MBON Acoustic Indices Study: Results Viewer

May River Estuary — Modeling Results & Validation

AUTHOR
MBON-USC Team

PUBLISHED
January 31, 2026

Overview

The question: Can acoustic indices — automated summaries of soundscape characteristics — predict biological activity in an estuary?

The short answer: Yes, for presence/absence. Less reliably for activity levels.

> **Elevator Pitch**
>
> Acoustic indices can tell you *whether* biological activity is happening (presence), but they can't reliably tell you *how much* (activity levels). Vessel detection is excellent (AUC = 0.92), biological presence detection is moderate (AUC ~0.75), and activity/count prediction doesn't generalize well across time periods.

Study Design

- **Data:** 13,102 observations across 3 monitoring stations (9M, 14M, 37M), 2 years (2018, 2021), at 2-hour resolution
- **Predictors:** 17 acoustic indices + temperature, depth, time of day, day of year, station
- **Responses:** 9 metrics — fish (activity, richness, presence), dolphins (burst pulse, echolocation, whistle, activity, presence), vessels (presence)
- **Models:** Generalized Additive Mixed Models (GAMMs) with AR1 autocorrelation structure

---

Model Summary

Convergence & Technical Soundness

All 9 models converged successfully.

Metric	AR1 (ρ)	AIC	Type
fish_activity	0.196	33,897	Count
fish_richness	0.185	22,472	Count
fish_presence	0.144	9,413	Binary
dolphin_burst_pulse	0.064	8,187	Count
dolphin_echolocation	0.128	30,765	Count

Metric	AR1 (ρ)	AIC	Type
dolphin_whistle	0.107	3,189	Count
dolphin_activity	0.129	32,367	Count
dolphin_presence	0.110	10,949	Binary
vessel_presence	0.077	6,961	Binary

AR1 (ρ) interpretation: Measures temporal “stickiness” — how much knowing the current state tells you about the next 2-hour window. Values of 0.06–0.20 indicate modest persistence, which is appropriate for these data.

Which Indices Predict Which Responses?

The table below shows which acoustic indices significantly predict each response ($p < 0.05$).

Index	Fish Act.	Fish Rich.	Fish Pres.	Dolph. BP	Dolph. Echo	Dolph. Whis.	Dolph. Act.	Dolph. Pres.	Vessel
ACI		✓	✓	✓					✓
BI			✓	✓	✓		✓	✓	✓
BioEnergy			✓					✓	
ECV					✓	✓	✓	✓	
EPS_KURT		✓	✓	✓	✓		✓		
EVNtCount		✓	✓				✓	✓	✓
EVNtMean			✓					✓	
H_Havrda			✓	✓			✓		✓
KURTt			✓					✓	✓
NBPEAKS				✓	✓		✓	✓	
ROU		✓	✓		✓		✓		
SKEWt		✓							
TFSD				✓	✓	✓	✓	✓	✓
VARt			✓						

Index	Fish Act.	Fish Rich.	Fish Pres.	Dolph. BP	Dolph. Echo	Dolph. Whis.	Dolph. Act.	Dolph. Pres.	Vessel
ZCR	✓	✓	✓			✓			✓
nROI		✓	✓	✓	✓		✓	✓	✓

Key Patterns

- **Broadly predictive indices:** TFSD (6/9), nROI (7/9), BI (6/9), EVNtCount (5/9)
- **Fish-specific:** ZCR predicts all 3 fish metrics but no dolphin metrics
- **Dolphin-specific:** ECV predicts dolphins (4/5) but not fish
- **Presence is easier to predict:** Fish presence has 12 significant predictors; fish activity has only 1

Results by Metric

fish_activity

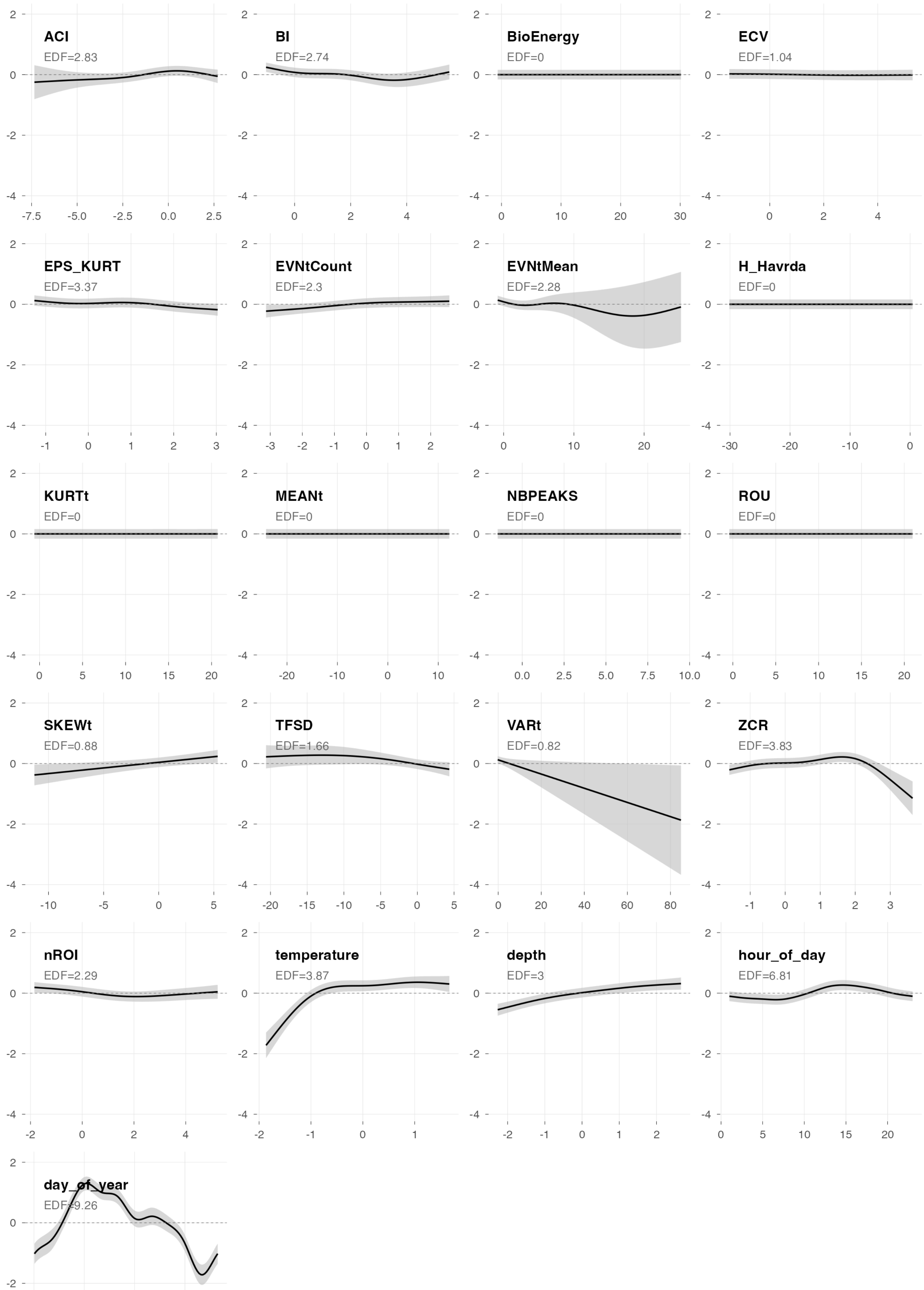
Response type: Count (negative binomial)

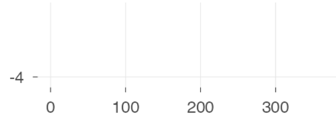
Key predictors: ZCR, hour_of_day, day_of_year

CV Performance: RMSE = 1.14, R^2 = 0.01 (poor generalization)

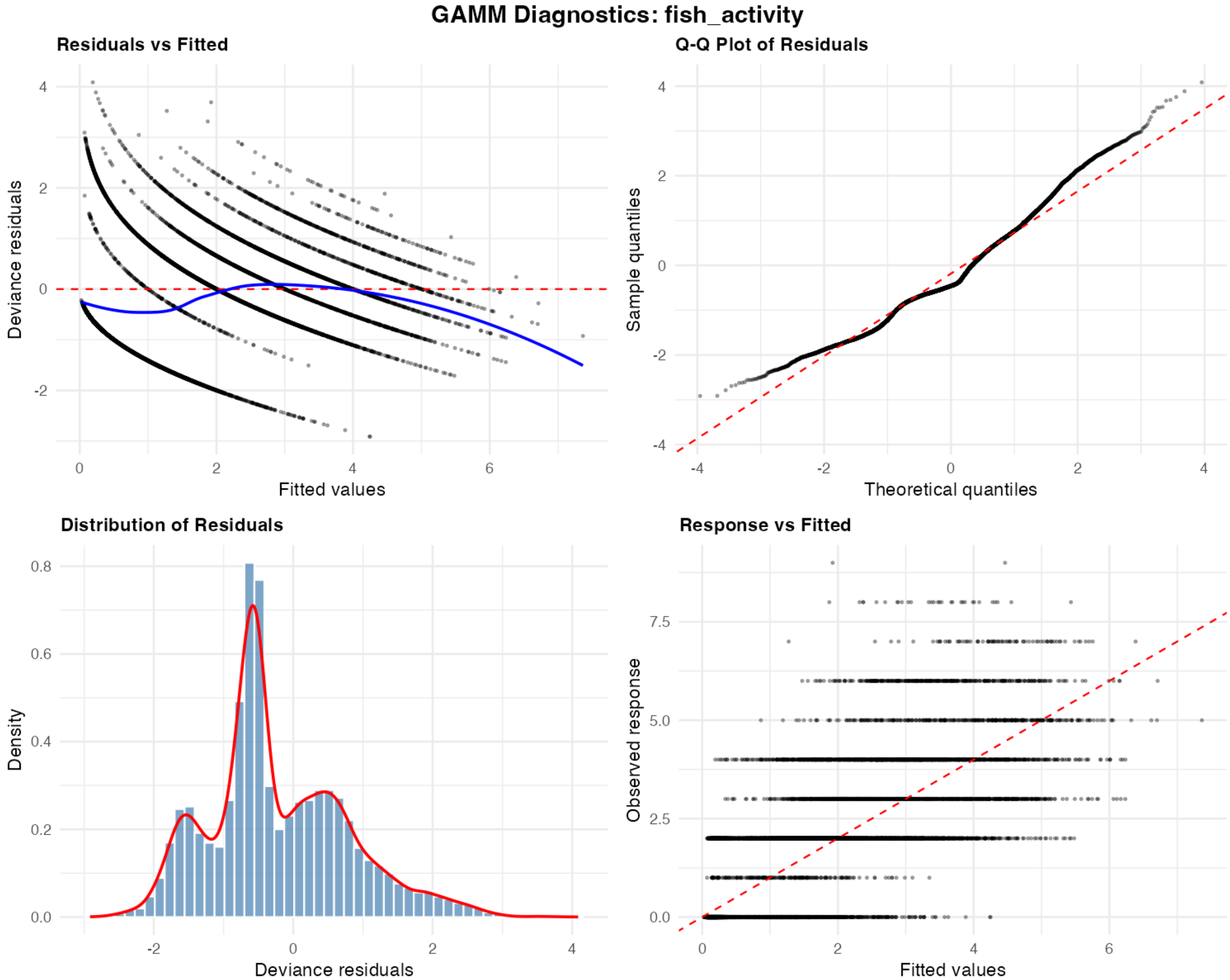
Smooth Terms Overview

fish_activity — Smooth Terms





Diagnostics



Interpretation

Fish activity shows strong diel patterns (hour_of_day EDF = 6.81) but weak index associations. Only ZCR is significant among the acoustic indices. The model fits the training data but doesn't generalize well to held-out weeks.

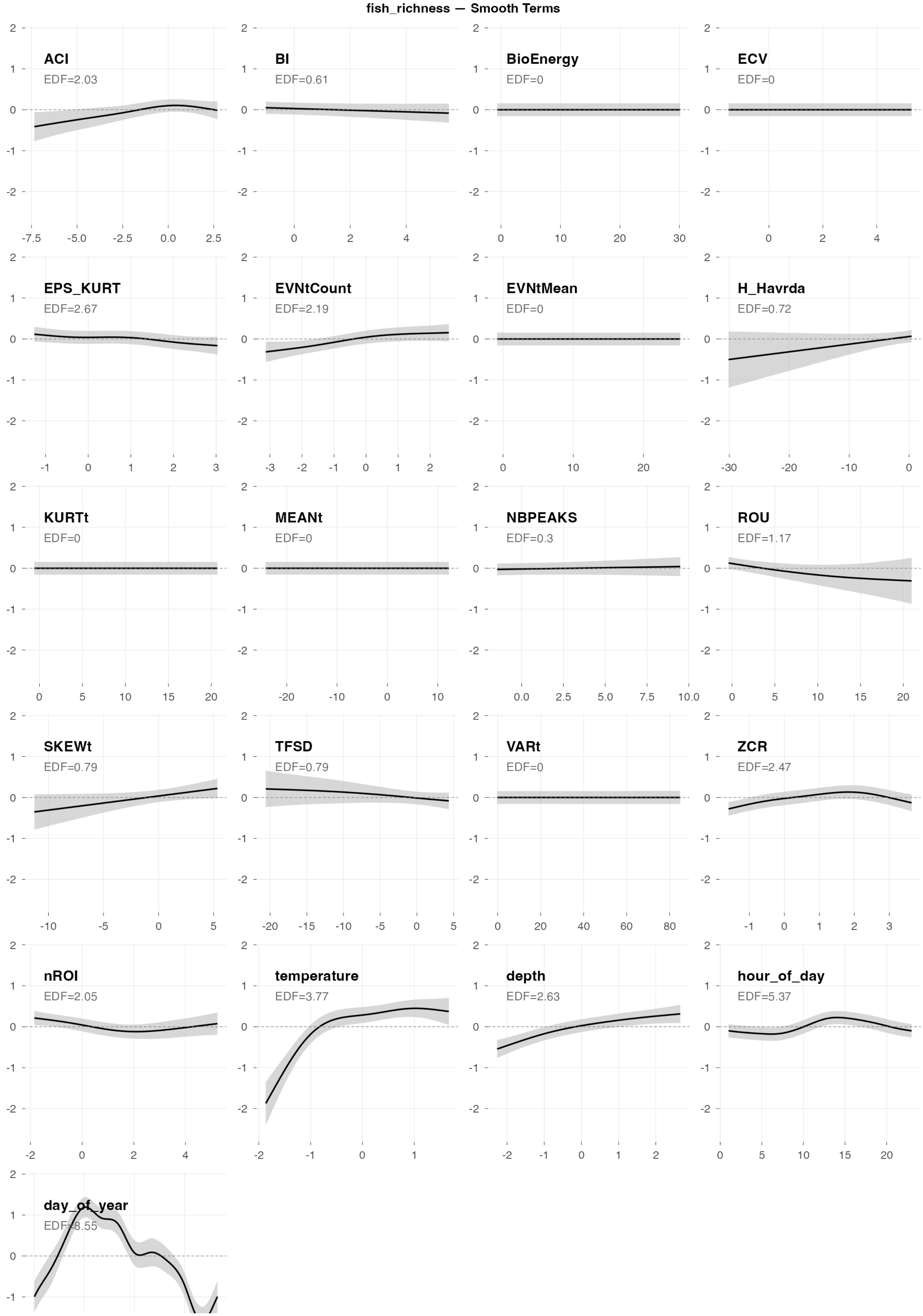
fish_richness

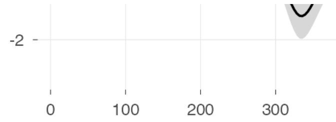
Response type: Count (negative binomial)

Key predictors: ACI, EPS_KURT, EVNtCount, ROU, SKEWt, ZCR, nROI, depth, hour_of_day

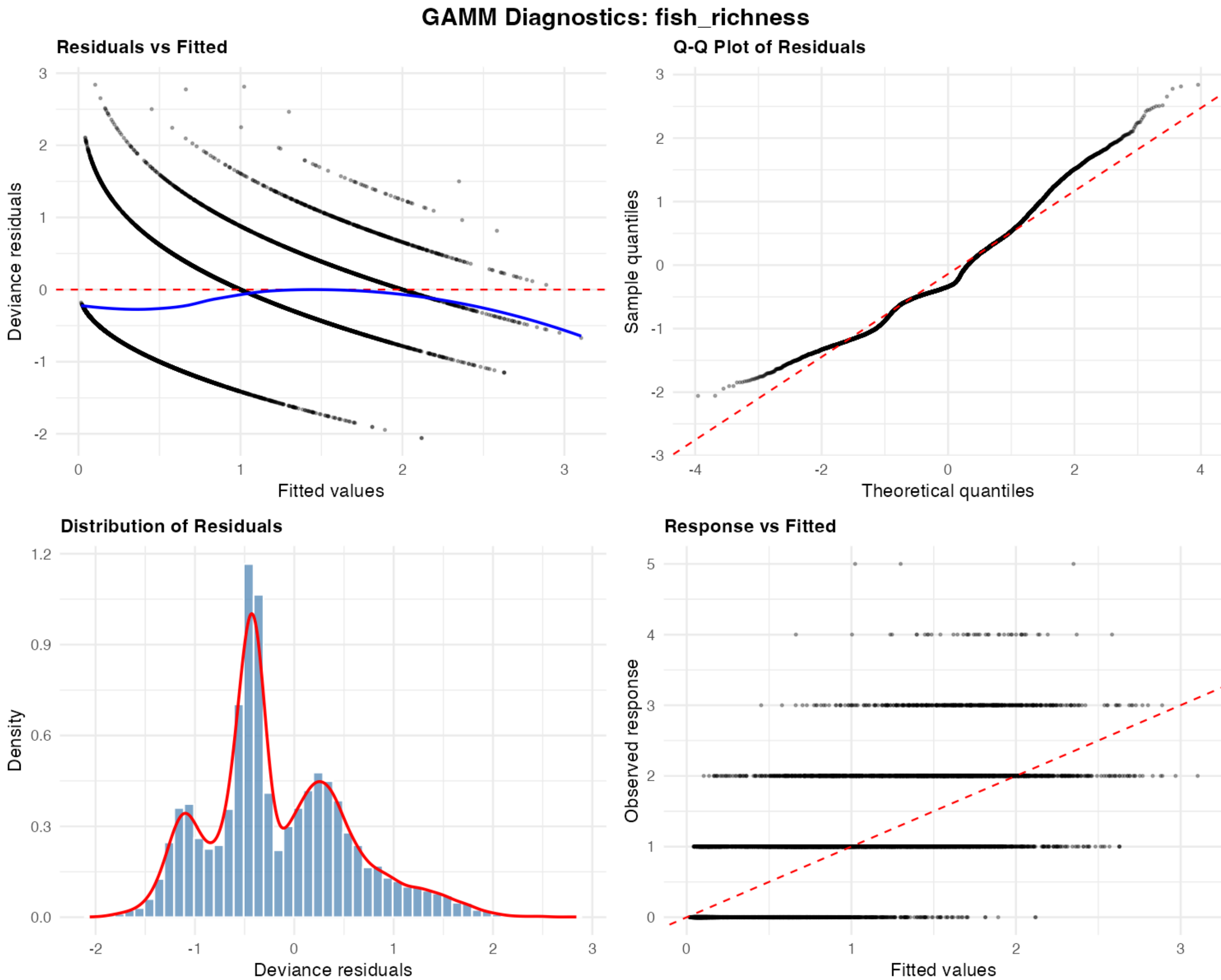
CV Performance: RMSE = 0.56, R^2 = 0.02 (low error, poor R^2)

Smooth Terms Overview


fish_richness — Smooth Terms
ACI
EDF=2.03
BI
EDF=0.61
BioEnergy
EDF=0
ECV
EDF=0
EPS_KURT
EDF=2.67
EVNtCount
EDF=2.19
EVNtMean
EDF=0
H_Havrda
EDF=0.72
KURTt
EDF=0
MEANt
EDF=0
NBPEAKS
EDF=0.3
ROU
EDF=1.17
SKEWt
EDF=0.79
TFSD
EDF=0.79
VARt
EDF=0
ZCR
EDF=2.47
nROI
EDF=2.05
temperature
EDF=3.77
depth
EDF=2.63
hour_of_day
EDF=5.37
day_of_year
EDF=8.55



Diagnostics



Interpretation

Fish richness has more significant index predictors than fish activity. Indices related to acoustic complexity (ACI, EVNtCount, nROI) are significant — consistent with the idea that more fish species create a more diverse soundscape. Depth (tidal proxy) is strongly significant: higher richness at high tide.

fish_presence

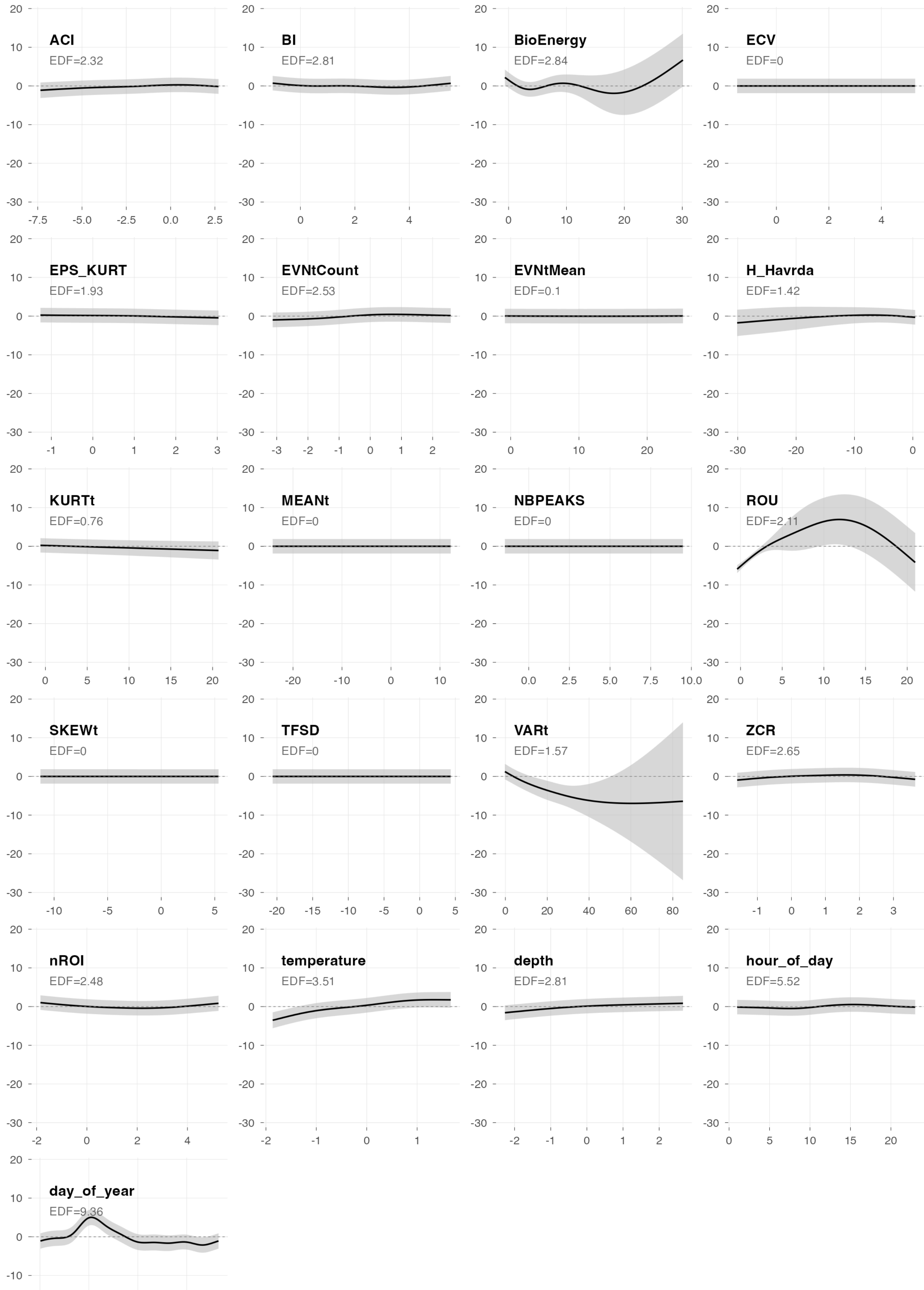
Response type: Binary (binomial)

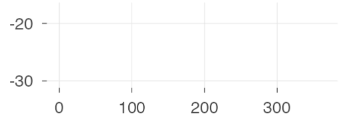
Key predictors: 12 of 17 indices significant, plus temperature, depth, hour_of_day, day_of_year, station

CV Performance: AUC = 0.75 (moderate — useful signal)

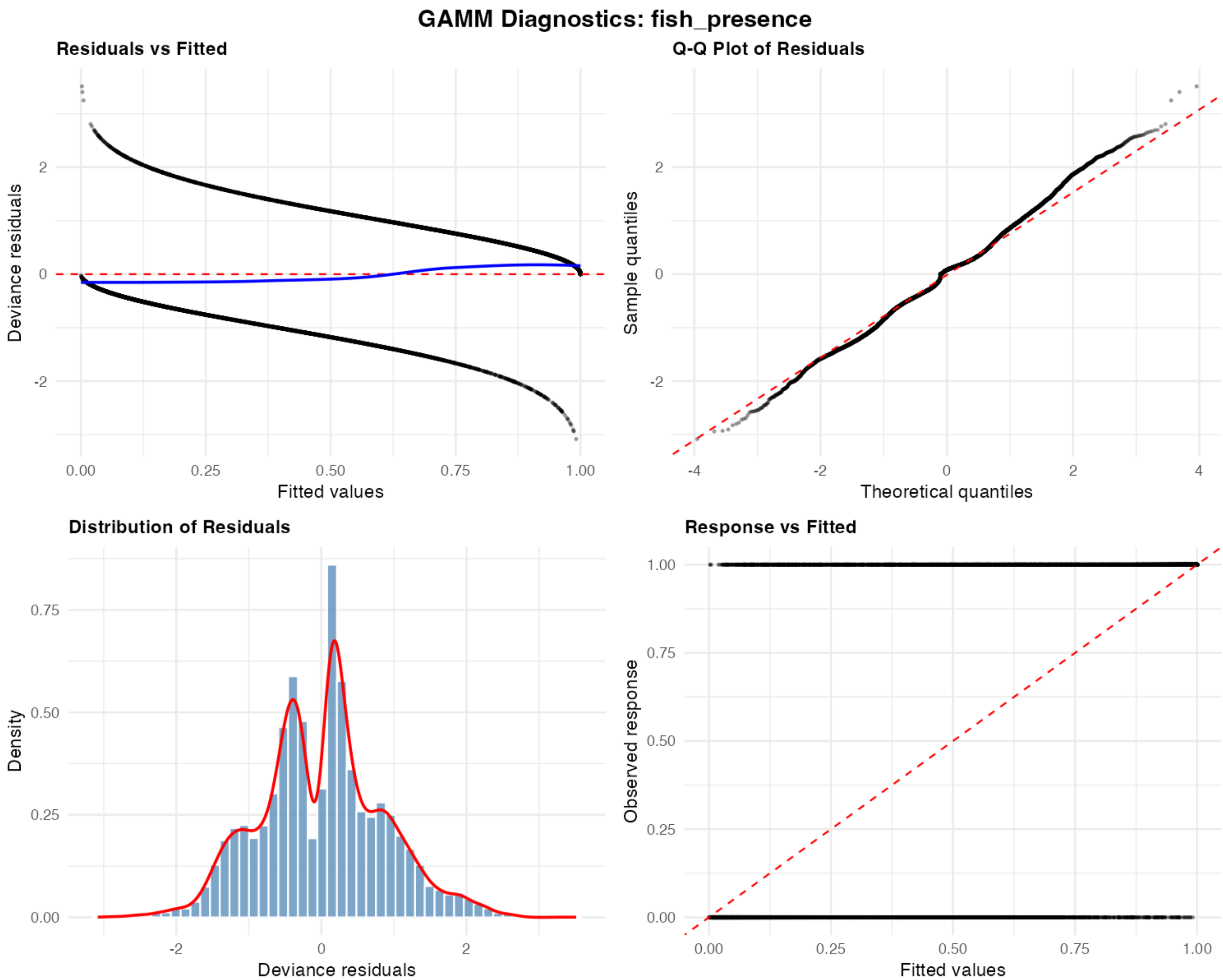
Smooth Terms Overview

fish_presence — Smooth Terms





Diagnostics



Interpretation

Fish presence is the most predictable fish metric. Many indices pick it up — detecting "something is there" is easier than quantifying how much. Strong seasonal pattern (day_of_year EDF = 9.36) with spring peak. The model generalizes moderately well (AUC = 0.75).

dolphin_burst_pulse

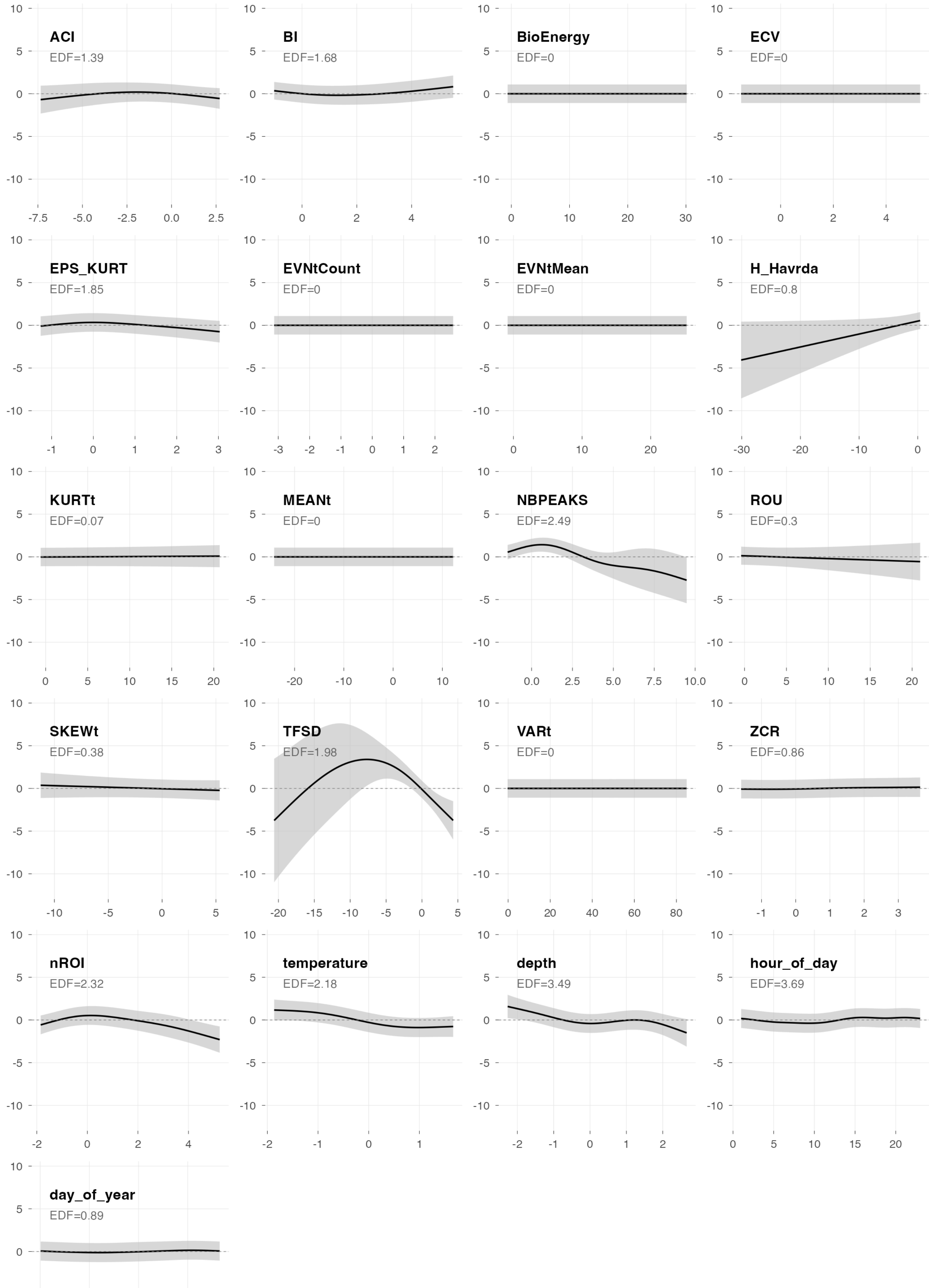
Response type: Count (negative binomial)

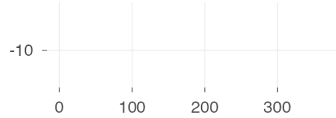
Key predictors: ACI, BI, EPS_KURT, H_Havrda, NBPEAKS, TFSD, nROI, temperature, depth, hour_of_day

CV Performance: RMSE = 1.54, R^2 = -0.13 (doesn't generalize)

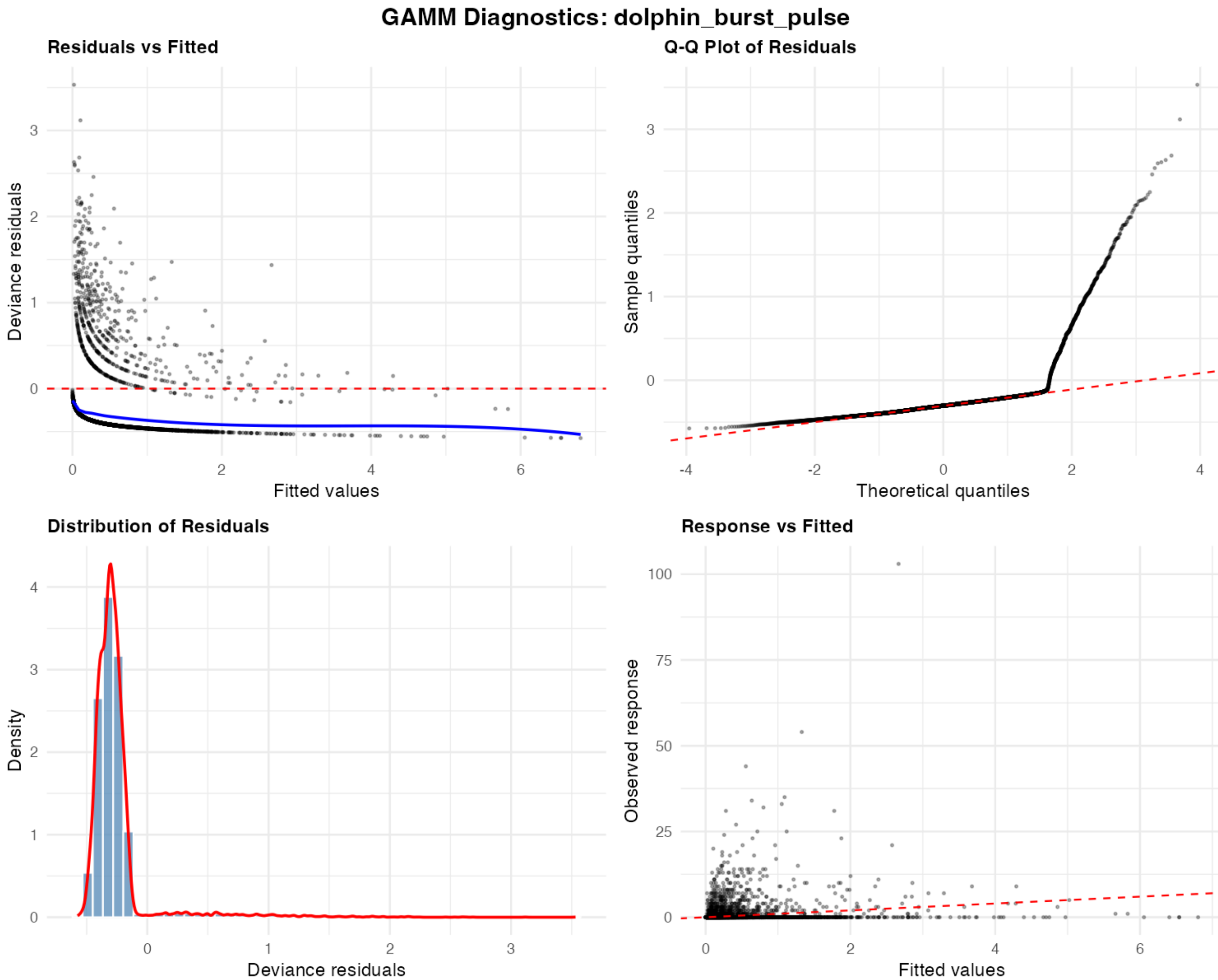
Smooth Terms Overview

dolphin_burst_pulse — Smooth Terms





Diagnostics



Interpretation

Burst pulses are foraging-related — discrete hunting events. Indices sensitive to impulsive sounds (TFSD, NBPEAKS, nROI) are predictive. Temperature is significant (more activity in cooler water). The model has the lowest AR1 (ρ = 0.06) — these are event-like, not persistent.

dolphin_echolocation

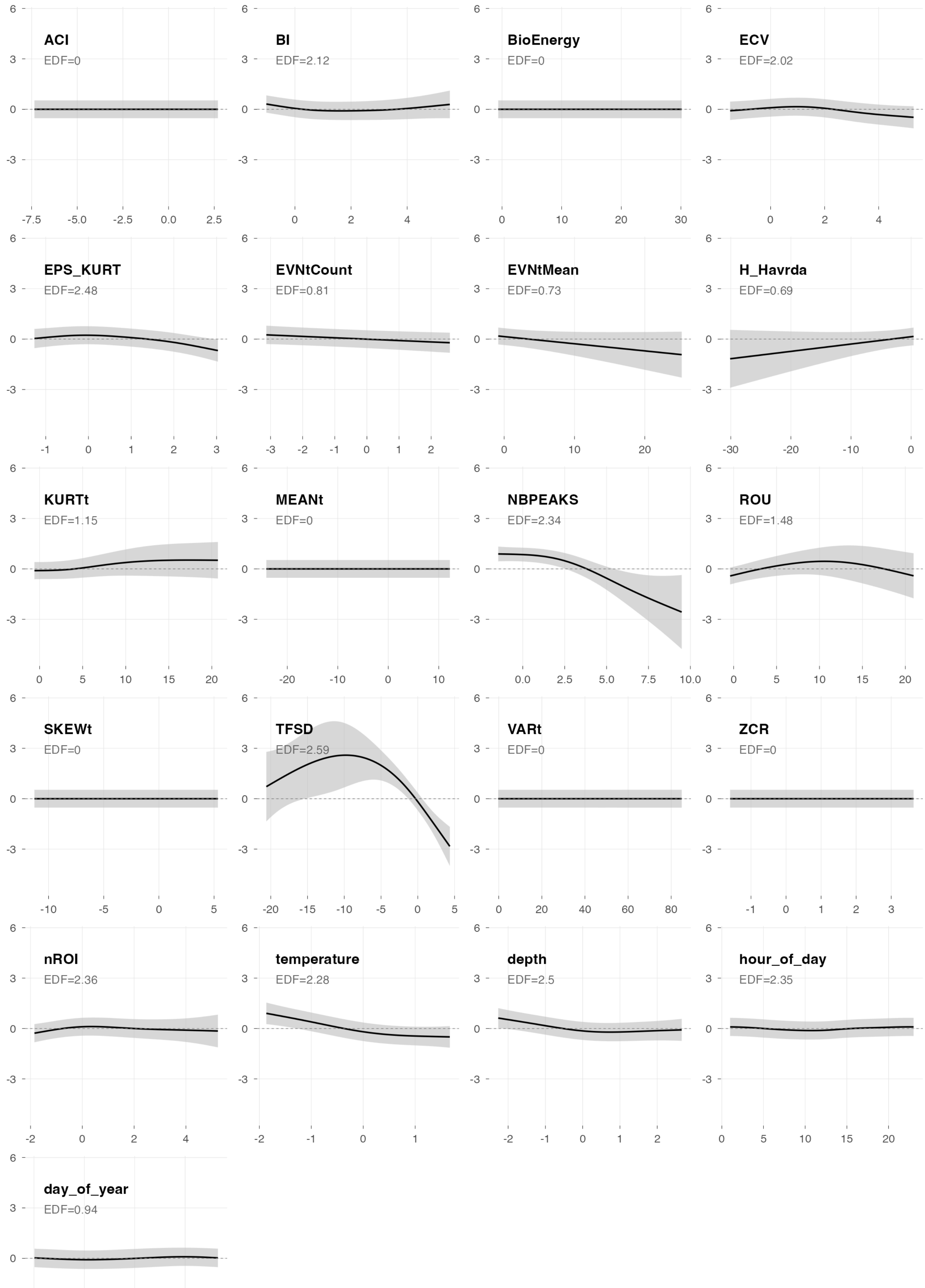
Response type: Count (negative binomial)

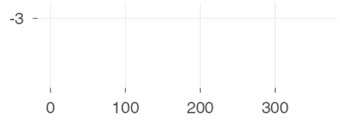
Key predictors: BI, ECV, EPS_KURT, NBPEAKS, ROU, TFSD, nROI, temperature, depth

CV Performance: RMSE = 11.0, R^2 = -455 (outlier folds dominating)

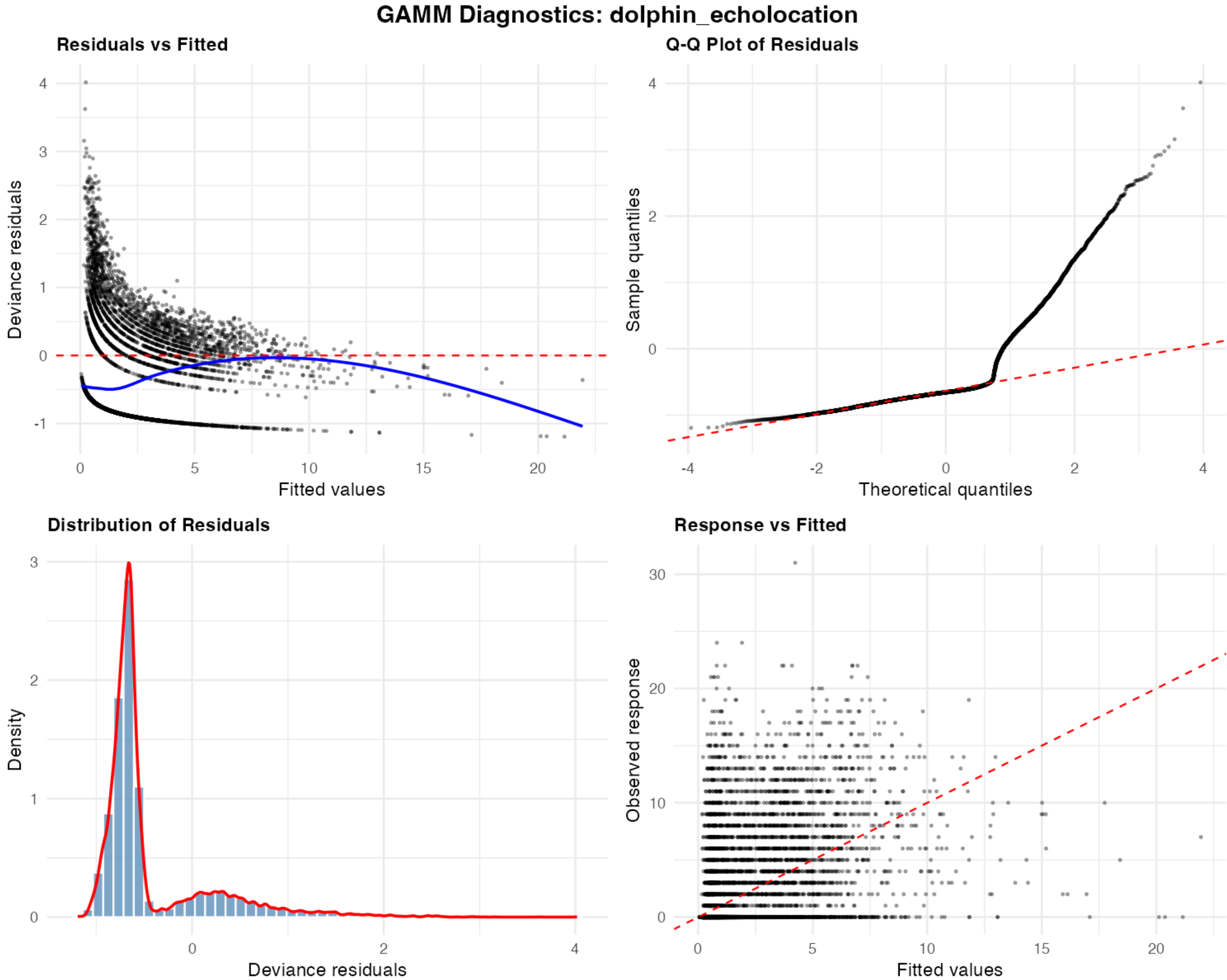
Smooth Terms Overview

dolphin_echolocation — Smooth Terms





Diagnostics



Interpretation

Echolocation clicks are broadband with distinctive temporal structure — reflected in significant BI, TFSD, NBPEAKS. The extremely negative R^2 in CV is due to a few outlier weeks with high activity; the model can't predict these spikes.

dolphin_whistle

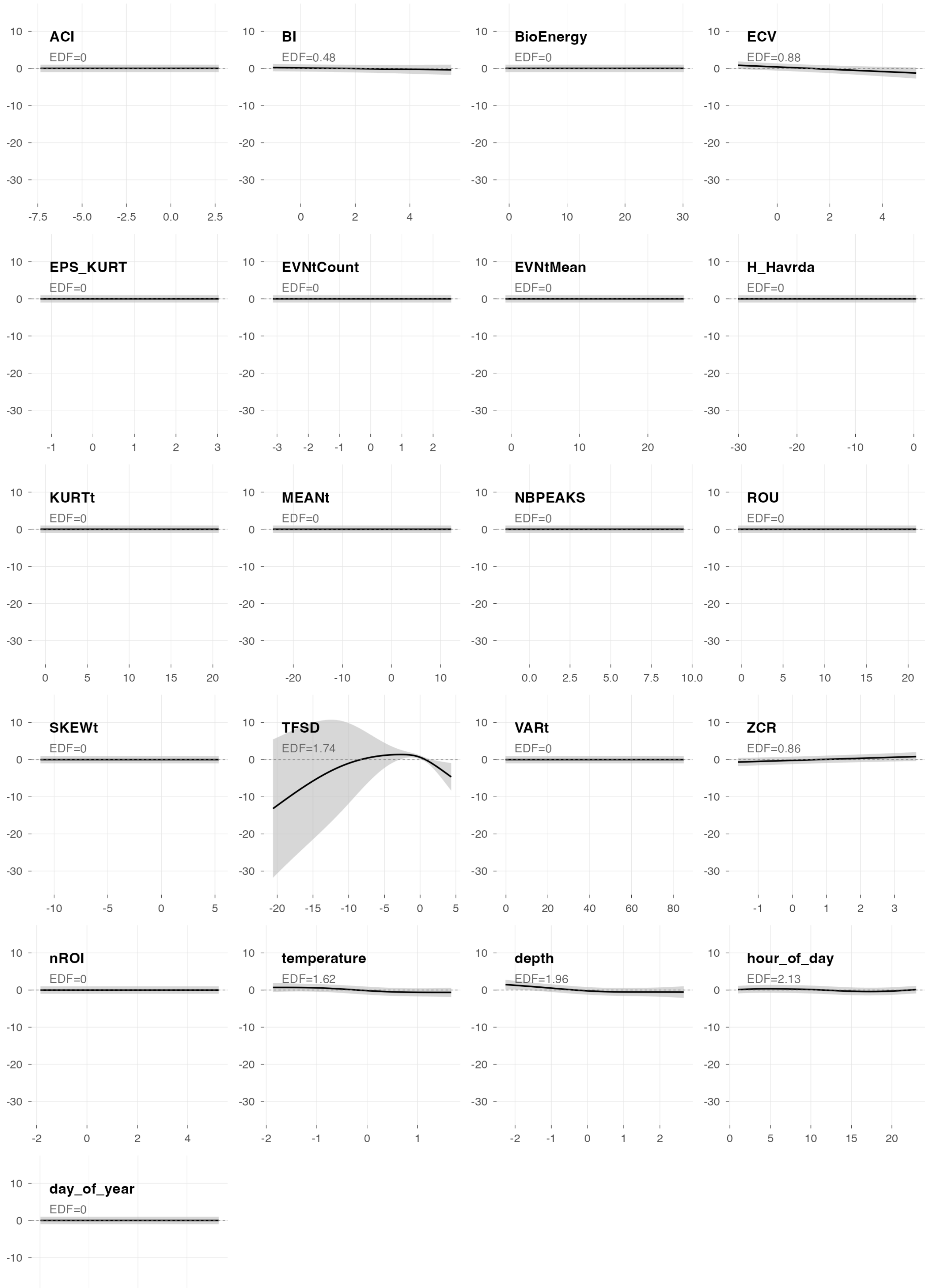
Response type: Count (negative binomial)

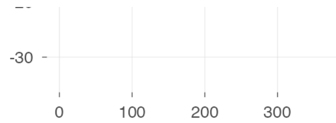
Key predictors: ECV, TFSD, ZCR, temperature, depth, hour_of_day, station

CV Performance: RMSE = 0.78, $R^2 = -\infty$ (rare events)

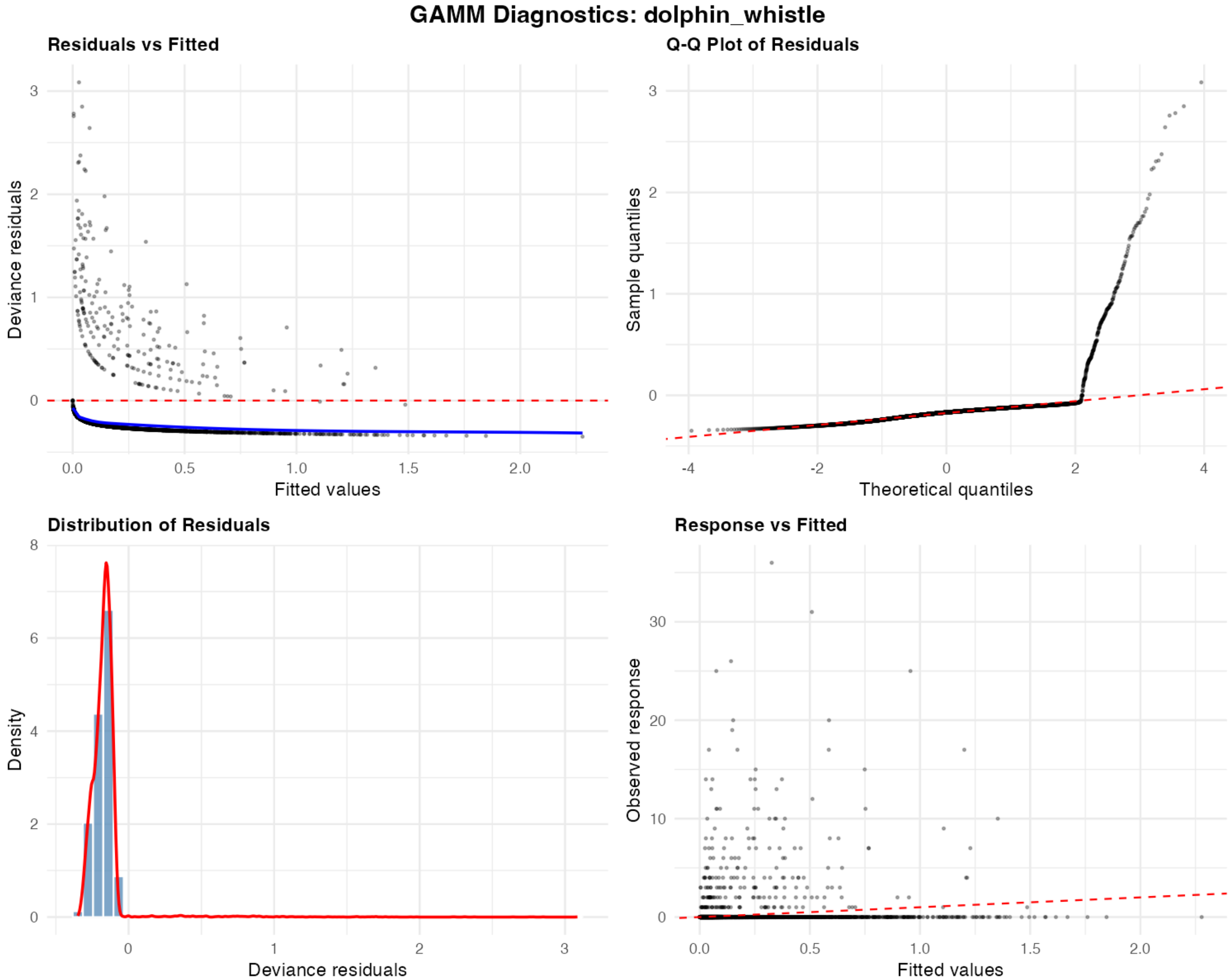
Smooth Terms Overview

dolphin_whistle — Smooth Terms





Diagnostics



Interpretation

Whistles are tonal social signals — only 3 indices (ECV, TFSD, ZCR) are significant. These capture spectral and temporal characteristics of whistles. The $R^2 = -\infty$ reflects that most observations are zeros (rare events); predicting *when* whistles occur is very hard.

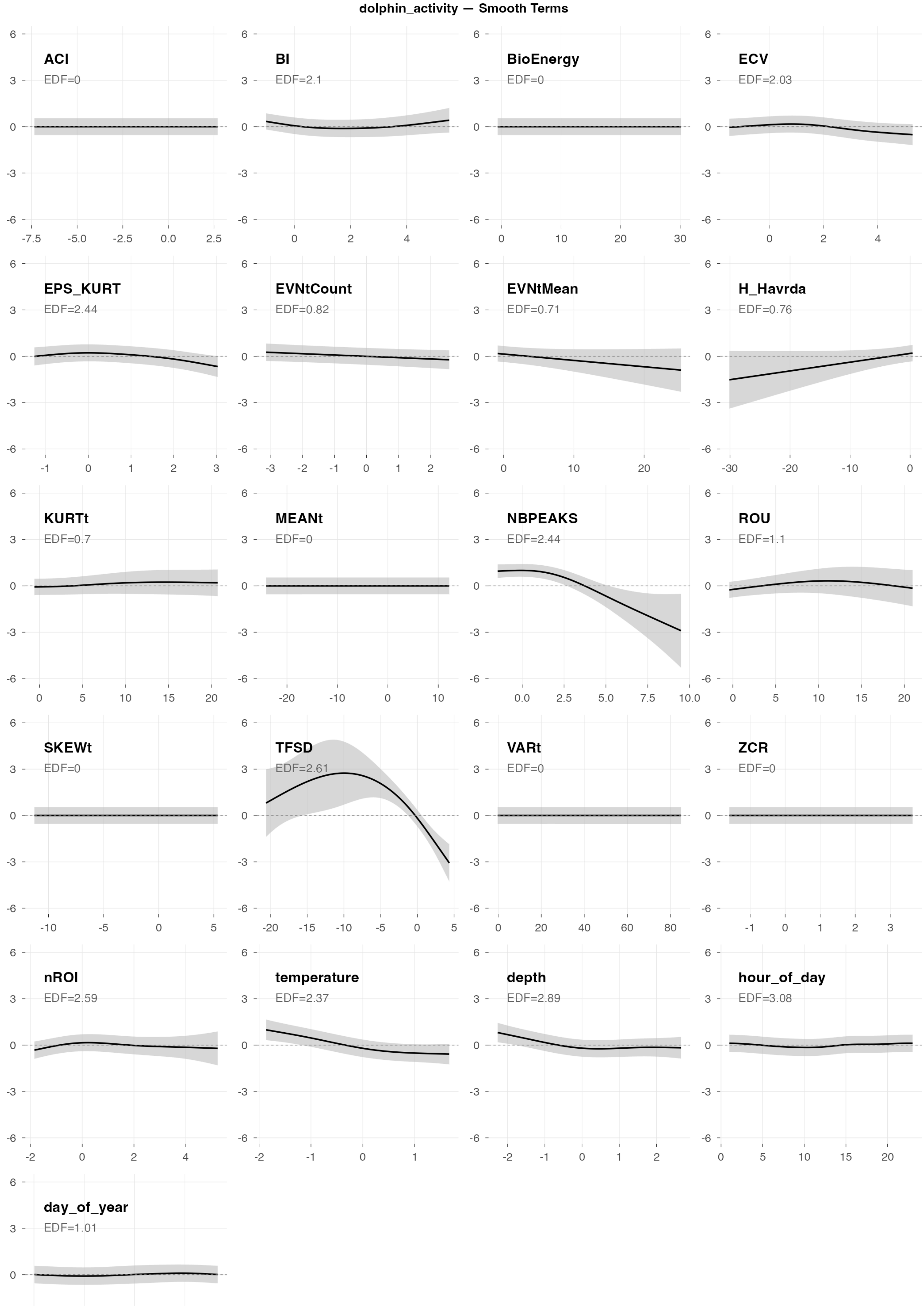
dolphin_activity

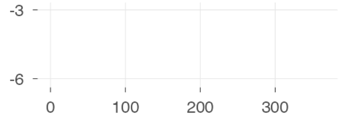
Response type: Count (negative binomial)

Key predictors: BI, ECV, EPS_KURT, EVNtCount, H_Havrda, NBPEAKS, ROU, TFSD, nROI, temperature, depth, hour_of_day, station

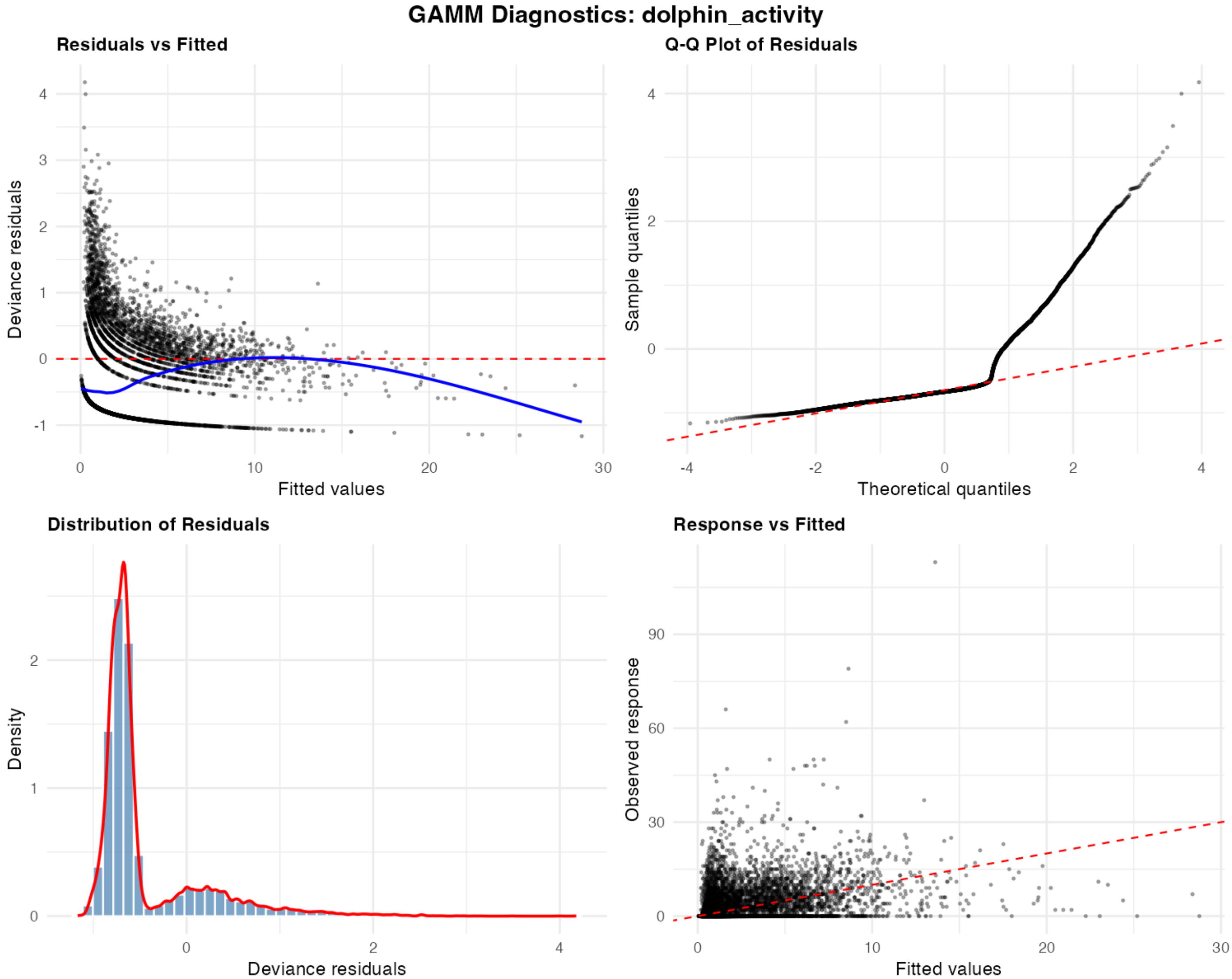
CV Performance: RMSE = 7.22, $R^2 = -48$ (poor generalization)

Smooth Terms Overview


dolphin_activity — Smooth Terms
ACI
EDF=0
BI
EDF=2.1
BioEnergy
EDF=0
ECV
EDF=2.03
EPS_KURT
EDF=2.44
EVNtCount
EDF=0.82
EVNtMean
EDF=0.71
H_Havrda
EDF=0.76
KURTt
EDF=0.7
MEANt
EDF=0
NBPEAKS
EDF=2.44
ROU
EDF=1.1
SKEWt
EDF=0
TFSD
EDF=2.61
VARt
EDF=0
ZCR
EDF=0
nROI
EDF=2.59
temperature
EDF=2.37
depth
EDF=2.89
hour_of_day
EDF=3.08
day_of_year
EDF=1.01



Diagnostics



Interpretation

Dolphin activity (total vocalizations) has many significant predictors — it aggregates across call types. Temperature is the dominant environmental driver (more activity in cooler water). Like other count models, it doesn't generalize well to held-out weeks.

dolphin_presence

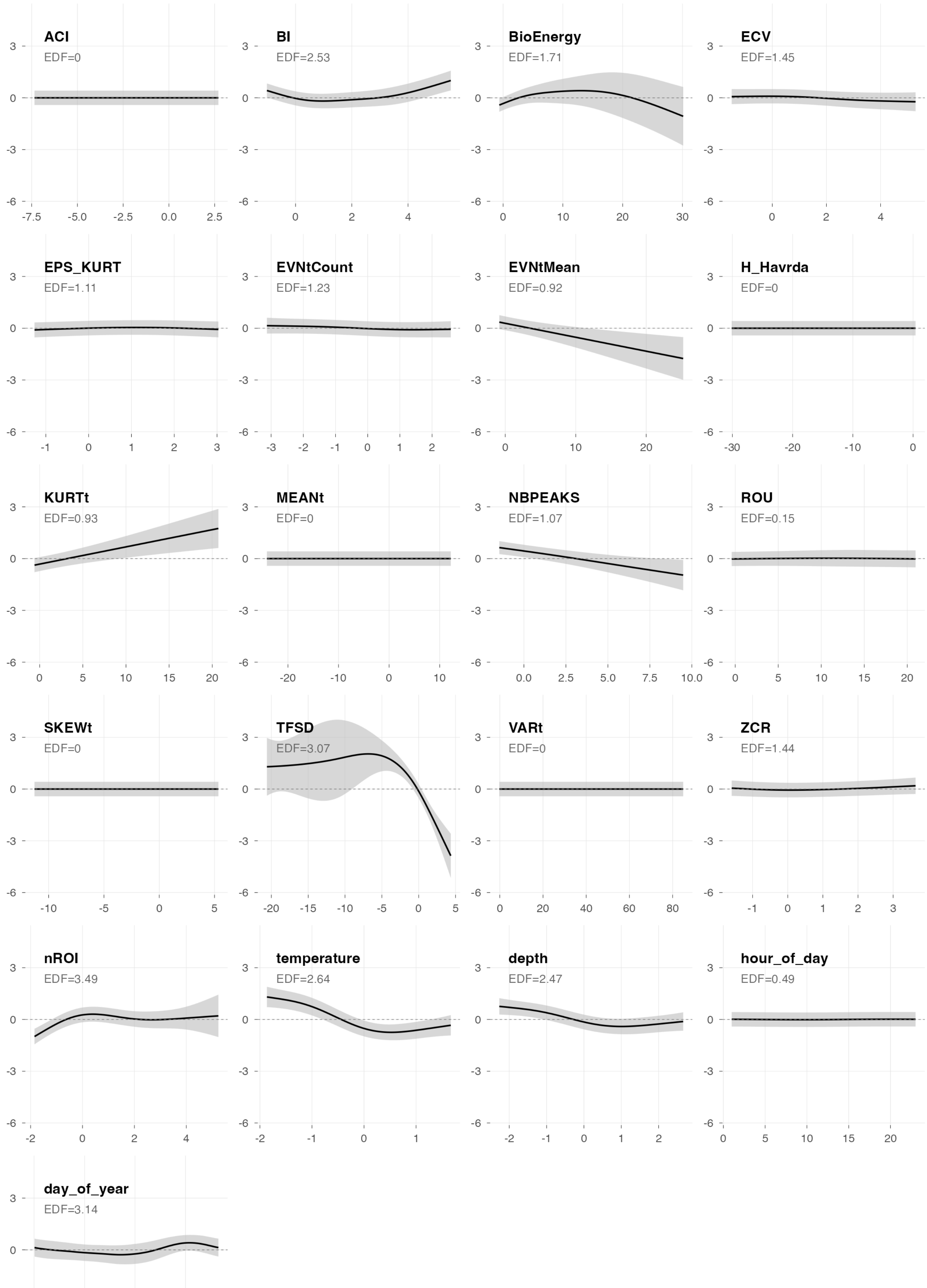
Response type: Binary (binomial)

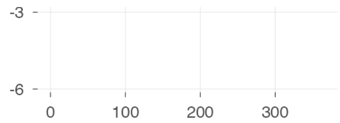
Key predictors: BI, BioEnergy, ECV, EVNtCount, EVNtMean, KURTt, NBPEAKS, TFSD, nROI, temperature, depth, station

CV Performance: AUC = 0.74 (moderate — useful signal)

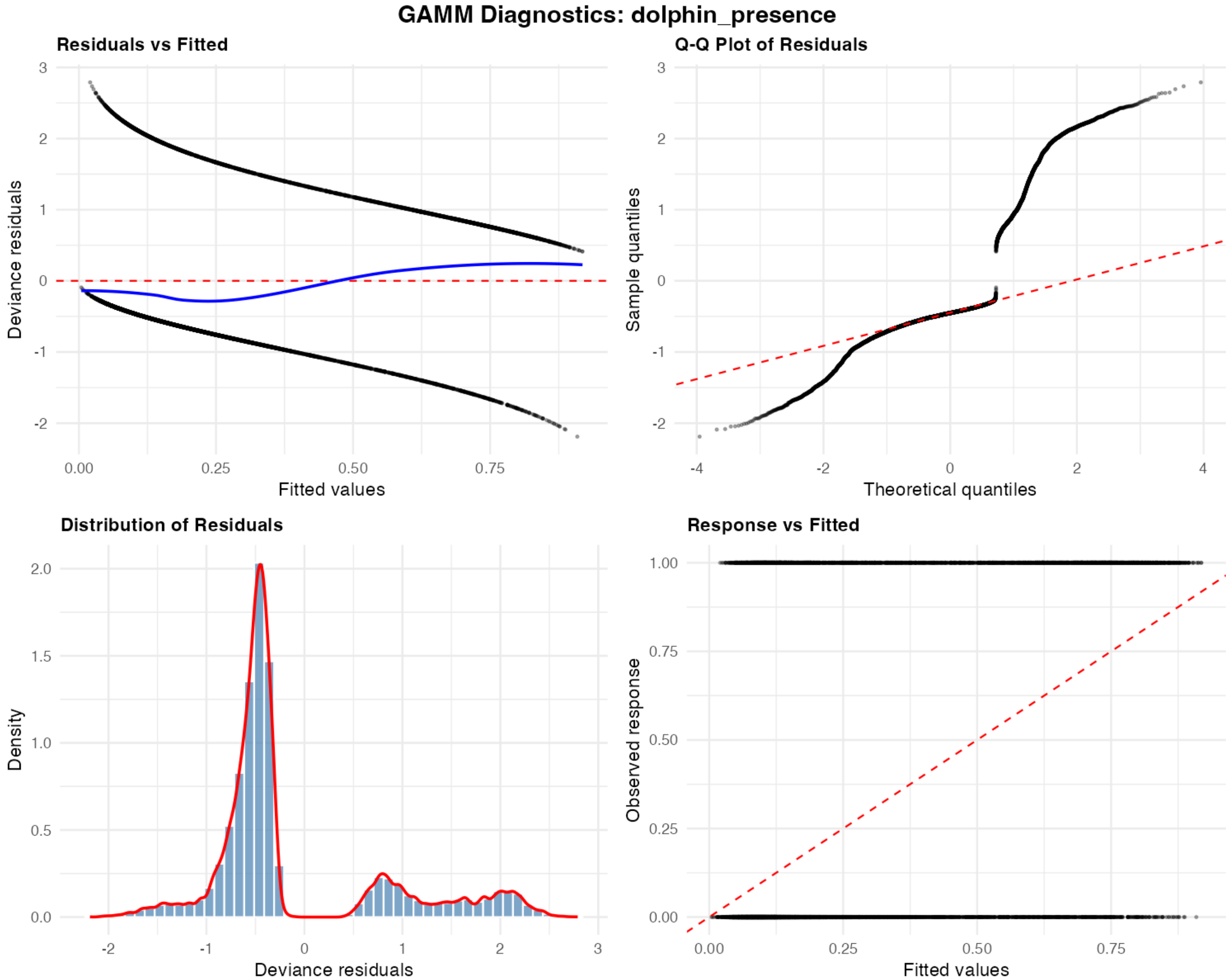
Smooth Terms Overview

dolphin_presence — Smooth Terms





Diagnostics



Interpretation

Dolphin presence is moderately predictable (AUC = 0.74), similar to fish presence. Many indices contribute. Temperature is significant — dolphins more likely present in cooler water. The binary framing (present/absent) is more predictable than count models.

vessel_presence

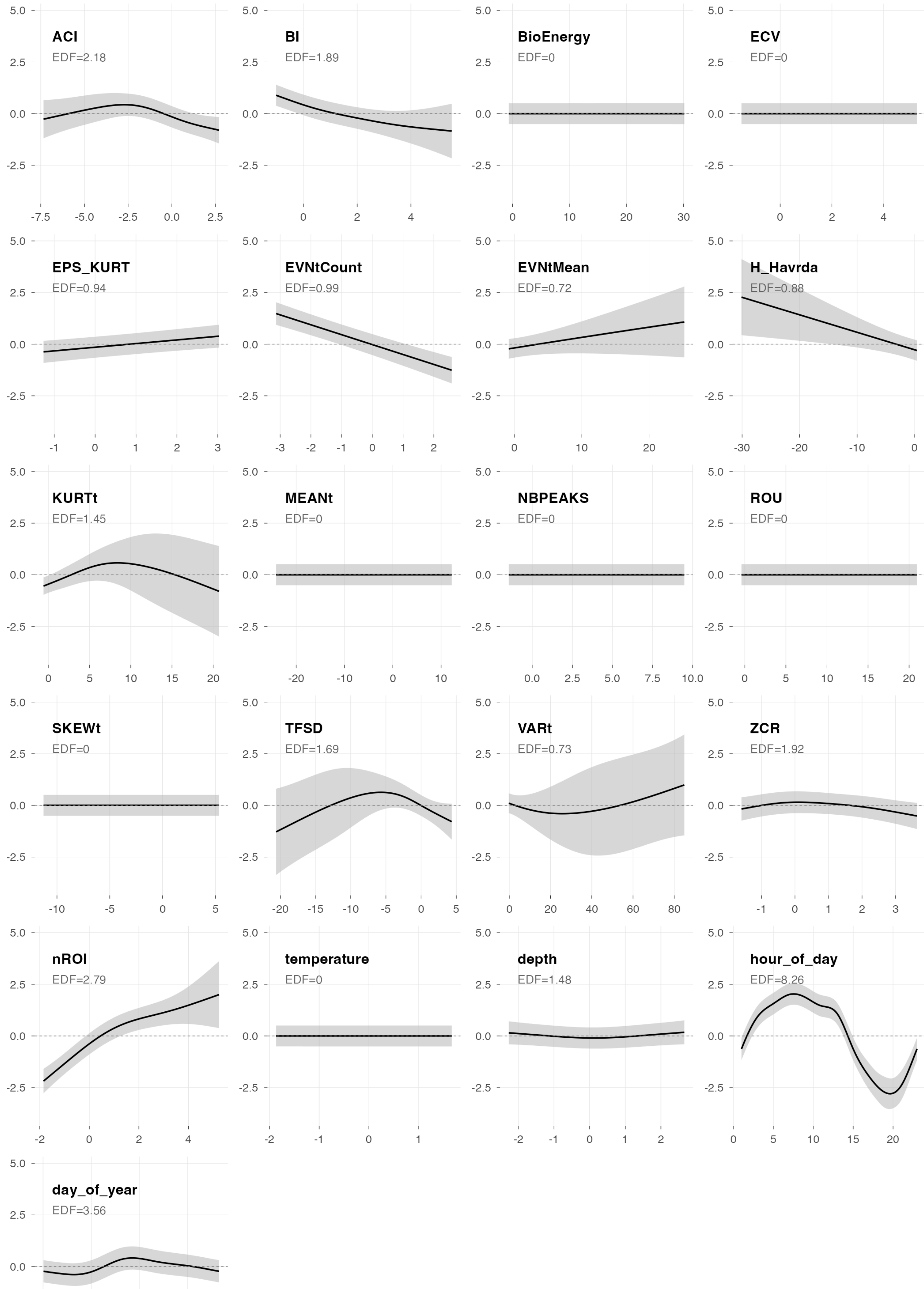
Response type: Binary (binomial)

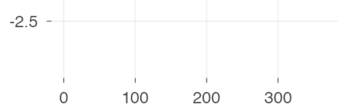
Key predictors: ACI, BI, EVNtCount, H_Havrda, KURTt, TFSD, ZCR, nROI, hour_of_day, day_of_year

CV Performance: AUC = 0.92 (excellent)

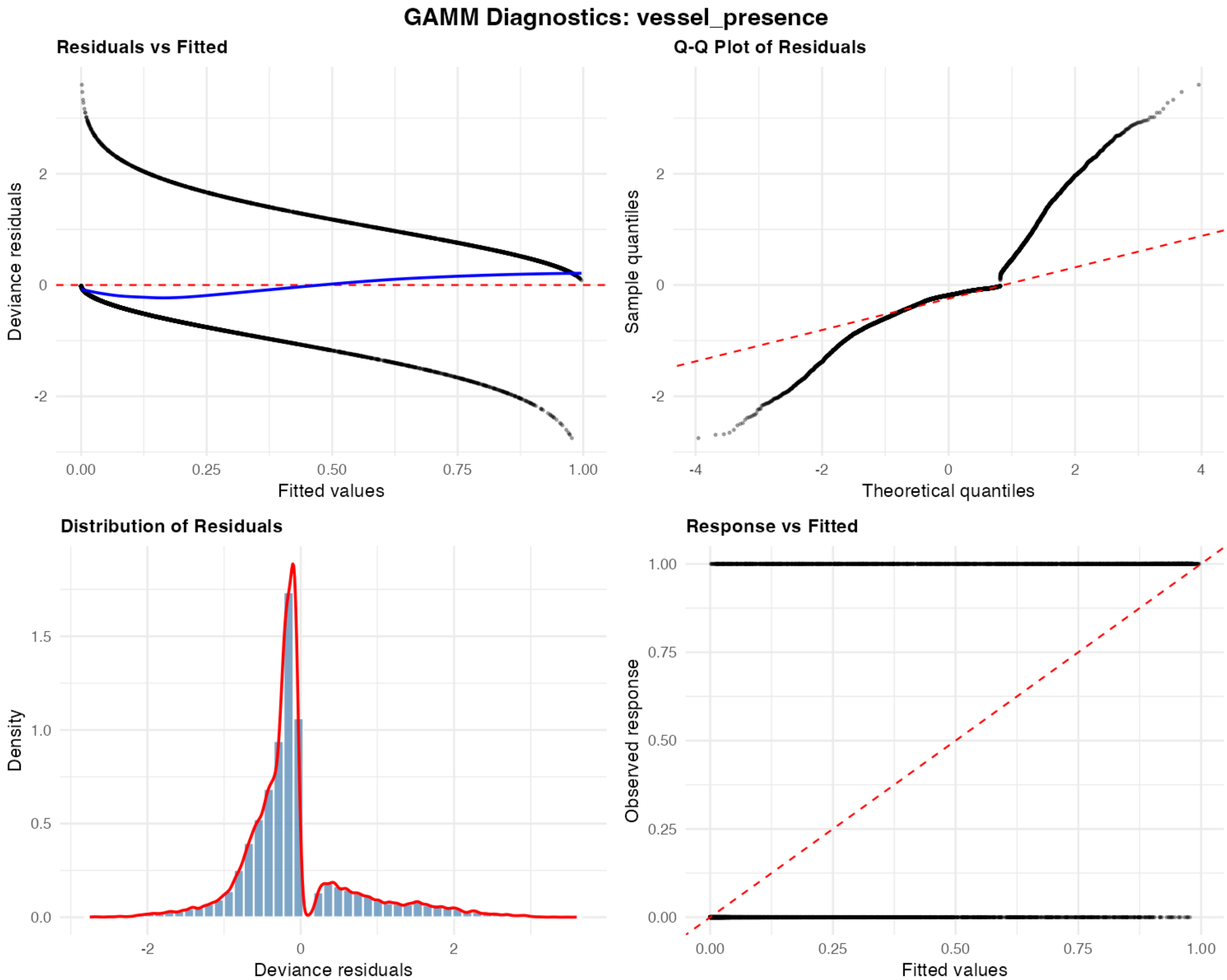
Smooth Terms Overview

vessel_presence — Smooth Terms





Diagnostics



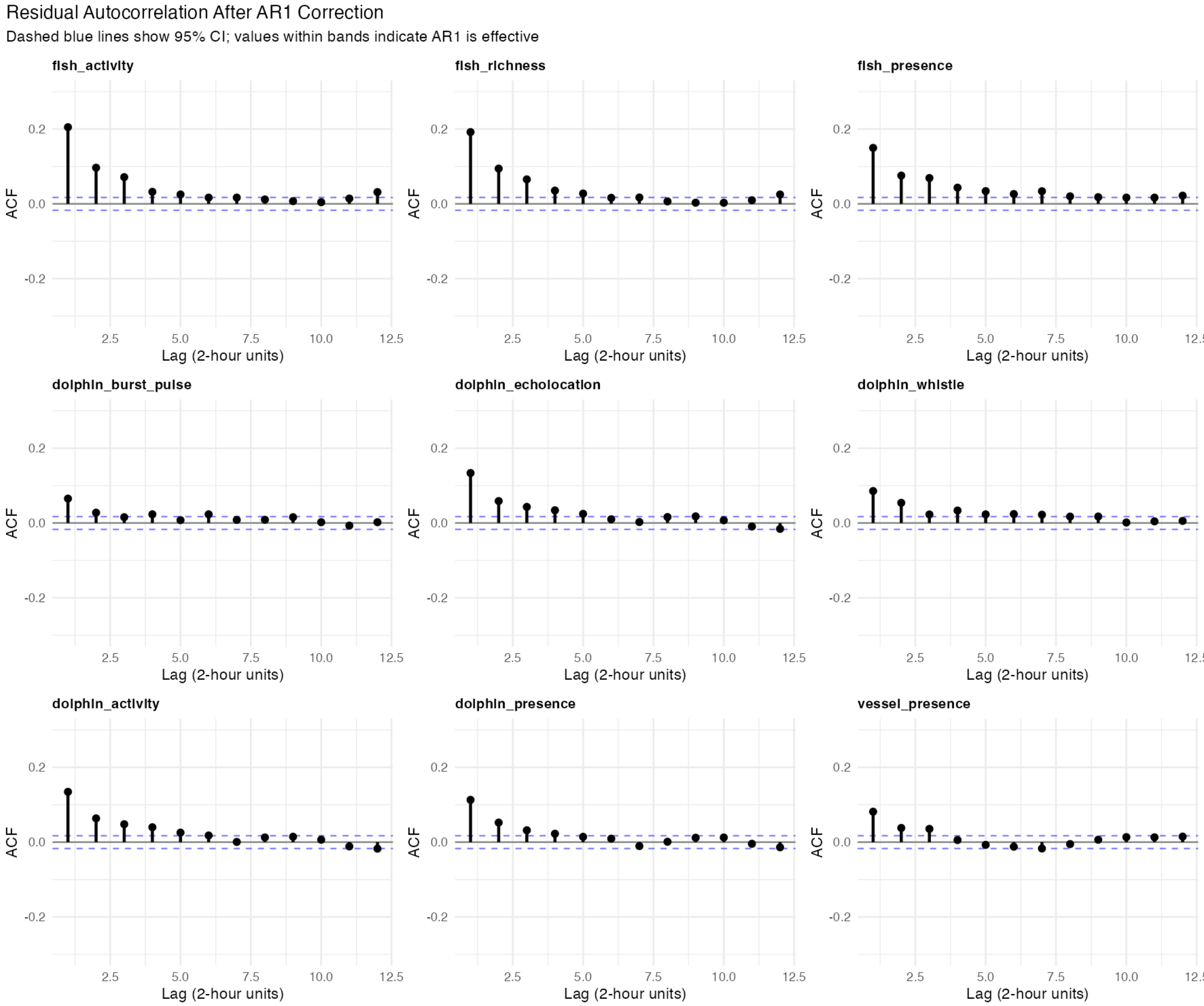
Interpretation

Vessel presence is the best-predicted metric. Boats are loud, broadband, and create distinctive temporal patterns — 8 indices are significant. Strong diel pattern (morning peak). Notably, temperature and depth are NOT significant — vessel detection is purely acoustic, not environmentally mediated.

Validation Results

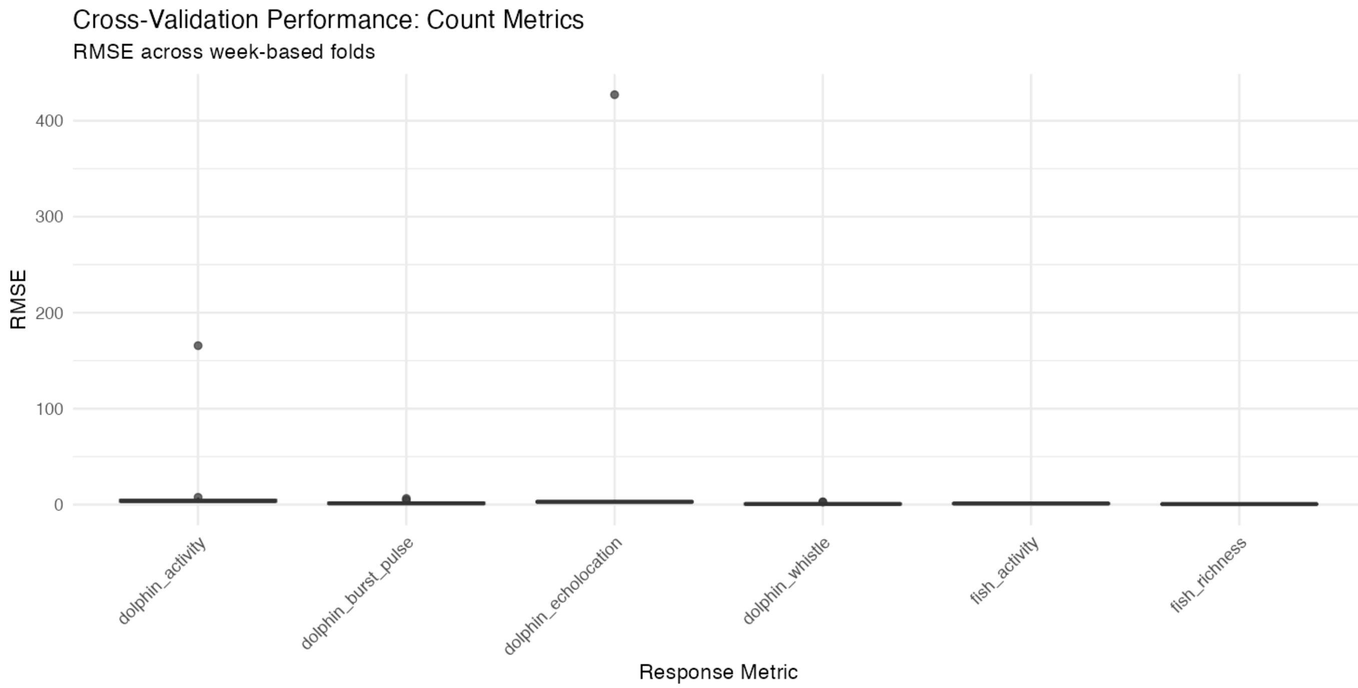
AR1 Autocorrelation Check

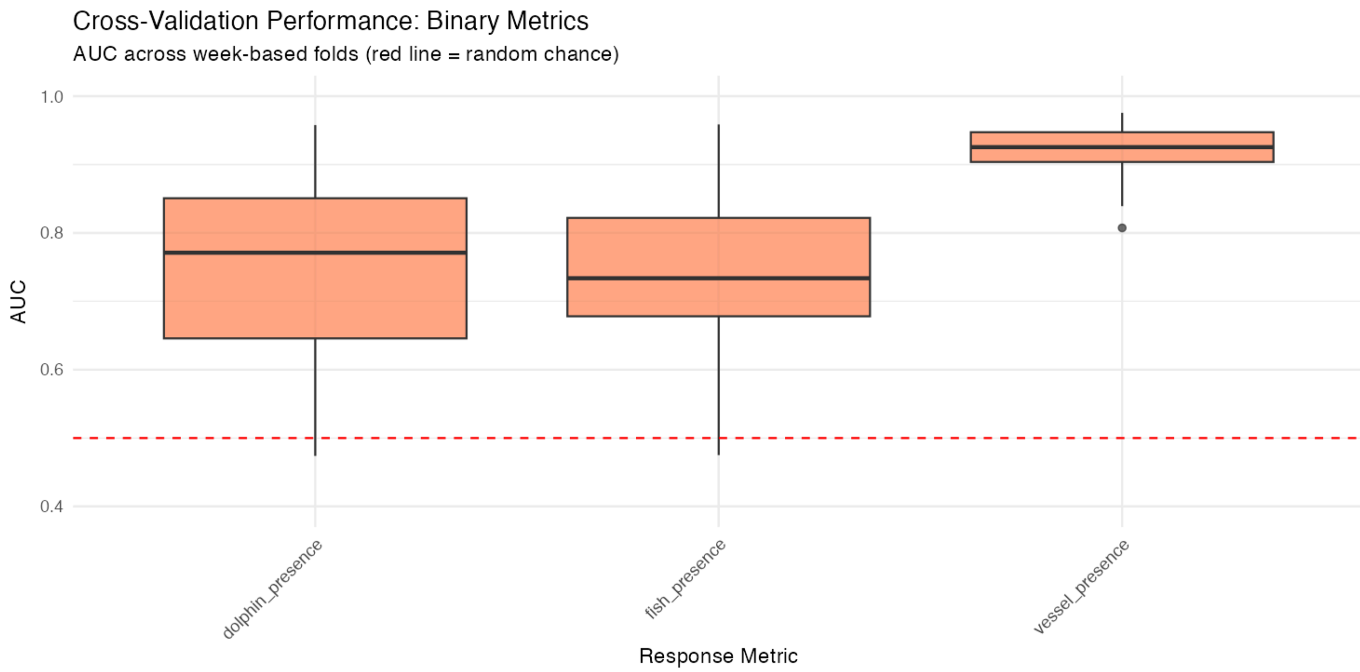
We checked if the AR1 correction removed temporal autocorrelation from residuals.



Finding: Residual ACF(1) ≈ estimated ρ, meaning the AR1 correction is not fully absorbing autocorrelation. This is a known limitation of the `bam()` approach. The autocorrelation levels are modest (0.06–0.20), so results are still interpretable but p-values should be treated conservatively.

Cross-Validation Performance





Binary Metrics (AUC)

Metric	Mean AUC	SD	Interpretation
vessel_presence	**0.92**	0.03	Excellent
fish_presence	0.75	0.11	Moderate
dolphin_presence	0.74	0.13	Moderate

Count Metrics (RMSE / R^2)

Metric	Mean RMSE	Mean R^2	Interpretation
fish_richness	0.56	0.02	Low error, poor R^2
fish_activity	1.14	0.01	Same
dolphin_whistle	0.78	$-\infty$	Rare events
dolphin_burst_pulse	1.54	-0.13	Poor
dolphin_echolocation	11.0	-455	Outlier folds
dolphin_activity	7.22	-48	Poor

Count Models Don't Generalize

The negative R^2 values for count models mean predictions are worse than simply predicting the mean. This is common in time series CV — weekly patterns don't transfer. The fitted relationships are real but shouldn't be used for prediction without caution.

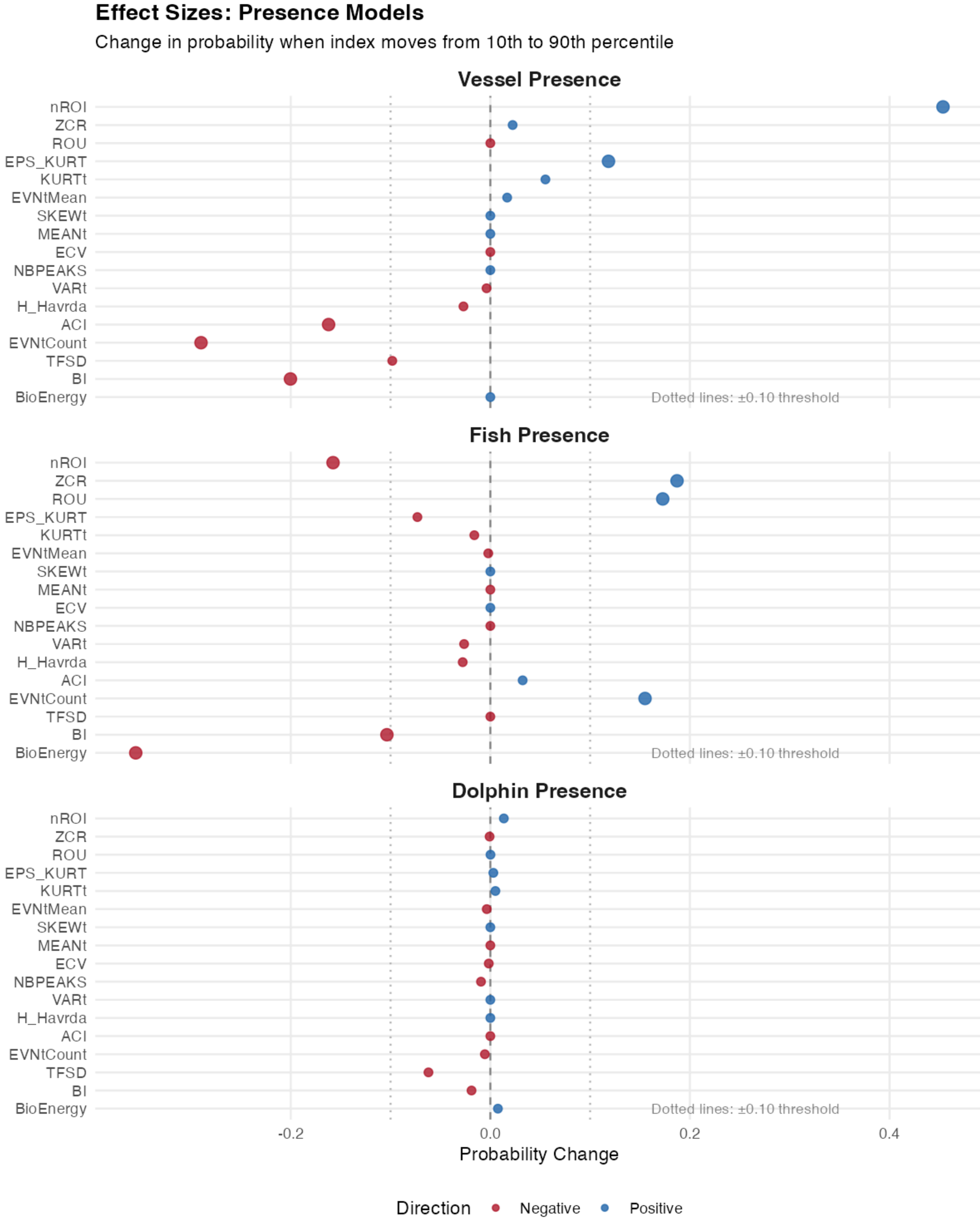
Diagnostic Checks

These checks provide context for interpreting the main results — which effects are practically meaningful, why count models struggle, and how to read the significance table.

Effect Sizes: What Actually Matters?

Statistical significance ($p < 0.05$) tells us an effect isn't zero. Effect size tells us whether it's large enough to matter. With 13,102 observations, even trivial effects can be "significant."

The plot below shows the change in detection probability when each index moves from its 10th to 90th percentile (holding other predictors at median values). Dotted lines mark ±0.10 as a rough threshold for "meaningful" effects.



Key findings:

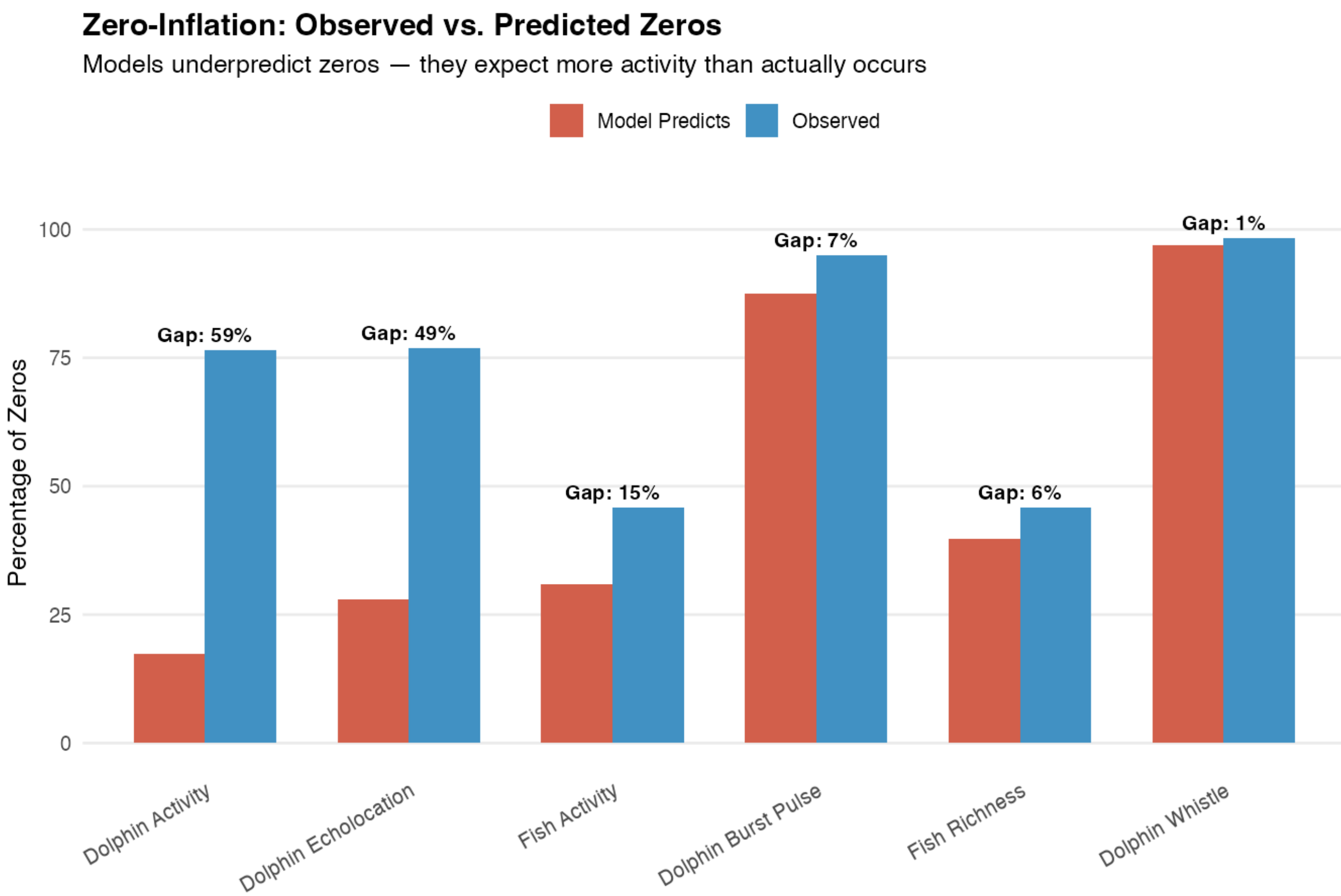
- **Vessel presence** has the largest practical effects — nROI alone shifts detection probability by +45 percentage points

- **Fish presence** has moderate effects from BioEnergy (−36 pp), ZCR (+19 pp), and ROU (+17 pp)
- **Dolphin presence** effects are all small (< 7 pp) despite statistical significance

This explains why vessel detection works so well (AUC = 0.92) — the index associations are genuinely strong. For dolphins, the indices provide weak signal even when "significant."

Zero-Inflation: Why Count Models Struggle

Count models assume a negative binomial distribution, but the data has far more zeros than this distribution expects. The models think activity should occur more often than it actually does.



Interpretation: Dolphins are rare and bursty — most 2-hour windows have zero detections. The Dolphin Activity model expects only 17% zeros but observes 77% (a 59 percentage point gap). This mismatch explains the poor cross-validation performance (negative R^2). The model structure doesn't fit the data-generating process.

For applications requiring activity counts, a zero-inflated model or hurdle model might be more appropriate. For presence/absence detection, the current binary models avoid this issue entirely.

Concurvity: Interpreting the Significance Table

Concurvity is the GAM equivalent of multicollinearity. High concurvity (> 0.8) means smooth terms explain similar variance — they're partially redundant.

Most index terms in these models have concurvity > 0.8. This means:

- The many "significant" indices in the results table aren't independent signals
- If ZCR predicts fish, correlated indices (ACI, EVNtCount) will also appear significant
- Effects can't be cleanly attributed to specific indices

Practical implication: Think of the indices as capturing "acoustic characteristics" broadly, not as 17 independent predictors. The indices that matter most are those with large effect sizes, not just statistical significance.

> **Diagnostic Summary**
>
> - **Effect sizes** show vessel and fish presence have practically meaningful index associations; dolphin effects are small even when statistically significant
> - **Zero-inflation** explains why count models don't generalize — too many zeros for the assumed distribution
> - **Concurvity** means significant indices are correlated, not independent predictors

---

Conclusions

What Works

1. **Vessel detection** (AUC = 0.92) — Acoustic indices reliably distinguish vessel presence. Useful for anthropogenic noise monitoring.

2. **Biological presence** (AUC ~0.75) — Fish and dolphin presence models show moderate predictive power. Indices contain real signal about whether activity is occurring.

3. **Index-response associations** — Multiple indices significantly predict biological metrics, with ecologically interpretable patterns (e.g., ZCR for fish, ECV for dolphins).

What Doesn't Work

1. **Activity/count prediction** — Models don't generalize to held-out weeks. Good for describing patterns, not for forecasting.

2. **Full AR1 correction** — Residual autocorrelation remains; interpret significance conservatively.

The Bottom Line

Acoustic indices are useful for **screening** ("is there activity?") but not for **quantification** ("how much activity?"). For presence/absence monitoring, they provide real value. For activity levels, they describe patterns but don't predict reliably.

Downloads

- Model Summary (CSV)
- CV Performance Summary (CSV)
- AR1 Validation (CSV)

- Effect Sizes (CSV)
- Zero-Inflation Check (CSV)

Generated: 2026-01-31