Structure in the Noise

Robust Representation Learning for Physical Time-Series Data

Elena Vasquez

Ph.D. Candidate, Department of Computer Science, Vertex University

Faculty Job Talk ▪ Department of Computer Science, Nimbus Institute of Technology ▪ November 12, 2026

Agenda

Roadmap for This Talk

Scientific Instruments Are Drowning in Noise

Motivation

Modern physical sensing networks – seismic arrays, climate buoys, wearable biosensors – generate high-frequency time series where the signal of interest is buried under sensor drift, environmental interference, and intermittent dropout [1].

- Deep models trained on clean benchmarks **overfit the noise floor**, not the underlying physics.
- Standard self-supervised objectives assume corruption is *random*, when it is highly **structured**.
- Apex Environmental Labs loses an estimated 22% of readings to this gap every field season.

Working Definition

A representation is *robust* if downstream performance degrades gracefully, not catastrophically, as corruption severity increases.

From Raw Signal to Robust Representation

A Structural View of the Problem

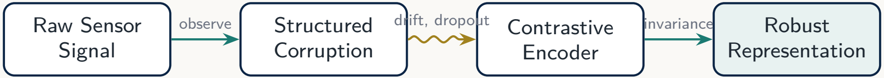


Figure 1: The corruption model is estimated from deployment logs rather than assumed, so the encoder learns invariances that match how instruments actually fail.

My Research Statement

Three Thrusts, One Question

Central Question

How do we build representations of physical time series that stay useful when the deployment environment stops looking like the training set?

1. **Noise-robust self-supervised learning** – encode structured, instrument-specific corruption directly into the pretraining objective.
2. **Uncertainty quantification under distribution shift** – give downstream users calibrated confidence, not just a point prediction.
3. **Label-efficient transfer** – make these representations useful to domain scientists who can label dozens of examples, not millions.

Contribution I: Structured Contrastive Denoising

Vasquez et al., 2024

Structured Contrastive Denoising (SCD) fits a lightweight *corruption model* from unlabeled deployment logs and uses it to generate contrastive pairs that mimic real instrument failure modes [2].

- Three-state corruption model: sensor drift, clipping, burst dropout.
- Pulls together views of the *same* signal under different corruptions.
- No labels needed at pretraining time – only raw logged readings.

Method	Meridian-Seismic F1	Apex-Climate F1	Nimbus-Bio F1	Robustness Drop ↓
Baseline CNN	0.71	0.68	0.74	18.4%
Standard SSL (SimCLR)	0.76	0.73	0.79	13.1%
Ours (SCD)	**0.85**	**0.82**	**0.88**	**5.6%**

Table 1: F1 under deployment-level corruption, averaged across three sites.

Contribution II: Calibrated Forecasting under Shift

Vasquez and Okafor, 2025

A robust representation is only useful if the model *knows when it does not know*. We combine conformal prediction with the encoder from Contribution I to produce calibrated intervals that hold up as corruption severity increases [3, 4].

- Wraps any point forecaster with distribution-free coverage guarantees.
- Calibration set is re-weighted using the encoder's own corruption-severity estimate.
- Deployed on six months of live Apex-Climate buoy data.

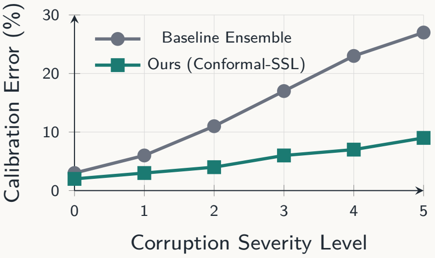
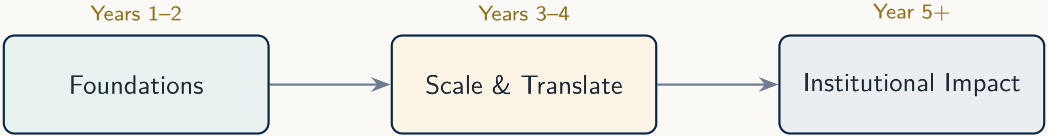


Figure 2: Calibration error stays flat under increasing corruption because the encoder's severity estimate re-weights coverage.

A Five-Year Research Program

From Robust Encoders to a Lab



Concrete Milestones

- **Years 1–2:** generalize the structured-corruption encoder to a multi-modal foundation model spanning seismic, climate, and biosensor streams.
- **Years 3–4:** field trial with Apex Environmental Labs; target an NSF CAREER proposal on label-efficient scientific monitoring.
- **Year 5+:** release `OpenSensor`, an open-source toolkit, and graduate the first two Ph.D. students working on this agenda.

Teaching Philosophy

Build Before You Prove

I teach the way I do research: students get their hands on a noisy, real dataset before they see the clean theorem that explains it.

- **Project-based assessment:** every course ends on an unseen dataset.
- **Scaffolded mentoring:** weekly office hours around a research log.
- **Inclusive access:** open-licensed notes and code, no paid software.

Course	Level	Description
CS 220: Data Structures (redesign)	Undergraduate	Labs rebuilt around real sensor logs
CS 331: Signals & Learning	Undergraduate	New elective bridging DSP and machine learning
CS 731: Robust Representation Learning	Graduate Seminar	Reading course built from this talk's agenda

Why Nimbus Institute of Technology

Fit

- The Signal Processing group here already builds the sensing hardware my encoders are trained on – an unusually short path from lab to field.
- Nimbus's field sites with Apex Environmental Labs are a ready-made deployment target for the Year 3–4 milestones on the previous slide.
- The Nimbus Data Science Institute's seed-grant program is a natural first step toward the CAREER proposal.

What I Bring

Two first-author contributions with code and data released, a concrete five-year plan, and a teaching record built around the same project-based philosophy this department already values.

Thank You

Questions Welcome

Structure in the Noise

Robust Representation Learning for Physical Time-Series Data

evasquez@vertex.edu

www.vertex.edu/~evasquez

github.com/evasquez-lab

Joint work with collaborators at Vertex University and Apex Environmental Labs.

References I

[1] Marcus Chen and Priya Sharma. "Failure Modes of Deployed Seismic Sensor Networks: A Five-Year Retrospective". In: *Synapse Review of Earth Sensing* 6.1 (2023), pp. 12–29.

[2] Elena Vasquez and Daniel Okafor. "Structured Contrastive Denoising for Physical Sensor Time Series". In: *Proceedings of the Meridian Conference on Machine Learning (MCML)*. Meridian Computing Society, 2024, pp. 214–227.

[3] Elena Vasquez and Daniel Okafor. "Calibrated Forecasting under Distribution Shift with Conformal Self-Supervised Representations". In: *Journal of Nimbus Data Science* 12.2 (2025), pp. 88–104.

[4] Sofia Diaz. "Distribution-Free Coverage Guarantees for Streaming Sensor Forecasts". In: *Transactions on Meridian Statistics* 19.4 (2025), pp. 455–471.