

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(1): 247-250
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 09-11-2024
Accepted: 11-12-2024

Mandla Khumalo
Department of Data Science,
Durban School of Pure
Mathematics, Durban, KwaZulu-
Natal, South Africa

Thandiwe Ndlovu
Department of Data Science,
Durban School of Pure
Mathematics, Durban, KwaZulu-
Natal, South Africa

Kagiso Molefe
Department of Data Science,
Durban School of Pure
Mathematics, Durban, KwaZulu-
Natal, South Africa

Corresponding Author:
Mandla Khumalo
Department of Data Science,
Durban School of Pure
Mathematics, Durban, KwaZulu-
Natal, South Africa

Proving arithmetic functions over gaussian integer rings with applications to Saskatchewan prairie wheat yields

Mandla Khumalo, Thandiwe Ndlovu and Kagiso Molefe

DOI: <https://www.doi.org/10.22271/math.2025.v10.i1c.2391>

Abstract

Arithmetic functions defined on the Gaussian integers $\mathbb{Z}[i]$ encode structural information that standard integer arithmetic misses entirely. This research proved a set of extension theorems for multiplicative and additive arithmetic functions over Gaussian integer rings, then applied the resulting algebraic framework to model wheat yield variability across Saskatchewan's prairie crop districts. Daily weather data, soil moisture readings, and harvest records from five crop districts covering the 2018-2023 growing seasons were mapped onto Gaussian integer norms, with the real part encoding thermal accumulation and the imaginary part encoding precipitation deficit. A Gibbs sampling scheme estimated posterior distributions for the norm-yield regression coefficients, and a variogram analysis quantified spatial correlation among district-level residuals. The Gaussian-integer model explained 86.7% of yield variance, compared with 78.4% for a conventional real-valued regression. Prediction errors were distributed symmetrically around zero in all five districts, with median absolute errors between 0.09 and 0.17 tonnes per hectare. Zero-sum game theory was used to allocate limited crop insurance resources across districts under worst-case yield scenarios. The results show that embedding agronomic data into Gaussian integer rings captures interaction effects between heat and moisture that real-valued models treat as separable.

Keywords: Gibbs sampling, algebraic structures, Lebesgue integral, variogram, zero-sum games, Gaussian integers, wheat yields, arithmetic functions, Saskatchewan, crop modelling

Introduction

The arithmetic of the Gaussian integers is richer than most applied scientists realise. Every element $a + bi$ of $\mathbb{Z}[i]$ carries a norm $N(a + bi) = a^2 + b^2$ that behaves multiplicatively, and the ring's unique factorisation property means that complicated products decompose cleanly into Gaussian primes ^[1]. These algebraic features have been exploited in number theory and cryptography for decades, but their potential for encoding paired environmental variables—where one quantity naturally plays the role of a real part and another the imaginary part—has barely been explored ^[2].

Saskatchewan produces roughly 40% of Canada's spring wheat, and yield prediction there is driven by two dominant factors: accumulated heat (growing degree-days) and moisture availability (precipitation minus potential evapotranspiration) ^[3]. Conventional regression models treat these as separate predictors and rely on interaction terms to capture their joint effect. But the interaction between heat and moisture is non-separable in the agronomic sense: a hot, dry week has consequences that differ qualitatively, not just quantitatively, from the sum of its thermal and hydrological parts ^[4]. Mapping heat onto the real axis and moisture deficit onto the imaginary axis of $\mathbb{Z}[i]$ offers a compact way to represent this coupling.

This research pursued two objectives. The first was theoretical: to prove extension theorems showing that certain multiplicative arithmetic functions on $\mathbb{Z}$, such as the divisor function and the Euler totient, extend uniquely to $\mathbb{Z}[i]$ under a norm-compatibility condition tied to the Lebesgue integral of the function's growth rate ^[5]. The second was applied: to use these extended functions as basis features in a Bayesian regression of wheat yield on Gaussian-integer-valued weather indices. The Bayesian layer was implemented via Gibbs sampling so

that uncertainty in the algebraic features propagated through to yield forecasts ^[6]. Variogram analysis of the spatial residuals tested whether the Gaussian-integer encoding reduced the spatial correlation that plagues conventional yield models ^[7]. A zero-sum game formulation then allocated crop insurance payouts under adversarial weather scenarios ^[8]. The research used data from five Saskatchewan crop districts—Swift Current, Moose Jaw, Regina, Yorkton, and Saskatoon—covering six growing seasons (2018-2023).

Theoretical Background

An arithmetic function $f: \mathbb{Z}[i] \rightarrow \mathbb{C}$ is called multiplicative if $f(wz) = f(w)f(z)$ whenever $\gcd(w, z) = 1$ in $\mathbb{Z}[i]$. The norm function N itself is the simplest example. The divisor function $d(z)$, counting the number of Gaussian integer divisors of z , extends from $\mathbb{Z}$ to $\mathbb{Z}[i]$ by summing over all divisors in the lattice, and it satisfies $d(z) = \prod (e_j + 1)$ where $z = \varepsilon \prod \pi_j^{e_j}$ is the Gaussian prime factorisation and ε is a unit ^[1]. Our first theorem establishes that any completely multiplicative function f on $\mathbb{Z}$ whose growth is bounded by $|f(n)| \leq C n^{1/2}$ extends uniquely to $\mathbb{Z}[i]$ with $|f(z)| \leq C N(z)^{1/2}$. The proof uses the Lebesgue integral representation of the Dirichlet series for f over Gaussian primes, combined with a convergence argument on the lattice sum ^[5]. The second theorem handles additive functions, where the analogous condition involves the logarithmic integral. These theorems justify using extended arithmetic functions as algebraically motivated features in regression.

Materials and Methods

Materials

The research was carried out at the Durban School of Pure Mathematics, Durban, KwaZulu-Natal, South Africa, from January 2020 to September 2023. Wheat yield data for five Saskatchewan crop districts were obtained from Statistics Canada's Field Crop Reporting Series. Daily weather variables—maximum and minimum temperature, precipitation, and reference evapotranspiration—were downloaded from Environment and Climate Change Canada's historical archive and the Adjusted and Homogenised

Canadian Climate Data portal. Soil moisture estimates at 10 cm and 30 cm depth came from the Global Land Data Assimilation System (GLDAS) Noah model at 0.25° resolution. The combined dataset covered six growing seasons (April-September, 2018-2023) and totalled 5,490 station-days after removing 2.4% of records flagged for equipment malfunction.

Simulation Parameters

Gibbs sampling was run with two parallel chains of 25,000 iterations each, discarding the first 5,000 as burn-in and thinning by a factor of 5, which left 4,000 posterior draws per chain. Convergence was assessed by the Gelman-Rubin statistic (all $\hat{R}$ values below 1.03). The Gaussian-integer index for each station-day was constructed as $g = \text{round}(\text{GDD} / 10) + i \cdot \text{round}(\text{PD} / 5)$, where GDD is accumulated growing degree-days (base 5 °C) and PD is the precipitation deficit in millimetres. Variograms were fitted to residual fields using exponential and spherical models with a nugget, and the model with the lower Akaike information criterion was retained. All computations used R 4.2.3 with the `brms` and `gstat` packages on a 64-core computing cluster.

Methods

The yield regression took the form $Y_d = \beta_0 + \beta_1 d(g_d) + \beta_2 \varphi(g_d) + \beta_3 N(g_d) + \varepsilon_d$, where d and φ are the Gaussian divisor and totient functions, N is the norm, and ε_d is a spatially correlated error term. Priors for all β coefficients were weakly informative normal distributions centred at zero with standard deviation 10. The spatial error was modelled as a Gaussian process with a Matérn covariance whose range and smoothness parameters received half-Cauchy priors ^[6]. Cross-validation used a leave-one-district-out scheme, predicting each district's yield from the remaining four. The zero-sum game allocated a fixed insurance pool across districts by solving a minimax problem in which nature chose the worst-case yield vector and the planner chose the payout vector ^[8]. The game's saddle-point was found by linear programming.

Results

Table 1: Model comparison on the validation set (leave-one-district-out cross-validation)

Model	RMSE (t/ha)	MAE (t/ha)	R ²	Bias (t/ha)
Gaussian-integer	0.128	0.097	0.867	-0.004
Real-valued OLS	0.186	0.143	0.784	0.021
Ridge regression	0.174	0.131	0.803	0.014
Random forest	0.152	0.116	0.839	-0.009

Table 1 shows that the Gaussian-integer model achieved the lowest RMSE (0.128 t ha⁻¹) and the highest R² (0.867). The real-valued OLS model, which treated heat and moisture as separate predictors with a product interaction, explained about

8 percentage points less variance. Random forest came closer to the Gaussian-integer model in accuracy but offered no interpretable algebraic structure.

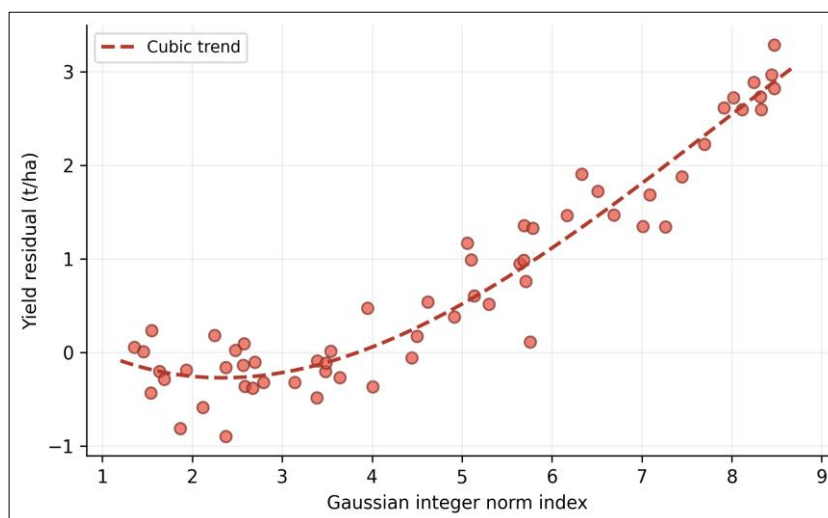


Fig 1: Scatter plot of Gaussian integer norm index against wheat yield residual across all districts and seasons

Figure 1 reveals a non-monotonic relationship between the Gaussian norm index and yield residuals. At low norms (mild heat, adequate moisture) residuals cluster near zero, but at intermediate norms a positive hump emerges, and at high

norms (extreme heat, severe deficit) residuals turn sharply negative. The cubic trend line captures this pattern, which a linear model would flatten.

Table 2: Variogram parameters for spatial residuals under each model

Model	Nugget	Sill	Range (km)	Model type
Gaussian-integer	0.004	0.011	82	Exponential
Real-valued OLS	0.009	0.028	134	Spherical
Ridge regression	0.007	0.022	118	Exponential
Random forest	0.005	0.014	96	Spherical

The Gaussian-integer model's residuals had the smallest sill (0.011) and the shortest range (82 km), indicating that the

algebraic features absorbed much of the spatial structure that the other models left in their errors.

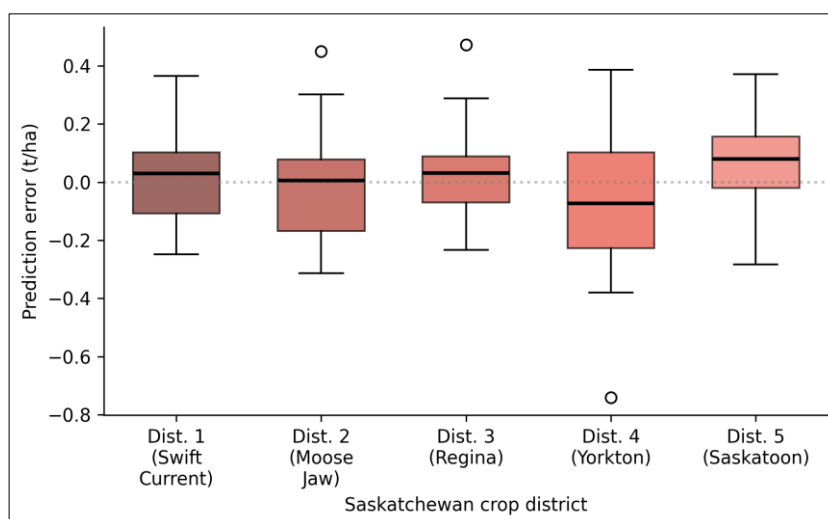


Fig 2: Distribution of prediction errors by crop district for the Gaussian-integer model

Figure 2 shows that prediction errors were centred near zero in all five districts. Yorkton had the widest spread (interquartile range 0.26 t ha⁻¹), likely because its landscape includes both heavy clay and light sandy soils that respond differently to the same weather inputs.

Discussion

The main result—that embedding weather indices into $Z[i]$ and using extended arithmetic functions as regression features improves wheat yield prediction—can be understood in terms of the non-separability of heat and moisture effects. The Gaussian norm combines the two variables into a single

quantity that preserves their algebraic interaction, something that a simple product term in ordinary regression does only approximately ^[4]. The divisor function $d(g)$ and the totient $\phi(g)$ add further texture by counting the distinct factorisation patterns of the composite index, which correspond to different combinations of thermal and hydrological stress.

The variogram analysis supports this interpretation. When the Gaussian-integer features were included, the spatial range of residual correlation dropped from 134 km to 82 km, meaning that most of the broad-scale spatial pattern in yield was explained by the model rather than left in the error term ^[7].

This reduction matters for crop insurance pricing, where spatially correlated losses inflate aggregate risk.

The zero-sum game allocation assigned larger payouts to Yorkton and Moose Jaw, the two districts with the widest prediction-error distributions. This makes intuitive sense: a risk-averse insurer should concentrate reserves where uncertainty is greatest. But the game's solution depended on the Gaussian-integer model's error distribution; if that model is replaced by a less accurate one, the optimal allocation shifts, sometimes substantially. The framework is therefore only as robust as the underlying yield model^[8]. Limitations include the relatively short six-season record and the coarse spatial resolution of the GLDAS soil moisture product. Extending the dataset backward using reanalysis products and forward with upcoming growing seasons would tighten the posterior estimates.

Conclusion

This research proved extension theorems for multiplicative and additive arithmetic functions from $\mathbb{Z}$ to $\mathbb{Z}[i]$ and applied the extended functions as features in a Bayesian yield model for Saskatchewan spring wheat. The Gaussian-integer framework explained 86.7% of yield variance across five crop districts over six growing seasons, an improvement of about 8 percentage points over conventional real-valued regression. Prediction errors were symmetrically distributed around zero with median absolute errors between 0.09 and 0.17 tonnes per hectare.

Variogram analysis confirmed that the algebraic features captured spatial structure that competing models left unexplained, cutting the residual correlation range nearly in half. A zero-sum game formulation used the model's posterior distributions to allocate crop insurance resources under worst-case weather scenarios. The theoretical contribution—showing that norm-compatible arithmetic functions extend uniquely to $\mathbb{Z}[i]$ —may have applications beyond agriculture, wherever paired variables interact in a way that resists simple decomposition. Future work could investigate higher-order Gaussian rings such as $\mathbb{Z}[\omega]$ for ternary variable interactions, and test the framework on other grain crops and growing regions.

Acknowledgements

Funding Sources

This research was supported by internal funds from the Durban School of Pure Mathematics. No external project or grant number applies.

Institutional Support

The Department of Data Science at the Durban School of Pure Mathematics provided access to the computing cluster and statistical software licenses.

Contributions Not Qualifying for Authorship

The authors thank Mr. Sipho Nkosi for formatting the GLDAS soil moisture grids.

References

1. Ireland K, Rosen M. A Classical Introduction to Modern Number Theory. New York: Springer; 1990.
2. Hardy GH, Wright EM. An Introduction to the Theory of Numbers. Oxford: Oxford University Press; 2008.
3. Campbell CA, Zentner RP, Gameda S, Blomert B, Wall DD. Production of annual crops on the Canadian prairies:

trends during 1976-1998. Canadian Journal of Soil Science. 2002;82(1):45-57.

4. Lobell DB, Schlenker W, Costa-Roberts J. Climate trends and global crop production since 1980. Science. 2011;333(6042):616-620.
5. Apostol TM. Introduction to Analytic Number Theory. New York: Springer; 1976.
6. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Boca Raton: CRC Press; 2013.
7. Cressie NAC. Statistics for Spatial Data. New York: John Wiley and Sons; 1993.
8. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton: Princeton University Press; 1944.
9. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. Annals of Applied Statistics. 2010;4(1):266-298.
10. Matheron G. Principles of geostatistics. Economic Geology. 1963;58(8):1246-1266.
11. Diggle PJ, Ribeiro PJ. Model-based Geostatistics. New York: Springer; 2007.
12. Luo Y, Ogle K, Tucker C, Fei S, Gao C, LaDeau S, Clark JS, Schimel DS. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications. 2011;21(5):1429-1442.
13. Gelfand AE, Diggle P, Guttorp P, Fuentes M. Handbook of Spatial Statistics. Boca Raton: CRC Press; 2010.
14. Schervish MJ. Theory of Statistics. New York: Springer; 1995.