African Crop Science Journal, Vol. 34, No. 3, pp. 651 - 674
pISSN 1021-9730/2026
eISSN 2072-6589/2026
© 2026, African Crop Science Society

African Crop Science Journal by African Crop Science Society is licensed under a Creative Commons Attribution 3.0 Uganda License.
Based on a work at www.ajol.info/
DOI: https://dx.doi.org/10.4314/acsj.v34i3.15

RELIABILITY OF RICE YIELD MONITORING UNDER CLIMATE NON-STATIONARITY IN UGANDA: A HYBRID MODELLING APPROACH

FIABILITÉ DU SUIVI DU RENDEMENT DU RIZ SOUS NON-STATIONNARITÉ CLIMATIQUE EN OUGANDE: UNE APPROCHE DE MODÉLISATION HYBRIDE

J. NDHOKERO [1,2], L. NKPORDEE [1] and A. GOPARAJU [1]

[1] Department of Mathematics and Statistics, Kampala International University, Kansanga, P. O. Box 20000, Kampala, Uganda
[2] National Crops Resources Research Institute, P. O. Box 7084, Kampala, Uganda
Corresponding author: ndhokero@gmail.com

(*Received 6 May 2026; accepted 26 May 2026*)

ABSTRACT

Climate change increasingly threatens the reliability of agricultural monitoring systems by altering the environmental conditions under which historical climate–yield relationships were established. In climate-sensitive smallholder production systems, monitoring models calibrated under past conditions may become unreliable as climatic patterns shift. This study assessed the reliability of statistical, machine-learning, and hybrid modelling approaches for monitoring rice yields in Uganda under climate non-stationarity. A 30-year district-level dataset (1995–2024) from Bugiri, Butaleja, Lira, and Nwoya was analysed using Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Random Forest (RF), Support Vector Machine (SVM), and a hybrid statistical–machine-learning framework. Reliability was evaluated using out-of-sample predictive performance, statistical comparison, residual bias diagnostics, spatial transferability, climatic stress robustness, and spatial vulnerability under heterogeneous perturbation conditions. The hybrid framework demonstrated the strongest overall monitoring reliability, with the lowest systematic bias (mean error = 0.18), stable residual behaviour, and strong spatial transferability. Although Random Forest showed marginally lower perturbation-related error escalation, the hybrid framework maintained superior multidimensional monitoring reliability overall. In contrast, ARIMAX exhibited substantial positive bias (mean error = 1.24) and marked deterioration under intensified climatic perturbations (61.7% increase in RMSE). Diebold–Mariano tests confirmed that ARIMAX had significantly poorer predictive performance than adaptive alternatives ($P < 0.05$). SPI, SPEI, and temperature-related variables emerged as influential predictors. These findings show that predictive accuracy alone is insufficient for evaluating agricultural monitoring under changing climatic conditions. Adaptive hybrid monitoring frameworks provide a stronger basis for climate-risk monitoring, early warning, and food security planning.

Key Words: Diebold-Mariano test, distribution shift, hydro-climatic stress, monitoring reliability, spatial transferability

RÉSUMÉ

Le changement climatique menace de plus en plus la fiabilité des systèmes de suivi agricole en modifiant les conditions environnementales sous lesquelles les relations historiques entre le climat et le rendement ont été établies. Dans les systèmes de production des petits exploitants sensibles au climat, les modèles de suivi étalonnés dans des conditions passées peuvent devenir peu fiables à mesure que les régimes climatiques changent. Cette étude a évalué la fiabilité des approches de modélisation statistique, d'apprentissage automatique (*machine learning*) et hybride pour le suivi des rendements du riz en Ougande sous non-stationnarité climatique. Un ensemble de données de 30 ans à l'échelle des districts (1995-2024) provenant de Bugiri, Butaleja, Lira et Nwoya a été analysé à l'aide d'un modèle autorégressif intégré de moyenne mobile avec variables exogènes (ARIMAX), de forêts d'arbres décisionnels (*Random Forest* - RF), de machines à vecteurs de support (SVM) et d'un cadre hybride associant statistiques et apprentissage automatique. La fiabilité a été évaluée à l'aide des performances de prédiction hors échantillon, de comparaisons statistiques, de diagnostics de biais résiduels, de la transférabilité spatiale, de la robustesse au stress climatique et de la vulnérabilité spatiale dans des conditions de perturbations hétérogènes. Le cadre hybride a démontré la plus forte fiabilité globale de suivi, avec le biais systématique le plus faible (erreur moyenne = 0,18), un comportement des résidus stable et une forte transférabilité spatiale. Bien que l'approche *Random Forest* ait montré une augmentation de l'erreur liée aux perturbations légèrement inférieure, le cadre hybride a maintenu une fiabilité de suivi multidimensionnelle supérieure dans l'ensemble. En revanche, le modèle ARIMAX a affiché un biais positif substantiel (erreur moyenne = 1,24) et une détérioration marquée sous des perturbations climatiques intensifiées (augmentation de 61,7 % de la RMSE). Les tests de Diebold-Mariano ont confirmé que le modèle ARIMAX présentait des performances prédictives significativement inférieures à celles des alternatives adaptatives ($P < 0,05$). Le SPI, le SPEI et les variables liées à la température sont apparus comme des prédicteurs influents. Ces résultats montrent que la précision prédictive seule est insuffisante pour évaluer le suivi agricole dans des conditions climatiques changeantes. Les cadres de suivi hybrides adaptatifs fournissent une base plus solide pour le suivi des risques climatiques, l'alerte précoce et la planification de la sécurité alimentaire.

Mots Clés: Test de Diebold-Mariano, décalage de distribution, stress hydro-climatique, fiabilité du suivi, transférabilité spatiale

INTRODUCTION

Climate change is increasingly reshaping environmental systems through rising temperatures, altered rainfall patterns, and more frequent extreme weather events. These changes affect ecological and agricultural processes by altering the relationships between environmental factors and system responses over time (Challinor *et al.*, 2014; Intergovernmental Panel on Climate Change (IPCC), 2023). In agricultural systems sensitive to climate, productivity depends on complex interactions among temperature, moisture availability, and management conditions. Extreme heat and drought can significantly lower yields (Ciais *et al.*, 2005; Lesk *et al.*, 2016; Zhao *et al.*, 2017).

Environmental monitoring systems are usually developed using historical data to track system behaviour and support forecasting, risk assessment, and management decisions. Such systems often rely, either explicitly or implicitly, on the idea that environmental relationships remain fairly stable over time. However, under climate non-stationarity, this assumption becomes harder to maintain. Changes in climate conditions may affect the strength, direction, or structure of climate-response relationships. This can reduce the reliability of models calibrated under historical circumstances (Milly *et al.*, 2008;

Koutsoyiannis, 2013; Haddeland *et al*., 2014). Structural breaks and shifts in environmental time series complicate inference when a single model structure is expected to remain valid across changing conditions (Hansen, 2001; Bai and Perron, 2003).

In predictive environmental modelling, declining performance under changing conditions can result from dataset shifts, covariate shifts, or concept drift, in which the statistical properties of the observed data change over time (Quionero-Candela *et al.*, 2009; Moreno-Torres *et al*., 2012; Subbaswamy and Saria, 2020). In time-series applications, standard validation methods may overestimate predictive ability when temporal dependence and structural changes are not properly accounted for (Bergmeir and Benítez, 2012; Bontempi *et al*., 2013). Therefore, reliable monitoring systems should be evaluated not just on predictive accuracy under historical conditions, but also on temporal stability, resilience to distributional shifts, and transferability across different environmental contexts.

Agricultural systems add complexity because crop responses to climate stress are often nonlinear, asymmetric, and influenced by threshold effects. Rice production is particularly relevant because yields are highly sensitive to hydro-climatic variability, especially rainfall fluctuations, drought stress, and temperature anomalies (Lobell and Burke, 2010; Lesk *et al.*, 2016; Zhao *et al*., 2017). Drought indicators such as the Standardised Precipitation Index (SPI) and the Standardised Precipitation Evapotranspiration Index (SPEI) are useful for identifying climate-related stress in crop production systems (McKee *et al*., 1993; Guttman, 1999; Vicente-Serrano *et al*., 2010). However, climate-yield relationships estimated under historical environmental conditions may become unreliable when applied beyond the calibration range (Lobell and Burke, 2010).

Eastern Africa has experienced increasing rainfall variability, recurrent droughts, and rising temperature anomalies, making the region crucial for assessing the reliability of monitoring under climate non-stationarity (Mubiru *et al*., 2012; Funk *et al*., 2015). In Uganda, these climatic disruptions have resulted in delayed rainfall onset, seasonal variability, recurring droughts, and sporadic flooding, especially in eastern and northern production zones where rice productivity is highly climate-sensitive (UBOS, 2021). These changing conditions pose an important practical question: can climate-sensitive agricultural monitoring systems remain reliable when environmental conditions increasingly differ from historical patterns?

Model selection brings important trade-offs in environmental monitoring. Traditional statistical methods, like autoregressive time-series models, provide understandable representations of temporal dependence but may struggle to capture nonlinear and interacting environmental responses (Lobell and Burke, 2010). Machine-learning methods offer greater flexibility for modelling complex nonlinear relationships and high-dimensional predictor interactions (Chlingaryan *et al*., 2018; Reichstein *et al*., 2019). However, their predictive reliability may decline when environmental conditions move far beyond the calibration domain because of concept drift and changing data distributions (Sugiyama and Kawanabe, 2012; Gama *et al*., 2014). Hybrid modelling approaches have emerged as a promising framework, combining structured temporal modelling with adaptive nonlinear learning (Dietterich, 2000; Reichstein *et al.*, 2019).

Reliable environmental monitoring has become increasingly important for climate services, early warning systems, climate-smart agriculture, and agricultural risk management (Lipper *et al*., 2014; Vaughan and Dessai, 2014). In smallholder farming systems,

monitoring outputs inform yield forecasting, climate-risk communication, adaptation planning, and weather-index insurance decisions. Since these applications rely on stable and understandable environmental signals, monitoring systems should be evaluated for operational resilience rather than just predictive performance (Nichols and Williams, 2006; Lindenmayer and Likens, 2010).

Despite growing interest in climate-impact modelling and predictive analytics, relatively few studies have directly examined monitoring reliability under changing climate conditions. Most climate-yield studies focus on predictive accuracy or long-term yield projections based on historical or scenario-based conditions, with limited attention to temporal stability, resilience to distributional shifts, and transferability across diverse agro-ecological settings. This creates an important analytical gap, as models that perform well under current conditions may lose reliability as climate-yield relationships evolve.

This study addresses that gap by evaluating the reliability of climate-yield monitoring in Uganda's smallholder rice systems under climate non-stationarity, using statistical, machine-learning, and hybrid modelling frameworks. Specifically, the study aims to: (i) assess monitoring reliability amid observed environmental variability; (ii) evaluate stability across diverse production environments; and (iii) examine robustness under climate-driven distribution shifts using RCP 4.5 and RCP 8.5 analogue perturbation scenarios as climatic stress sensitivity tests.

MATERIALS AND METHODS

Study area. The analysis used data from four major rice-producing districts in Uganda: Bugiri and Butaleja in Eastern Uganda, and Lira and Nwoya in Northern Uganda. These districts have different agro-ecological and production environments within Uganda's smallholder rice systems. Rice productivity here is significantly influenced by rainfall variability, temperature conditions, and drought-related moisture stress (Challinor *et al*., 2014; Ray *et al*., 2015; Babyenda *et al*., 2023).

Three rice production systems were examined: rain-fed lowland, irrigated lowland, and upland systems. Rain-fed lowland systems depend on seasonal rainfall and are especially susceptible to rainfall variability and drought. Irrigated lowland systems enjoy some control over water, but climatic fluctuations and water management challenges can still affect productivity. Upland systems rely mostly on rainfall and typically experience greater moisture stress and greater temperature variability. Differences in climate sensitivity between upland and lowland rice systems have been noted in other climate-sensitive agricultural areas where water availability is crucial for productivity and adaptation (Lobell and Burke, 2010; Zhao *et al*., 2017). Evidence from Ugandan rice systems shows that differing water-stress conditions affect how farmers manage their crops and the resulting yields across production environments (Ndhokero *et al.*, 2026).

The selected districts vary in rainfall patterns, temperature, and water availability. This makes them suitable for evaluating the reliability of climate–yield monitoring in diverse and changing climatic situations. It aligns with broader findings that climate variability affects crop productivity and can disrupt long-term stability in historical climate–yield relationships (Lobell and Burke, 2010; Challinor *et al*., 2014; Ray *et al*., 2015).

Monitoring data and environmental variables. A district-level dataset for 1995–2024 was created by combining climate observations, crop productivity records, and contextual production system information. Using multiple data sources increased temporal consistency and better represented climate–yield relationships across different smallholder rice production systems.

Climate data came from the NASA Prediction of Worldwide Energy Resources (NASA POWER) database. This database

offers satellite-derived and reanalysis-based meteorological data widely used in agricultural and environmental research (Stackhouse *et al.*, 2018). Agro-climatological variables, including total precipitation, minimum and maximum air temperature, relative humidity, and surface solar radiation, were extracted using the nasapower R package interface and compiled into annual district-level values for climate–yield analysis. Drought conditions were described using the Standardised Precipitation Index (SPI) and the Standardised Precipitation Evapotranspiration Index (SPEI). These indices capture precipitation anomalies and climatic water-balance stress, respectively (McKee *et al*., 1993; Vicente-Serrano *et al.*, 2010).

Data on rice production and yield were gathered from FAOSTAT (Food and Agriculture Organization Corporate Statistical Database), the Global Dataset of Historical Yields (GDHY), and national agricultural statistics from the Uganda Bureau of Statistics (UBOS) and the Ministry of Agriculture, Animal Industry and Fisheries (MAAIF). Merging these productivity datasets enhanced long-term yield consistency while reducing potential reporting bias from relying on a single source (Ray *et al*., 2015).

The final dataset included rice yield, climate variables, drought indices, and lagged predictors reflecting delayed climatic effects on crop productivity. Rice yield was standardised and expressed in metric tonnes per hectare (t ha^{-1}) to allow for comparability across districts and production systems.

Data processing and indicator construction. Continuous predictors were standardised before modelling to improve comparability among variables measured on different scales. This also reduced scale-dependent bias during model estimation. Few missing observations (<2%) were addressed using linear interpolation to maintain temporal continuity while minimising distortion of long-term climatic patterns. Similar preprocessing steps are commonly used in environmental time-series analysis, where maintaining temporal consistency is crucial for reliable findings (Bergmeir and Benítez, 2012; Hewamalage *et al*., 2021).

Lagged climatic predictors were included to account for the delayed impacts of environmental stress on rice productivity. Rainfall deficits, drought conditions, moisture anomalies, and temperature variability can influence crop performance across multiple growing seasons (Lobell and Burke, 2010; Lesk *et al*., 2016). Specifically, lagged SPI and temperature indicators were included to capture temporary carryover effects in climate–yield relationships under varying environmental conditions (Rowhani *et al*., 2011; Blanc, 2012). Adding lagged structures enhances the representation of temporal dependence and delayed climate–yield responses in climate-sensitive agricultural systems (Challinor *et al*., 2014).

These preprocessing steps improved consistency across indicators while reducing artefacts related to scale differences, structural changes, and temporal gaps that could affect statistical inferences under varying conditions (Hansen, 2001; Bai and Perron, 2003).

Predictive assessment framework. Predictive models were used to evaluate monitoring reliability under changing climate and environmental conditions. Four modelling setups representing different approaches to climate–yield monitoring were evaluated: Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX), Random Forest (RF), Support Vector Machine (SVM), and a hybrid statistical–machine-learning framework.

ARIMAX was included due to its established use in environmental and agricultural time-series analysis for representing temporal dependence and lagged climatic effects (Lobell and Burke, 2010). This model offers a clear statistical structure for analysing serial dependence in climate-sensitive productivity systems. However, traditional linear time-series models may become less

reliable under changing conditions when predictor-response relationships shift or complex environmental interactions arise (Milly *et al*., 2008). Model order selection was based on standard information criteria, including the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Random Forest and Support Vector Machine were included as flexible machine-learning methods capable of modelling nonlinear relationships and complex predictor interactions. Random Forest is particularly effective for handling nonlinear dependencies, variable interactions, and high-dimensional environmental datasets. Support Vector Machine provides flexibility in modelling complex response boundaries when data is limited (Chlingaryan *et al*., 2018; Rolnick *et al.*, 2022). Even though these methods often improve predictive performance in diverse environmental situations, their reliability can decline if predictor distributions shift outside the calibration range (Sugiyama and Kawanabe, 2012; Gama *et al.*, 2014).

To strengthen predictive reliability under changing environmental conditions, a hybrid statistical–machine-learning framework was developed by integrating the strengths of each modelling approach. For each observation at time t, ARIMAX (Equation 1), Random Forest (Equation 2), and Support Vector Machine (Equation 3) each generated separate predicted rice yields:

$$\hat{Y}_{ARIMAX,t} = f_{ARIMAX}(X_t) \quad \text{..... Equation 1}$$

$$\hat{Y}_{RF,t} = f_{RF}(X_t) \quad \text{..... Equation 2}$$

$$\hat{Y}_{SVM,t} = f_{SVM}(X_t) \quad \text{..... Equation 3}$$

Where:

X_t represents the climatic, hydro-climatic, and temporal predictor variables observed at time t. The final hybrid prediction (Equation 4) was obtained through a stacked ensemble formulation:

$$\hat{Y}_{Hybrid,t} = \beta_0 + \beta_1 \hat{Y}_{ARIMAX,t} + \beta_2 \hat{Y}_{RF,t} + \beta_3 \hat{Y}_{SVM,t} \quad \text{..... Equation 4}$$

Where:

$Y_{Hybrid,t}$ denotes the final hybrid prediction, β_0 is the intercept, and β_1, β_2 and β_3 are weighting coefficients estimated through the meta-learning regression model.

This ensemble design aimed to combine the clarity of structured temporal modelling with the adaptive flexibility of machine-learning approaches. ARIMAX captures temporal dependence and lagged effects, Random Forest handles nonlinear predictor interactions, and the Support Vector Machine is flexible for complex response structures. Integrating these complementary strengths was expected to improve predictive strength and monitoring reliability in changing conditions.

To assess the relative roles of climatic and temporal predictors, variable importance analysis was conducted using the Random Forest model instead of the hybrid framework. Random Forest was chosen because its importance method provides a clear way to assess predictor impact in nonlinear environmental modelling (Breiman, 2001). A model with 800 trees was fitted, and each predictor was randomly permuted to determine the resulting increase in prediction error. Higher importance scores indicated a stronger contribution to predictive performance within the Random Forest framework.

Monitoring reliability and validation strategy. Monitoring reliability was assessed across complementary dimensions to evaluate predictive performance, robustness, and adaptability under changing environmental conditions. These dimensions included out-of-sample monitoring diagnostics under observed climatic variability, statistical comparison of

predictive accuracy among competing models, assessment of residual behaviour and bias, comparative evaluation of monitoring reliability across modelling frameworks including spatial adaptability, robustness under scenario-based climatic stress conditions, and spatial vulnerability under heterogeneous perturbation scenarios.

Time-based data partitioning was used during model validation to preserve temporal order and minimise information leakage between training and test datasets. This approach aligns with recommended practice for climate and environmental time-series modelling (Bergmeir and Benítez, 2012; Hewamalage *et al*., 2021) and provides a more realistic assessment of predictive performance under evolving environmental conditions.

Statistical comparison of predictive performance. To determine whether observed differences in predictive performance among modelling approaches were statistically meaningful, pairwise comparisons of forecast errors were conducted using the Diebold–Mariano (DM) test (Diebold and Mariano, 1995). This test evaluates the null hypothesis that two competing models have equal predictive accuracy by comparing forecast losses.

Let $e_{1,t}$ and $e_{2,t}$ denote the prediction errors from two competing models at time t. The loss differential is defined as:

$$d_t = L(e_{1,t}) - L(e_{2,t}) \quad \text{..... Equation 5}$$

Where:

L(.) is a loss function. In this study, squared error loss was used, such that:

$$L(e_t) = e_t^2 \quad \text{..... Equation 6}$$

The null hypothesis of equal predictive accuracy is expressed as: $H_0: E(d_t) = 0$ against the alternative: $H_1: E(d_t) = \neq 0$

The Diebold–Mariano test statistic was computed as:

$$DM = \frac{\bar{d}}{\sqrt{v_{\bar{d}}}} \quad \text{..... Equation 7}$$

Where:

$$\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$$

is the sample mean loss differential across n out-of-sample observations, and v_d is the asymptotic variance of the mean loss differential. To account for temporal dependence and heteroskedasticity in the forecast error series, heteroskedasticity and autocorrelation consistent (HAC) estimators were used to estimate the variance:

$$v_d \approx \frac{1}{n}\left[\gamma_o + 2\sum_{k=1}^{h-1}\gamma_k\right] \quad \text{..... Equation 8}$$

Where :

Y_k is represents the k-th autocovariance of the loss differential series, and h denotes the forecast horizon (h=1 in this study).

Pairwise comparisons were conducted across all model combinations (ARIMAX vs RF, ARIMAX vs SVM, ARIMAX vs Hybrid, RF vs SVM, RF vs Hybrid, and SVM vs Hybrid). A statistically significant result indicates that one model has superior predictive accuracy relative to the other under the specified loss function. Statistical significance was assessed using asymptotic normal approximation at the 5% significance level.

Scenario-based climatic stress assessment. To examine monitoring robustness under changing environmental conditions, scenario-based climatic perturbations were introduced to simulate plausible departures from historical

climate patterns. These scenarios were used as sensitivity stress tests informed by Representative Concentration Pathway (RCP) 4.5 and RCP 8.5 climate-change narratives (van Vuuren *et al*., 2011; O'Neill *et al*., 2016), rather than as direct future yield prediction scenarios.

A uniform perturbation approach was applied across all study districts to enable consistent comparison of model behaviour under progressively intensified climatic stress. Under the moderate stress scenario (RCP 4.5 analogue), mean temperature was increased by 1.5°C, while annual rainfall was reduced by 2% relative to baseline observed conditions. Under the more severe stress scenario (RCP 8.5 analogue), temperature was increased by 2.5°C, accompanied by a 5% reduction in annual rainfall.

To maintain consistency between altered climatic conditions and associated hydro-climatic stress indicators, SPI and SPEI were proportionally adjusted to reflect increasing moisture stress under the perturbation scenarios. These adjustments were based on transformation rules linked to rainfall reduction magnitude, providing an approximate representation of intensified drought stress under shifted environmental conditions. The resulting perturbed predictor datasets were then supplied to the fitted monitoring models without re-estimation, allowing a direct assessment of predictive stability under environmental distribution shifts.

Monitoring robustness was evaluated by comparing prediction errors under perturbed conditions with baseline validation performance. Proportional increases in Root Mean Square Error (RMSE) were used to compare overall resilience across competing models, while district-level changes in RMSE were used to assess spatial vulnerability under heterogeneous climatic stress conditions.

Software implementation. All analyses were conducted in R version 4.3.2 (R Core Team, 2023), using the packages forecast, randomForest, and e1071 for time-series modelling, machine-learning analysis, and predictive evaluation. The analytical workflow included data preprocessing, model development, validation, statistical comparison of predictive performance, and climate perturbation analysis. Additional validation and diagnostic procedures were implemented through reproducible scripted workflows to ensure consistency across modelling stages.

Analytical scripts and associated procedures were archived to support reproducibility and transparency, consistent with recommended practices in computational environmental and statistical research (Peng, 2011; Stodden *et al*., 2016; Wilkinson *et al*., 2016).

RESULTS AND DISCUSSION

Monitoring reliability under observed environmental variability. Analysis of the 30-year district-level dataset (1995-2024) showed substantial interannual variation in rice productivity across Uganda's smallholder production systems, reflecting the strong sensitivity of yields to fluctuating hydro-climatic conditions. Observed and predicted rice yields varied across the competing modelling approaches (Fig. 1).

The hybrid statistical–machine-learning framework and Random Forest (RF) showed closer agreement with observed values, whereas the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model deviated more noticeably from the 1:1 reference line. Monitoring diagnostic assessment similarly revealed marked differences in predictive reliability under observed climatic variability (Table 1).

The hybrid framework exhibited the lowest systematic bias (Mean Error = 0.18), together with low residual dispersion and an approximately symmetric residual distribution, indicating strong monitoring reliability. Random Forest and Support Vector Machine (SVM) showed moderate bias (Mean Error = 0.53 and 0.61, respectively) and intermediate residual dispersion, suggesting moderate reliability

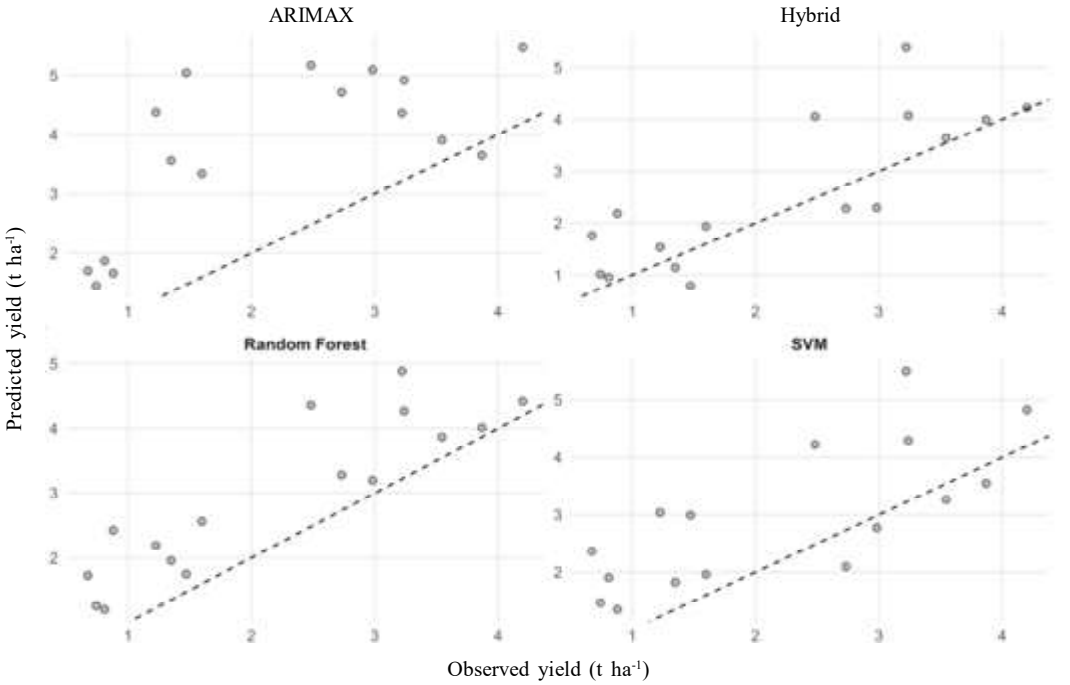

Figure 1. Observed versus predicted rice yields (t ha^{-1}) under observed climatic variability for ARIMAX, Random Forest (RF), Support Vector Machine (SVM), and the hybrid statistical–machine-learning framework. Closer alignment between observed and predicted values indicates improved monitoring reliability.

under observed climatic variability. In contrast, ARIMAX exhibited pronounced positive bias (Mean Error = 1.24), wider residual dispersion, and a clearly positively skewed residual distribution, pointing to substantial overestimation and reduced monitoring reliability.

These patterns suggest important structural limitations in conventional linear time-series monitoring approaches when applied to environmentally variable agricultural systems. The weaker performance and pronounced positive bias observed in ARIMAX imply that fixed-structure statistical models may struggle to adequately represent climate–yield relationships when agro-climatic conditions become unstable or increasingly nonlinear. Crop responses to climatic stress rarely follow simple linear pathways; rainfall variability, drought, and temperature anomalies often interact with agronomic management, soil conditions, hydrological constraints, and broader environmental heterogeneity in complex and threshold-dependent ways.

The more robust performance of the hybrid framework suggests that integrating structured temporal modelling with ensemble machine-learning approaches improves the representation of complex climate–yield relationships that may not be adequately captured by individual modelling frameworks alone. By preserving temporal dependence while accommodating nonlinear climatic interactions, the hybrid framework provides more stable predictive signals for operational decision-making. Systematic yield overestimation, as observed in the ARIMAX baseline, may have important practical consequences, including distorted climate-risk communication, weakened early warning credibility, and delayed food security responses.

TABLE 1. Monitoring diagnostic assessment under observed climatic variability

Model	Mean bias	Bias classification	Residual dispersion	Residual symmetry	Monitoring interpretation
ARIMAX	1.24	High	High	Positively skewed	Low monitoring reliability; systematic overestimation under climatic stress
Random Forest (RF)	0.53	Moderate	Moderate	Mild asymmetry	Moderate monitoring reliability under observed variability
Support Vector Machine (SVM)	0.61	Moderate	Moderate	Moderate asymmetry	Moderate monitoring reliability with stable error behaviour
Hybrid framework	0.18	Low	Low	Approximately symmetric	High monitoring reliability with reduced systematic bias

These findings are consistent with Lobell and Burke (2010), who reported that conventional statistical crop models may lose predictive reliability when environmental conditions move beyond historical calibration ranges. Challinor *et al*. (2014) similarly showed that crop responses to climatic stress often involve nonlinear thresholds and interacting sensitivities that rigid parametric structures may inadequately represent. Lesk *et al*. (2016) further demonstrated that climatic shocks can produce compound and asymmetric impacts on agricultural productivity, particularly in vulnerable smallholder systems.

More broadly, the hybrid framework's enhanced performance aligns with the environmental prediction literature, which shows that adaptive nonlinear approaches often outperform conventional statistical models when system responses are multidimensional, interacting, and environmentally heterogeneous (Chlingaryan *et al*., 2018; Reichstein *et al*., 2019). In the context of agricultural monitoring under climate non-stationarity, hybrid adaptive frameworks may therefore provide a stronger foundation for food security planning, climate-risk assessment, and operational early warning systems.

Statistical comparison of predictive performance among competing monitoring models. To determine whether the observed differences in monitoring performance reflected statistically meaningful differences in predictive accuracy, pairwise comparisons of out-of-sample forecast errors were conducted using the Diebold–Mariano (DM) test under a squared error loss function. The resulting comparative network revealed a clear statistical separation between conventional statistical modelling and adaptive monitoring frameworks (Fig. 2).

ARIMAX performed significantly worse than all competing models ($P < 0.05$), with substantial forecast error differentials relative to Random Forest, Support Vector Machine, and the hybrid framework. Among the adaptive

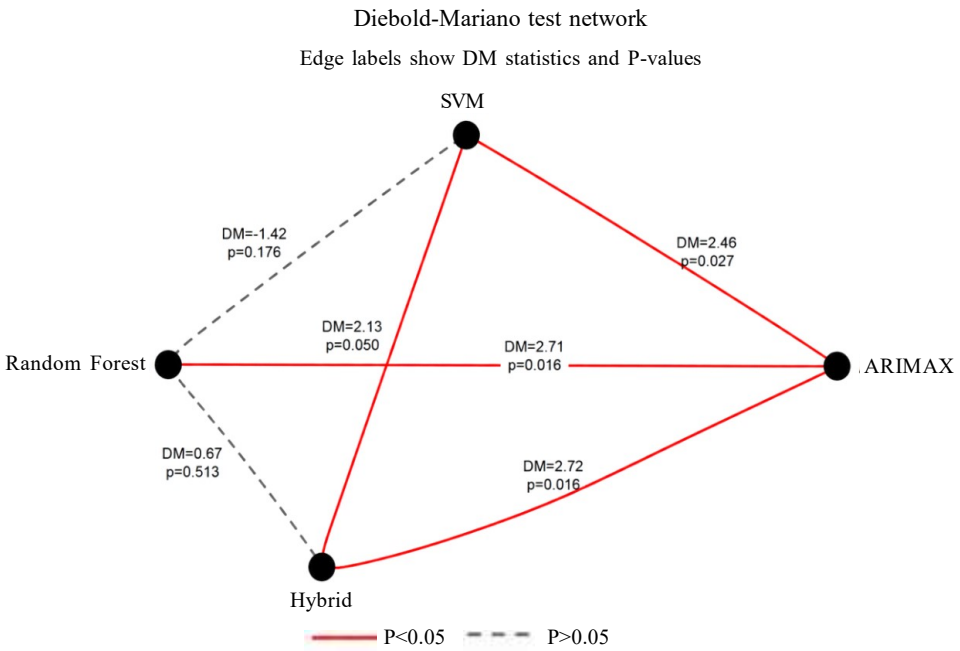

Figure 2. Pairwise statistical comparison of predictive accuracy among competing climate–yield monitoring models using the Diebold–Mariano test network. Edge labels show Diebold–Mariano (DM) statistics and associated p-values. Red solid lines indicate statistically significant differences in predictive accuracy ($P < 0.05$), while grey dashed lines indicate non-significant comparisons ($P \geq 0.05$).

models, no statistically significant differences were observed between Random Forest and the hybrid framework (DM = 0.6707; P = 0.5126) or between Random Forest and Support Vector Machine (P = 0.176), suggesting broadly comparable predictive performance under observed climatic variability. However, a statistically significant difference was observed between Support Vector Machine and the hybrid framework (DM = 2.1326; P = 0.0499), indicating a measurable predictive advantage for the hybrid configuration.

These results strengthen the interpretation that adaptive monitoring frameworks offer greater operational reliability than rigid statistical approaches in climate-sensitive agricultural systems. The consistent underperformance of ARIMAX suggests that fixed temporal structures are less able to represent climate–yield relationships characterised by nonlinearity, instability, and interacting environmental stressors. Adaptive approaches appear better positioned to accommodate climatic complexity and evolving system behaviour.

At the same time, statistical superiority under historically observed testing conditions should not be interpreted as definitive evidence of long-term operational reliability under changing environmental regimes. Although the Diebold-Mariano test provides a rigorous basis for comparing forecast accuracy (Diebold and Mariano, 1995), its conclusions remain conditional on the distributions of the environment represented in the evaluation sample. As predictor distributions shift or climate–response thresholds evolve, models that perform well under baseline conditions may undergo structural degradation.

Predictive adequacy under baseline conditions should therefore be viewed as necessary, but not sufficient, for reliable operational monitoring. Model selection should

also account for temporal dependence, spatial transferability, and resilience to climatic perturbation rather than relying exclusively on comparative predictive accuracy. This interpretation is consistent with broader predictive modelling literature showing that hybrid and ensemble approaches often outperform individual models by combining complementary predictive strengths while reducing dependence on a single structural representation (Dietterich, 2000). Machine-learning approaches have likewise demonstrated advantages in environmental prediction where nonlinear dependencies, multidimensional climatic interactions, and heterogeneous response structures dominate system behaviour (Chlingaryan *et al.*, 2018; Reichstein *et al.*, 2019).

As hydro-climatic conditions increasingly depart from historical norms across sub-Saharan Africa, monitoring systems based solely on conventional statistical architectures may face a heightened risk of performance failure. This challenge is well recognised through the concepts of dataset shift, covariate drift, and concept drift (Moreno-Torres *et al.*, 2012; Sugiyama and Kawanabe, 2012; Gama *et al.*, 2014), reinforcing the need for agricultural monitoring systems with adaptive modelling capacity under non-stationary environmental change.

Key climatic drivers influencing predictive performance in the Random Forest framework. Permutation-based variable importance analysis from the standalone Random Forest model identified a clear hierarchy of climatic, atmospheric, and temporal predictors associated with predictive performance (Table 2). Although these importance scores are specific to the Random Forest framework, they provide useful insight into the relative influence of key environmental drivers within the broader climate–yield monitoring context.

Temporal autoregressive memory emerged as the dominant predictive component, with lagged rice yield exhibiting the highest relative importance score (31.42). Hydro-climatic stress indicators followed as important environmental predictors, with the Standardised Precipitation Index (SPI) ranking second (22.87), mean temperature third (18.65), and the Standardised Precipitation Evapotranspiration Index (SPEI) fourth (15.43). Rainfall (11.28), relative humidity (7.94), and solar radiation (5.61) contributed comparatively less to predictive performance.

The prominence of lagged rice yield suggests that predictive capacity in smallholder rice systems is strongly shaped by temporal persistence. Current productivity appears to reflect not only contemporaneous climatic

TABLE 2. Relative importance of climatic and lagged predictors in the Random Forest monitoring framework

Rank	Predictor	Importance score	Monitoring interpretation
1	Lagged rice yield	31.42	Temporal persistence in rice productivity
2	SPI	22.87	Moisture anomaly signal
3	Mean temperature	18.65	Thermal stress sensitivity
4	SPEI	15.43	Climatic water-balance stress
5	Rainfall	11.28	Precipitation variability influence
6	Relative humidity	7.94	Atmospheric moisture effect
7	Solar radiation	5.61	Energy-related climatic influence

Note: Importance scores were derived from permutation-based Random Forest variable importance; higher values indicate stronger relative contribution to predictive monitoring performance

conditions, but also carryover effects linked to prior production performance, soil fertility dynamics, management continuity, technological inertia, and delayed biological responses to environmental stress. This highlights the importance of incorporating temporal memory into agricultural monitoring systems rather than treating crop outcomes as purely immediate climatic responses.

The importance of SPI and SPEI further suggests that hydro-climatic stress indicators provide more informative signals for monitoring reliability than raw precipitation measures alone. In smallholder systems with constrained adaptive capacity, these indices may better capture agricultural drought conditions and climatic water-balance stress. The substantial contribution of mean temperature similarly indicates that thermal stress plays an important role in climate–yield dynamics, likely through nonlinear physiological responses, shortened crop development periods, and increased evapotranspiration demand that intensifies moisture stress. The comparatively lower importance of solar radiation and relative humidity suggests that, although atmospheric energy balance remains relevant, its direct contribution is secondary to drought stress, thermal thresholds, and temporal persistence.

These findings align with broader evidence showing that climate extremes affect agricultural productivity through interacting rather than isolated pathways. Challinor *et al.* (2014) and Lesk *et al*. (2016) similarly demonstrated that climatic shocks often operate through compound mechanisms, while Rowhani *et al*. (2011) documented substantial yield impacts in East Africa arising from concurrent rainfall and temperature anomalies. The importance of drought and thermal indicators is especially relevant in Uganda's heterogeneous agro-ecological landscape, where water stress and climatic variability strongly shape smallholder production outcomes (Mubiru *et al*., 2012; Babyenda *et al*., 2023; Ndhokero *et al*., 2026). More broadly, the context-dependent nature of these climate–yield relationships supports the findings of Blanc (2012) and Ciais *et al*. (2005), reinforcing the value of flexible hybrid monitoring systems capable of capturing localised threshold effects and heterogeneous climatic sensitivities.

Bias and residual behaviour. Residual diagnostic analysis revealed clear differences in error structure among the competing models that were not fully captured by aggregate performance metrics (Fig. 3).

The conventional ARIMAX framework exhibited a broad residual density distribution with pronounced positive skewness, indicating substantial systematic bias and a persistent tendency to overestimate rice yields under adverse climatic conditions. Random Forest and Support Vector Machine (SVM) substantially reduced residual dispersion relative to the linear baseline, although SVM retained some distributional asymmetry. The hybrid framework produced the narrowest residual distribution, centred close to zero, indicating minimal systematic bias and stronger tracking consistency under observed climatic variability.

These residual patterns extend the earlier model comparisons by showing that the advantage of the hybrid framework is not limited to average predictive accuracy, but also reflects meaningful improvements in error structure. The pronounced positive skewness observed in ARIMAX is particularly problematic for agricultural risk management, as persistent upward bias during drought or thermal stress periods may lead to underestimation of production risk. In contrast, the narrower and approximately symmetric residual distribution produced by the hybrid framework suggests that integrating structured temporal modelling with ensemble machine-learning approaches improves directional balance and overall monitoring reliability. An error distribution centred near zero strengthens the credibility of environmental signals used in operational risk detection and decision support.

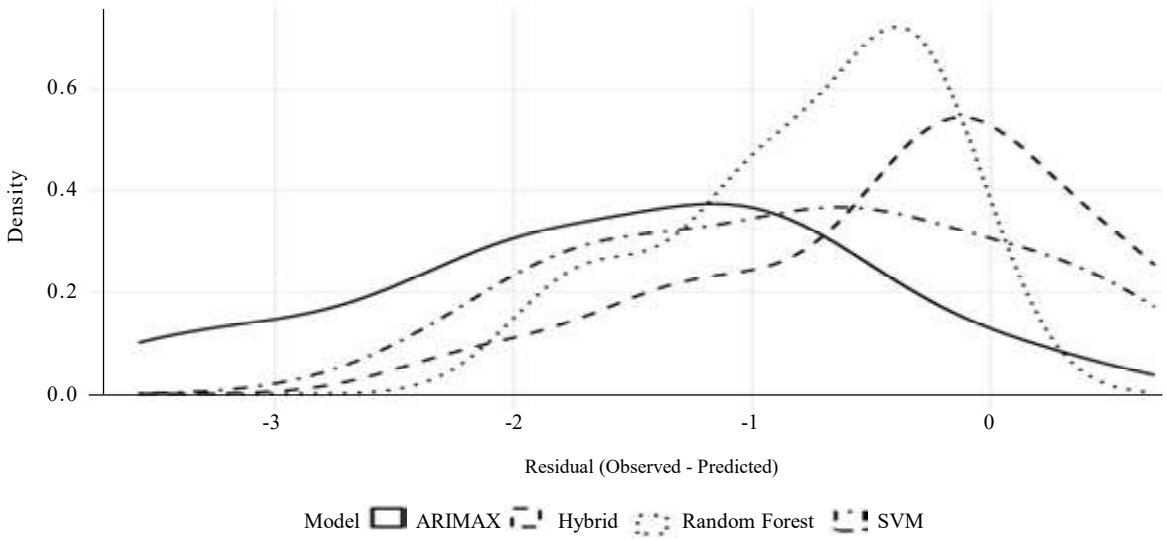

Figure 3. Residual density distributions (observed minus predicted rice yield) for ARIMAX, Random Forest (RF), Support Vector Machine (SVM), and the hybrid statistical–machine-learning framework. Narrower distributions centred around zero indicate reduced prediction bias and improved monitoring reliability.

The findings also underscore the importance of evaluating residual diagnostics alongside conventional accuracy metrics when assessing monitoring systems, since acceptable average performance may still conceal operationally important directional bias. This interpretation is consistent with the environmental monitoring principles advanced by Nichols and Williams (2006) and Lindenmayer and Likens (2010), who emphasised that reliable monitoring systems require not only predictive accuracy but also unbiased and interpretable error behaviour to support effective decision-making.

The structural weaknesses observed in ARIMAX under environmentally variable conditions are also consistent with Lobell and Burke (2010), who cautioned that conventional statistical crop models may degrade when environmental conditions extend beyond historical calibration domains. Conversely, the capacity of adaptive algorithms to correct more complex error structures aligns with broader environmental machine-learning literature (Chlingaryan *et al*., 2018; Reichstein *et al*., 2019). By preserving temporal dependence while reducing systematic overestimation, the hybrid framework appears better suited to operational food security monitoring and climate adaptation planning under climate non-stationarity.

Integrated monitoring reliability across modelling frameworks. Integrated comparative assessment revealed clear differences in operational monitoring reliability across the competing modelling frameworks, demonstrating that model adequacy cannot be judged solely on aggregate predictive accuracy (Table 3).

The integrated assessment indicates that the hybrid framework provides the strongest overall balance of monitoring reliability across the evaluated dimensions. It combines high predictive accuracy, reduced systematic bias, and strong spatial transferability, while maintaining robust performance under climatic perturbation. Although Random Forest demonstrated slightly greater resilience under perturbation stress, its performance across the broader evaluation dimensions remained less consistent than that of the hybrid framework. Support Vector Machine (SVM) exhibited moderate overall reliability, whereas the

TABLE 3. Integrated assessment of monitoring reliability across competing modelling frameworks

Model	Accuracy classification	Bias classification	Spatial transferability	robustness Perturbation	Overall monitoring reliability
ARIMAX	Low	High	Low	Low	Low
RF	Moderate	Moderate	Moderate	High	Moderate
SVM	Moderate	Moderate	Moderate	Low	Moderate
Hybrid framework	High	Low	High	High	High

Note: Classifications are based on comparative interpretation of empirical performance across predictive diagnostics, residual behaviour, district-level transferability, and climatic perturbation stress testing. Higher classifications indicate comparatively stronger monitoring reliability

conventional ARIMAX baseline consistently performed least favourably, highlighting its operational limitations under non-stationary environmental conditions.

District-level analysis further reinforced the spatial transferability of the hybrid framework across Uganda's heterogeneous rice-growing environments. The hybrid framework maintained comparatively stable prediction performance across the study districts despite marked differences in agro-ecological conditions, hydrological characteristics, and exposure to climatic stress.

These findings suggest that monitoring reliability is inherently multidimensional and cannot be adequately judged by predictive fit alone. Reliable monitoring systems must also demonstrate acceptable bias behaviour, spatial transferability, and resilience under environmentally shifted conditions. A model that performs adequately under historically observed testing conditions may still prove operationally unreliable if it exhibits localised instability, systematic directional bias, or sensitivity to changing predictor distributions. This concern is especially relevant in temporally dependent agricultural systems, where conventional validation approaches may overstate model adequacy by overlooking chronological structure, autocorrelation, and evolving environmental relationships.

The comparatively balanced performance of the hybrid framework suggests that combining complementary modelling paradigms offers practical advantages for climate-sensitive agricultural monitoring. Structured autoregressive components preserve interpretable temporal memory and represent delayed crop-system responses, while machine-learning components provide the flexibility needed to capture nonlinear climatic thresholds and heterogeneous environmental dynamics that may be poorly represented by conventional statistical models alone.

This interpretation is consistent with methodological evidence showing that conventional validation metrics may obscure model degradation under temporal and environmental shifts (Bergmeir and Benítez, 2012; Hewamalage *et al.*, 2021). More broadly, the improved performance of the hybrid framework aligns with foundational work on ensemble modelling (Dietterich, 2000) and environmental predictive modelling under non-stationary conditions (Reichstein *et al.*, 2019), both of which support the integration of complementary modelling approaches to improve predictive robustness. Given the environmental heterogeneity of sub-Saharan African rice-growing systems, adaptive hybrid monitoring frameworks may therefore provide

a more stable and operationally transferable basis for climate-risk monitoring, food security planning, and agricultural decision support.

Monitoring robustness under climate perturbation stress scenarios. Model performance deteriorated across all competing frameworks under scenario-based climatic stress conditions, indicating reduced monitoring reliability as environmental conditions shifted beyond historically observed ranges (Fig. 4). However, the magnitude of deterioration differed substantially among modelling approaches, revealing important contrasts in perturbation resilience and operational robustness.

The conventional ARIMAX baseline showed marked progressive error escalation as climatic stress intensified, with Root Mean Square Error (RMSE) increasing by 61.7% between the moderate (RCP 4.5) and severe (RCP 8.5) perturbation scenarios. By contrast, the hybrid framework and Random Forest demonstrated stronger resilience, with error increases of 33.5 and 28.7%, respectively. Although Support Vector Machine (SVM) exhibited relatively low absolute error under the moderate perturbation scenario, it showed the greatest deterioration under the more severe climate shift, with a 68.8% increase in RMSE (Table 4).

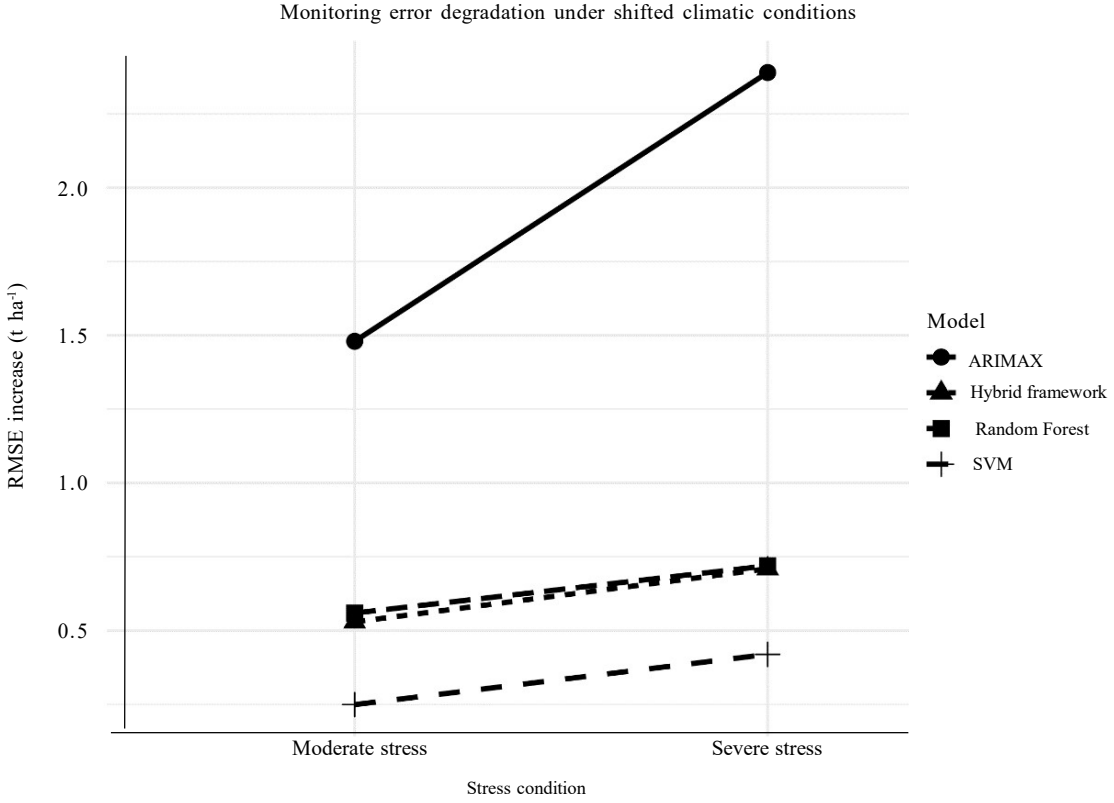

Figure 4. Comparative deterioration in monitoring performance under shifted climatic stress conditions. Increases in Root Mean Square Error (RMSE) are shown relative to baseline validation performance under moderate and severe scenario-based climatic perturbations analogous to RCP 4.5 and RCP 8.5 stress conditions. Lower RMSE increases indicate greater resilience and improved monitoring robustness under changing environmental conditions.

TABLE 4. Predictive robustness under RCP-analogue climatic perturbation stress scenarios

Model	RMSE (RCP 4.5 analogue)	RMSE (RCP 8.5 analogue)	Error increase (RCP 4.5→ 8.5) (%)	Robustness interpretation
ARIMAX	1.48	2.39	61.7	Low perturbation resilience; substantial structural degradation under intensified climatic stress
RF	0.55	0.71	28.7	High perturbation resilience; relatively stable predictive performance under climate stress
SVM	0.26	0.43	68.8	Low resilience under severe perturbation despite strong moderate-scenario performance
Hybrid framework	0.53	0.71	33.5	High perturbation resilience; stable performance under shifted climatic conditions

Note: RMSE = Root Mean Square Error; RCP = Representative Concentration Pathway. RMSE values under RCP 4.5 and RCP 8.5 were derived from perturbed scenario datasets and are not directly comparable with baseline validation RMSE values. Lower percentage increases indicate greater robustness to climatic perturbation stress

These results indicate that strong predictive performance under historically observed climatic conditions does not necessarily translate into operational reliability under altered environmental regimes. Models developed under fixed structural assumptions may lose predictive relevance when environmental distributions shift beyond their calibration domain, particularly when interacting climatic stressors trigger nonlinear crop responses not represented in historical observations.

The pronounced degradation observed in ARIMAX and SVM under severe perturbation conditions highlights the operational risks associated with relying on rigid modelling architectures for long-term climate-sensitive monitoring. Their mathematical structures appear less adaptable when required to extrapolate under shifted climatic conditions. Although Random Forest showed marginally greater perturbation resilience, the hybrid framework maintained a stronger overall balance between baseline predictive performance, bias control, spatial transferability, and robustness under climatic stress.

These findings reinforce the importance of stress-testing agricultural monitoring systems under plausible climatic perturbations rather than relying exclusively on historical validation baselines. This interpretation aligns with Milly *et al*. (2008), who argued that stationarity can no longer be assumed as a reliable foundation for environmental management under climate change, a concern further reinforced by recent climate assessments (IPCC, 2023). The observed degradation under shifted climatic conditions is also consistent with predictive modelling literature on dataset shift, covariate drift, and concept drift, which shows that model performance may deteriorate when predictor–response relationships evolve over time (Moreno-Torres *et al*., 2012; Sugiyama and Kawanabe, 2012; Gama *et al*., 2014). By combining strong baseline performance with comparatively robust resilience under climatic perturbation, the hybrid framework appears to provide a more dependable basis for climate-

risk monitoring, food security planning, and adaptation decision support under environmental non-stationarity.

Spatial vulnerability under perturbed climatic conditions. Spatial variation in predictive degradation under scenario-based climatic perturbations revealed clear differences in the transferability of the competing monitoring frameworks. Although all models experienced some decline in performance under shifted climatic conditions, both the magnitude and spatial distribution of error increases varied substantially across study districts (Table 5).

The conventional ARIMAX baseline exhibited pronounced spatial instability, particularly in Lira, where predictive error increased sharply under both moderate and severe perturbation scenarios (ΔRMSE = 2.73 and 3.99, respectively). Support Vector Machine (SVM) remained relatively stable in Bugiri and Butaleja but deteriorated substantially in Lira under severe perturbation (ΔRMSE = 0.75). Random Forest demonstrated strong resilience in Lira and Nwoya but showed notable vulnerability in Bugiri, where predictive error increased to 1.34 under RCP 8.5. By contrast, the hybrid framework maintained comparatively moderate error shifts across most districts, indicating stronger spatial transferability under climatic perturbation. However, district-level differences in RMSE shifts should be interpreted cautiously because scenario responses may reflect both local predictor structure and model sensitivity to perturbed inputs.

These findings suggest that monitoring vulnerability under climatic stress is spatially heterogeneous rather than geographically uniform. The pronounced localised breakdown of ARIMAX indicates that fixed-structure statistical models are particularly vulnerable when predictor–response relationships vary across agro-ecological zones. Although non-parametric models generally improved spatial adaptability, the uneven performance of Random Forest in Bugiri suggests that standalone machine-learning models may still be sensitive to localised environmental anomalies or unfamiliar climatic conditions.

The hybrid framework's more consistent performance suggests that combining temporal structure with complementary ensemble machine-learning approaches improves resilience to geographically heterogeneous environmental shifts. Anchoring machine-learning flexibility within an autoregressive framework appears to reduce the risk of localised structural failure while maintaining more stable monitoring performance across diverse production environments.

These findings reinforce the importance of evaluating regional transferability when assessing climate-sensitive monitoring systems. This interpretation aligns with Meyer and Pebesma (2021), who showed that predictive reliability declines when models

TABLE 5. District-level shifts in monitoring error (ΔRMSE) under climate perturbation stress scenarios

District	ARIMAX (RCP 4.5 / RCP 8.5)	Hybrid framework (RCP 4.5 / RCP 8.5)	Random Forest (RCP 4.5 / RCP 8.5)	SVM (RCP 4.5 / RCP 8.5)
Bugiri	0.36 / 0.51	0.86 / 1.01	1.11 / 1.34	0.03 / 0.20
Butaleja	0.97 / 1.62	0.17 / 0.09	0.01 / 0.45	0.04 / 0.17
Lira	2.73 / 3.99	0.59 / 0.96	0.02 / 0.05	0.50 / 0.75
Nwoya	0.52 / 0.87	0.17 / 0.27	0.10 / 0.17	0.11 / 0.35

Note: Values represent increases in Root Mean Square Error (ΔRMSE) under scenario-based climatic perturbations relative to baseline model performance; higher values indicate greater monitoring vulnerability under shifted environmental conditions

extrapolate beyond the environmental bounds of their calibration domain. The improved spatial stability of the hybrid framework is also consistent with broader environmental modelling literature suggesting that integrated modelling approaches can enhance transferability across heterogeneous ecosystems (Reichstein *et al*., 2019).

Given the marked geographic variability in rainfall regimes, hydrological conditions, and soil constraints across sub-Saharan African smallholder systems (Blanc, 2012; Ray *et al*., 2015), spatial robustness is a critical requirement for operational agricultural monitoring. This is particularly relevant in Uganda's rice-growing systems, where exposure to drought and water-stress constraints varies considerably across districts (Mubiru *et al*., 2012; Babyenda *et al*., 2023; Ndhokero *et al*., 2026). Adaptive hybrid monitoring frameworks may therefore provide a more equitable and operationally dependable basis for regional climate-risk communication and food security planning.

Implications for rice production and climate-risk management. The comparative assessment across predictive diagnostics, statistical comparisons, residual behaviour, spatial transferability, and climate perturbation stress testing demonstrates that modelling architecture directly shapes the operational reliability of agricultural monitoring systems. The hybrid framework consistently showed stronger multidimensional performance, characterised by low systematic bias (Mean Error = 0.18), strong spatial transferability, and comparatively stable error escalation under severe climatic perturbation (33.5%). In contrast, the conventional ARIMAX baseline exhibited substantial structural degradation under both variable and shifted climatic conditions. These findings suggest that model selection is not merely a computational choice, but a determinant of the reliability of early warning systems, climate-risk communication, and downstream agricultural decision-support applications.

Conventional monitoring systems based exclusively on fixed parametric assumptions may therefore provide unstable operational guidance under climate non-stationarity. Because smallholder rice productivity remains highly sensitive to interacting hydro-climatic stressors, systematic overestimation under adverse conditions may obscure emerging production risks, weaken early warning credibility, and delay timely intervention. The stronger multidimensional performance of the hybrid framework suggests that agricultural monitoring is more reliable when structured temporal modelling is combined with complementary machine-learning approaches capable of capturing complex and evolving climate–yield dynamics.

The observed spatial heterogeneity further suggests that geographically uniform national monitoring systems may inadequately capture local production realities. Rice-growing districts differ substantially in agroecological conditions, hydrological constraints, production system characteristics, and exposure to climatic stress. Evidence from Ugandan rice systems indicates that farmer management responses vary according to localised water-stress conditions, reinforcing the need for monitoring systems capable of accommodating regional heterogeneity rather than relying on uniform national assumptions (Ndhokero *et al.*, 2026).

The perturbation analyses also emphasise the importance of evaluating monitoring systems under environmental conditions extending beyond historically observed climatic ranges. Strong predictive performance under baseline conditions does not guarantee operational reliability under altered climatic regimes. Periodic stress-testing and recalibration are therefore likely to become essential components of robust monitoring governance as climatic baselines continue to evolve.

These implications align with the broader climate services and adaptation literature. Reliable agricultural monitoring is essential for translating climatic information into actionable decision support, particularly in smallholder systems where adaptive capacity is constrained, and decision windows are narrow (Vaughan and Dessai, 2014). Climate-smart agriculture similarly depends on credible environmental intelligence to inform timely adaptation, resilience planning, and productivity management under climatic uncertainty (Lipper *et al.*, 2014).

From a policy perspective, reliable monitoring is increasingly important for climate-risk financing, including index-based agricultural insurance and related risk-transfer mechanisms (Greatrex *et al*., 2015; Carter *et al.*, 2017). Because such systems depend on accurate environmental triggers, model bias or predictive instability may result in inappropriate risk pricing, poorly calibrated payout thresholds, and weakened institutional trust. Adaptive hybrid monitoring frameworks may therefore provide a more durable foundation for extension services, targeted adaptation investment, climate-risk communication, and evidence-based food security planning under accelerating environmental uncertainty.

CONCLUSION

Monitoring reliability in climate-sensitive smallholder rice production systems varies considerably across modelling approaches under environmentally dynamic and non-stationary conditions. The present findings indicate that the adaptive hybrid statistical–machine-learning framework offers a more reliable foundation for agricultural monitoring by combining structured temporal modelling with complementary ensemble machine-learning approaches. This combination improves the capacity to track evolving climate–yield relationships across heterogeneous production environments.

Conventional statistical monitoring approaches based on fixed linear temporal assumptions exhibit lower reliability when climatic conditions exceed the range represented during model calibration. While standalone machine-learning models provide greater flexibility in capturing complex relationships, their performance remains uneven across complementary dimensions of reliability, particularly under geographically heterogeneous and perturbed climatic conditions. Integrating temporal structure with complementary ensemble machine-learning approaches, therefore, offers a more stable and operationally robust framework for monitoring climate-sensitive agricultural systems under uncertainty.

The results further show that predictive accuracy alone does not provide a sufficient basis for judging operational monitoring adequacy. Reliable agricultural monitoring requires broader evaluation criteria that consider directional bias, spatial transferability, and resilience under environmental stress. Models that perform well under historically observed conditions may not necessarily retain comparable reliability as climatic regimes continue to evolve.

For Uganda's smallholder rice systems, adaptive monitoring frameworks have direct practical relevance for climate-risk management, early warning systems, climate services, index-based agricultural insurance, and evidence-informed adaptation planning. More broadly, agricultural monitoring systems should be viewed as adaptive decision-support infrastructure rather than static predictive instruments. Periodic stress testing, recalibration, and contextual validation will be essential to sustain monitoring reliability amid ongoing climate non-stationarity.

ACKNOWLEDGEMENT

This study forms part of the PhD research of the first author in Applied Statistics at Kampala

International University, Uganda. The authors acknowledge Kampala International University and the National Agricultural Research Organisation (NARO) for institutional support that facilitated this study.

REFERENCES

Babyenda, P., Kabubo-Mariara, J. and Odhiambo, S. 2023. Climate variability and agricultural productivity in Uganda. *African Journal of Agricultural and Resource Economics* 18:14-38. https://doi.org/10.53936/afjare.2023.18(1).2

Bai, J. and Perron, P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18:1-22. https://doi.org/10.1002/jae.659

Bergmeir, C. and Benítez, J.M. 2012. On the use of cross-validation for time series predictor evaluation. *Information Sciences* 191:192-213. https://doi.org/10.1016/j.ins.2011.12.028

Bontempi, G., Ben Taieb, S. and Le Borgne, Y.A. 2013. Machine learning strategies for time series forecasting. In: Aufaure, M.A. and Zimányi, E. (Eds.). *Business intelligence*. Lecture Notes in Business Information Processing 138:62-77. https://doi.org/10.1007/978-3-642-36318-4_3

Breiman, L. 2001. Random forests. Machine Learning 45(1):5-32. https://doi.org/10.1023/A:1010933404324

Carter, M., de Janvry, A., Sadoulet, E. and Sarris, A. 2017. Index insurance for developing country agriculture: A reassessment. *Annual Review of Resource Economics* 9:421-438. https://doi.org/10.1146/annurev-resource-100516-053352

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R. and Chhetri, N. 2014. A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change* 4:287-291. https://doi.org/10.1038/nclimate2153

Chlingaryan, A., Sukkarieh, S. and Whelan, B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture* 151:61-69. https://doi.org/10.1016/j.compag.2018.05.012

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M. and Valentini, R. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437:529-533. https://doi.org/10.1038/nature03972

Diebold, F.X. and Mariano, R.S. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13(3):253-263. https://doi.org/10.1080/07350015.1995.10524599

Dietterich, T.G. 2000. Ensemble methods in machine learning. In: Kittler, J. and Roli, F. (Eds.). *Multiple classifier systems*. Springer, pp. 1-15. https://doi.org/10.1007/3-540-45014-9_1

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. 2015. The climate hazards infrared precipitation with stations: A new environmental record for monitoring extremes. *Scientific Data* 2:150066. https://doi.org/10.1038/sdata.2015.66

Gama, J., Þliobaitë, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. 2014. A survey on concept drift adaptation. *ACM Computing Surveys* 46:1-37. https://doi.org/10.1145/2523813

Greatrex, H., Hansen, J.W., Garvin, S., Diro, R., Blakeley, S., Le Guen, M., Rao, K.N. and Osgood, D.E. 2015. Scaling up index insurance for smallholder farmers: Recent evidence and insights. CCAFS Report No. 14. CGIAR. https://cgspace.cgiar.org/handle/10568/53101

Guttman, N.B. 1999. Accepting the standardised precipitation index: A calculation algorithm. *Journal of the American Water Resources Association* 35:311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Voss, F., Wisser, D., and Kabat, P. 2014. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences USA* 111(9):3251-3256. https://doi.org/10.1073/pnas.1222475110

Hansen, B.E. 2001. The new econometrics of structural change: Dating breaks in US labour productivity. *Journal of Economic Perspectives* 15:117-128. https://doi.org/10.1257/jep.15.4.117

Hewamalage, H., Bergmeir, C. and Bandara, K. 2021. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting* 37:388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

IPCC. 2023. Climate change 2023: Synthesis report. https://doi.org/10.59327/IPCC/AR6-9789291691647

Koutsoyiannis, D. 2013. Hydrological variability under climate change. *Hydrological Sciences Journal* 58:1170-1197. https://doi.org/10.1080/02626667.2013.804626

Lesk, C., Rowhani, P. and Ramankutty, N. 2016. Influence of extreme weather disasters on global crop production. *Nature* 529:84-87. https://doi.org/10.1038/nature16467

Lindenmayer, D.B. and Likens, G.E. 2010. The science and application of ecological monitoring. *Biological Conservation* 143:1317-1328. https://doi.org/10.1016/j.biocon.2010.02.013

Lipper, L., Thornton, P., Campbell, B.M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., Hottle, R., Jackson, L., Jarvis, A., Kossam, F., Mann, W., McCarthy, N., Meybeck, A., Neufeldt, H., Remington, T., Sen, P.T. and Sessa, R. 2014. Climate-smart agriculture for food security. *Nature Climate Change* 4:1068-1072. https://doi.org/10.1038/nclimate2437

Lobell, D.B. and Burke, M.B. 2010. On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology* 150:1443-1452. https://doi.org/10.1016/j.agrformet.2010.07.008

McKee, T.B., Doesken, N.J. and Kleist, J. 1993. The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*. pp. 179–184.

Meyer, H. and Pebesma, E. 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution* 12:1620-1633. https://doi.org/10.1111/2041-210X.13650

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. 2008. Stationarity is dead: Whither water management? *Science* 319:573-574. https://doi.org/10.1126/science.1151915

Moreno-Torres, J.G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N.V. and Herrera, F. 2012. A unifying view on dataset shift in classification. *Pattern Recognition* 45: 521-530. https://doi.org/10.1016/j.patcog.2011.06.019

Mubiru, D.N., Komutunga, E., Agona, A., Apok, A. and Ngara, T. 2012. Characterising agrometeorological climate risks and uncertainties: Crop production in Uganda. *South African Journal of Science* 108(3/4):1-11. https://doi.org/10.4102/sajs.v108i3/4.470

Ndhokero, J., Nkpordee, L. and Goparaju, A. 2026. Farmer management responses to contrasting water stress in upland and lowland rice systems of Uganda. *African*

Crop Science Journal 34(2):383-396. https://doi.org/10.4314/acsj.v34i2.11

Nichols, J.D. and Williams, B.K. 2006. Monitoring for conservation. *Trends in Ecology and Evolution* 21:668-673. https://doi.org/10.1016/j.tree.2006.08.007

O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K. and Sanderson, B.M. 2016. The Scenario Model Intercomparison Project (Scenario MIP) for CMIP6. *Geoscientific Model Development* 9: 3461-3482. https://doi.org/10.5194/gmd-9-3461-2016

Peng, R.D. 2011. Reproducible research in computational science. *Science* 334:1226–1227. https://doi.org/10.1126/science.1213847

Quionero-Candela, J., Sugiyama, M., Schwaighofer, A. and Lawrence, N.D. 2009. Dataset shift in machine learning. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262170055.001.0001

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Ray, D.K., Gerber, J.S., MacDonald, G.K. and West, P.C. 2015. Climate variation explains a third of global crop yield variability. *Nature Communications* 6:5989. https://doi.org/10.1038/ncomms6989

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566:195-204. https://doi.org/10.1038/s41586-019-0912-1

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Slavin Ross, A., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J. and Bengio, Y. 2022. Tackling climate change with machine learning. *ACM Computing Surveys* 55(2):1-96. https://doi.org/10.1145/3485128

Rowhani, P., Lobell, D.B., Linderman, M. and Ramankutty, N. 2011. Climate variability and crop production in Tanzania. *Agricultural and Forest Meteorology* 151:449-460. https://doi.org/10.1016/j.agrformet.2010.12.002

Stackhouse, P.W., Gupta, S.K., Cox, S.J., Zhang, T., Mikovitz, J.C., and Hinkelman, L.M. 2018. NASA POWER: Prediction of Worldwide Energy Resources data archive. NASA Langley Research Center Atmospheric Science Data Center. https://power.larc.nasa.gov/

Stodden, V., McNutt, M., Bailey, D.H., Deelman, E., Gil, Y., Hanson, B., Heroux, M.A., Ioannidis, J.P.A. and Taufer, M. 2016. Enhancing reproducibility for computational methods. *Science* 354:1240-1241. https://doi.org/10.1126/science.aah6168

Subbaswamy, A. and Saria, S. 2020. From development to deployment: Dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 21:345-352. https://doi.org/10.1093/biostatistics/kxz041

Sugiyama, M. and Kawanabe, M. 2012. Machine learning in non-stationary environments: Introduction to covariate shift adaptation. The MIT Press. https://doi.org/10.7551/mitpress/9780262017091.001.0001

Uganda Bureau of Statistics (UBOS). 2021. Statistical Abstract 2021. Kampala, Uganda: Uganda Bureau of Statistics.

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J. and Rose, S.K. 2011. The representative concentration pathways: An overview. *Climatic Change* 109:5-31. https://doi.org/10.1007/s10584-011-0148-z

Vaughan, C. and Dessai, S. 2014. Climate services for society. *Wiley Interdisciplinary Reviews: Climate Change* 5:587-603. https://doi.org/10.1002/wcc.290

Vicente-Serrano, S.M., Beguería, S. and López-Moreno, J.I. 2010. A multiscalar drought index sensitive to global warming. *Journal of Climate* 23:1696-1718. https://doi.org/10.1175/2009JCLI2909.1

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E. and Brock, B. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3:160018. https://doi.org/10.1038/sdata.2016.18

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C. and Asseng, S. 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences USA* 114(35):9326-9331. https://doi.org/10.1073/pnas.1701762114