SonarKAN: Sonar Kolmogorov-Arnold Network for Disentangling Passive Sonar Signatures

Xiaoliang Chen

chenxiaoliang@soundai.com

SoundAI Technology https://orcid.org/0009-0003-7060-9532

Research Article

Keywords:

Posted Date: February 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8942861/v1

Additional Declarations: The authors declare no competing interests.

SonarKAN:Sonar Kolmogorov-Arnold Network for Disentangling Passive Sonar Signatures

Xiaoliang Chen

SoundAI Technology, Beijing, China
chenxiaoliang@soundai.com

February 23, 2026

Abstract

Data-driven neural models for passive sonar can be accurate yet difficult to interpret and diagnose because propagation and source/receiver effects become entangled. We propose Physics-aligned Sonar Kolmogorov–Arnold Network (SonarKAN), an architecture that aligns Kolmogorov–Arnold networks with the additive structure of the passive sonar equation. By constraining the network to B-spline univariate pathways, SonarKAN explicitly disentangles transmission loss from source spectral signatures. Furthermore, we introduce a physics-informed ridge projection that injects spreading-loss priors to accelerate convergence. Controlled simulations demonstrate accurate component recovery and substantially improved robustness to label noise compared to a multilayer perceptron (MLP) baseline.

1 Introduction

Many underwater acoustic inference tasks exhibit a low-dimensional structure. In the logarithmic (dB) domain, passive sonar relations express received levels and detection margins as sums and differences of physically interpretable terms (source level, transmission loss, noise level, directivity, and detection threshold) [9]. Generic deep networks, by contrast, often mix these contributions inside latent feature representations, which complicates attribution, model diagnostics, and domain transfer across environments. This study formulates a physics-aligned Sonar Kolmogorov–Arnold Network (SonarKAN). SonarKAN maps each input variable through a dedicated, directly inspectable univariate pathway and combines pathways by addition. Each pathway is parameterized by compactly supported B-splines, providing smoothness control, locality, and direct visual inspection of learned components [3]. We further propose a physics-informed initialization that projects a baseline transmission-loss model into the spline space and fixes the additive gauge via centering [4]. A controlled surrogate

aligned with the passive sonar equation is used to assess component recovery, convergence behavior, and robustness under additive label noise.

2 SonarKAN formulation

2.1 From Kolmogorov–Arnold representations to additive edge models

The Kolmogorov–Arnold representation theorem states that any continuous multivariate function on a compact domain can be represented as a finite superposition of continuous univariate functions and addition [7]. A canonical form is

$$f(\boldsymbol{x}) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right), \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$ and Φ_q and $\varphi_{q,p}$ are continuous univariate functions. Recent Kolmogorov–Arnold Networks (KANs) employ this principle, replacing fixed activation functions with learnable univariate edge functions [8]. SonarKAN adopts this edge-function approach but imposes an explicit additive decomposition to ensure physical interpretability. In a single-layer setting, we adopt the following structured additive model:

$$\hat{y}(\boldsymbol{x}) = b + \sum_{i=1}^{d} \varphi_i(x_i), \tag{2}$$

where d is the number of input variables, each φ_i is a learnable univariate function, and b is a scalar offset. Equation (2) is a nonlinear generalization of generalized additive models (GAMs) [4] and neural additive models [1]; however, SonarKAN uniquely parameterizes each φ_i using a compactly supported spline basis.

2.2 B-spline parameterization of univariate edge functions

Each univariate component is represented as a residual linear term plus a spline expansion:

$$\varphi(x) = ax + c + \sum_{m=0}^{M-1} w_m \, B_{m,p}(x), \tag{3}$$

where $\{B_{m,p}\}$ are B-spline basis functions of degree p on a knot vector $\boldsymbol{t} = (t_0, \ldots, t_{M+p})$ [3]. Each physical input is mapped affinely to $x \in [0, 1]$ prior to spline evaluation. Open-uniform (clamped) knots are used on $[0, 1]$: $t_0 = \cdots = t_p = 0$, $t_M = \cdots = t_{M+p} = 1$, with interior knots uniformly spaced. The

Cox–de Boor recursion defines the basis functions [2, 3] as

$$B_{m,0}(x) = \begin{cases} 1, & t_m \le x < t_{m+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

$$B_{m,p}(x) = \frac{x - t_m}{t_{m+p} - t_m} B_{m,p-1}(x) + \frac{t_{m+p+1} - x}{t_{m+p+1} - t_{m+1}} B_{m+1,p-1}(x), \tag{5}$$

with the convention that terms with zero denominators are treated as zero. At the right endpoint $x = t_{M+p}$ (i.e., $x = 1$ under normalization), we adopt the standard convention that the final knot span is closed on the right so that the basis functions form a partition of unity on $[0, 1]$ (consistent with standard spline-library conventions). B-splines have compact support, i.e., $B_{m,p}(x) = 0$ outside $[t_m, t_{m+p+1}]$ [3]. For fixed degree p, at most $p + 1$ basis functions are nonzero at any x, which yields locality in both function representation and parameter influence.

2.3 Approximation and identifiability

Approximation property. Let $g \in C^{p+1}([0, 1])$ and let $S_{p,M}$ denote the space spanned by degree-p B-splines with M bases on open-uniform knots with maximum knot span h. Standard spline approximation results imply that there exists $s \in S_{p,M}$ such that [3]

$$\|g - s\|_\infty \le C\, h^{p+1}\, \|g^{(p+1)}\|_\infty, \tag{6}$$

for a constant C independent of h. For an additive ground truth $g(\boldsymbol{x}) = \sum_i g_i(x_i)$, the approximation error of Eq. (2) is bounded by the sum of the corresponding univariate errors, providing an explicit accuracy–interpretability trade-off through M and p.

Identifiability (centering constraints). The decomposition in Eq. (2) is identifiable only up to additive constants: for any component index i and any constant δ, the transformation $\varphi_i \leftarrow \varphi_i + \delta$ and $b \leftarrow b - \delta$ leaves $\hat{y}$ unchanged. As in classical additive modeling [4], identifiability is enforced by centering each component on a reference measure (in practice, a dense grid or the empirical training distribution) and absorbing the removed mean into b.

3 Physics-consistent architecture via aligned sonar-equation

A standard passive sonar relation in decibels can be expressed as [9]

$$\mathrm{SE}(r, f, \theta) = \mathrm{SL}(f) - \mathrm{TL}(r, f) - \mathrm{NL}(f) + \mathrm{DI}(\theta) - \mathrm{DT} \tag{7}$$

where r is source–receiver range, f is frequency, and θ is bearing. Here SL is source level, TL is transmission loss, NL is noise level, DI is directivity index,

and DT is the detection threshold. Equation (7) is additive in the dB domain and motivates an architecture with term-wise pathways. For inputs (r, f, θ), SonarKAN takes the form

$$\hat{y}(r, f, \theta) = b + \varphi_r(r) + \varphi_f(f) + \varphi_\theta(\theta), \tag{8}$$

with the interpretation that φ_r captures the dominant range dependence of $-\mathrm{TL}$, φ_f captures the net spectral term $\mathrm{SL}(f) - \mathrm{NL}(f)$, φ_θ captures $\mathrm{DI}(\theta)$, and b captures the constant offset associated with DT.

Remark on separability. In general, $\mathrm{TL}(r, f)$ is not strictly separable, particularly in shallow-water waveguides where coherent multipath interference induces strong frequency–range coupling (e.g., striation patterns governed by the waveguide invariant) [6]. Such interference structures depend on phase terms like $\exp(i\Delta k_{nm} r)$ where Δk_{nm} varies with frequency, violating the additive assumption $\varphi_r(r) + \varphi_f(f)$. However, the additive approximation remains physically justified when the quantity of interest is an *incoherently averaged* intensity (energy-flux density) or a band-/ensemble-averaged transmission loss, for which interference cross-terms tend to decorrelate and average out [6]. When fine-grained phase-coherent features are required, one may introduce a low-order residual interaction (e.g., a limited $\varphi_{rf}(r, f)$ term) or extend SonarKAN with additional Kolmogorov–Arnold layers. The controlled surrogate in Sec. 5 is constructed to have a dominant range-dependent transmission-loss component, allowing Eq. (2) to accurately reconstruct the underlying components.

Figure 1 summarizes the SonarKAN topology, the spline-based edge mechanism, and the term-wise mapping to Eq. (7).

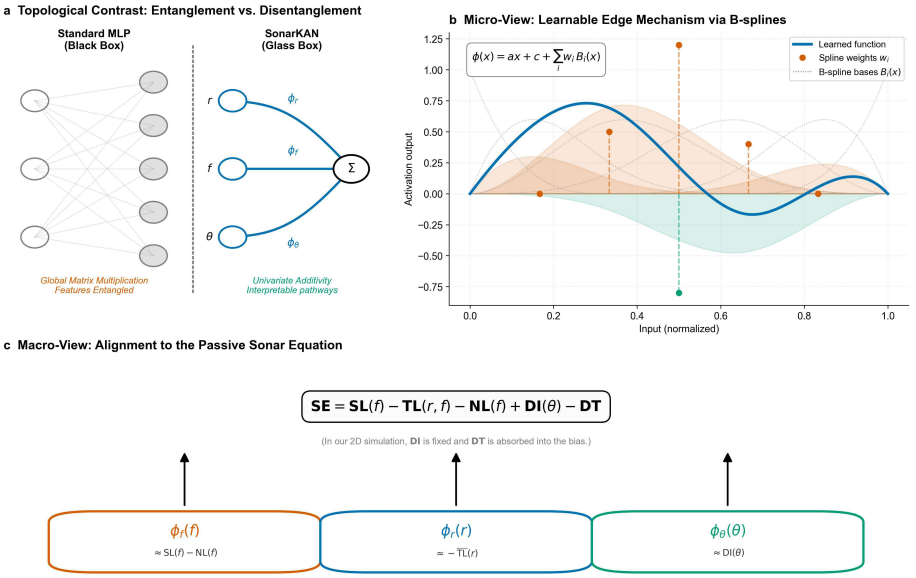

Figure 1: Operational principle of SonarKAN. (a) *Topology:* SonarKAN replaces dense mixing with additive pathways, enabling disentanglement by construction. (b) *Micro-mechanism:* each edge implements a learnable univariate function as a residual linear term plus a B-spline expansion (Eqs. (3)–(5)). (c) *Physical alignment:* via a first-layer identity mapping, SonarKAN aligns term-wise with the passive sonar equation (Eq. (7)), enabling direct interpretation of φ_r, φ_f, and φ_θ.

4 Physics-informed initialization as spline projection

Because SonarKAN uses explicit univariate representations, baseline physical models can be directly injected into selected components. For the range pathway, a standard first-order transmission-loss baseline is spherical spreading plus absorption [9]:

$$\mathrm{TL}_{\mathrm{base}}(r; f_c) = 20 \log_{10}\left(\frac{r}{r_0}\right) + \alpha\left(\frac{f_c}{10^3}\right) \frac{r}{10^3}, \tag{9}$$

where r is in meters, $r_0 = 1\,\mathrm{m}$ is the reference distance, f_c is a nominal center frequency in Hz, and $\alpha(\cdot)$ is an absorption coefficient in dB/km with frequency argument expressed in kHz. Any additive constant arising from the choice of r_0 is absorbed into b under the centering constraints in Sec. 2.3. The $20 \log_{10}(r/r_0)$ term follows from spherical spreading of pressure amplitude, $p(r) \propto 1/r$, together with the decibel definition for amplitudes [9]. In our experiment, α is

computed using the Thorp approximation [10]:

$$\alpha(f) = \frac{0.11 f^2}{1+f^2} + \frac{44 f^2}{4100+f^2} + 2.75 \times 10^{-4} f^2 + 0.003 \quad \text{dB/km}, \tag{10}$$

with f in kHz.

Projection onto spline coefficients. To initialize φ_r as an approximation to $-\mathrm{TL}_{\mathrm{base}}$, normalized range values $\{x_n\}_{n=1}^{N}$ are sampled on $[0,1]$ and the B-spline basis matrix $\boldsymbol{B} \in \mathbb{R}^{N \times M}$ is evaluated at $\{x_n\}$. Let $\boldsymbol{\theta} = [a \; c \; w_0 \; \ldots \; w_{M-1}]^\top$ collect the parameters of Eq. (3) and define the design matrix $\boldsymbol{A} = [\boldsymbol{x} \; \boldsymbol{1} \; \boldsymbol{B}]$. A ridge-regularized least-squares projection [5] is computed in coefficient space:

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_2^2 \;=\; (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^\top \boldsymbol{y}, \tag{11}$$

where $\boldsymbol{y}$ samples $-\mathrm{TL}_{\mathrm{base}}$ on the grid and $\lambda > 0$ is a small regularization parameter. The resulting component is then mean-centered to satisfy the identifiability constraints in Sec. 2.3.

5 Controlled simulation study

5.1 Sonar-equation-aligned surrogate

To validate interpretability and physics alignment under controlled conditions, we employ a two-input surrogate model with labels defined as:

$$y(r, f) = \mathrm{SL}_0 - \mathrm{TL}(r) + S(f) + \epsilon, \tag{12}$$

where SL_0 is a constant source level, $\mathrm{TL}(r)$ is a range-dependent transmission-loss term, $S(f)$ is a frequency-dependent spectral signature term, and ϵ is additive Gaussian noise in dB. The surrogate is constructed to satisfy the additive separability assumed by Eq. (2): TL depends only on r (absorption evaluated at fixed f_c), while S depends only on f. An optional extension (released with the code) introduces weak r–f coupling in the multipath residual and performs incoherent band averaging before conversion to dB, enabling stress-tests beyond strict separability. Ranges are sampled as $r \sim \mathcal{U}([100\,\mathrm{m}, 5000\,\mathrm{m}])$ and frequencies as $f \sim \mathcal{U}([1\,\mathrm{kHz}, 5\,\mathrm{kHz}])$. The ground-truth components are

$$\mathrm{TL}(r) = 20 \log_{10}\left(\frac{r}{r_0}\right) + \alpha\left(\frac{f_c}{10^3}\right) \frac{r}{10^3} + A \sin\left(\frac{2\pi r}{P} + \phi\right) \exp\left(-\frac{r}{R_d}\right), \tag{13}$$

$$S(f) = S_0 - D \exp\left(-\frac{(f - f_0)^2}{2\sigma^2}\right), \tag{14}$$

where $\alpha(\cdot)$ is given by Eq. (10) (frequency argument in kHz), $r_0 = 1\,\mathrm{m}$, and f_c is a fixed center frequency in Hz. The sinusoidal residual models a range-localized interference pattern.

Unless stated otherwise, the parameters are $\mathrm{SL}_0 = 180\,\mathrm{dB}$, $f_c = 3000\,\mathrm{Hz}$, $(A, P, R_d, \phi) = (4\,\mathrm{dB},\ 450\,\mathrm{m},\ 5000\,\mathrm{m},\ \pi/4)$, and $(S_0, D, f_0, \sigma) = (15\,\mathrm{dB},\ 12\,\mathrm{dB},\ 3500\,\mathrm{Hz},\ 150\,\mathrm{Hz})$. In all plots, components are reported after mean-centering to respect identifiability.

5.2 Models, training, and robustness metric

SonarKAN is fit in the additive form $\hat{y}(r, f) = b + \varphi_r(r) + \varphi_f(f)$ using cubic splines ($p = 3$) with $M = 23$ bases per component and open-uniform knots on $[0, 1]$. Two variants are considered to isolate the effect of physical priors: (i) *SonarKAN (physics init)* initializes φ_r by the spline projection in Eq. (11) using a 256-point range grid and $\lambda = 10^{-6}$, while (ii) *SonarKAN (random init)* uses the same architecture with default parameter initialization. During training, centering constraints are enforced after each optimizer step on a fixed reference grid (200 points per variable) and the removed constants are compensated in b.

Data and optimization. For each random seed, $N_{\mathrm{train}} = 2000$ noiseless training samples and $N_{\mathrm{test}} = 1000$ noiseless test samples are drawn uniformly over the ranges stated above. All models are trained for 200 epochs using full-batch Adam with learning rate 2×10^{-2} and mean-squared error loss. As a baseline, a small multilayer perceptron (MLP) with architecture $2 \to 16 \to 1$ (ReLU activation) is trained under the same optimizer and epoch budget. Reported curves show mean±standard deviation over five random seeds.

Parameter counts. With d inputs and M spline bases per component, the one-layer SonarKAN in Eq. (2) has $d(M + 2) + 1$ trainable parameters (each φ_i has M spline weights plus (a, c), and b adds one). For $(d, M) = (2, 23)$ this yields 51 parameters. The MLP baseline has 65 parameters.

Noise-robustness metric. To quantify robustness under label corruption, a label signal-to-noise ratio (SNR) in dB is defined and the label-noise variance is set as

$$\sigma_\epsilon^2 = \frac{\mathrm{Var}(y)}{10^{\mathrm{SNR}/10}}. \tag{15}$$

Here $\mathrm{Var}(y)$ is computed from the noise-free labels under the same sampling distribution (in practice, from the clean training set used in our experiments). For each SNR level, noisy test labels are generated and the test root-mean-square error (RMSE) is reported.

5.3 Results

Figure 2 reports the controlled surrogate results. SonarKAN recovers the spreading-plus-absorption trend in φ_r and captures the imposed range-localized oscillation, while φ_f recovers the spectral notch in Eq. (14). Physics-informed initialization provides a clear head start in optimization and converges faster than both a

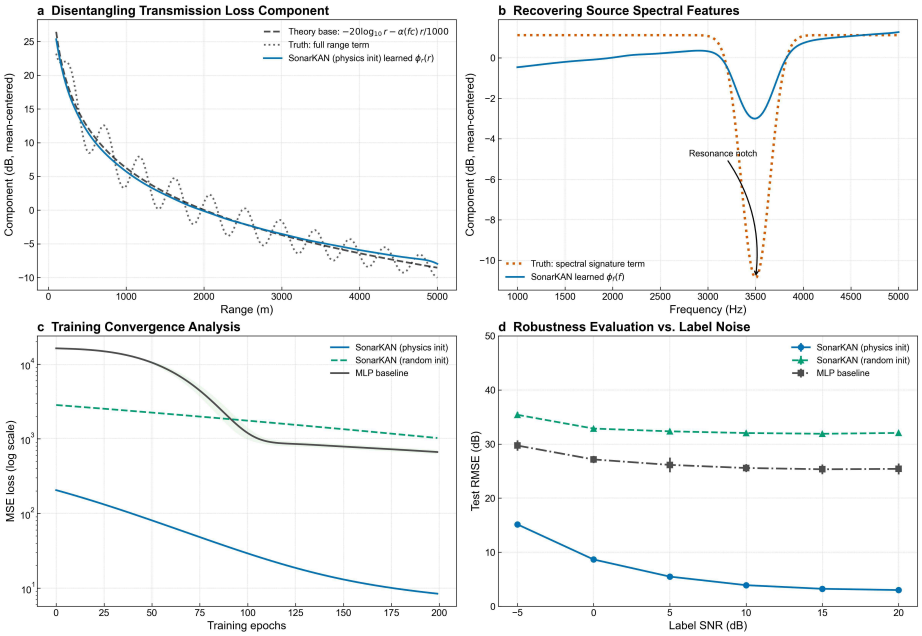

Figure 2: Controlled surrogate validation (mean-centered components). (a) Learned range component $\varphi_r(r)$ compared with the theoretical base term $-20\log_{10}(r/r_0) - \alpha(f_c/10^3)\, r/10^3$ and the full ground-truth range term in Eq. (13). (b) Learned frequency component $\varphi_f(f)$ compared with the ground-truth spectral signature term in Eq. (14). (c) Training loss versus epochs (mean±standard deviation over seeds) for SonarKAN with physics-informed initialization, SonarKAN with random initialization, and the MLP baseline. (d) Test RMSE versus label SNR computed using Eq. (15); error bars indicate ±1 standard deviation across seeds.

random-initialized SonarKAN and the MLP under identical training budgets. Across the SNR sweep, SonarKAN exhibits lower test RMSE than the MLP baseline, indicating improved robustness under additive label noise.

6 Discussion and limitations

SonarKAN is designed for settings where the target quantity in the dB domain admits an approximately additive decomposition with respect to physically meaningful variables (e.g., range and frequency). In practical passive-sonar processing, this additivity is often most defensible for band-averaged or incoherently averaged level estimates, where modal/eigenray interference microstructure is suppressed by averaging in frequency, time, or environment [6]. In strongly coherent shallow-water regimes, SonarKAN should therefore be interpreted as learning an effective coarse-grained decomposition rather than a

phase-resolving model of $\mathrm{TL}(r, f)$. When coherent range–frequency striations are diagnostically important, SonarKAN can be extended in a controlled manner by adding a low-order interaction term (e.g., $\varphi_{rf}(r, f)$) or by composing multiple Kolmogorov–Arnold layers. Such extensions preserve interpretability while relaxing strict separability.

Acknowledgements

We appreciate the guidance from professors at the Institute of Acoustics (Chinese Academy of Sciences) and engineers from China State Shipbuilding Corporation Limited.

Author Declarations

This work was supported by the Outstanding Engineer Growth Program. The authors declare no competing interests.

Data Availability

Code, configuration files, and scripts to reproduce the modelings and simulations including Figs. 1–2 are available at `https://github.com/soundai2016/SonarKAN`.

References

[1] Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. *Advances in neural information processing systems*, 34:4699–4711, 2021.

[2] Maurice G Cox. The numerical evaluation of b-splines. *IMA Journal of Applied mathematics*, 10(2):134–149, 1972.

[3] Carl De Boor. On calculating with b-splines. *Journal of Approximation theory*, 6(1):50–62, 1972.

[4] Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical science*, 1(3):297–310, 1986.

[5] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[6] Finn B Jensen, William A Kuperman, Michael B Porter, Henrik Schmidt, and Alexandra Tolstoy. *Computational ocean acoustics*, volume 2011. Springer, 2011.

[7] Andrei Nikolaevich Kolmogorov. On the representations of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Dokl. Akad. Nauk USSR*, volume 114, pages 953–956, 1957.

[8] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024.

[9] Xavier Lurton. *An introduction to underwater acoustics: principles and applications*. Springer Science & Business Media, 2002.

[10] William H Thorp. Analytic description of the low-frequency attenuation coefficient. *The Journal of the Acoustical Society of America*, 42(1):270–270, 1967.