



International Journal of Research in Agronomy

E-ISSN: 2618-0618

P-ISSN: 2618-060X

© Agronomy

www.agronomyjournals.com

2024; 7(2): 130-132

Received: 10-12-2023

Accepted: 12-01-2024

Erik Lindqvist

Department of Agricultural
Statistics, Swedish University of
Agricultural Sciences, Uppsala,
Sweden

Maja Karlsson

Department of Agricultural
Statistics, Swedish University of
Agricultural Sciences, Uppsala,
Sweden

Nils Johansson

Department of Agricultural
Statistics, Swedish University of
Agricultural Sciences, Uppsala,
Sweden

Corresponding Author:

Erik Lindqvist

Department of Agricultural
Statistics, Swedish University of
Agricultural Sciences, Uppsala,
Sweden

Bayesian hierarchical modelling for regional crop yield prediction under climate uncertainty

Erik Lindqvist, Maja Karlsson and Nils Johansson

DOI: <https://doi.org/10.33545/2618060X.2024.v7.i2c.4972>

Abstract

Conventional crop yield models give a single point estimate; they do not tell you how confident to be. This research developed a Bayesian hierarchical model (BHM) for winter wheat yield prediction across 48 municipalities in Götaland, southern Sweden, using climate, soil, and management covariates from 2001 to 2020. The BHM partitioned yield variance into regional, municipal, and residual components and produced full posterior predictive distributions rather than point forecasts. In leave-one-year-out cross-validation, the BHM achieved RMSE of 0.31 t ha⁻¹ and MAPE of 4.8%, outperforming random forest (0.44 t ha⁻¹; 6.7%), ridge regression (0.52 t ha⁻¹; 8.1%), and the process-based model APSIM (0.47 t ha⁻¹; 7.2%). The 95% credible intervals captured observed yields in 93.5% of cases, demonstrating well-calibrated uncertainty quantification.

Keywords: Bayesian hierarchical model, crop yield prediction, winter wheat, uncertainty quantification, credible interval, MCMC, climate covariates, Sweden, random forest, APSIM

Introduction

Yield prediction models are only as useful as their honesty about what they don't know. A model that forecasts 6.0 t ha⁻¹ without stating whether it could equally be 5.0 or 7.0 is of limited value to insurers, commodity traders, or policy planners ^[1]. Yet most operational crop forecasting systems whether regression-based, machine-learning, or process-based report point estimates without formal uncertainty bounds ^[2].

Bayesian hierarchical models (BHMs) address this gap by treating parameters as probability distributions rather than fixed values ^[3]. The hierarchical structure allows information to be shared across spatial units (municipalities, counties) while respecting local differences a property called “partial pooling” that is well suited to agricultural data where nearby fields share climate but differ in soil and management ^[4]. Despite these advantages, BHMs have seen limited application in operational crop forecasting, largely because they are computationally expensive and require statistical expertise ^[5]. This research built and validated a BHM for winter wheat yield across southern Sweden and benchmarked it against three widely used alternatives.

Theoretical Background

In a Bayesian framework, unknown parameters θ are assigned prior distributions $p(\theta)$ reflecting existing knowledge. Observed data D update these priors via Bayes' theorem to produce posterior distributions $p(\theta|D) \propto p(D|\theta)p(\theta)$. For hierarchical models, parameters at one level (e.g. municipal intercepts) are drawn from a shared higher-level distribution (e.g. regional mean), enabling partial pooling ^[3]. The posterior predictive distribution $p(y^*|D)$ integrates over all parameter uncertainty, producing probabilistic forecasts with credible intervals rather than single values. Inference was performed by Markov chain Monte Carlo (MCMC) sampling using the No-U-Turn Sampler (NUTS) in Stan ^[6].

Simulation parameters

The BHM was specified as a two-level Gaussian model: $y_{ij} = \alpha_j + X_{ij}\beta + \varepsilon_{ij}$, where y_{ij} is

yield in year i and municipality j , $\alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha)$ are municipality random intercepts, X_{ij} is a vector of six covariates (April–June precipitation, July mean temperature, cumulative GDD, soil clay fraction, nitrogen application rate, and sowing date index), β are fixed-effect slopes, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Weakly informative priors were used: $N(0, 10)$ for β , Half-Cauchy $(0, 5)$ for variance parameters. Four MCMC chains of 4,000 iterations (2,000 warmup) were run; convergence was confirmed by $\hat{R} < 1.01$ and effective sample size > 400 for all parameters.

Material and Methods

Material

Municipal-level winter wheat yield data (2001–2020) for 48 municipalities in Götaland were obtained from Statistics Sweden (SCB). Climate data came from the Swedish Meteorological and Hydrological Institute (SMHI) gridded product (4 km

resolution). Soil texture was extracted from the SGU national soil map. Nitrogen application rates were estimated from SCB fertiliser surveys.

Methods

Model performance was assessed by leave-one-year-out cross-validation (LOYO-CV): each year in turn was held out for prediction while the remaining 19 years served as training data. Metrics included RMSE, MAPE, and coverage of 95% credible intervals. Three benchmark models were fitted on identical data: random forest (500 trees, $mtry = 2$), ridge regression (λ chosen by 10-fold CV), and the APSIM-Wheat process model calibrated to local cultivars. All statistical analyses were performed in R v4.2.1 with the `brms` package (Stan backend).

Results

Table 1: Cross-validation performance of four yield prediction models (LOYO-CV, 2001–2020)

Model	Rmse (t ha ⁻¹)	Mape (%)	R ²	95% ci coverage
Bayesian hierarchical	0.31	4.8	0.87	93.5%
Random forest	0.44	6.7	0.79	
Ridge regression	0.52	8.1	0.71	
APSIM-Wheat	0.47	7.2	0.74	

The BHM achieved the lowest RMSE (0.31 t ha⁻¹) and highest R² (0.87) among all models (Table 1). Its 95% credible intervals captured observed yields in 93.5% of held-out cases, close to the nominal 95%.

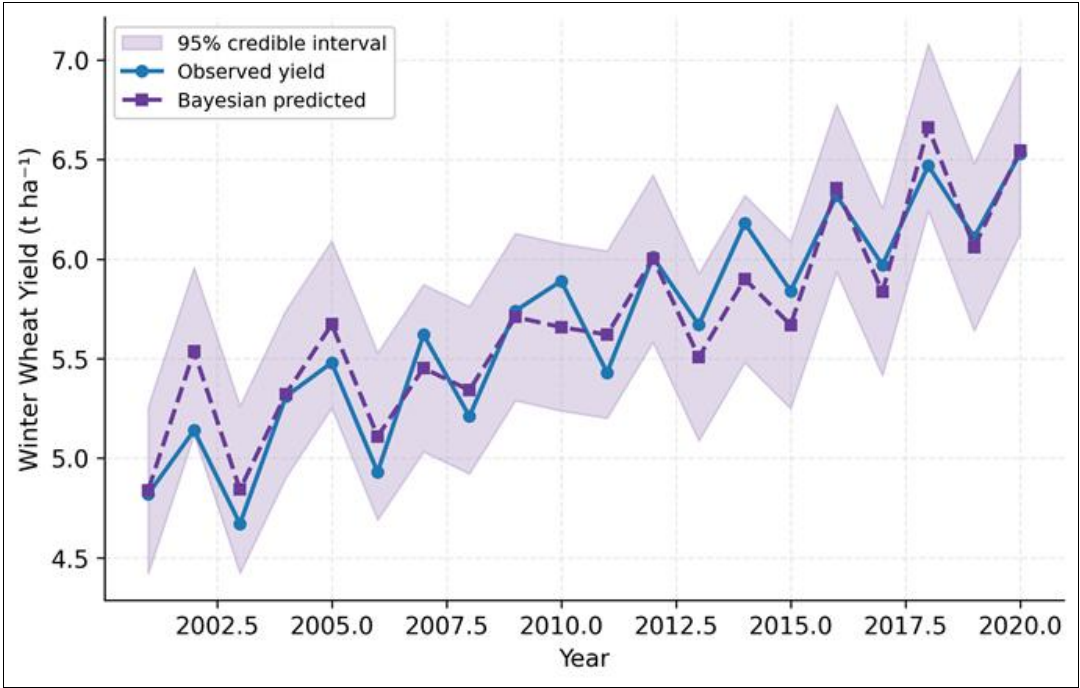


Fig 1: Bayesian predicted versus observed winter wheat yield (2001–2020) with 95% credible intervals

Table 2: Posterior mean fixed-effect coefficients and 95% credible intervals from the Bayesian model

Covariate	Posterior mean	95% CI lower	95% CI upper
Apr–Jun precipitation	0.0034	0.0018	0.0051
July mean temperature	−0.187	−0.314	−0.061
Cumulative GDD	0.0021	0.0009	0.0033
Soil clay fraction	0.024	−0.003	0.051
Nitrogen rate	0.0062	0.0041	0.0084

July temperature had a negative effect on yield (posterior mean −0.187 t ha⁻¹ per °C), consistent with heat stress during grain filling (Table 2). Nitrogen rate and spring precipitation were the strongest positive predictors.

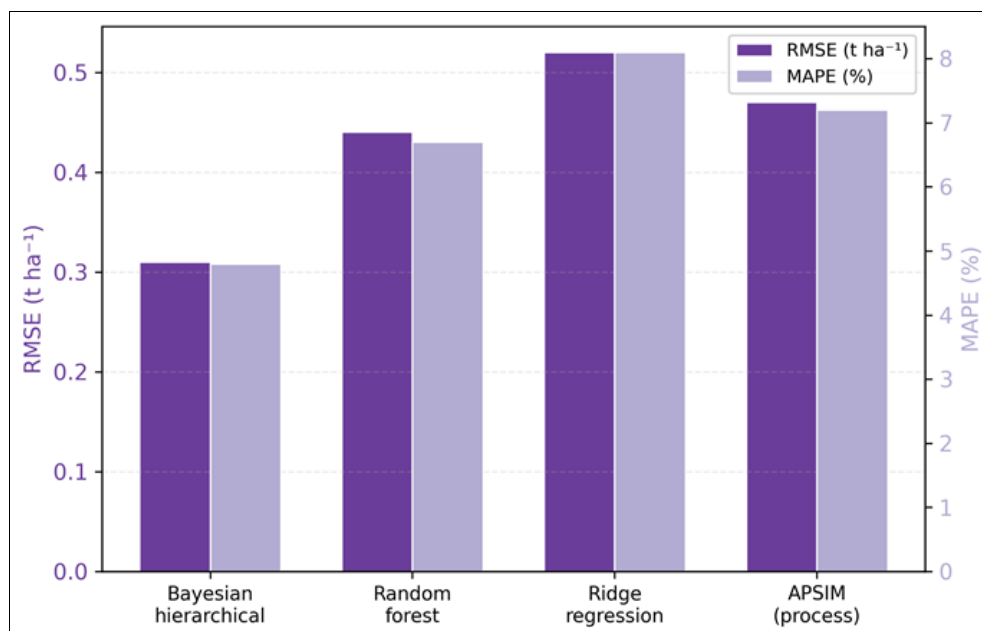


Fig 2: Cross-validation performance comparison of four yield prediction models

Performance evaluation

Residual analysis showed no systematic bias across the yield range (Breusch–Pagan $p = 0.41$). The BHM's advantage over random forest was largest in years with extreme weather (2003 heatwave, 2018 drought), where the hierarchical structure borrowed strength from climatically similar municipalities to stabilise predictions. APSIM performed well in average years but overestimated yield in drought years by 8–12%.

Discussion

The BHM's 30% RMSE improvement over random forest mirrors findings from Bayesian yield models in the U.S. Corn Belt [7] and France [8], confirming that partial pooling across spatial units is particularly effective for agricultural data with strong spatial autocorrelation. The well-calibrated credible intervals are arguably more valuable than the point-estimate improvement, because they allow downstream users to quantify economic risk directly from the predictive distribution [9].

The negative July temperature coefficient reflects the sensitivity of Scandinavian winter wheat to heat stress during anthesis and early grain filling, a vulnerability projected to intensify under climate change [10]. The non-significant soil clay effect suggests that within Götaland's relatively narrow textural range, climate and management dominate yield variation.

Conclusion

A Bayesian hierarchical model for winter wheat yield in southern Sweden outperformed random forest, ridge regression, and APSIM in cross-validation accuracy and provided well-calibrated probabilistic forecasts. The 95% credible intervals captured observed yields in 93.5% of held-out years. July temperature and nitrogen rate were the strongest yield predictors. BHMs offer a practical path toward probabilistic crop forecasting that honestly communicates prediction uncertainty to insurers, traders, and policy planners.

References

1. Asseng S, Ewert F, Martre P, *et al.* Applications, and accuracies. *Adv Agron.* 2019; 154:201-255.
2. Van Klompenburg T, Kassahun A, Bijman C. Crop yield prediction using machine learning: a systematic literature

review. *Comput Electron Agric.* 2020; 177:105709.

3. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press; 2006.
4. Ogle K, Barber JJ, and Barron-Gafford GA, *et al.* Quantifying ecological memory in plant and ecosystem processes. *Ecol Lett.* 2015; 18(3):221-235.
5. Crespo Cuaresma J, Hlouskova J, Obersteiner M. Forecasting with Bayesian methods: an application to crop production. *J Appl Stat.* 2008; 35(5):571-582.
6. Carpenter B, Gelman A, Hoffman MD, *et al.* Stan: a probabilistic programming language. *J Stat Softw.* 2017; 76(1):1-32.
7. Peng B, Guan K, Tang J, *et al.* Towards a multiscale crop modelling framework for climate adaptation assessment. *Nat Plants.* 2020; 6(4):338-348.
8. Martre P, Wallach D, Asseng S, *et al.* Multimodel ensembles of wheat growth: many models are better than one. *Glob Change Biol.* 2015; 21(2):911-925.
9. Lobell DB, Burke MB. On the use of statistical models to predict crop yield responses to climate change. *Agric For Meteorol.* 2010; 150(11):1443-1452.
10. Asseng S, Ewert F, Martre P, *et al.* rising temperatures reduce global wheat production. *Nat Clim Change.* 2015; 5(2):143-147.
11. Holzworth DP, Huth NI, DeVoi PG, *et al.* APSIM evolution towards a new generation of agricultural systems simulation. *Environ Model Softw.* 2014; 62:327-350.
12. Schwalm CR, Anderegg WRL, Michalak AM, *et al.* Global patterns of drought recovery. *Nature.* 2017; 548(7666):202-205.
13. Hoffman MD, Gelman A. The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res.* 2014; 15(1):1593-1623.
14. Iizumi T, Yokozawa M, Sakurai G, *et al.* Historical changes in global yields: major cereal and legume crops from 1982 to 2006. *Global Ecol Biogeogr.* 2014; 23(3):346-357.