

Symbolic Regression for Half-hourly Methane Fluxes Across Three Asian Rice Paddies under Contrasting Water Management

Jef Zerrudo^{1,2,4,6} Laurent Bataille¹ Ronald Hutjes¹ Bart Kruijt¹ Björn Ole Sander³ Caesar Arloo Centeno³ Reiner Wassmann⁴ Ronaldo Saludes⁵ Moises Dorado⁵ Pompe Sta Cruz⁶

¹Wageningen University & Research, NL ²DOST-PAGASA, PH ³IRRI, Los Baños, PH ⁴formerly IRRI ⁵UPLB, PH ⁶formerly UPLB

Poster presented at NCGG-10, Utrecht, 15–17 June 2026

Abstract

Rice paddies are a major source of atmospheric methane (CH_4), yet the dominant environmental control on flux differs sharply between water-management regimes — continuous flooding versus alternate wetting and drying (AWD) — so no single fixed equation can represent paddy CH_4 across systems in crop models. We used symbolic regression (PySR) to discover interpretable, regime-specific CH_4 equations directly from half-hourly eddy-covariance data at three Asian rice sites: JP-MSE (Japan, continuous flooding, 2012; 6,760 records), PH-IR (Philippines, aerobic AWD, 2016; 3,620 records), and KOR-CRK (South Korea, aggressive AWD, 2018). After GAM-based collinearity screening (5–9 predictors per site), we ran PySR (12 runs, 2000 iterations, maximum complexity 35), selected the highest- R^2 candidate on an 80/20 hold-out, and validated with day-grouped 5-fold cross-validation, which withholds entire days to respect half-hourly autocorrelation. Hold-out R^2 for the most-complex forms reached 0.69 / 0.74 / 0.52 (JP-MSE / PH-IR / KOR-CRK), but day-grouped CV R^2 was consistently lower (0.42 ± 0.13 / 0.39 ± 0.36 / 0.49 ± 0.11), and the most complex equation was not always best — PH-IR’s mid-complexity form generalised better than its highest-accuracy form. Critically, water management reshaped the functional form itself: temperature gated by hydrology under continuous flooding (JP-MSE), cumulative water depth (AUC) under aerobic AWD (PH-IR), and air temperature under aggressive AWD (KOR-CRK) — each regime selecting its own functional “grammar.” These regime-specific equations provide a foundation for daily-aggregated re-fits to couple CH_4 into the ORYZA crop model.

Discovered equations

JP-MSE — continuous flooding

Auto-best (C15, $R^2=0.65$): $\hat{\text{CH}}_4 = 0.140 \exp(0.131 T_v) / [1.319 - \tanh(-0.00567/(h \cdot P_r - 0.909))]$

Pareto-knee (C17, $R^2=0.64$): $\hat{\text{CH}}_4 = 0.196 \exp(0.119 T_v) / [1.367 - \tanh(-0.00485/(h \cdot P_r - 0.909)) + \text{buoy}_{T_s T_a}]$

Mid (C19, $R^2=0.68$): $\hat{\text{CH}}_4 = \tanh(0.141) \exp(0.131 (T_v - \tanh(u \cdot \text{VPD}))) / [1.341 - \tanh(-0.00611/(h \cdot P_r - 0.909))]$

Best (C32, $R^2=0.69$): $\hat{\text{CH}}_4 = 0.176 \exp(0.131 (T_v - \tanh(u \cdot \text{VPD}))) / D + \text{buoy}_{T_s T_a}$
 $D = \text{buoy}_{T_s T_a} + \tanh(e^{-0.651} + 0.00611/(h \cdot P_r - 0.909)) + 1.367 - \tanh(h \cdot v / (h \cdot P_r - 1.007))$

PH-IR — aerobic AWD

Auto-best (C4, $R^2=0.39$): $\hat{CH}_4 = \exp(0.00629 \text{ AUC})$

Pareto-knee (C10, $R^2=0.59$): $\hat{CH}_4 = \exp(0.00459 (\text{AUC} + 0.01426 \text{ SR} \cdot T_s)) + 1.062$

Mid (C19, $R^2=0.73$): $\hat{CH}_4 = [0.414 + \exp(0.00913 (\sqrt{\text{SR} \cdot T_s} + 0.323 \text{ AUC}))] / \exp(\tanh(0.00291 \text{ SR} \cdot \text{HOD}_{\sin} - 0.840))$

Best (C35, $R^2=0.74$): $\hat{CH}_4 = N/D + \sqrt{(0.594 + \exp(0.404 \text{ h} \cdot \text{VPD}))} - 0.112,$

$N = \exp(0.00435 (\text{h} \cdot \sin\text{TOD} + \sqrt{\text{SR} \cdot T_s} + \text{AUC})) + 0.0205;$

$D = \exp(\tanh(0.00403 (\text{SR} \cdot \text{HOD}_{\sin} - 0.0203 \text{ SR} \cdot T_s))) - 0.0205 \text{ h} \cdot \sin\text{TOD}$

KOR-CRK — aggressive AWD

Auto-best (C4, $R^2=0.46$): $\hat{CH}_4 = \exp(0.0974 T_{\text{air}})$

Pareto-knee (C6, $R^2=0.48$): $\hat{CH}_4 = \exp(0.121 T_{\text{air}} - 0.648)$

Mid (C20, $R^2=0.51$): $\hat{CH}_4 = \exp(0.106 (T_{\text{air}} - 0.000744 \text{ SR} \cdot v)) + (T_{\text{air}} + 3.80) \text{ dayhr} (-0.00771) - \text{h} \cdot \text{VPD}/0.277$

Best (C35, $R^2=0.52$): $\hat{CH}_4 = 0.00931 \text{ dayhr} (\text{h} \cdot P_r - T_{\text{air}}) - \text{h} \cdot \text{VPD} + \exp(0.107 T_{\text{air}}) - \tanh(\sqrt{\text{SR}}) - 5.86 \times 10^{-5} T_{\text{air}} (\text{SR} \cdot v - \text{SR}) - \tanh(\text{h} \cdot \text{VPD} T_{\text{air}}) - 2 \text{ h} \cdot \text{VPD}$

C_n denotes expression complexity; R^2 is the 80/20 hold-out value. \hat{CH}_4 = predicted half-hourly CH_4 flux. The four tiers per site are PySR Pareto-front candidates (simplest → most accurate).