

Article

Not peer-reviewed version

Soft Model Selection and Structured Assembly of Observables for Micro–Macro Transitions, Operational Relativity, and Emergent Thermodynamics

[Ricardo Adonis Caraccioli Abrego](#) *

Posted Date: 28 February 2026

doi: 10.20944/preprints202602.1931.v1

Keywords: model selection; softmax; loss; entropic regularization; information geometry; decoherence; coarse-graining; emergent thermodynamics; NISQ; mixture of experts



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Soft Model Selection and Structured Assembly of Observables for Micro–Macro Transitions, Operational Relativity, and Emergent Thermodynamics

Ricardo Adonis Caraccioli Abrego

Department of Electrical Engineering, Universidad Nacional Autónoma de Honduras (UNAH); ricardo.caraccioli@unah.edu.hn

Abstract

An operational meta-model is presented for transitions among quantum, classical, relativistic, and thermodynamic descriptions without forcing a single master state space. Unification is performed at the level of observable predictions: each formalism produces an output in a common space \mathcal{Y} defined by a feature map Φ (moments, spectra, correlations, or other functionals). Convex weights are assigned via a standard soft selection rule (softmax/Gibbs form) from losses, with entropic regularization and a complexity penalty (AIC/BIC/MDL) to reduce bias toward overly expressive models. Physics-facing priors are encoded through calibrated dimensionless knobs for decoherence/classicality, relativistic severity, and thermodynamicity, yielding a regime-aware gating layer. A simple out-of-catalog diagnostic (surprise/residual monitoring) is included to flag persistent mismatch that may indicate missing observables or missing models. A minimal case study template (harmonic oscillator with a bath) and an acid test in the NISQ regime (critical decoherence) are outlined as reproducible validation pathways.

Keywords: model selection; softmax; loss; entropic regularization; information geometry; decoherence; coarse-graining; emergent thermodynamics; NISQ; mixture of experts

1. Motivation, positioning, and scope

“Essentially, all models are wrong, but some are useful.”

— George E. P. Box [1]

This manuscript does not propose a new physical theory, nor does it claim to unify quantum mechanics, classical dynamics, relativity, and thermodynamics at the level of a single state space. Instead, it proposes an *operational* meta-model that assembles predictions in a shared observable space. The contribution is therefore methodological: a disciplined way to combine existing regime-specific models, quantify confidence through soft weights, and detect when the available model catalog is insufficient.

Contributions (Incremental, Operational)

The contributions are modest and practical:

1. A common observable space \mathcal{Y} is formalized via a feature map Φ , and regime transitions are treated as decisions in \mathcal{Y} (rather than forcing incompatible state spaces into a single master model).
2. A standard entropically-regularized soft selection rule (softmax/Gibbs form) is used to obtain convex weights from losses, augmented with a complexity penalty to mitigate overfitting.
3. A small set of physics-inspired knobs (decoherence/classicality, relativistic severity, thermodynamicity) is used as *operational priors* that bias weights toward plausible regimes.
4. A simple out-of-catalog diagnostic (surprise/residual monitoring) is proposed to flag when the chosen Φ and the candidate model set fail to explain data.

Relation to Prior Work

The weighting mechanism is closely related to well-known ideas such as Bayesian model averaging, Gibbs posteriors/PAC-Bayes methods, and mixture-of-experts gating. The goal is not to replace these frameworks, but to provide an explicit *physics-facing* operational interface: (i) a shared observable space \mathcal{Y} tailored to measurement constraints, (ii) regime priors encoded by calibrated knobs, and (iii) a simple diagnostic for model-set inadequacy. In this sense, the novelty is primarily in the *packaging and operationalization* for micro–macro and regime-transition problems.

Scope and Limitations

This framework is only as good as (a) the chosen observable map Φ and (b) the candidate model catalog. The knobs (ξ, r, θ) are not fundamental constants; they are operational summaries that must be calibrated for each experimental context. Accordingly, the method should be read as a principled decision layer on top of existing models, not as a substitute for domain-specific mechanistic modeling.

2. Level of Unification: the Output Space \mathcal{Y} and the Map Φ

Let X be the vector of observables to be predicted. Each formalism \mathcal{M}_i produces a natural object Z_i (trajectory, distribution, spectrum, quantum channel, etc.). A feature map is defined as

$$\Phi: \mathcal{Z} \rightarrow \mathcal{Y}, \quad \hat{X}_i := \Phi(Z_i) \in \mathcal{Y}.$$

Typical examples:

- $\Phi(P) = (\mu, \sigma^2, \kappa_3, \dots, \kappa_m)$ for a distribution P .
- $\Phi(S) = (S(\omega_1), \dots, S(\omega_m))$ for a sampled spectrum.
- $\Phi(C) = (C(\tau_1), \dots, C(\tau_m))$ for sampled correlations.

In practice, $\mathcal{Y} \approx \mathbb{R}^m$.

2.1. The Choice of \mathcal{Y} is not Neutral (an Explicit Operational Hypothesis)

Selecting Φ and \mathcal{Y} encodes which degrees of freedom are considered relevant in the micro–macro transition. In that sense, the coarse-graining implicit in Φ is an operational hypothesis: it assumes that certain functionals (moments, spectra, correlations) capture what is essential for the regime of interest. This choice should be guided by invariances, scales, and measurement constraints.

3. Assembly: Convex Combination and Structure-Preserving Operators

3.1. Base Meta-Predictor (Convex)

A minimal assembler is defined by

$$\hat{X} = \sum_i k_i \hat{X}_i, \quad \sum_i k_i = 1, \quad k_i \geq 0. \quad (1)$$

3.2. Structured assembly (when constraints exist)

When the observable has structure (e.g., normalized distributions, PSD matrices, density matrices), a naive linear combination may violate constraints (positivity, normalization, etc.). In such cases, a structure-preserving assembly operator is used:

$$\hat{X} = \mathcal{A}(\{\hat{X}_i\}, \mathbf{k}), \quad (2)$$

where \mathcal{A} is chosen to preserve structure. Examples:

- **Distributions:** convex mixture in the probability simplex (preserves normalization and non-negativity).
- **Density matrices:** convex mixture of states $\rho = \sum_i k_i \rho_i$ (preserves positivity and trace 1).
- **Constrained observables:** projection Π onto a valid convex set: $\hat{X} = \Pi(\sum_i k_i \hat{X}_i)$.

Equation (1) is used as the minimal reference, and (2) as its natural extension.

4. Weights from Loss: Soft Selection with Regularization and Complexity

Let $E_i \geq 0$ be a total loss associated with model i . Define:

$$k_i = \frac{\exp(-\lambda E_i)}{\sum_j \exp(-\lambda E_j)}, \quad \lambda > 0. \quad (3)$$

4.1. Convex Interpretation (Minimum Loss + Entropy)

Rule (3) solves

$$\min_{k \in \Delta} \sum_i k_i E_i + \frac{1}{\lambda} \sum_i k_i \ln k_i, \quad (4)$$

where $\Delta = \{k_i \geq 0, \sum_i k_i = 1\}$ and the second term is entropic regularization.

4.2. Loss Definition in \mathcal{Y}

If a reference X^* exists (measurement or “fine” simulation),

$$E_i^{\text{data}} = d(\hat{X}_i, X^*).$$

Common examples:

$$d_{\text{RMSE}}(a, b) = \sqrt{\frac{1}{m} \sum_{\ell=1}^m (a_{\ell} - b_{\ell})^2}, \quad d_{\text{KL}}(P \parallel Q) = \sum_x P(x) \ln \frac{P(x)}{Q(x)}.$$

4.3. Complexity Penalty (AIC/BIC/MDL)

To reduce bias toward overly complex models,

$$E_i = E_i^{\text{data}} + \beta \Omega_i, \quad (5)$$

where Ω_i can be AIC/BIC/MDL [2–4].

4.4. Knobs as Calibrated Parametrizations

The functional forms used for (ζ, r, θ) are chosen for monotonicity, boundedness, and numerical stability. Other monotone parametrizations would be equally valid. In applications, these mappings should be calibrated (or learned) from validity data, measurement noise levels, and known scale separations.

5. Physics-Facing Priors from Knobs

Define dimensionless knobs (examples; operational use only):

$$\zeta = \Gamma t_{\text{obs}} \quad (\text{decoherence / classicality}), \quad (6)$$

$$\beta_S = \frac{v^2}{c^2}, \quad \beta_G = \frac{2GM}{rc^2}, \quad r = 1 - \exp[-(\beta_S + \beta_G)] \quad (\text{relativistic severity}), \quad (7)$$

$$\theta = \frac{t_{\text{obs}}}{\tau_{\text{mix}}} N_{\text{eff}} \quad (\text{thermodynamicity}). \quad (8)$$

Let $q = e^{-\zeta}$. A minimal factorized prior is:

$$w_{QR} = qr, \quad w_Q = q(1-r), \quad w_R = (1-q)r, \quad w_C = (1-q)(1-r), \quad (9)$$

with $\sum w. = 1$.

5.1. Combining Prior + Loss

$$k_i = \frac{w_i \exp(-\lambda E_i)}{\sum_j w_j \exp(-\lambda E_j)}. \quad (10)$$

6. Hardening: Knob Coupling and Out-of-Catalog Monitoring

6.1. Coupling Among Knobs (Validity Covariance)

Knobs (ξ, r, θ) are often correlated in real systems. Let $\mathbf{p} := (\xi, r, \theta)^\top$ and introduce a covariance matrix $\Sigma_{\mathbf{p}} \succ 0$. A Mahalanobis-type penalty is defined by

$$E_{\text{cpl}}(\mathbf{p}) := \frac{1}{2}(\mathbf{p} - \boldsymbol{\mu})^\top \Sigma_{\mathbf{p}}^{-1}(\mathbf{p} - \boldsymbol{\mu}), \quad (11)$$

where $\boldsymbol{\mu}$ is an operational center (possibly context-dependent).

6.2. Hardened Total Loss

$$E_i := E_i^{\text{data}} + \alpha E_{\text{cpl}}(\mathbf{p}) + \beta \Omega_i, \quad (12)$$

and weights follow (10).

6.3. Surprise as a Diagnostic (Not a Physical Law)

Define residual loss $E_{\text{res}} := d(\hat{X}, X^*)$ and

$$S := E_{\text{res}} - \sum_i k_i E_i^{\text{data}}. \quad (13)$$

The surprise indicator is intended as an engineering diagnostic summarizing persistent mismatch between assembled predictions and reference observations. No claim is made that it is, by itself, a thermodynamic entropy production law; any thermodynamic reading should be treated as a hypothesis to be tested in open-system settings.

7. Minimal Case Study Template and a NISQ Acid Test

7.1. Minimal Reproducibility Checklist

Any case study should report: (i) the definition of Φ and the dimension of \mathcal{Y} , (ii) the metric d used for loss, (iii) the candidate models compared, (iv) the calibration procedure for (λ, α, β) and for knobs, and (v) weight trajectories k_i with residual monitoring.

7.2. NISQ Acid Test: Critical Decoherence

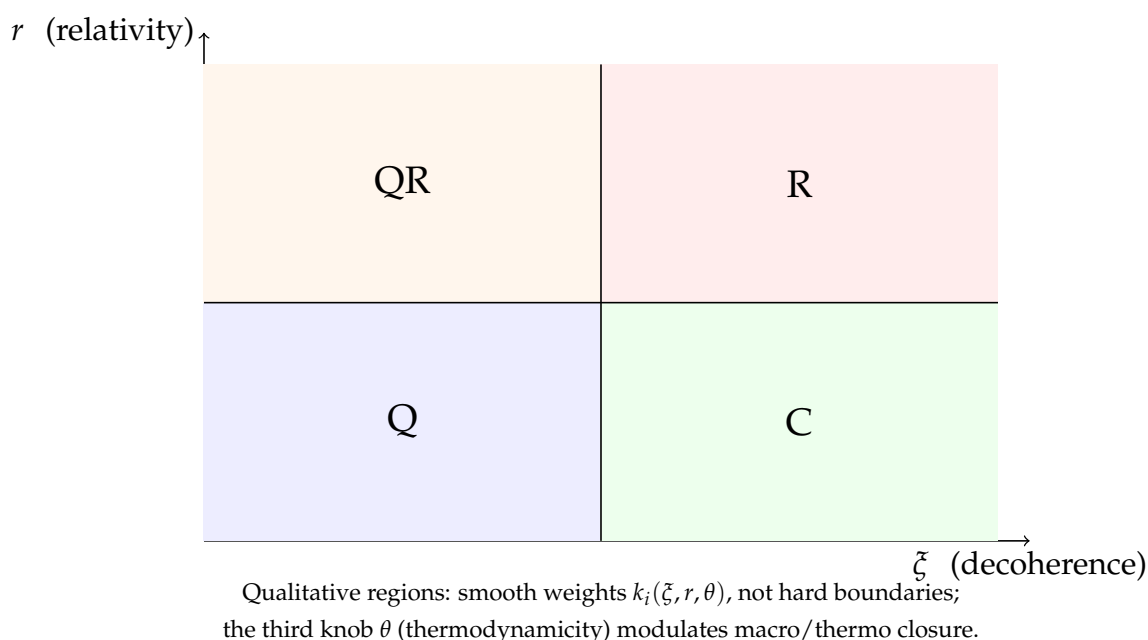
For a qubit with dephasing characterized by T_2 ,

$$\xi \approx \frac{t_{\text{obs}}}{T_2}, \quad r \approx 0,$$

and θ depends on effective mixing/noise. The operational critical time is defined by

$$t_{\text{crit}} : k_Q(t_{\text{crit}}) = k_{\text{th}}, \quad k_{\text{th}} \in (0, 1) \text{ (e.g. 0.5)}.$$

8. Qualitative Regime Figure



9. Conclusions

An operational meta-model for regime-aware prediction assembly has been presented. The framework combines (i) a shared observable space, (ii) soft loss-based weighting with regularization and complexity control, (iii) physics-facing regime priors via calibrated knobs, and (iv) a diagnostic for persistent out-of-catalog mismatch. The proposal is deliberately incremental: it is intended as a practical decision layer on top of existing mechanistic models, with clear calibration and validation hooks.

References

1. G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, 1987.
2. H. Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control **19**(6), 716–723 (1974).
3. G. Schwarz, *Estimating the dimension of a model*, Annals of Statistics **6**(2), 461–464 (1978).
4. J. Rissanen, *Modeling by shortest data description*, Automatica **14**(5), 465–471 (1978).
5. J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, *Bayesian model averaging: A tutorial*, Statistical Science **14**(4), 382–417 (1999).
6. O. Catoni, *PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*, IMS Lecture Notes–Monograph Series, 2007.
7. S.-I. Amari, *Information Geometry and Its Applications*, Springer, 2016.
8. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, *Adaptive mixtures of local experts*, Neural Computation **3**(1), 79–87 (1991).
9. W. H. Zurek, *Decoherence, einselection, and the quantum origins of the classical*, Rev. Mod. Phys. **75**, 715–775 (2003).
10. H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Oxford University Press, 2002.
11. E. T. Jaynes, *Information Theory and Statistical Mechanics*, Phys. Rev. **106**, 620–630 (1957).
12. R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, 2001.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.