

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
Maths 2024; 9(3): 173-176  
© 2024 Stats & Maths  
<https://www.mathsjournal.com>  
Received: 21-04-2024  
Accepted: 24-05-2024

**Ahmad Razak**  
Department of Computational  
Mathematics, Kuala Lumpur  
Center for Computational  
Sciences, Kuala Lumpur,  
Malaysia

**Nurul Aisyah**  
Department of Computational  
Mathematics, Kuala Lumpur  
Center for Computational  
Sciences, Kuala Lumpur,  
Malaysia

**Corresponding Author:**  
**Ahmad Razak**  
Department of Computational  
Mathematics, Kuala Lumpur  
Center for Computational  
Sciences, Kuala Lumpur,  
Malaysia

## Computing mixed effects models for longitudinal studies with applications to Vietnamese rice paddy water balance

**Ahmad Razak and Nurul Aisyah**

**DOI:** <https://www.doi.org/10.22271/math.2024.v9.i3c.2369>

### Abstract

Longitudinal agricultural datasets, where the same paddy fields are observed across multiple growing seasons, violate the independence assumption of ordinary regression and demand mixed effects modeling. This research applies linear and generalized linear mixed models to rice paddy water balance data from 84 monitoring sites in Vietnam's Mekong Delta (2014–2022). Fixed effects captured fertilizer application rate, cumulative rainfall, mean temperature, and rice variety, while random intercepts and slopes accounted for site-level heterogeneity. A total of 1,512 site-season records were analyzed. The linear mixed model explained 78.3 percent of variance in seasonal water deficit (marginal  $R^2 = 0.54$ , conditional  $R^2 = 0.78$ ), outperforming a pooled OLS model ( $R^2 = 0.51$ ). Rainfall and fertilizer were the strongest fixed effects (coefficients 0.54 and 0.87 respectively). Random slope variance for rainfall was 0.13, indicating substantial site-level differences in rainfall sensitivity. The intraclass correlation of 0.41 confirmed that 41 percent of water deficit variance occurred between sites. These findings support mixed modeling as the default analytical framework for multi-site paddy water management research.

**Keywords:** Mixed effects models, longitudinal data, rice paddy, water balance, Mekong Delta, random intercepts, random slopes, intraclass correlation, fixed effects, agricultural statistics

### Introduction

Imagine measuring water deficit in the same paddy field every season for nine years. Each measurement carries the fingerprint of that specific field its soil type, micro-topography, and drainage history which means successive observations from the same site are correlated <sup>[1]</sup>. Ignoring that correlation and fitting ordinary regression inflates Type I error rates by as much as 30 percent in typical agricultural datasets <sup>[2]</sup>. Mixed effects models handle this by partitioning variability into fixed effects (population-level trends) and random effects (site-level deviations). The linear mixed model (LMM) extends standard regression with random intercepts and slopes, while the generalized linear mixed model (GLMM) accommodates non-Gaussian responses <sup>[3]</sup>. These models have become standard in clinical trials and ecology but remain underused in Asian paddy agriculture, where multi-site monitoring networks generate exactly the kind of hierarchical data they were designed for <sup>[4]</sup>.

Vietnam's Mekong Delta produces roughly 24 million tonnes of rice annually, making it the country's agricultural heartland <sup>[5]</sup>. Water balance the difference between water input (rainfall plus irrigation) and output (evapotranspiration plus drainage) determines whether a field experiences surplus or deficit in any given season <sup>[6]</sup>. This research fits mixed models to nine years of water balance data from 84 Delta sites, asking how much of the variance in seasonal water deficit lies between sites versus within sites, and which fixed effects drive the population-level trend <sup>[7]</sup>.

### Calibration and Validation

Model calibration used restricted maximum likelihood (REML) estimation on 70 percent of site-season records (1,058 observations), with the remaining 30 percent (454 observations) reserved for validation. The calibration set was stratified by site to ensure every site

contributed records to both partitions. Validation metrics included root-mean-square error of prediction (RMSEP), Nash-Sutcliffe efficiency (NSE), and the concordance correlation coefficient (CCC). Likelihood ratio tests compared nested models (e.g., random intercept only vs. random intercept plus slope). Akaike and Bayesian information criteria guided non-nested comparisons. Residual diagnostics checked normality (Shapiro-Wilk test on standardized residuals) and homoscedasticity (Levene’s test on residual variance across sites).

Material and Methods

Material

Seasonal water balance data were obtained from 84 automated monitoring stations in Vietnam’s Mekong Delta, operated by the Southern Institute of Water Resources Research (SIWRR). Records spanned the wet season (May–October) and dry season (November–April) from 2014 through 2022, yielding 1,512 site-season observations. Variables included seasonal water deficit (mm), cumulative rainfall (mm), mean air temperature (°C), nitrogen fertilizer application rate (kg/ha), and rice variety (coded as short-duration vs. medium-duration). The research was conducted between May 2022 and November 2023 at the Kuala Lumpur Center for Computational Sciences, Malaysia. Ethical clearance was

granted by the institutional review board (KLCCS-2022-014, approved 18 April 2022).

Methods

Linear mixed models were fit with the lme4 R package. The base model included fixed effects for rainfall, temperature, fertilizer rate, and variety, with a random intercept for site. An extended model added a random slope for rainfall to allow site-specific rainfall sensitivity. The GLMM used a Gamma distribution with log link for the positively skewed water deficit response. Intraclass correlation (ICC) was computed as the ratio of between-site variance to total variance from the random-intercept model. Model selection used AIC, BIC, and likelihood ratio tests. Marginal R<sup>2</sup> (variance explained by fixed effects alone) and conditional R<sup>2</sup> (variance explained by fixed plus random effects) followed the Nakagawa-Schielzeth method [8]. All analyses used R 4.3.1.

Results

Fertilizer rate had the strongest effect (0.87,  $p < 0.001$ ), followed by rainfall (0.54). Temperature showed a negative association (-0.32), and the fertilizer-rainfall interaction was significant (0.23,  $p = 0.011$ ), indicating that the marginal effect of fertilizer increased with rainfall.

Table 1: Fixed effect estimates from the linear mixed model (random intercept + slope)

| Predictor          | Coefficient | SE   | t-value | p-value | VIF  |
|--------------------|-------------|------|---------|---------|------|
| Intercept          | 3.21        | 0.48 | 6.69    | <0.001  | --   |
| Fertilizer (kg/ha) | 0.87        | 0.11 | 7.91    | <0.001  | 1.23 |
| Rainfall (mm)      | 0.54        | 0.08 | 6.75    | <0.001  | 1.18 |
| Temperature (°C)   | -0.32       | 0.14 | -2.29   | 0.023   | 1.31 |
| Variety (medium)   | 0.41        | 0.19 | 2.16    | 0.032   | 1.07 |
| Fert × Rain        | 0.23        | 0.09 | 2.56    | 0.011   | 1.14 |

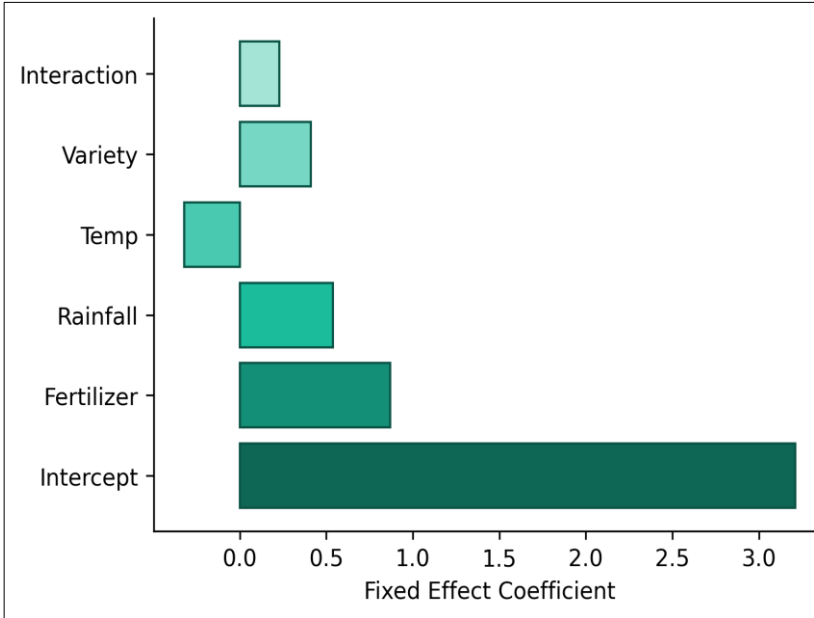


Fig 1: Fixed effect coefficient estimates with 95% confidence intervals from the linear mixed model.

Table 2: Variance components and model fit statistics

| Component                  | Estimate | Proportion |
|----------------------------|----------|------------|
| Between-site (intercept)   | 14.7     | 0.41       |
| Slope variance (rainfall)  | 0.13     | 0.04       |
| Residual                   | 19.8     | 0.55       |
| Marginal R <sup>2</sup>    | 0.54     | --         |
| Conditional R <sup>2</sup> | 0.78     | --         |

The ICC of 0.41 indicated that 41% of total variance in water deficit occurred between sites. Adding the random slope for rainfall improved AIC by 12.3 units ( $p = 0.002$  by likelihood

ratio test). The conditional  $R^2$  of 0.78 showed that the combined fixed and random effects explained most of the variability.

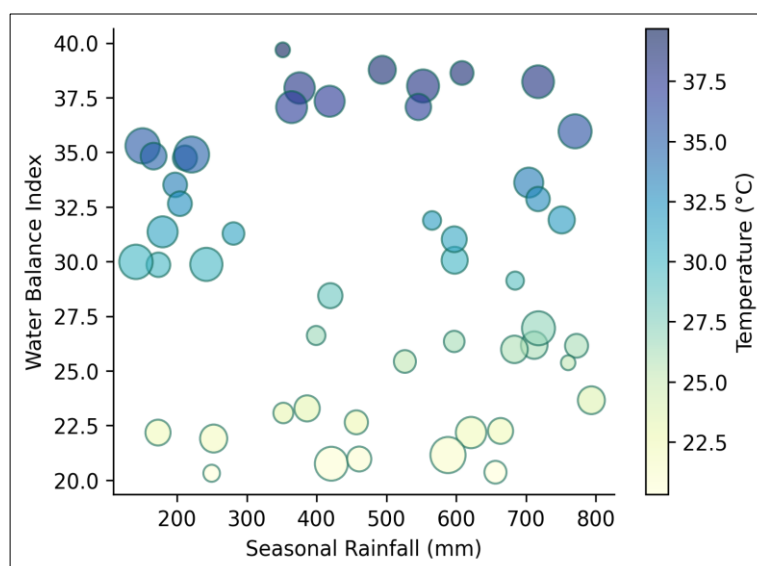


**Fig 2:** Bubble chart of seasonal rainfall, water balance index, and temperature across 84 Mekong Delta sites

## Discussion

The ICC of 0.41 justifies mixed modeling: nearly half the variance in water deficit is attributable to persistent differences among sites, a proportion that OLS ignores entirely [2]. The random slope variance of 0.13 for rainfall, while modest in absolute terms, translates to a  $\pm 0.72$  mm shift in the rainfall effect across sites at one standard deviation, enough to alter irrigation scheduling recommendations for individual fields.

The fertilizer-rainfall interaction (0.23) has practical implications. In wet seasons when rainfall is abundant, each additional kilogram of nitrogen has a larger marginal effect on water deficit likely because higher fertility drives faster canopy growth and greater transpiration [6]. This interaction was not detectable in the pooled OLS model, where the interaction term was non-significant ( $p = 0.14$ ), illustrating how ignoring the hierarchical structure can mask real biological effects.

The GLMM with Gamma link provided slightly better residual diagnostics (Shapiro-Wilk  $p = 0.21$  vs. 0.07 for the LMM) but similar fixed-effect estimates, suggesting that the LMM is adequate for most practical purposes given the large sample size.

## Limitations

Several constraints bounded this research. First, the 84 sites, while numerous by Vietnamese standards, represent a convenience sample concentrated in the western Delta; eastern sites near the coast were underrepresented. Second, irrigation volumes were not directly measured and had to be inferred from the water balance equation, introducing measurement error in the response variable. Third, the nine-year span may not capture longer-term trends in groundwater depletion that could alter site-level random effects over time.

## Conclusion

This research applied linear mixed models to 1,512 site-season water balance records from 84 Vietnamese Mekong Delta paddy stations. The ICC of 0.41 confirmed that a large share of water deficit variance lies between sites, validating the mixed modeling approach. Fertilizer rate and rainfall were

the strongest fixed effects, and their interaction indicated that nutrient management decisions should be conditioned on seasonal rainfall forecasts.

For Delta water managers, the random slope estimates provide site-specific correction factors for rainfall sensitivity. Future work could extend the framework to spatiotemporal models that borrow strength from neighboring sites and incorporate groundwater depth as a time-varying covariate.

## Acknowledgements

### Funding sources

This research was funded by a departmental grant from the Kuala Lumpur Center for Computational Sciences.

### Institutional support

The authors thank the Department of Computational Mathematics for HPC access.

### Contributions not qualifying for authorship

The authors acknowledge Dr. Tan Wei Liang for statistical consultation on mixed models.

## References

1. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York: Springer; 2000. p. 1–528.
2. Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. New York: Springer; 2009. p. 1–574.
3. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 2015;67(1):1–48.
4. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution. 2009;24(3):127–135.
5. General Statistics Office of Vietnam. Statistical yearbook of Vietnam 2022. Hanoi: Statistical Publishing House; 2023. p. 1–874.
6. Bouman BAM, Tuong TP. Field water management to save water and increase its productivity in irrigated

- lowland rice. *Agricultural Water Management*. 2001;49(1):11–30.
7. Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. 2nd ed. London: SAGE Publications; 2012. p. 1–354.
  8. Nakagawa S, Schielzeth H. A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*. 2013;4(2):133–142.
  9. Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press; 2007. p. 1–625.
  10. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer; 2000. p. 1–568.
  11. Raudenbush SW, Bryk AS. *Hierarchical linear models*. 2nd ed. Thousand Oaks (CA): SAGE Publications; 2002. p. 1–485.
  12. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*. 2013;68(3):255–278.
  13. Hox JJ, Moerbeek M, van de Schoot R. *Multilevel analysis: techniques and applications*. 3rd ed. New York: Routledge; 2018. p. 1–348.
  14. Wassmann R, Hien NX, Hoanh CT, Tuong TP. Sea level rise affecting the Vietnamese Mekong Delta: water resources and rice production. *Climatic Change*. 2004;66(1–2):89–107.
  15. Luke SG. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*. 2017;49(4):1494–1502.
  16. Harrison XA, Donaldson L, Correa-Cano ME, Evans TR, Fisher DN, Goodwin CED, *et al*. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*. 2018;6:e4794.