Supplementary file

Title: Drivers of greenhouse gas emissions in agricultural soils: The effect of residue management and soil type

Dharmendra Singh[12], Sangeeta Lenka[2]*, Rameshwar S kanwar[3], Shashi S. Yadav[1], Madhumonti Saha[2], Abhijit Sarkar[2], Dinesh Kumar Yadav[2], M Vassanda Coumar[2], Narendra Kumar Lenka[2*], Tapan Adhikari[2], Priyanka Yadon[2], Vijay Gami[2]

Affiliations of author

[1]Department of Soil Science and Agriculture Chemistry, College of Agriculture, Gwalior, Madhya Pradesh, India.
[2]ICAR- Indian Institute of Soil Science, Nabibagh, Bhopal, Madhya Pradesh-462038, India.
[3]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, USA-50011

*Corresponding authors: sangeeta.lenka@icar.gov.in; nklenka74@gmail.com

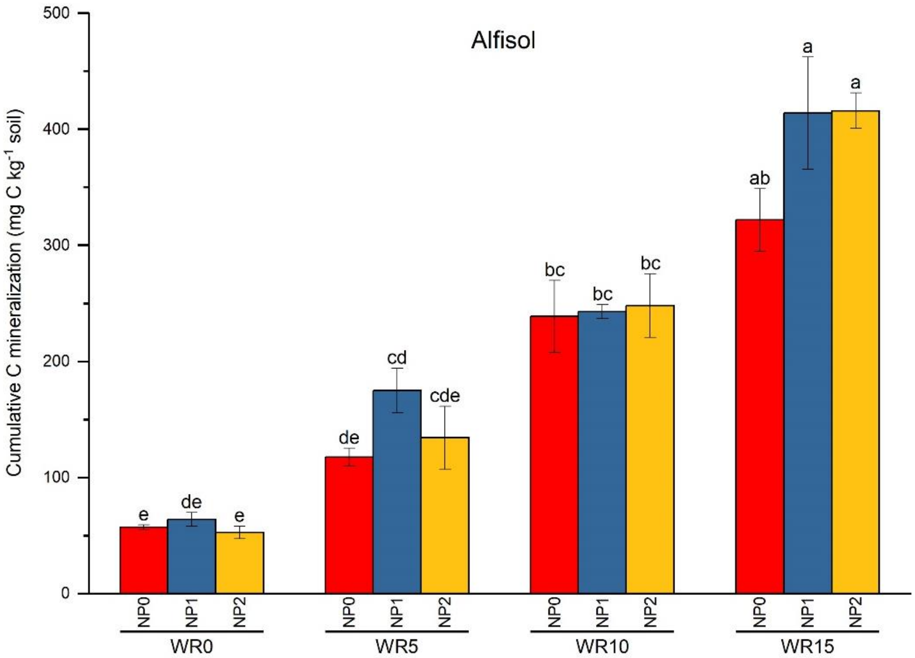
Methodology

Statistical analysis

In this study, Partial Least Squares (PLS) regression modeling was employed to assess the relative effects of management variables (residue and nutrient input, soil moisture), post-harvest soil parameters (NO_3-N, NH_4-N, MBC, labile C, and alkaline phosphatase), and inherited soil properties (pH, CEC, clay, sand, silt, total C, total N, legacy P, and K) on greenhouse gas emissions (CO_2, CH_4, N_2O). PLS was selected due to its ability to handle multicollinearity among explanatory variables and reduce dimensionality by decomposing the independent (X) and dependent (Y) variables into orthogonal latent factors. In Origin Pro 2024b, the model's parameters were optimized by selecting the number of latent factors based on the minimum root predicted residual error sum of squares (PRESS) during cross-validation, ensuring that the model was neither overfitted nor underfitted. Validation was carried out using k-fold cross-validation, which improved the model's predictive robustness. A Variable Importance in Projection (VIP) score threshold of >0.8 was used to identify key drivers of soil GHG emissions, in line with Gómez-Gener et al. (2018). This combination of parameter settings and validation methods allowed for the reliable identification of the most influential factors driving emissions, ensuring accurate and generalizable results.

Diagnostic plots in Origin Pro 2024b play a crucial role in validating and selecting the optimal Partial Least Squares (PLS) model by visually assessing model performance and fit. Key diagnostic plots such as Residual vs. Fitted and Predicted vs. Observed plots help evaluate how well the model predictions align with the actual data. For instance, a Residual vs. Fitted plot helps detect patterns in residuals, indicating potential model misspecification or non-linearity. In contrast, a Predicted vs. Observed plot clearly represents the model's predictive accuracy, with points ideally clustering along a 45-degree line if the model fits well.

Additionally, Variable Importance in Projection (VIP) scores are visualized to highlight the contribution of each predictor variable, enabling the identification of key drivers with VIP > 0.8. Scree plots, which display the explained variance for each latent factor, aid in selecting the optimal number of components by showing where the marginal gain in explained variance diminishes. Lastly, Root Mean Square Error of Prediction (RMSEP) and cross-validation plots help fine-tune the model by comparing predicted vs. actual performance across validation datasets. These diagnostic plots collectively ensure that the PLS model is robust, well-calibrated, and interpretable.



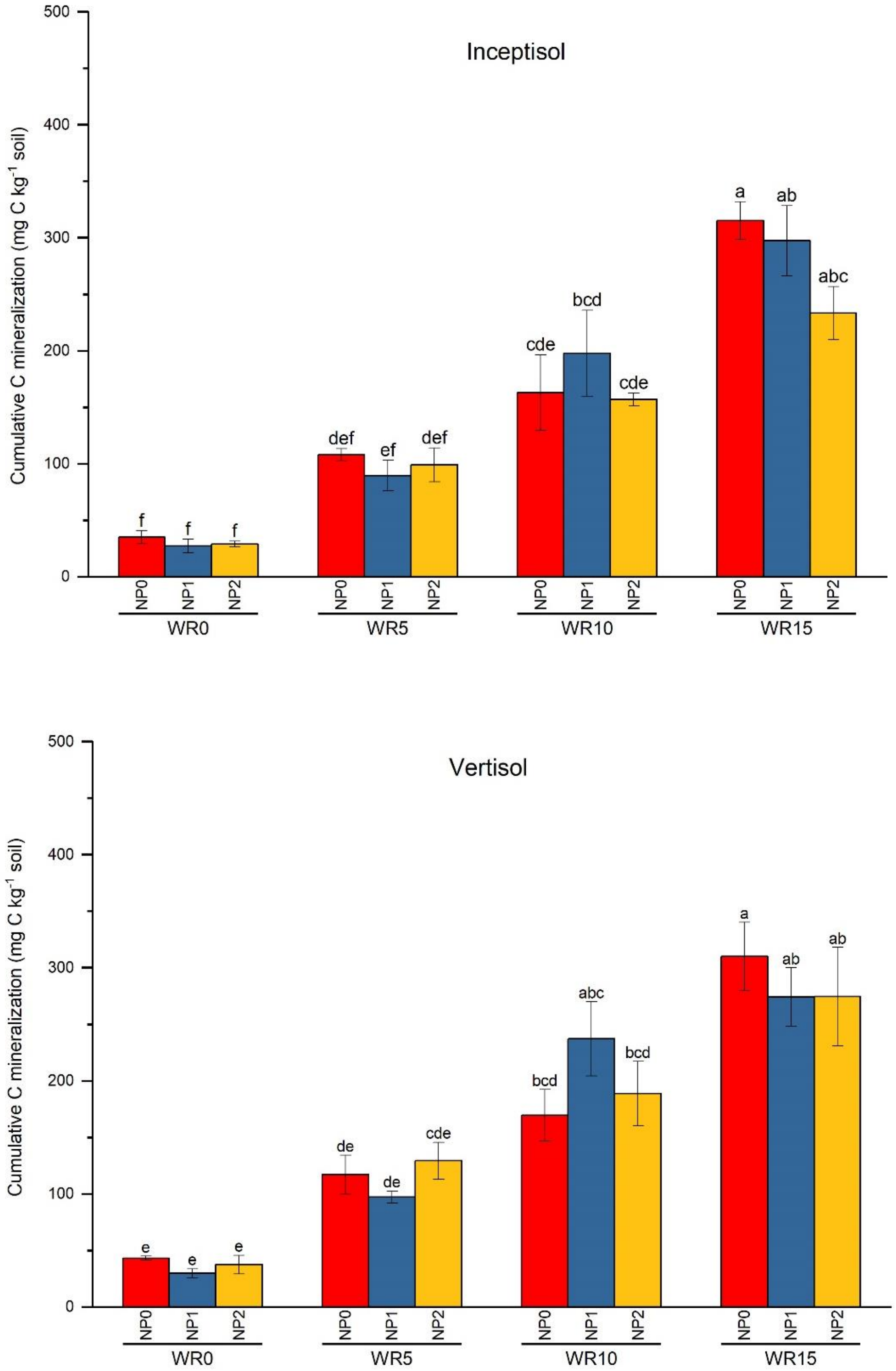


Fig. S1. Effect of wheat residue (WR0, WR5, WR10 and WR15) and nutrient (NP0, NP1 and NP2) inputs in three soil types (Vertisol, Alfisol, and Inceptisol) on total cumulative C mineralization (mg C kg^{-1} soil) over 96 days of incubation. Vertical bars represent the mean ± standard error (n=3). Different lower-case letters indicate significant differences among treatments at α<0.05.

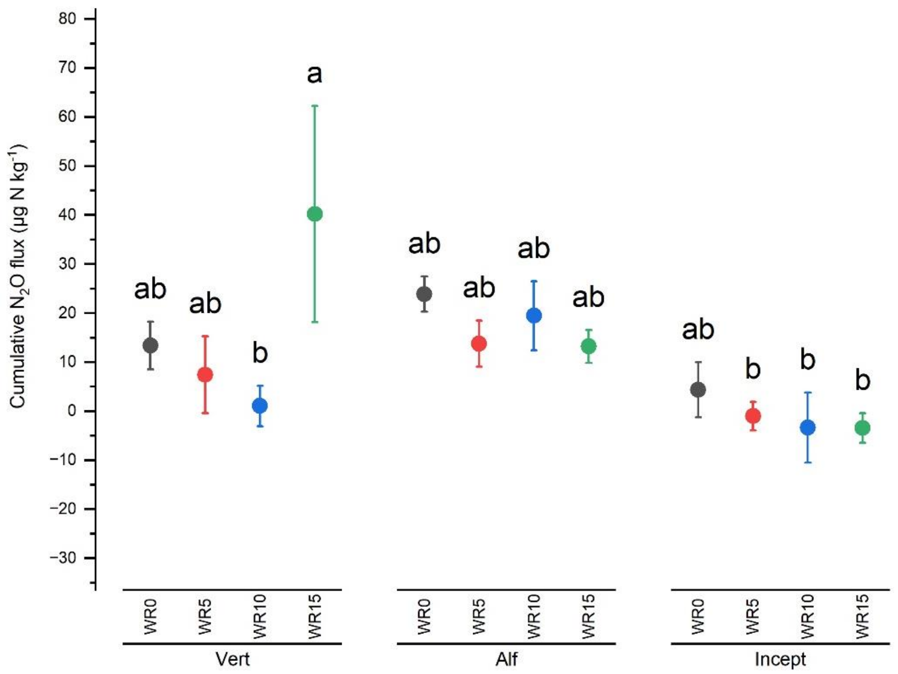


Fig. S2. Effect of wheat residue (WR0, WR5, WR10 and WR15) input and soil type (Vertisol: Vert; Alfisol: Alf; Inceptisol: Incept) on total cumulative N_2O flux (µg N kg^{-1} soil) averaged across nutrient input over 96 days of incubation. Vertical bars represent the mean ± standard error (n=3). Different lower-case letters indicate significant differences among treatments at α<0.05.

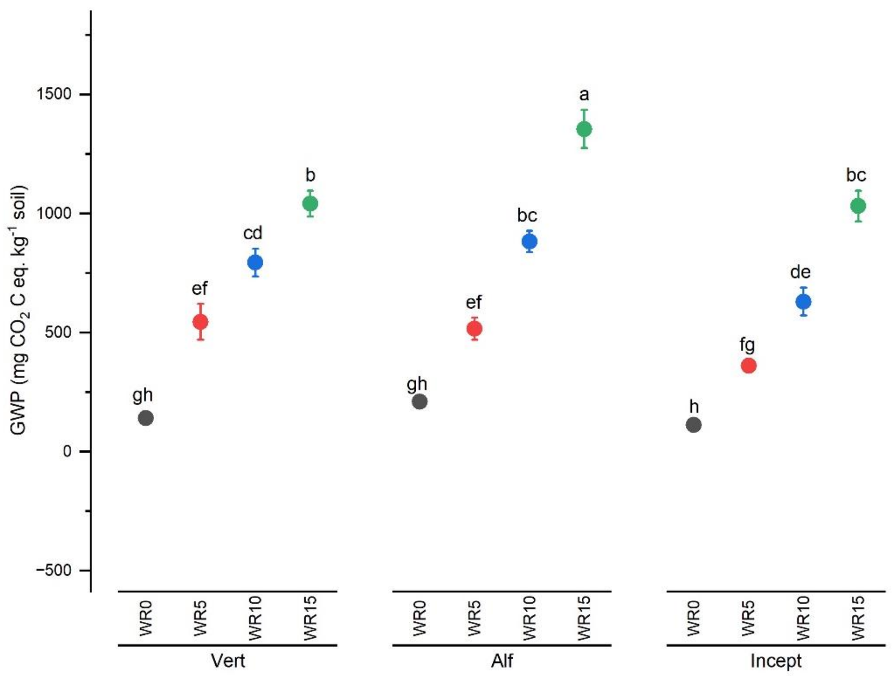
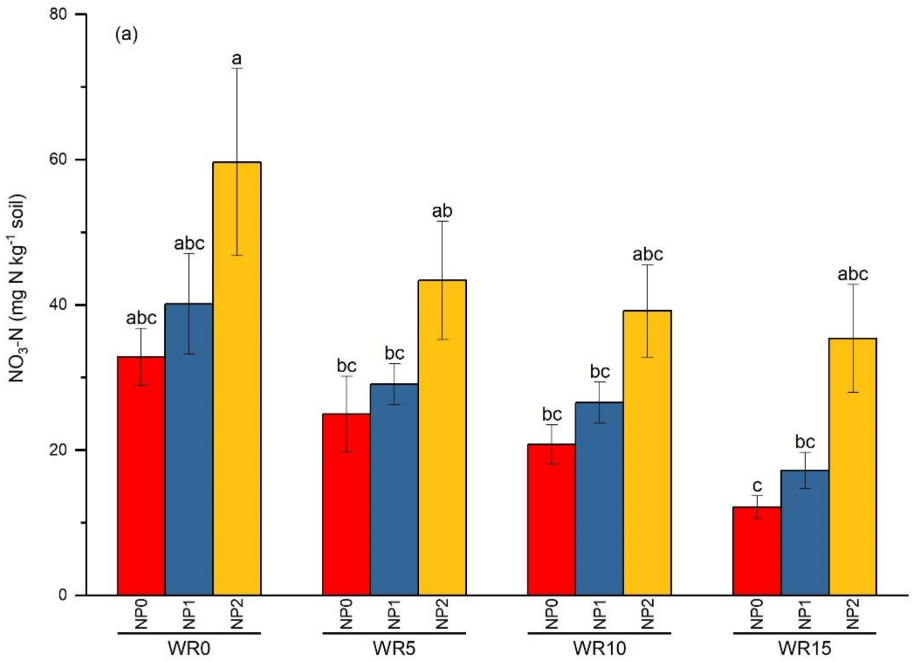


Fig. S3. Effect of wheat residue (WR0, WR5, WR10 and WR15) input and soil type (Vertisol: Vert; Alfisol: Alf; Inceptisol: Incept) on global warming potential (mg CO_2-C eq. kg^{-1} soil) averaged across nutrient input over 96 days of incubation. Vertical bars represent the mean ±

standard error (n=3). Different lower-case letters indicate significant differences among treatments at α<0.05.



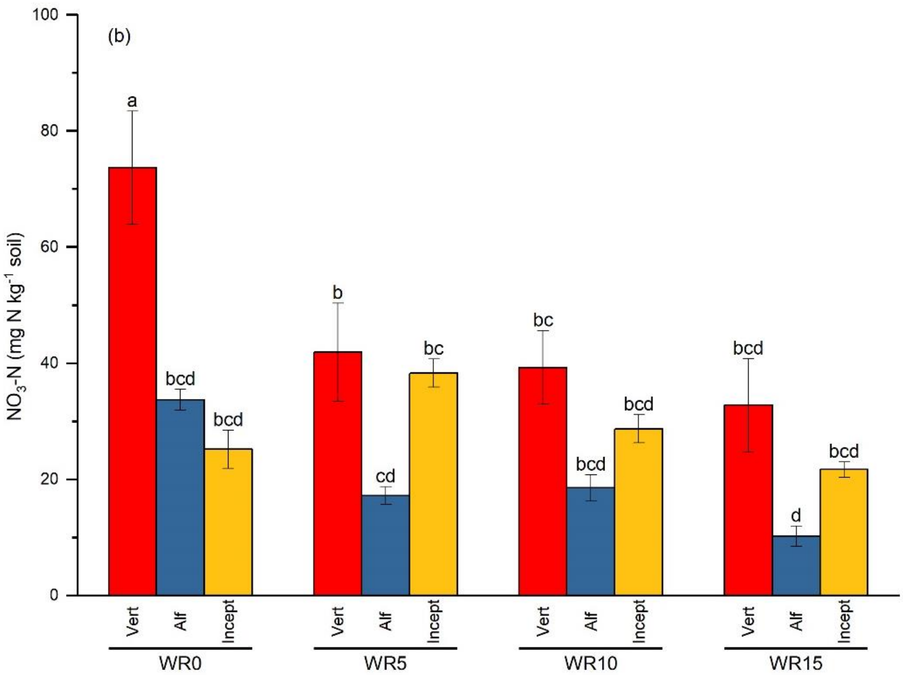


Fig. S4. Interaction effect of (a) wheat residue (WR) and nutrient (NP) input, and (b) wheat residue and soil type on post incubation soil NO_3-N (mg kg^{-1} soil). Vertical bars represent the mean ± standard error (n=3). Different lower-case letters indicate significant differences among treatments at α<0.05.

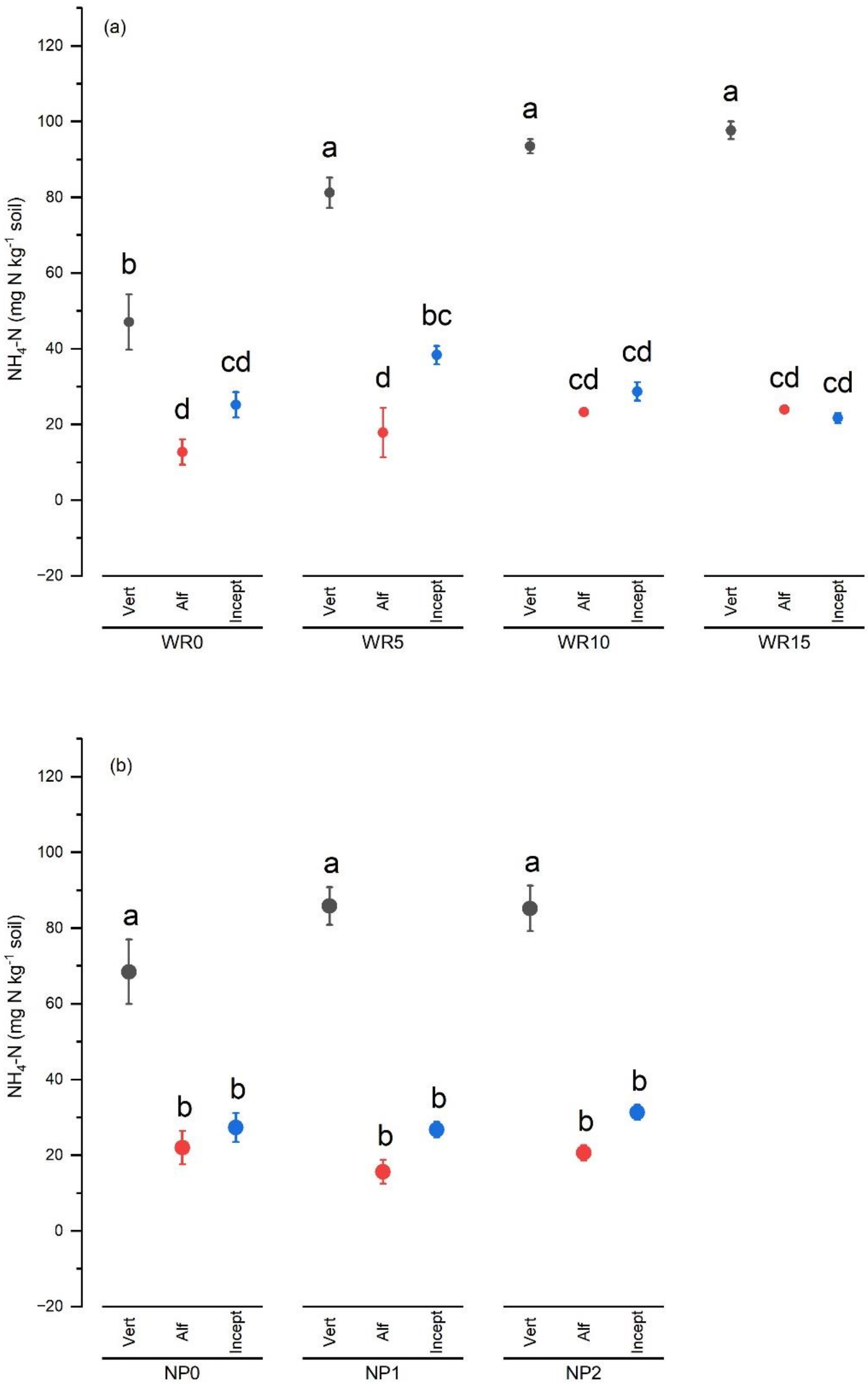
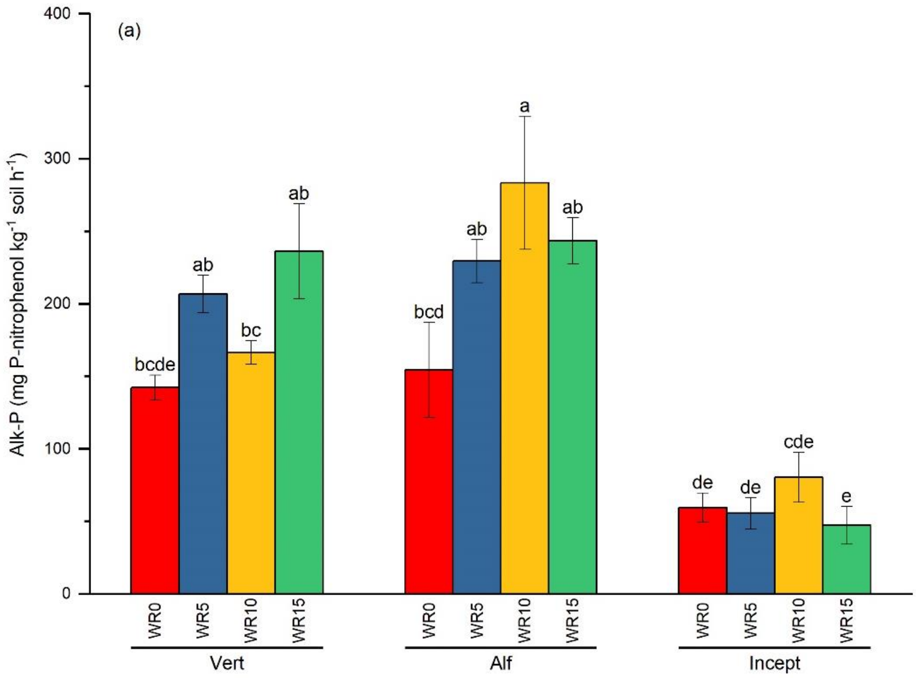


Fig. S5. Interaction effect of (a) wheat residue and soil type, nutrient input and (b) soil type on post incubation soil NH_4-N (mg kg^{-1} soil). Vertical bars represent the mean ± standard error (n=3). Different lower-case letters indicate significant differences among treatments at $\alpha<0.05$.



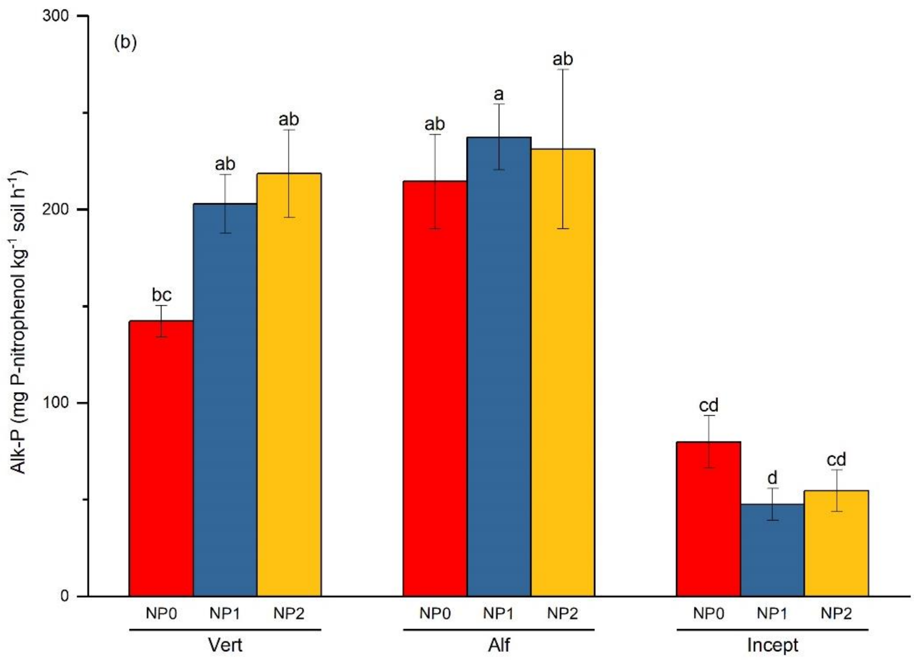


Fig. S6. Interaction effect of (a) wheat residue and soil type averaged across nutrient input, (b) nutrient input and soil type averaged across residue input on post incubation soil alkaline phosphatase (mg kg^{-1} soil). Vertical bars represent the mean ± standard error (n=3). Different lower-case letters indicate significant differences among treatments at $\alpha<0.05$.

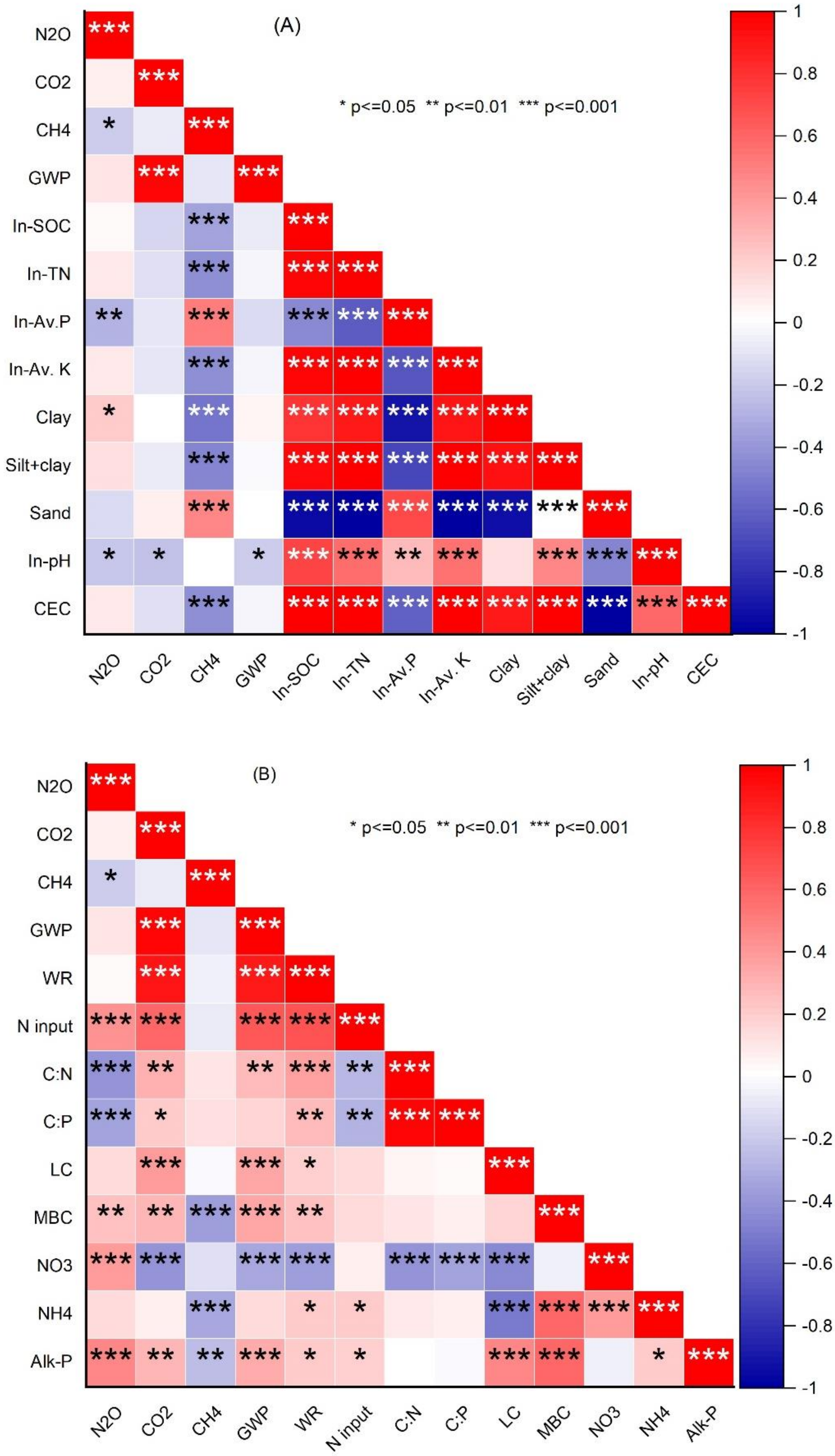


Fig. S7. Pearson correlation between (A) Natural drivers inherited soil properties and responses (CO_2, CH_4, N_2O, and GWP) (B) Anthropogenic variables (wheat residue: WR; nitrogen input; C: N: ratio of carbon and nitrogen; C: P: ratio of carbon and phosphorus, labile soil carbon: LC; microbial biomass carbon: MBC; nitrate N: NO_3-N; ammoniacal N: NH_4-N and alkaline phosphatase: alk-P) and responses (CO_2, CH_4, N_2O, and GWP).

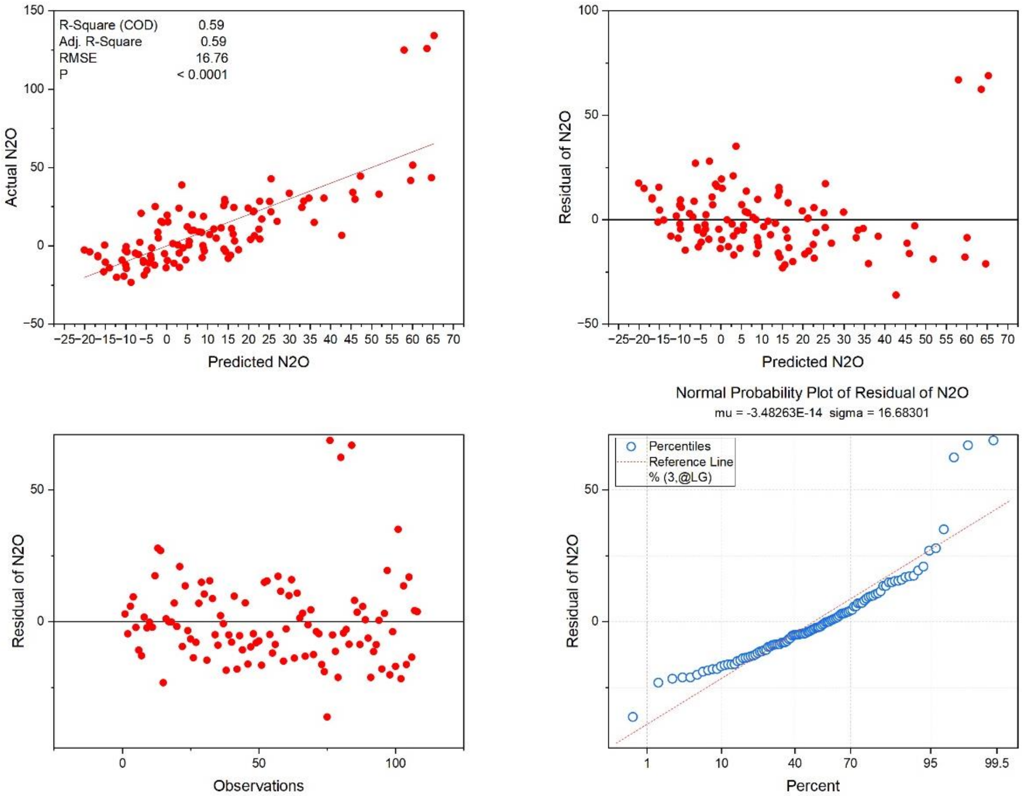


Fig. S8. Diagnostic plots of simulating N_2O (μg N kg^{-1} soil) in the partial least square regression model.

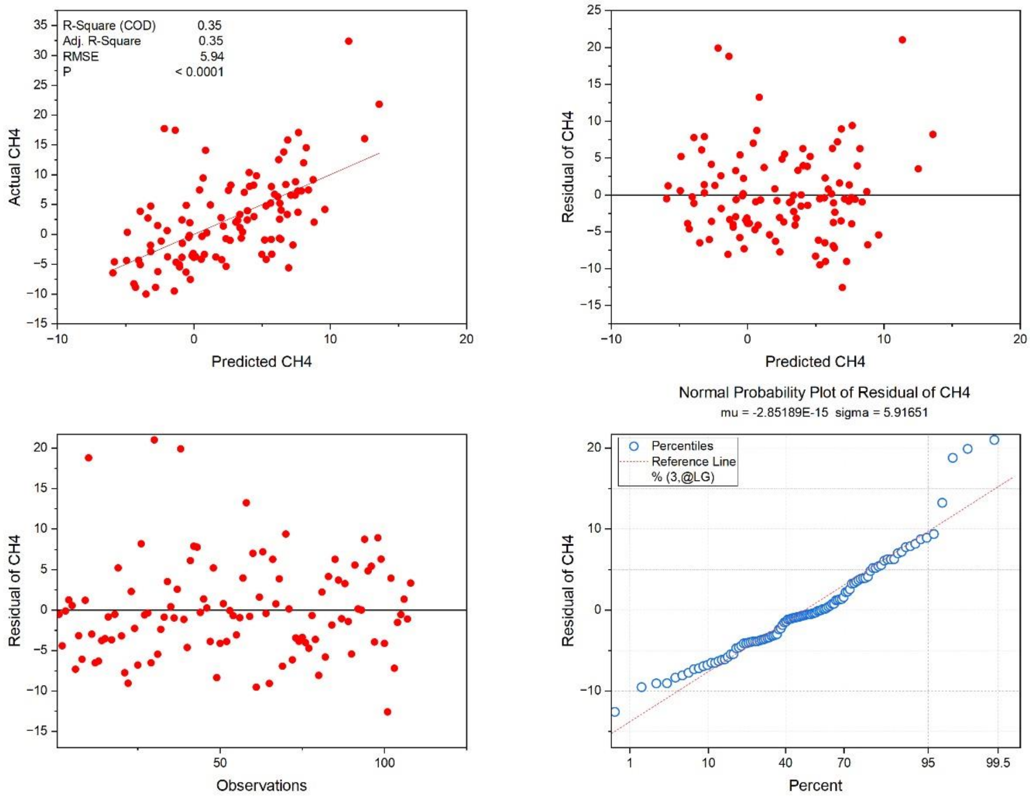


Fig. S9. Diagnostic plots of simulating CH_4 (μg C kg^{-1} soil) in the partial least square regression model.

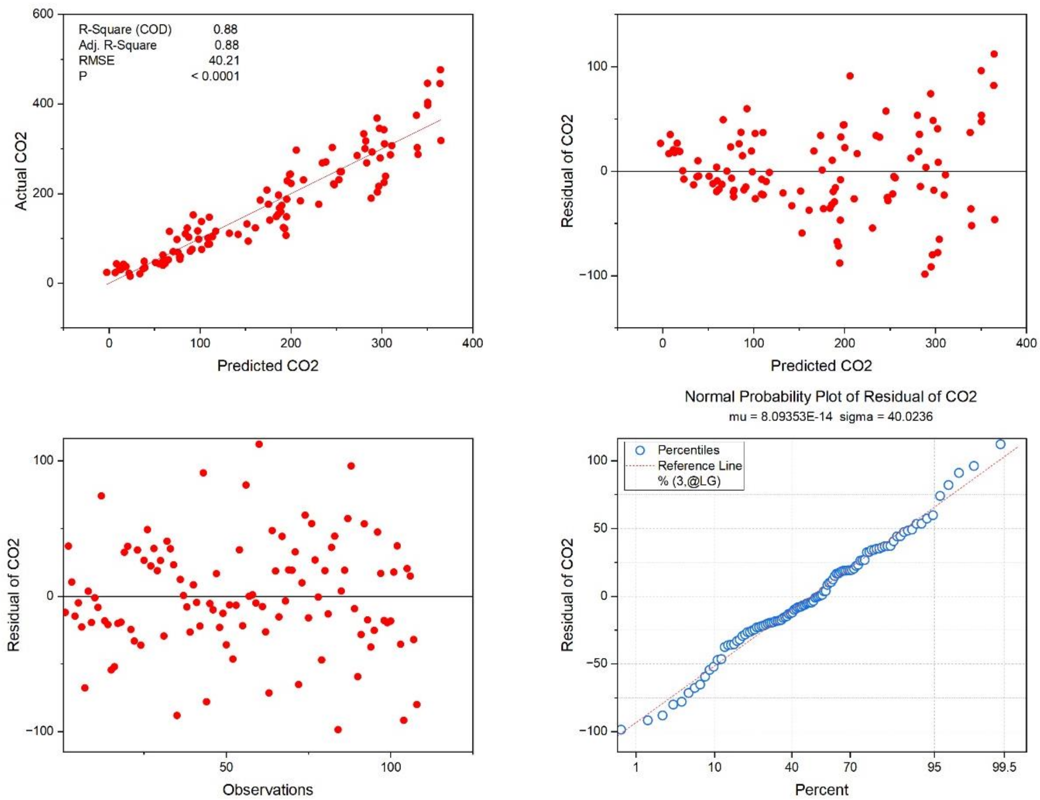


Fig. S10. Diagnostic plots of simulating CO_2 (mg C kg^{-1} soil) in the partial least square regression model.

Table S1

Nutrient stoichiometry of the residue returned treatments (WR0, WR5, WR10, and WR15) under the three nutrient levels (NP0, NP1 and NP2).

	NS0	NS1	NS2
	C/N/P	C/N/P	C/N/P
WR0	100 : 8.33 : 2.00	< 100 : 8.33 : 2.00	< 100 : 8.33 : 2.00
WR5	100 : 1.26 : 0.05	100 : 8.33 : 2.00	100 : 25 : 6
WR10	100 : 1.26 : 0.05	100 : 4.17 : 1.0	100 : 12.5 : 3
WR15	100 : 1.26 : 0.05	100 : 2.78 : 0.67	100 : 8.33 : 2.00
WR0	= 100 : 8.33 : 2.00	< 100 : 8.33 : 2.00	< 100 : 8.33 : 2.00
WR5	> 100 : 8.33 : 2.00	= 100 : 8.33 : 2.00	< 100 : 8.33 : 2.00
WR10	> 100 : 8.33 : 2.00	> 100 : 8.33 : 2.00	< 100 : 8.33 : 2.00
WR15	> 100 : 8.33 : 2.00	> 100 : 8.33 : 2.00	= 100 : 8.33 : 2.00

Table S2

Summary of ANOVA indicating source effects on soil CO_2, N_2O, CH_4 emissions, global warming potential (GWP), and relevant post incubation soil properties over 96 days of incubation.

Source of variation	DF	N_2O	CH_4	CO_2	GWP	Labile C	MBC	NO_3 -N	NH_4-N	Alk-P
WR	3	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Soil type	2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
NP	2	<0.0001	0.83857	0.29397	0.15248	0.33387	<0.0001	<0.0001	0.00122	0.01782
WR x Soil type	6	<0.0001	0.00154	0.03656	0.00747	0.0005	<0.0001	<0.0001	<0.0001	<0.0001
WR x NP	6	<0.0001	0.34313	0.75713	0.33204	0.1021	<0.0001	0.22332	0.0001	<0.0001
Soil type x NSL	4	<0.0001	0.05389	0.11187	0.00638	0.56593	<0.0001	<0.0001	<0.0001	<0.0001
WR x Soil type x NP	12	<0.0001	0.62866	0.10201	0.04756	0.01548	<0.0001	0.00851	<0.0001	<0.0001

Note: DF: degrees of freedom, WR: wheat residue input, NP: nitrogen and phosphorus input to balance the stoichiometry of C/N/P

Table S3

Variance explained for X and Y effects in the PLS model for estimation of N_2O flux.

Number of Factors	Variance Explained for X Effects (%)	Cumulative X Variance (%)	Variance Explained for Y Responses (%)	Cumulative Y Variance (%)
1	28.62	28.62	34.78	34.78
2	33.63	62.25	10.37	45.16
3	12.80	75.05	5.58	50.74
4	9.23	84.28	4.04	54.78
5	9.42	93.69	0.81	55.59
6	2.59	96.29	1.36	56.95
7	1.11	97.39	1.21	58.16
8	1.42	98.81	0.34	58.49
9	0.45	99.26	0.69	59.18

Table S4

Variance explained for X and Y effects in the PLS model for estimation of CH_4 flux.

Number of Factors	Variance Explained for X Effects (%)	Cumulative X Variance (%)	Variance Explained for Y Responses (%)	Cumulative Y Variance (%)
1	46.28057	46.28057	24.91708	24.91708
2	19.39828	65.67885	4.75391	29.67099
3	8.72634	74.40519	3.49346	33.16444
4	9.21816	83.62335	1.78169	34.94613
5	5.24263	88.86599	0.30251	35.24864

Table S5

Variance explained for X and Y effects in the PLS model for estimation of CO_2 flux.

Number of Factors	Variance Explained for X Effects (%)	Cumulative X Variance (%)	Variance Explained for Y Responses (%)	Cumulative Y Variance (%)
1	19.44	19.44	69.74	69.74
2	39.18	58.62	7.45	77.19
3	18.89	77.51	4.31	81.50
4	7.09	84.60	3.14	84.64
5	8.64	93.24	1.69	86.32
6	1.07	94.30	1.31	87.63