Annals of Actuarial Science (2026), 1–25
doi:10.1017/S1748499526100372

ORIGINAL RESEARCH PAPER

Quantifying and pricing crop yield risk under climate change: a hierarchical model forecasting perspective

Ning Zhang[1], Jin Yang[1], Zhuoqun Xie[2] and Daning Bi[1]

[1]College of Finance and Statistics, Hunan University, China; and [2]School of Finance, Shanghai University of Finance and Economics, China
Corresponding author: Daning Bi; Email: daningbi@hnu.edu.cn

(Received 16 September 2025 UTC; revised 02 April 2026 UTC; accepted 01 June 2026 UTC)

Abstract

Accurate and internally coherent crop-yield forecasts are important for agricultural risk management, crop-insurance ratemaking, and regional risk assessment under climate variability. However, crop yields are influenced by high-dimensional and strongly correlated weather conditions, while forecasts produced at different spatial levels often violate aggregation constraints. Existing studies focus on yield prediction within individual regions and pay limited attention to weather-informed forecasting, hierarchical coherence, and insurance-oriented risk measurement. This paper develops an integrated framework for hierarchical crop-yield forecasting and risk assessment by combining dimensionality reduction for high-dimensional weather variables, probabilistic forecasting, and forecast reconciliation. Using county- and state-level spring and winter wheat yields in Montana from 1982 to 2022, we compare alternative base forecasting models and reconciliation methods under scenarios with and without weather information. Forecast performance is evaluated using point and probabilistic scoring rules, and the reconciled predictive distributions are used to construct scenario-based measures of downside yield risk. The results show that incorporating weather information and hierarchical reconciliation improves the quality and coherence of hierarchical yield forecasts. The resulting probabilistic forecasts provide a basis for loss-rate estimation, cross-county risk comparison, and spatial risk mapping and also support crop-insurance ratemaking under a retain–cede game between private insurers and the government.

Keywords: agricultural production forecasting; climate risk measures; forecast consistency; ratemaking; risk zoning

1. Introduction

Climate change poses growing challenges to agricultural production (Luchtenbelt *et al.*, 2024). Droughts, floods, storms, and other weather-related hazards can directly disrupt crop growth and reduce yields, while increasing climatic variability further amplifies production uncertainty. For staple crops such as maize, wheat, and rice, yields are expected to fluctuate more substantially under changing climatic conditions (Rezaei *et al.*, 2023), and extreme events may even result in complete crop failure. These developments have made climate risk management increasingly important for safeguarding agricultural production and supporting policy responses at broader regional scales. This need is particularly evident in agricultural systems characterized by strong spatial aggregation. Crop production is unevenly distributed across counties, and output in major producing counties contributes disproportionately to overall state-level production. Consequently, state-level climate risk management cannot rely solely on aggregate information, nor can county-level decisions be made in isolation. Effective allocation of disaster-prevention resources and climate adaptation measures, therefore, requires forecasts that are not only accurate

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Institute and Faculty of Actuaries. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

at the county level but also coherent with state-level production outcomes. This naturally calls for hierarchical forecasting approaches that ensure cross-level consistency. Existing studies have mainly examined the effects of weather conditions on crop yields within individual regions and have shown that wheat is particularly sensitive to weather conditions at different stages of the growing cycle (Tan *et al.*, 2017; Zhao *et al.*, 2017; Sun *et al.*, 2021; Meng *et al.*, 2022). Some studies have developed yield forecasting models using a limited number of key weather variables, such as temperature and precipitation during the growing season (Liu *et al.*, 2019; He *et al.*, 2021; Bangelesa *et al.*, 2023; Zhang *et al.*, 2023), while others have focused on variable selection methods (Bai & Ng, 2002; Li *et al.*, 2021). Despite these efforts, it remains difficult to forecast current-year crop yields accurately under changing climatic conditions, quantify yield risk, and manage weather-related risk when climatic disasters occur.

Quantitative measurement of crop yield risk presents unique challenges. First, the limited availability of yield data constrains empirical modeling and robust statistical inference of yield distributions, as yield and loss observations are typically recorded only once per year and may occasionally suffer from missing values (Ramsey, 2020). Second, the presence of spatial and temporal correlations in the data-generating process complicates crop yield forecasting. Failure to account for these dependencies may lead to biased or misleading estimates of yield risk (Ramsey, 2020; Liu & Ker, 2021). Meanwhile, weather-related risk plays a critical role in both crop yield forecasting and risk management. Research has shown that incorporating weather covariates, rather than relying solely on historical yield data, can substantially improve the accuracy and reliability of crop yield models (Lobell *et al.*, 2011; Kleiber *et al.*, 2011). However, integrating weather variables into crop-yield forecasting and risk assessment models presents several significant challenges. On the one hand, weather data are rich and high-dimensional, which makes dimensionality reduction challenging, particularly given the typically short time series of available yield data (Li *et al.*, 2021). On the other hand, systemic weather risks, which exhibit spatial correlations across large geographic regions, often result in highly variable year-to-year losses, thereby complicating the accurate quantification of crop yield risk (Zhu *et al.*, 2018). Moreover, coherence in yield statistics matters not only for official reporting but also for insurance design and risk management. The U.S. National Agricultural Statistics Service (NASS) places considerable weight on consistency in reported yields across counties, states, and the nation. Through benchmarking procedures and hierarchical aggregation constraints, published estimates are made internally consistent across reporting levels and comparable across commodities under common definitions (Summary and Disclosure Methodology Branch, 2023). Coherence is also important at the contract level. Area-based crop insurance programs such as the Group Risk Plan (GRP) base indemnities on a county yield index rather than on an individual producer's realized yield.[1] As a result, contract outcomes depend on an external and publicly reported measure that cannot be manipulated by individual policyholders. For insurers, premium rating and risk management, therefore, require yield data that are consistent across levels of aggregation so that county-level pricing remains aligned with portfolio accounting at the state and national levels (Ker *et al.*, 2016). In this setting, coherence is not only a statistical property of yield data. It is also essential to use area yield indices in pricing, reserving, and reinsurance. Only when yield measures are coherent within the hierarchy can insurers connect county-level pricing decisions to portfolio loss distributions, economic capital assessment, and reinsurance pricing in a unified framework (Meyers, 2001; Meyers *et al.*, 2003; Chong *et al.*, 2023; Guan *et al.*, 2023).

In hierarchical yield forecasting, it is necessary to construct separate models for each level of the hierarchy. However, discrepancies among these models may lead to inconsistencies across forecast values across hierarchical levels, thereby violating the principle of hierarchical coherence. Forecast reconciliation is a statistical technique for adjusting hierarchical time-series forecasts to ensure consistency with the underlying hierarchical structure (Wickramasuriya *et al.*, 2019).

[1] https://www.rma.usda.gov/group-risk-plan

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

When addressing the forecasting problem of hierarchical time series, it remains standard practice to model and predict the output at each level independently. Although there is a theoretical aggregation relationship among forecasts at different levels, using distinct optimal models across levels often yields incoherent results that fail to preserve this structure. Forecast reconciliation methods offer a systematic solution to this issue by aligning forecasts across levels in a statistically optimal manner. One of the earliest reconciliation approaches introduced in the literature is the Bottom-Up (BU) method, which aggregates forecasts from the most disaggregated (bottom) level to obtain higher-level forecasts (Orcutt *et al*., 1968; Dunn *et al*., 1976). Building upon the Generalized Least Squares (GLS) reconciliation framework proposed by Hyndman *et al*. (2011), Wickramasuriya *et al*. (2019) introduced the Minimum Trace (MinT) reconciliation method, which improves the estimation of the inverse covariance matrix to achieve more accurate forecast reconciliation. The MinT approach obtains coherent forecasts for hierarchical time series by minimizing the total variance of forecast errors after reconciliation. To date, this method has been successfully applied in various domains, including electricity demand forecasting (Taieb *et al*., 2021) and population mortality projections (Li *et al*., 2019). However, it has yet to be explored in the context of agricultural production forecasting.

However, a key limitation of point forecasts is their inability to convey uncertainty about future outcomes. As a result, they provide no insight into the potential variability or risk associated with forecast errors, thereby limiting their usefulness in decision-making. To address this issue, probabilistic forecasts, which represent future quantities of interest as full probability distributions rather than single-point estimates, have been increasingly adopted across a wide range of fields. A widely used framework for generating density forecasts independently across hierarchical time series has been introduced in previous work (Hyndman *et al*., 2011). This approach enables large-scale probabilistic forecasting and ensures that the forecast distribution of any aggregated series equals the convolution of the distributions of its components, thus preserving hierarchical coherence. Moreover, point forecasts are inherently limited to providing a single estimate for each time step. In contrast, when the incoherent base forecast distributions are assumed to be jointly Gaussian, it is possible to fully leverage the relationship between point forecasts and forecast errors to construct prediction intervals and evaluate forecast performance using proper scoring rules or loss functions. This highlights the added value of probabilistic forecasts in capturing forecast uncertainty and enhancing decision support. Applications include meteorology (Leutbecher, 2019), energy (Taieb *et al*., 2016), and retail forecasting (Berry *et al*., 2020).

The objective of this study is to develop an integrated measurement framework to improve crop yield forecasting and risk assessment, thereby supporting more effective agricultural risk management and facilitating multi-regional risk zoning. To achieve this, we propose integrating high-dimensional weather variables, after applying dimensionality reduction techniques, into a probabilistic forecasting framework. This approach aims to improve both the accuracy and hierarchical coherence of hierarchical crop yield predictions. Specifically, this paper has the following marginal contributions. First, this study develops an ensemble yield forecasting model that identifies key weather-related risk factors and evaluates forecasting accuracy at multiple levels. The findings provide empirical support and methodological guidance for the design of crop insurance products and the formulation of data-informed weather risk management strategies in agriculture. Second, the proposed risk measurement framework contributes a generalizable approach to crop yield risk management. Utilizing the median of probabilistic forecasts offers an effective means of capturing and optimizing tail risks, thereby enhancing the robustness of agricultural risk assessment across different crop types. Moreover, when weather data are scarce or unavailable, forecast reconciliation techniques can partially compensate for the resulting loss in accuracy, thereby enabling reliable yield predictions while reducing data acquisition costs. Finally, the framework was applied to real crop yield insurance ratemaking by conducting a retain–cede rating game between a private insurer and the government (Ker & McGowan, 2000; Harri *et al*., 2011; Zhang, 2017; Ramsey, 2020; Liu & Ker, 2021; Zhu, 2024).

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

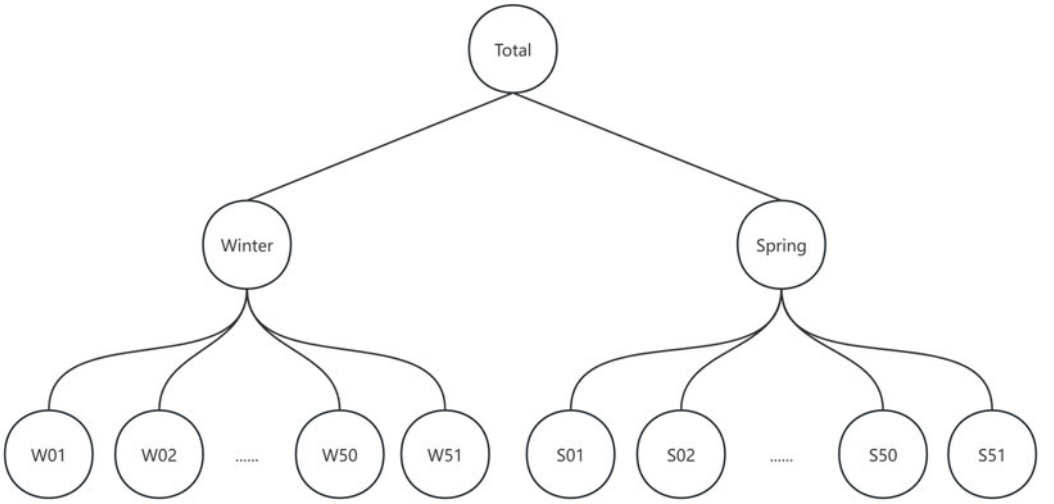


Figure 1. Hierarchical structure of crop yield.

The remainder of this paper is organized as follows. Section 2 describes the wheat and weather data used in this paper and constructs a hierarchical time series of two types of wheat in Montana. Section 3 starts from the point forecast and probability forecast models and introduces in detail the forecasting framework. In Section 4, an empirical study of two types of wheat in Montana is conducted based on the forecast models above. In Section 5, the proposed forecasting framework is applied to a crop insurance rating game. Section 6 concludes the article.

2. Data

Our empirical analysis is based on county-level wheat yield data and county-level monthly climate data for Montana, United States. These two datasets differ substantially in both temporal frequency and spatial structure. Wheat yield is measured annually, consistent with the harvest cycle, whereas climate variables are observed monthly. At the same time, the yield data exhibit a hierarchical structure that can be aggregated from the county level to crop-specific and state-level series, while the climate data are defined at the county administrative level. The key data-construction challenge is therefore to transform high-frequency county-level climate observations into county-year measures that align with annual wheat yield outcomes while preserving spatial consistency. In addition, because the resulting climate covariates are high-dimensional, we further apply principal component analysis (PCA) in the subsequent modeling stage to reduce dimensionality.

2.1 U.S. wheat data

Wheat yield data used in this study were obtained from the Quick Stats database maintained by the National Agricultural Statistics Service of the United States Department of Agriculture (USDA NASS).[2] The empirical dataset covers 51 Montana counties, and both spring wheat and winter wheat are cultivated across the state. Accordingly, the yield data have a natural hierarchical structure with three levels: state, wheat category (spring versus winter), and county. Figure 1 illustrates this hierarchy, from total state yield to disaggregated county-level series. In the empirical analysis, annual spring wheat yield is denoted by $\boldsymbol{y}_s$, annual winter wheat yield by $\boldsymbol{y}_w$, and aggregate annual wheat yield by $\boldsymbol{y}_T$. This structure allows us to analyze wheat dynamics at both disaggregated

[2]https://quickstats.nass.usda.gov/

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

and aggregated levels.[3] Because the original yield dataset contains missing observations, we use interpolation methods to impute the missing values.

2.2 U.S. weather data

The county-level monthly climate data for Montana are drawn from NASA's Prediction of Worldwide Energy Resources (POWER) project.[4] The dataset provides a distinct monthly climate time series for each county over 1982–2022. Climate variables are therefore defined at the county administrative level, and their realized values are county-specific, ensuring spatial alignment with county-level yield observations. These monthly climate observations are aggregated into county-year, crop-specific growing-season measures before being matched to annual yield data by county and year. Because spring wheat and winter wheat differ in their biological growing cycles, climate variables are constructed separately for each crop using crop-specific growth windows. For winter wheat, the relevant growth period spans October of year $t-1$ through June of year t (9 months), whereas for spring wheat it spans March through August of year t (6 months). Table 1 lists all climate variables included in the analysis. Accordingly, we construct 174 climate variables for spring wheat and 290 for winter wheat, each based on monthly observations within the corresponding growth period.

3. Model and methodology

This section outlines the overall framework for forecasting agricultural yield and quantifying yield risk under climate change. Section 3.1 reviews common modeling approaches that leverage mixed-frequency data for yield prediction. Section 3.2 discusses methods for achieving coherence in forecasts across spatial hierarchies. Section 3.3 introduces key concepts used to generate probabilistic forecast samples for quantifying yield risk under climate change. Section 3.4 describes the modeling pipeline and implementation steps.

The classical literature primarily employs yield decomposition models to empirically assess agricultural productivity. These models separate historical yields into two components: weather-driven yields and trend yields. Building on this foundation, we reconstruct the analytical framework (Coulombe *et al.*, 2021) as follows:

$$\min_{g\in\mathcal{G}}\left\{\sum_{t=1}^{T}\left(y_t-\hat{y}_t\right)^2\right\}=\min_{g\in\mathcal{G}}\left\{\sum_{t=1}^{T}\left(y_t-T(y_t)-g\left(\boldsymbol{X}_t\right)\right)^2\right\},$$

where y_t denotes the observed historical yield, and $\hat{y}_t$ represents the fitted yields. The yield is decomposed into two components, the trend yield $T(y_t)$, and the weather-induced yield $g(\boldsymbol{X}_t)$, where the function g, selected from a functional space $\mathcal{G}$, maps the transformed climate covariates $\boldsymbol{X}_t$ to yield outcomes. Based on this framework, we present three modeling approaches that differ in the degree of decomposition applied to yield dynamics.

3.1 Base forecasting methods

3.1.1 ARIMA

External weather conditions are critical determinants of wheat production. Incorporating weather variables into the ARIMA framework yields a dynamic regression model (ARIMAX), which improves forecasting accuracy by leveraging exogenous climatic information.

Assuming there are u weather factors, represented by y_t as the wheat yield at time point t, ϕ_i as the coefficient of the auto-regressive term, B as the regression operator, θ_j as the coefficient of the

[3] Additional visualizations of state-level, crop-level, and county-level yield data are reported in the Online Appendix A.
[4] https://power.larc.nasa.gov/

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 1. Climate-related variables

Variable	Description
sshk	Sunshine duration for month *k* (hours)
sddk	Number of sand-dust storm days in month *k* (days)
sndk	Number of snow-covered days in month *k* (days)
dt00k	Number of days in month *k* with daily minimum temperature ≤ 0 °C (days)
dx32k	Number of days in month *k* with daily maximum temperature > 0 °C (32°F) (days)
dx70k	Number of days in month *k* with daily maximum temperature > 21.1°C (70°F) (days)
dx90k	Number of days in month *k* with daily maximum temperature > 32.2°C (90°F) (days)
dptk	Daily mean dew point temperature averaged over all days in month k (°C)
snddepthk	Mean snow depth for month *k* (cm)
snwk	Total snowfall for month *k* (mm)
emxpk	Maximum daily precipitation in month *k* (mm)
emxtk	Maximum daily maximum temperature in month *k* (°C)
prcpk	Total precipitation (rain + melted snow) for month *k* (mm)
snowk	Total snow amount in month *k* (mm)
tavgk	Daily average temperature over all days in month *k* (°C)
tmaxk	Daily maximum temperature averaged over all days in month *k* (°C)
tmink	Daily minimum temperature averaged over all days in month *k* (°C)
tsk	Skin temperature aggregated from month 1 to *k* (°C)
ws2 mk	2-meter wind speed aggregated from month 1 to *k* (m/s)
t2mdewk	2-meter dew point temperature aggregated from month 1 to *k* (°C)
gwektopk	Top-layer soil wetness aggregated from month 1 to *k*
ws2mmaxk	Max 2 m wind speed aggregated from month 1 to *k* (m/s)
ws2mmink	Min 2 m wind speed aggregated from month 1 to *k* (m/s)
k	Calendar month, that is, $k = 1, 2, \ldots, 12$ for Jan–Dec

moving average term, $\mathbf{X}_{k,t}$ as the k-th weather factors at time point t, and β_k as the regression coefficient corresponding to the weather factors $\mathbf{X}_{k,t}$, the ARIMAX model of crop yield with weather factors is represented as follows:

$$y_t = \frac{(1 + \sum_{j=1}^{q} \theta_j B^j)\varepsilon_t + \sum_{k=1}^{u} \beta_k X_{k,t}}{(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d},$$

where p and q are the order of the model, d is the differential order, u is the number of weather variables, and ε_t is the error term.

3.1.2 Mixed frequency data sampling (MIDAS) regression

Mixed frequency data sampling (MIDAS) regression is designed to accommodate time series data observed at different temporal resolutions. Unlike traditional models that require data aggregation or latent factor extraction, MIDAS allows high-frequency covariates to directly inform the prediction of a low-frequency target variable. In this study, we employ a fixed single-equation MIDAS regression with an autoregressive component, where the dependent variable is measured annually, and the explanatory variables are sampled five times per year. The model is specified as follows:

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i y_t + \gamma \sum_{k=1}^{q} \Phi\left(k; \theta\right) L_{HF}^{q} X_{ct-k} + \varepsilon_t,$$

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

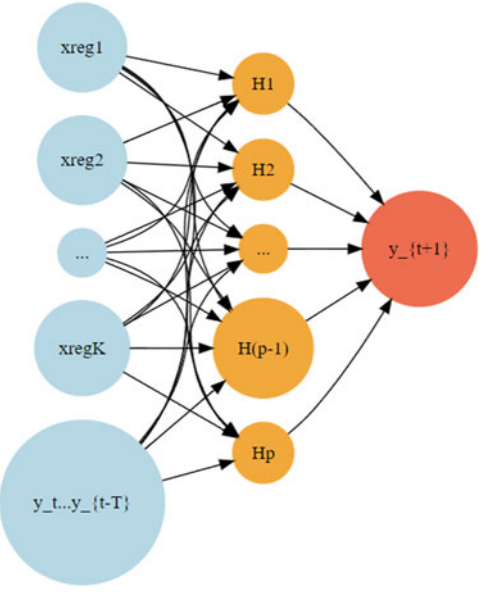


Figure 2. Feed-forward neural networks for forecasting crop yield.

where L_{HF} denotes the high-frequency lag operator, q is the high-frequency alignment indicator, and $\Phi(k;\theta)$ isthe weighting operator applied to the mixed-frequency weather factors X_{ct-k}. The subscript c in $ct-k$ denotes the high-frequency position relative to the low-frequency period t. Moreover, α denotes the intercept term of the model, while γ represents the overall coefficient (or scaling parameter) associated with the weighted high-frequency explanatory component. Accordingly, the model employed in this article is a MIDAS-AR model.

3.1.3 Neural network models

Extensive research has been conducted on a wide range of machine learning algorithms, which offer flexible modeling tools aimed at minimizing expected prediction loss. Among these, feed-forward neural networks (FFNN) are particularly effective in capturing complex nonlinear relationships between climate variables and agricultural yields, without requiring a priori specification of functional forms. Compared to traditional linear models and tree-based approaches, FFNN offer greater flexibility in learning intricate patterns from high-dimensional data. This advantage becomes especially critical when dealing with heterogeneous and uncertain climate conditions. Given that the dataset used in this study incorporates exogenous weather variables, we adopt the FFNN framework to examine the potential of machine learning in enhancing the accuracy of real-time crop yield forecasting.

Neural networks, based on the structural principles of the human brain, are computational frameworks that reproduce the architecture of biological neural systems. Among them, the FFNN is one of the most commonly used types of artificial neural networks. Figure 2 provides an example of FFNN. A defining characteristic of FFNN is their acyclic structure, meaning connections between nodes do not form feedback loops. The network processes crop yield historical observations through successive hidden layers, where each neuron applies a weighted transformation followed by a nonlinear activation function. In addition to yield history, the model incorporates exogenous weather factors as supplementary inputs. These weather factors (xreg1-xregK) serve as forecasts that capture the influence of climatic conditions on crop yield. The outputs from the hidden layers are combined to generate the final forecasts of crop yield in the output layer.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

3.2 Hierarchical forecast reconciliation

Accurate forecasting of wheat yields in a hierarchical system requires forecasts that are not only individually reasonable at each level but also mutually consistent with the underlying aggregation structure. Forecast reconciliation provides a statistical framework for adjusting a set of base forecasts so that coherence is achieved across spatial or categorical hierarchies (Zellner & Tobias, 2000). In this paper, a hierarchical time series is defined as an n-dimensional multivariate time series subject to known linear aggregation constraints, where lower-level county series aggregate to higher-level series such as crop-specific totals and overall yield.

Let $\boldsymbol{y}_t \in \mathbb{R}^m$ be a vector comprising observations from all time series in the hierarchy at time t, and $\boldsymbol{b}_t \in \mathbb{R}^n$ be a vector comprising observations of only the most disaggregated (bottom-level) series. The full hierarchy can be written as

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t.$$

When the summing matrix $\boldsymbol{S} \in \mathbb{R}^{m\times n}$ linearly aggregates base-level observations to higher hierarchy levels. Define $\boldsymbol{A} \in \mathbb{R}^{(m-n)\times n}$ as an $(m-n)\times n$ aggregation matrix with $n = n_a + n_b$, and I_n is an n-dimensional identity matrix, then

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I_n} \end{bmatrix}_{m\times n}.$$

For example, according to the hierarchical structure of wheat in Montana, $n_a = n_b = 51$, $n = 102$, $m = 105$, $\boldsymbol{y}_t = [\boldsymbol{y}_T, \boldsymbol{y}_S, \boldsymbol{y}_W, \boldsymbol{y}_s^1, \ldots, \boldsymbol{y}_s^{51}, \boldsymbol{y}_w^1, \ldots, \boldsymbol{y}_w^{51}]'$ is a series matrix of wheat yields at all levels, $\boldsymbol{b}_t = [\boldsymbol{y}_s^1, \ldots, \boldsymbol{y}_s^{51}, \boldsymbol{y}_w^1, \ldots, \boldsymbol{y}_w^{51}]'$ is a series matrix of two wheat yield observations in 51 counties. The aggregation matrix A can be written as

$$\boldsymbol{A} = \begin{bmatrix} \overbrace{1 \cdots 1}^{n=102}\ \overbrace{1 \cdots 1}^{} \\ \overbrace{1 \cdots 1}^{n_a=51}\ \overbrace{0 \cdots 0}^{n_b=51} \\ \underbrace{0 \cdots 0}_{n_a=51}\ \underbrace{1 \cdots 1}_{n_b=51} \end{bmatrix}_{3\times 102}.$$

The h-step-ahead base forecasts are generated independently across the hierarchical structure, thus necessitating a reconciliation process to adjust them through cross-level information integration, which enhances alignment with observed values while maintaining aggregation coherence.

Let $\widehat{\boldsymbol{y}}_{t+h|t}$ represent the h-step-ahead base forecasts for each series in the structure, and $\widetilde{\boldsymbol{y}}_{t+h|t}$ represent the h-step-ahead forecast of each time series after reconciliation. All linear reconciliation methods can be expressed as

$$\widetilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}\widehat{\boldsymbol{y}}_{t+h|t}, \tag{1}$$

where $\boldsymbol{S}$ is the summing matrix and $\boldsymbol{G}$ is the weight matrix, $\boldsymbol{SG}$ can be treated as the projection matrix. The transformation, in the most general case, is a mapping $\psi: \mathbb{R}^n \to \mathfrak{s}$, as $\widetilde{\boldsymbol{y}}_{t+h} = \psi\left(\widehat{\boldsymbol{y}}_{t+h}\right)$. The mapping ψ can also be represented as a combination of transformations $s \circ g$. The $s \circ g$ in this combination corresponds to the multiplied projection matrix $\boldsymbol{SG}$ by the left in equation (1). There are several common reconciliation methods currently available, which are summarized in Table 2.

3.3 Probabilistic forecasts

3.3.1 Forecasting process

Crop yields exhibit both spatial and temporal dependencies, largely driven by the evolving nature of systemic weather risk (Liu & Ker, 2021). Adjacent counties often display similar yield patterns

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 2. Summary of reconciliation choices for G

Reconciliation method	G
BU	$\left(\mathbf{0}_{n\times(m-n)}, \boldsymbol{I}_{n\times n}\right)$
OLS (Hyndman *et al.*, 2011)	$\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'$
MinT(Sam) (Wickramasuriya *et al.*, 2019)	$\left(\boldsymbol{S}'\widehat{\boldsymbol{\Sigma}}_{\text{Sam}}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\widehat{\boldsymbol{\Sigma}}_{\text{Sam}}^{-1}$
MinT(Shr) (Wickramasuriya *et al.*, 2019)	$\left(\boldsymbol{S}'\widehat{\boldsymbol{\Sigma}}_{\text{Shr}}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\widehat{\boldsymbol{\Sigma}}_{\text{Shr}}^{-1}$

Note: $\boldsymbol{\Sigma}_{\text{Sam}}$ and $\boldsymbol{\Sigma}_{\text{Shr}}$ are the sample and shrinkage (Wickramasuriya *et al.*, 2019) covariance matrix, respectively, of one-step-ahead in-sample base forecast errors.

due to shared climatic and geographic characteristics (Annan *et al.*, 2014; Du *et al.*, 2018). To accurately quantify yield risk, it is essential to account for both the time-series and cross-sectional dependencies embedded within the hierarchical structure of yield data. Probabilistic forecasting is particularly valuable in this condition, as it explicitly characterizes uncertainty, improves the estimation of risk metrics, and supports more robust risk management decisions by evaluating a range of possible outcomes rather than relying on point forecasting (Petropoulos *et al.*, 2022). Accordingly, we aim to generate coherent probabilistic forecast samples of wheat yields at the county, variety, and state levels, ensuring consistency across aggregation layers.

To illustrate the construction of probabilistic forecast samples, we follow Panagiotelis *et al.* (2023) to define the following terms:

- Define the h-step-ahead forecasting vector $\hat{\boldsymbol{y}}_{\boldsymbol{T+h|T}} = (\hat{y}^1_{T+h|T}, \ldots, \hat{y}^N_{T+h|T})$ and assume $\hat{\boldsymbol{y}}_{\boldsymbol{T+h|T}} \sim \mathcal{N}\left(\widehat{\boldsymbol{y}}_{T+h|T}, \boldsymbol{W}_h\right)$, where $\boldsymbol{W}_h = \mathrm{E}\left[\boldsymbol{y}_{T+h} - \widehat{\boldsymbol{y}}_{T+h|T}\right]\left[\boldsymbol{y}_{T+h} - \widehat{\boldsymbol{y}}_{T+h|T}\right]^{\top}$;
- Let $\hat{y}^{n|m}_{T+h|T}$ denote the mth sample point of nth series of unreconciled forecasts generated from the predictive probabilistic distribution $\mathcal{N}\left(\widehat{\boldsymbol{y}}_{T+h|T}, \boldsymbol{W}_h\right)$;
- A sample of size M generated from $\mathcal{N}\left(\widehat{\boldsymbol{y}}_{T+h|T}, \boldsymbol{W}_h\right)$ is denoted by $\hat{Y}_{T+h}$, where $\hat{Y}_{T+h} = (\hat{y}^{|1}_{T+h}, \hat{y}^{|2}_{T+h}, \ldots, \hat{y}^{|M}_{T+h})$.

When forecasting wheat yield, we restrict our attention to the case of $h = 1$, although we can generalize to larger h using the recursive method (Hyndman & Athanasopoulos, 2021). In order to forecast and quantify wheat yield risks under weather changes accurately, we calculate and generate R-step-ahead forecasts (base forecast) for simulated samples based on the following algorithm.

This algorithm can make full use of information within the weather data to ensure that the R-step-ahead forecasts are the most accurate under the same model framework. We use this algorithm after obtaining the optimal forecasting model for crop yield.

3.3.2 Evaluation of probabilistic forecast accuracy

This section introduces energy score (ES) and variogram score (VS) as scoring rules to evaluate the accuracy of probabilistic forecasts. In addition, we examine how the inclusion of weather information influences the performance of probabilistic reconciliation forecasts.

Energy score (ES). The ES, a proper scoring rule for evaluating multivariate probabilistic forecasts, extends the Continuous Ranked Probability Score (CRPS) to the multivariate setting and is defined as follows:

$$\mathrm{ES}\,(P, \boldsymbol{z}) = \mathrm{E}_P\,||\boldsymbol{X} - \boldsymbol{z}\,||_2 - \frac{1}{2}\mathrm{E}_P\,||\boldsymbol{X} - \boldsymbol{X}^*\,||_2,$$

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Algorithm 1 Generating R–step–ahead forecasts $\hat{\boldsymbol{y}}_{T+h|T}$ by one–step–ahead forecasts

procedure X-MODEL($\boldsymbol{\theta}, X$)
  Input: model parameters $\boldsymbol{\theta}$; weather data X
  Output: one–step–ahead forecasts
  for $t = T : T + R$ **do**
    for $i = 1 : N$ **do**
      $X^{'}_{k,t} \leftarrow$ reducing dimensionality X
      for $k = 1$ to K **do**
        Fit $f^i(\boldsymbol{\theta})$ model to $X^{'}_{k,t}$
        Calculate and store the AIC of $f^i(\boldsymbol{\theta})$
      end for
      $\hat{f}^i_{|t} \leftarrow \text{argmin}_{f^i(\boldsymbol{\theta})}(\text{AIC})$
    end for
    $\hat{y}^i_{t+1|t} \leftarrow \hat{f}^i_{|t}(\boldsymbol{\theta}, X^{'}_{k,t+1})$
  end for
  return R–step–ahead forecasts $(\hat{y}^1_{T+1|T}, \ldots, \hat{y}^N_{T+1|T}),(\hat{y}^1_{T+2|T}, \ldots, \hat{y}^N_{T+2|T}),\ldots,(\hat{y}^1_{T+R|T}, \ldots, \hat{y}^N_{T+R|T})$
end procedure

where $\boldsymbol{z} \in \mathbb{R}^{m\times 1}$ is the observation vector, $\boldsymbol{X}, \boldsymbol{X}^*$ are independent copies drawn from the distribution P, and E $||\boldsymbol{X}||_2$ is finite. Monte Carlo methods can be used when the analytical expressions for these expectations are not readily available:

$$\text{ES}\,(\hat{P}, z) = \frac{1}{N}\sum_{i=1}^{N} ||x_i - z||_2 - \frac{1}{2(N-1)}\sum_{i=1}^{N-1} ||x_i - x_{i+1}||_2,$$

where $x_1, x_2 \ldots, x_N \in \mathbb{R}^{m\times 1}$ is a collection of N random draws taken from the predictive distribution $\hat{P}$. This formulation is computationally more efficient than the multivariate extension of the empirical counterpart of the cumulative ranked probability score. However, it turns out that the discriminatory ability of the energy score to misspecified correlations could be limited.

Variogram score (VS). To mitigate some of the shortcomings of the ES, Scheuerer & Hamill (2015) introduced an alternative proper scoring rule that is based on pairwise differences between the components of an m-dimensional vector. Under the assumption that the absolute moment of order p is finite, the variogram score of order p is defined as follows:

$$\text{VS}(P, \boldsymbol{z}) = \sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\left(\left|z_i - z_j\right|^p - \text{E}_P\left|X_i - X_j\right|^p\right)^2,$$

where z_i denotes the ith component of $\boldsymbol{z}$, x_i denotes the ith component of a random vector x following distribution P, and w_{ij} is a non-negative weight assigned to the pair (i, j). Based on the simulation findings reported in Scheuerer & Hamill (2015) and Wickramasuriya (2024), the parameters are set to $p = 0.5$ and $w_{ij} = 1$ for all i, j.

3.4 Modeling pipeline and implementation steps

This subsection summarizes how the base forecasting models, forecast reconciliation methods, and probabilistic forecasting procedure are integrated into a unified empirical framework. The overall analysis combines county-level yield forecasting, hierarchical reconciliation, and probabilistic evaluation under two information settings, namely with weather information and without weather information.

The dataset combines monthly weather observations with annual wheat yield records. Owing to crop-specific differences in growing-season length, the weather feature set is high-dimensional, containing 290 weather variables for winter wheat and 174 for spring wheat, each aligned with the corresponding phenological cycle. To reduce dimensionality and alleviate multicollinearity in the

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

high-dimensional weather predictors, PCA is applied to the original weather variables, and the first five principal components, which cumulatively explain 90% of the total variance, are retained as the weather inputs for forecasting.

The forecasting process is organized hierarchically, with county-level spring and winter wheat yields at the bottom level, state-level spring and winter wheat yields at the intermediate level, and total wheat yield at the top level. This structure implies aggregation constraints across series. Accordingly, our forecasting framework proceeds in three main stages.

Step 1. We implement rolling-window forecasts with a fixed one-year forecasting horizon under two information settings: with weather information and without weather information. For example, the forecast for 2018 is generated using observations from 1982 to 2017. At each iteration, forecast accuracy is evaluated by comparing the root mean squared error (RMSE) across the three candidate base forecasting models, and the best-performing model is selected accordingly. Since these base forecasts are estimated separately for each series, they are not required to satisfy the aggregation constraints of the hierarchy at this stage.

Step 2. Under both information settings, we construct point forecasts for all series in the hierarchy based on the selected base forecasting model and then apply alternative forecast reconciliation methods to enforce aggregation coherence across different levels. The accuracy of the reconciled point forecasts is evaluated using RMSE. This step allows us to compare the point-forecast performance of different reconciliation methods under each information setting.

Step 3. Under both information settings, we generate probabilistic forecast samples using the proposed algorithm based on the selected base forecasting model. For each reconciliation method considered in Step 2, the corresponding probabilistic forecasts are reconciled within the hierarchy to obtain coherent predictive distributions. The accuracy of these reconciled probabilistic forecasts is evaluated using the ES and the VS. Finally, we jointly compare the alternative reconciliation methods under both information settings by combining the evidence from point forecast accuracy and probabilistic forecast accuracy. Specifically, RMSE is used to assess the quality of reconciled point forecasts, while ES and VS are used to assess the quality of reconciled probabilistic forecasts. This comprehensive evaluation helps identify the reconciliation method that provides the most satisfactory overall forecasting performance under each information setting, while also revealing whether the inclusion of weather information improves coherent point and probabilistic forecasts. In this sense, hierarchical coherence is enforced ex post through the reconciliation step rather than during the estimation of the base forecasting models, as illustrated in Figure 3.

4. Empirical study

4.1 Base forecast

In this subsection, we apply the ARIMA, FFNN, and MIDAS models to generate R-step-ahead base forecasts of wheat yields at hierarchical levels. The optimal base forecasting model is selected by incorporating the top five principal components of the weather variables, which together explain about 90% of their total variance. For comparison, we also report forecasts generated without weather information. Since the MIDAS model relies on external regressors (`xreg`), it cannot be employed in the absence of weather data. Here, ARIMA_w, FFNN_w, and MIDAS_w indicate the performance of the ARIMA, FFNN, and MIDAS models, respectively, with the inclusion of weather data.

As shown in Table 3, ARIMA emerges as the optimal model for 56 county-level series, while FFNN performs best for 46. At the category level, one series favors ARIMA and another FFNN, whereas at the state level, ARIMA again provides the most accurate forecasts. Overall, ARIMA appears to be the most effective base model in the absence of weather information.

The last three columns of Table 3 present model selection results when weather information is incorporated. Model performance exhibits notable heterogeneity across counties, where ARIMA

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 3. The number of regions with the best basic prediction accuracy of the three methods

Level	ARIMA	FFNN	Midas	ARIMA_w	FFNN_w	Midas_w
State-level	0	1	–	1	0	0
Category-level	1	1	–	1	1	0
County-level	**56**	46	–	**48**	37	17
Total	**57**	48	–	**50**	38	17

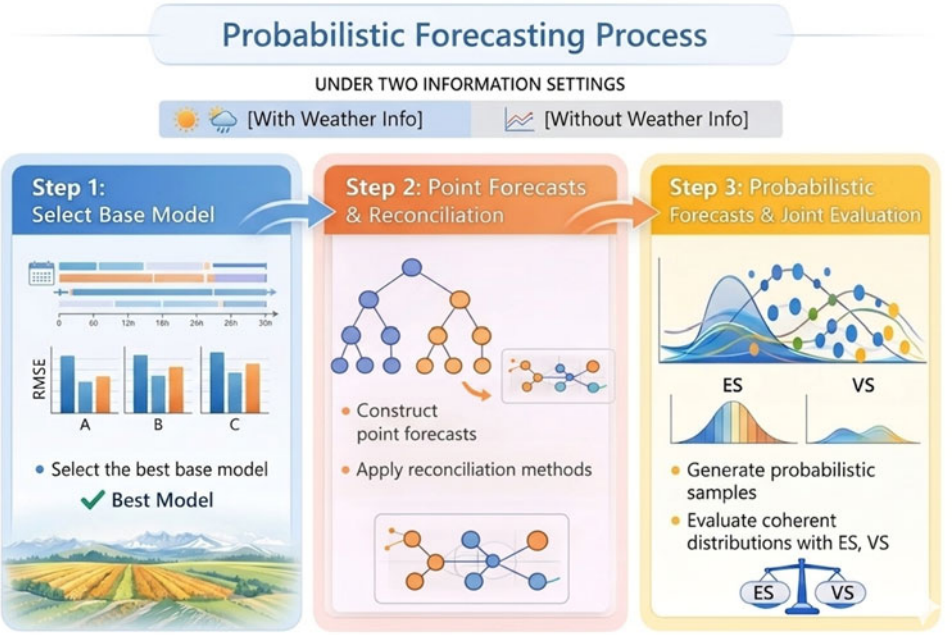


Figure 3. Analysis flowchart.

emerges as the best-performing model in 48 county-level series, followed by FFNN with 37 and MIDAS with 17. At the category level, ARIMA and FFNN each achieve optimal performance in one series. At the state level, ARIMA remains the sole model selected as optimal. Overall, based on the count of series where each model yields the lowest prediction error, ARIMA remains the most effective base forecasting model even when weather information is included.

The results indicate that ARIMA consistently outperforms the other two models of varying complexity in both scenarios with and without weather variables, and its superiority becomes particularly evident when weather information is incorporated.

The methodology first generates base forecasts of wheat yields at the state, category, and county levels using the ARIMA model. These forecasts are subsequently reconciled to ensure hierarchical consistency, yielding coherent hierarchical forecasts through the constrained optimization of the independently produced base forecasts.

4.2 Probabilistic forecast

In this section, the optimal base forecasting model is employed to generate reconciled wheat yield forecasts at the county, category, and state levels under a Gaussian framework. Base forecasts, with and without weather data, serve as benchmarks to quantify the percentage change in forecast accuracy for each reconciliation method. In a subsequent analysis, forecasts without weather data

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 4. Percentage improvements in RMSE for each method relative to the base forecast (with/without weather)

Method	States	Category	County	Method	States	Category	County
Base	–	–	–	Base_w	–	–	–
BU	−3.13	−1.14	**0.00**	BU_w	**−2.58**	**−5.03**	**0.00**
OLS	−0.85	1.08	1.85	OLS_w	0.38	−2.21	2.87
MinT(Sam)	**−3.70**	**−2.95**	0.05	MinT(Sam)_w	7.05	1.83	6.02
MinT(Shr)	−2.82	−1.08	0.06	MinT(Shr)_w	−1.63	−4.51	0.37

Note: RMSE improvements relative to base forecasts: left side without weather, right side with weather; Base row shows average RMSE, and entries below report percentage decreases (negative) or increases (positive) for reconciliation methods.

Table 5. Percentage improvements in base forecast RMSE with weather relative to no weather conditions

Method	States	Category	County	Method	States	Category	County
Base	–	–	–	Base_w	−2.51	2.77	−0.39
BU	−3.13	−1.14	0.00	BU_w	**−5.03**	**−2.39**	**−0.39**
OLS	−0.85	1.08	1.85	OLS_w	−2.14	0.61	2.47
MinT(Sam)	−3.70	−2.95	0.05	MinT(Sam)_w	4.36	4.66	5.60
MinT(Shr)	−2.82	−1.08	0.06	MinT(Shr)_w	−4.10	−1.86	−0.03

Note: RMSE improvements over the base model without weather: Base row shows average RMSE; entries below indicate percentage decreases (negative) or increases (positive) for reconciliation methods.

serve as the benchmark to evaluate relative performance. The results are further assessed using established scoring rules.

Table 4 reports the percentage changes in RMSE of each reconciliation method relative to its corresponding base forecasts. The left panel presents results without weather information, whereas the right panel reports results when weather variables are included in the base models. Under this definition, negative values indicate reductions in RMSE and therefore improvements in forecast accuracy, whereas positive values indicate increases in RMSE and therefore deteriorations relative to the base forecasts. For example, at the county level without weather information, MinT(Sam) records a value of 0.05%, which implies a very slight increase in RMSE rather than an improvement. By contrast, the BU method reproduces the bottom-level base forecasts by construction and therefore shows essentially no change at the county level. Across the higher hierarchical levels, however, MinT(Sam) generally performs better than the alternative reconciliation methods, suggesting that, in the absence of weather information, reconciliation can still improve coherence and predictive accuracy by exploiting cross-series dependence within the hierarchical structure.

Table 5 addresses a different question. Instead of comparing reconciled forecasts with their corresponding base forecasts, it reports the percentage change in RMSE when weather variables are included in the base models relative to the case without weather information. Therefore, this table isolates the incremental contribution of weather information to forecasting performance. The results show that the inclusion of weather variables substantially improves the underlying base forecasts, as evidenced by the reductions in forecast errors relative to the no-weather case. In this setting, the BU method achieves the strongest overall performance across several hierarchical levels. This pattern is intuitive because BU preserves the bottom-level base forecasts exactly, and once weather information has already improved those county-level forecasts, there is less room for additional gains from reconciliation adjustments. Excluding BU, the MinT methods still provide useful reconciliation benchmarks, but the contrast between MinT(Sam) and MinT(Shr) becomes more apparent after weather integration.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

A concise explanation for the discrepancy between MinT(Sam) and MinT(Shr) is that the two methods rely on different estimators of the forecast error covariance matrix. MinT(Sam) uses the sample covariance matrix directly, whereas MinT(Shr) employs a shrinkage estimator that is typically more stable in finite samples. After weather variables are incorporated, part of the common variation across counties is already captured by the weather factors, so the residual dependence across series tends to weaken. At the same time, the base forecasts become more accurate, leaving less systematic structure for reconciliation to exploit. Under these conditions, the sample covariance matrix used by MinT(Sam) may become relatively noisy or ill-conditioned, which can reduce reconciliation efficiency. By contrast, MinT(Shr) regularizes covariance estimation and is therefore less sensitive to estimation noise, leading to more stable performance when weather information is included. This pattern is consistent with the insight of Wickramasuriya *et al*. (2019) that the effectiveness of MinT depends critically on the quality of the estimated forecast error covariance matrix. Taken together, Tables 4 and 5 play complementary roles: Table 4 evaluates the effectiveness of reconciliation relative to the base forecasts, whereas Table 5 evaluates the additional benefit of incorporating weather information into the forecasting framework.

These findings suggest that the effectiveness of reconciliation methods in achieving forecast coherence is inherently context-dependent, and no single approach uniformly dominates across all forecasting environments. In particular, under small-sample conditions, BU may outperform MinT because it preserves the accuracy of the base forecasts at the lower levels of the hierarchy.[5] Correspondingly, after weather variables are incorporated into the forecasting models, MinT tends to deliver smaller gains in forecast accuracy than the BU approach.[6]

This interpretation helps explain the pattern reported in Table 5 in two respects. First, in the absence of weather variables, the county-level base forecasts are relatively weak. Since BU leaves the bottom-level forecasts unchanged by construction, its aggregate-level performance is correspondingly limited and is generally inferior to that of MinT. By contrast, MinT is able to improve forecast accuracy at higher levels of the hierarchy by effectively exploiting residual cross-correlations across series. Second, once weather variables are incorporated, the county-level base forecasts become substantially more accurate, which in turn leads to notable gains at the aggregate level under the BU approach. In this setting, the advantage of preserving the improved bottom-level forecasts becomes more pronounced. At the same time, the effectiveness of MinT reconciliation is reduced, because the weakened residual correlations constrain its ability to further improve forecast accuracy through covariance-based adjustments.

After analyzing the results of R-step-ahead forecasts, we use the method in Section 3 to generate probabilistic forecasting samples, where we set the number of simulated samples for each series to 2000. After we have simulated the wheat yield samples for all series, we evaluate their accuracy by ES and VS scoring rules in Section 3.

[5]Wickramasuriya *et al.* (2019) show that MinT reconciliation performs optimally when base forecasts are relatively inaccurate and residual correlations can be reliably estimated from historical data. In these contexts, corresponding to the scenario uninformed by weather variables, MinT enhances forecast coherence by exploiting residual covariances to capture cross-series dependencies. However, when weather variables are incorporated, the accuracy of the base forecast increases substantially, and residual correlations across regional yield series diminish. This reduction in residual dependence limits MinT's ability to exploit cross-series error structures, weakening its advantage. Moreover, under small-sample conditions, the estimation of residual covariance matrices suffers from the so-called "curse of dimensionality," which further reduces the effectiveness of MinT and may render BU a more robust reconciliation strategy.

[6]The reduced effectiveness of MinT reconciliation in the presence of weather information can be attributed to two main factors. First, reconciliation techniques such as MinT are designed to implicitly capture the influence of latent variables, like weather, when they are omitted from base forecasting models. In such cases, the effects of these omitted variables manifest as correlated residuals across regional forecasts. Given that certain weather conditions often affect multiple regions similarly, residuals from base models across counties may exhibit high cross-sectional correlation. MinT exploits this correlation structure via the residual covariance matrix to adjust and improve forecast coherence. Second, when weather variables are explicitly included in the base models, the explanatory power of the forecasts improves and the residual correlations across regions diminish. This weakens MinT's ability to extract a useful signal from residual dependencies, limiting its effectiveness in enhancing forecast accuracy.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 6. Scoring results

Weather conditions	Method	ES ($\times 10^7$)	VS ($\times 10^{24}$)
Model	Base	4.619	5.135
	BU	4.047	3.945
	OLS	4.610	5.083
	MinT(Sam)	**3.905**	**3.623**
	MinT(Shr)	4.096	4.025
Model under weather	Base	4.457	4.391
	BU	**3.590**	**3.043**
	OLS	4.393	4.279
	MinT(Sam)	3.876	3.301
	MinT(Shr)	3.801	3.330

The results for the ES and VS are reported in Table 6. Consistent with the conclusions drawn from the *R*-step-ahead forecast evaluations incorporating weather variables, it is observed that, with the exception of the ARIMA model, which attains the lowest VS under the MinT(Sam) reconciliation, all other models achieve their lowest scores under the same reconciliation method across the techniques considered. In addition, both ES and VS scores for MinT(Shr) and MinT(Sam) are consistently lower than those for OLS, corroborating the findings reported in Wickramasuriya (2024). These results collectively indicate that the BU method approach exhibits superior performance under the Gaussian assumption. Moreover, the BU method demonstrates consistent improvements across multiple scoring metrics, underscoring its robustness and effectiveness in enhancing probabilistic forecast accuracy within hierarchical time series structures.

Similarly, ES and VS are derived from probabilistic forecast reconciliation results incorporating weather variables. As shown in Table 6 above, the scores for most methods are generally lower after incorporating weather-related variables, indicating improved predictive performance compared to forecasts without weather inputs. Upon further examination, the BU method yields the lowest ES and VS, indicating the most substantial improvement in forecast accuracy for hierarchical time series among the reconciliation approaches considered. The performances of MinT and OLS are ranked second and third, respectively, with both methods yielding relatively higher scores. It is also important to emphasize that, regardless of the reconciliation method employed, all approaches consistently enhance the forecast accuracy of hierarchical time series forecasts when compared to the unreconciled probabilistic forecasts.

In summary, whether for point forecast reconciliation or probabilistic reconciliation under the Gaussian assumption, the optimal reconciliation technique should be selected based on the specific context. If the objective is solely to forecast crop yields, BU is recommended when weather information is available. While climate change simulations become more realistic, the associated climate data are costly, technically demanding, and difficult to obtain. Consequently, MinT is preferable in the absence of weather data, as acquiring such data entails high costs, substantial technical requirements, and considerable implementation challenges. Similarly, if the goal is to assess the extent to which the hierarchical series conforms to the Gaussian assumption, BU remains the appropriate choice.

4.3 Quantification of yield risks under climate variations

This subsection develops a parsimonious scenario-based sensitivity analysis framework to assess how weather fluctuations affect county-level wheat yield risk and to support cross-county risk ranking. Because weather variables are numerous and highly correlated, we first reduce the

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

weather vector to a small set of orthogonal factors using PCA. To ensure both interpretability and computational tractability, we implement two complementary perturbation schemes (Campolongo *et al.*, 2007; Borgonovo & Plischke, 2016).

(i) Principal-component-level perturbations (factor-space stress testing): PC1 and PC2 are perturbed separately using multiplicative stress factors. The factors 0.8 and 1.2, 0.5 and 1.5, and 0.2 and 1.8 correspond to shocks of ±20%, ±50%, and ±80%, respectively. This produces internally coherent weather regimes in the reduced factor space. With two principal components and six stress settings for each, the design yields 12 principal-component-level perturbation scenarios (2 components × 6 settings).

(ii) Variable-level perturbations based on the ten highest-loading variables (targeted screening in the original variable space): For each county-year observation, we identify the ten original weather variables with the largest absolute loadings on PC1 and PC2. Each of these variables is then perturbed in turn, holding all other inputs fixed, and the perturbed observations are projected back into the principal-component space. This yields a structured one-at-a-time screening design that focuses on the most influential weather variables. Applying this procedure to ten variables for each of the two principal components under six stress settings produces 120 variable-level perturbation scenarios (2 components × 10 variables × 6 settings).

For each perturbation scenario, we re-estimate the probabilistic yield forecast using the procedure in Subsection 4.2. We then compute a county-year risk measure, denoted by r, from the predictive distribution under stressed weather conditions. Specifically, wheat yield risk is defined as the average relative weather-related yield reduction rate:

$$r = \frac{\sum_t^T \min\left(y^{\text{Recon}}_{t,\text{new climate}} - y^{\text{Recon}}_{t,\text{climate}},\ 0\right)}{\sum_t^T y^{\text{Recon}}_{t,\text{climate}}},$$

where $y^{\text{Recon}}_{t,\text{climate}}$ denotes the median probabilistic reconciled forecast in year t under baseline climate conditions, and $y^{\text{Recon}}_{t,\text{new climate}}$ denotes the corresponding forecast under perturbed climate conditions. A smaller value of this metric indicates higher yield risk. In the Montana wheat-yield hierarchy, the BU method is optimal. Since BU preserves the original county-level forecasts, county-level yields remain unchanged before and after reconciliation for both wheat types; nevertheless, this metric is retained to evaluate yield risk under simulated climate perturbations.

Figure 4 shows the county-level spatial distribution of wheat-yield downside risk in Montana under four representative scenarios in which PC1 and PC2 are scaled to 0.5 and 1.5. Gray areas denote counties excluded from the empirical analysis. The maps reveal marked spatial heterogeneity across counties and wheat types. Spring-wheat risk is highly concentrated in a west-central to southwestern cluster, whereas winter-wheat risk is more spatially dispersed, although its highest-risk counties are still concentrated mainly in western and west-central Montana.

The effects of climate perturbations differ across PCs and wheat varieties. For spring wheat, the main hotspot remains broadly stable across scenarios, but the intensity varies with the perturbed component: the strongest hotspot appears under the 0.5×PC1 scenario, while under PC2, the risk becomes more pronounced in the 1.5×PC2 scenario. For winter wheat, the pattern remains more diffuse, with repeated concentrations of higher risk in western Montana and scattered pockets in interior counties. The 1.5×PC1 scenario produces the clearest localized winter-wheat hotspot, whereas PC2-driven scenarios generate milder but more widespread risk. Overall, spring-wheat vulnerability is more localized, whereas winter-wheat risk is broader but less sharply clustered.

Additional results from the principal-component-level perturbation analysis are reported in Appendix B (in the Supplementary Materials). Across the six perturbation scenarios, three patterns emerge. First, the spatial ordering of vulnerable counties is fairly stable. Second, the response

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

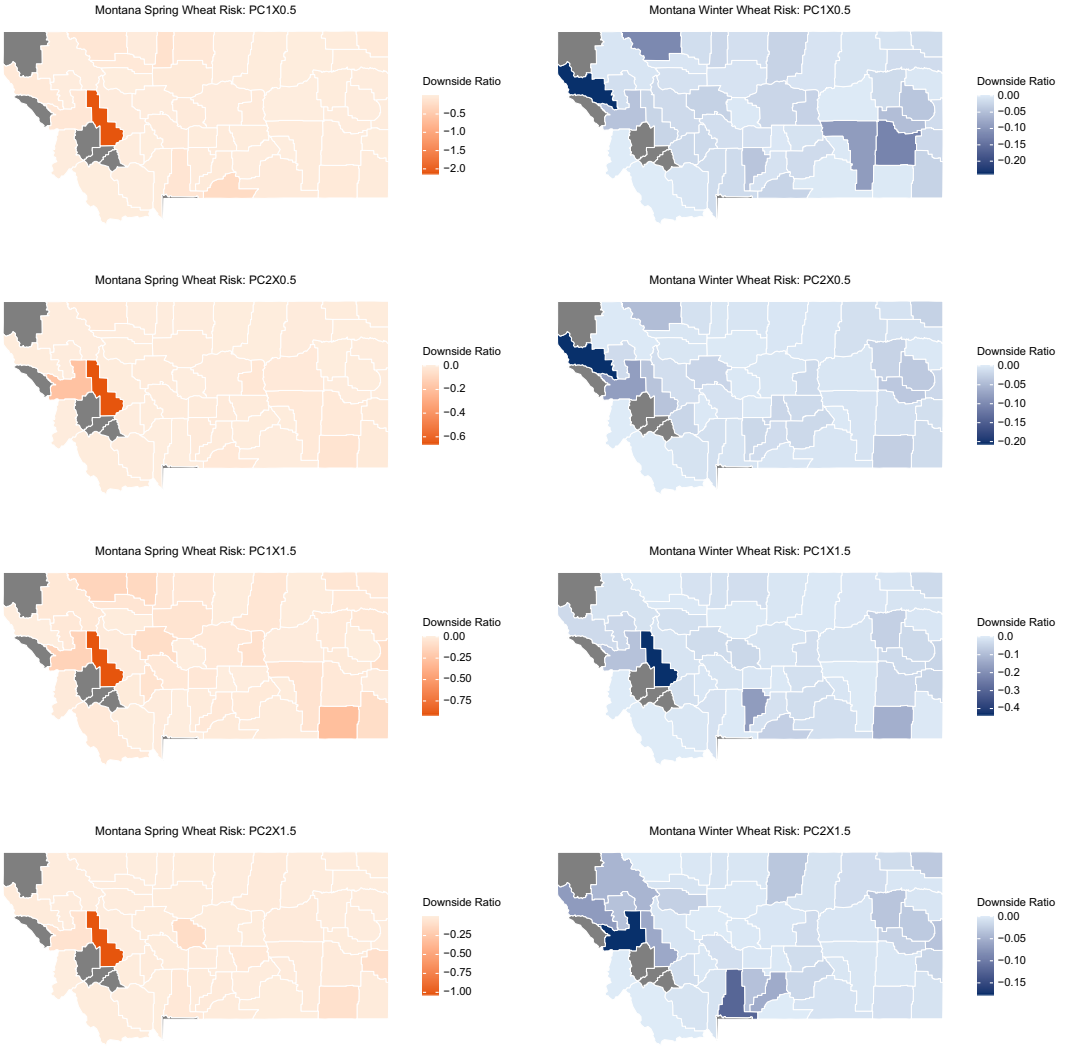


Figure 4. Montana county spring/winter wheat risk maps(±50%).
Note: The darker colors represent greater risk, and the gray areas were not included in the empirical analysis.

to climate perturbations is asymmetric: negative shocks mainly reinforce the existing pattern, whereas positive shocks, especially to PC1, tend to intensify and broaden the spatial reach of risk. Third, shocks to PC1 generally induce greater and more systematic changes than shocks to PC2, suggesting that PC1 captures the dominant weather-related source of yield vulnerability in Montana.

Results from the variable-level perturbation analysis are reported in Appendix C (in the Supplementary Materials). For spring wheat, the highest-risk counties remain concentrated in a relatively stable west-central to southwestern core across most scenarios, with differences mainly reflected in the extent of spatial spillovers under stronger shocks or particular variables. For winter wheat, the maps show a more spatially dispersed and often multi-center pattern, with recurrent hotspots in western or northwestern Montana and isolated pockets in interior or eastern counties. Perturbations related to PC1 tend to produce sharper localized spikes, whereas those related to PC2 generate a broader but milder spread.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 7. Risk ranking of counties in spring wheat yield, winter wheat yield, and total wheat yield in overall PC shocks (the higher ranking represents greater risk)

County	Spring_rank	Winter_rank	Total_rank	County	Spring_rank	Winter_rank	Total_rank
Beaverhead	41	19	30	Meagher	23	40	34
Big Horn	37	44	47	Missoula	16	1	1
Blaine	49	17	33	Musselshell	28	31	37
Broadwater	20	43	35	Park	15	8	13
Carbon	1	13	4	Petroleum	11	24	17
Carter	21	39	31	Phillips	17	47	28
Cascade	12	26	18	Pondera	18	45	29
Chouteau	13	51	26	Powder River	2	25	8
Custer	10	5	5	Powell	4	12	7
Daniels	29	9	16	Prairie	46	10	22
Dawson	47	6	11	Ravalli	35	22	27
Fallon	3	20	9	Richland	31	50	48
Fergus	26	41	38	Roosevelt	39	33	44
Flathead	36	28	36	Rosebud	6	16	12
Gallatin	48	35	49	Sanders	27	2	2
Garfield	43	42	50	Sheridan	24	21	24
Glacier	7	11	10	Stillwater	19	15	19
Golden Valley	30	48	46	Sweet Grass	22	7	14
Hill	14	32	25	Teton	50	4	6
Jefferson	44	30	45	Toole	5	36	15
Judith Basin	9	37	21	Treasure	45	49	51
Lake	34	23	32	Valley	40	29	42
Lewis and Clark	33	38	43	Wheatland	25	46	40
Liberty	8	34	20	Wibaux	32	3	3
Madison	51	18	41	Yellowstone	42	27	39
McCone	38	14	23				

Taken together, the principal-component-level and variable-level perturbation analyses provide a consistent picture of wheat-yield risk in Montana. Both reveal pronounced spatial heterogeneity: spring-wheat risk is localized in a persistent inland hotspot, whereas winter-wheat risk is more spatially diffuse. The PC-level analysis indicates that PC1 plays a stronger role than PC2 in shaping county-level downside exposure, while the variable-level analysis highlights substantial heterogeneity in local intensity and spillover patterns across individual climate variables.

Tables 7 and 8 show that county-level wheat-yield risk rankings remain broadly stable across the two perturbation frameworks. Missoula ranks first in total wheat-yield risk under both approaches, and seven of the ten highest-risk counties are common to both tables, indicating a robust spatial concentration of wheat-yield risk.

At the same time, meaningful county-level differences remain. Relative to the principal-component-level ranking, the variable-level analysis places Dawson, Ravalli, and Sheridan higher in the total-risk distribution, while Carbon, Glacier, and Powell move downward; treasure shows one of the largest rank shifts. The results also suggest that in counties such as Missoula and Sanders, severe winter-wheat risk contributes substantially to total wheat-yield risk. Overall, the two approaches indicate a stable aggregate pattern of spatial vulnerability while revealing local

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 8. Risk ranking of counties in spring wheat yield, winter wheat yield, and total wheat yield in Top-10 loading shocks (the higher ranking represents greater risk)

County	Spring_rank	Winter_rank	Total_rank	County	Spring_rank	Winter_rank	Total_rank
Beaverhead	42	21	30	Meagher	39	20	28
Big Horn	41	47	51	Missoula	19	1	1
Blaine	48	30	46	Musselshell	20	11	13
Broadwater	25	44	41	Park	45	25	38
Carbon	4	28	12	Petroleum	7	24	18
Carter	9	39	24	Phillips	11	40	25
Cascade	10	42	27	Pondera	28	43	44
Chouteau	13	51	33	Powder River	2	27	10
Custer	1	17	4	Powell	5	18	16
Daniels	27	15	20	Prairie	40	14	21
Dawson	44	2	3	Ravalli	30	4	6
Fallon	3	8	2	Richland	17	46	35
Fergus	23	49	45	Roosevelt	34	36	43
Flathead	46	32	47	Rosebud	6	12	15
Gallatin	50	9	11	Sanders	18	6	7
Garfield	33	31	36	Sheridan	21	7	8
Glacier	12	50	31	Stillwater	16	16	17
Golden Valley	15	23	22	Sweet Grass	29	13	19
Hill	14	33	26	Teton	49	5	9
Jefferson	47	35	50	Toole	8	37	23
Judith Basin	26	34	37	Treasure	31	26	29
Lake	43	22	34	Valley	37	29	39
Lewis and Clark	38	41	48	Wheatland	22	48	40
Liberty	32	38	42	Wibaux	24	3	5
Madison	51	19	32	Yellowstone	35	45	49
McCone	36	10	14				

reallocations of risk across counties, highlighting the importance of mitigation strategies that address both aggregate climate-factor movements and local weather-related factors.

5. Application to crop insurance rating games

In this section, we apply the proposed forecasting framework to crop insurance ratemaking. Our objective is to examine whether the proposed framework exhibits greater efficiency than the RMA ratemaking approach described in Section 5.1. To this end, we employ a simple game-theoretic model, which has been widely adopted in the literature, to compare the accuracy of the two methods (Ker & McGowan, 2000; Harri *et al*., 2011; Zhang, 2017; Ramsey, 2020; Liu & Ker, 2021; Zhu, 2024).

Typically, the determination of insurance premium rates is a two-stage process (Zhu *et al*., 2011). First, the historical trend of crop yield data is estimated, whether the data exhibit a deterministic or stochastic trend (Skees & Reed, 1986; Kaylen & Koroma, 1991; Goodwin & Hungerford, 2015), and the residuals from the trend estimation are adjusted for heteroskedasticity (Harri *et al*., 2011). Second, the adjusted yield data are used to estimate the yield distribution, which then serves as the basis for premium rate calculation. Among these steps, trend fitting and residual adjustment are the key procedures for achieving accurate pricing.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

5.1 RMA methodology

The United States Department of Agriculture's Risk Management Agency (RMA) manages the expenditures of crop insurance in the U.S. Currently, RMA estimates the county-level historical yield data with a two-knot linear spline (Liu & Ker, 2021):

$$y_t = \theta_1 + \theta_2 t + \delta_1 d_1(t - k_1) + \delta_2 d_2(t - k_2) + \varepsilon_t,$$

where d_1 and d_2 denote indicator functions that capture these structural breaks, defined such that $d_1 = 1$ when $t \geq k_1$ and $d_2 = 1$ when $t \geq k_2$, and both are zero otherwise. The parameters k_1 and k_2 represent the locations of the two structural break points (or knots), which are selected within the range $(1 + \bar{k}, \ldots, T - \bar{k})$. After the trend estimation, we denote the residuals as $\hat{\varepsilon}_t$, and the fitted values as $\hat{y}_t$. The degree of heteroscedasticity is estimated via Harri *et al.* (2011):

$$\ln\left(\hat{\varepsilon}_t^2\right) = \alpha + \gamma \ln\left(\hat{y}_t\right) + \nu_t,$$

where ν_t is an i.i.d. error term with $E(\nu_t) = 0$ and $\mathrm{Var}(\nu_t) = \sigma_\nu^2$. After that, yields are adjusted based on a one-step-ahead forecast ($\hat{y}_{T+1}$) and the estimated heteroscedasticity coefficient ($\hat{\gamma}$):

$$\hat{y}_t^* = \hat{y}_{T+1} + \hat{\varepsilon}_t \left(\frac{\hat{y}_{T+1}}{\hat{y}_t}\right)^{\frac{\hat{\gamma}}{2}}.$$

The adjusted yields can then be used to estimate conditional yield distributions or generate empirical premium rates:

$$R_{T+1,i}^{RMA} = \frac{1}{T}\sum_{t=1}^{T} \max\left\{0, \frac{\lambda y_{T+1,i}^* - \hat{y}_{t,i}^*}{\lambda y_{T+1,i}^*}\right\},$$

where λ is the coverage ratio such that $\lambda y_{T+1,i}^*$ is the yield guarantee of insurance.

5.2 Crop insurance rating games

In this section, we consider a crop insurance rating game to better determine the insurance premium rates for state-level wheat data in Montana. First, we introduce the formula for determining the insurance premium rate. The premium rate is calculated as the ratio between the expected indemnity (I) and the insurance liability (L). In other words, under a given coverage level, the premium rate can be interpreted as the expected loss per unit of insurance liability. Let $\lambda \in (0, 1)$ be the coverage ratio and i be the county index, then the liability can be computed as $L = \lambda E(y_i)$. If the realized yield falls short of L, then the insurance contract is triggered and the indemnity is calculated as the shortfall in yield; that means $I = \max(0, L - y_i)$. Thus, the premium rate R_i can be written as follows:

$$R_i = \frac{E(I)}{L} = \frac{E(\max(0, \lambda E(y_i) - y_i))}{\lambda E(y_i)}.$$

In our forecasting framework, the premium rate R_i is calculated to reflect its dynamic adjustment to weather variations as:

$$R_i = \frac{E(\max(0, \lambda(1 + E(r))\widetilde{\boldsymbol{y}}_i^{0.5} - \widetilde{\boldsymbol{y}}_i))}{\lambda(1 + E(r))\widetilde{\boldsymbol{y}}_i^{0.5}},$$

$\widetilde{\boldsymbol{y}}_i$ represents the probabilistic reconciliation forecast result for the county, and $\widetilde{\boldsymbol{y}}_i^{0.5}$ denotes the median of the probabilistic reconciliation forecast. The coverage ratio λ, treated as an adjustable model parameter, is incorporated together with the expected value of r under varying weather conditions, denoted as $E(r)$. Accordingly, $\lambda(1 + E(r))\widetilde{\boldsymbol{y}}_i^{0.5}$ represents the yield guarantee after yield risk-based adjustments. This term represents a risk-adjusted yield threshold that incorporates the

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

expected climate risk $E(r)$ and the adjustment coefficient λ. In this study, λ is set to be 90%, reflecting a conservative risk protection level. When the yield $\widetilde{\boldsymbol{y}}_i$ falls below this yield guarantee, it indicates that a climate-induced yield loss has occurred. In other words, the guarantee value defines the minimum yield level that can be ensured under climate uncertainty, or the yield level that can be achieved with high confidence after risk adjustment. It is analogous to a "minimum guaranteed return" in finance or insurance, representing a climate risk–adjusted guaranteed yield in the agricultural context.

In this insurance pricing game, there are two types of participants: the U.S. Department of Agriculture (USDA), which adopts the RMA method, and private insurers, which employ our proposed forecasting framework. During the game, the private insurer transfers a portion of its policy portfolio to the USDA. Specifically, when the insurer's premium rate is lower than the RMA rate ($R_i < R_i^{RMA}$), the insurer retains the policy; conversely, when the insurer's premium rate is higher than the RMA rate ($R_i > R_i^{RMA}$), the insurer cedes the policy to the government. Therefore, the private insurer strategically updates its risk rating system to engage in adverse selection against the USDA.

In the preceding part, the proposed game-theoretic pricing mechanism was introduced. Its effectiveness relative to the RMA method is evaluated by the accuracy of the resulting pricing, quantified using loss ratios. In the end, loss ratios are calculated for both the insurer's portfolio and the government's portfolio. If the insurer's internal pricing system based on the proposed forecasting framework is more accurate than the RMA method, the insurer's loss ratio should be lower. The loss ratio (LR) for a set of policies, denoted by Ω, is defined as follows (Ker *et al.*, 2016):

$$\mathrm{LR}_{\Omega} = \frac{\sum_{i\in\Omega} \max\left(0, \hat{y}_i^G - y_i\right)}{\sum_{i\in\Omega} R_i^{RMA}},$$

where R_i^{RMA} is the RMA estimated premium rate, $\hat{y}_i^G$ is the yield guarantee for policies ($\hat{y}_i^G = \lambda y_{T+1,i}^*$ for RMA, and $\hat{y}_i^G = \lambda(1 + E(r))\widetilde{\boldsymbol{y}}_{i|0.5}$ for our model), and y_i is the actual yield.

We calculate premium rates for spring wheat and winter wheat across 51 counties using one-step-ahead rolling forecasts. Tables 9 and 10 report the rating-game results under two weather-adjustment schemes, namely the overall PC specification and the top-10 loading specification. The results show that the two adjustment schemes perform very similarly: for both spring wheat and winter wheat, they generate comparable retention rates and nearly identical mean loss ratios for the private insurer. On average, retention remains economically meaningful, at around 35%–38% across crop types, and the private insurer's mean loss ratios are also very close under the two schemes. More importantly, both adjusted pricing schemes generally improve upon the RMA benchmark. Although in a few individual years the private insurer's loss ratio may exceed that of the government program, the average private loss ratio is lower than the corresponding RMA loss ratio for both spring wheat and winter wheat, while still preserving a nontrivial share of retained counties. Hence, the evidence suggests that both the overall-PC and top-10-loading adjustments provide robust and practically similar improvements over the RMA-based pricing approach.

6. Conclusion

This study proposes a unified forecasting framework for crop yield prediction and climate risk stratification. A hierarchical time series structure, encompassing county, category, and state levels, is constructed to capture the spatial and structural complexity of agricultural production. Forecast reconciliation is employed to ensure hierarchical coherence and improve forecast accuracy across all levels. Moreover, probabilistic forecasting under Gaussian assumptions facilitates simulation-based assessment of climate yield loss rates and the ranking of regional yield risks. This forecasting approach not only enables the evaluation of the overall accuracy of hierarchical time series forecasts but also enhances the credibility of the risk assessment derived from these forecasts. This

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Table 9. Crop insurance rating game (overall PC)

Category	Year	Number of counties	Retained by private (%)	Mean LR_private	Mean LR_RMA
Spring wheat	2015	51	15.7	1.326	6.961
Spring wheat	2016	51	35.3	2.818	5.635
Spring wheat	2017	51	19.6	2.179	4.707
Spring wheat	2018	51	52.9	0.93	0.619
Spring wheat	2019	51	68.6	2.432	1.715
Spring wheat	2020	51	41.2	0.439	0.589
Spring wheat	2021	51	31.4	4.043	10.883
Winter wheat	2015	51	37.3	1.859	2.448
Winter wheat	2016	51	41.2	1.741	1.614
Winter wheat	2017	51	27.5	3.133	3.941
Winter wheat	2018	51	29.4	2.222	1.973
Winter wheat	2019	51	43.1	0.612	0.743
Winter wheat	2020	51	31.4	1.718	1.991
Winter wheat	2021	51	35.3	1.69	3.135

Table 10. Crop insurance rating game (Top-10 loading)

Category	Year	Number of counties	Retained by private (%)	Mean LR_private	Mean LR_RMA
Spring wheat	2015	51	15.7	1.308	6.961
Spring wheat	2016	51	39.2	2.762	5.635
Spring wheat	2017	51	21.6	2.182	4.707
Spring wheat	2018	51	49	0.923	0.619
Spring wheat	2019	51	64.7	2.405	1.715
Spring wheat	2020	51	41.2	0.437	0.589
Spring wheat	2021	51	29.4	3.966	10.883
Winter wheat	2015	51	37.3	1.92	2.448
Winter wheat	2016	51	45.1	1.769	1.614
Winter wheat	2017	51	25.5	3.132	3.941
Winter wheat	2018	51	27.5	2.1	1.973
Winter wheat	2019	51	45.1	0.619	0.743
Winter wheat	2020	51	35.3	1.675	1.991
Winter wheat	2021	51	39.2	2.018	3.135

allows for the identification of high-risk areas under different climate variation scenarios and provides a foundation for evidence-based climate risk management.

Empirical analysis using county-level spring and winter wheat data from Montana, the U.S., demonstrates the effectiveness of the proposed framework. The results indicate that incorporating hierarchical reconciliation and weather-adjusted forecasts significantly improves the accuracy of the base forecast. Moreover, the yield risk assessment framework effectively distinguishes regional patterns of yield vulnerability, showing that spring wheat risk is more spatially concentrated, whereas winter wheat risk is more dispersed across counties. When applied to the crop insurance rating game, the proposed forecasting framework, relative to the RMA method, enables insurers to effectively diversify part of their risk in the premium determination for each type of wheat, while simultaneously providing valuable guidance for policymakers. These findings carry significant implications for policymakers, insurers, and reinsurers. More accurate crop yield forecasts and

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

climate risk classifications can support the design of more precise insurance schemes, optimize subsidy allocation, and facilitate the implementation of adaptive strategies to mitigate agricultural risks under changing climatic conditions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1748499526100372.

Data availability statement. Replication materials are available on request from the authors. The data and code that support the findings of this study are available from the corresponding author, Daning Bi, upon reasonable request.

Funding statement. This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grant No. 72201093. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. The authors declare none.

References

Annan, **F.**, **Tack**, **J.**, **Harri**, **A.**, & **Coble**, **K.** (2014). Spatial pattern of yield distributions: Implications for crop insurance. *American Journal of Agricultural Economics*, **96**(1), 253–268.

Bai, **J.**, & **Ng**, **S.** (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**(1), 191–221.

Bangelesa, **F.**, **Pollinger**, **F.**, **Sponholz**, **B.**, **Mapatano**, **M.**, **Hatloy**, **A.**, & **Paeth**, **H.** (2023). Statistical-dynamical modeling of the maize yield response to future climate change in West, East and Central Africa using the regional climate model REMO. *Science of The Total Environment*, **905**, 167265.

Berry, **L.**, **Helman**, **P.**, & **West**, **M.** (2020). Probabilistic forecasting of heterogeneous consumer transaction–sales time series. *International Journal of Forecasting*, **36**(2), 552–569.

Borgonovo, **E.**, & **Plischke**, **E.** (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, **248**(3), 869–887.

Campolongo, **F.**, **Cariboni**, **J.**, & **Saltelli**, **A.** (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, **22**(10), 1509–1518.

Chong, **W.**, **Feng**, **R.**, & **Jin**, **L.** (2023). Holistic principle for risk aggregation and capital allocation. *Annals of Operations Research*, **330**(1), 21–54.

Coulombe, **P.**, **Marcellino**, **M.**, & **Stevanović**, **D.** (2021). Can machine learning catch the covid-19 recession? *National Institute Economic Review*, **256**, 71–109.

Du, **X.**, **Hennessy**, **D. A.**, **Feng**, **H.**, & **Arora**, **G.** (2018). Land resilience and tail dependence among crop yield distributions. *American Journal of Agricultural Economics*, **100**(3), 809–828.

Dunn, **D. M.**, **Williams**, **W. H.**, & **DeChaine**, **T. L.** (1976). Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association*, **71**(353), 68–71.

Goodwin, **B.**, & **Hungerford**, **A.** (2015). Copula-based models of systemic risk in US agriculture: Implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, **97**(3), 879–896.

Guan, **Y.**, **Tsanakas**, **A.**, & **Wang**, **R.** (2023). An impossibility theorem on capital allocation. *Scandinavian Actuarial Journal*, **2023**(3), 290–302.

Harri, **A.**, **Coble**, **K. H.**, **Ker**, **A. P.**, & **Goodwin**, **B. J.** (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, **93**(3), 707–717.

He, **X. H.**, **Luo**, **H. T.**, **Qiao**, **M. J.**, **Tian**, **Z. H.**, & **Zhou**, **G. S.** (2021). Yield estimation of winter wheat in China based on CNN-RNN network. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, **37**(17), 124–132.

Hyndman, **R.**, **Ahmed**, **R.**, **Athanasopoulos**, **G.**, & **Shang**, **H.** (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, **55**(9), 2579–2589.

Hyndman, **R.**, & **Athanasopoulos**, **G.** (2021). *Forecasting: Principles and practice*. OTexts.

Kaylen, **M.**, & **Koroma**, **S.** (1991). Trend, weather variables, and the distribution of US corn yields. *Applied Economic Perspectives and Policy*, **13**(2), 249–258.

Ker, **A.**, & **McGowan**, **P.** (2000). Weather-Based adverse selection and the US Crop Insurance Program: The private insurance company perspective. *Journal of Agricultural and Resource Economics*, **25**(2), 386–410.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Ker, A., Tolhurst, T., & Liu, Y. (2016). Bayesian estimation of possibly similar yield densities: Implications for rating crop insurance contracts. *American Journal of Agricultural Economics*, **98**(2), 360–382.

Kleiber, W., Raftery, A., & Gneiting, T. (2011). Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *Journal of the American Statistical Association*, **106**(496), 1291–1303.

Leutbecher, M. (2019). Ensemble size: How suboptimal is less than infinity? *Quarterly Journal of the Royal Meteorological Society*, **145**, 107–128.

Li, H., Li, H., Lu, Y., & Panagiotelis, A. (2019). A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance: Mathematics and Economics*, **86**, 122–133.

Li, H., Porth, L., Tan, K., & Zhu, W. (2021). Improved index insurance design and yield estimation using a dynamic factor forecasting approach. *Insurance: Mathematics and Economics*, **96**, 208–221.

Liu, J. M., He, X. T., Wang, P. X., & Huang, J. X. (2019). Early prediction of winter wheat yield with long time series meteorological data and random forest method. *Transactions of the Chinese Society of Agricultural Engineering*, **35**(6), 158–166.

Liu, Y., & Ker, A. (2021). Simultaneous borrowing of information across space and time for pricing insurance contracts: An application to rating crop insurance policies. *Journal of Risk and Insurance*, **88**(1), 231–257.

Lobell, D., Schlenker, W., & Roberts, C. (2011). Climate trends and global crop production since 1980. *Science*, **333**(6042), 616–620.

Luchtenbelt, H., Doelman, J., Bos, A., Daioglou, V., Jägermeyr, J., Müller, C., Stehfest, E., & van Vuuren, D. (2024). Quantifying food security and mitigation risks consequential to climate change impacts on crop yields. *Environmental Research Letters*, **20**(1), 014001.

Meng, H. Y., Wang, Z. L., Yao, P., Qian, L., Chen, C., Luo, Y. Y., & Ju, X. L. (2022). Comparative study of different meteorological yield separation methods in waterlogging disaster assessment. *Chinese Journal of Eco-Agriculture*, **30**(6), 976–989.

Meyers, G. (2001). The cost of financing insurance. In Casualty actuarial society forum (p. 221).

Meyers, G., Klinker, F., & Lalonde, D. (2003). The aggregation and correlation of insurance exposure. In Casualty actuarial society forum (pp. 15–82).

Orcutt, G. H., Watts, H. W., & Edwards, J. B. (1968). Data aggregation and information loss. *American Economic Review*, **58**(4), 773–787.

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., & Hyndman, R. (2023). Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. *European Journal of Operational Research*, **306**(2), 693–706.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, **38**(3), 705–871.

Ramsey, A. (2020). Probability distributions of crop yields: A Bayesian spatial quantile regression approach. *American Journal of Agricultural Economics*, **102**(1), 220–239.

Rezaei, E. E., Webber, H., Asseng, S., Boote, K., Durand, J. L., Ewert, F., Martre, P., & MacCarthy, D. S. (2023). Climate change impacts on crop yields. *Nature Reviews Earth & Environment*, **4**(12), 831–846.

Scheuerer, M., & Hamill, T. M. (2015). Variogram-Based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, **143**(4), 1321–1334.

Skees, J., & Reed, M. (1986). Rate making for farm-level crop insurance: Implications for adverse selection. *American Journal of Agricultural Economics*, **68**(3), 653–659.

Summary, Estimation and Disclosure Methodology Branch, National Agricultural Statistics Service, Methodology Division (2023). The yield forecasting program of NASS. NASS Research Reports 343273, United States Department of Agriculture, National Agricultural Statistics Service.

Sun, S., Yang, X. G., Zhang, Z. T., Zhao, J., & Liu, Z. J. (2021). Impacts of different grades of drought on winter wheat yield in North China Plain. *Transactions of the Chinese Society of Agricultural Engineering*, **37**(14), 69–78.

Taieb, S., Huser, R., Hyndman, R., & Genton, M. (2016). Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid*, **7**(5), 2448–2455.

Taieb, S., Taylor, J., & Hyndman, R. (2021). Hierarchical probabilistic forecasting of electricity demand with smart meter data. *Journal of the American Statistical Association*, **116**(533), 27–43.

Tan, K. Y., Wu, D. R., & Zhao, H. R. (2017). Trend analysis of temperature conditions over different growth periods of winter wheat under climate warming in North China Plain. *Chinese Journal of Agrometeorology*, **38**(6), 333.

Wickramasuriya, S. (2024). Probabilistic forecast reconciliation under the Gaussian framework. *Journal of Business & Economic Statistics*, **42**(1), 272–285.

Wickramasuriya, S., Athanasopoulos, G., & Hyndman, R. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, **114**(526), 804–819.

Zellner, A., & Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, **19**(5), 457–465.

Zhang, K., Gong, L. Q., He, Z., & Huo, P. C. (2023). Crop yield prediction based on 3DCNN-TCN model. *Microelectronics & Computer*, **40**(10), 83–89.

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press

Zhang, **Y. Y.** (2017). A density-ratio model of crop yield distributions. *American Journal of Agricultural Economics*, **99**(5), 1327–1343.

Zhao, **C.**, **Liu**, **B.**, **Piao**, **S. L.**, **Wang**, **X. H.**, **Lobell**, **D. B.**, **Huang**, **Y.**, **Huang**, **M. T.**, **Yao**, **Y. T.**, **Bassu**, **S.**, **Ciais**, **P.**, **Durand**, **J-L.**, **Elliott**, **J.**, **Ewert**, **F.**, **Janssens**, **I. A.**, **Li**, **T.**, **Lin**, **E.**, **Liu**, **Q.**, **Martre**, **P.**, **Müller**, **C.**, . . ., **Asseng**, **S.** (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of The National Academy of Sciences of The United States of America*, **114**(35), 9326–9331.

Zhu, **W.** (2024). A deep factor model for crop yield forecasting and insurance ratemaking. *North American Actuarial Journal*, **28**(1), 57–72.

Zhu, **W.**, **Tan**, **K.**, **Porth**, **L.**, & **Wang**, **C.** (2018). Spatial dependence and aggregation in weather risk hedging: A Lévy subordinated hierarchical Archimedean copulas (LSHAC) approach. *ASTIN Bulletin: The Journal of the IAA*, **48**(2), 779–815.

Zhu, **Y.**, **Goodwin**, **B.**, & **Ghosh**, **S.** (2011). Modeling yield risk under technological change: Dynamic yield distributions and the US Crop Insurance Program. *Journal of Agricultural and Resource Economics*, **36**(1), 192–210.

Cite this article: Zhang N, Yang J, Xie Z and Bi D (2026). Quantifying and pricing crop yield risk under climate change: a hierarchical model forecasting perspective, *Annals of Actuarial Science*, 1–25. https://doi.org/10.1017/S1748499526100372

https://doi.org/10.1017/S1748499526100372 Published online by Cambridge University Press