

Operational Carbon-Flux Forecasting for Peatlands

A foundation-model approach to 96-hour net ecosystem exchange prediction, validated against a reproducible benchmark.

First edition · 2026 · Pre-launch

ABSTRACT

MAZE produces 96-hour, hourly forecasts of net ecosystem exchange (NEE) for peatland sites from the recent history of a site's own flux signal. The forecasting engine is a fine-tuned time-series foundation model. On a held-out blanket-bog peatland it attains an aggregate R^2 of 0.60 across the full 96-hour horizon; on a structurally different forest site it reaches 0.73–0.77, where conventional machine-learning models lose 37–54% of their skill. Critically, the checkpoint served in production is the one that was validated — the system users receive is the validated model, not an approximation of it. Every forecast carries a calibrated uncertainty band and an explicit flag on the hours least likely to be reliable. This document sets out the method, the validation evidence, and — as importantly — the current limits of what MAZE claims.

MAZE

“We’ll find a way.”

SECTION 1

The monitoring gap

Net ecosystem exchange is the net of two large, opposing fluxes — photosynthetic uptake of carbon and its release through ecosystem respiration. Over a peatland it swings between net uptake and net release on daily and seasonal cycles, governed by water-table depth, temperature, light and the slow biogeochemistry of waterlogged peat. Whether a given site is currently a sink or a source is therefore not a fixed property; it is a moving quantity that has to be tracked.

Direct measurement is possible but scarce. Eddy-covariance flux towers measure NEE accurately, yet they are expensive, sparse, and absent from the overwhelming majority of sites that a restoration programme or a carbon-project portfolio actually needs to understand. Process-based ecosystem models exist but are slow to configure and calibrate per site. For the people who develop, finance and verify peatland carbon projects — increasingly under standards such as the UK Peatland Code — the practical need is a forward-looking estimate of flux at a specific site, on a timescale relevant to management, without instrumenting every hectare.

There is also a machine-learning obstacle that any deployable product must confront. Models trained to predict flux in one ecosystem typically degrade sharply when applied to another — precisely the situation faced by a service that must forecast across many heterogeneous real-world sites rather than the one it was tuned on. A method that only works on the ecosystem it was trained on is not an operational tool. Robustness across site types is not a nice-to-have; it is the product.

SECTION 2

Approach

MAZE is built on a time-series **foundation model** — an 80-million-parameter transformer (the TEMPO family) pretrained on a large, structurally diverse corpus of time series. Rather than learning peatland flux from scratch, MAZE adapts these pretrained temporal representations to NEE forecasting. This is the same shift that reshaped language and vision: a general model with broad prior structure, specialised to a domain, tends to generalise where a narrowly-trained model overfits.

Forecasting from the flux signal alone

MAZE forecasts NEE from a single input channel: the site's own recent NEE history. A 14-day (336-hour) lookback produces a 96-hour (4-day) forecast. It does not require aligned meteorological feeds, satellite covariates, or site-specific instrumentation beyond the flux history itself. The pretrained representations carry enough structure to recover the implicit meteorological drivers from the dynamics of the flux alone. This is a deliberate operational choice: it makes the model deployable at sites that lack rich, well-aligned sensor networks — which is most of them.

Two adapted models, matched to the site

MAZE deploys two adapted checkpoints. A **universal** model, adapted on peatland/wetland flux records, is the default and serves bog and fen sites. An **ecosystem-specific forest** model, further adapted on forest flux, serves forest sites. MAZE selects the appropriate model for a site's ecosystem type. The adaptation corpus, training procedure and hyperparameters are proprietary and are not disclosed here; the validation that follows is designed to be judged on results, not on the recipe.

Preprocessing and uncertainty

Incoming flux histories are gap-filled hierarchically and standardised using statistics fixed at training time and applied unchanged at inference, so no information from the site being forecast leaks into preprocessing. Each forecast is produced as a small Monte-Carlo ensemble (following the dropout-as-Bayesian-approximation principle), which yields a per-hour uncertainty band and an explicit **high-uncertainty flag** — the hours whose forecast spread exceeds a set multiple of the median are marked for review rather than presented with false confidence.

SECTION 3

Validation

MAZE’s skill is measured on flux-tower sites **entirely withheld** from model adaptation. This spatial hold-out matters: evaluating a flux model on later data from a site it was trained on is a well-known way to flatter performance, because the model can lean on that site’s own autocorrelation. Reporting only on unseen sites means the figures below reflect genuine transfer to a new location, which is what deployment actually demands. Reference flux is NEE_{VUT_REF} from the harmonised FLUXNET2015 archive, computed with community-standard friction-velocity filtering.

THE CENTRAL GUARANTEE

The checkpoint MAZE serves **in production is the checkpoint that was validated** — the deployment path does not silently diverge from the version that was measured. In validation, that checkpoint reproduces its published benchmark to within $0.05 R^2$, with a largest observed deviation of $0.001 R^2$ across the four model×site combinations tested.

Table 1 shows the production model–ecosystem pairings against benchmark. All four model×site combinations were validated and every one reproduced to within the $\pm 0.05 R^2$ bound; the pairings below are those MAZE actually serves.

Model	Site (ecosystem)	R^2 (MAZE)	R^2 (benchmark)	ΔR^2
Universal	UK-AMo (blanket-bog peatland)	0.598	0.599	-0.001
Universal	SE-Htm (spruce forest)	0.727	0.728	-0.001
Forest	SE-Htm (spruce forest)	0.766	0.766	+0.000

Table 1 — The served checkpoint reproduces the validation benchmark. Aggregate R^2 over the 96-hour horizon, across the held-out test sites, measured in deterministic validation. ΔR^2 is MAZE minus benchmark.

Skill on the primary peatland target

On the held-out blanket-bog peatland (UK-AMo), the universal model attains an aggregate R^2 of 0.60 across the four-day horizon. Skill is highest in the near term and decays with lead time, as any honest forecast must — Table 2 reports the full profile. MAZE does not hide this decay: the per-hour uncertainty band widens accordingly, so a user can see where in the horizon the forecast is strong and where it weakens.

Lead time	R ²	RMSE	MAE
1 hour	0.881	1.28	0.79
6 hours	0.698	2.01	1.26
12 hours	0.695	2.05	1.25
24 hours	0.680	2.06	1.24
48 hours	0.595	2.37	1.47
72 hours	0.564	2.39	1.54
96 hours	0.474	2.66	1.70

Table 2 — Per-horizon skill, universal model on the blanket-bog peatland (UK-AMo). RMSE and MAE in $\mu\text{mol m}^{-2} \text{s}^{-1}$. The aggregate R² of 0.60 pools all lead times.

Robustness across ecosystems

The clearest evidence that MAZE is an operational tool rather than a single-site curiosity is what happens when the model is moved to an ecosystem unlike its training data. Conventional models — gradient boosting, random forests, recurrent networks — were given a substantial advantage in this comparison: the full multivariate input of meteorological drivers and satellite vegetation bands, where MAZE saw only the univariate flux signal. They nonetheless lost between 37% and 54% of their skill moving from the peatland to the forest site. MAZE’s foundation model did the opposite: its skill **rose** by 21.6%. It also held near-zero forecast bias on the forest, where the baselines developed large systematic bias — evidence the pretrained representations accommodate a new ecosystem instead of imposing the prior one.

Model	Input	Change in R ² , peatland → forest
MAZE (universal foundation model)	Univariate NEE	+21.6% (skill increases)
LSTM network	Multivariate + satellite	−37.3%
XGBoost	Multivariate + satellite	−39.9%
Random forest	Multivariate + satellite	−53.5%

Table 3 — Cross-ecosystem transfer. Change in aggregate R² when the same model is moved from the held-out peatland (UK-AMo) to the held-out forest (SE-Htm). Conventional models degrade; the foundation model improves.

How uncertainty behaves

Forecast uncertainty is not uniform, and MAZE surfaces where it concentrates. Under hot, high-vapour-pressure-deficit conditions — the meteorological regimes that drive the most volatile flux — forecast spread rises to roughly three to four times its baseline, and summer uncertainty runs up to about four times winter levels. In exactly those conditions the per-hour band widens and the high-uncertainty flag fires. The product is built to tell a user when to trust it less, which for a verification audience is more valuable than a single headline number.

SECTION 4

What the product delivers

- **A 96-hour, hourly NEE forecast** for a given peatland site, generated from the site’s recent flux history.
- **A calibrated uncertainty band** on every hour of the forecast, reflecting model confidence rather than a flat error bar.

- **Explicit high-uncertainty flags** identifying the hours least likely to be reliable, so they can be treated with appropriate caution.
- **The deployed-equals-validated guarantee:** the served model reproduces its benchmark to within $0.05 R^2$.

Forecasts in production are generated as a Monte-Carlo ensemble to produce the uncertainty band; the benchmark figures in Section 3 use a single deterministic pass to compare like-for-like against the established benchmark. The expected difference between the two from ensemble averaging is small — on the order of $0.02 R^2$ — and is not separately quantified in this edition.

SECTION 5

Limitations & current scope

What a method does not yet claim matters as much as what it does, particularly for a verification audience. MAZE states its limits plainly.

- **The validated horizon is 96 hours.** Skill decays with lead time — from $R^2 \approx 0.88$ at one hour to ≈ 0.47 at 96 hours on the primary peatland site. MAZE reports per-hour uncertainty so this is visible, not buried.
- **MAZE forecasts hourly flux, not an annual carbon balance.** Integrating short-horizon skill into a defensible annual net figure — the quantity the Peatland Code ultimately requires — needs additional methodology. It is on the roadmap and is **not** part of the current product.
- **Validation to date covers specific ecosystems.** Headline figures come from a blanket-bog peatland and, for the forest model, a spruce forest. Broadening site and ecosystem coverage with the same benchmark discipline is ongoing.
- **Forecasts are driven by recent flux history.** They assume a reasonably representative recent record is available and do not anticipate abrupt regime changes — for example a sudden management intervention — occurring outside the lookback window.

SECTION 6

Roadmap

The current product delivers validated 96-hour operational forecasts. The next major line of work extends the methodology to a **defensible annual carbon balance** — the Peatland Code use case — carrying the same validation-against-benchmark discipline through to the annual figure rather than asserting it. Alongside this, MAZE is expanding ecosystem and site coverage, validating each addition against held-out flux-tower data before it informs the product.

IN ONE LINE

MAZE forecasts peatland carbon flux four days ahead, tells you how much to trust each hour, and serves exactly the model it validated. The annual carbon balance is the next chapter — and it will arrive the same way: measured, not claimed.

REFERENCES

Selected references

- Pastorello, G., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7, 225.
- Reichstein, M., et al. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, 11, 1424–1439.
- Cao, D., et al. (2024). TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting. *International Conference on Learning Representations (ICLR)*.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *International Conference on Machine Learning (ICML)*.

IUCN UK Peatland Programme. The Peatland Code — the voluntary standard for UK peatland restoration carbon projects.

© 2026 MAZE. This document describes methodology and validation results for an unreleased product and is provided for evaluation. Figures are drawn from held-out validation against the FLUXNET2015 archive. mazeeco.com