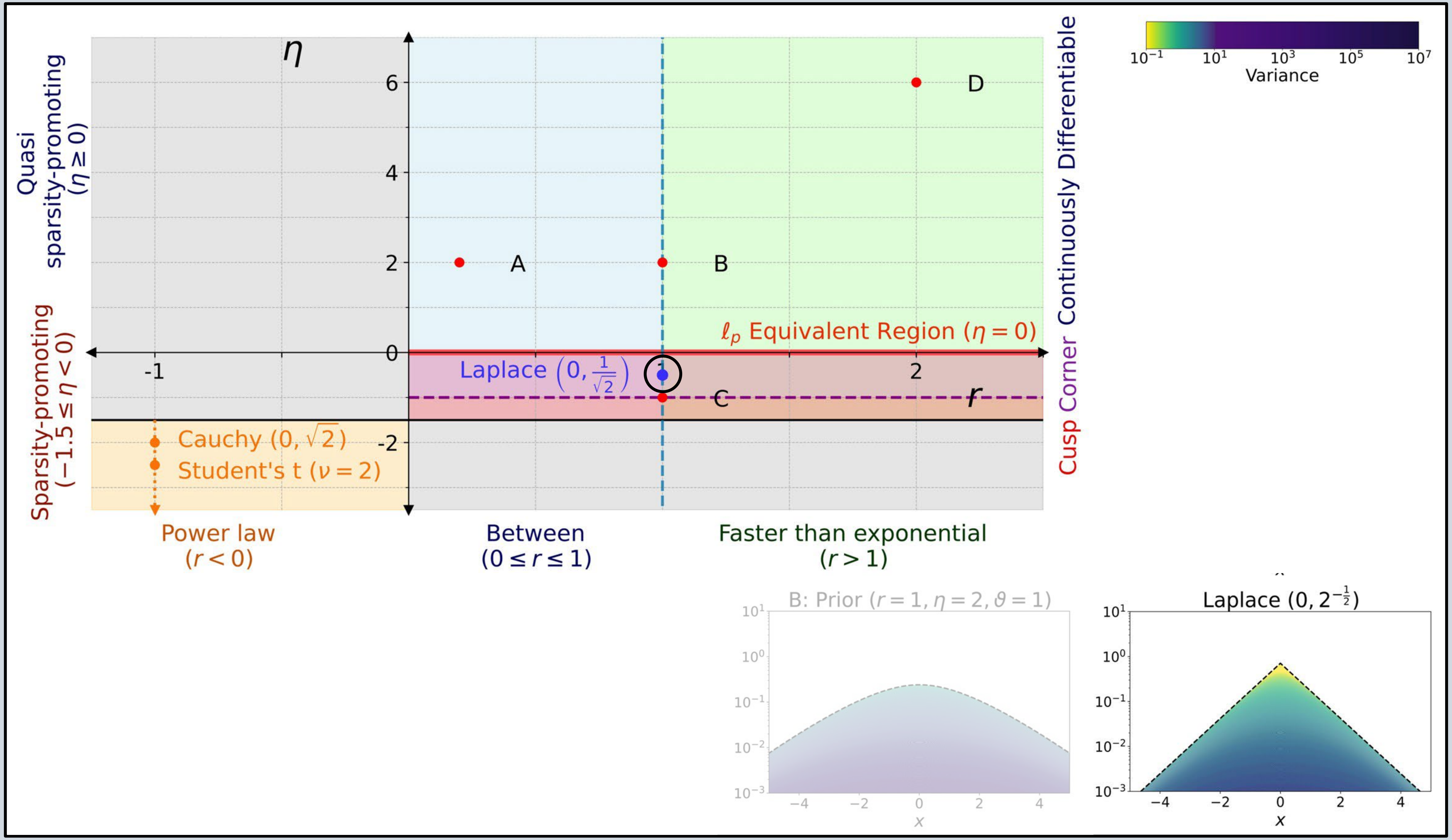

This model is a **generalized-gamma scale mixture of normals.**



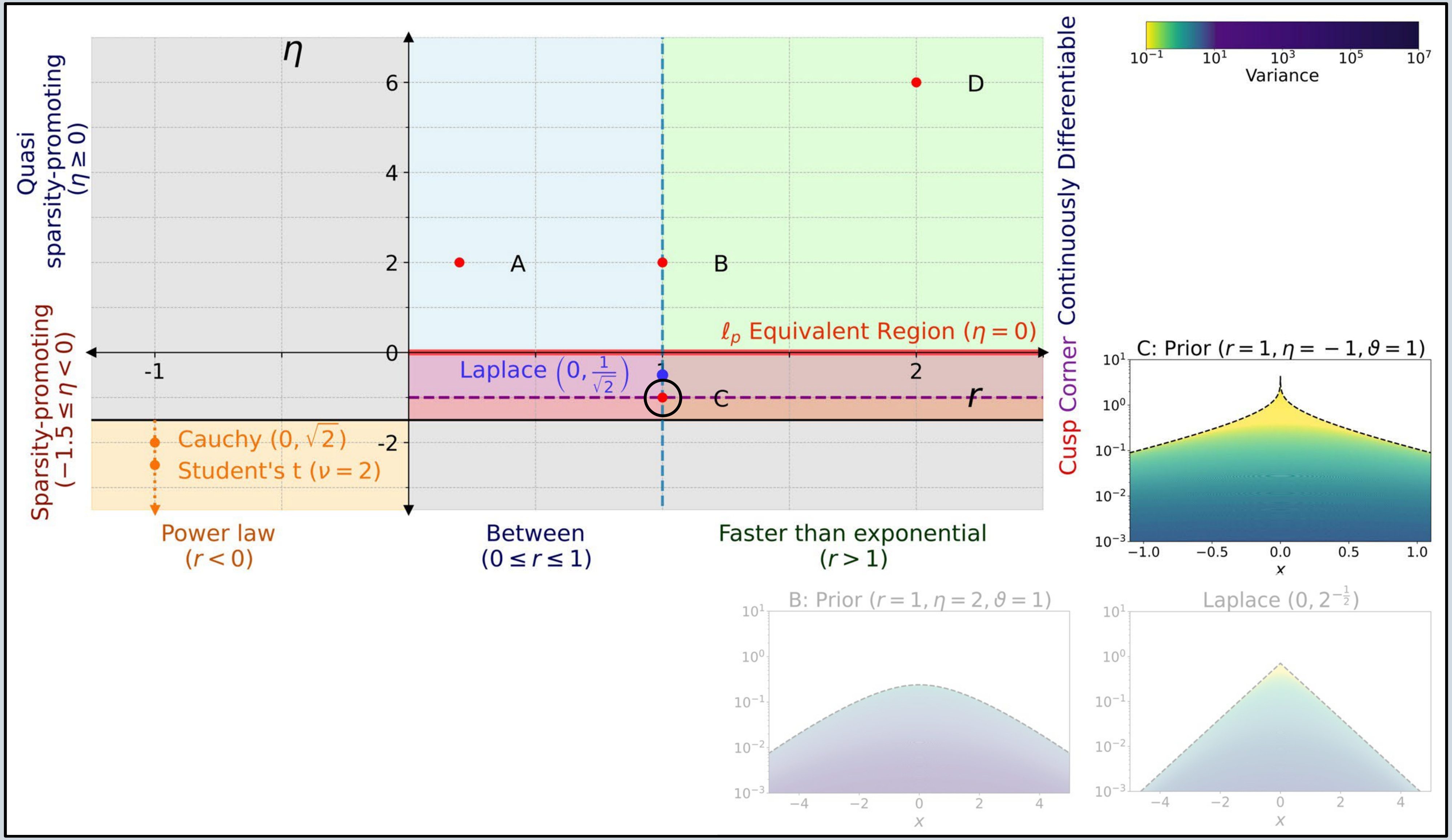

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.



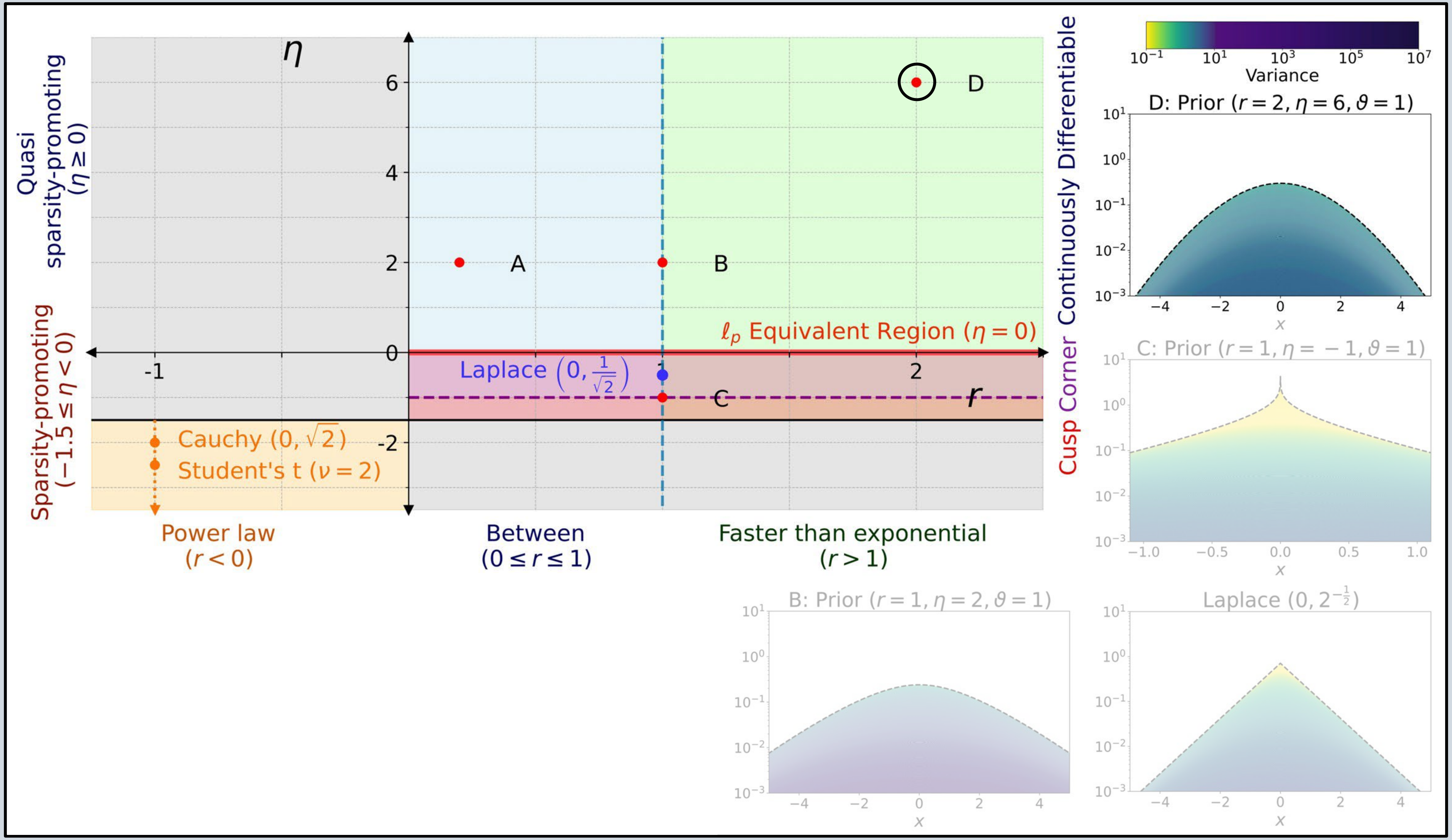

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.



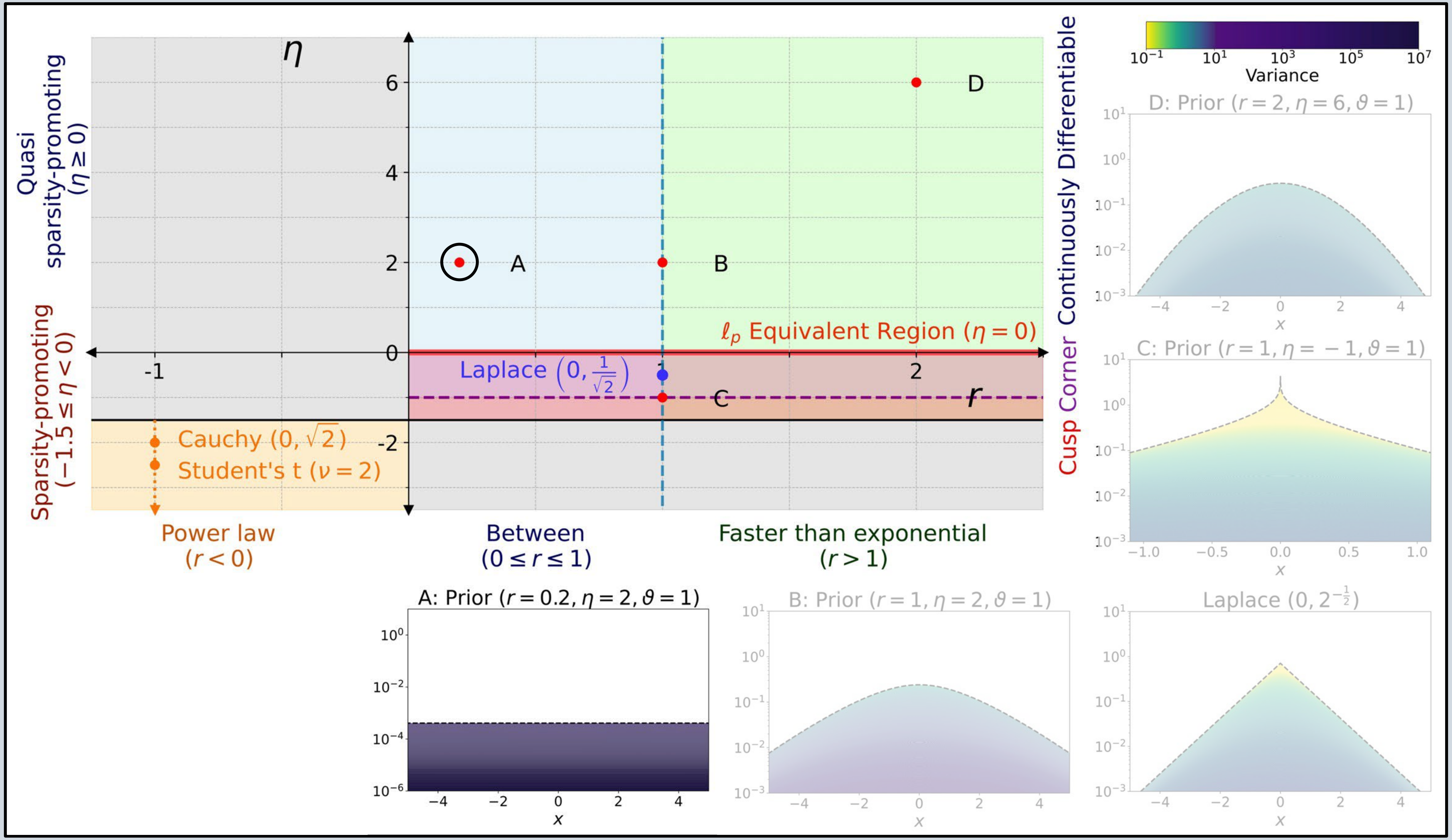

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.



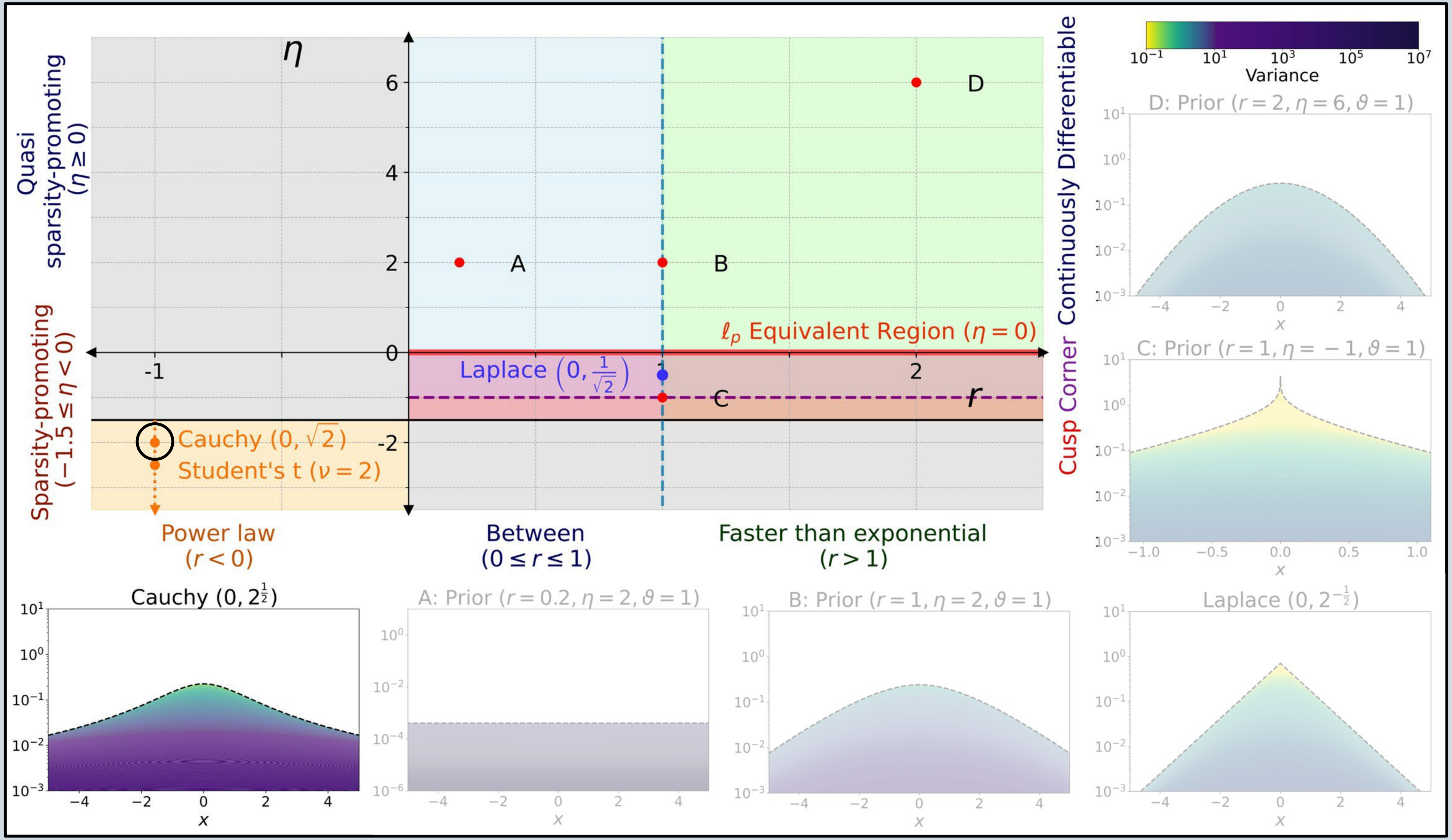

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.



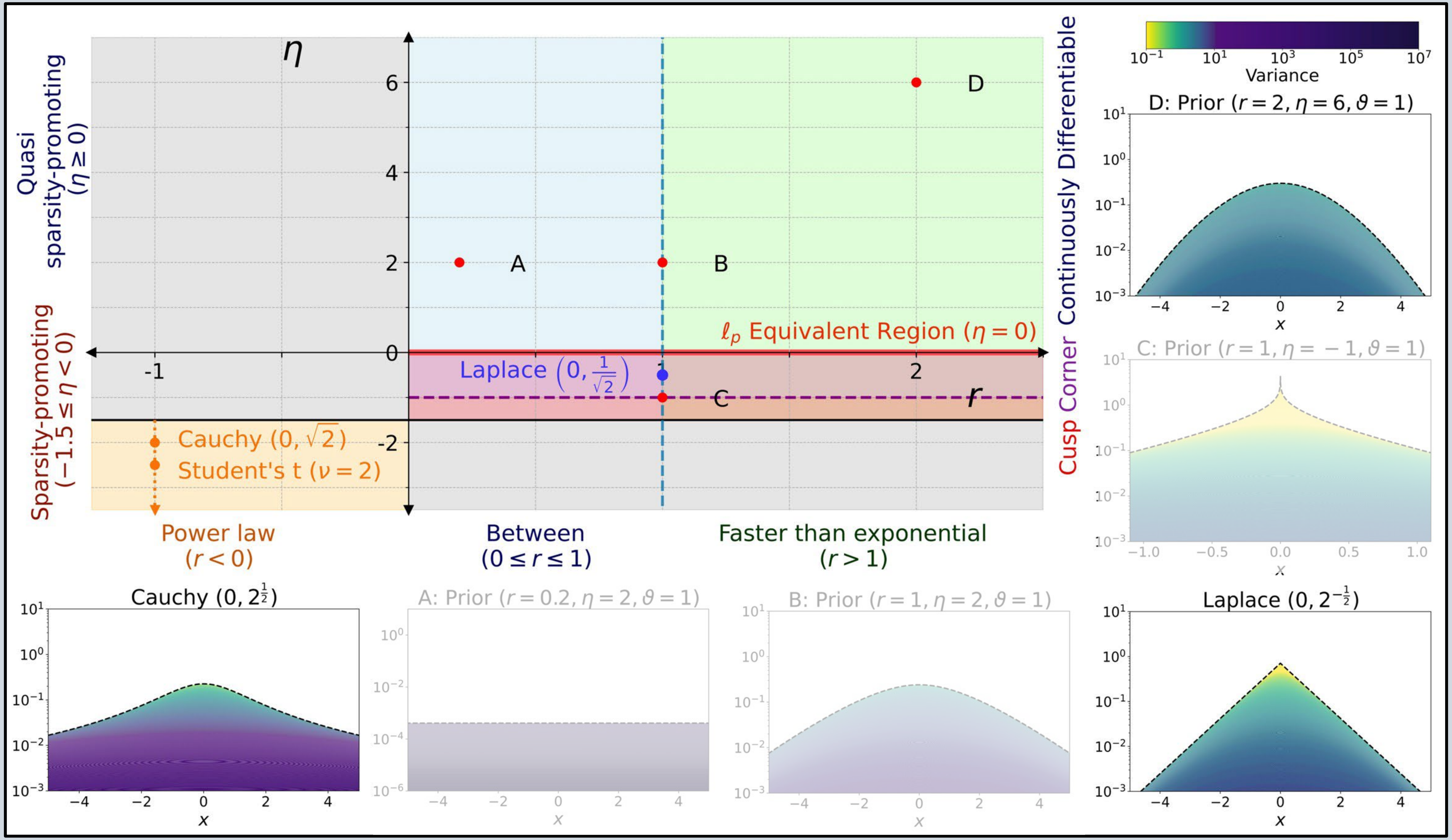

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.



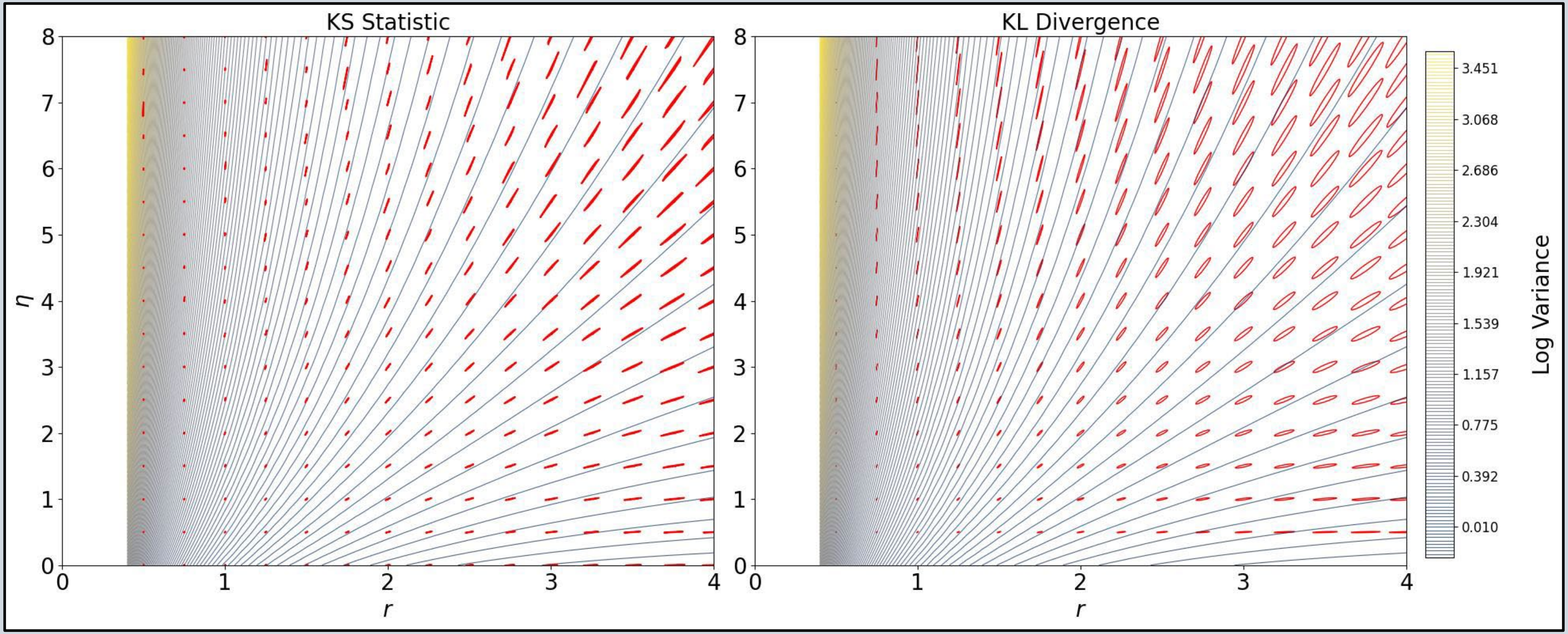

The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.




The shape parameters provide **separate control over the bulk ($\eta$)** and **tails ($r$)** of the prior.

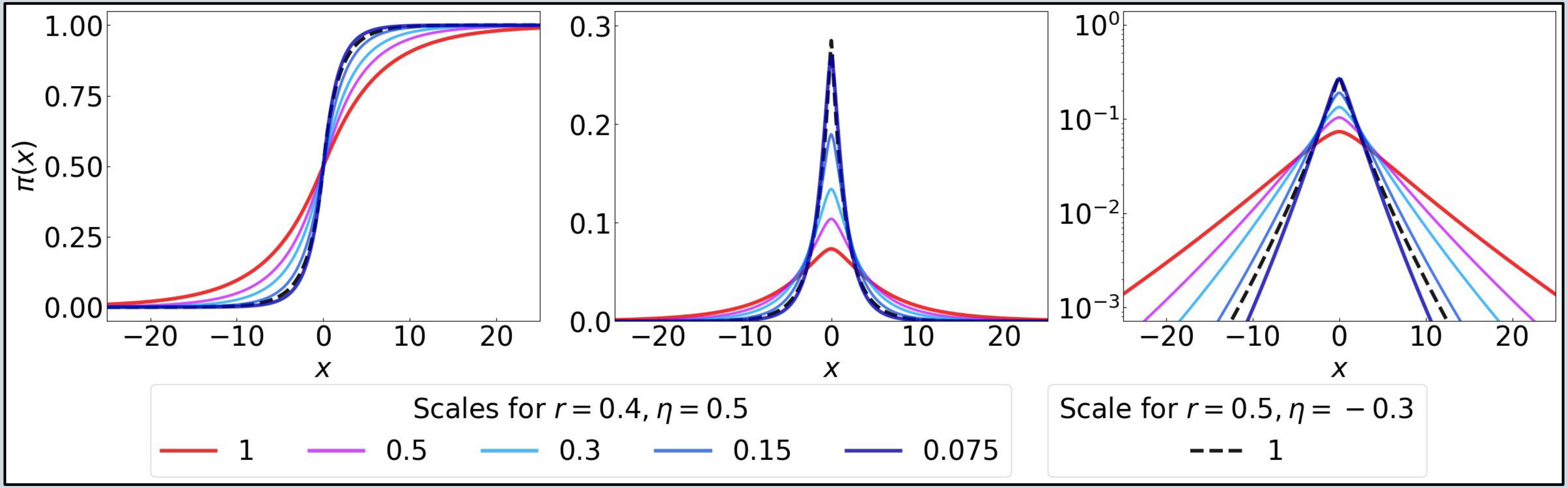



**t-distributions, Laplace distributions (any $\ell_p$ ),** and **Gaussian distributions** are special cases.




Given fixed scale, the shape parameters are identifiable.

For fixed scale, the prior varies most (in KL or KS) along directions which change its variance.




Variations in scale can approximately account for changes in the shape parameters.

Collectively, $\boldsymbol{r, \eta, \vartheta}$ **are approximately unidentifiable**

| Title | Year | Examples | Motivation |
|---|---|---|---|
| Hypermodels in the Bayesian imaging framework [SM11] | 2008 | Synthetic | Computational |
| Conditionally Gaussian hypermodels for cerebral source localization [SM5] | 2009 | Synthetic | Application |
| Sparse Bayesian image Restoration [SM2] | 2010 | Classic | Application |
| A hierarchical Krylov–Bayes iterative inverse solver for MEG with physiological preconditioning [SM6] | 2015 | Synthetic | Application |
| Bayes meets Krylov: Statistically inspired preconditioners for CGLS [SM8] | 2018 | Synthetic | Computational |
| Brain activity mapping from MEG data via a hierarchical Bayesian algorithm with automatic depth weighting [SM7] | 2019 | Real and Synthetic | Application |
| Hierarchical Bayesian models and sparsity: $\ell^2$-magic [SM12] | 2019 | Synthetic | Computational |
| Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors [SM10] | 2020 | Synthetic | Computational |
| Sparsity promoting hybrid solvers for hierarchical Bayesian inverse problems [SM9] | 2020 | Synthetic | Computational |
| Overcomplete representation in a hierarchical Bayesian framework [SM20] | 2020 | Synthetic | Computational |
| A variational inference approach to inverse problems with gamma hyperpriors [SM1] | 2022 | Synthetic | Computational |
| Hierarchical ensemble Kalman methods with sparsity-promoting generalized Gamma hyperpriors [SM17] | 2022 | Synthetic | Computational |
| Hierarchical Ensemble Kalman Methods with Sparsity-Promoting Generalized Gamma Hyperpriors [SM17] | 2022 | Synthetic | Computational |
| Sparsity promoting reconstructions via hierarchical prior models in diffuse optical tomography [SM19] | 2023 | Synthetic | Computational |
| Sequential image recovery using joint hierarchical Bayesian learning [SM23] | 2023 | Real and Synthetic | Computational |
| Leveraging joint sparsity in hierarchical Bayesian learning [SM14] | 2024 | Synthetic | Computational |
| Efficient sampling for sparse Bayesian learning using hierarchical prior normalization [SM15] | 2025 | Synthetic | Computational |
| Efficient sparsity-promoting MAP estimation for Bayesian linear inverse problems [SM18] | 2025 | Classic | Computational |

Despite a growing literature, **there is limited empirical validation** of the prior model.

Validation has often been limited to small sets of classical or synthetic examples. Examples using real-world data have validated a full Bayesian pipeline.

***Direct validation of the prior in isolation of an inverse solver/likelihood Is missing.***

# Motivating Question

**Does the mixture model fit real-world imaging datasets?**

1. If so, under what representations?
2. And, for what range of parameter values?

To adopt the prior model in absence of a first-principles justification we should trust that:

1. It fits large, real-world imaging datasets that vary in source, subject, and representation
2. It provides substantially better fits than simpler standard models
3. It’s parameters can be identified accurately via empirical Bayes or the inferential process is robust to misspecification errors.

# Data

SOURCES, REPRESENTATIONS, DATA TREATMENT

# Data Sets

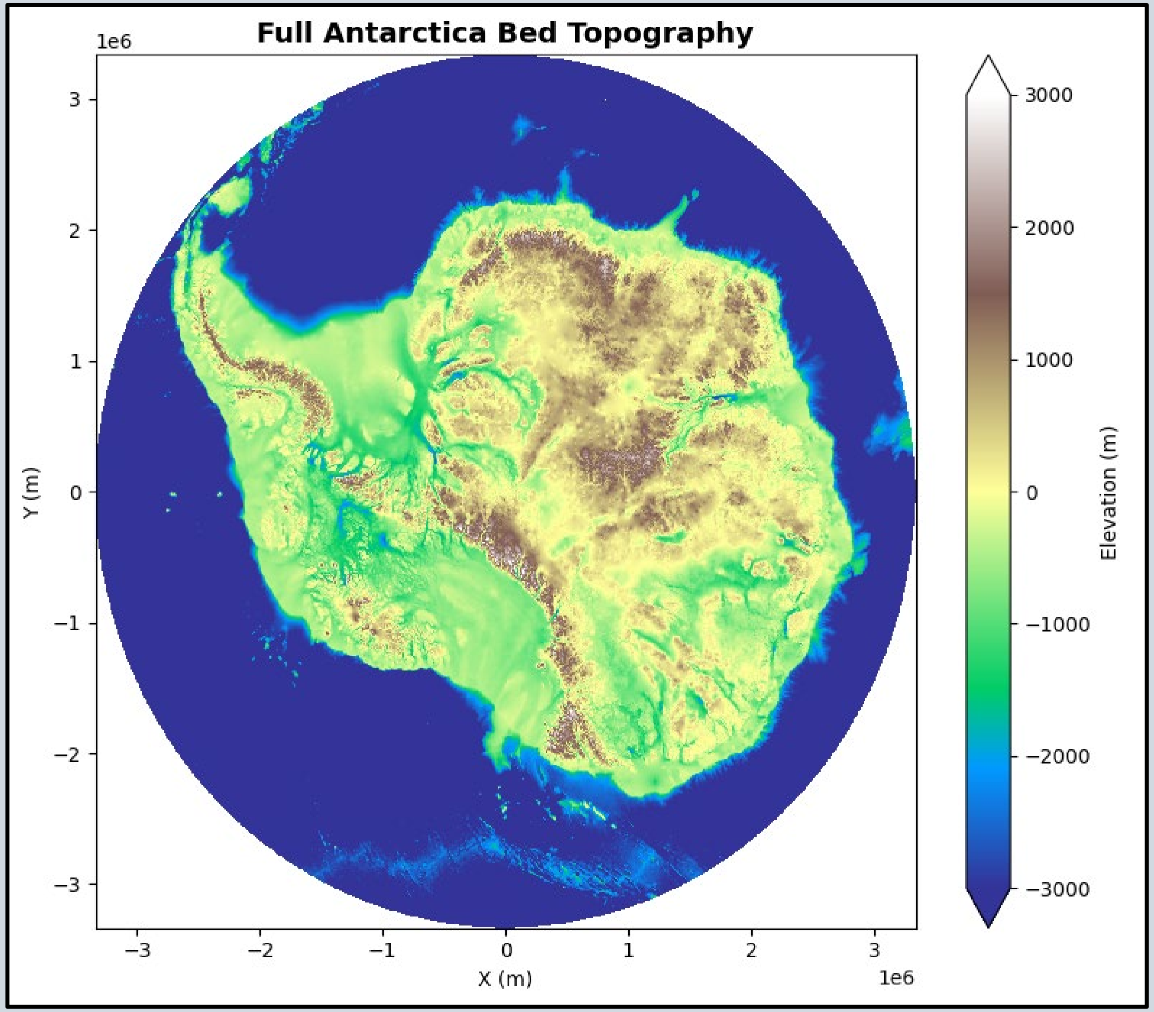
1. Satellite Images
2. Natural Images
3. Brain MRI (synthetic)

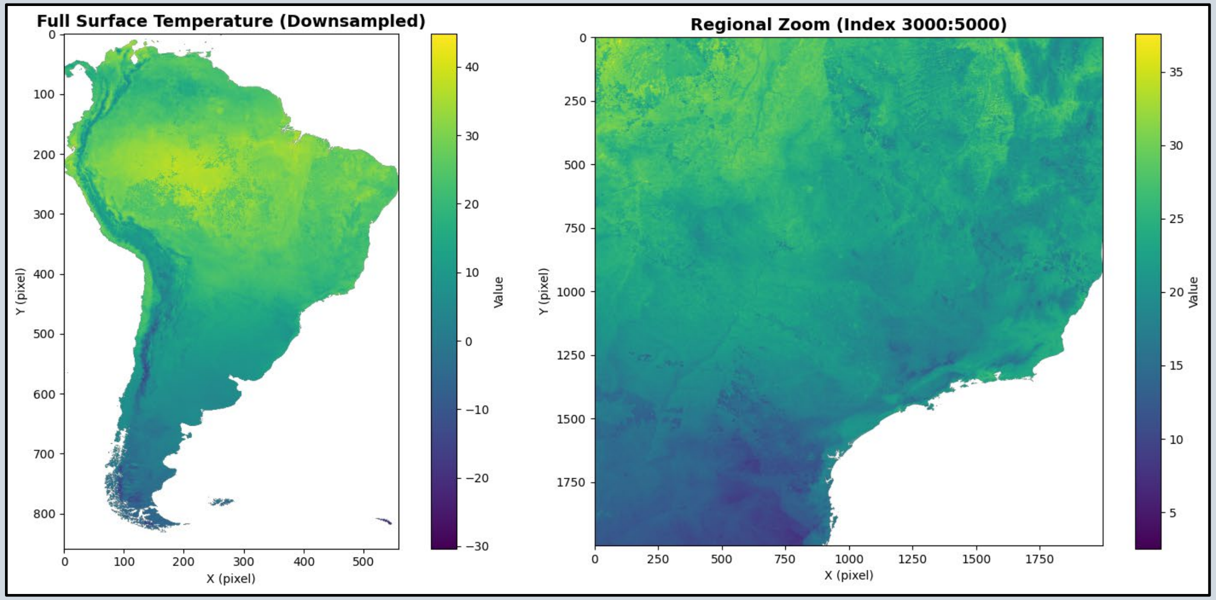
Contrast to classical image dataset.

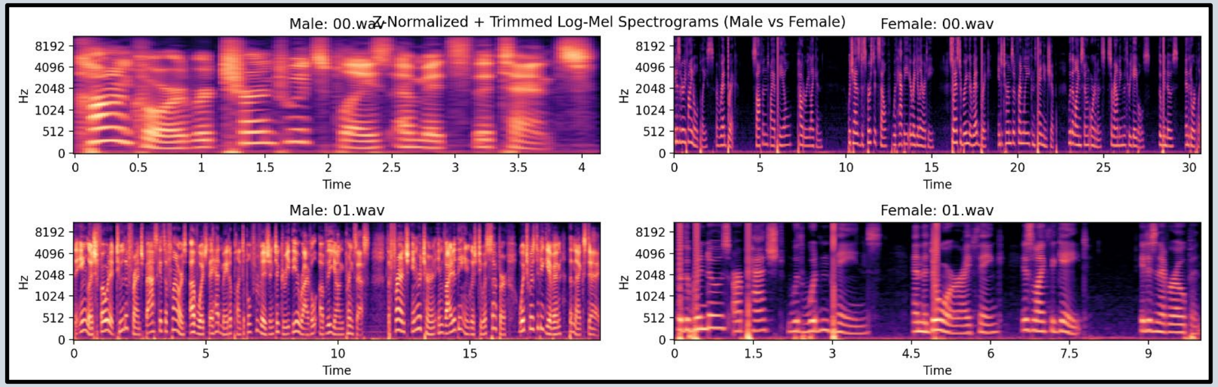
| Name | # Images | Description |
|---|---|---|
| agriVision | 4500 | Agricultural Vision, central US farmlands ($256 \times 256$) [17] |
| pastis | 1590 | Panoptic Agricultural Satellite Time Series, farm fields in France ($128 \times 128$) [46] |
| spaceNet | 3401 | Multi-Sensor All Weather Mapping Dataset, centered in Rotterdam ($400 \times 400$) [41] |
| coco | 4050 | Common Objects in Context, split into indoor/outdoor subsets ($256 \times 256$) [31] |
| segmentAnything | 7072 | Natural image dataset, compiled for training segmentation models ($512 \times 512$) [30] |
| syntheticMRI2D | 3000 | 2D synthetic MRI brain images, split into axial ($333 \times 234$), coronal ($244 \times 234$), and sagittal slices ($174 \times 234$) [39] |
| syntheticMRI3D | 100 | 3D synthetic MRI brain images ($160 \times 224 \times 160$) [39] |

Ongoing study of geospatial data (Antarctic bedmap, surface temperature, magnetic anomalies) and audio (speech).


Remote Sensing
Natural
Medical
Classical
FISHERMAN'S WHARF
U.S. AIR FORCE



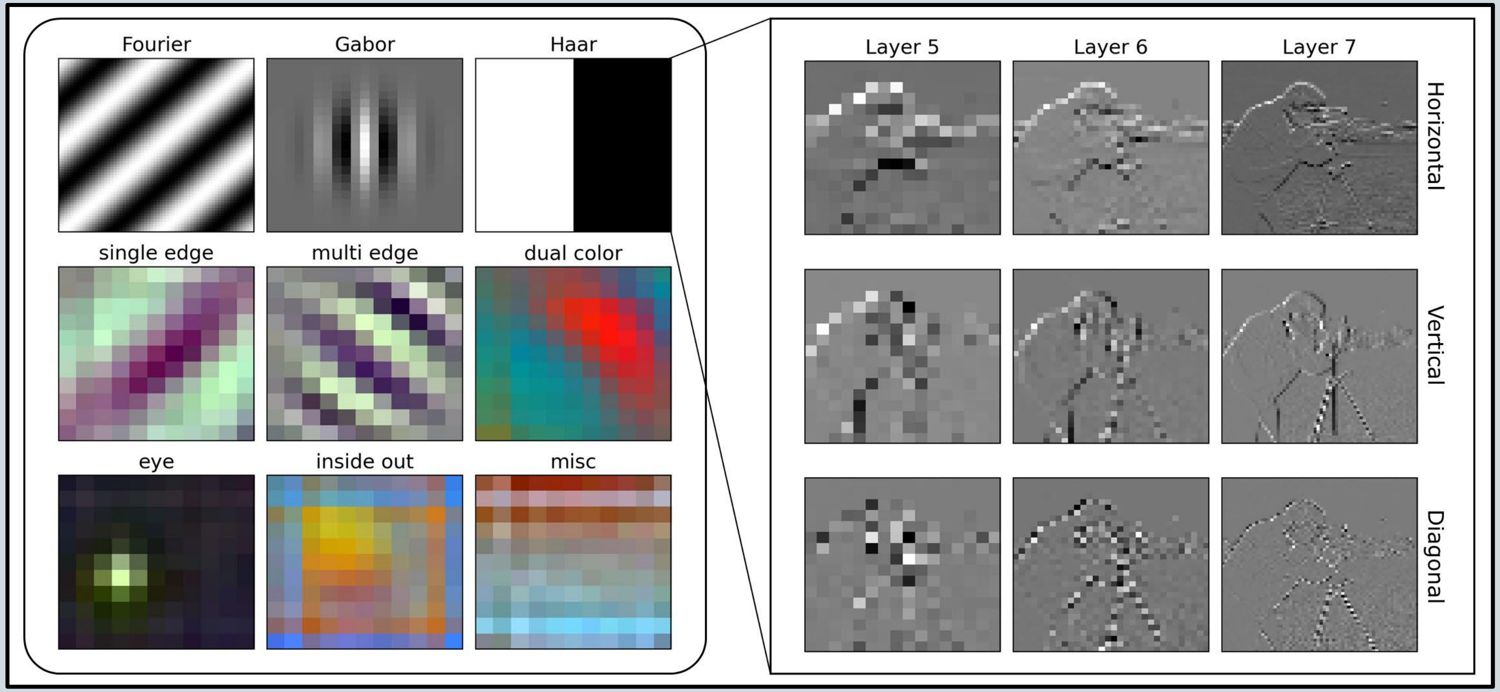

Example geospatial and audio data.

Work on these datasets is ongoing. **Preliminary results are consistent with the completed studies.**

# Representations

Images:

1. Fourier
2. Gabor Wavelet
3. Haar Wavelet
4. AlexNet Filter Bank




Audio:

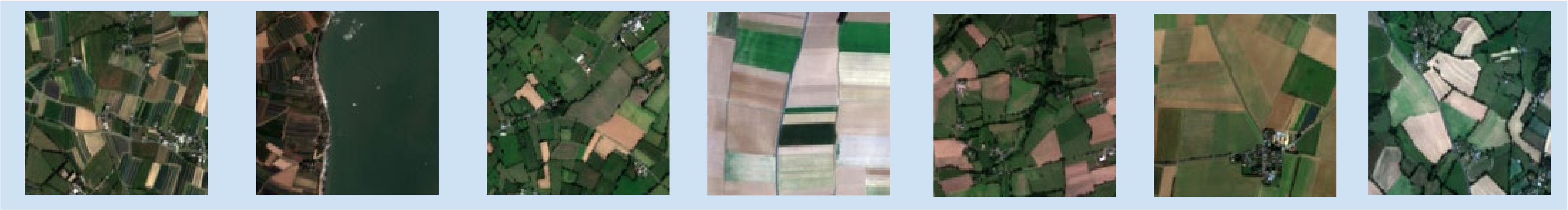
1. Fourier
2. Short-time Fourier
3. Continuous Wavelet
4. Erblet

# Data Treatment

1. Standardize (center, scale) and symmetrize (augment with reflections)
2. Transform by convolving with a filter bank (change representation)
3. Group the resulting coefficients and assign all coefficients within a group the same prior parameters.

Step (3.) treats all coefficients assigned to a group as if identically distributed.

Original:



Augmented:

Grouping coefficients is equivalent to **augmenting the dataset**, where all coefficients assigned to a group are randomly exchanged.

Some coefficients are **approximately exchangeable** (e.g. waves of similar wavelength, coefficients that can be exchanged by rotations or reflections), so modeling the augmented data introduces **negligible** misspecification errors.

Otherwise, **augmentations introduce misspecification errors**.

We accept an augmentation if those errors could:

1. Be corrected by a likelihood,
2. Produce artifacts that could be corrected post-hoc,
3. Produce artifacts that do not interfere with the user objectives.

And, if grouping:

1. Combines coefficients that would plausibly be modeled as identical, and
2. Significantly reduces the number of separate coefficients to model.

| Image Augmentation | Transform | Coefficient Augmentation | Types of Misspecification Error |
|---|---|---|---|
| Translation | Fourier | Group real and imaginary parts | Negligible error due to approximate exchangeability. |
| Spatial Permutation | Haar<br>Gabor<br>Learned | Group by orientation per layer<br>Group by filter | Likelihood constrained misspecification due to collaging effect resolved by data. |
| Rotation (90°) | Haar | Group horizontal and vertical details | Negligible error due to approximate exchangeability. |
| Rotation (Arbitrary) | Fourier<br>Gabor | Group by frequency components<br>Group over orientation per filter | Arbitrary rotations are Irrelevant Artifacts. |
| Scaling | Fourier | Group similar frequency components | Negligible error due to approximate exchangeability. |
| Symmetrization | Haar<br>Gabor | Append negated coefficients | Recoverable Artifacts after inspection |

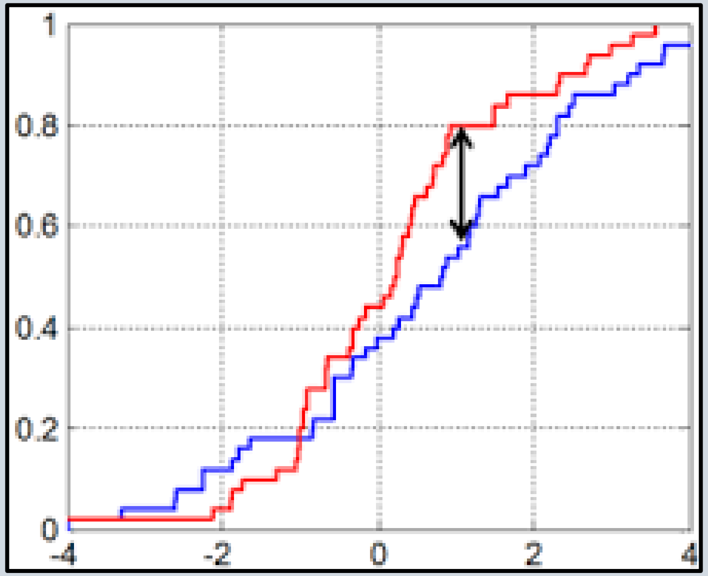
# Methods and Results

INDEPENDENCE, HYPOTHESIS TESTING (KS), FIT EVALUATION

# Test Procedure

For each coefficient group we:

1. **Sweep over outlier rejection thresholds** and estimate empirical variance, then standardize the data to each empirical variance. This is equivalent to selecting a series of **scale parameters.**
2. Run a grid **search over the shape parameters for each choice of scale parameter**
3. **Evaluate fit** using the one-sample **Kolmogorov-Smirnov (KS) distance (practical significance)**
4. **Test the statistical significance** of observed discrepancies using a **KS test**
5. Return the regions of shape and scale parameters that pass (3) or pass (4)



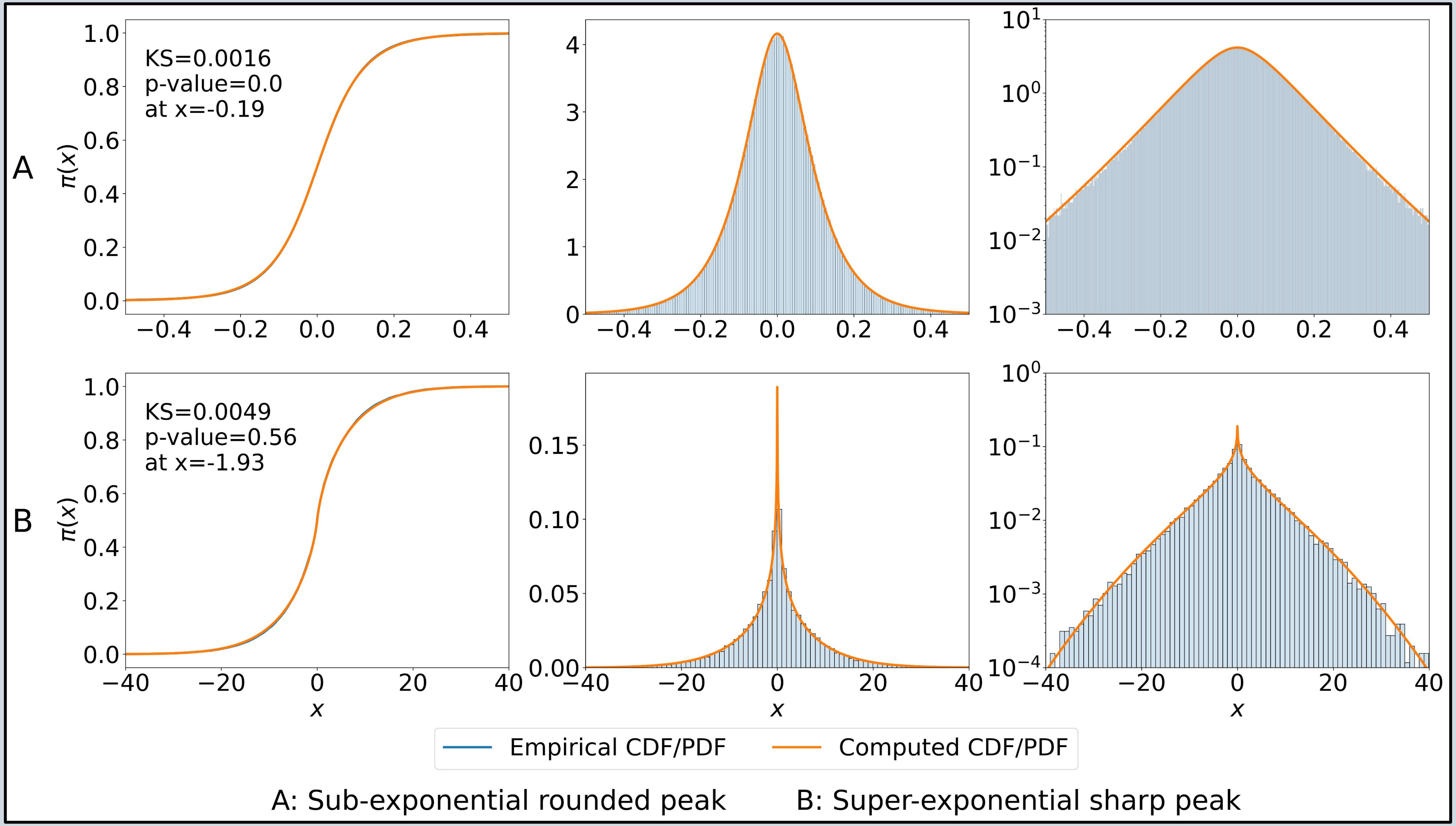
# Test Procedure

We **do not apply any multiple testing correction** since we don't believe any data actually comes from our model.

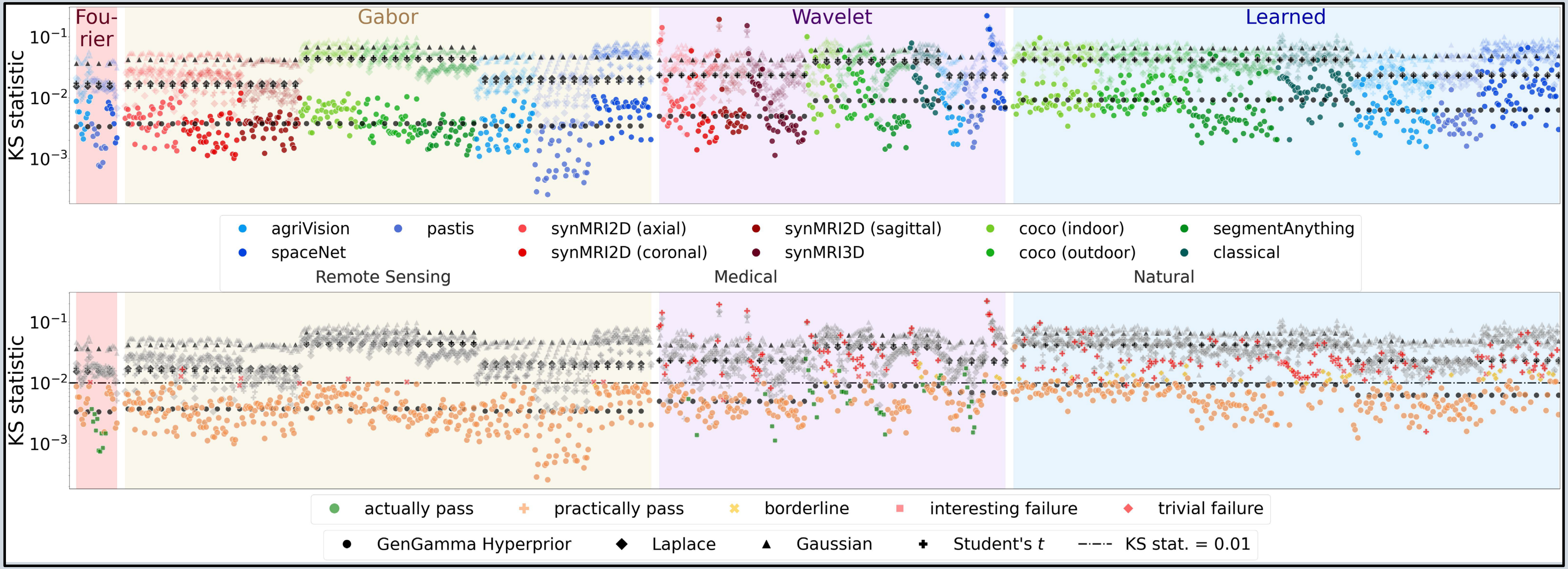
- When the fit passes step (4) we interpret it as an example where the discrepancy between the model and the data is too small to falsify the model.
- If the statistical standard is stringent, then discrepancies that are too small to falsify the model are likely too small to produce inferential errors.

We check all borderline cases by eye **to classify model failures** and to ensure our fit standards are conservative.

- Statistical Pass = p-value greater than 0.05, more than 100 samples
- Practical Pass = KS statistic smaller than 0.01, corroborated by eye

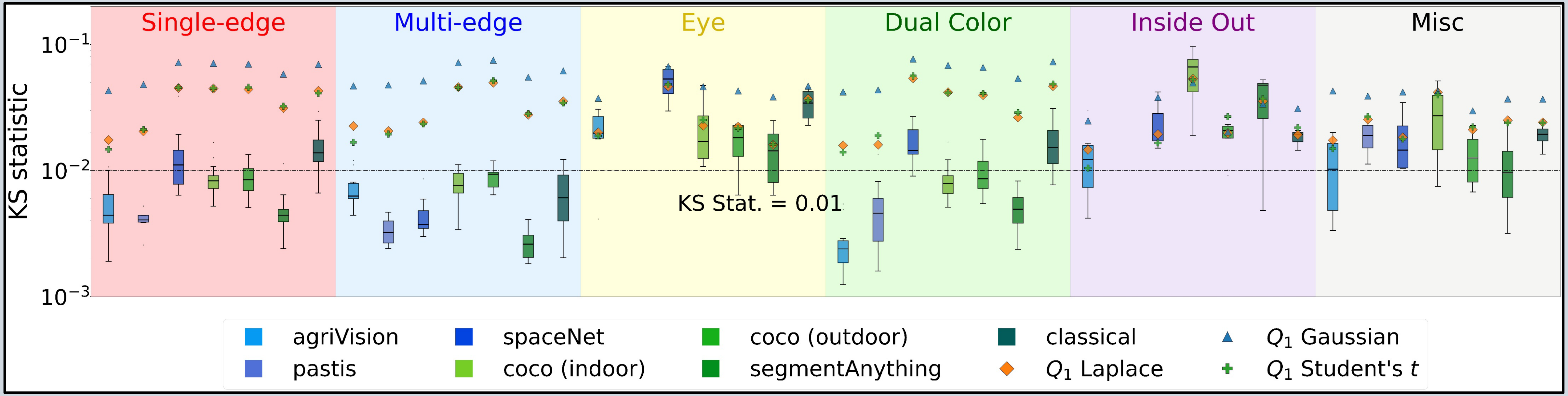



The hierarchical **prior can achieve high fidelity fits for non-standard distributions**.



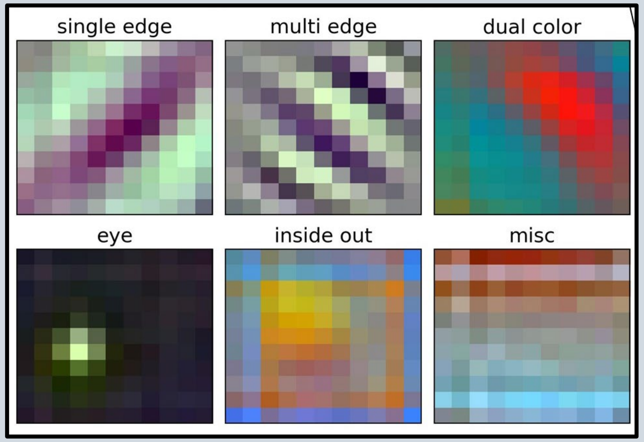

The KS statistic **is often small (< 0.02),** and in almost all cases **an order of magnitude smaller than for standard models**.

In most cases, **discrepancies are statistically significant**, but **small**.




Tend to see better fits for filters that are:

1. **wavelet like** (edge detectors),
2. **smooth** (Gabor > Haar),
3. **higher frequency** (though trends vary by dataset and filter).

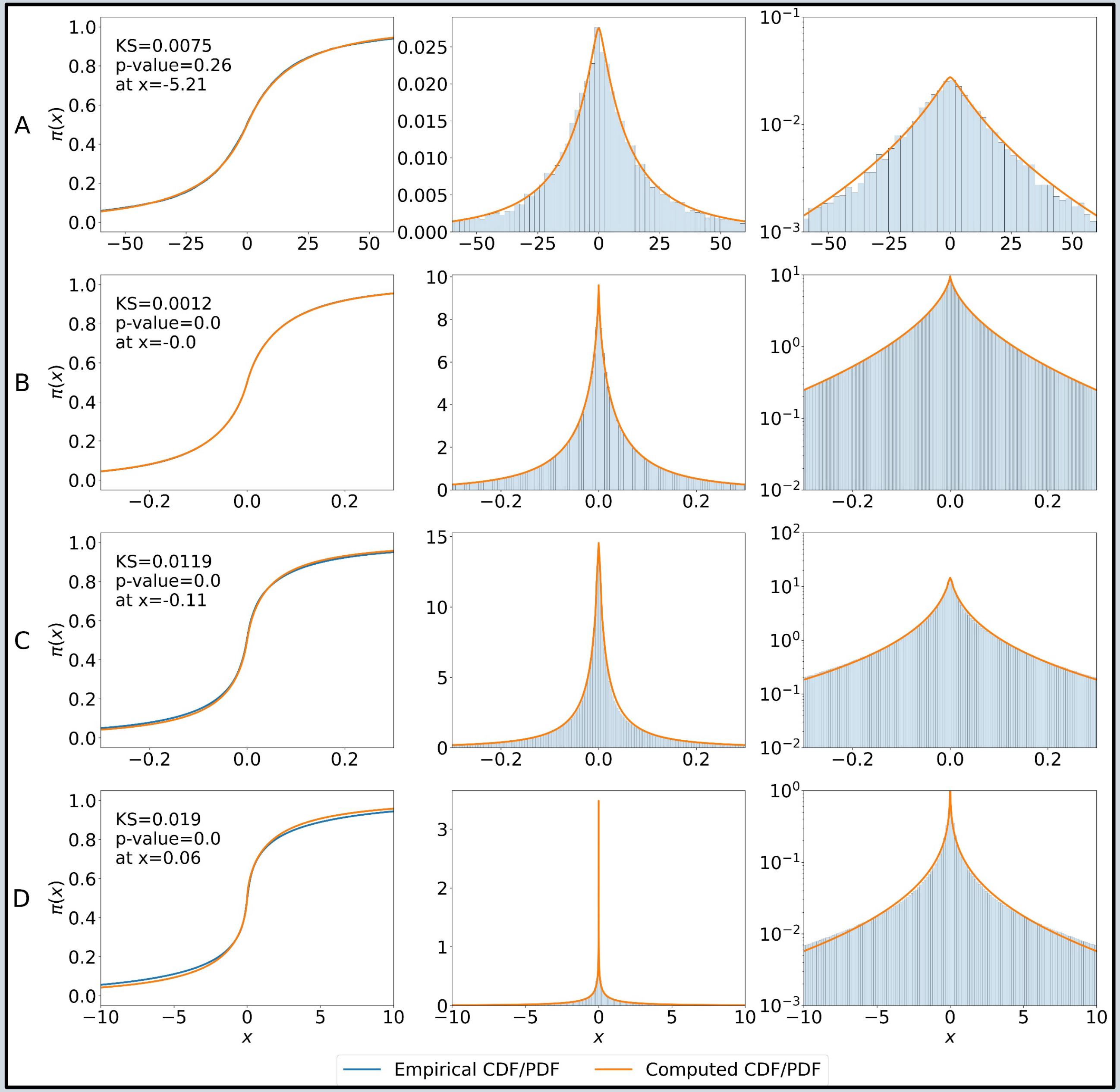


**Example Fit Outcomes:**

- **A:** statistically pass
- **B:** practically pass
- **C:** borderline case (rejected)
- **D:** Interesting failure (rejected)

In many cases (e.g. C, D) we **rejected fits that would be unlikely to produce inferential errors.**

Categorized borderline cases by hand to ensure thresholds for acceptance were defined conservatively.

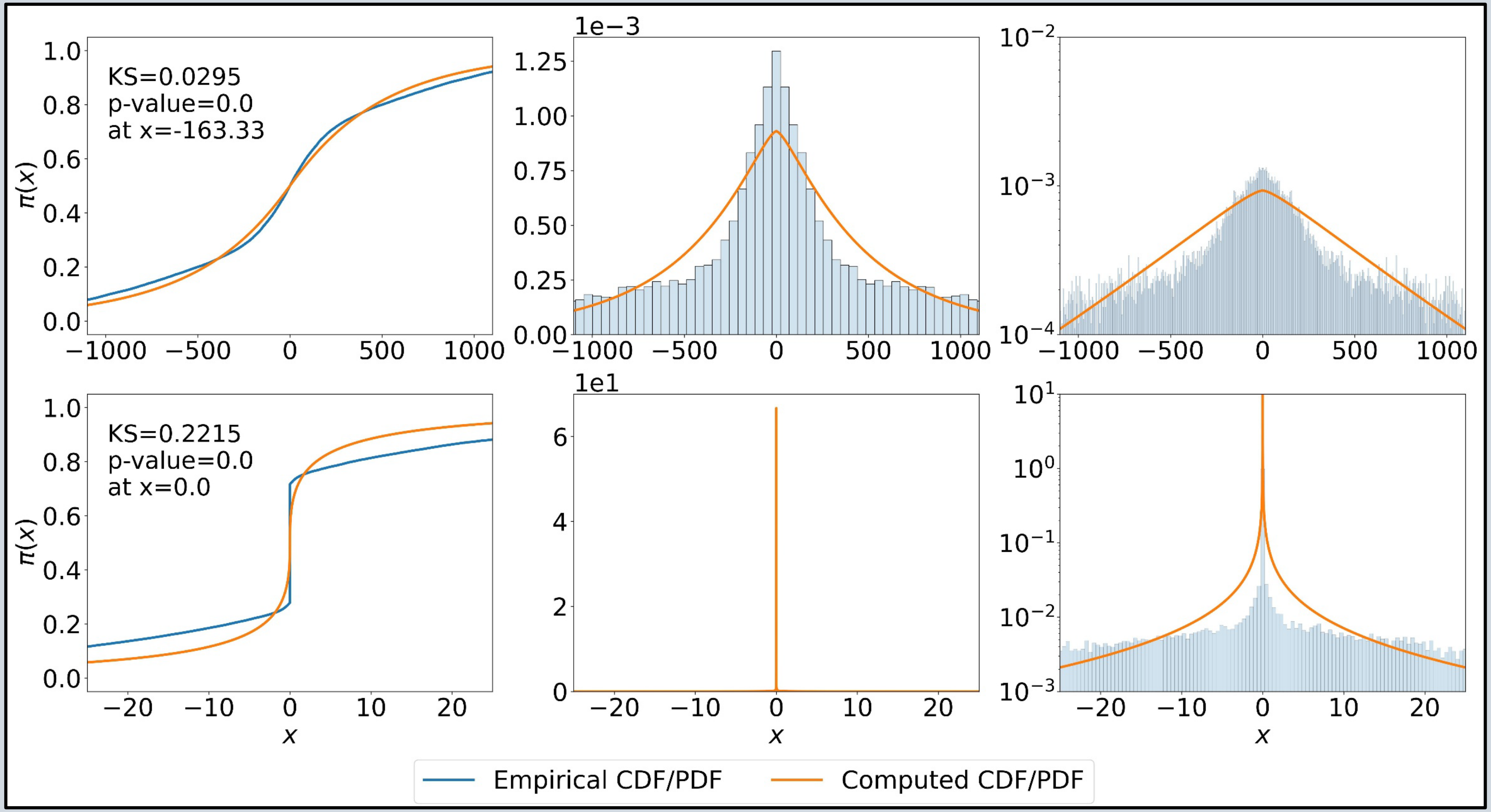


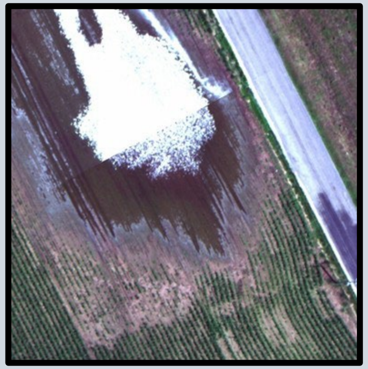
| Transform | Dataset | Median Samples | Median KS Stat. | Stat. Pass $\alpha = 0.05$ (%) | Combined Pass (%) |
|---|---|---|---|---|---|
| Fourier | agriVision | 2.6e6 | 0.005 | 2.3 | 95.5 |
| | pastis | 2.6e5 | 0.002 | 90.9 | 100.0 |
| | spaceNet | 1.8e6 | 0.005 | 0.0 | 100.0 |
| | Remote Sensing | 5.5e5 | 0.003 | 34.2 | 98.4 |
| Gabor | synMRI2D (axial) | 1.8e8 | 0.005 | 0.0 | 95.2 |
| | synMRI2D (coronal) | 1.4e8 | 0.003 | 0.0 | 95.2 |
| | synMRI2D (sagittal) | 9.7e7 | 0.004 | 0.0 | 97.6 |
| | Medical | 1.4e8 | 0.004 | 0.0 | 96.0 |
| | coco (indoor) | 6.6e7 | 0.006 | 0.0 | 97.0 |
| | coco (outdoor) | 6.6e7 | 0.004 | 0.0 | 95.8 |
| | segmentAnything | 2.6e8 | 0.003 | 0.0 | 100.0 |
| | standardTesting | 2.4e6 | 0.008 | 0.0 | 57.1 |
| | Natural | 6.6e7 | 0.004 | 0.0 | 94.5 |
| | agriVision | 2.6e8 | 0.003 | 0.0 | 100.0 |
| | pastis | 1.6e7 | 0.001 | 0.6 | 100.0 |
| | spaceNet | 1.6e8 | 0.007 | 0.0 | 92.9 |
| | Remote Sensing | 1.6e8 | 0.003 | 0.2 | 97.6 |

| | | | | | |
|---|---|---|---|---|---|
| Haar | synMRI2D (axial) | 4.2e6 | 0.007 | 0.0 | 76.2 |
| | synMRI2D (coronal) | 3.7e6 | 0.004 | 9.5 | 80.9 |
| | synMRI2D (sagittal) | 2.4e6 | 0.005 | 0.0 | 81.0 |
| | synMRI3D | 1.6e6 | 0.005 | 7.1 | 69.0 |
| | Medical | 2.9e6 | 0.005 | 4.8 | 75.3 |
| | coco (indoor) | 2.6e5 | 0.013 | 25.0 | 50.0 |
| | coco (outdoor) | 3.9e5 | 0.011 | 25.0 | 53.1 |
| | segmentAnything | 1.8e6 | 0.004 | 25.0 | 85.2 |
| | standardTesting | 9.2e3 | 0.013 | 9.5 | 47.6 |
| | Natural | 4.1e5 | 0.008 | 24.0 | 62.6 |
| | agriVision | 1.7e6 | 0.005 | 15.3 | 100.0 |
| | pastis | 1.5e5 | 0.004 | 57.1 | 100.0 |
| | spaceNet | 1.6e6 | 0.012 | 4.7 | 53.1 |
| | Remote Sensing | 7.0e5 | 0.007 | 24.0 | 84.4 |
| Learned | coco (indoor) | 2.0e8 | 0.009 | 0.0 | 64.5 |
| | coco (outdoor) | 2.0e8 | 0.010 | 0.0 | 57.8 |
| | segmentAnything | 7.9e8 | 0.005 | 0.0 | 81.2 |
| | standardTesting | 7.1e6 | 0.014 | 0.0 | 24.1 |
| | Natural | 2.0e8 | 0.009 | 0.0 | 58.2 |
| | agriVision | 7.9e8 | 0.005 | 0.0 | 73.3 |
| | pastis | 4.9e7 | 0.004 | 0.0 | 93.8 |
| | spaceNet | 4.8e8 | 0.013 | 0.0 | 42.9 |
| | Remote Sensing | 4.8e8 | 0.006 | 0.0 | 66.2 |

**Headline Numbers:**

- Median KS = **0.005**,
- % Statistically Pass = **7.3%** for median sample size 60 million
- % Practically Pass = **83.5%** at KS < 0.01 standard



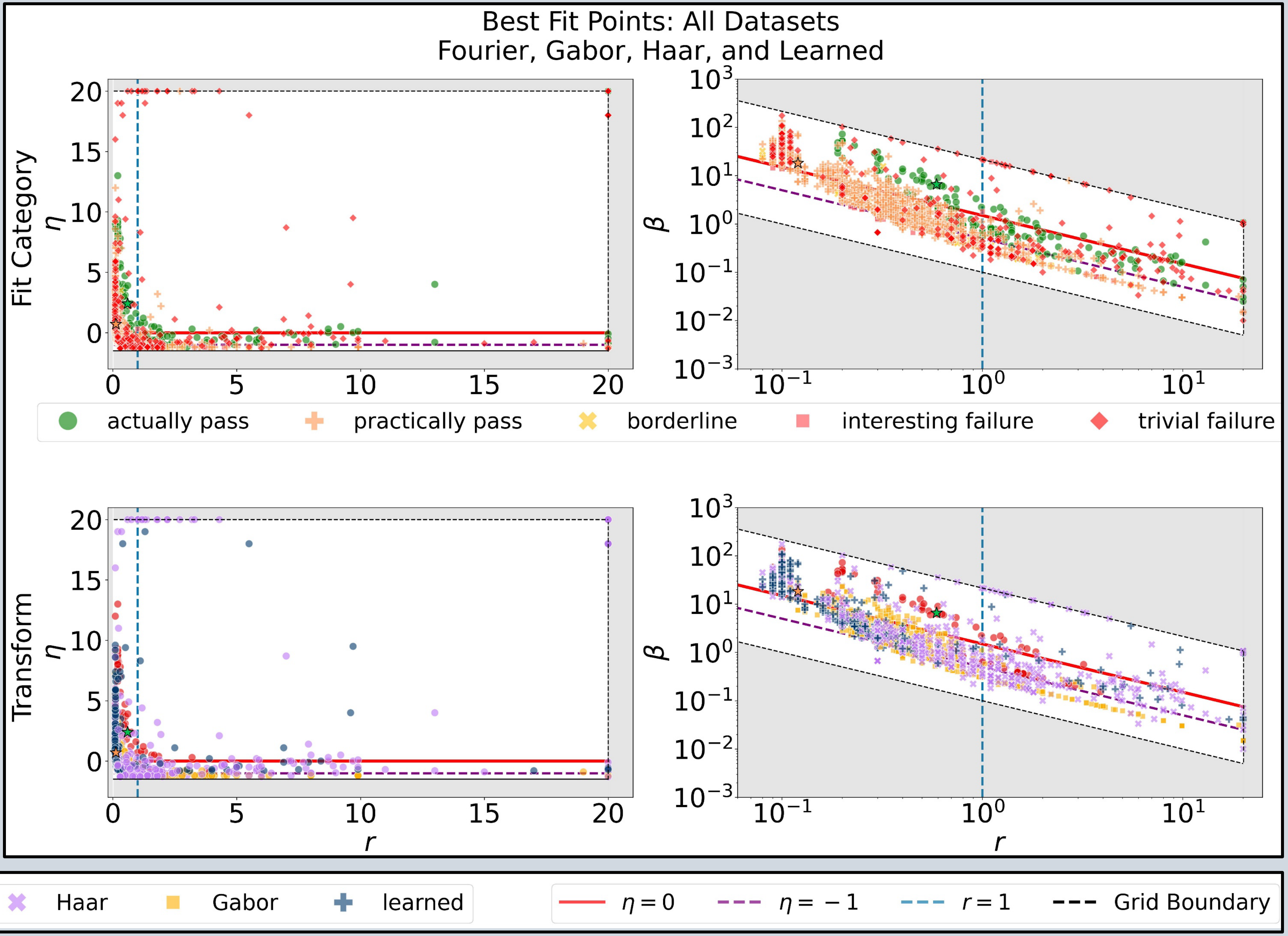



Not all examples fit.

Best fit shape parameters vary widely.

Are **not limited to small $\boldsymbol{\eta}$** (e.g. $\ell_p$ priors).

**Shape parameters are poorly identified if scale can vary.**



# Discussion

LIMITATIONS, FUTURE/ONGOING WORK, PREPRINT

# Limitations and Future Work

1. The KS statistic is relatively insensitive to tail behavior, cannot precisely resolve $r$
   - Run an intersection test using KS and Anderson-Darling
2. Expand to study other image data sets (geospatial) and signals (audio)

**Thank you for your attention and for the invitation to attend!**

Preprint (repository and bibliography):

*Do Generalized Gamma Scale Mixtures of Normals Fit Large Image Datasets?*

*Brandon Marks, Yash Dave, Zixun Wang, Hannah Chung, Riya Patwa, Simon Cha, Michael Murphy, and Alexander Strang. (2025).*

[https://arxiv.org/abs/2512.17038](https://arxiv.org/abs/2512.17038)

# References


D. Calvetti, et al. Conditionally Gaussian hypermodels for cerebral source localization, SIAM Journal on Imaging Sciences, 2 (2009), pp. 879–909. https://epubs.siam.org/doi/10.1137/080723995

D. Calvetti, et al. Brain activity mapping from MEG data via a hierarchical Bayesian algorithm with automatic depth weighting , Brain topography, 32 (2019), pp. 363–393. https://link.springer.com/article/10.1007/s10548-018-0670-7

D. Calvetti, et al. Sparse reconstruction from few noisy data: analysis of hierarchical bayesian models with generalized hyperpriors, Inverse Problems, 36 (2020 ) https://par.nsf.gov/servlets/purl/10183041

D. Calvetti, et al. Hierarchical Bayesian models and sparsity: $\ell 2$-magic, Inverse Problems, 35 (2019), p. 035003 https://iopscience.iop.org/article/10.1088/1361-6420/aaf5ab/pdf

J. Lindbloom et al. Efficient sparsity-promoting map estimation for bayesian linear inverse problems, Inverse Problems, 41 (2025), p. 025001 https://iopscience.iop.org/article/10.1088/1361-6420/ada17f

Y. Xiao and J. Glaubitz, Sequential image recovery using joint hierarchical bayesian learning, Journal of Scientific Computing, 96 (2023) https://arxiv.org/abs/2206.12745

Sainte Fare Garnot Vivienand Landrieu Loic. PASTIS - Panoptic Segmentation of Satellite Image Time Series. 1.0, Zenodo, 22 June 2021, https://zenodo.org/records/5012942

David, Wilson, et al. "Agriculture-Vision: A Large Aerial Image Database for Agricultural Pattern Analysis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. https://arxiv.org/abs/2001.01306

Shermeyer, J., et al. SpaceNet 6: Multi-Sensor All Weather Mapping Dataset. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 768-777. https://arxiv.org/abs/2004.06500

W. H. L. Pinaya, et al. Brain imaging generation with latent diffusion models, 2022 https://arxiv.org/abs/2209.07162

T. Lin, et al. Microsoft COCO: common objects in context, CoRR, abs/1405.0312 (2014). https://arxiv.org/abs/1405.0312

A. Kirillov, et al. Segment anything, (2023). https://arxiv.org/abs/2304.02643